# A theory of the visual motion coding in the primary visual cortex [1]

Zhaoping Li
Computer Science Department, Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong

Published in *Neural Computation* vol. 8, no.4, p705-30, May, 1996.

**Abstract**

This paper demonstrates that much of visual motion coding in primary visual cortex can be understood from a theory of efficient motion coding in the multiscale representation. The theory predicts that cortical cells can have a spectrum of directional indices, be tuned to different directions of motion, and have spatio-temporally separable or inseparable receptive fields (RF). The predictions also include the following correlations between motion coding and spatial, chromatic, and stereo codings: the preferred speed is larger when the cell receptive field size is larger, the color channel prefers smaller speed than the luminance channel, and both the optimal speeds and the preferred directions of motion can be different for inputs from different eyes to the same neuron. These predictions agree with experimental observations. In addition, this theory makes predictions that have not been experimentally investigated systematically and provides testing ground for the efficient multiscale coding framework. These predictions are: (1) if nearby cortical cells of a given preferred orientation and scale prefer opposite directions of motion and have quadrature RF phase relationship with each other, then they will have the same directional index; (2) a single neuron can have different optimal motion speeds for opposite motion directions of monocular stimuli, and (3) a neuron's ocular dominance may change with motion direction if the neuron prefers opposite directions for inputs from different eyes.

[1]Work supported by the Research Grant Council of Hong Kong.

## 1. Introduction

Primary visual cortical cells sensitive to motion and selective to motion directions have been observed physiologically since the works of Hubel and Wiesel (1959, 1962). Simple cells are found to be tuned to directions of motion to various degrees in addition to their selectivities to orientation, spatial frequency, ocular origin, etc ( Holub and Morton-Gibson 1981, Foster, Gaska, Nagler, and Pollen 1985, Reid, Soodak, Shapley 1991, DeAngelis, Ohzawa, and Freeman 1994). This paper demonstrates that many of the motion sensitive/directional selective properties in cortical simple cells can be understood as consequences of efficient coding of visual inputs in a multiscale framework. Such an understanding provides detailed predictions of the simple cell spatio-temporal receptive field (RF) properties. These predictions can be compared with known observations or experimentally tested.

Efficiency of information representation has long been advocated as the coding principle for early stages of sensory processing (Barlow 1961). This is because the natural signals have structures and regularities. Visual inputs, for example, have correlated signals in image pixels, making some input signals largely predictable from others. Such regularities make pixel-by-pixel input representation highly redundant or inefficient, in the sense that the same information is signalled wastefully many times through different neural channels. An efficient code with reduced redundancy not only gives coding and neural implementation economy, but also arguably provides cognitive advantages (Barlow 1961) due to the knowledge of input statistics, which has to be inherent in the code to reduce the redundancy.

One of the most noticable visual input redundancies is the pairwise pixel-pixel correlations. Concentrating on such redundancy, several recent works have formulated efficient coding in the language of information theory or decorrelation/factorial codes modified appropriately under noise (Srinivasan Laughlin and Dubs 1982, Linsker 1989, Atick and Redlich 1990, Bialek, Ruderman, and Zee 1991, Nadal and Parga 1993). In particular, efficient coding has provided a theory of retinal processing and predicted the spatio-chromatic receptive fields of the retinal ganglion cells agreeing with those observed physiologically (Srinivasan et al 1982, Atick and Redlich 1990, Atick, Li, and Redlich 1992).

There are other types of regularities in natural images that we believe the visual system beyond the retina takes advantage of. One such regularity is translation and scale invariance, namely, the image of an object at one location or distance can predict much of the image of the same object at another location or distance. It has recently been proposed that one of the preprocessing goals of the early visual cortex is, without compromising coding efficiency, to produce a representation where actions of translation and scaling are manifested or factored out towards object invariance (Li and Atick 1994a). The resulting, so-called *multiscale*, representation remaps the visual field into multiple retinotopic maps identical in all respects except for the densities and RF sizes of their sampling nodes. This representation is also a step towards redundancy reduction when it is followed by attentional mechanisms to compensate the manifested translation and scaling changes to produce object invariant neural activity patterns (Li and Atick 1994ab).

Efficient coding in the multiscale representation has predicted many of the simple cell RF properties in the spatial, chromatic, and stereo domains (Li and Atick 1994ab). These predictions include the simple cell selectivities to orientation, spatial frequency, color, ocular origin,

disparity, as well as the particular frequency tuning bandwidth, phase quadrature structure between neighboring cells, and spatio-chromatic-stereo interactions in cell selectivities observed experimentally. The theoretical understanding further aided the study of the visual system by motivating experimental tests of some predictions which had not been investigated experimentally (Li, 1995, Anzai, DeAngelis, Ohzawa, and Freeman 1994). However, the temporal input dimension was ignored in these earlier theoretical works (Li and Atick 1994ab, Li 1995). The current work demonstrates that including the temporal dimension enables the same framework to additionally predict simple cell motion sensitivities and directional selectivities that have been observed or can be tested experimentally.

The primary visual cortex is likely to have other functions in mind in addition to the goal of efficiency and invariance. Previous works (Li and Atick 1994ab, Li, 1995) did not take into account other possible cortical functions and were limited to only linear coding mechanisms as approximations. This necessarily led to unexplained cortical phenomena and quantitative disagreements between reality and theoretical predictions (see discussion). Extending the previous approach to motion coding, the current work has the same limitations. However, it helps to explore the potential and limitations of the efficient coding framework and provide a testing ground for it with additional predictions which have not yet been experimentally investigated.

Various neurophysiological, psychophysical, and computational motion models have been proposed (e.g., Reichardt 1961, Torre and Poggio 1978, Marr and Ullman 1981, van Santen and Sperling 1984, Adelson and Bergen, 1985, Watson and Ahumada 1985). They are mostly designed to model the neuronal mechanisms underlying directionality or to provide computational algorithms for visual motion detection and computation. Some of them (e.g., Reichardt 1961, Torre and Poggio 1978, Marr and Ullman 1981, van Santen and Sperling 1984) have highly non-linear components at an early stage, either to ensure directionality or to compute motion velocity. Physiological observations, however, reveal essentially linear mechanisms underlying directionality in simple cells (Reid, Soodak, and Shapley, 1991, Jagadeesh, Wheat, Ferster 1993). Motion models of Adelson and Bergen (1985) and Watson and Ahumada (1985) do include linear components before a latter-stage non-linearity and are designed for motion sensing or detection within the constraints of known physiological and psychophysical observations. The current work derives the motion coding using a linear mechanism from the requirement of efficient multiscale representation, without *a priori* specifying the purpose of visual motion computation or selectivity. Its predictions include some that have not been experimentally investigated in addition to ones that agree with known observations. A special case from the derivations will be shown to resemble the linear components in the models of Adelson and Bergen (1985) and Watson and Ahumada (1985).

The next section presents the theoretical formulation of the efficient motion coding in the multiscale representation. Section 3 explores the predicted RFs and correlations between motion coding and codings in the space, color, and stereo domains, to compare them with experimental observations. Section 4 summarizes the results and discusses the limitations and desired experimental tests of the theory.

## 2. Efficient motion coding in a multiscale representation

Visual input is inefficient because the input $S(x,t)$, assumed to be of zero mean for simplicity,

at retina location $x$ and time t is correlated with $S(x',t')$ by the amount

$$R_{x,t;x',t'} \equiv \quad < S(x,t)S(x',t') >, \tag{1}$$

where $< \cdot >$ denotes average over inputs of the visual environment. Without loss of generality, the retina is taken as one-dimensional. Visual inputs are assumed to be statistically translation invariant and reflection symmetric, such that $R_{x,t;x',t'} = R_{x+a,t+\tau;x'+a,t'+\tau} \equiv R(x-x',t-t') = R(\pm(x-x'),\pm(t-t'))$. Then $R$ can also be characterized by its Fourier transform[2] $R(f_j,\omega) = \frac{1}{2\pi\sqrt{N}} \sum_x \int_{-\infty}^{\infty} dt R(x,t) e^{-if_j x - i\omega t} = R(\pm f_j, \pm\omega)$, which is also the average input power in frequency $(f_j,\omega)$. Here $N$ is the total number of input units covering a visual space $x \in (0,N)$ with unit grid spacing.

Under noiseless conditions, a more efficient code $O(j,t)$ can be constructed within the linear coding scheme by a transform $O(j,t) = \sum_x \int_{-\infty}^{\infty} dt' K(j,t;x,t') S(x,t')$ such that the outputs are decorrelated $< O(j,t)O(j',t') >= \delta_{jj'}\delta(t-t')$. If higher order input correlations are ignored, such decorrelated outputs $O(j,t)$ imply that no information is redundantly sent through different output units or at different times. The code $O(j,t)$ is thus efficient. One should note that we require both spatial and temporal decorrelation, in contrast to the mere spatial decorrelation when the temporal dimension was ignored (Li and Atick 1994a). The temporal dimension cannot be treated like the spatial dimension because of causality. In addition, visual object scale invariance does not extend from space to time. Accordingly, the multiscale coding, which is necessitated in the spatial domain by the scale invariance (Li and Atick 1994a), has no *a priori* reason to be applied temporally[3].

A special efficient code $O_j$ is obtained by passing $S(x,t)$ through a spatial filter $K_x^{f_j}$, to achieve spatial decorrelation, and a temporal filter $K_t^{f_j}$, to achieve temporal decorrelation:

$$S(x,t) \quad \rightarrow \quad S(f_j,t) \equiv \sum_x K_x^{f_j}(x) S(x,t) \rightarrow O(f_j,t) \equiv \int_{-\infty}^{\infty} dt' K_t^{f_j}(t-t') S(f_j,t') \tag{2}$$

$$K_x^{f_j}(x) \quad \equiv \quad \frac{1}{\sqrt{N}} e^{-if_j x} \qquad \text{each } j \text{ has a different spatial frequency } f_j \tag{3}$$

$$K_t^{f_j}(t-t') \quad \equiv \quad \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega R^{-1/2}(f_j,\omega) e^{-i\omega(t-t') - i\phi(f_j,\omega)} \tag{4}$$

$$K(j,t,x,t') \quad \equiv \quad K^{f_j} \equiv K_t^{f_j}(t-t') K_x^{f_j}(x) \tag{5}$$

where $\phi(f_j,\omega) = -\phi(f_j,-\omega)$ is chosen such that the temporal filter $K_t^{f_j}(t-t')$ is causal ( $K_t^{f_j}(t<0) = 0$) and has minimum temporal spread[4] for each $j$.

[2]The same symbol $R$ is used for the correlation function $R(x,t)$ as well as its Fourier Transform $R(f,\omega)$. The arguments $(x,t)$ or $(f,\omega)$ specifies the actual function concerned. Such practices are used throughout the paper for some other functions and variables as well to avoid proliferation of notation.

[3]There may be *a posteriori* reasons for multiscale in time, for instance, to compute motion velocity (Grzywacz and Yuille 1990, see Section 3). However, at least in the primary visual cortex, the selectivity to temporal scale is much poorer (Holub and Morton-Gibson 1981 and Foster et al 1985) than that to spatial scale.

[4]Define $A^{f_j}(t) = \frac{1}{\pi} \int_0^{\infty} d\omega R^{-1/2}(f_j,\omega) e^{-i\omega(t-t') - i\phi(f_j,\omega)}$, taking envelope$(t)$ and phase$(t)$ as the amplitude and phase of $A^{f_j}(t)$, then $K_t^{f_j}(t) = \text{envelope}(t)\cos(\text{phase}(t))$. The minimum temporal spread of $K_t^{f_j}$ is defined when $\int_{-\infty}^{\infty} dt(t-\bar{t})^2 \text{envelope}(t)$, where $\bar{t} = \int_{-\infty}^{\infty} dt \text{envelope}(t)$, is minimum

Decorrelation in $O(j,t)$ can be verified as follows. The signal $S(f_j,t)$ is the spatial Fourier transform of $S(x,t)$ for spatial frequency $f_j$. Accordingly, $S(f_j,t)$ and $S^*(f_{j'},t)$, for $f_j \neq f_{j'}$ are decorrelated from each other in a translationally invariant system

$$\begin{aligned} < S(f_j,t)S^*(f_{j'},t') > &= \frac{1}{N}\sum_{x,x'} < S(x,t)S(x',t') > e^{-if_jx+if_{j'}x'} \\ &= \frac{1}{N}\sum_{x,x'} R(x-x',t-t')e^{-if_j(x-x')}e^{-i(f_j-f_{j'})x'} \\ &= R(f_j,t-t')\delta_{jj'}. \end{aligned} \tag{6}$$

where superscript $*$ denotes complex conjugate. Each $S(f_j,t)$ is a temporally correlated signal, which is temporally decorrelated by the transform $S(f_j,t) \rightarrow O(f_j,t) = \int_{-\infty}^{\infty} dt' K_t^{f_j}(t-t')S(f_j,t')$ with the temporal whitening filter $K_t^{f_j}$, which has Fourier transform $K_t^{f_j}(\omega) = R^{-1/2}(f_j,\omega)$:

$$\begin{aligned} < O(f_j,t)O^*(f_j,t') > &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\tau d\tau' K_t^{f_j}(t-\tau-\tau')K_t^{f_j}(t'-\tau)R(f_j,\tau') \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega K^{f_j}(\omega)R(f_j,\omega)K^{f_j}(-\omega)e^{i\omega(t-t')} \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega e^{i\omega(t-t')} = \delta(t-t') \end{aligned} \tag{7}$$

The spatio-temporal RF $K^{f_j} = K_x^{f_j}K_t^{f_j}$ for this efficient code $O_j = K^{f_j}S$ is however not spatially local or retinotopic, simply because $K^{f_j}$ contains a spatial Fourier wave $K_x^{f_j}$ which is non-local[5]. In addition $K_x^{f_j}$ is unique for each output $j$ with unique frequency $f_j$, requiring a unique RF for each output cell. However, other efficient codes can be constructed from this one (Li and Atick 1994a) by any unitary transform $\mathbf{U}$ (where the bold-faced $\mathbf{U}$ denotes a matrix, $\mathbf{U}\mathbf{U}^\dagger = \mathbf{1}$, and $\mathbf{U}^\dagger \equiv (\mathbf{U}^*)^\mathbf{T}$) with $O_j \rightarrow \sum_{j'} U_{jj'}O_{j'}$ and $K^{f_j} \rightarrow \sum_{j'} U_{jj'}K^{f'_j}$. Decorrelation is preserved in the new code since $\sum_{jj'} U_{ij}(U_{i'j'})^* < O_jO_{j'}^* >= \delta_{ii'}$. As noted in the introduction, it was propsed that the goal of the cortex is to construct a multiscale representation, which is also spatially local, retinotopic, and translationally invariant, in the sense that the RF of each cell is the same as that of many other cells in the same scale except for the RF center locations. An efficient code of this multiscale nature is achieved (see Li and Atick 1994a for details) by combining the original filters $K^{f_j}$ or outputs $O(f_j,t)$ within each frequency band $f^a < |f_j| \leq f^{a+1} = 3f^a$ by a unitary transform $\mathbf{U^a}$ in that band:

$$O(f_j,t) \rightarrow O_n^a(t) = \sum_{f^a<|f_j|\leq f^{a+1}} \mathbf{U^a}_{nj}O(f_j,t) \tag{8}$$

$$K^{f_j} \rightarrow K_n^a = \sum_{f^a<|f_j|\leq f^{a+1}} \mathbf{U^a}_{nj}K^{f_j} \tag{9}$$

where $a$ indicates the spatial scale or frequency band, and $K_n^a$ is the spatiotemporal RF for the $n^{th}$ output unit in that scale, $n = 1,2,...,N^a \propto (f^{a+1}-f^a)$. As is shown in the Appendix, the

[5]The RFs for this code are not real, but this representation is used for mathematical convenience and it does not affect the final results.

general spatiotemporal receptive fields of such nature are:

$$\mathbf{K_n^a}(x;t-t')$$

$$\propto \sum_{f^a<f\le f^{a+1}} \int_0^\infty d\omega K(f,\omega)(A^+\cos(\phi(x)+\phi(t))+A^-\cos(\phi(x)-\phi(t))) \quad (10)$$

$$= \sum_{f^a<f\le f^{a+1}} \int_0^\infty d\omega K(f,\omega)((A^++A^-)\cos(\phi(x))\cos(\phi(t)) + (A^--A^+)\sin(\phi(x))\sin(\phi(t))) \quad (11)$$

with $$\phi(x) = f(x_n^a - x) - \pi n/2 + \phi^x \quad (12)$$

$$\phi(t) = \omega(t-t') + \phi(f,\omega) + \phi^t \approx \omega(t-t'-t_p) + \phi^t \quad (13)$$

$$(A^+, A^-, \phi^x, \phi^t) = \begin{cases} (A_e^+, A_e^-, \phi^x, \phi_e^t) & \text{if } n \text{ is even} \\ (A_e^-, A_e^+, \phi^x, \phi_o^t) & \text{if } n \text{ is odd} \end{cases} \quad (14)$$

where $K(f,\omega) \equiv R^{-1/2}(f,\omega)$ denotes the spatio-temporal sensitivity of the filters, $(A^+)^2 + (A^-)^2 = 1$, $x_n^a = (N/N^a)n$ or $x_n^a = (N/N^a)(n + n \bmod 2)$ is the RF center[6] in the unit of the visual input grid size, and $t_p > 0$ approximates the filter latency which is determined by $\phi(f,\omega)$. The five parameters $(A_e^+, A_e^-, \phi^x, \phi_e^t, \phi_o^t)$ specify the RFs for all neural units $n = 1, 2, ..., N^a$, and different choices of them give different, but equivalently efficient, coding representations. The RF centers $x_n^a$ of the neural units $n = 1, 2, ..., N^a$ in this scale $a$ are distributed over the whole visual space $x \in (0, N)$ with the Nyquist sampling rate — the number of neurons $N^a$ in this scale is proportional to the bandwidth $f^{a+1} - f^a$, with two neurons covering every two sampling periods $2N/N^a$ (Figure 1).

Let us examine $K_n^a(x;t-t')$, the spatiotemporal RF of the $n^{th}$ unit in the $a^{th}$ scale. It is selective to spatial frequencies $f \in (f^a, f^{a+1})$ and all temporal frequencies $\omega$ with a sensitivity proportional to $K(f,\omega)$. The spatial and temporal part of the filter are embodied in $(\cos(\phi(x)), \sin(\phi(x)))$ and $(\cos(\phi(t)), \sin(\phi(t)))$, respectively. When $x = x_n^a$, the RF spatial phase is $\phi(x) = -\pi n/2 + \phi^x$ for all $f$ — phase coherence — implying $x_n^a$ as the RF center. The RF amplitude reaches its peak at $x = x_n^a$ and quickly decays as $x$ moves away from it. This phase coherence and a finite bandwidth $f \in (f^a, f^{a+1})$ ensures the filter locality with a spatial spread $\Delta x \sim 1/(f^{a+1} - f^a)$ by the uncertainty principle. Similarly, the temporal phase coherence, $\phi(t) \approx$ constant for all $\omega$, is achieved at $t - t' = t_p$, the temporal latency, as implied by the temporal locality of the filter[7]. Translation invariance is achieved since the RFs are the same for every second neural units $n$ and $n+2$, $K_n^a(x;t-t') = -K_{n+2}^a(x - (x_{n+2}^a - x_n^a);t-t')$, except for a shift in the RF centers and up to a polarity change. (It is not possible to have

[6] Here both choices, $x_n^a = (N/N^a)n$ and $x_n^a = (N/N^a)(n + n \bmod 2)$, are valid. In Li and Atick 1994a however, only the first choice is given.

[7] As we stated earlier, $\phi(f,\omega)$ is chosen to make $K_t^f(t)$ causal and have minimum temporal spread, implying the temporal coherence $\omega\tau_p + \phi(f,\omega) \approx$ *constant* for all $\omega$ given a $f$. A change $\phi(f,\omega) \to \phi(f,\omega) + \alpha$ still satisfies the requirement; and $\phi(f,\omega) \to \phi(f,\omega) - \omega\tau$ for $\tau > 0$ merely prolongs the filter latency $\tau_p \to \tau_p + \tau$. Although the minimum latency $\tau_p = \tau_p^{min}$ depends on $f$, it is possible to choose $t_p$ as the largest $\tau_p^{min}$ within a limited band $(f^a, f^{a+1})$, such that $fx_p + \omega t_p + \phi(f,\omega) \approx$ *constant* can be satisfied for $x_p = 0$ or any $x_p$ without compromising causality. Since $K(f,\omega)$ varies very little within the limited band (see later), $t_p$ can be very close to the shortest latency for every spatial frequency component $f$ in the band. Similar temporal phase structures have been observed in experiments (Hamilton et al 1989).

the same RF for every neuron in an efficient code when the spatial frequency bandwidth in the cortex is larger than one octave (Li and Atick 1994a)). Note that a drifting grating $\cos(fx + \text{spatial phase} \pm (\omega t + \text{temporal phase}))$ has a drifting velocity $v = \pm\omega/f$. Our neurons then respond to the two motion directions with relative amplitudes $A^+$ and $A^-$, respectively, and have a directional index $D.I. \equiv \frac{||A^+|-|A^-||}{|A^+|+|A^-|}$.

Variations of $(A_e^{\pm}, \phi^x, \phi_e^t, \phi_o^t)$ generate a whole family of spatiotemporal RFs of various directionality and RF phases. At one extreme when $A^+ = A^-$:

$$\mathbf{K_n^a}(x; t-t') \propto \sum_{f^a < f \le f^{a+1}} \int_0^\infty d\omega K(f,\omega)\cos(\phi(x))\cos(\phi(t)) \approx K_n^a(x) K^a(t-t') \quad (15)$$

$$\text{where} \quad \begin{array}{rcl} K^a(t-t') & \equiv & \int_0^\infty d\omega K(f^{peak},\omega)\cos(\phi(t)) \\ K_n^a(x) & \equiv & \sum_{f^a < f \le f^{a+1}} \cos(\phi(x)) \end{array}$$

where $f^{peak} = \sqrt{f^a f^{a+1}}$. This neuron is not at all directionally selective, as intuitively expected from equation (15) which approximates the filters as spatiotemporally separable[8]. From here on, such filters will be viewed as separable. The other extreme case is when $A^+ = 1$ and $A^- = 0$:

$$\begin{aligned} \mathbf{K_n^a}(x; t-t') &\propto \sum_{f^a < f \le f^{a+1}} \int_0^\infty d\omega K(f,\omega)(\cos(\phi(x))\cos(\phi(t)) - \sin(\phi(x))\sin(\phi(t))) \\ &= \sum_{f^a < f \le f^{a+1}} \int_0^\infty d\omega K(f,\omega)\cos(\phi(x)+\phi(t)) \end{aligned}$$

which is selective to only one motion direction. This neuron has a spatiotemporally inseparable RF composed of a pair of spatio-temporal separable filters, $\sum_{f^a < f \le f^{a+1}} \int_0^\infty d\omega K(f,\omega)$ $\cos(\phi(x))\cos(\phi(t))$ and $\sum_{f^a < f \le f^{a+1}} \int_0^\infty d\omega K(f,\omega)\sin(\phi(x))\sin(\phi(t))$, with quadrature phase relationship between the two components in both space and time dimensions. Such neurons have been proposed as the first stage components in the motion models of Adelson and Bergen (1985) and Watson and Ahumada (1985). All degrees of directionality are possible as $A^+/A^-$ changes. The choice of $\phi^x$ on the other hand dictates the spatial phase $\phi_n(x = x_n^a) = -\pi n/2 + \phi^x$ of the RF, e.g., $\phi^x = 0$ gives even or odd RF (bar or edge detectors) depending on $n$; While the value of $\phi^t$ gives the RF temporal phases at the peak of the temporal response $t = t' + t_p$.

In contrast to other motion models, the current work in addition predicts a precise relationship between the RFs of the neighboring neural units (i.e., the $n^{th}$ and $(n+1)^{th}$ units, see Fig. 1):

$$(A^+/A^-)_n = (A^-/A^+)_{n+1} \quad \text{opposite motion direction preferences} \quad (16)$$

$$(D.I.)_n = (D.I.)_{n+1} \quad \text{same directional index} \quad (17)$$

$$\phi_n(x_n^a) = \phi_{n+1}(x_{n+1}^a) + \pi/2 \quad \text{quadrature relationship between spatial phases} \quad (18)$$

(see equations (14) and (12)). When testing this neighbor relationship in the cortex, one should choose neurons that (1) have the same optimal spatial frequency since they belong to the same scale $a$, (2) are tuned to the same orientation as implied by our one-dimensional mathematical

[8]This approximation is valid when $K(f,\omega)$ is a smooth function of $f$ and changes little within a limited frequency range $(f^a, f^{a+1})$ (see next section).

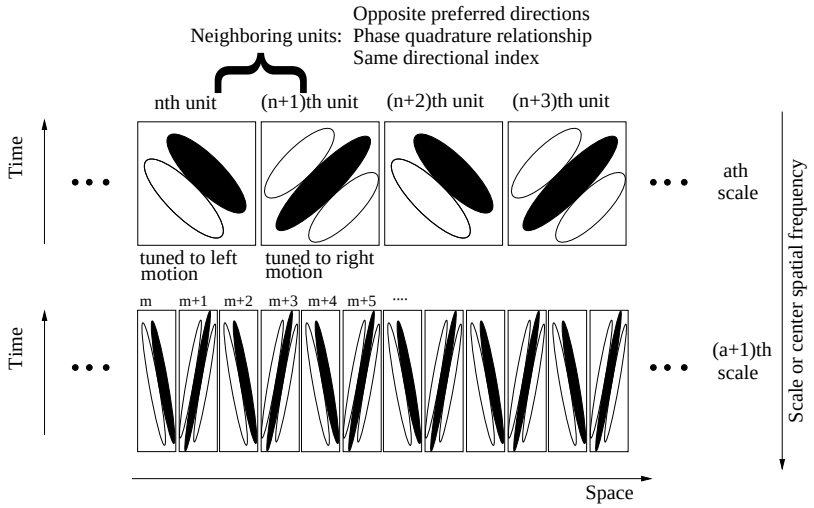


Figure 1: Schematic illustration of the efficient motion coding in the multiscale representation. The RFs for the two neighboring scales are shown. The space and time are in horizontal and vertical directions respectively. A perfectly oriented bar or edge in space-time implies complete cell directionality, as is used in this figure for example. The slope and sign of the orientation correspond to the preferred speed and direction of motion. Note that (1) the neighboring units have the quadrature phase relationship, opposite preferred directions of motion, and the same directional index, (2) the preferred motion speed decreases as the cell RF size decreases (see section 3). The RF centers of the neighboring units are displaced by a distance comparable to the RF sizes, this displacement is exaggerated in the figure to avoid RF overlap for clear illustration.

treatment, (3) have RF centers $x_n^a = (N/N^a)n$ or $(N/N^a)(n + n \mod 2)$ which are displaced by $x_{n+1}^a - x_n^a = 0$, or $N/N^a$, or $2N/N^a$, comparable to that of the efficient Nyquist sampling period $N/N^a$, which is roughly half a grating period of the optimal spatial frequency. Hence one should distinguish the *neighboring units* mentioned here from the anatomical neighboring cells in the cortex, which could be tuned to different optimal spatial frequencies and orientations, etc. (see discussion). Furthermore, the two observed cells in experiments may be neighboring units ($n^{th}$ and $n + 1^{th}$ units) *as well as* second neighboring units ($n^{th}$ and $(n + 2)^{th}$ units) to each other, if their RF center displacements are not carefully monitored. In such cases, the two cells' directional preferences are likely to be the same as well as opposite. This is observed physiologically, where neighboring cells tend to prefer the same, and sometimes opposite, but fewer times orthogonal, directions of motion (Berman, Wilkes, and Payne 1987). It is desirable to test whether those preferring opposite directions and having quadrature phase relationships also have the same directional index. In addition, one has the following observation from the theory. Given any scale and orientation, there is no need to have *two* filters tuned to opposite directions at *each* spatial location in order to have a complete representation. An efficient code needs an average of only one filter tuned to one direction per sampling interval — a pair of quadrature filters tuned to opposite directions for every two Nyquist sampling intervals (see discussion).

To illustrate the spatiotemporal RF, we need the knowledge of $K(f, \omega)$, which depends

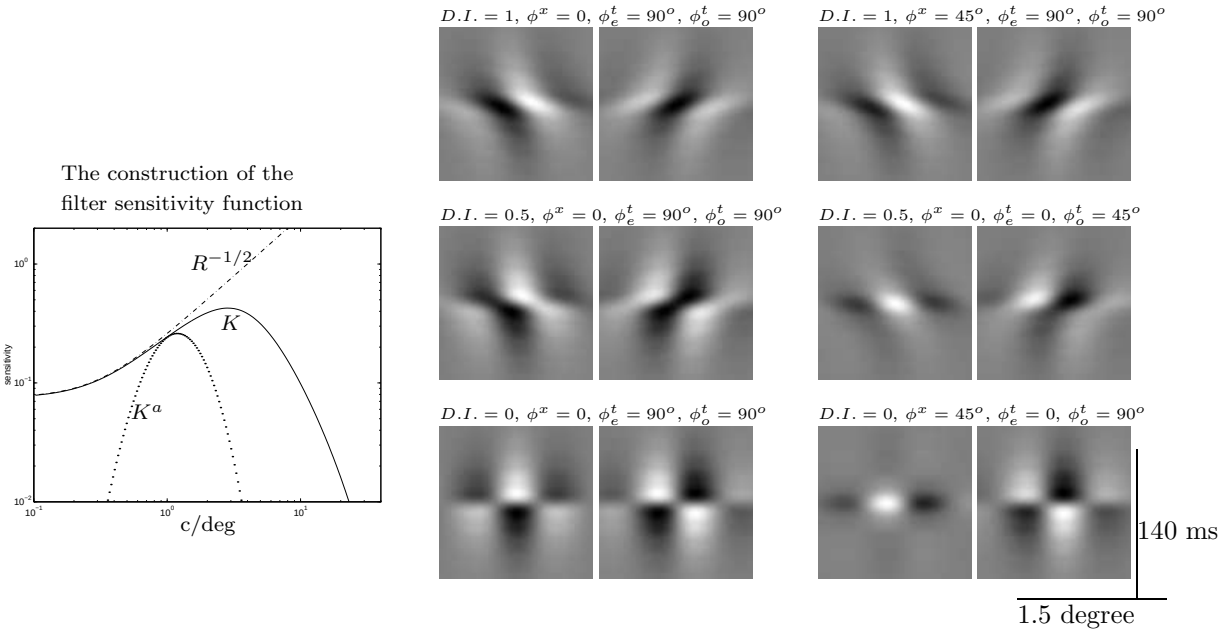


Figure 2: On the left is the filter sensitivity $K$ which deviates from the noiseless case $K_{noiseless} = R^{-1/2}$ at high frequency, where the signal $R$ is weak, in order to smooth out noise. $K^a$ is the sensitivity for scale $a$ centered around $f = 1$ c/deg. The temporal dimension is ignored for clarity in the plot. On the right are RF examples for pairs of neighboring units, next to each other with the left one as the even unit of the pair, each under the parameter value set $(D.I., \phi^x, \phi^t_e, \phi^t_o)$ which generates them. The space and time are in horizontal and vertical directions respectively, and each RF is centered at the RF center $(x,t) = (x^a_n, t' + t_p)$. The gray levels depict the filter amplitudes, gray for near zero amplitudes, white and black for larger positive and negative amplitudes respectively. The preferred spatial frequency is $f^{peak} = 1$ c/deg. A perfectly oriented bar or edge in space-time implies complete directionality, and a spatiotemporal separability implies non-directionality. Note (1) the neighboring units have the quadrature phase relationship, opposite preferred directions, and the same directional index D.I., (2) changes in the RFs as $D.I.$ decreases in left column, (3) differences in RF phases between left and right column pairs of the same directionality. These RFs are obtained by the approximation $\mathbf{K^a_n}(x,t) \approx (A^+ + A^-)K_x(x)K_t(t) + (A^- - A^+)\tilde{K}_x(x)\tilde{K}_t(t)$, where $K_t(t) \propto \int_0^\infty dw K^a(f^{peak},\omega)\cos(\phi(t))$, $\tilde{K}_t(t) \propto \int_0^\infty dw K^a(f^{peak},\omega)\sin(\phi(t))$, $K_x(x) \propto \int_0^\infty df K^a(f,\omega^{peak})\cos(\phi(x))$, $\tilde{K}_x(x) \propto \int_0^\infty df K^a(f,\omega^{peak})\sin(\phi(x))$. This approximation and figure format are used in other figures of this paper as well.

on $R(f,\omega)$ and should be modified under noise. Different measurements have suggested that in natural scenes $R(f,\omega=0) \propto 1/f^2$ (Field 1987) and $R(f=0,\omega) \propto 1/\omega^2$ (Dong and Atick 1994). Without additional knowledge, this paper models $R(f,\omega) \propto (f^2 + \xi^2\omega^2)^{-1}$, where $\xi = 0.4$ cycle·second/degree is chosen to give a final contrast sensitivity $K(f,\omega)$ peaking around 8 Hz for low $f$ as observed psychophysically (Nakayama 1985). As one will see below, the qualitative results in this paper depend only on $R(f,\omega)$ decaying with increasing $f$ and $\omega$. Hence the exact $R(f,\omega)$ or $\xi$ is not crucial. The complete decorrelation requires $K(f,\omega) = (R(f,\omega))^{-1/2}$ (see equation (10)), which increases with $(f,\omega)$ to amplify the lower signal power $R(f,\omega)$ at higher $(f,\omega)$. This leads to undesirable noise amplification when at high $(f,\omega)$ the weak signal $R(f,\omega)$ is overwhelmed by noise $R_N$, which is assumed to be white noise and therefore $R_N =$ constant over $(f,\omega)$. A noise smoothing strategy[9] is employed to lower $K(f,\omega)$ whenever the

[9]Noise smoothing gives $K(f,\omega) \approx M(f,\omega)K(f,\omega)|_{noiseless}$ (following Atick and Redlich 1992, Li and Atick 1994b, and noise smoothing also follows from information theoretical arguments), where $M \propto R(f,\omega)/(R(f,\omega)+$

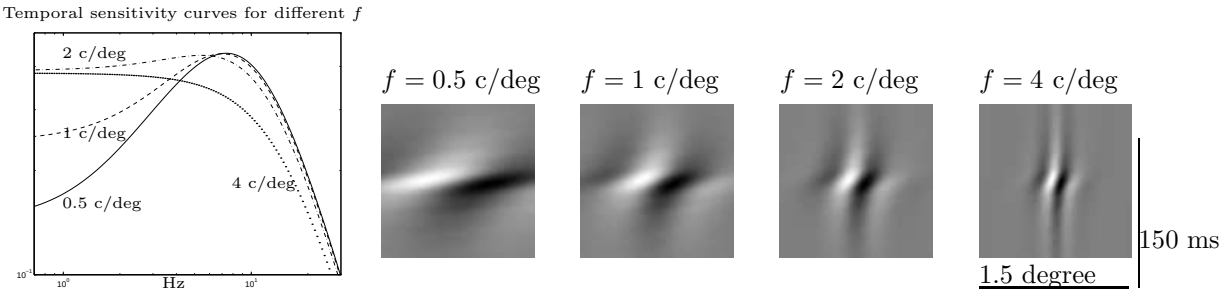


Figure 3: Changes of temporal sensitivities and spatiotemporal receptive fields with the optimal spatial frequency $f$. The filter orientation in space-time has a steeper slope as $f$ increases, implying decreasing preferred motion speeds. Parameters used: $D.I. = 1$, $\phi^x = 0$, and $\phi^t = 90^o$.

signal-to-noise $R(f,\omega)/R_N$ is small. The generic feature of $K(f,\omega)$ is (Figure 2):

$$K(f,\omega) \quad \text{increases with } f,\omega \text{ when } R(f,\omega) \gg R_N \text{ at small } (f,\omega) \tag{19}$$

$$K(f,\omega) \quad \text{decreases with } f,\omega \text{ when } R(f,\omega)/R_N \text{ is small at large } (f,\omega) \tag{20}$$

$K(f,\omega)$ peaks at some intermediate $(f,\omega)$, where the signal $R(f,\omega)$ starts to be overwhelmed by noise. Hence if $R = S^2/(f^2+\xi^2\omega^2)$, then $K(f,\omega)$ peaks at lower $(f,\omega)$ for smaller $S^2$. In the multiscale representation, we further model (Figure 2) $\sum_{f^a<|f|\le f^{a+1}} K(f,\omega)$ by $\int_0^\infty df K^a(f,\omega)$, where $K^a(f,\omega) = K(f,\omega)\exp(-(\log(f/f^{peak})/\sigma)^2/2)$ and $\sigma = \log(\sqrt{3})$ is to model a 1.6 octave bandwidth (Li and Atick 1994a) of the frequency selective channel with optimal frequency $f^{peak} = \sqrt{f^a f^{a+1}}$. Figure 2 illustrates some examples of the spatio-temporal RFs of neighboring cells using these models.

## 3 Correlation between motion coding and visual codings in space, color, and stereo

We explore additional predictions from the motion coding theory to compare them with experimental observations or subject them to experimental tests. This can be carried out by studying the correlations between motion coding and codings in space, color, and stereo. It was shown in section 2 that for signal power $R = S^2/(f^2+\xi^2\omega^2)$, the peak sensitivity $K(f^{peak},\omega^{peak})$ will occur at a lower frequency $(f^{peak},\omega^{peak})$ when the signal power $R$, or $S^2$, is smaller. In particular, the temporal sensitivity curve $K_t^f(\omega) \equiv K(f,\omega)$ for each spatial frequency $f$ also peaks at some $\omega = \omega^{peak}(f)$. Hence $\omega^{peak}(f)$ decreases as $S^2$ decreases or $f$ increases. This has immediate consequences on cross-channel coding correlations when one notices that the signal power magnitude $R$ depends on the frequency $f$, on whether the signal is achromatic or ocularly opponent, etc., as will be shown below.

*Correlation between spatial coding and motion coding*

This theory thus predicts that the cell optimal speed decreases with increasing optimal spatial frequency (see Figure 3), as observed in experiments (Holub and Morton-Gibson 1981, Foster et al 1985). This is because for a neuron with optimal spatiotemporal frequency $(f^{peak},\omega^{peak})$, the preferred motion speed is roughly $v \sim \omega^{peak}/f^{peak}$. The prediction follows since both $1/f^{peak}$ and, from the argument above, $\omega^{peak}$, decrease with increasing $f$. The model $R(f,\omega) = S^2/(f^2+\xi^2\omega^2)$ gives a slowly decreasing or roughly constant $\omega^{peak}(f)$ for a range of low spatial frequencies $f$ (Fig. 3), suggesting a roughly inverse relationship $v \sim 1/f^{peak}$. At a

$R_N$) is a low pass smoothing filter. In detail, what is used in the paper are: $M(f,\omega) = (R/(R+1))\exp[-(f/f_c)^{1.4}]$, $K(f,\omega) \propto M(M^2(R+1)+1)^{-1/2}$, $R = 16.0/(f^2+\xi\omega^2+f_\nu^2)$, $f_\nu = 0.3$ c/deg, $f_c = 22$ c/deg.

higher $f$, whose exact value depends on the signal-to-noise or $S^2$, $\omega^{peak}(f)$ starts to decrease sharply with $f$, and temporal sensitivity $K_t^f(\omega)$ becomes significantly low-pass and $v$ approaches zero. The same trend of $\omega^{peak}(f)$ is observed physiologically (Holub and Morton-Gibson 1981) and psychophysically (Kelly 1979). The physiologically measured $\omega^{peak}$ varies from cell to cell by up to a factor around 10 for a given cell optimal $f$ (Holub and Morton-Gibson 1981, Foster et al 1985). Such variations can not be accounted for by the present theory, and may serve other computational purposes, e.g., Grzywacz and Yuille (1990) have used them for velocity computation. However, cortical cells have a wide, around 3 octaves, temporal frequency bandwidth (Holub and Morton-Gibson 1981, Foster et al 1985). This width is comparable to, and likely contributed to, the measured spread in $\omega^{peak}(f)$.

Another prediction is:

$$\frac{\text{best sensitivity to contrast reversal grating}}{\text{sensitivity to drifting grating of preferred direction}} \begin{cases} = 1 & \text{if } D.I. = 0 \\ \rightarrow 0.5 & \text{as } D.I. \rightarrow 1 \end{cases} \tag{21}$$

This stems from equations (10) and (11), which suggest gains of $\propto 1/2(|A^+| + |A^-|)$ and $\propto A^{\pm}$, respectively, to the two grating types. Psychophysically, the detection threshold for counter-phased gratings is almost twice of that for drifting gratings over a wide spatio-temporal frequency range (Levinson and Sekular 1975, Watson, Thompson, Murphy, and Nachmias 1980). These observations were explained by noting that two half-contrast drifting gratings of opposite directions sum to a full contrast reversal grating (see Burr 1991). The current prediction, however, is on a single cell level and relies on the assumed linear mechanisms. Significant cortical non-linearity (Reid et al 1991) should give a quantitatively different reality, however, but the trend of decreasing ratio above with increasing directional index should still hold and can be tested.

*Correlation between color and motion coding*

This theory also predicts a smaller optimal speed for the chromatic channel (see Figure 4), since the chromatic signal power (hence $\omega^{peak}_{chromatic}(f)$) is smaller than luminance signal power (or $\omega^{peak}_{luminance}(f)$). This is consistent with the observation that the perceived motion slows down dramatically at isoluminance (Cavanagh, Tyler, and Favreau (1984)). The color channel is traditionally viewed as insensitive to motion (see Nakayama 1985). However, there are recent psychophysical and physiological evidences of chromatic contribution to motion detection (Dobkins and Alright 1992, 1994). At a single striate cortical cell level, chromatic and luminance signals are multiplexed (see Li and Atick 1994a). Accordingly, the actual motion sensitivity in a single color selective cell is complicated, and should depend on whether stimuli is isoluminant or not.

*Correlation between stereo and motion coding*

Stereo coding (Li and Atick 1994b, Li 1995) is composed of ocular summation (the input summation from the two eyes) and ocular opponency (the input difference between the two eyes) channels. Let $\mathbf{K_{sum}}(x,t)$ and $\mathbf{K_{opp}}(x,t)$ be the RFs for the summation and opponency channels, respectively. We have

$$\mathbf{K_c}(x,t) \propto \int_0^\infty df \int_0^\infty d\omega K_c(f,\omega)(A_c^+ \cos(fx + \omega t + \phi_c^+) + A_c^- \cos(fx - \omega t + \phi_c^-)) \tag{22}$$

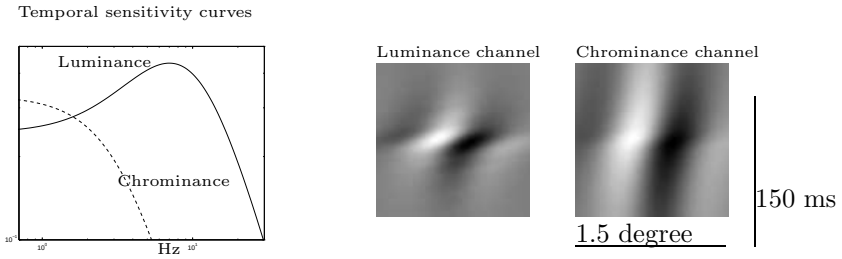


Figure 4: Temporal sensitivity and spatiotemporal RFs for luminance and chrominance channels. Parameters used $D.I. = 1$, $\phi^x = 0$, and $\phi^t = 90^o$, $f^{peak} = 1$c/deg, and the signal power in the chromatic channel is 4% of that in the luminance channel. A smaller optimal motion speed in the chromatic channel is apparent.

for $c = sum, opp$. Here all phase contributions, e.g., $fx_n^a$, and $\phi(f,\omega)$ (see equation (10), that do not depend on $(x,t)$ are summed into variables $\phi^{\pm}$. (The subscript $n$ for the neuron and superscript $a$ for scale are omitted for clarity). The binocular RFs in a cortical cell are (Li and Atick 1994b, Li 1995) $\mathbf{K_l}(x,t) = \mathbf{K_{sum}}(x,t) + \mathbf{K_{opp}}(x,t)$ for the left and $\mathbf{K_r}(x,t) = \mathbf{K_{sum}}(x,t) - \mathbf{K_{opp}}(x,t)$ for the right eyes:

$$\mathbf{K_{eye}}(x,t) = \int_0^\infty df \int_0^\infty \quad d\omega \quad (K_{eye}^+(f,\omega)\cos(fx + \omega t + \phi_{eye}^+) \tag{23}$$

$$+K_{eye}^-(f,\omega)\cos(fx - \omega t + \phi_{eye}^-)) \tag{24}$$

with $eye = l, r$. Here $K_{eye}^+(f,\omega)$ and $K_{eye}^-(f,\omega)$ are the monocular sensitivities to stimuli of opposite motion directions. The questions are: what are the directionality and the optimal speed $v_{eye}^{\pm} \sim \omega_{eye}^{\pm peak}/f_{eye}^{\pm peak}$ for each eye, and how do they depend on the ocular origin and the motion direction? Here $(f_{eye}^{\pm peak}, \omega_{eye}^{\pm peak})$ is the frequency to reach peak sensitivity in the curve $K_{eye}^{\pm}(f,\omega)$.

Let $\mathbf{A_c^{\pm}} = A_c^{\pm} e^{i\phi_c^{\pm}}$ and $\mathbf{K_{eye}^{\pm}}(f,\omega) = K_{eye}^{\pm}(f,\omega)e^{i\phi_{eye}^{\pm}}$, we have

$$\begin{aligned} \mathbf{K_l^{\pm}}(f,\omega) &= K_{sum}(f,\omega)\mathbf{A_{sum}^{\pm}} + K_{opp}(f,\omega)\mathbf{A_{opp}^{\pm}} \\ \mathbf{K_r^{\pm}}(f,\omega) &= K_{sum}(f,\omega)\mathbf{A_{sum}^{\pm}} - K_{opp}(f,\omega)\mathbf{A_{opp}^{\pm}} \end{aligned} \tag{25}$$

The signal power for ocular summation and opponency are $R_{sum} = (1+r)R(f,\omega)$ and $R_{opp} = (1-r)R(f,\omega)$, respectively, where $0 < r < 1$ is the input ocular correlation normalized by the self-correlation within a single eye[10]. The inequality $R_{sum} > R_{opp}$ immediately gives a larger optimal speed in the summation channel $v_{sum} > v_{opp}$, and in addition, the channel sensitivities $K_{sum}(f,\omega)$ and $K_{opp}(f,\omega)$ should differ and they are not simply related by a gain factor: $K_{sum}(f,\omega) \not\propto K_{opp}(f,\omega)$. Consequently by equation (25), $K_{eye}^+(f,\omega) \not\propto K_{eye}^-(f,\omega)$ and $K_l^{\pm}(f,\omega) \not\propto K_r^{\pm}(f,\omega)$. This means, in a single neuron, the RFs for the two eyes can differ in detailed form as well as in overall sensitivity, and the contrast sensitivity curves $K_{eye}^{\pm}(f,\omega)$ of two directions also differ by more than a gain factor. Accordingly, this theory predicts (1) the

[10]The temporal dimension of visual inputs was ignored in the earlier works (Li and Atick 1994b, Li 1995) and the ocular signal powers were denoted as $(1 \pm r(f))R(f)$. Here we simply assume that the generalization $(1 \pm r(f))R(f) \to (1 \pm r(f,\omega))R(f,\omega)$ holds approximately.

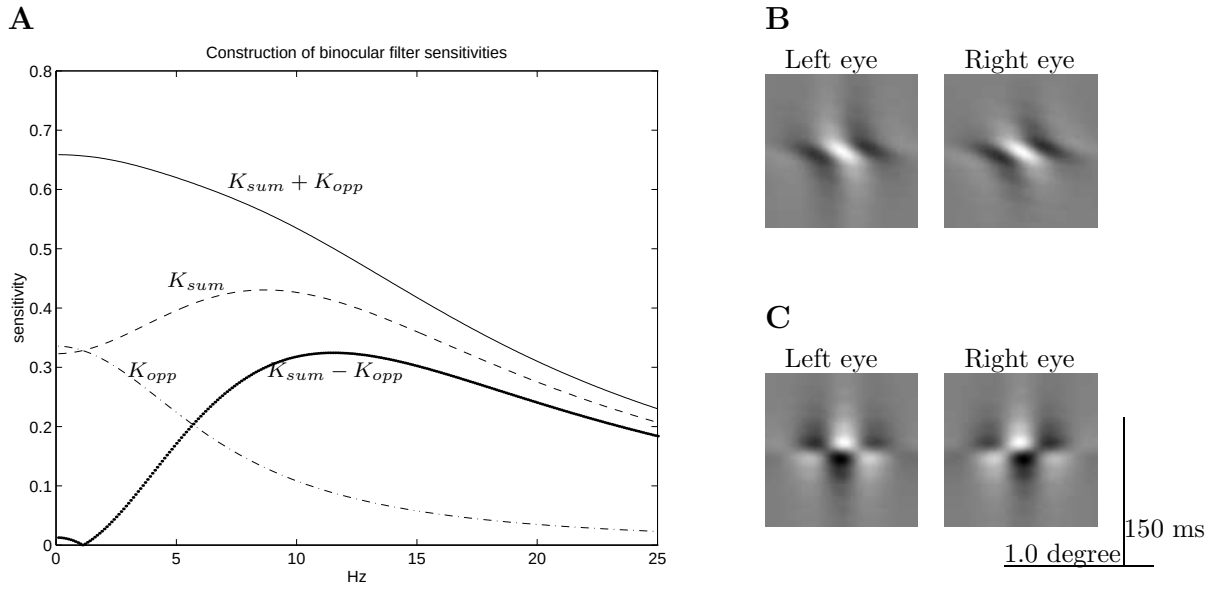


Figure 5: Interaction between motion and stereo coding. **A**: Temporal sensitivity functions for the ocular summation $K_{sum}$, ocular opponency $K_{opp}$, $K_{sum}+K_{opp}$, and $|K_{sum}-K_{opp}|$ channels for spatial frequency $f^{peak}=2$ cycles/degree, which is used in **B** and **C**. Here the binocular correlation used is $r(f^{peak}) = 0.96e^{-f^{peak}/(15c/deg)}$. **B**: An example of different preferred velocities for the two eyes (see text). **C**: An example of different preferred directions of motion for the two eyes (see text). It is not difficult to see that the optimal speeds for opposite directions of motion in the same eye are also different, and the ocular dominance changes for this neuron with motion directions.

optimal speed $v_{eye}^{\pm} \sim \omega_{eye}^{\pm peak}/f_{eye}^{\pm peak}$ can vary with eye origin *and* the motion direction; (2) some neurons change their preferred motion direction with ocular origin or change their ocular dominance with motion direction; (3) the directional index $\frac{||K_{eye}^{+}|-|K_{eye}^{-}||}{|K_{eye}^{+}|+|K_{eye}^{-}|}$ for monocular stimuli can vary with frequency $(f,\omega)$ of the drifting grating presented, as observed physiologically (Reid et al 1991), since $K_{eye}^{+}/K_{eye}^{-}$ is not a constant of $(f,\omega)$.

To illustrate the predictions, consider first the example when $\mathbf{A_{sum}^{+}} = \mathbf{A_{opp}^{+}} = \mathbf{A_{sum}^{-}} = -\mathbf{A_{opp}^{-}} = \mathbf{A}$ (Fig 5A, 5C). Then

$$\mathbf{K_l^{\pm}}(f,\omega) = \mathbf{A}(K_{sum}(f,\omega) \pm K_{opp}(f,\omega)) \quad (26)$$

$$\mathbf{K_r^{\pm}}(f,\omega) = \mathbf{A}(K_{sum}(f,\omega) \mp K_{opp}(f,\omega)) = \mathbf{K_l^{\mp}}(f,\omega) \quad (27)$$

Although both the summation and opponency channels are non-directional, this cell has a directional RF when considering either eye alone since $K_l^{+} > K_l^{-}$ and $K_r^{-} > K_r^{+}$, but the preferred direction changes with the eye. In addition, the ocular dominance changes with motion direction since $K_l^{+} > K_r^{+}$ (left-dominant) but $K_l^{-} < K_r^{-}$ (right-dominant) by equation (27), which implies that a direction change is equivalent to a ocular origin change for this cell. Furthermore, the optimal motion speed for the left eye for example is larger in the negative direction $v_l^{-} > v_l^{+}$. This is because the temporal sensitivity curve is a low-pass in the positive direction $K_l^{+}(\omega) = K_{sum}(\omega) + K_{opp}(\omega)$ but a band-pass in the negative direction $K_l^{-}(\omega) = K_{sum}(\omega) - K_{opp}(\omega)$ (Fig. 5A), giving $\omega_l^{-peak} > \omega_l^{+peak}$ although the preferred spatial frequencies $f \in (f^a, f^{a+1})$ are roughly the same for the two directions. Another example (Fig. 5A, 5B) is when $\mathbf{A_{sum}^{+}} = \mathbf{A_{opp}^{+}} = \mathbf{1}$ and $\mathbf{A_{sum}^{-}} = \mathbf{A_{opp}^{-}} = \mathbf{0}$. The the ocular summation/opponency, and hence the left/right eye, channels are completely directional. The monoc-

ular RFs have sensitivities $K^+_{l,r} = K_{sum} \pm K_{opp}$ by equation (25). This cell thus changes its optimal speed with ocular origin just as the cell in the previous example does with motion direction (within a single eye). The predicted ocular differences in preferred speeds and directions of motion have been observed physiologically (Beverley and Regan 1973, Poggio 1992, DeAngelis et al 1994), and such neurons can sense object motion in depth. The predicted changes in the monocular optimal speed with motion direction as well as the ocular dominance changes with direction can be experimentally tested.

## 4. Summary and Discussion

This paper demonstrates that efficient coding in the multiscale representation can account for many experimental observations of motion and directional sensitivity in simple cells of the primary visual cortex. A whole spectrum of neural directional indices and different degrees of RF spatiotemporal separability are predicted. In addition, the cortical motion coding is predicted to correlate with the codings in space, color, and stereo domain. Explicitly, the theory predicts that the cell preferred speeds decrease with their increasing optimal spatial frequencies, can differ for the two eyes in the same neuron, are much slower in the color sensitive channel, and that the two eyes in the same neuron can prefer opposite directions of motion. These predictions agree with physiological or psychophysical observations (Beverly & Regan 1973, Holub and Morton-Gibson 1981, Cavanagh et al 1984, Dobkins and Albright 1993, 94, Foster et al 1985, Reid et al 1991, Poggio 1992, DeAngelis et al 1994). Furthermore, the theory gives testable predictions that have not been experimentally investigated systematically. These predictions are: (1) if two nearby neurons prefer the same optimal spatial frequency, same orientation, and opposite motion directions, and have quadrature RF phase relationship, then they should have the same directional index; (2) a single neuron can have different optimal speeds for opposite directions of motion presented monocularly; and (3) a neuron's ocular dominance may change with motion direction when opposite directional preferences occur for inputs from different eyes.

A special class of predicted neurons by this theory resembles the linear units in the motion models by Adelson and Bergen (1985) and Watson and Ahumada (1985). While these computational models are constructed with the goal of motion or velocity computation within the constraints of physiology and psychophysics, the present theory derives from the efficient coding in the multiscale representation without *a priori* requiring motion sensing or computation. The efficient coding framework provides the following additional features not present in the previous models: (1) given spatial orientation and scale, a requirement of only *one pair* of phase quadrature filters preferring opposite directions for every *two* Nyquist sampling intervals in the visual field; (2) a mechanism relating RF properties to input signal powers, leading to additional predictions on the correlation between the motion coding and the spatial, chromatic, and stereo coding. The formulation by equation (11) is similar to the model by Hamilton, Albrecht, and Geisler (1989), except that the former is derived from efficient coding principles while the latter is constructed to fit the experimental data.

The current theory uses linear approximation for cortical coding mechanisms. The significant cortical non-linearity, such as those that facilitate the neural responses to preferred motion directions and inhibit the responses to non-preferred directions, as observed by Reid et al (1991), will lead to quantitative discrepancies between the theory and experiments. However, one can

make the following observations. First, this work focuses on efficiency by reducing redundancy between different neural units. Efficiency can be enhanced by using a proper non-linear transfer function at the single neuron level to achieve maximum information within a limited dynamic range, i.e., histogram equalization, as was done by Laughlin (1981). Such non-linearity would be within a single cell and may be similar to the action potential generation mechanism. It does not affect the cell's directional preference and the receptive field significantly, and is still within the goal of efficiency. Experimental evidence (Jagadeesh et al. 1993) also suggested that most of the non-linearity in simple cell motion selectivity originates from the action potential generation. Second, coding efficiency is always with respect to a particular visual environment, characterized by, e.g., signal-to-noise ratio or adaptation levels. To maintain efficiency, environmental changes should lead to coding changes which are necessarily of non-linear mechanisms, e.g., gain control or normalization (e.g., Heeger 1993), and may involve interactions between output neurons. The current work does not include such non-linearity because it focuses on *what* an efficient code should be but not *how* it is developed or adapted. Thirdly, the RF characterizes only the effective transform from visual inputs to cortical outputs, it does not exclude the possible contribution from the cortical feedback interactions (Douglas and Martin 1992) which could play a significant role in the actual receptive field construction. This argument is apparent in the linear approximation although the reality is most likely non-linear. Having said these, one should note that other visual functions beyond efficiency are likely to contribute to the cortical nonlinearity which can not be understood by the current framework.

A lack of precise knowledge of the natural input power spectrum in the temporal domain makes most theoretical predictions non-quantitative. In any case, the quantitative predictions would also depend on the signal-to-noise used in particular experiments.

This theory has further considerable limitations. The derivations in section 2 and Appendix implicitly assume that there are as many input units (retina ganglion cells) as output units (primary visual cortical cells). Under that assumption, an efficient representation (10) should have only one particular parameter set $(A_e^{\pm}, \phi^x, \phi_e^t, \phi_o^t)$, permitting only one directional index for all cells and two receptive field phase values in quadrature of each other (at least when considering cells preferring the same orientation and scale). In fact, when spatial and stereo codings are also included (Li and Atick 1994a,b), it then follows that there should be only two (orthogonal) choices of preferred orientation as well as one ocular dominance index and two optimal disparity values for each spatial scale and orientation. In reality, however, there are about 40 times as many cortical cells in V1 as retinal ganglion cells (Barlow 1981) and a spectrum of directional indices, preferred orientations and disparities, and ocular dominance indices in a single cortex (Hubel and Wiesel 1974, Berardi, Bisti, Cattaneo, Fiorentini, and Maffei 1982, Berman et al 1987). The reasons for the cortical cell proliferation and their extent are beyond the scope of the efficient coding theory. However, given a larger cortical cell population compared to that of the retinal ganglion cells, the efficient coding theory can be generalized and still applied. Essentially the same cell RF properties can be obtained, either in a statistical mechanics framework by Nadal and Parga (1993, and Nadal private communication 1995) or non-statistically (Li, in preparation). Briefly, when the output units are many times more numerous than the sensory input units, efficient coding will produce many different copies of the codes like the ones in section 2. Each copy carries less information than it would if the

cell population were smaller, and the representations in different copies are not decorrelated but the overall representation is still the most efficient given the larger population. However, different copies can have different code parameters (e.g.,$A_e^{\pm}, \phi^x, \phi_e^t, \phi_o^t$ if considering only the motion coding) and can thus generate a whole spectrum of RF properties observed in the cortex.

Many of the predictions from the efficient coding framework, such as the cell quadrature phase structures, the spatial frequency bandwidth, the color selective blob cells, and the correlation between spatial and stereo coding, some of which rely heavily on the efficient coding assumption, agree with experimental observations (see Li and Atick 1994ab and references there in). In addition, the theoretical framework has already provided predictions which had not been experimentally investigated and have been subsequently confirmed in experiments (Li 1995, Anzai et al 1994). These facts give credibility to efficient coding as a useful framework for understanding at least some of the primary visual cortical processings. The current work, with some of its predictions not yet experimentally investigated, provides more testing grounds to explore the strength and limitations of the efficient coding framework. In particular, the test on the prediction of neighboring motion sensitive units is crucial to the theory. This is because the confirmation of this prediction requires the neighboring cells to (1) have the same directional index if (2) they are in quadrature phase relationship, (3) have the same optimal spatial frequency and orientation and (4) prefer the opposite motion directions. To simultaneously satisfy these conditions would be difficult if the neural properties were randomly assigned. Note that conditions (2) and (3) are to reduce or eliminate the probability that the two neighboring cells might be from different efficient copies in the output cell population. This is because different efficient copies are likely to have different orientation preferences and center spatial frequencies, and in addition, the output neurons in different copies may be anatomically close but have no *a priori* reasons for a fixed relationship in their RF properties.

**Acknowledgement** I wish to thank Edward H. Adelson for very helpful discussions and comments on the draft, and Ning Qian, Christopher Kolb, Wyeth Bair, Joachim Braun, and the two referees for carefully reading the manuscript and very useful comments.

# References

[1] Adelson E. H. and Bergen J. R. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* Vol. 2, No. 2 1985. p. 284 - 299.

[2] Anzai, A., DeAngelis, G. C., Ohzawa, I., and Freeman R. D. (1994) Private communication

[3] Atick, J. J. and Redlich, A. N. Towards a theory of early visual processing. *Neural Comp.*, **2**, 308–320, 1990. also: What does the retina know about natural scenes? *Neural Comp.*, **4**, 196–210, 1992.

[4] Atick, J. J., Li, Z. and Redlich, A. N. 1992. Understanding retinal color coding from first principles. *Neural Comp.*, **4**, 559–572.

[5] Barlow, H.B. 1961. Possible principles underlying the transformation of sensory messages *Sensory Communication* ed W. A. Rosenblith (Cambridge, MA: MIT Press).

[6] Barlow H. B. 1981 The Ferrier lecture, 1980: Critical limiting factors in the design of the eye and visual cortex *Proceedings of the Royal Society, London* B, 212, 1-34.

[7] Berardi N., Bisti S., Cattaneo A., Fiorentini A., and Maffei L. Correlation between the preferred orientation and spatial frequency of neurones in visual areas 17 and 18 of the cat *J. Physiol.* (1982), 323. 603-618.

[8] Berman N. E. J., Wilkes M. E. and Payne B. R. Organization of orientation and direction selectivity in areas 17 and 18 of cat cerebral cortex *J. of Neurophysiol.* Vol. 58. No. 4. p676-699 1987.

[9] Beverley, K I. and Regan, D. Evidence for the existence of neural mechanisms selectively sensitive to the direction of movement in space *J. Physiol.* **235** 17-29, (1973).

[10] Bialek W., Ruderman, D. L., and Zee A. Optimal sampling of natural images: A design principle for the visual system? in *Advances in Neural Information Processing 3.* R. Lippmann, J. Moody, and D. Touretzky, eds. pp. 363-369, Morgan Kaufmann, San Mateo, CA, 1991.

[11] Burr D. C. Human sensitivity to flicker and motion, Vision and Visual Dysfunction Vol. 5: Limits of vision. Eds Kulikowski, J. J. Walsh V. and Murray I. J. The Macmillian Press 1991.

[12] Cavanagh P., Tyler C. W., Favreau O. E. Perceived velocity of moving chromatic gratings. *J. Optical Society America* 1, p. 893-899, 1984.

[13] DeAngelis G. C., Ohzawa I. and Freeman R. Neuronal mechanisms underlying stereopsis: how do simple cells in the visual cortex encode binocular disparity? *Perception* in press 1994.

[14] Dobkins K. R., Albright T. D. What happens if it changes color when it moves?: Psychophysical experiments on the nature of chromatic input to motion detectors. *Vision Res.* Vol. 33, No. 8, p-1019-1036, 1993, see also Dobkins K. R., Albright T. D. What happens if it changes color when it moves?: Neurophysiological experiments on the nature of chromatic input to macaque area MT. *J. Neurosci.* Vo. 14(8), p.4854-4870, 1994.

[15] Douglas, R. J. and Martin, K. A. C. Exploring cortical microcircuits: A combined anatomical, physiological, and computational approach. In *Single neuron computation* T. McKenna, J. Davis, and S. F. Zornetzer (eds.), 381-412, Academic Press, Orlando, Florida, USA.

[16] Dong D-W, and Atick, J.J. Temporal decorrelation: a theory of lagged and nonlagged cells in the lateral geniculate nucleus. Submitted for publication 1994.

[17] Field, D. J. 1987. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am.*, **A 4**, 2379–2394.

[18] Foster K, H, Gaska J. P., Nagler, M., Pollen D. A., Spatial and Temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the Macaque monkey *J. Physiol.* (1985) **365** 331-363.

[19] Grzywacz N. M. and Yuille A. L. A model for the estimate of local image velocity by cells in the visual cortex *Proc. R. Soc. Lond.* B 239, 129-161 (1990).

[20] Heeger, D. J. Modeling simple cell direction selectivity with normalized, half-squared, linear operators. *J. Neurophysiol.* 70, 1885-1898, 1993.

[21] Holub R. A. and Morton-Gibson M. Response of Visual Cortical neurons of the cats to moving sinusoidal gratings: response-contrast functions and spatiotemporal interactions. it J. Neurophysiol. **46** 1981 1244- 1259

[22] Hubel, D. H. and Wiesel, T. N. Receptive fields of single neurones in the cat's visual cortex. *J. Physiol. Lond.* 148: 574-591, 1959. and, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. Lond* 160:106-154, 1962.

[23] Hubel, D. H. and Wiesel, T. N. Uniformity of monkey striate cortex: a parallel relationship between field size, scatter, and magnification factor. *J. Comp. Neurol* 158: 295-305, 1974.

[24] Jagadeesh, B, Wheat H. S., Ferster D. Linearity of summation of synaptic potentials underlying direction selectivity in simple cells of the cat visual cortex *Science* 262:1901-1904 (1993).

[25] Kelly D. H. (1979) Motion and vision II. Stabilized spatio-temporal threshold surface. *J. Opt. Soc. Am.* 69, 1340-1349.

[26] Laughlin S. B. A simple coding procedure enhances a neuron's information capacity. *Z. Naturf.* 36c. 910-912, 1981.

[27] Levinson, E. and Sekuler, R. (1975) The independence of channels in human vision selective for direction of movement. *J. Physiol.* 250, p 347-366.

[28] Li, Zhaoping and Atick, J. J. 1994a. Towards a theory of the striate cortex. *Neural Comp.*, **6**, 127-146, (1994a).

[29] Li, Zhaoping and Atick, J. J. 1994b "Efficient stereo coding in the multiscale representation" *Network* Vol.5 1-18. (1994b)

[30] Li, Zhaoping "Understanding ocular dominance development from binocular input statistics", *The neurobiology of computation* (Proceeding of computational neuroscience conference 1994), P. 397-402. Ed. J. Bower, Kluwer Academic Publishers, 1995.

[31] Linsker, R. 1989 An application of the principle of maximum information preservation to linear systems, in *Advances in Neural Information Processing 1.* D. Touretzky, ed., pp. 186-194. Morgan Kaufmann, San Mateo CA.

[32] Marr D. and Ullman S. 1981, Directional selectivity and its use in early visual processing. *Proc. R. Soc. Lond. B* 211, 151- 180.

[33] Nadal, J.-P., Parga, N. Information processing by a perceptron in an unsupervised learning task. *Network: Computation in Neural Systems* Vol. 4, no. 3. p295-312. 1993.

[34] Nakayama K. Biological image motion processing: A review *Vision Res.* Vol. 25, No. 5, pp 625-660. 1985.

[35] Poggio, G. F. 1992. Physiological basis of ste reoscopic vision in J. R. Cronly-Dillon Ed. *Vision and Visual Dysfunction*. Vol. 9 *Binocular Vision* E. Reg an Ed. CRC Press, Inc.

[36] Reichardt W., Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In *Sensory Communication* ed. Rosenblity W. A., Wiley, New York, 1961.

[37] Reid R. C. , Soodak R. E., and Shapley R. M. Directional selectivity and spatiotemporal structure of receptive fields of simple cells in cat striate cortex. *J. Neurophysiol.* Vol 66 No. 6, 1991 p.505- 529.

[38] Srinivasan M. V., Laughlin S. B., and Dubs A. Predictive coding: a fresh view of inhibition in the retina *Proc. R. Soc. Lond.* B. 216, 427-459, 1982.

[39] Torre, V., and Poggio T. 1978 A synaptic mechanism possibly underlying directional selectivity to motion. *Proc. R. Soc. Lond. B* 202, 409-416.

[40] van Santen J. P. H. and Sperling G. 1984 A temporal covariance model of human motion perception. *J. Opt. Soc. Am. A.* 1, 451-473.

[41] Watson A. B. and Ahumada Al. Model of human visual-motion sensing *J. Opt. Soc. Am. A* Vol. 2 No. 2 1985. p. 322- 341.

[42] Watson A. B., Thompson, P. G., Murphy, B, J. and Nachmias, J. (1980) Summation and discrimination of gratings moving in opposite directions Vision Res. 20, 341-347.

## Appendix

This appendix is to show that equations (10)- (14) depict the general spatiotemporal RFs in a multiscale linear efficient code that is translation invariant, temporally causal, and spatiotemporally local (i.e., the RFs have finite and minimum spatiotemporal span to ensure finite synaptic connection length, retinotopy, and minimum delay in information extraction.)

It was shown (Li and Atick 1994a) that the unitary matrix required to combine the non-local filters in a spatial frequency band in equation (9) is

$$U^a_{nj} = \begin{cases} \frac{1}{\sqrt{N^a}} e^{i(f_j x^a_n - \pi n/2 + \phi)} & \text{if } f_j > 0 \\ \\ \frac{1}{\sqrt{N^a}} e^{-i(|f_j| x^a_n - \pi n/2 + \phi)} & \text{if } f_j < 0 \end{cases} \tag{28}$$

for $n = 1, 2, ...., N^a$ and arbitrary $\phi$. Combining (28), (9), (3), (4), and (5), we have

$$\mathbf{K^a_n}(x, t-t') \propto \sum_{f^a<f\leq f^{a+1}} \int_0^\infty d\omega K(f,\omega)\cos(f(x^a_n - x) - \pi n/2 + \phi)\cos(\omega(t-t') + \phi(f,\omega)) \quad (29)$$

This is exactly the spatiotemporally separable filters in equation (11) when $A^+ = A^-$. This RF is the most local spatially as implied by the phase coherence at $x = x^a_n$ and a finite bandwidth $f \in (f^a, f^{a+1})$. There is translation invariance $\mathbf{K^a_n}(x, t-t') = -\mathbf{K^a_{n+2}}(x - (x^a_{n+2} - x^a_n), t-t')$ between every second unit, and quadrature RF phase relationship between neighbors $\mathbf{K^a_n}$ and $\mathbf{K^a_{n+1}}$. These RF similarities give the best translation invariance possible in a scale of more than 1 octave bandwidth in the cortex (Li and Atick 1994a). The RF centers $x^a_n = (N/N^a)n$ (or $x^a_n = (N/N^a)(n + n \bmod 2)$), for $n = 1, 2, ..., N^a$, are distributed over the input visual field $x \in (0, N)$.

To obtain the general RF, we note that any changes in spatial phase $\phi$ in equation (29) and the temporal phase $\phi(f,\omega) \to \phi(f,\omega) + \beta$ for any $\beta$ will not compromise efficiency, spatial locality, translation invariance, causality, and the minimum temporal latency and spread. A code of such kind is denoted by $K^a_n(\cdot|\phi,\beta)$. Equation (9) states that a desired RF has to be a linear combination of the filters $K^{f_j}$ of equation (5)), with $\phi(f,\omega) \to \phi(f,\omega) + \beta$ for any $\beta$ and $f_j \in (f^a, f^{a+1})$. In particular $K^a_n(\cdot|\phi,\beta)$ is so constructed. The desired causality and locality additionally require the general RF to be composed of only causal and most local filters: $\mathbf{K^a_n} = \sum_{\phi,\beta} w(\phi,\beta)\mathbf{K^a_n}(\cdot|\phi,\beta)$, where $w(\phi,\beta)$ is a weight function.

Note that equation (29) can also be written as

$$\begin{aligned}\mathbf{K^a_n}(x; t-t') \quad \propto \quad & \sum_{f^a<f\leq f^{a+1}} \int_0^\infty d\omega K(f,\omega)(\cos((f(x^a_n - x) - \pi n/2) + (\omega(t-t') + \phi(f,\omega)) + \phi^+) \\ & + \cos((f(x^a_n - x) - \pi n/2) - (\omega(t-t') + \phi(f,\omega)) + \phi^-)) \qquad (30)\end{aligned}$$

with $\phi^+ = \phi + \beta$ and $\phi^- = \phi - \beta$. Then the general RF $\mathbf{K^a_n} = \sum_{\phi,\beta} w(\phi,\beta)\mathbf{K^a_n}(\cdot|\phi,\beta)$ is

$$\begin{aligned}\mathbf{K^a_n} \quad \propto \quad & \sum_{f^a<f\leq f^{a+1}} \int_0^\infty d\omega K(f,\omega)(A^+\cos((f(x^a_n - x) - \pi n/2) + (\omega(t-t') + \phi(f,\omega)) + \phi^+) \\ & + A^-\cos((f(x^a_n - x) - \pi n/2) - (\omega(t-t') + \phi(f,\omega)) + \phi^-)) \qquad (31)\end{aligned}$$

with $A^\pm e^{\phi^\pm} = \sum_{\phi,\beta} w(\phi,\beta)e^{\phi\pm\beta}$. The best possible translation invariance $\mathbf{K^a_n}(x, t-t') = (\pm)\mathbf{K^a_{n+2}}(x - (x^a_{n+2} - x^a_n), t-t')$ requires the same $(A^\pm, \phi^\pm) = (A^\pm_e, \phi^\pm_e)$ for all the even $n$ and $(A^\pm, \phi^\pm) = (A^\pm_o, \phi^\pm_o)$ for all the odd units. Hence, the general RFs are within this class of filters and decorrelation between outputs $O^a_n = \int_{-\infty}^\infty dx \int_{-\infty}^\infty dt' \mathbf{K^a_n}(x^a_n - x, t-t')S(x,t')$, i.e., $< O^a_n(t)O^a_m(t') >= \delta_{tt'}\delta_{nm}$, will restrict the choices of the parameters $(A^\pm_e, A^\pm_o, \phi^\pm_e, \phi^\pm_o)$.

The output correlation is

$$\begin{aligned}< O^a_n(t_1)O^a_m(t_2) > \quad = \quad & \int_{-\infty}^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty dx_1dx_2dt'_1dt'_2K^a_n(x^a_n - x_1, t_1 - t'_1) \\ & < S(x_1,t'_1)S(x_2,t'_2) > K^a_m(x^a_m - x_2, t_2 - t'_2) \\ = \quad & \int_{-\infty}^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty dx_1dx_2dt'_1dt'_2K^a_n(x^a_n - x_1, t_1 - t'_1) \\ & R(x_1 - x_2, t'_1 - t'_2)K^a_m(x^a_m - x_2, t_2 - t'_2) \\ = \quad & \int_{-\infty}^\infty df \int_{-\infty}^\infty d\omega K^a_n(f,\omega)R(f,\omega)K^{a\,*}_m(f,\omega)e^{if(x^a_n - x^a_m) + i\omega(t_1 - t_2)}\end{aligned}$$

where $K_n^a(f,\omega)$ is the Fourier transform of $K_n^a(x_n^a - x, t-t') \equiv \int_{-\infty}^{\infty} df \int_{-\infty}^{\infty} d\omega K_n^a(f,\omega) e^{if(x_n^a - x) + i\omega(t-t')}$. Hence for $f^a < |f| \leq f^{a+1}$

$$K_n^a(f,\omega) = K(f,\omega) \cdot \begin{cases} A_n^+ e^{i(-\pi n/2 + \phi(f,|\omega|) + \phi_n^+)} & \text{if } f > 0,\ \omega \geq 0 \\ A_n^+ e^{-i(-\pi n/2 + \phi(f,|\omega|) + \phi_n^+)} & \text{if } f < 0,\ \omega < 0 \\ A_n^- e^{i(-\pi n/2 - \phi(f,|\omega|) + \phi_n^-)} & \text{if } f > 0,\ \omega < 0 \\ A_n^- e^{-i(-\pi n/2 - \phi(f,|\omega|) + \phi_n^-)} & \text{if } f < 0,\ \omega \geq 0 \end{cases} \tag{32}$$

where $(A_n^{\pm}, \phi_n^{\pm}) = (A_e^{\pm}, \phi_e^{\pm})$ or $(A_o^{\pm}, \phi_o^{\pm})$ when $n$ is even or odd. Since $K(f,\omega) = R^{-1/2}(f,\omega)$, denoting complex conjugate by *c.c.*, we have

$$\begin{aligned} < O_n^a(t_1) O_m^a(t_2) > &= \sum_{f^a < f \leq f^{a+1}} \int_0^{\infty} d\omega (A_n^+ A_m^+ e^{if(x_n^a - x_m^a) - i\text{sgn}(f)\pi(n-m)/2 + i\omega(t_1-t_2) + i(\phi_n^+ - \phi_m^+)} + c.c.) \\ &+ (A_n^- A_m^- e^{if(x_n^a - x_m^a) - i\text{sgn}(f)\pi(n-m)/2 - i\omega(t_1-t_2) + i(\phi_n^- - \phi_m^-)} + c.c.) \end{aligned} \tag{33}$$

where $\text{sgn}(f) = 1$ or $-1$ when $f > 0$ or $f < 0$, respectively.

To continue, we note that $U^a$, as given in equation (28), is a unitary matrix, hence, $\sum_j U_{nj}^a (U_{mj}^a)^* = \delta_{nm}$. Then

$$\sum_{f^a < |f| \leq f^{a+1}} e^{i(f(x_n^a - x_m^a) - \text{sgn}(f)\pi(n-m)/2)} = \sum_{f^a < f \leq f^{a+1}} e^{i(f(x_n^a - x_m^a) - \pi(n-m)/2)} + c.c. \propto \delta_{nm} \tag{34}$$

Hence, $\sum_{f^a < f \leq f^{a+1}} e^{i(f(x_n^a - x_m^a) - \pi(n-m)/2)} \equiv i\rho$ is a pure imaginary number when $n \neq m$. Similarly $\int_0^{\infty} d\omega e^{-i\omega(t-t')} \equiv i\eta$ is a pure imaginary number when $t \neq t'$. By the definition (Li and Atick 1994a) of $U^a$, $f_j = 2\pi j/N$ (radian/grid) with $j = j^a + 1, j^a + 2, ...., j^{a+1}$, $N^a = 2(j^{a+1} - j^a)$. Then

$$\begin{aligned} &\sum_{f^a < f \leq f^{a+1}} e^{i(f(x_n^a - x_m^a) - \pi(n-m)/2)} \\ &= e^{-i\pi(n-m)/2 + if_{j^a+1}(x_n^a - x_m^a)} \sum_{j=0}^{N^a/2-1} (\exp(i(x_n^a - x_m^a)2\pi/N))^j \\ &= e^{-i\pi(n-m)/2 + if_{j^a+1}(x_n^a - x_m^a)} (1 - \exp(i\pi(x_n^a - x_m^a)N^a/N))/(1 - \exp((i2\pi(x_n^a - x_m^a)/N))) \\ &= \begin{cases} 0 & \text{if } n = m + 2l \neq n, \text{ since } x_n^a = (N/N^a)n \text{ or } (N/N^a)(n + n \bmod 2); \\ 1/2 & \text{(up to a normalization constant), if } n = m; \\ i\rho & \text{otherwise.} \end{cases} \end{aligned}$$

This gives decorrelation $< O_n^a(t_1) O_m^a(t_2) >= 0$ in equation (33) for all $m = n + 2k \neq n$ for any $(A^{\pm}, \phi^{\pm})$. While for $n = m$, we have the temporal decorrelation: $< O_n^a(t_1) O_n^a(t_2) > \propto \int_{-\infty}^{\infty} d\omega e^{i\omega(t_1 - t_2)} \propto \delta_{t_1 t_2}$

When $m = n + 2k + 1$, we differentiate two situations: (1) $t_1 = t_2$ and (2) $t_1 \neq t_2$. Their respective decorrelation require

$$\begin{aligned} < O_n^a(t) O_{n+2k+1}^a(t) > &= (A_n^+ A_m^+ i\rho e^{i(\phi_n^+ - \phi_m^+)} + c.c.) + (A_n^- A_m^- i\rho e^{i(\phi_n^- - \phi_m^-)} + c.c.) \\ &= -\rho(A_n^+ A_m^+ \sin(\phi_n^+ - \phi_m^+) + A_n^- A_m^- \sin(\phi_n^- - \phi_m^-)) = 0, \end{aligned} \tag{35}$$

$$
\begin{aligned}
< O_n^a(t_1) O_{n+2k+1}^a(t_2 \neq t_1) > &= 0 \\
&= (A_n^+ A_m^+ i\rho \int_0^\infty d\omega e^{i\omega(t_1-t_2)+i(\phi_n^+-\phi_m^+)} + c.c.) \\
&+(A_n^- A_m^- i\rho \int_0^\infty d\omega e^{-i\omega(t_1-t_2)+i(\phi_n^--\phi_m^-)} + c.c.) \\
&= (A_n^+ A_m^+ i\rho i\eta e^{i(\phi_n^+-\phi_m^+)} + c.c.) - (A_n^- A_m^- i\rho i\eta e^{i(\phi_n^--\phi_m^-)} + c.c) \\
&= -\rho\eta(A_n^+ A_m^+ \cos(\phi_n^+ - \phi_m^+) - A_n^- A_m^- \cos(\phi_n^- - \phi_m^-)). \qquad (36)
\end{aligned}
$$

Combining equations (35) and (36) gives $A_n^+ A_m^+ e^{i(\phi_n^+-\phi_m^+)} = A_n^- A_m^- e^{-i(\phi_n^--\phi_m^-)}$ for $m = n+2k+1$. Hence we have $A_e^\pm = \gamma A_o^\mp$, for $\gamma = \pm 1$, and $\phi_e^+ + \phi_e^- = \phi_o^+ + \phi_o^-$. It then concludes that the general efficient spatiotemporal code is of the form in equations (10) - (14) and is determined by 5 parameters $(A_e^+, A_e^-, \phi^x \equiv (\phi^+ + \phi^-)/2, \phi_e^t \equiv (\phi_e^+ - \phi_e^-)/2, \phi_o^t \equiv (\phi_o^+ - \phi_o^-)/2)$.