

# Capacity and Energy Cost of Information in Biological and Silicon Photoreceptors

PAMELA ABSHIRE, STUDENT MEMBER, IEEE, AND ANDREAS G. ANDREOU, MEMBER, IEEE

## Invited Paper

*We outline a theoretical framework to analyze information processing in biological sensory organs and in engineered microsystems. We employ the mathematical tools of communication theory and model natural or synthetic physical structures as microscale communication networks, studying them under physical constraints at two different levels of abstraction. At the functional level, we examine the operational and task specification, while at the physical level, we examine the material specification and realization. Both levels of abstraction are characterized by Shannon's channel capacity, as determined by the channel bandwidth, the signal power, and the noise power. The link between the functional level and the physical level of abstraction is established through models for transformations on the signal, physical constraints on the system, and noise that degrades the signal.*

*As a specific example, we present a comparative study of information capacity (in bits per second) versus energy cost of information (in joules per bit) in a biological and in a silicon adaptive photoreceptor. The communication channel model for each of the two systems is a cascade of linear bandlimiting sections followed by additive noise. We model the filters and the noise from first principles whenever possible and phenomenologically otherwise. The parameters for the blowfly model are determined from biophysical data available in the literature, and the parameters of the silicon model are determined from our experimental data.*

*This comparative study is a first step toward a fundamental and quantitative understanding of the tradeoffs between system performance and associated costs such as size, reliability, and energy requirements for natural and engineered sensory microsystems.*

## I. INTRODUCTION

Biological sensory organs operate at performance levels set by fundamental physical limits [1], under severe constraints of size, weight, structural composition, and energy resources. Consider the mammalian retina, a thin structure (approximately 500  $\mu\text{m}$ ) at the back of the eyeball and an outpost of

the brain. It serves as both a sensor and a preprocessor for the visual cortex. Nobel laureate David Hubel writes [2]:

The eye has often been compared to a camera. It would be more appropriate to compare it to a TV camera attached to an automatically tracking tripod—a machine that is self-focusing, adjusts automatically for light intensity, has a self cleaning lense, and feeds into a computer with parallel processing capabilities so advanced that engineers are only just starting to consider similar strategies for the hardware they design. . .No human inventions, included computer assisted cameras can begin to rival the eye [2, p. 33].

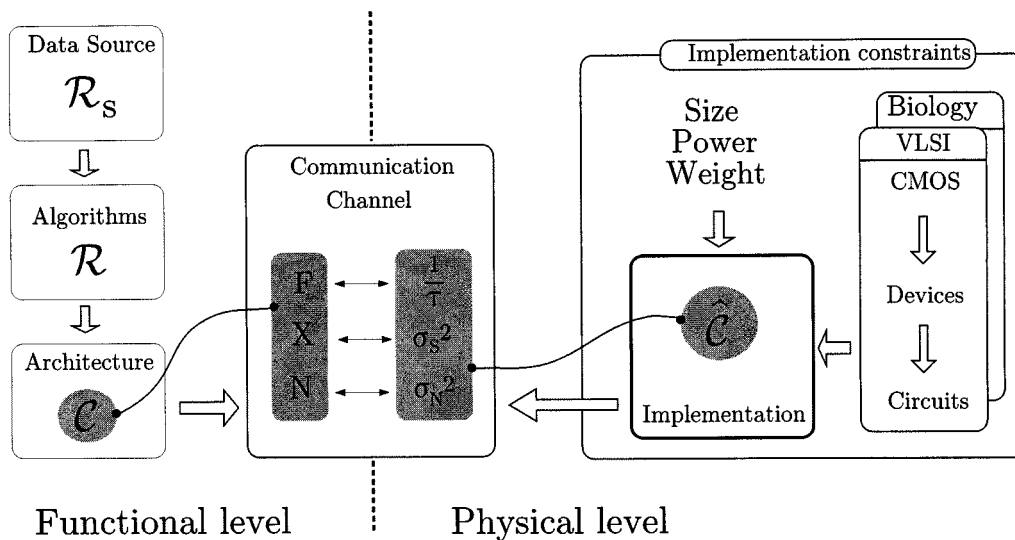
Natural sensory microsystems are indeed engineering marvels of heterogeneous integration and structural complexity across different physical scales, from molecules, to cells, to networks [3]. Higher level function in neural structures is equally remarkably effective and efficient. The total power consumed by the awake human brain is about 10 W [4], [5]. It is believed that the existence of such efficient computational structures results from an evolutionary necessity toward systems that are truly autonomous, and thus constrained to operate under strict bounds of size and weight and at temperatures favorable for the development of life as we know it ( $T \approx 300$  K). The power supply in biological systems is approximately 100 mV, only a few  $kT/q$ , while the bandwidth of the constituent components at the macroscopic scale is only a few hundred hertz, comparable and sometimes even lower than the bandwidth of the signals that have to be processed. Typical current levels in biological structures are in the pico to nano ampere range [6].

While biologists and neuroscientists proceed to elucidate the principles of information processing in biological systems, engineers have sought to abstract such knowledge in human engineered microsystems [7], [8]. Carver Mead coined the term *neuromorphic* electronic systems over a decade ago [9] to characterize mixed analog/digital (A/D) very large scale integration (VLSI) [7], [10] systems that abstract their organization from biology. Then, the silicon

Manuscript received February 1, 2001; revised May 29, 2001. This work was supported by a DARPA/ONR MURI N00014-95-1-0409 with Boston University on Automated Sensing and Vision Systems and by an NSF Graduate Fellowship.

The authors are with the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218 USA.

Publisher Item Identifier S 0018-9219(01)06742-1.



**Fig. 1.** Sensory microsystems represented as communication channels, and studied at two levels of abstraction: Functional level (left) and Physical level (right).  $\mathcal{R}_s$  is the information rate of the source,  $\mathcal{R}$  is the information rate at the output of the algorithm,  $\mathcal{C}$  is the channel capacity,  $F$  is the bandwidth of the channel,  $N$  is the noise power and  $X$  is the signal power.

retina was the vision of the future. One of the early silicon retina architectures has seen a metamorphosis and is now referred to as a CMOS active pixel sensor (APS) imager [11]. Then, the MOS logarithmic photoreceptor [12] was on the forefront of research and it was the core of many of the original silicon retinae designs. Now, the basic transducer structure with compressive nonlinearity is called a nonintegrating wide dynamic range photodetector. Today, plain imagers as well as more sophisticated mixed A/D VLSI sensory systems on a chip are becoming commodities as microprocessors were a decade ago [13].

In this paper, we present work that employs quantitative tools from communication and information theory to study biological sensory organs and silicon microsystems. We analyze the functional properties and performance of the blowfly photoreceptor and compare it to a biologically inspired silicon model. For both systems we construct a communication channel model that incorporates physical transformations between input and output, as well as degradation by noise. Such models allow us to investigate tradeoffs between cost and performance and provide a starting point for investigation of the energy efficiency of early visual information processing.

## II. SENSORY MICROSYSTEMS AS COMMUNICATION NETWORKS

In the following sections, we briefly outline our mathematical framework. We model sensory microsystems as microscale communication networks [14]. These communication networks are implemented under physical constraints including operational constraints such as available power resources, size, and weight constraints (see Fig. 1). We study sensory microsystems at two levels of abstraction, the *functional level* and the *physical level*.

The two levels of abstraction, the functional level and the physical level, relate to each other through the canonical

model of a communication channel. Both levels are characterized by the channel capacity  $\mathcal{C}$ , which is determined by the channel bandwidth  $F$ , the signal power  $X$ , and the noise power  $N$ . The link between the functional level and the physical level is established through physical models for transformations on the signal such as bandwidth limitations  $F \mapsto (1/\tau)$  as determined by time constant  $\tau$ , constraints on the system such as signal power  $X \mapsto \sigma_s^2$ , and noise corrupting the signal path  $N \mapsto \sigma_N^2$ .

At the *functional level*, we consider the task and operational specification for the microsystem, including data sources (prior distributions), algorithms, architectures, and task requirements (output distributions), as well as functional constraints, such as redundancy reduction, information maximization, and fidelity requirements (distortion, signal-to-noise ratio, ...).

**Data sources** are described probabilistically. The message to be communicated is selected from a set of possible messages; that selection process can be random or can depend on previous selections. Any data source is characterized by its entropy, or self-information, rate, which is determined by the variety of possible messages and the frequency at which they are presented.  $\mathcal{R}_s$  is the information rate of the source.

**Algorithms** are characterized by the information rate  $\mathcal{R}$  at the output, which is determined both by the data source and by the data processing requirements. The information rate of the algorithm is a quantitative specification for the computational effort that is expected from the implementation of the system.

**Architectures** can be investigated at the functional level of abstraction without considering implementation details. Alternatives can be explored that satisfy the specification of the processing problem (data source and algorithm). An architecture is a hierarchical composition of communication channels and is characterized by channel capacity  $\mathcal{C}$  at the

output. The channel coding theorem ensures that information transmission rates up to the channel capacity are possible, so for rates less than capacity  $\mathcal{R} < \mathcal{C}$ , it is possible for the architecture to implement the required algorithm [15].

Many authors over the years have employed redundancy reduction or information maximization as an organizing or optimizing principle, for both biological and engineered systems. Fidelity requirements constrain the set of acceptable alternative architectures.

At the *physical level*, we consider the physical specification and instantiation for the microsystem, including implementation technologies, devices, circuits, and environment (temperature, operating conditions, . . .), as well as implementation constraints, such as models for signal power ( $\sigma_S^2$ ), physical and phenomenological models for noise power ( $\sigma_N^2$ ), physical and phenomenological models for component bandwidth determined by the dynamics ( $1/\tau$ ), models for size and weight constraints, and models for power acquisition  $\mathcal{P}_a$  and dissipation  $\mathcal{P}_d$  and energy storage  $\mathcal{P}_s$ .

Atick and Redlich [16], Van Hateren [17], and Boahen [18] discuss the design of optimal filter cascades—point to point network architectures—for processing sensory information. These designs are examples of synthesis: starting from the functional specification of maximizing mutual information given noise characteristics and proceeding to derive optimal transfer functions for the filters. Similarly, Furth and Andreou [19] outline a synthesis procedure to derive the optimal filterbank—fan-out network architecture—for speech signals under output coding constraints. Huck *et al.* [20] and Fales *et al.* [21] recently proposed a comprehensive framework to study visual processing systems, cameras, and digital signal processing algorithms as communication channels. They show how information theory can be used in a very basic and fundamental way to codesign the hardware and software components of a vision system for optimal performance.

The work presented here complements the above work in that we seek to relate the functional properties of information processing to the statistical properties of signals from natural sources and to the physical properties of the information processing medium. We construct detailed channel models for sensory hardware at the *circuit* and *device* level. Our formulation also includes the important physical constraints of size and available power resources.

#### A. Data Sources

In sensory microsystems, the input data originate as signals transduced by sensors interacting with the environment. In many cases, this data takes on real values: for example, the acoustic pressure or light intensity as a continuous function of time at a particular location, but not necessarily, as for example the presence or absence of a light flash. We consider these sets of measurements to be **random variables**, each taking on discrete or continuous values associated with the repeated outcomes of a particular experiment.

Discrete random variables can be obtained either directly as the outcome of an experiment or as a mapping on contin-

uous value data produced by an experiment. In this case, the real value data of signals are projected through a quantizer  $\Gamma$ , to a random variable  $X$  that generates symbols from the set  $S_X = \{x_1, x_2, x_3, \dots, x_N\}$  with probability  $P(x_i) = p_i$ . The projection  $\Gamma$  defines the domain of the problem and provides a starting point for analysis.

The information content (in bits) for the random variable  $X$  is given by Shannon's entropy  $H(X)$  [15] defined as follows:

$$H(X) = - \sum_{i=1}^N p_i \log_2 p_i. \quad (1)$$

$H(X)$  is the average uncertainty in the symbols generated by the source. The most improbable symbols have the highest information content. The source that produces output symbols with equal probability achieves the maximum entropy [15].

The joint and conditional entropies of two random variables with joint distribution  $p(x, y)$  and conditional distribution  $p(x|y)$  are defined similarly to the entropy definition in (1):

$$H(X, Y) = - \sum_{x \in S_X} \sum_{y \in S_Y} p(x, y) \log_2 p(x, y) \quad (2)$$

$$H(X|Y) = - \sum_{x \in S_X} \sum_{y \in S_Y} p(x, y) \log_2 p(x|y) \quad (3)$$

where  $S_X = \{x_1, x_2, \dots, x_M\}$  and  $S_Y = \{y_1, y_2, \dots, y_N\}$  denote the support sets for random variables  $X$  and  $Y$ . The *chain rule* provides the relationship between the individual entropies  $H(X)$  and  $H(Y)$  and the joint and conditional entropies  $H(X, Y)$ ,  $H(X|Y)$ , and  $H(Y|X)$ :

$$H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y) \quad (4)$$

where the conditional entropy  $H(Y|X)$  represents the entropy of the output when the input is known. The conditional entropy  $H(X|Y)$  is often referred to as the equivocation. It is the average ambiguity in the input when the output is known. The entropic measure of information employed for the discrete source can be extended to the case of continuous distributions [15], and in that case is referred to as the *differential entropy*.

The discrete entropy and differential entropy provide the average information per symbol, or observation of the data. Sometimes we are interested in sensor data at more than one instant of time; sometimes we observe sensor data over a period of time. In this case, we describe the observed random variables  $X_i$ , indexed by time  $i$ , as a **stochastic process**  $\{X_i\}$ .

The **entropy rate** per transmission of a stochastic process  $\{X_i\}$  is defined by

$$H(\mathcal{X}) = \lim_{n \rightarrow \infty} \frac{1}{n} H(X_1, X_2, \dots, X_n). \quad (5)$$

When the stochastic process is random, a sequence of independent and identically distributed (i.i.d.) random vari-

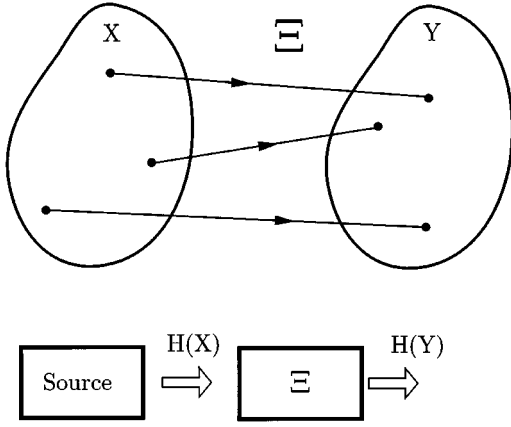


Fig. 2. An algorithm defined as a deterministic mapping  $\Xi$ .

ables, then the entropy rate is just the entropy  $H(\mathcal{X}) = H(X_1)$ . We define the **rate** of information generated by the source per unit of time as

$$\mathcal{R}_s = f_s H(\mathcal{X}) \quad (6)$$

where  $f_s$  is the frequency of transmissions and has units of frequency (Hz) and  $\mathcal{R}_s$  has units of bits per second.

### B. Algorithms

We define an **algorithm** as a deterministic mapping of an input random variable in the domain set  $S_X$  to an output random variable in the range set  $S_Y$  (see Fig. 2).  $\Xi$  could represent a linear matrix transformation  $y = Mx$ , a Boolean logic function  $y = F(x)$ , a lookup operation  $y_i = T[x_i]$ , or any other deterministic digital or analog signal processing function.

The information transmitted through the input–output (I/O) mapping is known as the *mutual information*  $I(X; Y)$ . It is the information we learn about the input by observing the output, or vice versa. It can be computed by subtracting the remaining (conditioned) uncertainty about the input  $H(X|Y)$  from the original (unconditioned) uncertainty about the input  $H(X)$ . Mutual information  $I(X; Y)$  is defined as follows:

$$\begin{aligned} I(X; Y) &\equiv \sum_{x \in S_X} \sum_{y \in S_Y} p(x, y) \log_2 \left( \frac{p(x, y)}{p(x)p(y)} \right) \quad (7) \\ &= H(X) - H(X|Y) = H(Y) - H(Y|X). \quad (8) \end{aligned}$$

The rate  $\mathcal{R}$  of information flow per unit time at the output of the algorithm is given by

$$\mathcal{R} = f_s I(X; Y) \quad (9)$$

where  $f_s$  is the temporal frequency of symbol generation at the source.

In general, the mutual information  $I(X; Y)$  is a function of both the input  $X$  and the transformation; hence, **the rate  $\mathcal{R}$  is an intrinsic property of the algorithm and the input data** and does not depend on implementation details. Whereas the differential entropy for continuous random variables depends on the coordinate system in which the variables are defined,

the mutual information does not share this difficulty with the entropy. It is the difference between two differential entropies, and dependence on coordinate system cancels out.

The data processing inequality [22] guarantees that deterministic mappings cannot generate information. If the random variables  $X$ ,  $Y$ , and  $Z$  form a Markov chain  $X \rightarrow Y \rightarrow Z$ , then the mutual information between  $Z$  and  $X$  cannot be greater than the mutual information between  $Y$  and  $X$ :  $I(X; Y) \geq I(X; Z)$ . When the equality condition holds, we classify the algorithm as an *information transfer* or *communication* process. Information transfer implies an invertible mapping, thus ensuring that the equivocation is zero ( $H(X|Y) = 0$ ), the mutual information is equal to the input entropy ( $I(X; Y) = H(X)$ ), and the input entropy is preserved ( $I(X; Y) = H(Y)$ ).<sup>1</sup>

Algorithms where the inequality is strict  $I(X; Y) > I(X; Z)$  will be referred to as *information transform* processes. The latter case is consistent with the intuitive notion of data processing to reduce the entropy and, hence, is classified as a *computation* process.

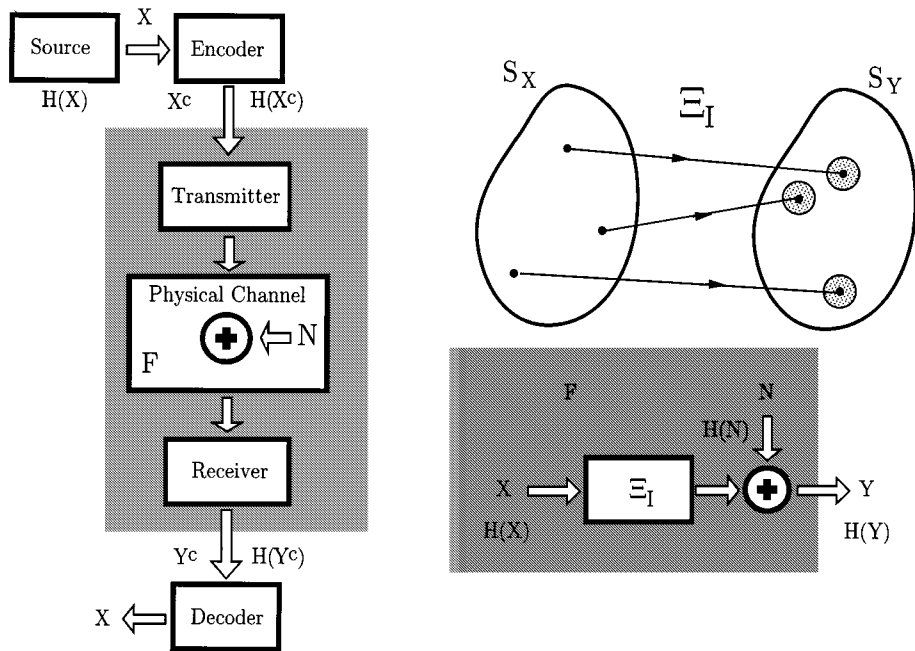
The distinction between transform and transfer algorithms is somewhat superficial, as the transfer processes are realizations in a subspace of the transform space where the mapping between the input and the output is invertible. Furthermore, the inevitable presence of noise in most physical systems makes it impossible to maintain a strictly invertible mapping. In practice, this undermines any distinction between a transfer process and a transform process. However, we find that the distinction between the two kinds of processes is not only intuitively appealing but provides a convenient assessment of the task requirements and an indication of appropriate performance metrics for analysis and synthesis.

### C. Communication Channel

A communication channel is a mathematical abstraction that separates the elementary act of communication into its constituent components for easier study. Macroscale physical communication systems such as radio links, copper wires, and optical links as well as magnetic and optical storage systems can be modeled using the general block diagram of a communication channel, shown on the left in Fig. 3.

The message to be communicated begins at the source that generates random variable  $X$  with entropy rate  $H(\mathcal{X})$ . The encoder has a dual role of matching the transmission to the source and to the channel, sometimes described separately as source coding and channel coding. First, in the role of source encoder, it removes redundancy in the input data stream, lowering the symbol frequency while maintaining the information rate, thus improving the efficiency of information transmission. Then in the role of channel coding, it further encodes the data to provide immunity from noise introduced in the physical channel. In this context the encoder could add redundancy to improve the reliability of data transmission through the physical channel. The encoded data  $X^c$  are converted into a suitable physical representation for transmission

<sup>1</sup>Of course, changes of variable still affect the differential entropy, and for continuous random variables one must specify the coordinate system in which to measure the information.



**Fig. 3.** Block diagram of a communication system abstracted into its basic components (left). The canonical model of a communication channel with signal transformation  $\Xi$  and additive noise  $N$  (right).

through the physical channel by the transmitter. The process of transmission is reversed at the receiver, whose input is a noisy and distorted version of the transmitted data. The receiver reconstructs the input to the channel, producing an estimate of the encoded data  $Y^c = \hat{X}^c$ . The decoder inverts the mapping of the encoder to produce an estimate of the original input; ideally  $\hat{X} = X$ .

For communication at the physical level, the encoding and decoding functions are often simple, sometimes so simple that we omit the *encoding* and *decoding* blocks. Many physical communication systems can be adequately described using the additive noise model shown on the right in Fig. 3. The mapping from input to output is the deterministic function  $\Xi$ . Uncertainty in the system is modeled as an additive noise contribution  $Z$  at the output of the channel, such that  $Y = \Xi[X] + Z$ , where  $X$  and  $Y$  are the input and output signals respectively. The information rate through the system is limited by the channel bandwidth  $F$  and the noise power  $N$  corrupting the signal. When the mapping  $\Xi$  is the identity transformation, Fig. 3 represents the canonical additive noise channel.

One of the most important contributions of information theory is the channel coding theorem. It's not obvious in advance that error-free transmission is possible over noisy channels, but this is precisely what we gain from the channel coding theorem. The maximum information rate  $\mathcal{R}$  of error-free transmission through a communication channel is limited by the channel capacity  $\mathcal{C}$  [15]. Conversely, any information rate  $\mathcal{R}$  is achievable with probability of error approaching zero as long as

$$\mathcal{R} < \mathcal{C}. \quad (10)$$

Although this is a remarkable result, the proof provides no guidance for finding the best coding strategy. This is a matter of trial and error and cleverness. Practical codes exist today that enable communication near the Shannon limit, albeit with heavy computational demands on the decoder.

The channel capacity is determined by maximizing the mutual information  $I(X; Y)$  between the input and the output

$$\mathcal{C} \equiv \max_{p(x)} I(X; Y) = \max_{p(x)} [H(Y) - H(Y|X)]. \quad (11)$$

For the additive noise channel  $Y = X + Z$ , this becomes

$$\mathcal{C} = \max_{p(x)} H(Y) - H(Z). \quad (12)$$

A **Gaussian channel** is an additive noise channel where the noise is a random Gaussian process. For a Gaussian channel, the transmission bandwidth and the ratio of the signal power to noise power are sufficient to determine the channel's capacity to transmit information. The noise entropy per second is  $H(Z) = F \log_2 \pi c N$ , where  $F$  is the bandwidth of transmission over the channel and  $N = \sigma_Z^2$  is the noise power. With no constraints on the input, we are free to choose input levels spaced arbitrarily far apart, and there is little trouble in distinguishing the outputs. The capacity is unbounded. However, unconstrained inputs are physically implausible; the earth is finite, and certainly our microsystems are. We'll resolve this difficulty by imposing a power limitation  $P$  on the input signal, so that  $1/n \sum_{i=1}^n x_i^2 \leq P$ . The Gaussian distribution maximizes the entropy for a fixed variance, so the output entropy per second is bounded by the entropy of a Gaussian distribution

with the same output power  $H(Y) \leq F \log_2 \pi e(P + N)$ . This leads to the channel capacity of a Gaussian channel given by

$$C = F \log_2 \left( 1 + \frac{P}{N} \right) \quad (13)$$

where  $P$  and  $N$  are the average signal and noise power, respectively, and the capacity  $C$  is given in units of bits per second. This result applies to channels with an average signal power constraint  $P$  and additive white Gaussian noise of power  $N$ .

More generally, when the noise is colored, the capacity is given by

$$C = \max_{X(f): \sigma_X^2 \leq P} \int_0^\infty \log_2 \left( 1 + \frac{X(f)}{N(f)} \right) df \quad (14)$$

where  $P$  is the input power constraint,  $X(f)$  is the power spectral density of the signal (given by the Fourier transform of the autocorrelation), and  $N(f)$  is the power spectral density of the noise. The signal that maximizes the capacity can be found using the water-filling analogy; typically the input-referred noise power spectral density is somewhat cup-shaped, and then finding the total output power spectral density simply corresponds to putting a fixed amount of water in the cup. The basic idea is that signal energy is concentrated at frequencies where noise is low [22], [23].

The capacity of a channel depends only on the channel properties and signal constraints, such as bandwidth, noise, and constraints on the signal values; it does not depend on the specific details of any particular task for which the channel may be used. Although it is straightforward to define task-dependent measures of performance, it is appealing to study the *maximum* information rate, or channel capacity, especially for peripheral sensory systems communicating data that are used for many different tasks.

#### D. Network Architecture

Network architecture analysis can proceed at the functional level of abstraction without reference to implementation. At the functional level, an architecture  $\mathcal{A}^{IP}$  is a network of channels organized in a hierarchy such that  $\mathcal{A}^{IP} \equiv \{\Xi_1, \Xi_2, \Xi_3 \dots \Xi_N\}$  where  $N$  is the total number of the constituent channels.

The calculation of channel capacity for architectures with arbitrary topologies is an open problem with theoretical results available only for special cases and specific architecture topologies [22]. We now proceed with the analysis in the frequency domain for Gaussian channels. We consider filtering transformations of the functional form  $\xi(f)$ , with power spectral density of the additive Gaussian noise  $N(f)$  and signal power spectral density  $X(f)$ . The filtering transformations  $\xi(f)$  are assumed to be noiseless. Idealized filtering operations are simply coordinate transformations. They are invertible one-to-one mappings and mutual information is preserved [15].

The network architecture consisting of a series of point-to-point channels is mathematically tractable and provides a

sufficiently rich framework for a wide range of applications, including the photoreceptor models considered in this paper.

The canonical model for a cascade network architecture is mathematically described by (15) and (16) below. The signal power  $X_n(f)$  at any stage  $n$  is transformed through a sequence of linear filters  $\xi_i(f)$ . The noise power  $N_n(f)$  is the summed power of  $m$  independent, additive Gaussian noise sources  $N_j(f)$  that are also transformed by sequences of linear filters. Explicitly, the signal and noise at stage  $n$  are given by

$$X_n(f) = \prod_{i=1}^n |\xi_i(f)|^2 X_{in}(f) \quad (15)$$

$$N_n(f) = \sum_{j=1}^m \prod_{i=k_j}^n |\xi_i(f)|^2 N_j(f) \quad (16)$$

where  $X_{in}(f)$  is the power spectrum of the input signal, the noise  $N_j(f)$  from independent source  $j$  enters at stage  $k_j$ , and the output signal is the last signal in the cascade  $Y(f) \equiv X_N(f)$ .

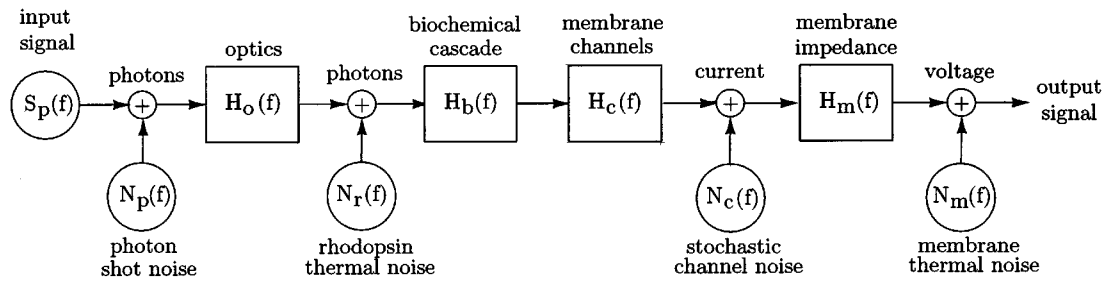
In the next section, we will employ cascade network architecture models for the blowfly and silicon photoreceptors to determine their channel capacity and, thus, relate their functional performance to the implementation details.

### III. PHOTORECEPTOR INFORMATION CAPACITY

#### A. Blowfly Photoreceptor

The visual system of the fly has been extensively studied by physiologists. Vision in the blowfly *Calliphora* begins with two compound eyes that are each composed of a hexagonal array of ommatidia. Each ommatidium contains eight photoreceptors that receive light through a facet lens and respond in graded fashion to the incident light. Electrical signals from the photoreceptor cells project to cells in the lamina and the medulla. In this investigation, we focus on the photoreceptors R1–6, which project to large monopolar cells in the lamina. The fly receives behaviorally relevant information as light reflected or emitted from objects in the environment. Photons are guided through the optics of the compound eye to the photoreceptors. Absorption of photons activates photosensitive pigments in the photoreceptor cells. The activated pigments trigger a cascade of biochemical reactions that produce “messenger” molecules. These messengers cause ion channels in the photoreceptor membrane to open. The open channels provide a membrane conductance, which allows an ionic current to flow that changes the membrane voltage. This voltage change propagates down a short axon to the synaptic terminal in the lamina. In the discussion that follows, we investigate the signals transduced through photoreceptors that project onto a single large monopolar cell, ignoring any spatial or spectral aspects of information flow in the system.

Information processing in the early visual system of the fly involves transformations between different physical degrees of freedom: photons, conformational state of proteins, concentrations of various chemical messengers, current, voltage.



**Fig. 4.** Communication network architecture for the blowfly photoreceptor. Signal transformations are modeled as filters linearized about an operating point, and noise sources are assumed to be independent and additive.

The goal of the above processes is to communicate information from one physical structure to another while preserving the message. We model these transformations as a cascade of communication channels that have bandwidth limitations.

Each of these transformations is associated with changes in the signal itself and with the inevitable introduction of noise. This begins even before transduction, as the arrival times of the photons are randomly distributed. Other sources of noise include the thermal activation of rhodopsin, the stochastic nature of channel transitions, and Johnson noise resulting from membrane impedance. We model each noise source as an independent, additive contribution to the channel.

The structure of the communication network architecture for the blowfly photoreceptor is shown in Fig. 4. The architecture is a cascade of transfer functions, with each transfer function representing a distinct physical process that together comprise phototransduction as described above. Each transfer function is linearized about an operating point, which is determined by the mean intensity of the incident light. Each noise source contributes independent, additive noise at the location in the system depicted. We determine the amplitude and power spectrum for each noise source, including photon shot noise, rhodopsin thermal noise, stochastic channel noise, and membrane thermal noise. The transfer functions and noise sources are modeled from first principles when possible and phenomenologically otherwise. While the cells under study exhibit nonlinearity at very low light levels or for large signals, they have been studied extensively as linear systems, and their linear properties are well documented in the literature [24]. Modeling the transfer functions as linear systems will be accurate when the variance of the signal is sufficiently small that the operating point remains fixed. This requirement is approximately satisfied for white noise stimulation protocols as in [25].

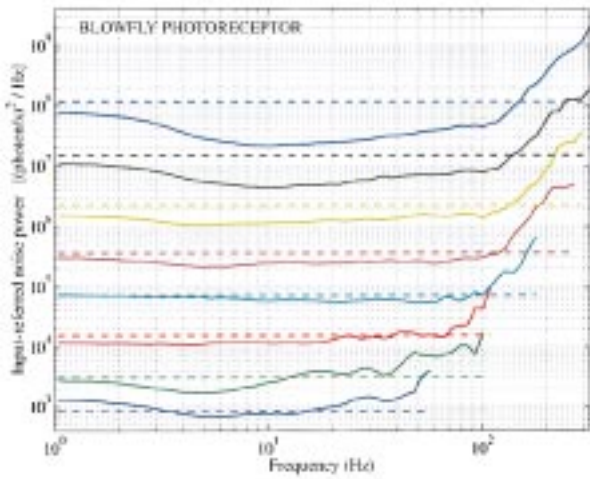
With a communication channel model and its relation to the structure established, we proceed to describe the components of the model. The spectral density of the input photon signal depends on the environment. Shot noise of the input photon stream increases with the mean intensity. A pupil mechanism within the photoreceptor cell attenuates the photon stream; this adaptive mechanism extends the dynamic range of the cell. Absorption of photons causes a conformational change in the photosensitive pigment, rhodopsin, which triggers a cascade of biochemical reac-

**Table 1**  
Summary of Model Equations

| Signal                   | $S_p(f)$ determined by environment  |
|--------------------------|---|
| Photon shot noise        | $N_p(f) = 2I$   |
| Optics                   | $H_o^2(f) = C_o(I)$   |
| Rhodopsin thermal noise  | $N_r(f) = 2 \cdot 10^{-3}$  |
| Biochemical cascade      | $H_b^2(f) = \frac{k_b^2}{1 + (2\pi f t_b)^2 n_b + 1}$                                       |
| Membrane current         | $H_c^2(f) = (V_m - E_{ch})^2$   |
| Stochastic channel noise | $N_c(f) = \frac{4N\tau_c^2(V_m - E_{ch})^2 n_{oc}(1 - n_{oc})\tau_c}{1 + (2\pi\tau_c f)^2}$ |
| Transfer impedance       | $ H_m(f) ^2 =  V_{out}/I_{in} ^2$   |
| Johnson noise            | $N_m(f) = 4kT \text{Re}\{Z_{out}(f)\}$  |

tions. The photosensitive pigment can also be activated spontaneously by thermal fluctuations, and this spontaneous isomerization contributes shot noise to the signal (rhodopsin thermal noise). Many details of the biochemical cascade are unknown and remain an active area of investigation [26], so the biochemical cascade is modeled phenomenologically according to the adapting bump model [27]–[29]. The biochemical cascade is considered to be a “bump” of membrane conductance in response to each activated rhodopsin molecule, with the gamma function providing a reasonable match to the empirically measured transfer function. This membrane conductance allows an ionic current to flow. Membrane conductance is determined by the average number of open ion channels, and fluctuations in the number of open channels produce noise in the membrane current. The membrane current generates a voltage across the cell membrane. The photoreceptor cell is an extended structure, so transfer of the signal within the cell is modeled using a cable model with three compartments as in [30]. The membrane model consists of a capacitance in parallel with light-gated channels, weakly active (linearized) potassium channels, and leakage channels, each in series with its own reversal potential (battery). The real part of the membrane impedance contributes thermal noise to the membrane voltage.

For the sake of brevity, more detailed descriptions are omitted and the main equations are summarized in Table 1. More details about the model components and parameters can be found elsewhere [31]–[34]. The parameters of the



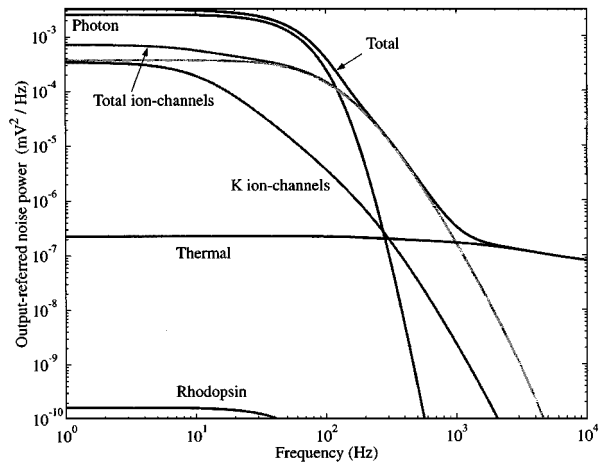
**Fig. 5.** Input-referred noise power for the blowfly photoreceptor as a function of frequency, from data in [35]. The different traces represent different background light levels, from top to bottom corresponding to [500 000 160 000 50 000 16 000 5 000 1 600 500 160] effective photons/s and covering a range from overcast daylight to streetlight. The dashed lines represent the total channel power (signal plus noise) after waterfilling for each operating point.

optical attenuation, the biochemical transfer function, the membrane channels, and membrane impedance have been estimated from physiological data [35], [36]; the procedure and parameters are summarized elsewhere [31], [34]. All parameters of the model are constrained using reported empirical data on blowfly physiology.

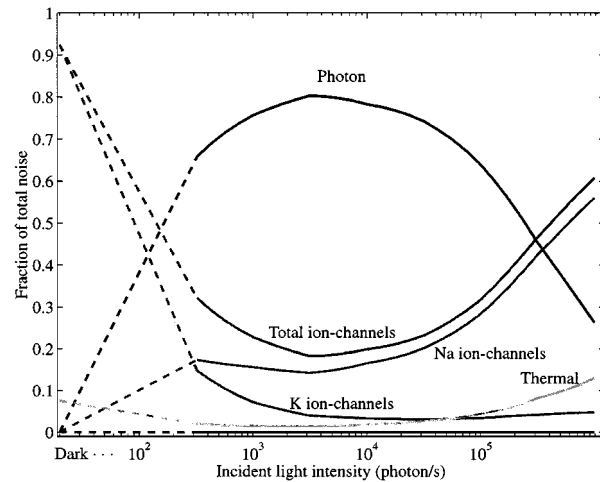
The input-referred noise power spectral density for the blowfly photoreceptor as a function of frequency is shown in Fig. 5. The input noise was derived from the experimentally measured transfer functions and noise in [35]. The different traces represent different light levels, from top to bottom corresponding to [500 000 160 000 50 000 16 000 5 000 1 600 500 160] effective photons/s,<sup>2</sup> a range from overcast daylight to streetlight. The dashed lines result from the computation of capacity; they represent the total channel power (signal plus noise) after waterfilling for each operating point.

From the above data and our model, we can determine the dominant noise sources that limit the rates of information transmission. Fig. 6 shows the output-referred noise, i.e., voltage noise at the photoreceptor axon, for an intensity of 16 000 effective photons/s. Over the frequency range of physiological interest the dominant noise sources are photon shot noise and stochastic channel noise, with thermal noise only relevant at high frequencies and rhodopsin noise irrelevant at this light level. Fig. 7 shows the fraction of total noise power at the output as a function of the incident light intensity for each of the noise sources in the model. This figure summarizes results similar to Fig. 6 over a range of light intensities, with the case of no light indicated by the dashed lines and “dark” condition on the left. Photon shot noise and stochastic channel noise are clearly the dominant noise sources in the system.

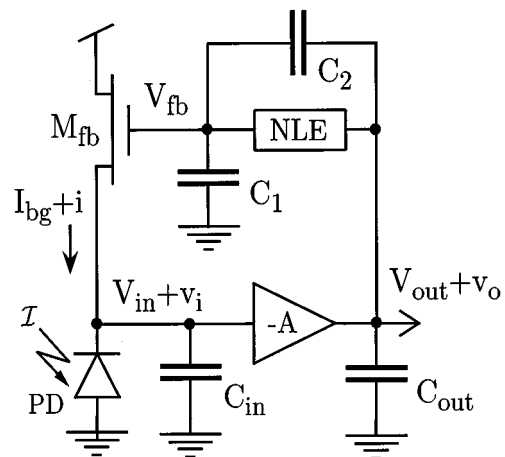
<sup>2</sup>Effective photons/s are a unit of intensity that is experimentally calibrated at low light levels by counting discrete bumps in the membrane voltage which result from single photon absorptions.



**Fig. 6.** Individual contributions of independent noise sources to output noise spectral density as a function of frequency, at an incident intensity of 16 000 effective photons/s.



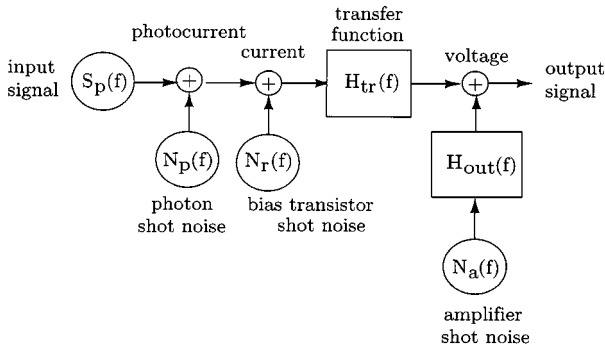
**Fig. 7.** Fraction of total noise power contributed by each noise source as a function of incident light intensity, with the case of no light indicated by the dashed lines and “dark” condition on the left.



**Fig. 8.** Adaptive photoreceptor circuit of Delbrück and Mead [12].

### B. Silicon Model

We now consider the Delbrück silicon photoreceptor [12], shown in Fig. 8. This circuit provides an analog output that



**Fig. 9.** Communication network architecture for the silicon photoreceptor. Signal transformations are modeled as filters linearized about an operating point, and noise sources are independent and additive.

has low gain for static signals and high gain for transient signals about an operating point. This adaptational strategy allows the output to represent a large dynamic range while retaining sensitivity to small inputs. We study the linearized behavior about an operating point, as we did for the blowfly photoreceptor.

This circuit has been considered in detail in [12]; the results will be given briefly here.

The structure of the communication network architecture for the silicon model of the blowfly photoreceptor is shown in Fig. 9. As for the blowfly photoreceptor, it is modeled as a cascade of transfer functions, with each transfer function linearized about an operating point.

The transfer function  $H_{tr}(s)$  of the feedback circuit, from input photocurrent  $i$  to output voltage  $v_o$ , is given by

$$H_{tr}(s) = \frac{\frac{A}{G}}{(\tau_i s + 1)(\tau_o s + 1) + \frac{A g_{mn}}{G} \frac{g + s C_2}{g + s(C_1 + C_2)}} \quad (17)$$

where  $A$  is the gain of the output amplifier,  $G$  is the conductance at the input node,  $\tau_i$  is the time constant at the input node,  $\tau_o$  is the time constant of the output amplifier,  $g_{mn}$  is the transconductance of transistor  $M_{fb}$ , and  $g$  is the conductance of the adaptive element NLE. The transfer function from current to voltage  $v_o$  at the output node is given by

$$H_{out}(s) = \frac{B}{\tau_o s + 1} \quad (18)$$

where  $1/B$  is the conductance at the output node. The parameters of these transfer functions will depend on the bias conditions as well as device and circuit parameters.

In this work, we use a minimal model for the noise of a MOS transistor that includes just the shot noise. This component of the noise will be independent of device parameters or specifics of the fabrication process. Each MOS transistor and diode will contribute current shot noise  $N(f) = 2qI$ , where  $q$  is the elementary charge and  $I$  is the current through the transistor. At the input node, there will be photon shot noise and current shot noise contributed by the bias transistor, for a total of  $4qI_{bg}$ , where  $I_{bg}$  is the photocurrent determined by the background illumination. At the output node, noise will be contributed by all three transistors of the feedback amplifier, for a total of  $6qI_b$ , where  $I_b$  is the bias current

for the amplifier. The noise contributed by the adaptive element is neglected, since the current through the element is small. The total input-referred noise from these sources will be

$$N_{tot}(f) = 4qI_{bg} + \left| \frac{H_{out}}{H_{tr}} \right|^2 6qI_b. \quad (19)$$

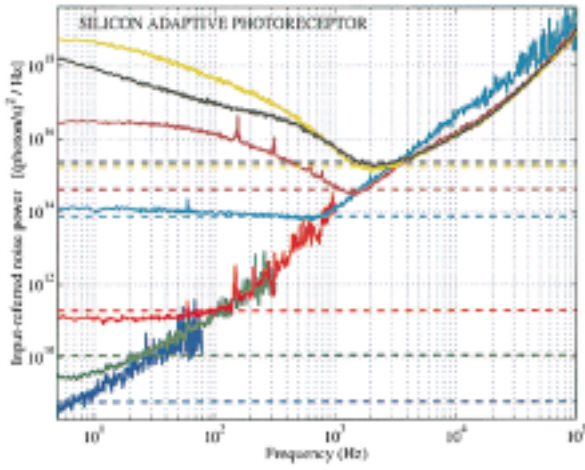
When the silicon photoreceptor is biased at high current densities, the noise sources at the input node dominate; otherwise current shot noise of the transistors in the feedback amplifier plays a significant role.

We experimentally determined the transfer characteristics and noise power spectral density of the adaptive silicon photoreceptor, and the input-referred noise characteristics are shown in Fig. 10. The input light was provided by a light-emitting diode (LED) held in position with a custom-machined fixture over the packaged chip. To measure transfer characteristics from input light to output voltage, the LED was biased with a constant dc voltage plus a small ac signal. A spectrum analyzer computed the transfer function as the ratio of the power spectra of the output signal and the input signal. To measure noise characteristics, the LED was biased with a constant dc voltage only. The noise power spectral density is the spectrum of the output signal for a constant input. Different levels of background light were obtained using neutral density filters (Kodak Wratten, Eastman Kodak Co., Rochester, NY) placed in between the LED fixture and the chip.

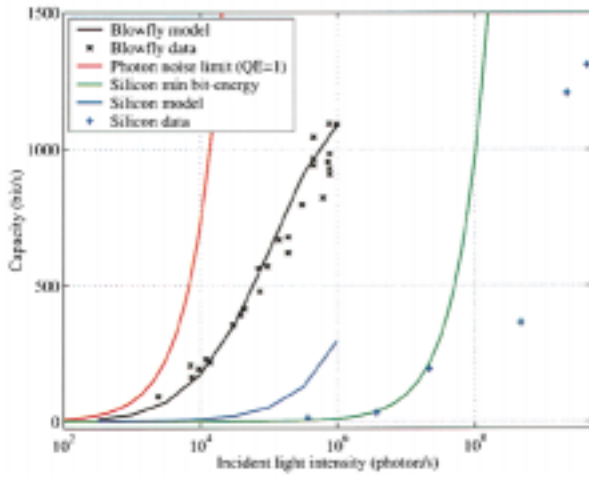
### C. Silicon versus Biology

From the input-referred noise power densities, we compute the information capacity for both the blowfly and silicon photoreceptors as a function of the incident light intensity. In order for the capacity to be well defined, we must specify how to constrain the signal power. We assume a fixed contrast power of 0.1 for all light levels, where contrast is defined as the normalized intensity. This value of contrast power was used in [35] and [24] to measure the empirical transfer functions and noise spectra and in [25] to estimate channel capacity. A contrast power of 0.1 is representative of natural scenes [37] and falls within the linear response range of the cells.

Fig. 11 shows a quantitative comparison of the capacity of both the blowfly and silicon photoreceptors as a function of incident light intensity. The capacity for our model of the blowfly photoreceptor is shown as the solid black line, with empirical estimates from [25] shown as black  $\times$ 's. The capacity for our model of the silicon photoreceptor is shown in Fig. 11 along with empirical estimates from our experimental data, shown as blue  $+$ 's. The three solid curves are for different bias conditions: the red line shows the maximum capacity [the photon noise limit for a detector with quantum efficiency (QE) of 100%], the blue line shows the capacity when the silicon photoreceptor uses the same power as the blowfly photoreceptor, and the green line shows the capacity when the silicon photoreceptor is biased for minimum bit energy, as discussed in the next section. Interestingly, the capacity of the blowfly photoreceptor is higher for lower light



**Fig. 10.** Input-referred noise power for the blowfly photoreceptor as a function of frequency, derived from our experimental data. The different traces represent different background light levels, corresponding to the +’s in Fig. 11, with intensity increasing from bottom to top. The dashed lines represent the total channel power (signal plus noise) after waterfilling for each operating point.



**Fig. 11.** Information capacity as a function of incident intensity for blowfly photoreceptor computed from our model (black) and estimated from experimental data (×’s) [25], and for the silicon photoreceptor under two bias conditions: using the same power as the blowfly (blue), and when biased for minimum bit-energy (green). The experimentally derived information capacity of the silicon photoreceptor is shown in +’s. Unmodeled noise sources account for the discrepancy between predicted and measured capacity. The photon noise limit is plotted in red.

levels but there is a crossover point at high light levels beyond which the silicon photoreceptor can achieve higher capacity.

Our models also allow us to determine the dominant noise sources that limit the rates of information transmission for both silicon and biology. Over the frequency range of physiological interest, from dc to a few hundred hertz, the dominant noise sources for the blowfly photoreceptor are photon shot noise and stochastic channel noise. At this time, the only sources modeled for the silicon photoreceptor are photon shot noise and current shot noise, and that is why there is a discrepancy between the model and the experimentally derived capacity. Flicker noise in the MOS transistors, shot

noise from leakage currents, and instrumentation noise, not included in the theoretical model of the silicon photoreceptor, are likely to be the dominant sources of noise which limit the performance of the silicon photoreceptor.

#### IV. ENERGY COST OF INFORMATION

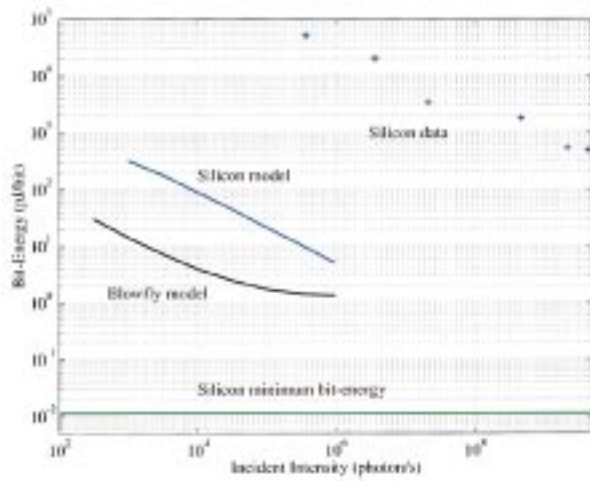
Having developed a measure for performance (i.e., information capacity), we must specify cost in order to quantify the efficiency and understand the tradeoff between performance and cost. We take the cost to be the power dissipated in transducing signals. In general, biological and VLSI systems are dissipative physical structures; signals are communicated by the flow of ions or other chemical substances, and some driving force must power this flow. Therefore, communication and computation require the dissipation of energy.

We calculate the power dissipation for the blowfly photoreceptor as the free energy lost in transducing the signals. At this time, we consider only the dissipation due to current flow across the membrane; we do not model dissipation due to the biochemical cascade, synaptic transmission, or support processes such as protein synthesis. Our resulting model for power dissipation is the membrane current times the potential difference between the membrane voltage and the reversal potential, integrated over the surface area of the cell. We approximate this by considering the photoreceptor to be isopotential. There is current flow across the membrane even in darkness, which results in power consumption without signal transduction. This membrane current increases with background light intensity, but not much because of the adaptation of the biochemical cascade. The cost of information processing in the blowfly retina has been reported to be as high as  $10^7$  ATP per bit [38], and our work predicts similar costs. We calculate the power dissipation of the adaptive silicon photoreceptor as the bias and signal current that must be sourced from the power supply, multiplied by the power supply voltage.

Our measure of performance, the capacity, and our measure of cost, the power, allow us to define a measure for efficiency, the bit-energy, which is simply the ratio between the power dissipated and the information capacity. The bit-energy gives the minimum power required to transmit a single bit of information through the system

$$BE = \frac{P}{C}. \quad (20)$$

The bit energy for the blowfly photoreceptor and the adaptive silicon photoreceptor are shown in Fig. 12. The bit-energy for the blowfly photoreceptor varies from  $\approx 20$  pJ/bit for low intensities to  $\approx 2$  pJ/bit for high intensities. As discussed above, the bit energy for the silicon photoreceptor depends on the bias of the feedback amplifier. The green curve is the minimum bit energy, while the blue line corresponds to the bit energy when the silicon photoreceptor is biased to consume the same power as the blowfly photoreceptor. The bit energy at maximum capacity occurs for an arbitrarily large bias current and an arbitrarily large bit-energy; it is off the scale of the ordinate in Fig. 12.



**Fig. 12.** Bit-energy as a function of incident background light intensity for the blowfly photoreceptor (black) and for the silicon photoreceptor at minimum bit energy (green) and when biased to use the same power as the blowfly (blue). The experimentally determined bit-energy of the silicon photoreceptor is shown in  $\times$ 's.

The capacity of the blowfly photoreceptor is higher for lower light levels but there is a crossover point beyond which the silicon photoreceptor has higher capacity. The bit energy of the silicon photoreceptor depends on the bias current: it can be made arbitrarily high by increasing the bias current, but when it is restricted to use as much power as the blowfly photoreceptor, the silicon photoreceptor provides a lower capacity (and a higher bit energy). It is interesting to note that the silicon photoreceptor can obtain a lower bit energy than the blowfly photoreceptor. The bias current that provides minimum bit energy is approximately half the photocurrent. At this bias condition, the capacity is very low in comparison to the blowfly photoreceptor, as shown in Fig. 11.

Traditional engineering philosophy might suggest biasing the feedback amplifier of the silicon photoreceptor with as high a current as possible, to maximize the capacity and provide the highest gain. Our work suggests that this is not the most efficient approach when power consumption is a design constraint. While bit-energy could provide a standard for comparing the efficiency of communication among different technologies [39], this is not necessarily the only relevant performance criterion. Minimum latency or minimum information rate in the system are other constraints that may need to be satisfied during operation.

## V. DISCUSSION

### A. Information Theory and Biology

The effectiveness and efficiency of biological systems stems partly from exploiting *prior* knowledge about the problems that they encounter [40]. Such knowledge, in the form of *internal models*, reflects the statistical properties of the natural environments in which they function. The exploitation of such prior knowledge plays an important role in both evolutionary development of neural structures and adaptation mechanisms during system function.

In the 50 years since Shannon's first probabilistic formulation of information theory [15], its concepts and tools have been employed many times to quantify information processing in neural systems from a functional perspective, at different levels of description from neurons, to networks, and behavior.

Attneave [41] and Barlow [42] introduced the idea that representation in neural systems is guided by coding principles from information theory. Since then, information theoretic principles have motivated both theoretical and experimental studies of neural coding. The early qualitative observations about neural representation have been expanded and developed more rigorously by many researchers, including Atick [43], Linsker [44], and Laughlin [45]. Other studies employ abstract mathematical models to investigate coding strategies for single spiking neurons (Mackay and McCulloch, [46]; Stein, [47]; Levy and Baxter, [48]). A rich literature has emerged using measurements from neural systems to quantify information rate under specific experimental conditions (Eckhorn and Pöpel, [49]; Theunissen and Miller, [50]; Rieke *et al.*, [51]; Buračas *et al.*, [52]).

The above cited theoretical work, employing communication and information theoretic analysis to study neural systems, has contributed significantly to understanding coding strategies, neural computation, and the reliability of neural signals. However, most previous work, with the notable exception of Manwani and Koch [53], has considered I/O relationships only at the functional level for "black box" neural systems. While this functional level of description illustrates principles of coding, computation, and reliability, no previous work has proceeded to describe neural systems at an implementation level. The components of neural systems constitute a technology that merits investigation at an implementation level, in terms of transduction mechanisms and fundamental and practical noise sources. We believe that biophysical understanding is rich enough to support this approach, and that it offers insight into the mechanisms hidden within the box.

### B. Information Theory and Microsystems

Ralph Landauer first attempted to relate microsystems and technologies to information and communication theory. Landauer [54] discussed the ultimate limitations of computing and employed the entropy reduction in a logical calculation as an energy bound—for each bit of information discarded there should be associated energy costs comparable to  $kT$ . He also suggested that all one-to-one mappings could be reversed and therefore, in principle, should not consume energy.

Mead and Conway [55] further elaborated on the view of computation as entropy reduction and introduce the notion of *logical entropy* and *spatial entropy*. Logical entropy  $S_L$  is related to the logical operations performed on a given assemblage of data while *spatial entropy*  $S_S$  is a measure of the data being in the wrong place. It requires communication operations to get the data in the right place, in order

to do the logical operations. Those channels we have characterized as *transfer* channels correspond to spatial entropy, requiring communication operations, and those channels we have characterized as *transform* channels relate to logical entropy, requiring computation operations.

Hosticka [56] employed information theoretic arguments to compare different signal representations for signal processing in electronic circuits; our work in [57] and [39] improved and extended the work of Hosticka. Information theoretic measures of complexity for computation—logical reduction of entropy—have been discussed in [58] and [59]. Different groups have used information theoretic ideas to analyze analog [19] and digital synchronous [60] and asynchronous [61] VLSI circuits.

Another seminal contribution in this field is the work of Shanbhag [62] aimed at deriving fundamental limits in power dissipation for digital VLSI circuits, working at the circuit level. The work presented in this paper takes his work one step further by considering analog and physical channels (other than electronic) which are modeled at more fundamental levels of description and which relate the information processing performance to the fundamental physical properties of signal transduction and corruption by noise.

## VI. CONCLUSION

In this paper, we have presented a basic methodology that enables a quantitative and fundamental link between function and structure in sensory microsystems. Our theoretical framework can be employed to analyze the tradeoffs between information processing performance and power costs, and to understand which noise sources can be neglected under which operating conditions. As a specific example, we have performed a comparative study between biophysical and silicon models of an adaptive photoreceptor.

From an engineering perspective, the methodology presented here offers a possible path to study the impact of technological advances in microsystem technologies. Information processing in physical microstructures proceeds as forces act deterministically on physical entities of charge and mass. Accordingly, energy in information bearing degrees of freedom (signal) must be available to do useful work with minimal coupling to noninformation bearing degrees of freedom (noise). The scaling of dynamics, signal power, and noise power as a function of the size of the structure ultimately determines the classes of physical devices that are suitable for reaching gigascale integration limits.

From a science perspective, the theoretical framework presented in this paper offers a fundamental and quantitative approach to study the scaling of the information capacity of physical structures as a function of the physical dimension and properties for different biological, mechanical, chemical, and electrical substructures. Such a study will elucidate further the evolutionary principles of biological systems that incorporate heterogeneous elements at a hierarchy of scales from molecules, to cells, networks, and behavior within a physical environment.

## ACKNOWLEDGMENT

The authors acknowledge the work of Simon Laughlin and Paul Furth, which provided inspiration. The authors also acknowledge Marc Cohen, whose thoughtful commentary has improved this work.

## REFERENCES

- [1] W. Bialek, "Physical limits to sensation and perception," *Ann. Rev. Biophys. Biophys. Chem.*, vol. 16, pp. 455–478, 1987.
- [2] D. Hubel, *Eye, Brain and Vision*. New York: Scientific American Library, 1988.
- [3] P. Churchland and T. Sejnowski, *The Computational Brain*. Cambridge, MA: MIT Press, 1991.
- [4] L. Aiello and P. Wheeler, "The expensive-tissue hypothesis," *Curr. Anthropol.*, vol. 36, pp. 199–201, Apr. 1995.
- [5] A. Hakeem, G. Sandoval, M. Jones, and J. Allman, *Brain and Life Span in Primates*, 4th ed. New York: Academic, 1996, ch. 5.
- [6] B. Hille, *Ionic Channels of Excitable Membranes*, 2nd ed. Sunderland, MA: Sinauer, 1992.
- [7] C. A. Mead, *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley, 1989.
- [8] C. Koch, "Seeing chips: Analog VLSI circuits for computer vision," *Neural Comput.*, vol. 1, pp. 184–200, 1989.
- [9] C. A. Mead, "Neuromorphic electronic systems," *Proc. IEEE*, vol. 78, pp. 1629–1635, Oct. 1990.
- [10] R. Sharpshkar, J. Kramer, G. Indiveri, and C. Koch, "Analog VLSI architectures for motion processing: From fundamental limits to system applications," *Proc. IEEE*, vol. 84, pp. 969–987, July 1996.
- [11] E. Fossum, "CMOS image sensors: Electronic camera-on-a-chip," *IEEE Trans. Electron Devices*, vol. 44, no. 10, pp. 1689–1698, 1997.
- [12] T. Delbrück and C. A. Mead, "Analog VLSI phototransduction by continuous-time, adaptive, logarithmic photoreceptor circuits," *Caltech CNS Mem.* 30, 1996.
- [13] E. Vittoz, "Present and future industrial applications of bio-inspired VLSI systems," in *Proc. 1999 Microneuron Conf.*, Los Alamitos, CA, 1999, pp. 2–11.
- [14] A. G. Andreou and P. Abshire, "Sensory communication microsystems: analysis and synthesis framework," Dept. of Elect. and Comput. Eng., Johns Hopkins Univ., Baltimore, MD, JHU-ECE-01/01, 2001.
- [15] C. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Oct. 1948.
- [16] J. Atick and N. Redlich, "What does the retina know about natural scenes," *Neural Comput.*, vol. 4, no. 2, pp. 196–210, 1992.
- [17] J. H. van Hateren, "A theory of maximizing sensory information," *Biol. Cybern.*, vol. 68, pp. 23–29, 1992.
- [18] K. Boahen, "Retinomorph vision systems: Reverse engineering the vertebrate retina," Ph.D. dissertation, California Inst. Technology, 1997.
- [19] P. Furth and A. Andreou, "A design framework for low power analog filter banks," *IEEE Trans. Circuits Syst. I*, vol. 42, pp. 966–971, Nov. 1995.
- [20] F. Huck, C. Fales, and Z. Rahman, "An information theory of visual communication," *Philos. Trans. R. Soc. London A*, vol. 354, pp. 2193–2248, 1996.
- [21] C. Fales, F. Huck, R. Alter-Gartenberg, and Z. Rahman, "Image gathering and digital restoration," *Philos. Trans. R. Soc. London A*, vol. 354, pp. 2249–2287, 1996.
- [22] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [23] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10–21, Jan 1949.
- [24] M. Juusola, R. O. Uusitalo, and M. Weckström, "Transfer of graded potentials at the photoreceptor-interneuron synapse," *J. Gen. Physiol.*, vol. 105, pp. 117–148, 1995.
- [25] R. R. de Ruyter van Steveninck and S. B. Laughlin, "The rate of information transfer at graded-potential synapses," *Nature*, vol. 379, pp. 642–645, Feb. 1996.
- [26] S. Chyb, P. Raghun, and R. C. Hardie, "Polyunsaturated fatty acids activate the *Drosophila* light-sensitive channels TRP and TRPL," *Nature*, vol. 397, pp. 255–259, Jan. 1999.
- [27] F. Wong and B. W. Knight, "Adapting-bump model for eccentric cells of *Limulus*," *J. Gen. Physiol.*, vol. 76, pp. 539–557, Nov. 1980.

- [28] F. Wong, B. W. Knight, and F. A. Dodge, "Dispersion of latencies in photoreceptors of *Limulus* and the adapting-bump model," *J. Gen. Physiol.*, vol. 76, pp. 517–537, Nov. 1980.
- [29] —, "Adapting-bump model for ventral photoreceptors of *Limulus*," *J. Gen. Physiol.*, vol. 79, pp. 1089–1113, June 1982.
- [30] J. H. van Hateren, "Electrical coupling of neuro-ommatidial photoreceptor cells in the blowfly," *J. Comp. Physiol. A*, vol. 158, pp. 795–811, 1986.
- [31] P. Abshire and A. G. Andreou, "Relating information capacity to a biophysical model of the blowfly retina," Dept. of Elect. and Comput. Eng., Johns Hopkins Univ., Baltimore, MD, JHU/ECE-98-13, 1998.
- [32] —, "Relating information capacity to a biophysical model for blowfly retina," in *Proc. 1999 Int. Joint Conf. Neural Networks*, Washington, DC, July 1999.
- [33] —, "Information capacity of the blowfly retina," in *Proc. 33rd Conf. Information Sciences and Systems*, vol. 2, Baltimore, MD, Mar. 1999.
- [34] —, "A communication channel model for information transmission in the blowfly photoreceptor," *Biosyst. J.*, to be published.
- [35] M. Juusola, E. Kouvalainen, M. Järvillehto, and M. Weckström, "Contrast gain, signal-to-noise ratio, and linearity in light-adapted blowfly photoreceptors," *J. Gen. Physiol.*, vol. 104, pp. 593–621, Sept. 1994.
- [36] M. Juusola and M. Weckström, "Band-pass filtering by voltage-dependent membrane in an insect photoreceptor," *Neurosci. Lett.*, vol. 154, pp. 84–88, 1993.
- [37] S. B. Laughlin, "A simple coding procedure enhances a neurone's information capacity," *Zeitschrift für Naturforschung*, vol. 36c, pp. 910–912, 1981.
- [38] S. Laughlin, R. de Ruyter van Steveninck, and J. Anderson, "The metabolic cost of neural information," *Nature Neurosci.*, vol. 1, pp. 36–41, May 1998.
- [39] A. G. Andreou and P. Furth, "An information theoretic framework for comparing the bit-energy of signal representations at the circuit level," in *Low-Voltage/Low-Power Integrated Circuits and Systems*, E. Sanchez-Sinencio and A. G. Andreou, Eds. Piscataway, NJ: IEEE Press, 1998, ch. 8.
- [40] H. B. Barlow, "Unsupervised learning," *Neural Comput.*, vol. 1, no. 3, pp. 295–311, 1989.
- [41] F. Attneave, "Some informational aspects of visual perception," *Psych. Rev.*, vol. 61, no. 3, pp. 183–193, 1954.
- [42] H. B. Barlow, "Possible principles underlying the transformation of sensory messages," in *Sensory Communication*, W. A. Rosenblith, Ed. Cambridge, MA: MIT Press, 1961, pp. 217–234.
- [43] J. J. Atick, "Could information theory provide an ecological theory of sensory processing?," *Network* 3, pp. 213–251, 1992.
- [44] R. Linsker, "From basic network principles to neural architecture: Emergence of spatial-opponent cells, Emergence of orientation-selective cells, and Emergence of orientation columns," *Proc. Nat. Acad. Sci. USA*, vol. 83, 1986.
- [45] S. B. Laughlin, "Matching coding, circuits, cells, and molecules to signals: General principles of retina design in the fly's eye," *Progr. Retinal Eye Res.*, vol. 13, pp. 165–196, 1994.
- [46] D. M. MacKay and W. S. McCulloch, "The limiting information capacity of a neuronal link," *Bull. Math. Phys.*, vol. 14, pp. 127–135, 1952.
- [47] R. B. Stein, "The information capacity of nerve cells using a frequency code," *Biophys. J.*, vol. 7, pp. 797–826, 1967.
- [48] W. B. Levy and R. A. Baxter, "Energy efficient neural codes," *Neural Comput.*, vol. 8, pp. 531–543, 1996.
- [49] R. Eckhorn and B. Pöpel, "Rigorous and extended application of information theory to the afferent visual system of the cat. I. Basic concepts, and II. Experimental results," *Kybernetik*, vol. 16 and 17, 1974.
- [50] F. E. Theunissen and J. P. Miller, "Representation of sensory information in the cricket cercal sensory system. II. Information theoretic calculation of system accuracy and optimal tuning-curve widths of four primary interneurons," *J. Neurophysiol.*, vol. 66, pp. 1690–1703, Nov. 1991.
- [51] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code*. Cambridge, MA: MIT Press, 1997.
- [52] G. T. Buračas, A. M. Zador, M. R. DeWeese, and T. D. Albright, "Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex," *Neuron*, vol. 20, pp. 959–969, May 1998.
- [53] A. Manwani and C. Koch, "Detecting and estimating signals in noisy cable structures: I. Neuronal noise sources, and II. Information theoretical analysis," *Neural Comput.*, vol. 11, pp. 1797–1873, Nov 1999.
- [54] R. Landauer, "Irreversibility and heat generation in the computing process," *IBM J. Res. Develop.*, vol. 5, pp. 183–191, Jan. 1961.
- [55] C. Mead and L. Conway, *Introduction to VLSI Systems*. Reading, MA: Addison-Wesley, 1980.
- [56] B. Hosticka, "Performance comparison of analog and digital circuits," *Proc. IEEE*, vol. 73, pp. 25–29, Jan. 1985.
- [57] P. Furth and A. Andreou, "Comparing the bit-energy of continuous and discrete signal representations," in *Proc. 4th Workshop Physics and Computation (PhysComp96)*, Boston, MA, Nov. 1996, pp. 127–133.
- [58] G. J. Chaitin, "Information theoretic computational complexity," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 10–15, Jan. 1974.
- [59] Y. Abu-Mostafa, "Complexity of information extraction," Ph.D. dissertation, California Inst. Technology, 1983.
- [60] R. M. D. Marculescu and M. Pedram, "Information theoretic measures of energy consumption at register transfer level," in *Proc. ACM/IEEE Int. Symp. Low Power Design*, 1995.
- [61] J. Tierno, "An energy-complexity model for VLSI computation," Ph.D. dissertation, California Inst. Technology, 1995.
- [62] N. R. Shanbhag, "A mathematical basis for power-reduction in digital VLSI systems," *IEEE Trans. Circuits Syst.*, vol. 44, pp. 935–951, 1997.



**Pamela Abshire** (Student Member, IEEE) was born in North Carolina in 1970. She received the B.S. degree in physics with honors from the California Institute of Technology in 1992. She received the M.S. degree in electrical and computer engineering from Johns Hopkins University, Baltimore, MD, in 1997, and is pursuing the Ph.D. degree there.

Between 1992 and 1995, she worked as a Research Engineer in the Bradycardia Research Department of Medtronic, Inc., Minneapolis, MN.

Her research interests include information theory for physical systems, noise theory for electronic, photonic, and biological systems, sensory information processing, and algorithm, circuit, and microsystem design for low power applications.

She is the recipient of a National Science Foundation Graduate Fellowship and a Customer-Focused Quality Award from Medtronic.

**Andreas G. Andreou** (Member, IEEE) was born in Nicosia Cyprus in 1957. He received the Ph.D. degree in electrical engineering and computer science from Johns Hopkins University, Baltimore, MD, in 1986.

Between 1986 and 1989, he held post-doctoral fellow and associate research scientist positions in the electrical and computer engineering department while also a member of the professional staff at the Johns Hopkins Applied Physics Laboratory. Andreou became an Assistant Professor of electrical and computer engineering in 1989, Associate Professor in 1993, and Professor in 1996. In 1995 and 1997, he was a Visiting Associate Professor and Visiting Professor, respectively, in the computation and neural systems program at the California Institute of Technology. In 1999, he was appointed as the first Chair and Director of the computer engineering program at Johns Hopkins University. He is the co-founder of the Center for Language and Speech Processing at Johns Hopkins University. His research interests include electron devices, micropower integrated circuits, neuromorphic VLSI, sensory information processing and neural computation. He is a co-editor of the books *Low-Voltage/Low-Power Integrated Circuits and Systems* (Piscataway, NJ: IEEE Press, 1998) and *Adaptive Resonance Theory Microchips* (Norwell, MA: Kluwer, 1998).

Dr. Andreou is a recipient of a National Science Foundation Research Initiation Award. In 1989 and 1991, he was awarded the R.W. Hart Prize for his work on integrated circuits for space applications. He is the recipient of the 1995 and 1997 Myril B. Reed Best Paper Award and the 2000 IEEE Circuits and Systems Society Darlington Award. He is Associate Editor of the Journal IEEE TRANSACTIONS ON NEURAL NETWORKS and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—PART II.