

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

Zhaoping Li^{*}

Technical Report HKUST-CS95-10
April 21, 1995

^{*}Department of Computer Science
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong

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

Abstract

This paper demonstrates that much of the visual motion coding in the 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, are tuned to different directions of motion, and their receptive fields (RF) can be spatio-temporally separable or inseparable. The predictions also include the following correlations between the motion coding and spatial, chromatic, and stereo codings: the preferred speed is larger when the cell receptive field sizes are larger, the color channel prefers smaller speed than the luminance channel, both the optimal speeds and the preferred directions of motion can be different for the two eyes in 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) nearby cortical cells of a given preferred orientation and scale are to prefer opposite directions of motion, have the same directional index, and are of quadrature spatio-temporal phase relationships with each other; (2) a single neuron can have different optimal motion speeds for opposite motion directions of monocular stimuli to a given eye, and (3) a neuron's ocular dominance may change with motion direction if the two eyes prefer opposite motion directions.

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 orientations, spatial frequencies, and ocular origins 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 the cortical simple cells can be understood as consequences of efficient coding of visual inputs in the multiscale framework. Such an understanding provides detailed predictions of the simple cell spatio-temporal receptive field (RF) properties. These predictions can be compared with the known observations or experimentally tested.

Efficiency of information presentation has long been advocated as the coding principle for early stages of sensory processings (Barlow 1961). This is because the natural signals have structure and regularity. Visual inputs, for example, have correlated signals in image pixels, making some input signals predictable to some extent from the others. Such regularities make pixel-by-pixel input representation highly redundant or inefficient, in the sense that the same information is signalled more than once 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) by the input statistical knowledge, which has to be inherent in the code to reduce the redundancy.

One of the most noticable input redundancies is the pairwise pixel correlations in the

visual inputs. 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 (Linsker 1989, Atick and Redlich 1990, 1992, 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 (Atick and Redlich 1990, 1992 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 the 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 is recently 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 is 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 (see Li and Atick 1994b for discussions).

Efficient coding in the multiscale representation has predicted many of the simple cell RF properties in spatial, chromatic, and stereo domain (Li and Atick 1994ab). These predictions include the simple cell selectivities to orientation, spatial frequencies, color, ocular origins, disparities, as well as the particular frequency tuning bandwidth, phase

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

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 significant or 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 the non-linearity at a later stage and are designed for motion sensing or detection within the constraints of known physiological and psychophysical observations. The current work derives the visual motion coding in a linear mechanism from the requirement of

efficient coding in a multiscale representation, without a priori specifying a purpose of visual motion computation or selectivity. Its predictions include those that have not been experimentally investigated in addition to the ones that agree with known observations. A special case out of 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 illustrates the predicted spatiotemporal RF and correlations between motion coding and codings in the space, color, and stereo domains, to compare them with the 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

The visual input is inefficient because the input $S(x_n, t)$ at retina location x_n and time t is correlated with $S(x_m, t')$ by the amount

$$R_{x_n, t; x_m, t'} \equiv \langle S(x_n, t) S(x_m, t') \rangle, \quad (1)$$

where the bracket denotes ensemble average. 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 $R_{x_n, t; x_m, t'} = R_{x_n+a, t+\tau; x_m+a, t'+\tau} \equiv R(x_n - x_m, t - t') = R(\pm(x_n - x_m), \pm(t - t'))$. Then R can also be characterized by its Fourier transform¹ $R(f_j, \omega) = \frac{1}{2\pi} \sum_n \int dt R(x_n, t) e^{-if_j x_n - i\omega t} = R(\pm f_j, \pm \omega)$, which is also the average input

¹The 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, ω) specifies the actual function concerned. Such practices are used throughout the paper for some other functions and variables as well to avoid proliferation of notations.

power in frequency (f_j, ω) .

Under noiseless conditions, a more efficient code $O(j, t)$ can be constructed within the linear coding scheme by a transform $O(j, t) = \sum_x \int dt' K(j, t; x, t') S(x, t')$ such that the outputs are decorrelated $\langle O(j, t) O(j', t') \rangle = \delta_{jj'} \delta(t - t')$. 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 similarly as the spatial dimension because of causality. In addition, the scale invariance property, and hence the multiscale coding, in the spatial domain is not applied temporally.

A special efficient code is the one when O_j is obtained by passing $S(x, t)$ through a spatial filter $K_x^{f_j}$ and a temporal filter $K_t^{f_j}$, namely

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

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

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

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

where N is the total number of input units, $\phi(f_j, \omega) = -\phi(f_j, -\omega)$ is chosen such that the temporal filter $K_t^{f_j}(t - t')$ is causal ($K_t^{f_j}(t < 0) = 0$) and have minimum temporal spread² for each j .

²Define $A^{f_j}(t) = \frac{1}{\pi} \int_0^{\infty} d\omega R^{-1/2}(f_j, \omega) e^{-i\omega(t-t') - i\phi(f_j, \omega)}$, taking envelop(t) and phase(t) as the ampli-

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 $S^*(f_{j'}, t)$ should be decorrelated in a translationally invariant system

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

where superscript $*$ denotes complex conjugate. This ensures the decorrelation between $O(f_j, t)$ and $O^*(f_{j'}, t)$ for $f_j \neq f_{j'}$. Each $O(f_j, t)$ is in addition temporally decorrelated by the temporal whitening filter $K_t^{f_j}$, which has Fourier transform $K_t^{f_j}(\omega) = R^{-1/2}(f_j, \omega)$:

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

The spatio-temporal receptive fields K^{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-local³. In addition $K_x^{f_j}$ is unique for each output j with its unique frequency

tude and phase of $A^{f_j}(t)$, then $K_t^{f_j}(t) = \text{envelop}(t) \cos(\text{phase}(t))$. The minimum temporal spread of $K_t^{f_j}$ is defined as when $\int dt(t - \bar{t})^2 \text{envelop}(t)$, where $\bar{t} = \int dt \text{envelop}(t)$, is minimum

³The RF's for this code is not real, but this representation is used for mathematical convenience and it does not affect the final results.

f_j , requiring a unique RF for each output cell. However, other efficient codes can be constructed from this one by (Li and Atick 1994a) 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 $\sum_{jj'} U_{ij}(U_{i'j'})^* < O_j O_{j'}^* \geq \delta_{ii'}$. As noted in the introduction, the cortex is to construct a multiscale representation, which is also spatially local, retinotopic, and translationally invariant (in a sense detailed later) within each scale. An efficient code of the 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 $f^a \leq |f_j| < f^{a+1}$ by a unitary transform \mathbf{U}^a in that band:

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

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

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, $f^{a+1} = 3f^a$, $f^a = \frac{\pi N^a}{2N}$, and $N^a \propto (f^{a+1} - f^a)$ is the number of units in that scale. The required $N^a \times N^a$ matrix transform \mathbf{U}^a has elements (Li and Atick 1994a):

$$U_{nj}^a = \begin{cases} \frac{1}{\sqrt{N^a}} e^{i(f_j x_n^a - \pi n/2 + \theta)} & \text{if } f_j > 0 \\ \frac{1}{\sqrt{N^a}} e^{-i(|f_j| x_n^a - \pi n/2 + \theta)} & \text{if } f_j < 0 \end{cases} \quad (10)$$

for $n = 1, 2, \dots, N^a$ and arbitrary θ . Combining (10), (9), (3), (4), and (5), we have

$$K_n^a(x_n^a, t; x, t') \propto \frac{2}{\sqrt{N^a}} \int_{f^a}^{f^{a+1}} df \int_0^\infty d\omega K(f, \omega) \cos(f(x_n^a - x) - \pi n/2 + \theta) \cos(\omega(t - t') + \phi(f, \omega)) \quad (11)$$

where $K(f, \omega) \equiv R^{-1/2}(f, \omega)$ denotes the spatio-temporal sensitivity.

Let us examine $K_n^a(x_n^a, t; x, t')$, 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 ω with a sensitivity $K(f, \omega)$. The spatial and temporal part of the filter at each (f, ω) are embodied in the expressions $\cos(f(x_n^a - x) - \pi n/2 + \theta)$ and $\cos(\omega(t - t') + \phi(f, \omega))$, respectively. When $x = x_n^a$, the spatial phase of the RF is $-\pi n/2 + \theta$ for all f , implying x_n^a as the center of the receptive field. This RF center phase $-\pi n/2 + \theta$ states that neighboring units (n^{th} and $(n+1)^{th}$) have a quadrature phase difference, e.g., $\theta = 0$ gives edge and bar detectors (odd and even RF) for the two neighboring units, respectively. The choices of the RF centers x_n^a for an efficient code (i.e., a unitary \mathbf{U}^a) are⁴ $x_n^a = (N/N^a)n$ or $x_n^a = (N/N^a)(n + n\%2)$ where $n\%2$ is $(n \bmod 2)$ and is equal to 0 or 1 when n is even or odd, respectively. This means, the RF centers of the units in the code are located at, in units of N/N^a (roughly half a grating period of the preferred spatial frequency of the cells), $(0, 1, 2, \dots, N^a - 1)$ or $(0, 0, 2, 2, 4, 4, \dots, N^a - 2, N^a - 2)$ in the two cases, respectively, to cover the visual space with two units every two sampling intervals $2N/N^a$. The $(n+2)^{th}$ unit has a spatial phase $-\pi n/2 - \pi + \theta$ at its RF center and hence the same spatial RF as the n^{th} filter up to a polarity change and a spatial translation by a distance $\Delta x = 2N/N^a$. All the units

⁴Here both choices, $x_n^a = x_n^a = (N/N^a)n$ and $x_n^a = (N/N^a)(n + n\%2)$, are valid. In Li and Atick 1994a however, only the first choice is given.

however have the same temporal RF properties. A finite bandwidth (f^a, f^{a+1}) , by the uncertainty principle, gives spatially local receptive fields with a spatial extent roughly $\Delta x \sim 1/(f^{a+1} - f^a) \sim N/N^a$ on the order of magnitude, comparable to the displacement between the RF centers of the neighboring units in this code. One should distinguish the *neighboring units* (whose RF centers may be displaced by a distance comparable to their RF width) in this efficient code from the anatomical neighboring cells in the cortex. This is because we are concentrating here on one scale and one orientation (one-dimension) only at a time, while in the cortex, the anatomical neighboring cell could be tuned to another optimal spatial frequency and orientation and there are other factors that have to be considered (see discussions). From here on, $x_n^a - x \rightarrow x$, $t - t' \rightarrow t$, and $K_n^a(x_n^a, t; x, t') \rightarrow K_n^a(x; t)$ are taken for simplicity.

In fact, there can be different but equivalently efficient multiscale representations by different choices of θ . The specific θ dictates whether the receptive field $K_n^a(x; t)$ is even and odd for neighboring units (when θ is a multiple of $\pi/2$) or of any phase as long as the quadrature phase relationship between the neighboring units is preserved. Similarly, one can show that decorrelation is preserved by the phase change in the temporal filter $\phi(f_j, \omega) \rightarrow \phi(f_j, \omega) + \beta$ for any β , without compromising causality, temporal translation invariance (i.e., $K_t(t, t') = K_t(t + \tau, t' + \tau)$) as a physical system, and the minimum temporal latency spread of the filter. The general description for this class of filters is:

$$K_n^a(x, t) \propto 2 \int_{f^a}^{f^{a+1}} df \int_0^\infty d\omega K(f, \omega) \cos(fx - \pi n/2 + \theta) \cos(\omega t + \phi(f, \omega) + \beta) \quad (12)$$

$$\begin{aligned}
&= \int_{f^a}^{f^{a+1}} df \int_0^\infty d\omega K(f, \omega) (\cos(fx - \pi n/2 + \theta + \omega t + \phi(f, \omega) + \beta) \\
&\quad + \cos(fx - \pi n/2 + \theta - \omega t - \phi(f, \omega) - \beta))
\end{aligned} \tag{13}$$

$$\begin{aligned}
K_n^a(x, t) &\approx K_n^a(x) K^a(t) & K^a(t) &\equiv \int_0^\infty d\omega K(f^{peak}, \omega) \cos(\omega t + \phi(f^{peak}, \omega)) \\
K_n^a(x) &\equiv \frac{2}{\sqrt{N^a}} \int_{f^a}^{f^{a+1}} df \cos(fx - \pi n/2 + \theta)
\end{aligned} \tag{14}$$

where $f^{peak} = \sqrt{f^a f^{a+1}}$. If one notes that a drifting grating $\cos(fx + \text{spatial phase} \pm (\omega t + \text{temporal phase}))$ indicates a drifting velocity $v = \pm \omega/f$, then equation (13) states that this class of filters are non-directionally selective, since the first term makes it respond to motion in one direction and the second term to the opposite motion direction with the same amplitude. This is also intuitively expected from equation (14) which approximates the filter as spatiotemporally separable from equation (12), which is strictly speaking not spatio-temporally separable. This approximation is valid when $K(f, \omega)$ is a smooth function of f and changes little within a limited band (f^a, f^{a+1}) (see next section). From here on, such filters will be viewed as separable.

However, as is shown in the Appendix, the general class of efficient codes in the multiscale framework is in fact⁵,

$$\mathbf{K}_n^a = K_n^a(\theta, \beta) + (-1)^n q K_n^a(\theta', \beta') \tag{15}$$

for any $q \leq 1$, $\theta \neq \theta'$, and $\beta' - \beta = \pi/2$. Here, θ , θ' , and β (β'), in addition to q , are explicitly written as parameters for the representation and other arguments are omitted

⁵This is true up to a factor $1/(1+q^2)$, which I am ignoring for clarity. $q \leq 1$ is taken without loss of generality.

for clarity. One could recognize from this that a sum of two different spatiotemporally separable filters can give spatiotemporal inseparability and hence directionality in the new filter. This is indeed what happens. The bold-faced \mathbf{K}_n^a is to distinguish the general composite code from the spatio-temporally separable ones. $q = 0$ simply gives equation (12) as a special case. The other special case, when $q = 1$ and $\theta' - \theta = \pi/2$, gives

$$\begin{aligned}\mathbf{K}_n^a &\propto \int_{f^a}^{f^{a+1}} df \int_0^\infty d\omega K(f, \omega) \cos(fx - \pi n/2 + \theta) \cos(\omega t + \phi(f, \omega) + \beta) \\ &\quad + (-1)^n K(f, \omega) \sin(fx - \pi n/2 + \theta) \sin(\omega t + \phi(f, \omega) + \beta) \\ &= \int_{f^a}^{f^{a+1}} df \int_0^\infty d\omega K(f, \omega) \cos(fx - \pi n/2 + \theta + (-1)^n(\omega t + \phi(f, \omega) + \beta)) \quad (16)\end{aligned}$$

This (the n^{th}) unit thus responds to only one direction of motion, and is composed of a pair of spatio-temporal separable filters, $\int_{f^a}^{f^{a+1}} df \int_0^\infty d\omega K(f, \omega) \cos(fx - \pi n/2 + \theta) \cos(\omega t + \phi(f, \omega) + \beta)$ and $\int_{f^a}^{f^{a+1}} df \int_0^\infty d\omega (-1)^n K(f, \omega) \sin(fx - \pi n/2 + \theta) \sin(\omega t + \phi(f, \omega) + \beta)$, with quadrature phase relationship between the two components in both the spatial and temporal 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). Furthermore, the factor $(-1)^n$ in \mathbf{K}_n^a states that neighboring units (the n^{th} and $(n+1)^{th}$ units) respond to the opposite directions of motion, and the term $-\pi n/2$ in the spatial phase again leads to a quadrature phase relationship between the neighboring units.

In general when $q \leq 1$ and $\Delta\theta \equiv \theta' - \theta$, equations (15) and (13) lead to

$$\begin{aligned}\mathbf{K}_n^a(x; t) &\propto \int_{f^a}^{f^{a+1}} df \int_0^\infty d\omega K(f, \omega) (A^+ \cos(fx - \pi n/2 + \theta + \omega t + \phi(f, \omega) + \beta + \theta^+) \\ &\quad + A^- \cos(fx - \pi n/2 + \theta - \omega t - \phi(f, \omega) - \beta + \theta^-)) \quad (17)\end{aligned}$$

$$\begin{aligned}
&= \int_{f^a}^{f^{a+1}} df \int_0^\infty d\omega K(f, \omega) \\
&\quad (A^+ + A^-) \cos(\phi(x)) \cos(\phi(t)) + (A^- - A^+) \sin(\phi(x)) \sin(\phi(t))
\end{aligned} \tag{18}$$

$$A^\pm e^{i\theta^\pm} = 1 + (-1)^n q e^{i(\pm\pi/2 + \Delta\theta)} \quad (\text{see Figure 1}) \tag{19}$$

$$\begin{aligned}
\phi(x) &= fx - \pi n/2 + \theta + 1/2(\theta^+ + \theta^-) \\
\phi(t) &= \omega t + \phi(f, \omega) + \beta + 1/2(\theta^+ - \theta^-)
\end{aligned} \tag{20}$$

From equation (17), the neuron responds to the two directions of motion with relative amplitudes A^+ and A^- , respectively, and has a directional index $\frac{|A^+ - A^-|}{|A^+ + A^-|}$. Equation (19) states that the n^{th} and the $(n \pm 1)^{th}$ units have $A_n^+ = A_{n \pm 1}^-$, $\theta_n^+ = \theta_{n \pm 1}^-$. Consequently, the neighboring units prefer opposite motion directions, have the same directional index, and are again in phase quadrature of each other due to the $-\pi n/2$ term in $\phi(x)$. The neurons are non-directional when $q = 0$ and becomes more directional as $q \rightarrow 1$ and $\Delta\theta \rightarrow \pi/2$. One should note that in extending this analysis from 1-dimension to 2-dimension in space, these neighboring neurons prefer the same spatial orientation. Since the centers of the RFs in the code is spaced by $x_n^a = (N/N^a)n$ or $x_n^a = (N/N^a)(n + n\%2)$ for $n = 1, 2, \dots, N^a$, one has the following observation. 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 for every two sampling intervals of the scale, tuned to opposite directions and their RF centers superposed or displaced by one sampling interval of the scale (see discussion).

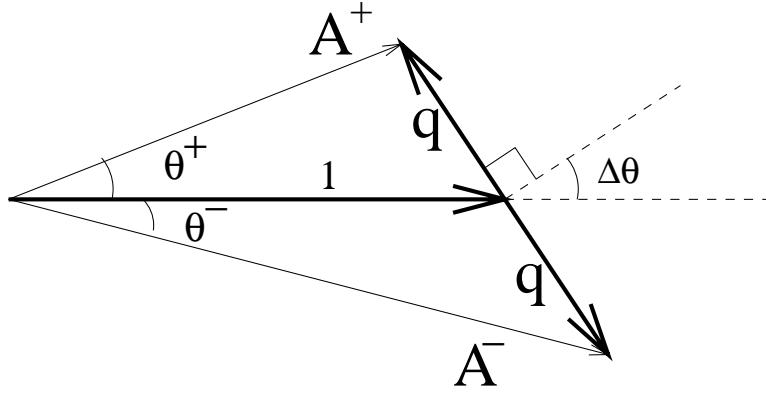


Figure 1: Demonstration of the vector construction of A^\pm from q and $\Delta\theta$

Equation (18) writes the filter again as a spatio-temporal quadrature pair with unequal amplitudes $|A^+ \pm A^-|$. This is similar to the model by Hamilton, Albrecht, and Geisler (1989), except that each term in the current theory is only strictly separable for one frequency component (f, ω) only. Note that this expression does not mean that the general equation (15) can be written with $\theta' - \theta = \pi/2$ since $\phi(t)$ changes with n in general because $\theta_n^+ \neq \theta_n^-$.

3 Theoretical spatiotemporal filters and comparison with the experimental observations

To illustrate the spatiotemporal RF, we need the knowledge of $K(f, \omega)$ and $\phi(f, \omega)$. Although $\phi(f, \omega)$ depends ultimately on $K(f, \omega)$ to ensure causality, the spatiotemporal locality of the filter implies a phase coherence, $fx_p + \omega t_p + \phi(f, \omega) \approx \phi_c = \text{constant}$ for all (f, ω) , at some location $(x = x_p, t = t_p)$ where the amplitude of the receptive field reaches its peak and quickly decays as (x, t) moves away from it⁶. Taking $x_p = 0$ and

⁶As we stated earlier, $\phi(f, \omega)$ is chosen to make $K_t^f(t)$ causal and have minimum temporal spread, implying the temporal coherence $\omega\tau_p + \phi(f, \omega) \approx \text{constant}$ for all ω given a f . A change $\phi(f, \omega) \rightarrow \phi(f, \omega) + \alpha$ still satisfies the requirement; and $\phi(f, \omega) \rightarrow \phi(f, \omega) - \omega\tau$ for $\tau > 0$ merely prolongs the filter

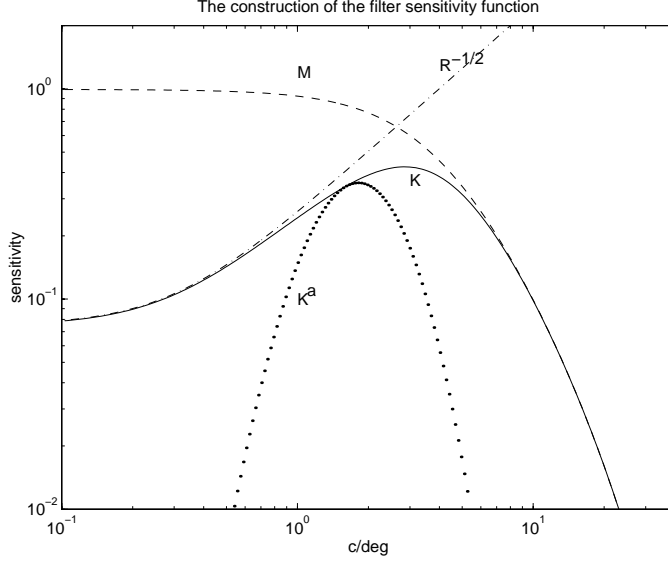


Figure 2: Demonstration of the construction of the receptive field sensitivity K and K^a from the power spectrum R and noise smoothing (and retinal MTF) function M . The temporal dimension is ignored for clarity in this figure.

$\beta + \phi_c \rightarrow \beta$, we have $\phi(t) = \omega(t - t_p) + \beta + 1/2(\theta^+ - \theta^-)$ in equation (20), and $t_p > 0$ can be viewed as the latency of the filter. Similar temporal phase structures have been observed in experiments (Hamilton et al 1989).

The exact $K(f, \omega)$ depends on $R(f, \omega)$ and should be modified under noise. Different measurements have suggested that $R(f, \omega = 0) \propto 1/f^2$ (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. The qualitative results do not depend crucially on the exact $R(f, \omega)$ as long as $R(f, \omega)$ decays with increasing f and ω . The complete decorrelation requires (see equation (11)) $K(f, \omega) =$

latency $\tau_p \rightarrow \tau_p + \tau$. Although the minimum latency $\tau_p = \tau_p^{min}$ depends on f , it is possible to choose t_p as the largest τ_p^{min} within a limited band (f^a, f^{a+1}) , such that $f x_p + \omega t_p + \phi(f, \omega) \approx constant$ 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.

$(R(f, \omega))^{-1/2}$ and causes $K(f, \omega)$ to increase with (f, ω) to amplify inputs of low signal power $R(f, \omega)$. This is undesirable in reality since the low input signal-to-noise ratio at high (f, ω) would give unacceptably amplified noise levels. A noise smoothing strategy $K(f, \omega) \approx M(f, \omega)K(f, \omega)|_{noiseless}$ (following Atick and Redlich 1992, Li and Atick 1994b, and noise smoothing also follows from information theoretical arguments), where $M \propto R(f, \omega)/(R(f, \omega) + R_N)$ and R_N is the noise power, is utilized to lower the amplitude of $K(f, \omega)$ whenever the signal-to-noise $R(f, \omega)/R_N$ is small⁷ especially at high (f, ω) . Let the noise power R_N be constant for all f, ω , the generic feature of $K(f, \omega)$ is (Figure 2):

$$K(f, \omega) \quad \text{increases with } f, \omega \text{ when } R(f, \omega) \gg R_N \text{ at small } (f, \omega) \quad (21)$$

$$K(f, \omega) \quad \text{decreases with } f, \omega \text{ when } R(f, \omega)/R_N \text{ is small at large } (f, \omega) \quad (22)$$

$K(f, \omega)$ reaches its peak at some intermediate value (f, ω) , where the signal $R(f, \omega)$ starts to be overwhelmed by noise. Hence, $K(f, \omega)$ will peak at smaller (f, ω) for smaller signal R . In particular, the temporal sensitivity curve $K_t^f(\omega) \equiv K(f, \omega)$ for each spatial frequency f also peaks at some $\omega = \omega^{peak}(f)$, and $\omega^{peak}(f)$ decreases if $R(f, \omega) \rightarrow R(f, \omega) - \Delta$ where $\Delta > 0$. This fact will lead to many motion coding features presented below, when spatio-temporal sensitivities $K(f, \omega)$ will be derived for channels tuned to different spatial frequencies, to luminance or chrominance, and to different ocular origins.

In the multiscale representation, we further model (Figure 2) $\int_{f_a}^{f_a+1} df K(f, \omega)$ 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 chan-

⁷In 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.

nel with optimal frequency $f^{peak} = \sqrt{f^a f^{a+1}}$. Collecting the parts together, the following model is used for illustrative purpose:

$$\mathbf{K}_{\mathbf{n}}^{\mathbf{a}}(x; t) \propto \int df \int_0^\infty d\omega K^a(f, \omega) \\ ((A^+ + A^-) \cos(\phi(x)) \cos(\phi(t)) + (A^- - A^+) \sin(\phi(x)) \sin(\phi(t))) \quad (23)$$

$$\phi(x) = fx - \pi n/2 + \theta \quad \phi(t) = \omega t + \beta + 1/2(\theta^+ - \theta^-) \\ A^\pm e^{i\theta^\pm} = 1 + (-1)^n q e^{i(\pm\pi/2 + \Delta\theta)} \quad (24)$$

Here $\theta + 1/2(\theta^+ + \theta^-) \rightarrow \theta$ is taken for simplicity since $\theta^+ + \theta^-$ does not change with n ; $t - t_p \rightarrow t$ is taken so that all filter illustrations are centered around the latency t_p of the temporal impulse response and should not be mistaken as non-causal. Fig. 3 illustrates some examples of the spatio-temporal RF's for different values of q , $\Delta\theta$, θ , and β . In each case, the neighboring units have quadrature phase relationships with each other, the opposite preferred directions of motion, the same optimal speed, and the same directionality index. Physiologically, there is evidence that the 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 the neighboring cells preferring opposite directions also have the same directionality index and are in quadrature phase relationship. In this theory, changes of the parameters $(q, \Delta\theta, \theta, \beta)$ lead to a whole spectrum of directionality indices and filter phases. The optimal speed and the quadrature structure remain fixed for each spatial scale.

Correlation between spatial coding and motion coding

This theory predicts that the cell optimal speed decreases with increasing optimal

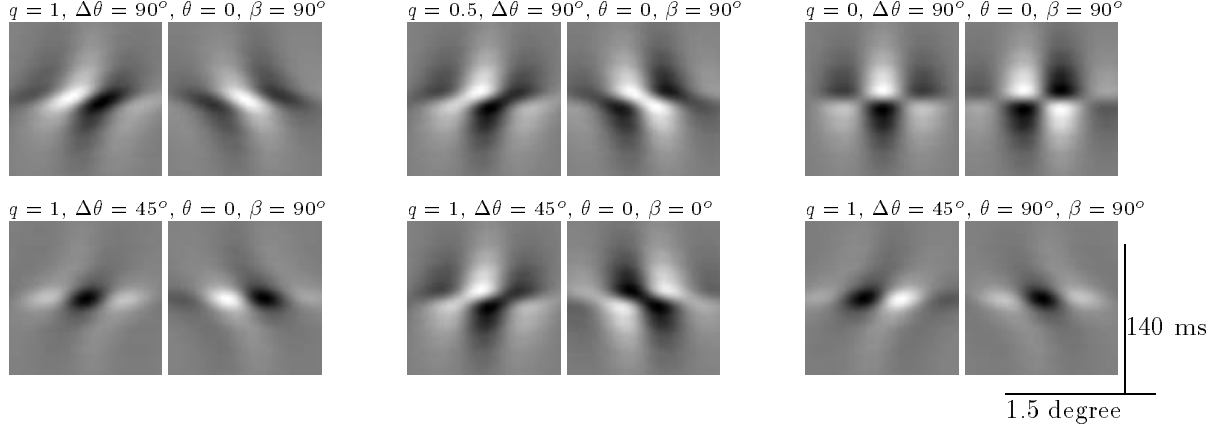


Figure 3: Examples of spatiotemporal RF's for neighboring units, next to each other, under the parameter value set $q, \Delta\theta, \theta, \beta$ which generates them. The space and time are in horizontal and vertical directions respectively. The picture gray levels depict the filter amplitudes, gray for near zero amplitudes, white and black for larger positive and negative amplitudes respectively. The preferred spatial frequency is $f^{peak} = 1$ cycle/degree. A perfectly oriented bar or edge in space-time implies complete cell directionality, the slope and sign of the orientation correspond to the preferred speed and direction of motion, and a spatiotemporal separability implies non-directionality. Note (1) the neighboring units have the quadrature phase relationship and opposite preferred directions of motion, (2) the reduction of directionality in the top row when $q = 1 \rightarrow 0$, (3) the bottom row cells have the same directionality but different receptive field forms. These figures are obtained by the approximation $\mathbf{K}_{\mathbf{n}}^{\mathbf{a}}(x, t) \approx (A^+ + A^-)K_x(x)K_t(t) + (A^- - A^+)\tilde{K}_x(x)\tilde{K}_t(t)$, where $K_t(t) \propto \int dw K^a(f^{peak}, \omega) \cos(\phi(t))$, $\tilde{K}_t(t) \propto \int dw K^a(f^{peak}, \omega) \sin(\phi(t))$, $K_x(x) \propto \int df K^a(f, \omega^{peak}) \cos(\phi(x))$, $\tilde{K}_x(x) \propto \int df K^a(f, \omega^{peak}) \sin(\phi(x))$. This approximation and figure format are used in other figures of this paper as well.

spatial frequency, as observed in experiments (Holub and Morton-Gibson 1981, Foster, Gaska, Nagler, and Pollen 1985). This is because for a neuron with optimal spatial frequency f^{peak} and optimal temporal frequency ω^{peak} , the preferred motion speed is roughly ω^{peak}/f^{peak} . Since the signal $R(f, \omega)$ decreases with increasing f , so should ω^{peak} , and the prediction follows. The model $R(f, \omega)$ used in this paper gives a slowly decreasing or roughly constant optimal temporal frequency $\omega^{peak}(f)$ for a range of low spatial frequencies f (Fig. 4), suggesting a roughly inverse relationship $v \sim 1/f$ between optimal speed and spatial frequency. At a higher frequency f , whose exact value depends on the signal-to-noise, $\omega^{peak}(f)$ starts to decrease sharply with f , and temporal sensitivity $K_t^f(\omega)$ becomes significantly low-pass. The same trend of $\omega^{peak}(f)$ is observed physiologically (Holub and

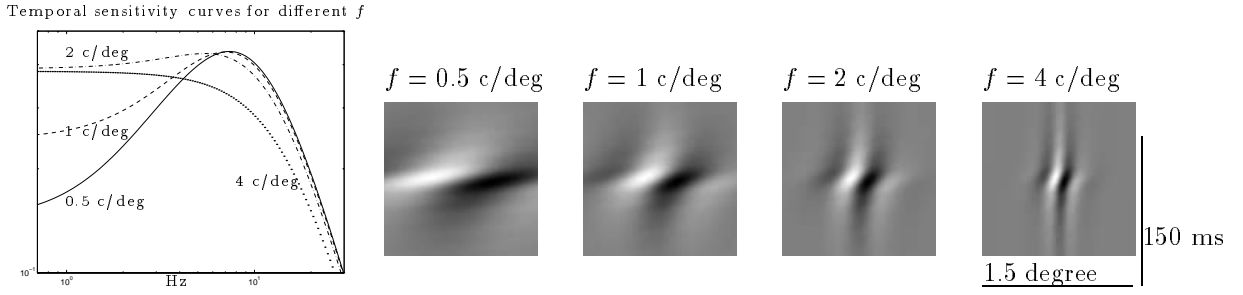


Figure 4: Changes of temporal sensitivities and spatiotemporal receptive fields with the optimal spatial frequency f . The filter orientation in space-time has a steeper slope as f increases, implying decreasing preferred motion speeds. Parameters used: $q = 1.0$, $\theta = 0^\circ$, $\Delta\theta = 90^\circ$, and $\beta = 90^\circ$.

Morton-Gibson 1981) and psychophysically (Kelly 1979). Fig.4 illustrates the changes of neural spatio-temporal RF's with their optimal spatial frequencies.

Another prediction is that the ratio, (best sensitivity to contrast reversal grating) : (sensitivity to drifting grating of preferred direction), is roughly one for neurons of small directional indices, but decreases to 0.5 as the directionality increases. This stems from equation (23), which suggests 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 non-linearity in cortical neurons can lead to a quantitative difference between the theory and the experiments (see Reid et al 1991). However, the trend of decreasing ratio (sensitivity to contrast reversal) : (sensitivity to drifting grating) with increasing directional index should still hold and can be tested.

Correlation between color and motion coding

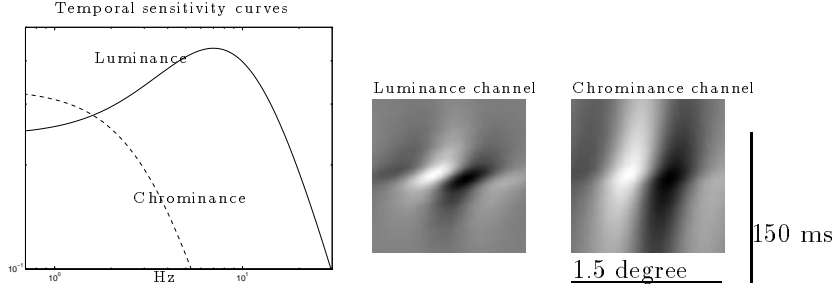


Figure 5: Temporal sensitivity and spatiotemporal RF's for luminance and chrominance channels. Parameters used $q = 1$, $\Delta\theta = 90^\circ$, $\beta = 90^\circ$, $\theta = 0$, $f^{peak} = 1\text{c/deg}$, and the signal power in the chromatic channel is 4% of that in the luminance channel. A smaller optimal motion speed in the chromatic channel is apparent.

Since the signal power in the chromatic channel is smaller than that of the luminance channel, the optimal temporal frequency ω^{peak} for the chromatic channel is also smaller given an optimal spatial frequency. Hence the theory predicts a smaller optimal speed for the chromatic channel (see Figure 5). Insensitivity to motion in the color channel has long been observed psychophysically (see Nakayama 1985). However, at a single 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 the motion stimuli is of equal luminance or not.

Correlation between stereo and motion coding

Stereo coding (Li and Atick 1994b, Li 1995a) is composed of ocular summation (the input summation from the two eyes) and ocular opponency (the input difference between the two eyes) channels. The subscript s and o are used to denote the summation and opponency channels respectively, the subscript n for the neuron and superscript a for scale will be omitted for clarity. The signal power in these two channels are $R_s = (1 + r)R(f, \omega)$ and $R_o = (1 - r)R(f, \omega)$, respectively, where $r < 1$ is the input ocular correlation

normalized by the self-correlation within a single eye⁸. The different signal powers $R_s > R_o$ suggest different spatiotemporal sensitivity functions $K_s(f, \omega)$ and $K_o(f, \omega)$ for the two channels. In particular, $K_s(f, \omega) \not\propto K_o(f, \omega)$ and the optimal speed in the opponency channel should be smaller $v_o < v_s$. Let $\mathbf{K}_s(x, t)$ and $\mathbf{K}_o(x, t)$ be the RFs for the two channels respectively, the binocular RFs in a cortical cell are (Li and Atick 1994b, Li 1995a) $\mathbf{K}_l(x, t) = \mathbf{K}_s(x, t) + \mathbf{K}_o(x, t)$ for the left and $\mathbf{K}_r(x, t) = \mathbf{K}_s(x, t) - \mathbf{K}_o(x, t)$ for the right eyes. Writing the filters in the form

$$\mathbf{K}_c(x, t) \propto \int df \int d\omega K_c(f, \omega) (A_c^+ \cos(fx + \omega t + \phi_c^+) + A_c^- \cos(fx - \omega t + \phi_c^-)) \quad (25)$$

$$\mathbf{K}_e(x, t) \propto \int df \int d\omega K_e^+(f, \omega) \cos(fx + \omega t + \phi_e^+) + K_e^-(f, \omega) \cos(fx - \omega t + \phi_e^-) \quad (26)$$

for $c = s, o$, and $e = l, r$. Then if $\mathbf{A}_c^\pm = A_c^\pm e^{i\phi_c^\pm}$ and $\mathbf{K}_e^\pm = K_e^\pm e^{i\phi_e^\pm}$, we have

$$\begin{aligned} \mathbf{K}_l^\pm(f, \omega) &= K_s(f, \omega) \mathbf{A}_s^\pm + K_o(f, \omega) \mathbf{A}_o^\pm \\ \mathbf{K}_r^\pm(f, \omega) &= K_s(f, \omega) \mathbf{A}_s^\pm - K_o(f, \omega) \mathbf{A}_o^\pm \end{aligned} \quad (27)$$

Here $K_e^+(f, \omega)$ and $K_e^-(f, \omega)$ are the spatiotemporal sensitivity functions for monocular stimuli in opposite motion directions. One first notes that the optimal motion speed for any direction presented to either eye is roughly $v_e^\pm \sim \omega^{peak} / f^{peak}$, where $(\omega^{peak}, f^{peak})$ is roughly where the sensitivity $K_e^\pm(f, \omega)$ reaches its peak. Since $K_s(f, \omega) \not\propto K_o(f, \omega)$, the above equations give $K_e^+(f, \omega) \not\propto K_e^-(f, \omega)$ and $K_l^\pm(f, \omega) \not\propto K_r^\pm(f, \omega)$, i.e., the sensitivity changes with motion direction or eye origin can not be accounted for by a constant gain

⁸The temporal dimension of visual inputs was ignored in the earlier works (Li and Atick 1994b, Li 1995a) and the ocular signal powers were denoted as $(1 \pm r(f))R(f)$. Here we simply assume that the generalization $(1 \pm r(f))R(f) \rightarrow (1 \pm r(f, \omega))R(f, \omega)$ holds approximately.

factor. Accordingly, the theory predicts that the optimal speed $v_e^\pm \sim \omega^{peak}/f^{peak}$ for monocular stimuli can vary with the eye origin *and* the motion direction (even within a single eye). In addition, since the preferred motion direction for any eye depends on whether $K_e^+ > K_e^-$, the fact $K_l^\pm(f, \omega) \not\propto K_r^\pm(f, \omega)$ makes it possible to obtain neurons whose preferred motion directions change with the ocular origin or whose ocular dominances change with motion directions, as demonstrated below. A final straightforward prediction is that, since K_e^+/K_e^- is not a constant of (f, ω) , the directional index $\frac{|K_e^+ - K_e^-|}{|K_e^+ + K_e^-|}$ for monocular motion stimuli can even vary with the frequency (f, ω) of the drifting grating presented, as observed physiologically (Reid et al 1991).

To illustrate the predictions, consider first the example when $\mathbf{A}_s^+ = \mathbf{A}_o^+ = \mathbf{A}_s^- = -\mathbf{A}_o^-$ (Fig 6A, 6C). Both the ocular summation and opponency RFs are non-directional. Then

$$\mathbf{K}_l^\pm(f, \omega) = \mathbf{A}_s^\pm(K_s(f, \omega) \pm K_o(f, \omega)) \quad (28)$$

$$\mathbf{K}_r^\pm(f, \omega) = \mathbf{A}_s^\pm(K_s(f, \omega) \mp K_o(f, \omega)) = \mathbf{K}_l^\mp(f, \omega) \quad (29)$$

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, this cell changes its ocular dominance with motion direction since $K_l^+ > K_r^+$ (left-dominant) but $K_l^- < K_r^-$ (right-dominant) by equation (29), which implies that a change of motion direction is equivalent to a change of ocular origin for this cell. Furthermore, consider the left eye for example, while the two motion directions have roughly the same preferred spatial frequency $f \in (f^a, f^{a+1})$, the temporal sensitivity curves are different $K_l^+(\omega) = K_s(\omega) + K_o(\omega) \neq K_l^-(\omega) = K_s(\omega) - K_o(\omega)$. In fact, this example gives the positive motion

direction a low-pass temporal sensitivity while the negative direction a band-pass (Fig.6A), leading to a larger optimal speed for the negative direction, $v_l^- > v_l^+$. Another example (Fig. 6A, 6B) is when $\mathbf{A}_s^+ = \mathbf{A}_o^+ = \mathbf{1}$ and $\mathbf{A}_s^- = \mathbf{A}_o^- = \mathbf{0}$. Both the ocular summation and the opponency channels are completely directional with the same RF phase. By equation (27) the ocular RFs are also completely directional with sensitivities $K_{l,r}^+ = K_s \pm K_o$. This cells 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 speed 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 physiological observations of motion and directional sensitivity in simple cells of the primary visual cortex. A whole spectrum of neural directionality indices are predicted, some neurons have spatiotemporal filters that are spatio-temporally separable and others do not. 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

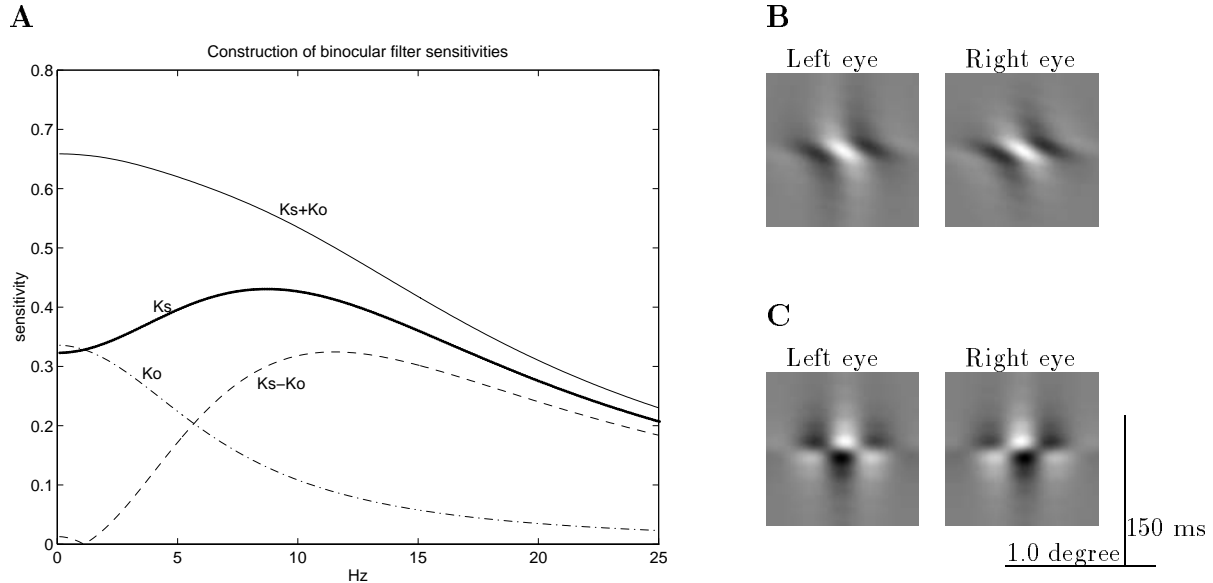


Figure 6: Interaction between motion and stereo coding. **A**: Temporal sensitivity functions for the ocular summation K_s , ocular opponency K_o , $K_s + K_o$, and $|K_s - K_o|$ channels for spatial frequency $f^{peak} = 2$ cycles/degree, which is used in **B** and **C**. Here the binocular correlation 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.

predictions agree with physiological or psychophysical observations (Beverly & Regan 1973, Holub and Morton-Gibson 1981, Foster et al 1985, Reid et al 1991, Poggio 1992, DeAngelis et al 1994). Furthermore, the theory gives additional predictions that have not been experimentally investigated systematically and can be tested. These predictions are: (1) given the optimal spatial frequency and orientation, the neighboring neural units prefer opposite directions of motion, have the same directional index, and have quadrature phase relationship with each other; (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 the two eyes prefer opposite motion directions.

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 by 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* sampling intervals of the visual field in that scale, instead of the *four* filters — a pair of phase quadrature filters for both motion directions — at *each* location (sampling interval) to compute motion energy (Adelson and Bergen 1985); (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.

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. In addition, the significant nonlinearity to facilitate the neural response in the preferred motion directions (Reid et al 1991), although it originates mostly from the generations of action potentials (Jagadeesh et al. 1993), can not be explained by this theory.

This theory has further considerable limitations. The assumption of efficient coding implies that there should be no more cortical cells than the input (the retinal ganglion) cells. An efficient representation (15) should have only one particular parameter set ($q, \theta, \beta, \theta'$), 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). There can be, however, different, but equivalently efficient (up to

the second order correlation), codes in different cortices, each with its own single directionality index and two quadrature filter phases. In reality, however, there are about 40 times as many cortical cells in V1 as retinal ganglion cells (Barlow 1981) and a spectrum of directionality indices in a single cortex (Berman, Wilkes, and Payne 1987, although it is not certain whether neurons tuned to one scale and orientation in a single cortex also have varying directional indices). This theory provides no explanation for the proliferation of cortical cells. In fact, the same limitation existed in earlier works of efficient coding in the multiscale representation applied to spatial, chromatic, and stereo coding (Li and Atick 1994a, b). A single copy of efficient spatial and stereo coding, for instance, should have only two choices of preferred orientations as well as one ocular dominance index and two optimal disparity values for each spatial scale and orientation — this is not true in reality (Hubel and Wiesel 1974, Berardi, Bisti, Cattaneo, Fiorentini, and Maffei 1982). Nevertheless, many of the predictions by this 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 1995a, 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. It is plausible that the primary visual cortex has many different representations, each is itself efficient. Since each efficient representation is complete, it is possible to construct different efficient representations from a single one when necessary. 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) prefer the opposite motion directions, (2) have the same directional index, and (3) to be in quadrature phase relationship, if they (4) have the same optimal spatial frequency and orientation. To simultaneously satisfy these conditions would be difficult if the neural properties were randomly assigned. In addition, the probable need of more filters to compute motion energy could provide one of the explanations for proliferation of cortical cells.

Acknowledgement I wish to thank Edward H. Adelson for very helpful discussions and comments on the draft, and Ning Qian, Christopher Kolb, Wyeth Bair, and Joachim Braun, 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. 1990. Towards a theory of early visual processing. *Neural Comp.*, **2**, 308–320.

- [4] Atick, J. J. and Redlich, A. N. 1992. What does the retina know about natural scenes?
Neural Comp., **4**, 196–210.
- [5] Atick, J. J., Li, Z. and Redlich, A. N. 1992. Understanding retinal color coding from first principles. *Neural Comp.*, **4**, 559–572.
- [6] Barlow, H.B. 1961. Possible principles underlying the transformation of sensory messages *Sensory Communication* ed W. A. Rosenblith (Cambridge, MA: MIT Press).
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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).
- [11] 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.

- [12] 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.
- [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] 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.
- [15] 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.
- [16] 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.
- [17] 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. *J. Neurophysiol.* **46** 1981 1244- 1259
- [18] 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.

- [19] 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.
- [20] 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).
- [21] Kelly D. H. (1979) Motion and vision II. Stabilized spatio-temporal threshold surface. *J. Opt. Soc. Am.* 69, 1340-1349.
- [22] Levinson, E. and Sekuler, R. (1975) The independence of channels in human vision selective for direction of movement. *J. Physiol.* 250, p 347-366.
- [23] Li, Zhaoping and Atick, J. J. 1994a. Towards a theory of the striate cortex. *Neural Comp.*, **6**, 127-146, (1994a).
- [24] Li, Zhaoping and Atick, J. J. 1994b “Efficient stereo coding in the multiscale representation” *Network* Vol.5 1-18. (1994b)
- [25] Li, Zhaoping “Understanding ocular dominance development from binocular input statistics”, presented at Computation and Neural Systems meeting 1994, to appear in *Computation and Neural Systems* Ed. J. Bower, Kluwer Academic Publishers, 1995a.
- [26] Li, Zhaoping “Efficient motion coding in the visual cortex” Submitted to “Machines that learn” workshop in Snowbird, Utah, April, 1995b.

- [27] 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.
- [28] Marr D. and Ullman S. 1981, Directional selectivity and its use in early visual processing. *Proc. R. Soc. Lond. B* 211, 151- 180.
- [29] 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.
- [30] Nakayama K. Biological image motion processing: A review *Vision Res.* Vol. 25, No. 5, pp 625-660. 1985.
- [31] Poggio, G. F. 1992. Physiological basis of stereoscopic vision in J. R. Cronly-Dillon Ed. *Vision and Visual Dysfunction*. Vol. 9 *Binocular Vision* E. Regan Ed. CRC Press, Inc.
- [32] Reichardt W., Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In *Sensory Communication* ed. Rosenblith W. A., Wiley, New York, 1961.
- [33] 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.
- [34] Torre, V., and Poggio T. 1978 A synaptic mechanism possibly underlying directional selectivity to motion. *Proc. R. Soc. Lond. B* 202, 409-416.

- [35] 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.
- [36] 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.
- [37] 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

We verify that the composite code (15) is indeed efficient, by proving that the output $O_n^a(t) = \sum_{x,t'} \mathbf{K}_n^a(x_n^a, t; x, t') S(x, t')$ is decorrelated $\langle O_n^a(t) O_m^a(t') \rangle = \delta_{nm} \delta(t - t')$. For clarity, the following short hand will be used sometimes in the proof: K as $K^a(\theta, \beta)$, K' as $K^a(\theta', \beta')$, \int as $\int \int \dots$, $\int df$ as $\int_{-f^{a+1}}^{-f^a} df + \int_{f^a}^{f^{a+1}}$, and $\langle SS \rangle$ as the correlation matrix between inputs. In addition, normalization constants are ignored when they are not essential. Since $\mathbf{K}_n^a = K_n + (-1)^n q K'_n$,

$$\begin{aligned}
\langle O_n^a(t) O_m^a(t') \rangle &= (\mathbf{K}^a \langle SS \rangle \mathbf{K}^{aT})_{n,t;m,t'} \\
&= (K \langle SS \rangle K^T)_{n,t;m,t'} + (-1)^{n+m} q^2 (K' \langle SS \rangle K'^T)_{n,t;m,t'} \\
&\quad + (-1)^m q (K \langle SS \rangle K'^T)_{n,t;m,t'} + (-1)^n q (K' \langle SS \rangle K^T)_{n,t;m,t'} \\
&= (1 + q^2) \delta_{nm} \delta_{tt'} + q((-1)^m (K \langle SS \rangle K'^T)_{n,t;m,t'} + (-1)^n (K' \langle SS \rangle K^T)_{n,t;m,t'})
\end{aligned}$$

where K (and similarly K' , \mathbf{K}_n^a) is seen here as a matrix with elements $(K)_{n,t,x,t'} = K_n^a(x_n^a, t; x, t')$.

It will be shown below that

$$\begin{aligned}
& (K^a(\theta, \beta) < SS > K^a(\theta', \beta')^T)_{n,t;m,t'} \\
& = \begin{cases} 0 & \text{if } n = m + 2l, \text{ where } l \neq 0 \text{ is an integer} \\ \cos(\Delta\theta) \cos(\Delta\beta) & \text{if } n = m \text{ and } t = t' \\ 2\eta \cos(\Delta\theta) \sin(\Delta\beta) & \text{if } n = m \text{ and } t \neq t' \\ 2\rho \sin(\Delta\theta) \cos(\Delta\beta) & \text{if } n \neq m + 2l, \text{ where } l \text{ is an integer, and } t = t' \\ 4\rho \sin(\Delta\theta) \eta \sin(\Delta\beta) & \text{if } n \neq m + 2l, \text{ where } l \text{ is an integer, and } t \neq t' \end{cases} \quad (30)
\end{aligned}$$

where $\Delta\theta = \theta' - \theta$ and $\Delta\beta = \beta' - \beta$, and ρ and η are constants depending on n, m, t, t' .

Note that $\Delta\beta \rightarrow -\Delta\beta$ and $\Delta\theta \rightarrow -\Delta\theta$ leads to $(K < SS > K'^T)_{n,t;m,t'} \rightarrow (K' < SS > K^T)_{n,t;m,t'}$. Thus

$$\begin{aligned}
& < O_n^a(t) O_m^a(t') > = (1 + q^2) \delta_{nm} \delta_{tt'} \\
& + \begin{cases} 0 & \text{if } n = m + 2l, \text{ where } l \neq 0 \text{ is an integer} \\ (-1)^n 2q \cos(\Delta\theta) \cos(\Delta\beta) & \text{if } n = m \text{ and } t = t' \\ 0 & \text{if } n = m \text{ and } t \neq t' \\ q(-1)^m 4\rho \sin(\Delta\theta) \cos(\Delta\beta) & \text{if } n \neq m + 2l \text{ and } t = t' \\ 0 & \text{if } n \neq m + 2l, \text{ where } l \text{ is an integer, and } t \neq t' \end{cases}
\end{aligned}$$

Decorrelation $< O_n^a(t) O_m^a(t') > = \delta_{nm} \delta(t - t')$ is achieved with $\Delta\beta = \pm\pi/2$.

To prove equation (30),

$$(K^a(\theta, \beta) < SS > K^a(\theta', \beta')^T)_{n,t;m,t'} = (K^a(\theta, \beta) < SS > K^a(\theta', \beta')^\dagger)_{n,t;m,t'}$$

$$\begin{aligned}
&= \int dx dx' d\tau d\tau' K_n^a(x_n^a - x; t - \tau) K_m^{a'}(x_m^a - x'; t' - \tau') R(x - x', \tau - \tau') \\
&= \int df d\omega K_n(f, \omega) K_m(-f, -\omega) R(f, \omega) e^{if(x_n^a - x_m^a) + i\omega(t - t')}
\end{aligned}$$

$K_n(f, \omega)$ can be obtained from equation (12):

$$K_n^a(x_n - x, t - t') = \int df \int d\omega K(f, \omega) e^{i(f(x_n - x) - \text{sgn}(f)(\pi n/2 - \theta) + \omega(t - t') + \phi(f, \omega) + \text{sgn}(\omega)\beta)} \quad (31)$$

where $\text{sgn}(x) = 1$ when $x \geq 0$ and $\text{sgn}(x) = 0$ otherwise. Hence

$$K_n(f, \omega) = K(f, \omega) e^{i(-\text{sgn}(f)(\pi n/2 - \theta) + \phi(f, \omega) + \text{sgn}(\omega)\beta)} \quad (32)$$

for $|f| \in (f^a, f^{a+1})$, and similarly for $K_m(f, \omega)$. Since $K(f, \omega) = R^{-1/2}(f, \omega)$, we have

$$\begin{aligned}
&(K^a(\theta, \beta) < SS > K^a(\theta', \beta')^T)_{n,t;m,t'} \\
&= \int df e^{i(f(x_n^a - x_m^a) - \text{sgn}(f)\pi(n-m)/2 - \text{sgn}(f)\Delta\theta)} \int d\omega e^{i(\omega(t - t') - \text{sgn}(\omega)\Delta\beta)}
\end{aligned}$$

To continue, we note the following. Since U^a , as given in equation (10), is a unitary matrix, hence, $\sum_j U_{nj}^a (U_{mj}^a)^* = \delta_{nm}$. Denoting complex conjugate by *c.c.*,

$$\int df e^{i(f(x_n^a - x_m^a) - \text{sgn}(f)\pi(n-m)/2)} = \int_{f^a}^{f^{a+1}} e^{i(f(x_n^a - x_m^a) - \pi(n-m)/2)} + c.c. \propto \delta_{nm} \quad (33)$$

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

with $j = j^a + 1, j^a + 2, \dots, j^{a+1}$, $N^a = 2(j^{a+1} - j^a)$. Then

$$\begin{aligned}
& \int_{f^a}^{f^{a+1}} df e^{i(f(x_n^a - x_m^a) - \pi(n-m)/2)} \\
&= e^{-i\pi(n-m)/2 + if_{j^a+1}(x_n^a - x_m^a)} \sum_{j=0}^{N^a/2-1} (\exp(i(x_n^a - x_m^a)2\pi/N))^j \\
&= e^{-i\pi(n-m)/2 + if_{j^a+1}(x_n^a - x_m^a)} (1 - \exp(i\pi(x_n^a - x_m^a)N^a/N)) / (1 - \exp(i2\pi(x_n^a - x_m^a)/N))
\end{aligned}$$

Noting that $x_n^a = (N/N^a)n$ or $x_n^a = (N/N^a)(n + n\%2)$, we have

$$\begin{aligned}
& \int_{f^a}^{f^{a+1}} df e^{i(f(x_n^a - x_m^a) - \pi(n-m)/2)} \\
&= \begin{cases} 0 & \text{if } n = m + 2l, \text{ where } l \neq 0 \text{ is an integer} \\ 1/2 & \text{(up to a normalization constant), if } n = m, \\ i\rho & \text{otherwise.} \end{cases}
\end{aligned}$$

The proof is concluded that, for l as a non-zero integer,

$$\begin{aligned}
& (K^a(\theta, \beta) < SS > K^a(\theta', \beta')^T)_{n,t;m,t'} \\
&= \int df e^{-i(f(x_n^a - x_m^a) + \text{sgn}(f)\pi(n-m)/2 + \text{sgn}(f)\theta)} \int d\omega e^{i(\omega(t-t') + \text{sgn}(\omega)\beta)} \\
&= (e^{-i\theta} \int_{f^a}^{f^{a+1}} e^{-i(f(x_n^a - x_m^a) + \pi(n-m)/2)} + c.c.) (e^{i\beta} \int_0^\infty d\omega e^{i\omega(t-t')} + c.c.) \\
&= \begin{cases} 0 & \text{if } n = m + 2l, \\ \cos(\Delta\theta) \cos(\Delta\beta) & \text{if } n = m \text{ and } t = t' \\ 2\eta \cos(\Delta\theta) \sin(\Delta\beta) & \text{if } n = m \text{ and } t \neq t' \\ 2\rho \sin(\Delta\theta) \cos(\Delta\beta) & \text{if } n \neq m + 2l \text{ and } t = t' \\ 4\rho\eta \sin(\Delta\theta) \sin(\Delta\beta) & \text{if } n \neq m + 2l, \text{ and } t \neq t' \end{cases}
\end{aligned}$$