Toward a unified theory of efficient, predictive, and sparse coding

Matthew Chalk[a,b,1], Olivier Marre[b], and Gašper Tkačik[a]

[a]Department of Physical Sciences, Institute of Science and Technology Austria, 3400 Klosterneuburg, Austria; and [b]Sorbonne Universités, Université de Pierre et Marie Curie Paris 06, INSERM, CNRS, Institut de la Vision, 75012 Paris, France

Edited by Charles F. Stevens, The Salk Institute for Biological Studies, La Jolla, CA, and approved November 20, 2017 (received for review June 22, 2017)

A central goal in theoretical neuroscience is to predict the response properties of sensory neurons from first principles. To this end, "efficient coding" posits that sensory neurons encode maximal information about their inputs given internal constraints. There exist, however, many variants of efficient coding (e.g., redundancy reduction, different formulations of predictive coding, robust coding, sparse coding, etc.), differing in their regimes of applicability, in the relevance of signals to be encoded, and in the choice of constraints. It is unclear how these types of efficient coding relate or what is expected when different coding objectives are combined. Here we present a unified framework that encompasses previously proposed efficient coding models and extends to unique regimes. We show that optimizing neural responses to encode predictive information can lead them to either correlate or decorrelate their inputs, depending on the stimulus statistics; in contrast, at low noise, efficiently encoding the past always predicts decorrelation. Later, we investigate coding of naturalistic movies and show that qualitatively different types of visual motion tuning and levels of response sparsity are predicted, depending on whether the objective is to recover the past or predict the future. Our approach promises a way to explain the observed diversity of sensory neural responses, as due to multiple functional goals and constraints fulfilled by different cell types and/or circuits.

neural coding | prediction | information theory | sparse coding | efficient coding

Sensory neural circuits perform a myriad of computations, which allow us to make sense of and interact with our environment. For example, neurons in the primary visual cortex encode information about local edges in an image, while neurons in higher-level areas encode more complex features, such as textures or faces. A central aim of sensory neuroscience is to develop a mathematical theory to explain the purpose and nature of such computations and ultimately, predict neural responses to stimuli from first principles.

The influential "efficient coding" theory posits that sensory circuits encode maximal information about their inputs given internal constraints, such as metabolic costs and/or noise (1–4); similar ideas have recently been applied in genetic and signaling networks (5, 6). While conceptually simple, this theory has been extremely successful in predicting a host of different neural response properties from first principles. Despite these successes, however, there is often confusion in the literature, due to a lack of consensus on (*i*) what sensory information is relevant (and thus, should be encoded) and (*ii*) the internal constraints (determining what information can be encoded).

One area of potential confusion is between different ideas of why and how neural networks may need to make predictions. For example, given low noise, efficient coding predicts that neurons should remove statistical dependencies in their inputs so as to achieve nonredundant, statistically independent responses (3, 4, 7–9). This can be implemented within a recurrent network where neurons encode a prediction error equal to the difference between their received inputs and an internally generated expectation, hence performing "predictive coding" (10–13). However, Bialek and coworkers (14, 15) recently proposed an alternative theory, in which neurons are hypothesized to preferentially encode sensory information that can be used to predict the future, while discarding other nonpredictive information (14–17). While both theories assume that neural networks make predictions, they are not equivalent: one describes how neurons should compress incoming signals, and the other describes how neurons should selectively encode only predictive signals. Signal compression requires encoding surprising stimuli not predicted by past inputs; these are not generally the same as predictive stimuli, which are informative about the future (16).

Another type of code that has been studied extensively is "sparse coding": a population code in which a relatively small number of neurons are active at any one time (18). While there are various reasons why a sparse code may be advantageous (19–21), previous work has shown that sparse coding emerges naturally as a consequence of efficient coding of natural sensory signals with a sparse latent structure (i.e., generated by combining many sensory features, few of which are present at any one time) (22). Sparse coding has been successful in predicting many aspects of sensory neural responses (23, 24), notably the orientation and motion selectivity of neurons in the primary visual cortex (25–29). Nonetheless, it is unclear how sparse coding is affected by other coding objectives, such as efficiently predicting the future from past inputs.

An attempt to categorize the diverse types of efficient coding is presented in *SI Appendix*, *Efficient Coding Models*. To consistently organize and compare these different ideas, we present a unifying framework based on the information bottleneck (IB) (30). In our work, a small set of optimization parameters

Significance

Sensory neural circuits are thought to efficiently encode incoming signals. Several mathematical theories of neural coding formalize this notion, but it is unclear how these theories relate to each other and whether they are even fully consistent. Here we develop a unified framework that encompasses and extends previous proposals. We highlight key tradeoffs faced by sensory neurons; we show that trading off future prediction against efficiently encoding past inputs generates qualitatively different predictions for neural responses to natural visual stimulation. Our approach is a promising first step toward theoretically explaining the observed diversity of neural responses.

Author contributions: M.C., O.M., and G.T. designed research, performed research, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

[1]To whom correspondence should be addressed. Email: matthewjchalk@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1711114115/-/DCSupplemental.

determines the goals and constraints faced by sensory neurons. Previous theories correspond to specific values of these parameters. We investigate the conditions under which different coding objectives have conflicting or synergistic effects on neural responses and explore qualitatively unique coding regimes.

Efficient Coding with Varying Objectives/Constraints

We consider a temporal stimulus, $x_{-\infty:t} \equiv (\ldots, x_{t-1}, x_t)$, which elicits neural responses, $r_{-\infty:t} \equiv (\ldots, r_{t-1}, r_t)$. We seek a neural code described by the probability distribution $p(r_t|x_{-\infty:t})$, such that neural responses within a temporal window of length τ encode maximal information about the stimulus at lag Δ given fixed information about past inputs (Fig. 1*A*). This problem can be formalized using the IB framework (30–32) by seeking a code, $p(r_t|x_{-\infty:t})$, that maximizes the objective function:

$$L_{p(r_t|x_{-\infty:t})} = I\left(R_{t-\tau:t}; X_{t+\Delta}\right) - \gamma I\left(R_t; X_{-\infty:t}\right), \quad [1]$$

where the first term (to be maximized) is the mutual information between the responses between $t-\tau$ and τ and the stimulus at time $t+\Delta$, while the second term (to be constrained) is the mutual information between the response at time t and past inputs (which we call the coding capacity, C). A constant, γ, controls the tradeoff between coding fidelity and compression. This objective function can be expanded as

$$L_{p(r_t|x_{-\infty:t})} = \langle \log p\left(x_{t+\Delta}|r_{t-\tau:t}\right) - \log p\left(x_{t+\Delta}\right) - \gamma \log p\left(r_t|x_{t-\infty:t}\right) + \gamma \log p\left(r_t\right)\rangle_{p(r,x)}. \quad [2]$$

Previously, we showed that, in cases where it is not possible to compute this objective function directly, one can use approximations of $p(x_{t+\Delta}|r_{t-\tau:t})$ and $p(r_t)$ to obtain a lower bound, $\tilde{L} \leq L$, that can be maximized tractably (31) (*SI Appendix*,*General Framework*).

From Eqs. **1** and **2**, we see that the optimal coding strategy depends on three factors: the decoding lag, Δ; the code length, τ; and the coding capacity, C (determined by γ). Previous theories of neural coding correspond to specific regions within the 3D parameter space spanned by Δ, τ, and C (Fig. 1*B*). For example, temporal redundancy reduction (3, 33) occurs (*i*) at low internal noise (i.e., high C), (*ii*) where the objective is to encode the recent past ($\Delta < 0$), and (*iii*) where information about the stimulus can be read out by integrating neural responses over time ($\tau \gg 0$). Increasing the internal noise (i.e., decreasing C) results in a temporally redundant "robust" code (34–37) (blue sphere in Fig. 1*B*). Recent work positing that neurons efficiently encode information about the future ($\Delta > 0$) looked exclusively at near-instantaneous codes, where $\tau \sim 0$ (red plane in Fig. 1*B*) (15, 38–40). Here, we investigate the relation between these previous works and focus on the (previously unexplored) case of neural codes that are both predictive ($\Delta > 0$) and temporal ($\tau > 0$) and have varying signal to noise (variable C) (black sphere in Fig. 1*B*).

To specialize our theory to the biologically relevant case, we later investigate efficient coding of natural stimuli. A hallmark of natural stimuli is their sparse latent structure (18, 22, 25, 26): stimulus fragments can be constructed from a set of primitive features (e.g., image contours), each of which occurs rarely (Fig. 1*C*). Previous work showed that, in consequence, redundancy between neural responses is minimized by maximizing their sparsity (*SI Appendix*, *Efficient Coding Models*) (22). Here, we investigated what happens when the objective is not to minimize redundancy but rather, to efficiently predict future stimuli given finite coding capacity.

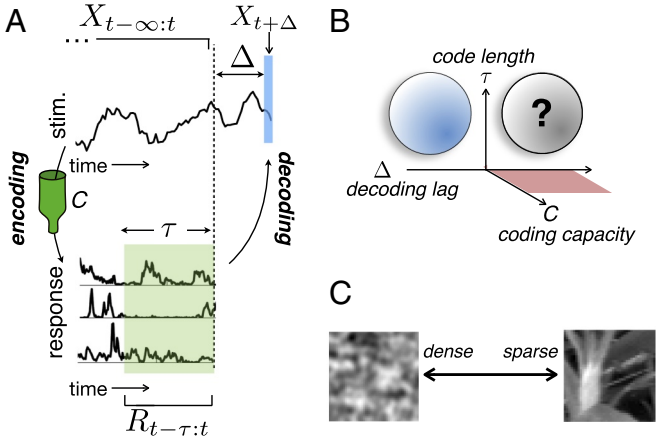

Fig. 1. Schematic of modeling framework. (*A*) A stimulus (stim.) (*Upper*) elicits a response in a population of neurons (*Lower*). We look for codes where the responses within a time window of length τ maximize information encoded about the stimulus at lag Δ, subject to a constraint on the information about past inputs, *C*. (*B*) For a given stimulus, the optimal code depends on three parameters: τ, Δ, and *C*. Previous work on efficient temporal coding generally looked at $\tau > 0$ and $\Delta < 0$ (blue sphere). Recent work posited that neurons encode maximal information about the future ($\Delta > 0$) but only treated instantaneous codes $\tau \sim 0$ (red plane). Our theory is valid in all regimes, but we focus in particular on $\Delta > 0$ and $\tau > 0$ (black sphere). (*C*) We further explore how optimal codes change when there is a sparse latent structure in the stimulus (natural image patch; *Right*) vs. when there is none (filtered noise; *Left*).

Results

Dependence of Neural Code on Coding Objectives. Our initial goal was to understand the influence of different coding objectives in the simplest scenario, where a single neuron linearly encodes a 1-d input. In this model, the neural response at time t is $r_t = \sum_{k=0}^{\tau_w} w_k x_{t-k} + \eta_t$, where $w = (w_0, \ldots, w_{\tau_w})$ are the linear coding weights and η_t is a Gaussian noise with unit variance.*

With 1-d stimuli that have Gaussian statistics, the IB objective function takes a very simple form:

$$L = -\frac{1}{2}\log\left\langle\left(x_{t+\Delta} - \sum_{k=0}^{\tau} u_k r_{t-k}\right)^2\right\rangle - \gamma\frac{1}{2}\log\left\langle r_t^2\right\rangle, \quad [3]$$

where $u = (u_0, \ldots, u_\tau)$ are the optimal linear readout weights used to reconstruct the stimulus at time $t+\Delta$ from the responses between $t-\tau$ and t. Thus, the optimal code is the one that minimizes the mean-squared reconstruction error at lag Δ, constrained by the variance of the neural response (relative to the noise variance).†

Initially, we investigated "instantaneous" codes, where $\tau = 0$, so that the stimulus at time $t+\Delta$ is estimated from the instantaneous neural response at time t (Fig. 2*A*). We considered three different stimulus types, which are shown in Fig. 2*B*. With a "Markov" stimulus (Fig. 2*B*, *Top* and *SI Appendix*, *Methods for Simulations in the Main Text*), with a future trajectory that depended solely on the current state, x_t, the neurons only needed to encode x_t to predict the stimulus at a future time, $x_{t+\Delta}$. Thus, when $\tau = 0$, we observed the trivial solution where $r_t \propto x_t$, irrespective of the decoding lag, Δ (Fig. 2 *C* and *D* and *SI Appendix*, Fig. S2*A*).

With a "two-timescale" stimulus constructed from two Markov processes that vary over different timescales (Fig. 2*B*, *Middle*), the optimal solution was a low-pass filter to selectively encode

* τ_w is the encoding filter length, not to be confused with τ, the decoding filter length.

† We omitted the constant stimulus entropy term, $\langle \log p(x_{t+\Delta})\rangle$, from Eq. **3** and the noise entropy term, $\langle \log p\left(r_t|x_{t-\infty:t}\right)\rangle$ [since with no loss of generality, we assume a fixed amplitude additive noise (32)].

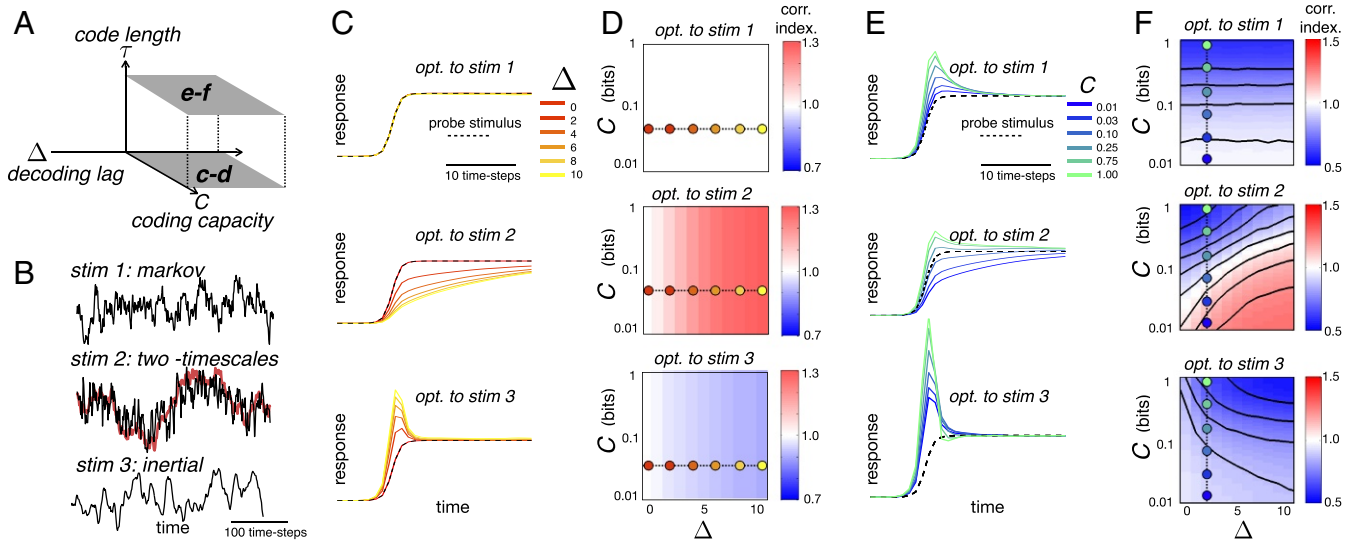

Fig. 2. Dependence of optimal code on decoding lag, Δ; code length, τ; and coding capacity, C. (*A*) We investigated two types of code: instantaneous codes, where $\tau = 0$ (*C* and *D*), and temporal codes, where $\tau > 0$ (*E* and *F*). (*B*) Training stimuli (stim.) used in our simulations. Markov stimulus: future only depends on the present state. Two-timescale stimulus: sum of two Markov processes that vary over different timescales (slow stimulus component is shown in red). Inertial stimulus: future depends on present position and velocity. (*C*) Neural responses to probe stimulus (dashed lines) after optimization (opt.) with varying Δ and $\tau = 0$. Responses are normalized by the final steady-state value. (*D*) Correlation (corr.) index after optimization with varying Δ and C. This index measures the correlation between responses at adjacent time steps normalized by the stimulus correlation at adjacent time steps (i.e., $\langle r_t r_{t+1} \rangle / \left\langle r_t^2 \right\rangle$ divided by $\langle x_t x_{t+1} \rangle / \left\langle x_t^2 \right\rangle$). Values greater/less than one indicate that neurons temporally correlate (red)/decorrelate (blue) their input. Filled circles show the parameter values used in *C*. (*E* and *F*) Same as *C* and *D* but with code optimized for $\tau \gg 0$. Plots in *E* correspond to responses to probe stimulus (dashed lines) at varying coding capacity and fixed decoding lag (i.e., $\Delta = 3$; indicated by dashed lines in *F*).

the predictive, slowly varying part of the stimulus. The strength of the low-pass filter increased monotonically with Δ (Fig. 2 *C* and *D* and *SI Appendix*, Fig. S2*A*).

Finally, with an "inertial" stimulus, with a future trajectory that depended on both the previous state, x_t, and velocity, $x_t - x_{t-1}$ (Fig. 2*B*, *Bottom*), the optimal solution was a high-pass filter so as to encode information about velocity. The strength of this high-pass filter also increased monotonically with Δ (Fig. 2 *C* and *D* and *SI Appendix*, Fig. S2*A*, *Bottom*).

With an instantaneous code, varying the coding capacity, C, only rescales responses (relative to the noise amplitude) so as to alter their signal-to-noise ratio. However, the response shape is left unchanged (regardless of the stimulus statistics) (Fig. 2*D*). In contrast, with temporally extended codes, where $\tau > 0$ (so the stimulus at time $t+\Delta$ is estimated from the integrated responses between time $t - \tau$ and t) (Fig. 2*A*), the optimal neural code varies with the coding capacity, C. As with previous efficient coding models, at high C (i.e., high signal-to-noise ratio), neurons always decorrelated their input, regardless of both the stimulus statistics and the decoding lag, Δ (to achieve nonredundant responses) (*SI Appendix*, *Efficient Coding Models*), while decreasing C always led to more correlated responses (to achieve a robust code) (*SI Appendix*, *Efficient Coding Models*) (36). However, unlike previous efficient coding models at low to intermediate values of C (i.e., intermediate to low signal-to-noise ratio), the optimal code was qualitatively altered by varying the decoding lag, Δ. With the Markov stimulus, increasing Δ had no effect; with the two-timescale stimulus, it led to low-pass filtering, and with the inertial stimulus, it led to stronger high-pass filtering.

Taken together, "phase diagrams" for optimal, temporally extended codes show how regimes of decorrelation/whitening (high-pass filtering) and of smoothing (low-pass filtering) are preferred depending on the coding capacity, C, and decoding lag, Δ. We verified that a qualitatively similar transition from low- to high-pass filtering is also observed with higher dimensional stimuli and/or more neurons. Importantly, we show that these phase diagrams depend in an essential way on the stimulus statistics already in the linear Gaussian case. We next examined what happens for non-Gaussian, high-dimensional stimuli.

Efficient Coding of Naturalistic Stimuli. Natural stimuli exhibit a strongly non-Gaussian statistical structure, which is essential for human perception (22, 41). A large body of work has investigated how neurons could efficiently represent such stimuli by encoding their nonredundant or independent components (4). Under fairly general conditions (e.g., that stimuli have a sparse latent structure), this is equivalent to finding a sparse code: a form of neural population code, in which only small fractions of neurons are active at any one time (22). For natural images, this leads to neurons that are selective for spatially localized image contours, with receptive fields (RFs) that are qualitatively similar to the RFs of V1 simple cells (25, 26). For natural movies, this leads to neurons selective for a particular motion direction, again similar to observations in area V1 (27).

However, an independent (sparse) temporal code has only been shown to be optimal (*i*) when the goal is to maximize information about past inputs (i.e., $\Delta < 0$) and (*ii*) at low noise (i.e., at high capacity; $C \gg 0$). We were interested, therefore, in what happens when these two criteria are violated: for example, when neural responses are optimized to encode predictive information (i.e., for $\Delta \geq 0$).

To explore these questions, we modified the objective function of Eq. **3** to deal with multidimensional stimuli and non-Gaussian statistics of natural images (*SI Appendix*, *General Framework*). Specifically, we generalized the second term of Eq. **3** to allow optimization of the neural code with respect to higher-order (i.e., beyond covariance) response statistics. This was done by approximating the response distribution $p(r)$ by a Student t distribution, with shape parameter, ν, learned directly from data (*SI Appendix*, Eq. **S5**) (31). Crucially, our modification permits—but does not enforce by hand—sparse

neural responses (42). For nonsparse, Gaussian stimuli, the IB algorithm returns $\nu \to \infty$, so that the Student t distribution is equivalent to a Gaussian distribution, and we obtain the results of the previous section; for natural image sequences, it replicates previous sparse coding results in the limit $\Delta < 0$ and $C \gg 0$ (*SI Appendix*, Fig. S5), without introducing any tunable parameters.

We investigated how the optimal neural code for naturalistic stimuli varied with the decoding lag, Δ, while keeping coding capacity, C, and code length, τ, constant. Stimuli were constructed from 10×10-pixel patches drifting stochastically across static natural images (Fig. 3*A*, *SI Appendix*, *Methods for Simulations in the Main Text*, and *SI Appendix*, Fig. S3). Gaussian white noise was added to these inputs (but not the decoded variable, $X_{t+\Delta}$) (*SI Appendix*, *Methods for Simulations in the Main Text*). Neural encoding weights were optimized with two different decoding lags: for $\Delta = -6$, the goal was to encode past stimuli, while for $\Delta = 1$, the goal was to predict the near future. Fig. 3*B* confirms that the codes indeed are optimal for recovering either the past ($\Delta = -6$) or future ($\Delta = 1$) as desired.

After optimization at both values of Δ, individual neurons were selective to local oriented edge features (Fig. 3 *C* and *D*) (25). Varying Δ qualitatively altered the temporal features encoded by each neuron, while having little effect on their spatial selectivity. Consistent with previous results on sparse temporal coding (27), with $\Delta = -6$, single cells were responsive to stimuli moving in a preferred direction as evidenced by spatially displaced encoding filters at different times (Fig. 3*C* and *SI Appendix*, Fig. S6 *A–C*) and a high "directionality index" (Fig. 3*E*). In contrast, with $\Delta = 1$, cells responded equally to stimuli moving in either direction perpendicular to their encoded stimulus orientation. This was evidenced by spatiotemporally separable RFs (*SI Appendix*, Fig. S6 *D–F*) and directionality indexes near zero. This qualitative difference between the two types of code for naturalistic movies was highly surprising, and we sought to understand its origins.

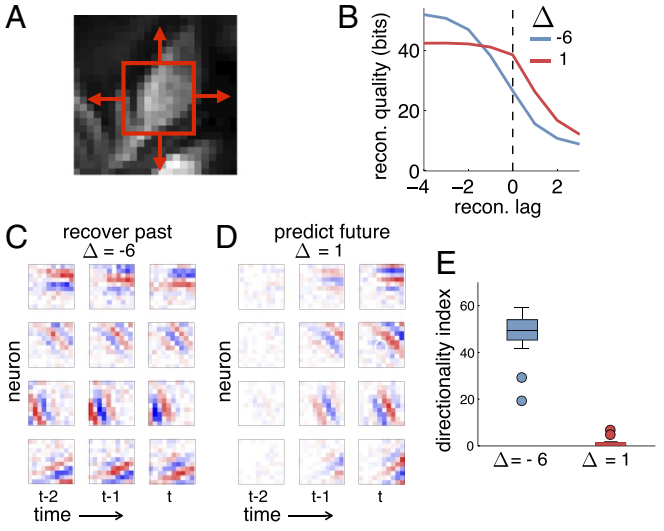

Fig. 3. Efficient coding of naturalistic stimuli. (*A*) Movies were constructed from a 10×10-pixel patch (red square), which drifted stochastically across static natural images. (*B*) Information encoded [i.e., reconstruction (recon.) quality] by neural responses about the stimulus at varying lag (i.e., reconstruction lag) after optimization with $\Delta = -6$ (blue) and $\Delta = 1$ (red). (*C*) Spatiotemporal encoding filters for four example neurons after optimization with $\Delta = -6$. (*D*) Same as *C* for $\Delta = 1$. (*E*) Directionality index of neural responses after optimization with $\Delta = -6$ and $\Delta = 1$. The directionality index measures the percentage change in response to a grating stimulus moving in a neuron's preferred direction vs. the same stimulus moving in the opposite direction.

Tradeoff Between Sparsity and Predictive Power. To gain an intuitive understanding of how the optimal code varies with decoding lag, Δ, we constructed artificial stimuli from overlapping Gaussian bumps, which drifted stochastically along a single spatial dimension (Fig. 4*A* and *SI Appendix*, *Methods for Simulations in the Main Text*). While simple, this stimulus captured two key aspects of the naturalistic movies. First, Gaussian bumps drifted smoothly in space, resembling stochastic global motion over the image patches; second, the stimulus had a sparse latent structure.

We optimized the neural code with Δ ranging from -2 to 2, holding the coding capacity, C, and code length, τ, constant. Fig. 4*B* confirms that highest performance was achieved when the reconstruction performance was evaluated at the same lag for which each model was trained. This simpler setup recapitulated the surprising result that we obtained with naturalistic stimuli: namely, when $\Delta < 0$, neurons were selective to a single preferred motion direction, while when $\Delta \geq 0$, neurons responded equally to stimuli moving from either direction to their RF (Fig. 4 *C* and *D*).

Predicting the future state of the stimulus requires estimating its current motion direction and speed. How is it possible then that optimizing the code for predictions ($\Delta > 0$) results in neurons being unselective to motion direction? This paradox is resolved by realizing that it is the information encoded by the entire neural population that counts, not the information encoded by individual neurons. Indeed, when we looked at the information encoded by the neural population, we did find what we had originally expected: when optimized with $\Delta > 0$, the neural population as a whole encoded significantly more information about the stimulus velocity than its position (relative to when $\Delta < 0$), despite the fact that individual neurons were unselective to motion direction (Fig. 4 *E* and *F*).

The change in coding strategy that is observed as one goes from encoding the past ($\Delta < 0$) to the future ($\Delta > 0$) is in part due to a tradeoff between maintaining a sparse code and cells responding quickly to stimuli within their RF. Intuitively, to maintain highly selective (and thus, sparse) responses, neurons first have to wait to process and recognize the "complete" stimulus feature; unavoidably, however, this entails a processing delay, which leads to poor predictions. This can be seen in Fig. 4 *G* and *H*, which shows how both the response sparsity and delay to stimuli within a cell's RF decrease with Δ. In *SI Appendix*, *Supplementary Simulations*, we describe in detail why this tradeoff between efficiency and prediction leads to direction-selective filters when $\Delta < 0$ but not when $\Delta > 0$ (*SI Appendix*, Fig. S7).

Beyond the effects on the optimal code of various factors explored in detail in this paper, our framework further generalizes previous efficient and sparse coding results to factors listed in *SI Appendix*, Table S1 and discussed in *SI Appendix*, *Supplementary Simulations*. For example, decreasing the capacity, C (while holding Δ constant at -2), resulted in neurons being unselective to stimulus motion (*SI Appendix*, Fig. S8*A*), with a similar result observed for increased input noise (*SI Appendix*, Fig. S8*B*). Thus, far from being generic, traditional sparse temporal coding, in which neurons responded to local motion, was only observed in a specific regime (i.e., $\Delta < 0$, $C \gg 0$, and low input noise).

Discussion

Efficient coding has long been considered a central principle for understanding early sensory representations (1, 3), with well-understood implications and generalizations (23, 37). It has been successful in predicting many aspects of neural responses in early sensory areas directly from the low-order statistics of natural

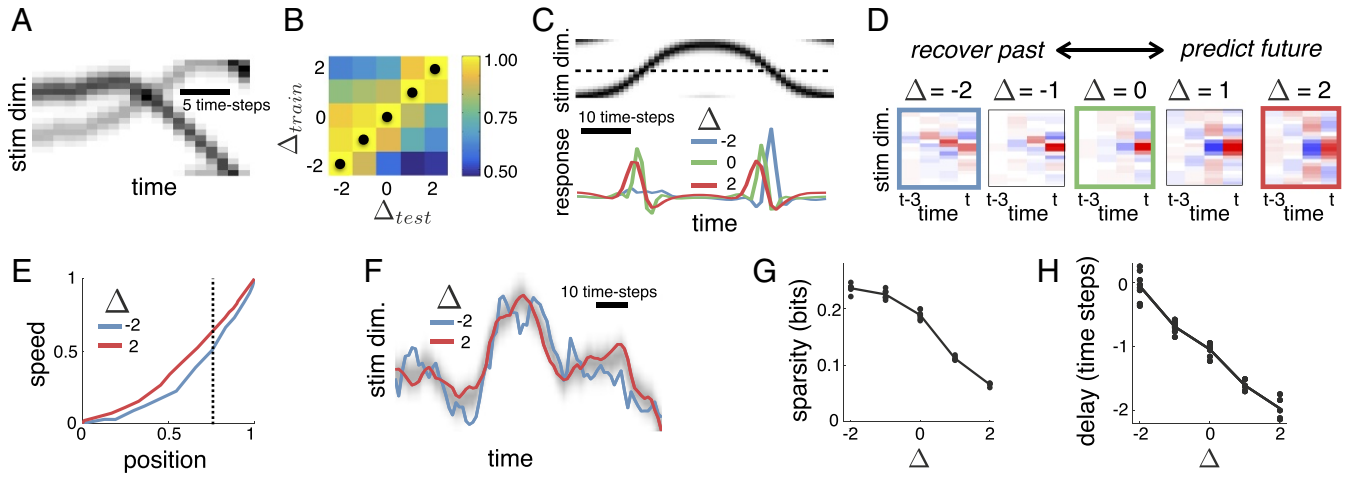

Fig. 4. Efficient coding of a "Gaussian-bump" stimulus. (*A*) Stimuli (stim.) consisted of Gaussian bumps that drifted stochastically along a single spatial dimension (dim.) (with circular boundary conditions). (*B*) Information encoded by neural responses about the stimulus at varying lag, Δ_{test}, after optimization with varying Δ_{train}. Black dots indicate the maximum for each column. (*C*) Response of example neuron to a test stimulus (*Upper*) and after optimization with $\Delta = -2$ (blue), $\Delta = 0$ (green), and $\Delta = 2$ (red; *Lower*). (*D*) Spatiotemporal encoding filters for an example neuron after optimization with different Δ. (*E*) Circular correlation between the reconstructed speed of a moving Gaussian blob and its true speed vs. the circular correlation between the reconstructed position and its true position obtained from neural responses optimized with $\Delta = \pm 2$ (red and blue curves). Curves were obtained by varying γ in Eq. **3** to find codes with different coding capacities. (*F*) Linear reconstruction of the stimulus trajectory obtained from neural responses optimized with $\Delta = \pm 2$ (red and blue curves). The full stimulus is shown in grayscale. While coding capacity was chosen to equalize the mean reconstruction error for both models (vertical dashed line in *E*), the reconstructed trajectory was much smoother after optimization with $\Delta = 2$ than with $\Delta = -2$. (*G*) Response sparsity (defined as the negentropy of neural responses) vs. Δ (dots indicate individual neurons; the line indicates population average). (*H*) Delay between stimulus presented at a neuron's preferred location and each neuron's maximum response vs. Δ.

stimuli (7, 22, 32, 43, 44) and has even been extended to higher-order statistics and central processing (45, 46). However, a criticism of the standard theory is that it treats all sensory information as equal, despite empirical evidence that neural systems prioritize behaviorally relevant (and not just statistically likely) stimuli (47). To overcome this limitation, Bialek and coworkers (14, 15) proposed a modification to the standard efficient coding theory, positing that neural systems are set up to efficiently encode information about the future given fixed information about the past. This is motivated by the fact that stimuli are only useful for performing actions when they are predictive about the future.

The implications of such a coding objective have remained relatively unexplored. Existing work only considered the highly restrictive scenario where neurons maximize information encoded in their instantaneous responses (15, 38, 40). In this case (and subject to some additional assumptions, such as Gaussian stimulus statistics and instantaneous encoding filters), predictive coding is formally equivalent to slow feature analysis (39). This is the exact opposite of standard efficient coding models, which (at low noise/high capacity) predict that neurons should temporally decorrelate their inputs (3, 33).

We developed a framework to clarify the relation between different versions of the efficient coding theory (14, 30, 31). We investigated what happens when the neural code is optimized to efficiently predict the future (i.e., $\Delta > 0$ and $\tau > 0$) (Fig. 1*B*). In this case, the optimal code depends critically on the coding capacity (i.e., signal-to-noise ratio), which describes how much information the neurons can encode about their input. At high capacity (i.e., low noise), neurons always temporally decorrelate their input. At finite capacity (i.e., mid to high noise), however, the optimal neural code varies qualitatively depending on whether the goal is to efficiently predict the future or reconstruct the past.

When we investigated efficient coding of naturalistic stimuli, we found solutions that are qualitatively different from known sparse coding results, in which individual neurons are tuned to directional motion of local edge features (27). In contrast, we found that neurons optimized to encode the future are selective for motion speed but not direction (Fig. 3 and *SI Appendix*, Fig. S6). Surprisingly, however, the neural population as a whole encodes motion even more accurately in this case (Fig. 4*E*). We show that these changes are due to an implicit tradeoff between maintaining a sparse code and responding quickly to stimuli within each cell's RF (Fig. 4 *G* and *H*).

It is notable that, in our simulations, strikingly different conclusions are reached by analyzing single-neuron responses vs. the population responses. Specifically, looking only at single-neuron responses would lead one to conclude that, when optimized for predictions, neurons did not encode motion direction; looking at the neural population responses reveals that the opposite is true. This illustrates the importance of population-level analyses of neural data and how, in many cases, single-neuron responses can give a false impression of which information is represented by the population.

A major challenge in sensory neuroscience is to derive the observed cell-type diversity in sensory areas from a normative theory. For example, in visual area V1, one observes a range of different cell types, some of which have spatiotemporally separable RFs and others do not (48, 49). The question arises, therefore, whether the difference between cell types emerges because different subnetworks fulfill qualitatively different functional goals. One hypothesis, suggested by our work, is that cells with separable RFs have evolved to efficiently encode the future, while cells with nonseparable RFs evolved to efficiently encode the past. More generally, the same hypothesis could explain the existence of multiple cell types in the mammalian retina, with each cell type implementing an optimal code for a particular choice of optimization parameters (e.g., coding capacity or prediction lag).

Testing such hypotheses rigorously against quantitative data would require us to generalize our work to nonlinear encoding and decoding models (*SI Appendix*, Table S1). Here, we focused on a linear decoder to lay a solid theoretical foundation and

permit direct comparison with previous sparse and robust coding models, which also assumed a linear decoder (25–27, 35, 36). In addition, a linear decoder forces our algorithm to find a neural code for which information can be easily extracted by downstream neurons performing biologically plausible operations. While the linearity assumptions simplify our analysis, the framework can easily accommodate nonlinear encoding and decoding. For example, we previously used a "kernel" encoding model, where neural responses are described by a nonparametric and nonlinear function of the input (31). Others have similarly used a deep convolutional neural network as an encoder (50).

As mentioned earlier, predictive coding has been used to describe several different approaches. Clarifying the relationship between inequivalent definitions of predictive coding and linking them mathematically to coding efficiency provided one of the initial motivations for our work. In past work, alternative coding theories are often expressed using very different mathematical frameworks, impeding comparison between them and sometimes leading to confusion. In contrast, by using a single mathematical framework to compare different theories—efficient, sparse, and predictive coding—we were able see exactly how they relate to each other, the circumstances under which they make opposing or similar predictions, and what happens when they are combined.

ACKNOWLEDGMENTS. This work was supported by Agence Nationale de Recherche (ANR) Trajectory, the French State program Investissements d'Avenir managed by the ANR [LIFESENSES: ANR-10-LABX-65], a European Commission Grant (FP7-604102), NIH Grant U01NS090501, and AVIESAN-UNADEV Grant (to O.M.), and Austrian Science Fund Grant FWF P25651 (to G.T.).

1. Attneave F (1954) Some informational aspects of visual perception. *Psychol Rev* 61:183–193.
2. Linsker R (1988) Self-organization in a perceptual network. *IEEE Computer* 21: 105–117.
3. Barlow HB (1961) Possible principles underlying the transformation of sensory messages. *Sensory Communication*, ed Rosenblith WA (MIT Press, Cambridge, MA), pp 217–234.
4. Simoncelli EP, Olshausen BA (2001), Natural image statistics and neural representation. *Ann Rev Neurosci* 24:1193–1216.
5. Tkacik G, Bialek W (2016) Information processing in living systems. *Ann Rev Condens Matter Phys* 7:89–117.
6. Bialek W (2012) *Biophysics: Searching for Principles* (Princeton Univ Press, Princeton), pp 353–468.
7. Atick JJ, Redlich AN (1992), What does the retina know about natural scenes? *Neural Comput* 4:196–210.
8. Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. *J Opt Soc Am A* 4:2379–2394.
9. Kersten D (1987) Predictability and redundancy of natural images. *J Opt Soc Am A* 4:2395–2400.
10. Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2:79–87.
11. Boerlin M, Deneve S (2011) Predictive coding of dynamical variables in balanced spiking networks. *PLoS Comp Biol* 7 e1001080.
12. Srinivasan MV, Laughlin SB, Dubs A (1982) Predictive coding: A fresh view of inhibition in the retina. *Proc R Soc B* 216:427–459.
13. Druckmann S, Hu T, Chklovskii DB (2017) A mechanistic model of early sensory processing based on subtracting sparse representations. *Adv Neural Inf Process Syst* 25:1979–1987.
14. Bialek W, De Ruyter Van Steveninck R, Tishby N (2006) Efficient representation as a design principle for neural coding and computation. *Proceedings of the IEEE International Symposium on Information Theory*, pp 659–663. Available at ieeexplore.ieee.org/abstract/document/4036045/. Accessed December 7, 2017.
15. Palmer SE, Marre O, Berry MJ II, Bialek W (2015) Predictive information in a sensory population. *Proc Natl Acad Sci USA* 112:6908–6913.
16. Salisbury J, Palmer S (2016) Optimal prediction in the retina and natural motion statistics. *J Stat Phys* 162:1309–1323.
17. Heeger DJ (2017) Theory of cortical function. *Proc Natl Acad Sci USA* 114:1773–1782.
18. Olshausen BA, Field DJ (2004) Sparse coding of sensory inputs. *Curr Op Neurobiol* 14:481–487.
19. Barlow HB (1972) Single units and sensation: A neuron doctrine for perceptual psychology? *Perception* 1:371–394.
20. Baum EB, Moody J, Wilczek F (1988) Internal representations for associative memory. *Biol Cybern* 59:217–228.
21. Field DJ (1994) What is the goal of sensory coding? *Neural Comput* 6:559–601.
22. Hyvärinen A, Hurri J, Hoyer PO (2009) *Natural Image Statistics* (Springer, Berlin).
23. Smith EC, Lewicki MS (2006) Efficient coding of natural sounds. *Nature* 439:978–982.
24. Theunissen FE (2003) From synchrony to sparseness. *Trends Neurosi* 26:61–64.
25. Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381:607–609.
26. Bell AJ, Sejnowski TJ (1997), The "independent components" of natural scenes are edge filters. *Vis Res* 37:3327–3338.
27. van Hateren JH, van der Schaaf A (1998) Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc Biol Sci* 265:359–366.
28. van Hateren JH, Ruderman DL (1998) Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proc R Soc Lond B Biol Sci* 265:2315–2320.
29. Olshausen BA (2002) Sparse codes and spikes. *Probabilistic Models of the Brain: Perception and Neural Function*, eds Rao RPN, Olshausen BA, Lewicki MS (MIT Press, Cambridge, MA), pp 257–272.
30. Tishby N,Pereira FC, Bialek W (1999) The information bottleneck method. arXiv:physics/0004057.
31. Chalk M, Marre O, Tkačik G (2016) Relevant sparse codes with variational information bottleneck. *Adv Neural Inf Process Syst* 29:1957–1965.
32. Chechik G, Globerson A, Tishby N, Weiss Y (2005) Information bottleneck for Gaussian variables. *J Machine Learn Res* 6:165–188.
33. Dan Y, Atick JJ, Reid RC (1996) Efficient coding of natural scenes in the lateral geniculate nucleus: Experimental test of a computational theory. *J Neurosci* 16: 3351–3362.
34. Karklin Y, Simoncelli EP (2011), Efficient coding of natural images with a population of noisy linear-nonlinear neurons. *Adv Neural Inf Process Syst* 24:999–1007.
35. Doi E, Lewicki MS (2005) Sparse coding of natural images using an overcomplete set of limited capacity units. *Adv Neural Inf Process Syst* 17:377–384.
36. Doi E, Lewicki MS (2014) A simple model of optimal coding for sensory systems. *PLoS Comput Bio* 10:e1003761.
37. Tkačik G, Prentice JS, Balasubramanian V, Schneidman E (2010) Optimal population coding by noisy spiking neurons. *Proc Nat'l Acad Sci USA* 107:14419–14424.
38. Creutzig F, Sprekeler H (2008), Predictive coding and the slowness principle: An information-theoretic approach. *Neural Comput* 20:1026–1041.
39. Berkes P, Wiskott L (2005) Slow feature analysis yields a rich repertoire of complex cell properties. *J Vis* 5:579–602.
40. Buesing L, Maass W (2010) A spiking neuron as information bottleneck. *Neural Comput* 22:1961–1992.
41. Oppenheim AV, Lim JS (1981) The importance of phase in signals. *Proc IEEE* 69: 529–541.
42. Olshausen BA, Millman KJ (2000) Learning sparse codes with a mixture-of-gaussians prior. *Advances in Neural Information Processing Systems*, 12, Ed Solla SA, Leen TK, Muller KR (MIT Press, Cambridge, MA), pp 841–847.
43. Doi E, et al. (2012), Efficient coding of spatial information in the primate retina. *J Neurosci* 32:16256–16264.
44. Balasubramanian V, Sterling P (2009) Receptive fields and functional architecture in the retina. *J Physiol* 587:2753–2767.
45. Tkačik G, Prentice JS, Victor JD, Balasubramanian V (2010), Local statistics in natural scenes predict the saliency of synthetic textures. *Proc Natl Acad Sci USA* 107:18149–18154.
46. Hermundstad AM, et al. (November 14, 2014), Variance predicts salience in central sensory processing. *eLife*, 10.7554/eLife.03722.
47. Machens CK, Gollisch T, Kolesnikova O, Herz AVM (2005) Testing the efficiency of sensory coding with optimal stimulus ensembles. *Neuron* 47:447–456.
48. DeAngelis GC, Ohzawa I, Freeman R (1995) Receptive-field dynamics in the central visual pathways. *Trends Neurosci* 18:451–458.
49. Priebe NJ, Lisberger SG, Movshon JA (2006) Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex. *J Neurosci* 26: 2941–2950.
50. Alemi A, Fischer I, Dillon JV, Murphy K (2016) Deep variational information bottleneck. arXiv:1612.00410.