A frequency-resolved mutual information rate and its application to neural systems

Davide Bernardi and Benjamin Lindner

You might find this additional info useful...

This article cites 46 articles, 21 of which can be accessed free at:
/content/113/5/1342.full.html#ref-list-1

Updated information and services including high resolution figures, can be found at:
/content/113/5/1342.full.html

Additional material and information about *Journal of Neurophysiology* can be found at:
http://www.the-aps.org/publications/jn

This information is current as of March 17, 2015.

A frequency-resolved mutual information rate and its application to neural systems

Davide Bernardi[1,2] **and Benjamin Lindner**[1,2]

[1]*Bernstein Center for Computational Neuroscience Berlin, Berlin, Germany; and* [2]*Physics Department, Humboldt University Berlin, Berlin, Germany*

Bernardi D, Lindner B. A frequency-resolved mutual information rate and its application to neural systems. *J Neurophysiol* 113: 1342–1357, 2015. First published December 4, 2014; doi:10.1152/jn.00354.2014.—The encoding and processing of time-dependent signals into sequences of action potentials of sensory neurons is still a challenging theoretical problem. Although, with some effort, it is possible to quantify the flow of information in the model-free framework of Shannon's information theory, this yields just a single number, the mutual information rate. This rate does not indicate which aspects of the stimulus are encoded. Several studies have identified mechanisms at the cellular and network level leading to low- or high-pass filtering of information, i.e., the selective coding of slow or fast stimulus components. However, these findings rely on an approximation, specifically, on the qualitative behavior of the coherence function, an approximate frequency-resolved measure of information flow, whose quality is generally unknown. Here, we develop an assumption-free method to measure a frequency-resolved information rate about a time-dependent Gaussian stimulus. We demonstrate its application for three paradigmatic descriptions of neural firing: an inhomogeneous Poisson process that carries a signal in its instantaneous firing rate; an integrator neuron (stochastic integrate-and-fire model) driven by a time-dependent stimulus; and the synchronous spikes fired by two commonly driven integrator neurons. In agreement with previous coherence-based estimates, we find that Poisson and integrate-and-fire neurons are broadband and low-pass filters of information, respectively. The band-pass information filtering observed in the coherence of synchronous spikes is confirmed by our frequency-resolved information measure in some but not all parameter configurations. Our results also explicitly show how the response-response coherence can fail as an upper bound on the information rate.

information transmission; information filter; neural variability; stochastic spike trains

SENSORY SYSTEMS SEND INFORMATION about external stimuli to the brain in the form of spike trains (Rieke et al. 1996; Borst and Theunissen 1999). Identifying what features of signals are selected by a neuron or neural population is important for understanding the functional physiology of any neural circuit. A basic classification scheme arises in the frequency domain: how much information does a neuron encode about fast or slow components of a stimulus? Put differently, does a neuron select information about high- or low-frequency bands of a time-dependent signal? Such a form of information filtering has been studied in the vestibular (Sadeghi et al. 2007; Massot et al. 2011), auditory (Rieke et al. 1995; Marsat and Pollack 2004), visual (Warland et al. 1997; Reinagel et al. 1999;

Passaglia and Troy 2004), and electrosensory systems (e.g., Chacron et al. 2003; Middleton et al. 2009; Neiman and Russell 2011). Theoretical studies have suggested different mechanisms for information filtering with respect to frequency components of stimuli at the cellular (Stein et al. 1972; Oswald et al. 2004; Vilela and Lindner 2009b; Lindner et al. 2009; Droste et al. 2013) and the network level (Middleton et al. 2009; Sharafi et al. 2013).

Selecting information with respect to frequency is particularly relevant if a neuron is subject to several sensory stimuli with power in different frequency bands, as, for instance, in the electrosensory system of electric fish. Information filtering in this case means that a neuron preferentially encodes information about a specific signal; Chacron et al. (2003) showed how this selectivity may be a primary sensory computation involved in distinguishing between communication signals by conspecifics and environmental signals such as generated by prey. Auditory receptors of grasshoppers encode preferentially lower frequency components of call-like signals (Machens et al. 2001), and also retinal ganglion cells were shown to act as low-pass information filters (Warland et al. 1997). In the vestibular system, sensory afferents can be classified based on the variability of the resting discharge. Sadeghi et al. (2007) point out that the distinction between regularly firing and irregularly firing afferents also corresponds to different information-filtering properties: regular afferents do not exhibit appreciable information filtering with respect to frequency whereas irregular afferents serve as high-pass information filters.

These results are based on an approximate measure of information transfer, the spectral coherence function. For Gaussian input signals, the coherence function can be used to extract a lower bound on the mutual information rate (MIR) (Bialek et al. 1993; Gabbiani 1996). The MIR is a rigorous way of quantifying the information transmission from a stimulus into a spike train without assumptions on the nature of the encoding. Strong et al. (1998) developed a procedure, the so-called "direct method," to estimate the MIR from neural data. Besides some practical issues, large data sets are needed and the numerical calculation is tricky (Paninski 2003), it does not indicate whether slow or fast components are preferentially encoded because it gives just a single number of bits per second. In contrast, the coherence-based lower bound approximation is a function of frequency. This property makes the spectral coherence alluring as a measure of information filtering. However, by using the coherence for the detection of information-filtering effects, researchers have tacitly assumed that the approximation works for all frequency bands equally

Address for reprint requests and other correspondence: B. Lindner, Humboldt Univ. Berlin, Physics Dept., Newtonstr. 15, 12489 Berlin, Germany (e-mail: benjamin.lindner@physik.hu-berlin.de).

well (or equally poorly), such that the filtering effects seen in the coherence are meaningful for the true information transfer. This assumption has never been previously justified. Moreover, it is not clear how to define "information filtering" without resorting to the coherence and the lower bound approximation. Here, we demonstrate the possibility of measuring the MIR in a frequency-resolved way. The numerical scheme we propose is based on the direct method (Strong et al. 1998) but allows us to determine how the number of bits per second the neuron transmits is distributed among different frequency bands of a Gaussian signal.

We apply this scheme to three model systems. We start with an inhomogeneous Poisson process (a broadband filter of information) for which we know analytical results (Bialek and Zee 1990) that we can compare to. For this system, our algorithm gives results in excellent agreement with the theoretical prediction. We then inspect the information filtering in the stochastic leaky integrate-and-fire (LIF) neuron. The LIF model is widely used in theoretical and simulation studies as a model for an integrator neuron, and it was shown that, in the lower bound approximation, it is a low-pass filter of information (Stein et al. 1972; Vilela and Lindner 2009b). Our results confirm the qualitative picture of these previous results, i.e., the LIF model transmits preferentially low-frequency components of a broadband stimulus. We then study information filtering in the spikes fired in synchrony by two uncoupled LIF neurons driven by a common Gaussian stimulus. This setup represents a simple yet general cartoon for coincidence detection, an information-processing scheme relevant both in sensory systems detecting temporal information (Carr 1993) and in higher processing stages (König et al. 1996). In the lower bound approximation, the synchronous output encodes information preferentially on an intermediate frequency range, thus acting as a band-pass filter of information (Middleton et al. 2009; Sharafi et al. 2013). Our findings agree with these previous results in some but not all parameter settings. We also compare our results to another frequency-resolved spectral measure, the response-response (RR) coherence, often used in the experimental literature (Borst and Theunissen 1999; Passaglia and Troy 2004; Marsat and Pollack 2004; Middleton et al. 2006; Krahe et al. 2008; Middleton et al. 2009; Massot et al. 2011; Neiman and Russell 2011; McGillivray et al. 2012) as an upper bound to the mutual information. However, the RR coherence is a true upper bound only under very restrictive assumptions on the nature of the noise in the system and also on the coding scheme (Borst and Theunissen 1999; Roddey et al. 2000). The use of the RR coherence as an upper bound was already criticized by Rozell and Johnson (2004), who illustrated its failure in a simple model (for another example, see Borst and Theunissen 1999). Our results show that the RR coherence can fail as an upper bound on the information transmission in the case of more realistic neuron models. Finally, we verify that our findings do not hinge on the particular choice of the LIF model by comparing it with the biologically more realistic exponential integrate-and-fire (EIF) model. The qualitative picture remains unchanged in all considered cases.

MATERIALS AND METHODS

Established Measures of Information Transmission

Spike trains and spectral measures. We consider a single spike train $x(t)$ encoding a time-dependent stationary input stimulus $s(t)$. A particular neural response is completely characterized by specifying the spike times t_i. A convenient mathematical representation of a spike train $x(t)$ is a sum of Dirac delta functions centered at the spike times (Gerstner and Kistler 2002):

$$x(t) = \sum_{i=1}^{n} \delta(t - t_i). \quad (1)$$

For the analysis of frequency-resolved information transmission Fourier transforms are typically employed, in particular those of the spike train and the signal:

$$\tilde{x}(f) = \int_0^T dt\, x(t)e^{2\pi ift}, \quad \tilde{s}(f) = \int_0^T dt\, s(t)e^{2\pi ift}, \quad (2)$$

where T is the time window. The frequency-dependent correlations between signal and spike train are measured by the coherence function:

$$C_{xs}(f) = \lim_{T \to \infty} \frac{|\langle \tilde{x}(f)\tilde{s}^*(f)\rangle|^2}{\langle |\tilde{x}(f)|^2\rangle\langle |\tilde{s}(f)|^2\rangle}, \quad (3)$$

the angular brackets indicate averaging over repeated trials. The limit $T \to \infty$ means in practice that the measurement time needs to be sufficiently long. The spectral coherence is the squared linear correlation coefficient between stimulus and response in the frequency domain and obeys $0 \leq C_{xs}(f) \leq 1$. For Gaussian signals, the spectral coherence provides a lower bound on the MIR between $s(t)$ and $x(t)$; see *Eq. 7* below. For a linear noiseless system, the coherence is equal to 1 for all frequencies and thus a perfect transmitter of information. Smaller values of $C_{xs}(f)$ can be due to nonlinearities or noise, both of which are present in neurons.

Information theory and the direct method. The information theory by Shannon (Shannon 1948; Papoulis and Pillai 2002) has been applied to quantify in a general and rigorous way the information transfer from an external stimulus into the neural code (Rieke et al. 1996; Strong et al. 1998; Borst and Theunissen 1999). In this section we briefly review first the central information measure of information theory, the MIR, and then the so-called "direct method," a procedure to compute the MIR directly from spike trains (Strong et al. 1998).

In any practical application, spike trains are sampled with a finite time precision. If the time resolution is small enough that at most one spike falls into a single time bin, we deal with bit sequences or "words" w_i of binary symbols (see Fig. 1A). If the probability of each word w_i is p_i, the Shannon entropy of the ensemble of sequences is defined as

$$H(x) = -\sum_i p_i\log_2 p_i, \quad (4)$$

where the sum runs over all possible words. These and other conditional probabilities used below can be in principle estimated from the frequencies of occurrence of the words, i.e., the relative frequencies of all possible bit sequences. If $H(x,T)$ indicates the entropy of a discretized spike train of length T, then the entropy per unit time in the long time limit, i.e., the entropy rate of the output spike train, is:

$$H'(x) = \lim_{T \to \infty} \frac{H(x, T)}{T}. \quad (5)$$

The MIR between s and x is defined as (Shannon 1948; Papoulis and Pillai 2002):

$$I'(s, x) = H'(x) - \langle H'(x|s)\rangle_s, \quad (6)$$

where $H'(x|s)$ is the entropy rate given a particular signal s. Averaging over all possible signals yields the term $\langle H'(x|s)\rangle_s$, i.e., the entropy rate

Fig. 1. Illustration of the direct method (Strong et al. 1998). *A*: a time-dependent signal *s(t)* is encoded into a spike train, identified by the spike times $\{t_i\}$. The encoding is not deterministic, because noise $\eta(t)$ is present in general. Spike trains are converted into binary sequences by introducing a time discretization. Segments of length $T = L\Delta t$ correspond to words of length *L*. In this example, $L = 10$ and the $2^{10} = 1,024$ possible words are labeled from 0 to 1,023, according to the binary number the string represents. IF, integrate-and-fire. *B*: estimation of the spike train entropy. Many stimuli are presented and the occurrences of each word are counted. A good estimate of the probability distribution for large *L* requires a very large data set. The entropy for a fixed word length is then calculated from *Eq. 4*. Finally, the entropy has to be calculated for increasing word length to calculate the entropy rate. Finite-size effects are difficult to treat as discussed in the APPENDIX. *C*: estimation of the noise entropy. The same frozen stimulus is presented multiple times. The entropy rate is then calculated as in *B*. This procedure has to be repeated for different frozen signals and then the average over the entropy rates has to be taken.
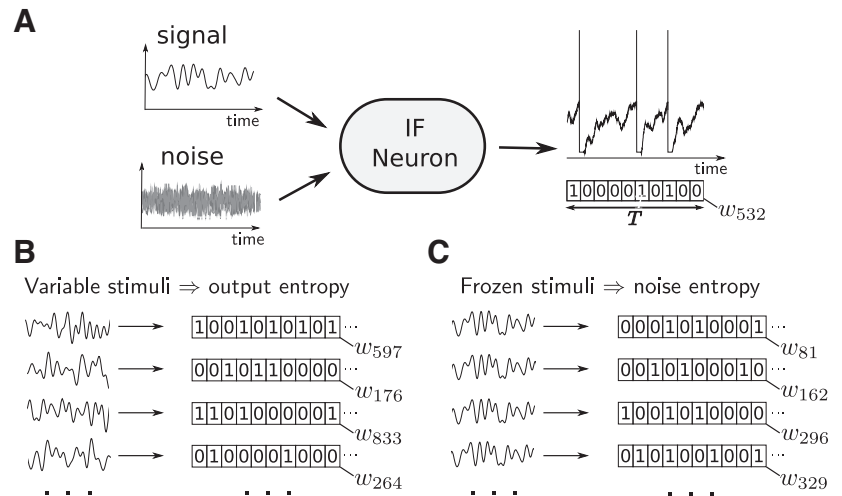
of the output given a frozen stimulus, averaged over the stimulus ensemble (noise entropy rate).

The basic idea of the direct method is simple: first, the probability distribution of all possible output bit sequences of length $T = L\Delta t$ is sampled from a large data set obtained by presenting different input stimuli (Fig. 1*B*). This probability distribution is used to estimate the entropy for different word lengths. The extrapolation for an infinite time window gives the first term in *Eq. 6*. To estimate the noise entropy rate, the same procedure is applied to a data set obtained by the repeated presentation of the same frozen stimulus, as in Fig. 1*C*. This gives $H'(x|s)$, which can be averaged over many frozen stimuli yielding the noise entropy $\langle H(x|s)\rangle_s$, i.e., the second term in *Eq. 6*.

Despite the simplicity of the basic idea, the practical implementation of the direct method is affected by some technical difficulties and requires large data sets. Further details on these issues and on our implementation of this method can be found in the APPENDIX.

Lower bound on the information rate. The direct method requires large data sets and the analytical calculation of the MIR is in most cases intractable. Thus, in many studies, a lower bound on the MIR (Bialek et al. 1993; Gabbiani 1996; Rieke et al. 1996; Borst and Theunissen 1999), which requires the sole calculation of the second-order statistics, is used instead:

$$I'_{LB}(x, s) = -\int_0^{f_c} df \log_2\left(1 - C_{xs}(f)\right). \quad (7)$$

This lower bound on the MIR is valid if the statistics of the input signal are Gaussian and is solely given in terms of the spectral coherence function in *Eq. 3*.

The integrand of *Eq. 7* is sometimes treated as a MIR density and plotted as representing the efficiency of information transfer vs. frequency (see, for example, Rieke et al. 1995; Reinagel et al. 1999; Passaglia and Troy 2004; Marsat and Pollack 2004; Sadeghi et al. 2007; Chacron et al. 2007; Krahe et al. 2008; Massot et al. 2011; Deemyad et al. 2012). We will denote this lower bound density with $i_{LB}(f)$:

$$i_{LB}(f) = -\log_2\left(1 - C_{xs}(f)\right). \quad (8)$$

Otherwise, the spectral coherence itself is used as a measure of information transmission at a particular frequency (for example, by Stein et al. 1972; Chacron et al. 2003; Oswald et al. 2004; Middleton et al. 2006, 2009, 2011; Lindner et al. 2009; Neiman and Russell 2011; McGillivray et al. 2012; Sharafi et al. 2013). This can be done because $i_{LB}(f)$ represents just a monotonic deformation of the spectral coherence and features in the shape of the coherence are reflected in the shape of $i_{LB}(f)$. The accuracy of this lower bound is unclear for most spiking neurons (Rozell and Johnson 2004) and in most applications the use of $i_{LB}(f)$ represents just the only available option. In

particular, it cannot be proven that the goodness of the lower bound approximation is independent of the frequency. Only in this case qualitative features in the lower bound would be meaningful for the true information transfer.

RR coherence. The RR coherence is used in the literature as another measure of information transmission with respect to frequency. The RR coherence is defined as the spectral coherence between two independent responses (indicated with x_1 and x_2) to the same stimulus:

$$C_{x_1x_2}(f) = \lim_{T \to \infty} \frac{\left|\langle \tilde{x}_1(f)\tilde{x}_2^*(f)\rangle\right|^2}{\langle|\tilde{x}_1(f)|^2\rangle\langle|\tilde{x}_2(f)|^2\rangle}, \quad (9)$$

where the same conventions as in *Eq. 3* apply. The RR coherence quantifies the fraction of the output power spectrum that can be reconstructed by means of an optimal (possibly nonlinear) model (Roddey et al. 2000). Note that the RR coherence at a specific frequency is not only affected by the signal power and the system's response at this frequency but also by global features of signal and response, such as overall signal strength and bandwidth. Furthermore, it is in general non-zero outside the frequency band of the signal.

Under the restrictive assumptions the RR coherence can be used to calculate an upper bound on the MIR (Borst and Theunissen 1999):

$$I'_{RR}(x, s) = -\int_0^\infty df \log_2\left(1 - \sqrt{C_{x_1x_2}(f)}\right). \quad (10)$$

In addition to technical assumptions on the statistics of the noise, this formula is a true upper bound on the MIR only if the system encodes information exclusively by the mean conditional response (Borst and Theunissen 1999). It is not clear how this assumption can be justified in general (Rozell and Johnson 2004), but because the RR coherence is used in experimental studies we compare our results to it as well.

Numerical Methods

Generation of input signals. Both the lower bound density formula and the frequency-resolved MIR, which we introduce in the first part of RESULTS, require the use of input signals with Gaussian statistics and a prescribed power spectrum $S_{ss}(f)$. The power spectrum of a signal indicates how its variance is distributed among frequencies and is defined as

$$S_{ss}(f) = \lim_{T \to \infty} \frac{\langle|\tilde{s}(f)|^2\rangle}{T}, \quad (11)$$

where *T* is the duration of the signal. Signals with prescribed power spectrum can be generated in the frequency domain and then transformed back into the time domain (Press et al. 2007). For every point in the frequency domain, the real and imaginary parts of the Fourier

transform of the signal are independently drawn from a Gaussian distribution with variance $T \cdot S_{ss}(f)/2$. This ensures that *Eq. 11* holds.

Generation of spike trains. Spike trains are represented as discrete arrays. In entropy calculations, they can be represented as sequences of any binary symbols (see APPENDIX). In calculation of spectral measures, a spike is represented as $1/\Delta t$, so that the fundamental property of the delta function is preserved and the time integral of a spike train gives the spike count. All neural models were implemented by self-developed C++ routines.

The Poisson neuron. The first considered model is a Poisson process encoding the signal into its time-dependent firing rate. The probability of observing k spikes in a short time interval Δt is given by the Poisson distribution $e^{-r(t)\Delta t}[r(t)\Delta t]^k/k!$ [assuming that the rate $r(t)$ does not vary significantly within the time bin]. If $r(t)\Delta t \ll 1$, the expansion in series can be truncated at the first order so that $r(t)\Delta t$ represents the probability of observing a spike in a time bin. The probability of finding more than one spike in a time bin is of higher order and is neglected. The spike trains can be simulated by generating a uniformly distributed random number $0 < \eta < 1$ at every simulation time step. If $\eta < r(t)\Delta t$, a spike is generated.

In our simulations, the instantaneous firing rate was set to $r(t) = r_0[1 + \varepsilon s(t)]$ where $s(t)$ is a Gaussian random input signal with zero average and unit variance and ε is a nondimensional parameter controlling the signal strength. To have some resemblance with a spike train we use the time resolution $\Delta t = 1$ ms, which is the minimal duration of an action potential. The MIR in this model turns out to be the very small difference of two large numbers, and this leads to problems in the numerical evaluation. Because the MIR is an increasing function of r_0, we chose a high but not completely unrealistic mean firing rate $r_0 = 100$ Hz. This also means that the condition $r(t)\Delta t = 0.1 \ll 1$ is not completely fulfilled, and we expect to see small deviations from the known formulas for the Poisson process. Our model can be considered a Bernoulli process with time-dependent probability or an inhomogeneous Poisson process with "finite-bin-size" corrections. For each frequency band, 100 different partial signals of ~2 s were repeated 75 million times.

The stochastic LIF model. In this simple yet widely used model, the subthreshold dynamics of the membrane potential include only the voltage-independent membrane leak conductances (for reviews, see Fourcaud and Brunel 2002 and Burkitt 2006). The subthreshold dynamics are governed by:

$$\tau_m \dot{v} = -v + \mu + \sqrt{2Dc}\xi_s(t) + \sqrt{2D(1-c)}\xi_n(t). \quad (12)$$

The model is complemented by the fire-and-reset rule:

$$v(t) = v_T \Rightarrow \begin{cases} \text{spike at } t \\ v(t + \tau_{abs}) = v_R \end{cases}$$

where τ_{abs} is the absolute refractory period. The constant μ in *Eq. 12* sets the resting potential of the model. The independent Gaussian processes $\xi_s(t)$ and $\xi_n(t)$ denote input signal and the neuronal noise, respectively. Both $\xi_s(t)$ and $\xi_n(t)$ have zero average, unit noise intensity, and a flat power spectrum; the parameter D sets the overall noise intensity. While for $\xi_s(t)$ a cutoff frequency of 500 Hz was chosen, for $\xi_n(t)$ no cutoff frequency was used. The parameter c is constrained to $0 \leq c \leq 1$ and represents the relative strength of the signal. We use nondimensional units for the membrane potential v and choose $v_T = 1$ for the threshold and $v_R = 0$ for the reset point (Vilela and Lindner 2009a).

A stochastic Euler procedure was used to integrate *Eq. 12*. The integration time step was set to a hundredth of the membrane time constant τ_m, for which the biologically plausible value of 10 ms was chosen. The integration time step was always kept at 0.1 ms independently of the choice of the time resolution for signal and spike trains. To approximate the stationary state, the initial value for the voltage was set equal to the last value of the previous trial. For the first trial, a random value between zero and one was drawn. Although the stationary voltage distribution is not uniform, the influence of this

initial condition extends only to the first interspike intervals of the first spike train, where the total number of interspike intervals generated in a simulation ranges from 10^8 to 10^{10}. The noise entropy was calculated by averaging on 105 different partial signals of ~2.4 s. Each signal was repeated 7 to 14 million times, about ten times fewer than in simulations of the Poisson model.

EIF model. The EIF model (Fourcaud-Trocmé et al. 2003) differs from the LIF model in the subthreshold dynamics:

$$\tau_m \dot{v} = -v + \Delta_T \exp\left(\frac{v-1}{\Delta_T}\right) + \mu + \sqrt{2Dc}\xi_s(t)$$
$$+ \sqrt{2D(1-c)}\xi_n(t). \quad (13)$$

The additional exponential term models the dynamics of sodium channels. The additional parameter Δ_T indicates the steepness of the activation, and for $\Delta_T \to 0$ the EIF model converges to an LIF model. We chose $\Delta_T = 0.1$, which is close to the value used by Badel et al. (2008), to reproduce the voltage traces of a pyramidal cell stimulated in vitro. In the EIF model, the spike is defined as a divergence of the voltage variable. In numerical simulations, the integration is interrupted at a finite value v_T and reset. We used $v_T = 2$ and repeated each signal of 3.3 s 12 million times.

Coding by synchrony. Two uncoupled stochastic LIF or EIF models driven by a shared Gaussian white noise input are simulated as described in the two previous paragraphs, and then the spikes fired in synchrony are extracted. To this end, one of the two spike trains was chosen as a reference spike train and a small coincidence time window τ_s was used as the coincidence criterion (Middleton et al. 2009). For every spike at t_i in the reference spike train, a function looks into the other spike train. If a spike is also present in the time bins within $[t_i \pm \tau_s/2]$, a synchronous spike is registered. All spikes used to generate this synchronous spike are then deleted, so that repeated assignments of the same spike of the reference spike train are prevented. The size of τ_s influences the fraction of spikes of the reference spike train that is kept in the synchronous output. To have a good trade-off between coherence height (enhanced by a large τ_s) and peakedness (enhanced by a small τ_s), the standard choice for the simulations was $\tau_s = 5$ ms. This value is also not very different from the typical duration of a postsynaptic potential generated by a fast excitatory synapse (Koch 1999), and it is therefore reasonable as a choice for the integration timescale for the generation of the synchronous output. For each frequency band of each simulation, 105 different partial signals (length 2 to 3 s) were repeated 11 to 15 million times.

Numerical calculation of entropy rates. Entropy rates are calculated from the spike trains via the direct method (Strong et al. 1998). Further details on our implementation of this method can be found in the APPENDIX.

RESULTS

A Frequency-Resolved Mutual Information

In the following, we present a scheme to measure information transmission in neural models driven by a Gaussian stimulus. It possesses the generality of the direct method, and it is frequency resolved as the lower bound derived from the spectral coherence. This scheme is only applicable if the stimulus ensemble is Gaussian, but this limitation is also valid for the lower bound formula and Gaussian noise stimuli are very frequently used in experimental and theoretical applications.

Any continuous time function (signal) lasting a time T and containing frequencies up to a cutoff frequency f_c can be represented by the $2Tf_c$ real-valued coefficients of its decomposition as a Fourier series. For a Gaussian process, all Fourier components are statistically independent and the joint distribution factorizes (Stratonovich 1963) for a sufficiently long time window. The

entropy of the ensemble is therefore the sum of the entropies of the distributions of each component. In other words, the frequency components of the signal can be considered independent sources of information. By exploiting this property, one can think of the signal as of a compound of many independent partial signals

$$s(t) = \sum_k s_k(t) \qquad (14)$$

and separately measure the MIR between each partial signal $s_k(t)$ and the spike train $x(t)$ while all other partial signals $s_{k'}(t)$ with $k' \neq k$ are present. To this end, the entropy rate of the response is computed keeping the vector components belonging to the frequency band $[f_k - \Delta f_{BW}/2, f_k + \Delta f_{BW}/2]$ of the partial signal $s_k(t)$ frozen, while the rest [those associated with $s_{k'}(t)$, $k' \neq k$] is randomly varied (Fig. 2A).

If averaging over many partial signals is performed, a frequency-dependent partial noise entropy rate is obtained:

$$N_k = \langle H'(x|s_k) \rangle_{s_k}. \qquad (15)$$

If this partial noise entropy rate is subtracted from the full entropy rate, we obtain the MIR between the full output and a subset of the input signal, the frequency band Δf_{BW} centered around f_k. To compare this measure to the lower bound density defined in *Eq. 8*, we divide the frequency-resolved MIR by the partial stimulus bandwidth:

$$i_k = \frac{H'(x) - N_k}{\Delta f_{BW}}. \qquad (16)$$

The shorthand name of MIR density will be used in the following to indicate the MIR per unit frequency *Eq. 16*, in analogy to the lower bound density $i_{LB}(f)$. The MIR density i_k is bounded from below by $i_{LB,k}$, the lower bound density averaged in the k-th frequency band

$$i_{LB,k} = \frac{-1}{\Delta f_{BW}} \int_{f_k - \Delta f_{BW}/2}^{f_k + \Delta f_{BW}/2} df \log_2(1 - C_{xs}(f)). \qquad (17)$$

Note that the range of integration $(0, f_c)$ is replaced by the limits of the k-th frequency band, because outside of it the

partial signal has no power and therefore the SR coherence is zero by definition. The difference between i_k and $i_{LB,k}$ indicates how good the coherence-based approximation is within the k-th frequency band. To compare i_k to the information rate estimate based on the RR coherence, we define

$$i_{RR,k} = \frac{-1}{\Delta f_{BW}} \int_0^\infty df \log_2\left(1 - \sqrt{C_{x_1 x_2}(f)}\right). \qquad (18)$$

Here x_1 and x_2 are two independent responses to the same partial signal $s_k(t)$ and not to the total signal $s(t)$. The infinite integration boundary in *Eq. 18* reflects the fact that information about the signal from the k-th frequency band can be nonlinearly encoded in all frequency bands of the output. Furthermore, note that in particular the integrand of *Eq. 18* may drastically differ from the commonly plotted RR coherence obtained by using the total signal.

As an illustration of how the method of decomposing the signal in effective signal and effective noise works in practice, consider a flat spectrum with cutoff frequency $f_c = 100$ Hz and take the frequency band 0–20 Hz as effective signal (Fig. 2). In the first realization all coefficients are randomly generated as described above. All coefficients of the part of the spectrum chosen as signal are stored and reused in the next realizations, while the rest of the Fourier transform is randomly generated from scratch. Four realizations of this procedure are plotted in Fig. 2B. The partial $i_{LB,k}$ signal (common part to all stimuli) is plotted in red.

Both i_k and $i_{LB,k}$ will in general depend on Δf_{BW}, the bandwidth of the partial signals, i.e., the frequency resolution of our method. This Δf_{BW} can be adapted to the features shown by the coherence function. However, practical difficulties limit Δf_{BW}. The principal reason is that Δf_{BW} is inversely proportional to the correlation time of the effective signal.[1] The signal correlation time has also an impact on the correlation time of

---

[1] One possible definition of the correlation time is the integral of the squared normalized correlation function $\tau_{corr} = \int_0^\infty d\tau C^2(\tau)/C^2(0)$, where the autocorrelation function $C(\tau) = \langle x(t + \tau)x(t)\rangle - \langle x(t)\rangle^2$ is connected to the power spectrum by the Wiener-Khinchin theorem. With this definition of correlation time one finds $\tau_{corr} = 1/(2\Delta f_{BW})$ for a box-shaped spectrum.
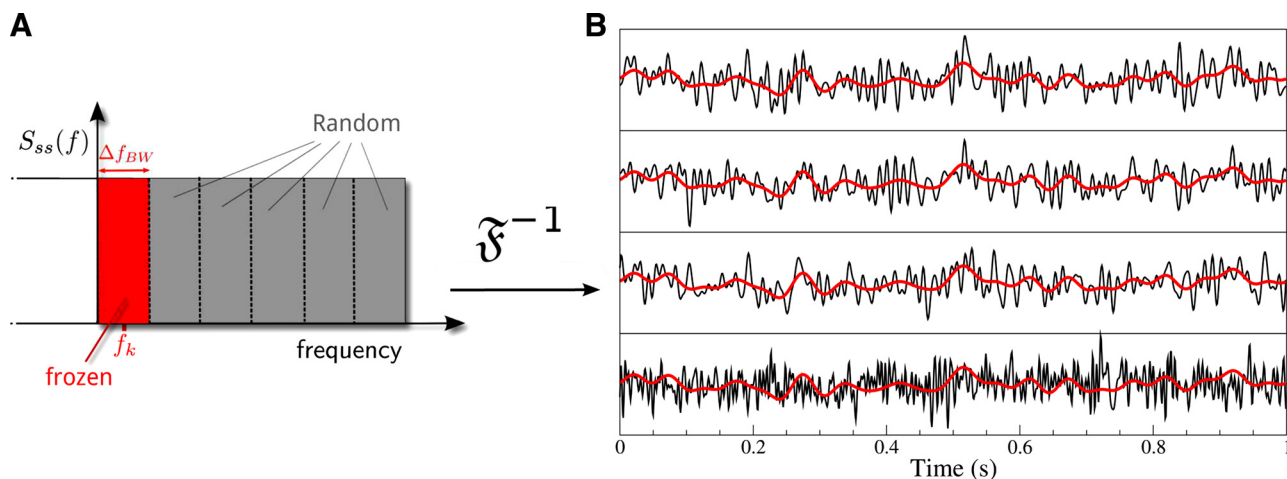

Fig. 2. A Gaussian input signal can be regarded as a superposition of independent frequency components. All components in the considered frequency band $[f_k - \Delta f_{BW}/2, f_k + \Delta f_{BW}/2]$ form the partial stimulus $s_k(t)$ (repeatedly presented to the system), while the rest (grey) is treated as effective noise (randomly varied at each trial). *A*: visualization of the procedure in the frequency domain. *B*: visualization of the procedure in the time domain. Four realizations of band-limited Gaussian white noise, cutoff frequency of 100 Hz (black thin lines). The frequency band 0–20 Hz is kept frozen so that the partial signal $s_k(t)$ (red thick line) is repeatedly presented.

the output spike trains. Longer correlation times in the output also require longer time windows in the extrapolation step of the direct method. Longer time windows require larger data sets and this means that a narrow Δf_{BW} causes the computational cost to grow significantly. A narrow Δf_{BW} also implies that N_k is very close to $H'(x)$, so that the relative error on i_k is larger. Because of these practical limitations, 30 Hz was the smallest Δf_{BW} we used in the simulations. In addition, on the functional level it is plausible that neural systems in many cases make only a coarse distinction between fast and slow stimuli and cannot separate different frequency components down to the single Hertz.

Frequency-resolvable and synergetic parts of the MIR. In general, the sum of the single $i_k\Delta f_{BW}$ (i.e., the information rates for the single frequency bands) is not equal to the total MIR I'. This means that there is a limit to the fraction of the MIR that can be ascribed to single frequency bands. This fraction depends on the bandwidth of the partial signals

$$\alpha(\Delta f_{BW}) = \frac{\sum_k i_k\Delta f_{BW}}{I'}. \quad (19)$$

The trivial case of a single frequency band or the nontrivial case of a Gaussian channel both imply $\alpha = 1$, i.e., the total information rate can be resolved with respect to the frequency bands. In general, for a nonlinear system $0 \leq \alpha \leq 1$. If $\alpha < 1$, the missing information is encoded in the synergy between different bands and cannot be ascribed to any particular band. Therefore, the difference $I' - I'_{LB}$ between the total MIR and the lower bound is made up of two contributions: the intraband discrepancy between the lower bound and the frequency-resolved MIR

$$\beta(\Delta f_{BW}) = \frac{\sum_k (i_k - i_{LB,k})\Delta f_{BW}}{I' - I'_{LB}}, \quad (20)$$

and the synergy between bands

$$\gamma(\Delta f_{BW}) = \frac{I' - \sum_k i_k\Delta f_{BW}}{I' - I'_{LB}}. \quad (21)$$

Both γ and β are normalized with respect to the difference $I' - I'_{LB}$, such that $\beta + \gamma = 1$.

Frequency-Resolved Information Transfer of the Poisson Model

Before studying systems whose nonlinear encoding properties are unknown, all algorithms were tested by considering a model for which analytical expressions are available. In the time coding framework (the whole discretized spike train identifies the response of the system), a suitable result is a formula for a Poisson neuron model with firing rate modulated by Gaussian noise, derived by Bialek and Zee (1990). The result is expressed as an expansion in series whose leading term is an integral over the signal power spectrum and coincides with the lower bound in *Eq. 7*.

Although the Poisson process keeps no memory of its own history, the time bins forming the system response are in general correlated because of the signal autocorrelation and this makes the analytical calculation of entropy rates difficult. However, if the signal cutoff frequency equals the maximum resolvable frequency $1/(2\Delta t)$, the signal is effectively white

and consecutive time steps are completely uncorrelated. In this special case, the time dependence can be ignored. This reduces the system to a Gaussian random number with uniform threshold, a simplified version of the setup of Stocks (2000).

Given a particular signal value s, the probability of observing a spike is $p_{1|s} = r_0\Delta t(1 + \varepsilon s)$ if $s \epsilon [-\varepsilon^{-1}, (1 - r_0\Delta t)/(r_0\Delta t \varepsilon)]$, while $p_{1|s} = 0$ and $p_{1|s} = 1$ if s is left or right from this interval, respectively (see *Generation of spike trains*). Furthermore, the probability of no spike is $p_{0|s} = 1 - p_{1|s}$. The MIR can be then expressed as:

$$I'(x, s) = \frac{1}{\Delta t}\int_{-\infty}^{+\infty} ds \left[p_{1|s}\log_2\left(\frac{p_{1|s}}{p_1}\right) + p_{0|s}\log_2\left(\frac{p_{0|s}}{p_0}\right) \right]\frac{e^{-s^2/2}}{\sqrt{2\pi}}, \quad (22)$$

where p_1 and p_0 indicate the probability of the presence or absence of a spike regardless of the signal value, respectively. They can be explicitly calculated by using that $p_0 = \int ds p_{0|s} e^{-s^2/2} ds$

$$p_0 = \int_{-\infty}^{-1/\varepsilon} ds 1 \cdot \frac{e^{-s^2/2}}{\sqrt{2\pi}} + \int_{-1/\varepsilon}^{\frac{1-r_0\Delta t}{r_0\Delta t\varepsilon}} ds(1 - r_0\Delta t(1 + \varepsilon s))$$

$$\frac{e^{-s^2/2}}{\sqrt{2\pi}} + \int_{\frac{1-r_0\Delta t}{r_0\Delta t\varepsilon}}^{+\infty} ds 0 \cdot \frac{e^{-s^2/2}}{\sqrt{2\pi}}$$

$$= \frac{1}{2}\text{erfc}\left(\frac{r_0\Delta t - 1}{\sqrt{2}r_0\Delta t\varepsilon}\right)(1 - r_0\Delta t) + \frac{1}{2}r_0\Delta t\,\text{erfc}\left(\frac{1}{\sqrt{2}\varepsilon}\right)$$

$$+ \frac{r_0\Delta t\varepsilon}{\sqrt{2\pi}}\left(\exp\left[-\frac{1}{2}\left(\frac{1-r_0\Delta t}{\varepsilon r_0\Delta t}\right)^2\right] - \exp\left[-\frac{1}{2}\varepsilon^{-2}\right]\right) \quad (23)$$

where we used the definition of the complementary error function (p_1 can be determined from $p_1 = 1 - p_0$). For an arbitrary value of the signal strength parameter ε, the MIR *Eq. 22* can be integrated numerically. For $\varepsilon \to 0$, the boundaries of the middle interval of integration go to infinity and we can expand all logarithms and integrate the resulting series term by term. To the lowest significant order in ε one finds:

$$I'(x, s) \approx \frac{r_0\varepsilon^2}{2\ln 2}\left(1 + \frac{r_0\Delta t}{(1 - r_0\Delta t)}\right), \quad (24)$$

which reduces to the result for the Poisson process by Bialek and Zee (1990) in the limit $r_0\Delta t \to 0$ (i.e., if our Bernoulli process approaches a true Poisson process). While *Eqs. 22* and *24* rest on the assumption of a completely white signal, the derivation by Bialek and Zee (1990) does not make assumptions on the temporal structure of the signal. Therefore, we expect

$$i_k \underset{\varepsilon\to 0}{\searrow} i_{LB,k} \quad (25)$$

and $i_k \approx i_{LB,k}$ for any weak Gaussian signal.

The situation of a weak signal was considered first. The signal strength parameter was set to the value of $\varepsilon = 0.2$. If the bandwidth of the effective stimulus spans the entire frequency range up to the Nyquist frequency, we can compare the MIR calculated via the direct method, which was 3.259 ± 0.02

bits/s, to the value resulting from numerical evaluation of the analytical expression *Eq. 22* of 3.269 bits/s. The two values are thus consistent with each other.

The signal was then split in 10 partial signals to measure the MIR in a frequency-resolved way. In Fig. 3*A* the relevant information measures are plotted. In this case $i_{RR,k}$ was omitted because it perfectly overlaps with $i_{LB,k}$. It can be seen that $i_k \approx i_{LB,k}$ as expected. In particular, we see that the Poisson neuron does not introduce any information filtering; all frequency components are equally well encoded and we can regard the Poisson process as a broadband information filter.

If the MIRs of the separate frequency bands are added up, we obtain the value $\sum_k i_k \Delta f_{BW} = 3.233 \pm 0.024$ bits/s, which is slightly lower than the total MIR found in the previous simulation, which used the whole frequency range as bandwidth of the effective stimulus. Discussion of this is postponed to the case of a stronger signal, in which this effect is more evident.

Another simulation was performed with a Gaussian "colored" noise: a modified Ornstein-Uhlenbeck process with a sharp cutoff at $f = f_c$, whose power spectrum is

$$S_{ss}(f) = \begin{cases} \dfrac{2D}{1 + (2\pi\tau f)^2} & \text{if } |f| < f_c \\ 0 & \text{if } |f| > f_c \end{cases}. \quad (26)$$

Here τ is the correlation time (which was set to 1 ms) and D is the noise intensity, which was set equal to $\pi\tau \cdot [2\arctan(2f_c\pi\tau)]^{-1}$ so that $s(t)$ has unit variance. In this case, the low-pass shape of the signal power spectrum is also reflected by $i_{LB}(f)$ (Fig. 3*B*). For both simulations it can be seen that $i_k = i_{LB,k}$ within the error bars as expected from *Eq. 25*.

A stronger white stimulus ($\varepsilon = 0.8$) is used in the next simulation while the other parameters are unchanged. Integration of the exact expression *Eq. 22* gave a value of 47.72 bits/s for the total MIR, which agrees with the result of the numerical simulation 47.65 ± 0.18 bits/s. Integration of the lower bound yields 40.73 bits/s. The variance of the total signal is large
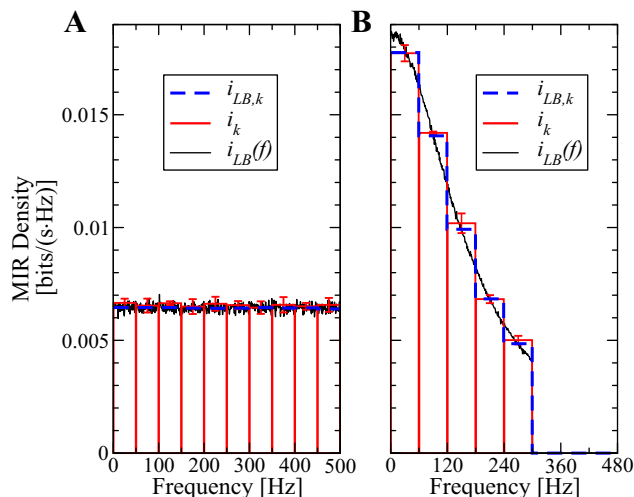

Fig. 3. The results for a Poisson neuron confirm the predictions of the theory. In *A*, the signal has a flat spectrum while in *B* the signal power is low-pass filtered and there is a sharp cutoff at f_c. In both cases, the mutual information rate (MIR) density is equal to the lower bound, as expected in case of a weak simulation (see text). Parameters: *A*: $\varepsilon = 0.2$, $f_c = 500$ Hz; *B*: $\varepsilon = 0.2$, $f_c = 300$ Hz, $\tau = 1$ ms, $D = 1.45 \times 10^{-3}$ s.

(0.64), and the lower bound is not as tight as in the previous case.

Next, the stimulus was split in five frequency bands. For each of the partial signals, the lower bound is much tighter than for the global one (not shown). Also, in this case the sum of the MIRs of the single frequency bands is smaller than the MIR of the total signal, a synergetic effect: other frequency bands are always present, but using them as signal rather than as noise increases the information transmission rate beyond the sum of the signal components, even if these signal components are independent (and their entropy is additive). This can be explained by noting that the unconditioned signal components are independent but not necessarily independent if conditioned on the spike train, as we show below. The MIR between the signal s and the spike train x can be defined in three ways (Shannon 1948; Papoulis and Pillai 2002):

$$I'(s, x) = H'(x) - \langle H'(x|s) \rangle_s = H'(s) + H'(x) \\ - H'(s, x) = H'(s) - \langle H'(s|x) \rangle_x, \quad (27)$$

The direct method is based on the first definition used until this point. However, in the following discussion of synergetic effects, it is more convenient to use the third definition in *Eq. 27*: the MIR defined as the entropy of the signal minus the entropy rate of the signal given a spike train, averaged over spike trains. Consider now the situation of a signal that can be decomposed into just two partial signals s_A and s_B, so that $s = (s_A; s_B)$. In this case, the MIR between the signal and the output can be written as:

$$I'(x, s) = H'((s_A; s_B)) - \langle H'((s_A; s_B)|x) \rangle_x = H'(s_A) \\ + H'(s_B) - \langle H'((s_A; s_B)|x) \rangle_x \quad (28)$$

where in the second equality we used that s_A and s_B are independent. Nothing can be said in general about the conditional independence, so that from the general property of the joint entropy (Shannon 1948) it follows

$$\langle H'((s_A; s_B)|x) \rangle_x \leq \langle H'(s_A|x) \rangle_x + \langle H'(s_B|x) \rangle_x$$

and inserting this into *Eq. 28* yields

$$I'(x, s) \geq I'(x, s_A) + I'(x, s_B). \quad (29)$$

This effect is visualized in Fig. 4*A*. The area of each colored bar represents the MIR of a single signal of bandwidth ranging from 0 to Δf_{BW}, while the dashed bars are the MIRs for each partial signal of bandwidth 100 Hz. It is clear that the area of each colored signal is larger than the sum of the partial signals in the same frequency range. These results also show why the MIR density should be intended just as a shorthand name for MIR per unit frequency, because it lacks the fundamental property of a true density. If the data of Fig. 4*A* are integrated with respect to frequency, the data shown in Fig. 4*B* are obtained. It can be seen that the MIR of each colored signal (blue squares) grows more rapidly than the cumulative sum of each partial signal (red circles), which remains very close to the lower bound (black crosses).

Frequency-Resolved Information Transfer of the Integrate-and-Fire Model

We now turn to the simple but biophysically more realistic LIF model. First numerical calculations of the coherence function for a LIF model stimulated by white Gaussian noise go
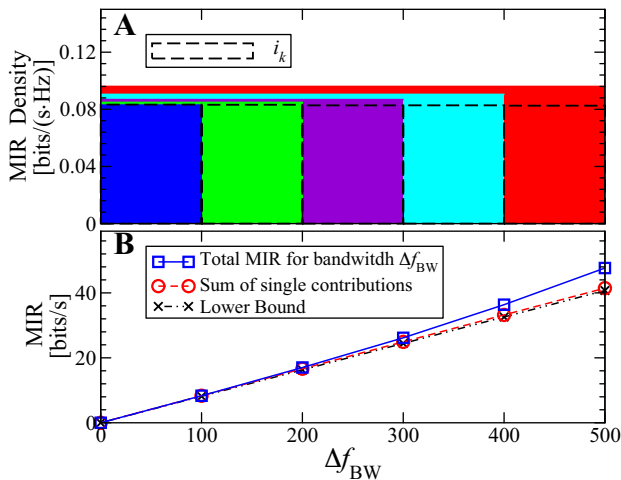
Fig. 4. Illustration of how the interaction between different bands carries synergetic information. *A*: MIR density i_k for each partial signal (black dashed line) of bandwidth 100 Hz and for a single band spanning the frequency range up to the *k*-th band (boxes of various colors). Each single band has an area (the MIR) that is larger than the sum of the areas of the partial signals within the same frequency band. *B*: squares represent the MIR for a single band spanning the frequency range up to the *k*-th band (integral of the boxes of various colors from *A*). Circles represent the cumulative sum of the MIR for single partial signals of bandwidth 100 Hz in the same frequency range (integrated MIR from the dashed boxes in *A*). Parameters: $\varepsilon = 0.8$, $f_c = 500$ Hz.

back to the 1970s (Stein et al. 1972); for a more recent study, see Vilela and Lindner (2009b). These studies showed that the coherence, and hence $i_{LB}(f)$, attain a global maximum at zero frequency. This seems to be so regardless of whether the neuron fires irregularly like a Poisson process or regularly like a pacemaker cell. To test if this is also true for the frequency-resolved MIR, we use extensive simulations of the model (typically 10^{10} spikes or more are generated) to determine the information rate of each partial signal of bandwidth $\Delta f_{BW} = 50$ Hz. We first consider a bias current $\mu < 1$ such that the deterministic model ($D = 0$) would not fire at all and in the stochastic model, consequently, all spikes are elicited by fluctuations (noise-driven regime).

The results of the simulations in Fig. 5*A1* show that $i_k \approx i_{LB,k}$ for moderate signal intensity, indicating that the lower bound approximation works very well. In the simulation shown in Fig. 5*A2*, the firing regime is the same, but the signal is very strong. It can be seen that at low frequencies the information transfer is ~20% larger than the lower bound, while at high frequencies the match is still close. In particular, a low-pass information filtering is still observed, being even more pronounced than in the lower bound approximation. Considering the high signal-to-noise ratio in the input ($c = 0.98$), the agreement of the lower bound with the true frequency-resolved MIR is still surprisingly good. The RR estimate of the information rate, $i_{RR,k}$, shows in all cases the same low-pass shape as the lower bound density $i_{LB,k}$. For moderate signal strength, $i_{RR,k}$ is close to the lower bound, whereas for strong signal amplitude it exceeds both the lower bound and the true MIR density, specifically at low frequencies.

Next, we study the effects of changing the time resolution used to sample the spike train and the input signal (not the time step used in the simulation of the neuron, which was always kept constant in all simulations). In Fig. 5*B1* the sampling rate for spike trains and signals was $\Delta t = 1$ ms, whereas in Fig. 5*B2*

the sampling rate was set to $\Delta t = 2$ ms. To prevent more than one spike to fall into the same time bin, a refractory period $\tau_{abs} = 2$ ms was introduced in both simulations; in addition, we chose a value of the signal strength that was in between the values used in Fig. 5, *A1* and *A2*. On closer inspection it can be seen that both lower bound and MIR density are slightly lower for the coarser time resolution (Fig. 5*B2*), but the two plots in Fig. 5, *B1* and *B2*, are barely distinguishable. This suggests that for this parameter choice the system is making little use of the timing precision on the millisecond scale. The plot in Fig. 5, *B1* and *B2*, also serves a consistency check: doubling the time resolution allows to use time windows for the extrapolation to infinite time, which is also almost twice as large. The fact that the results are consistent with each other represents a further control on the validity of the used numerical procedure.

Another question to explore is whether an interaction between different subsignals is present in the LIF model, similar to the one we observed in the case of the Poisson neuron. In Fig. 6, *A1* and *B1*, the bandwidth of one subsignal is progressively increased from 50 to 300 Hz. Other parameters are as in Fig. 5, *A1* and *A2*, respectively. The MIR of this single signal is plotted with blue squares as a function of its bandwidth Δf_{BW}. The red circles represent the cumulative sum of the MIR of the subsignals covering the same frequency range (from 0 to Δf_{BW}). This can be obtained from the integration of the MIR density i_k in Fig. 5, *A1* and *A2*, respectively. The black crosses correspond to the integral of $i_{LB}(f)$ in the range from 0 to Δf_{BW}. For a moderate signal (Fig. 6*A1*), the synergetic effects are very small and all curves are very close to each other. For a strong signal (Fig. 6*B1*), the synergetic effects become evident: the total MIR (blue squares) grows more rapidly than the cumulative sum of the MIR (red circles). The difference between the lower bound and the cumulative frequency-resolved MIR is mostly due to the lowest frequency band. Note that the ratio between the total MIR and the lower bound for the total range up to 300 Hz is ~1.4, a value on the same order
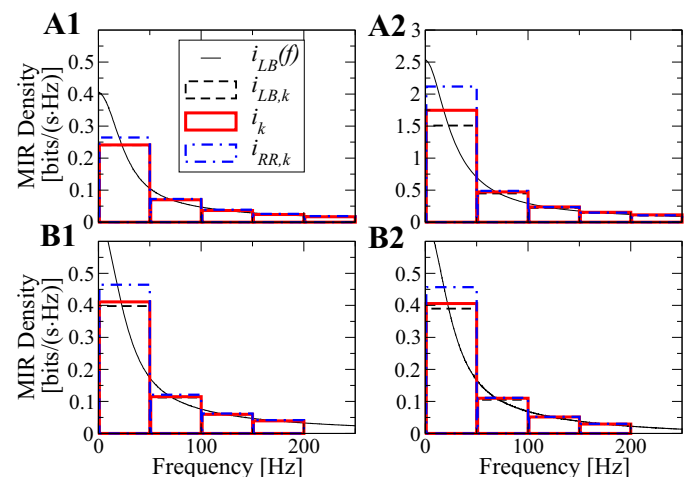

Fig. 5. Information transmission in a single leaky integrate-and-fire (LIF) model in the subthreshold regime is well described by the coherence-based approximation and does not depend much on time resolution. *A1*: weak signal. *A2*: strong signal. *B1*: intermediate signal strength, fine time resolution. *B2*: same as *B1* with coarser time resolution. Parameters: $\mu = 0.75$; *A1* and *A2*: $\tau_{abs} = 0$ ms; *B1* and *B2*: $\tau_{abs} = 2$ ms; *A1*: $D = 6 \times 10^{-4}$ s; *A2*: $D = 3.3 \times 10^{-3}$ s; *B1* and *B2*: $D = 7.6 \times 10^{-4}$ s; *A1*: $c = 0.34$; *A2*: $c = 0.98$; *B1* and *B2*: $c = 0.53$; *A1*, *A2*, and *B1*: $\Delta t = 1$ ms; *B2*: $\Delta t = 2$ ms. Error bars are within line thickness.
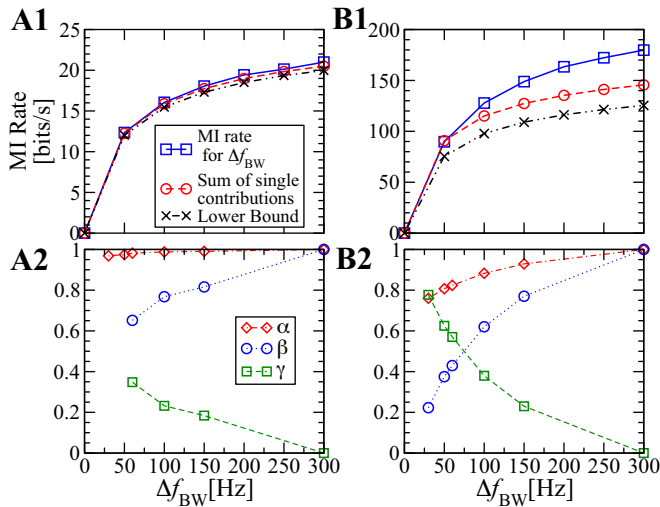
Fig. 6. Coding synergy in the LIF model. *A1* and *B1*: MIR for a total, containing all partial signals with lower frequency, signal (blue squares) and cumulative sum of partial signals (red circles). *A2* and *B2*: fraction of frequency-resolvable information α, fraction of the difference between full MIR and lower bound $I' - I'_{LB}$, which is explained by the intraband difference β and the synergy between bands γ. Parameters: $\mu = 0.75$, $\tau_{abs} = 0$ ms; *A*: $D = 6 \times 10^{-4}$ s; *B*: $D = 3.3 \times 10^{-3}$ s; *A*: $c = 0.34$; *B*: $c = 0.98$.

of magnitude observed in experimental studies (see Table 2 in Aldworth et al. 2011).

An important issue is how much of the total MIR can be resolved with respect to frequency. This fraction α, defined in *Eq. 19*, is shown in Fig. 6, *A2* and *B2*, by diamonds. In the case of a moderate signal (Fig. 6*A2*) the fraction is always above 97% ($\alpha > 0.97$). For a strong signal (Fig. 6*B2*), synergy is higher and α drops. Still, the fraction of frequency-resolvable MIR is always larger than 3/4 and is ~80% for $\Delta f_{BW} = 50$ Hz (the case plotted in Fig. 5*A2*). The difference between the full MIR I' and the full lower bound I'_{LB} consists of two contributions: on the one hand the sum of the differences between the frequency-resolved MI density i_k and the lower bound $i_{LB,k}$, on the other hand the synergetic information between frequency bands. We express these two contributions in terms of β and γ, defined in *Eqs. 20* and in *21* and plotted in Fig. 6, *A2* and *B2*, by circles and squares, respectively. For small frequency bandwidth and strong signal (Fig. 6*B2*), β seems to tend to zero (for weak stimuli, used in Fig. 6*A2* the values of β and γ at small Δf_{BW} are unreliable and therefore not shown). This would imply that for this model and in this limit the frequency-resolvable part of the MIR converges to the lower bound and the discrepancy between the lower bound and the total MIR is solely due to synergetic information. We note that the variance of the partial signal decreases with decreasing Δf_{BW} and thus in the limit $\Delta f_{BW} \to 0$ the assumptions of linear response theory for the transmission of one partial signal are perfectly matched (Vilela and Lindner 2009b) (the transmission of the total nonweak stimulus is still nonlinear).

So far, we have only considered the noise-induced firing regime of the stochastic LIF model. Results in the suprathreshold regime ($\mu > 1$) are very similar to the above findings. We have tested two different values of the baseline current $\mu = 1.3$ and $\mu = 2$ and two values of the relative signal strength $c = 0.34$ and $c = 0.98$. For all inspected parameter sets, the lower bound yields a very reasonable approximation to the frequency-

resolvable MIR (Fig. 7). In particular, we recover the low-pass information filtering found in the lower bound approximation. The RR estimate of the information rate, $i_{RR,k}$, shows a very similar behavior to the subthreshold case, being close to both MIR density and to the lower bound.

To summarize the results of this section, the performed simulations suggest that the linear encoding hypothesis and the lower bound approximation are reasonable for the LIF model driven by white Gaussian noise even in the case of a nonweak stimulus. In particular, the fraction of frequency-resolvable MIR is high and the model can be regarded as a low-pass filter of information. The RR estimate of the information rate is found to be close to the lower bound density and to the MIR. This suggests that a single LIF neuron encodes most of the information about the stimulus into its instantaneous firing rate.

Frequency-Resolved Information Transfer for the Synchronous Output of Two Neurons

Coincident, i.e., synchronous, spikes emitted by distinct neurons in a population are important in sensory systems detecting temporal information (Carr 1993) and seem to play an important role also in higher processing stages in the cortex (Koenig et al. 1996). Hence, in our context it is of interest how information with respect to frequency is encoded in the synchronous spikes of several neurons. An intriguing information-filtering effect has been observed by Middleton et al. (2009) for a population of neurons receiving a common stimulus: the coherence between the common stimulus and the spikes that neurons fire in synchrony is suppressed at low frequencies. The coherence and $i_{LB}(f)$ attain then a peak at a nonzero frequency f_{max}, and this means that, in the lower bound approximation, these synchronous spikes seem to encode preferentially information about higher frequency bands (a band-pass filter of information). We would like to explore if such a peak is also observed if the frequency-resolved MIR density is used as information measure. To this end, we study the simplest setup mimicking a postsynaptic cell acting as a coincidence detector
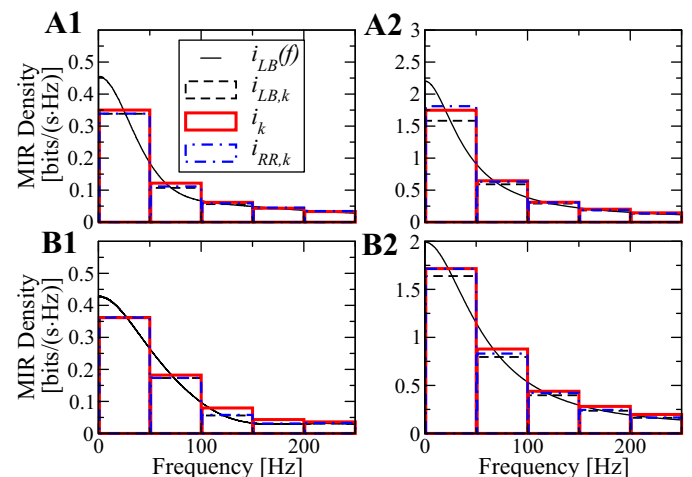

Fig. 7. Information filtering in the suprathreshold regime of a single LIF does not differ qualitatively from that in the subthreshold regime. *A1*: weak signal, lower baseline current. *A2*: strong signal, lower baseline current. *B1*: weak signal, higher baseline current. *B2*: strong signal, higher baseline current. Parameters: $\tau_{abs} = 2$ ms; *A1* and *A2*: $\mu = 1.3$; *B1* and *B2*: $\mu = 2.0$; *A1* and *B1*: $D = 6 \times 10^{-4}$ s; *A2* and *B2*: $D = 3.3 \times 10^{-3}$ s; *A1* and *B1*: $c = 0.34$; *A2* and *B2*: $c = 0.98$.
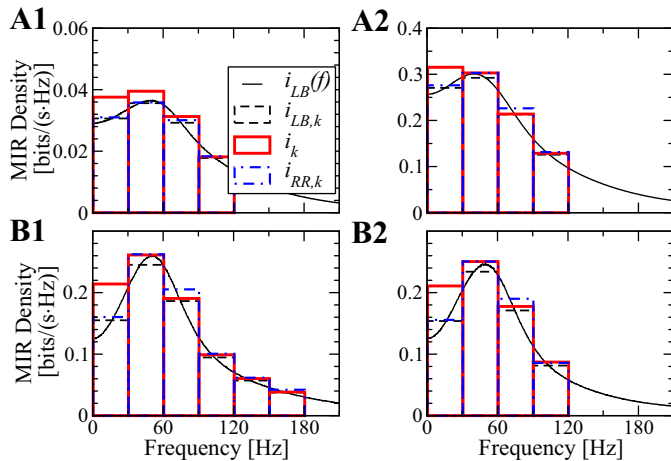
Fig. 8. For the synchronous spikes, the band-pass filtering is smoothed and is not present without an absolute refractory period. *A1*: for a weak signal and no refractory period, the band-pass filtering is weak. *A2*: for a stronger signal, the band-pass filtering disappears. *B1* and *B2*: with a refractory period the band-pass filtering is observed, although smoothed in the MIR compared with the lower bound. Parameters: $\mu = 1.3$; *A1*, *A2*, and *B1*: $\Delta t = 1$ ms; *B2*: $\Delta t = 2$ ms; *A1*: $D = 3.8 \times 10^{-4}$ s; *A2*, *B1*, and *B2*: $D = 6.6 \times 10^{-4}$; *A1* and *A2*: $\tau_{\mathrm{abs}} = 0$ ms; *B1* and *B2*: $\tau_{\mathrm{abs}} = 2$ ms; *A1*: $c = 0.09$; *A2*, *B1*, and *B2*: $c = 0.53$.

rather than an input integrator: the synchronous spikes of two uncoupled LIF units sharing the same input stimulus. A theory explaining the band-pass filtering observed in the spectral coherence of such a system has been put forward by Sharafi et al. (2013).

If no refractory period is included in the model, the optimal choice for μ and D is a moderately strong baseline current and a noise intensity that is smaller but still comparable to the value used in the simulation of the last section.

The bandwidth of the partial signals was set to the value of $\Delta f_{\mathrm{BW}} = 30$ Hz, because a larger bandwidth would not allow to resolve the peak. This bandwidth is the lowest value compatible with the technical restrictions (see *Numerical Methods*). In Fig. 8*A*, the results of the simulation for a weak (Fig. 8*A1*) and a moderate signal (Fig. 8*A2*) are shown. A much stronger signal was not considered, because it would cause the peak in $i_{\mathrm{LB}}(f)$ to disappear, as argued by Sharafi et al. (2013). In spite of the small signal variance, the MIR density is significantly above the lower bound in the frequency range 0–30 Hz for both signal intensities. In the other frequency bands, the i_k and $i_{\mathrm{LB},k}$ are close. Differently from what was observed in all simulations of a single LIF model, the profile of the MIR density in Fig. 8*A2* displays a qualitative difference from the lower bound, namely a low-pass information filtering (no peak). For the small intensity used for the data in Fig. 8*A1*, a weak maximum is still present but the difference between the first two bins is within the respective error bars. Interestingly, the MIR in the lower frequency bands exceeds $i_{\mathrm{RR},k}$. This offers an example of how this spectral measure is not a general upper bound on the information transfer.

Next, we explore the effect of adding an absolute refractory period to the dynamics of the two LIF neurons. An absolute refractory period also automatically prevents the rare, but possible, event of more than one spike falling into the same time bin and allows us to change the time resolution Δt. In Fig. 8, *B1* and *B2*, the results of the first simulation including a refractory period are shown. In Fig. 8, *B1* and *B2*, a time

resolution of 1 or 2 ms is used, respectively. The results look different from those of Fig. 8, *A1* and *A2*, because here a definite peak is observed in the MIR density for a moderate signal intensity, apparently due to the nonvanishing absolute refractory period. Most information is encoded in the band of 30–60 Hz compared with any other single band. The information-filtering effect, which was essentially absent for $\tau_{\mathrm{abs}} = 0$ in Fig. 8, *A1* and *A2*, is now apparent but it is less sharp than in the lower bound approximation. The similarity with the previous plots is that the difference between i_k and $i_{\mathrm{LB},k}$ is largest for the low-frequency bands. In Fig. 8*A2*, this difference was large enough to cause the peak to disappear, while here it is just smoothed. In the lowest frequency band, the actual information transfer is greater than what is predicted by the RR estimate, which follows the lower bound rather closely in this case. Doubling the time step has a small effect on the overall information transmission, as already seen for the single LIF model. Comparing Fig. 8, *B1* and *B2*, shows that both i_k and $i_{\mathrm{LB},k}$ undergo a small global decrease but the relative difference between the two quantities is almost unchanged.

Simulations of the system reveal that for $\tau_{\mathrm{abs}} > 0$, a more pronounced peak in $i_{\mathrm{LB}}(f)$ is observed for a higher baseline current μ. However, the results in Fig. 9, *A1* and *A2*, reveal that this is not the case for the MIR density i_k but quite to the contrary there is no peak in i_k at all. In Fig. 9, *A1* and *A2*, we display two simulation results with $\mu = 2$, which differ only in the time resolution Δt. In both cases the frequency-resolved MIR density decreases monotonically. Hence, the band-pass filtering effect is lost in this regime, and this finding does not change by varying the temporal resolution. The different time resolution causes a global decrease of the information transmission. This decrease is larger than for a single LIF neuron (cf. the red histogram in the first frequency bins of Fig. 5, *B1* and *B2*), suggesting that the time precision is more important in this encoding scheme. The results of these simulations demonstrate that the presence of an absolute refractory period enhances the suppression of the coherence at low frequencies but is not always sufficient to cause a peak in the MIR density
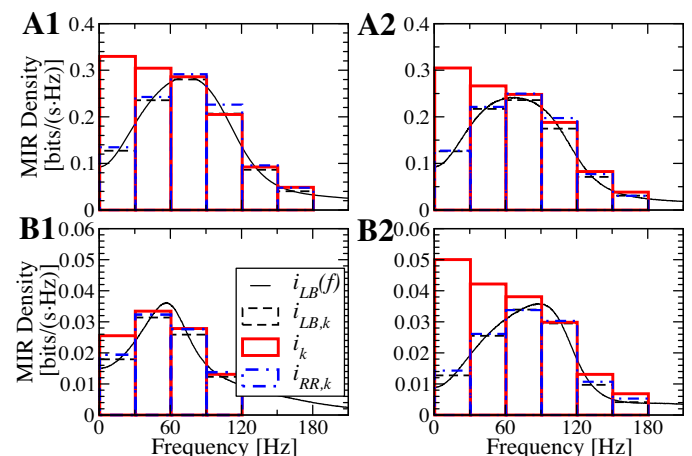

Fig. 9. For strong baseline current (high firing rate), band-pass filtering disappears for both fine (*A1*) and coarse (*A2*) time resolution. The difference between low and high baseline current remains also for a weak signal [low (high) current in *B1* (*B2*)]. Parameters: $\tau_{\mathrm{abs}} = 2$ ms; *A1*, *B1*, and *B2*: $\Delta t = 1$ ms; *A2*: $\Delta t = 2$ ms; *A1*, *A2*, and *B2*: $\mu = 2.0$; *B1*: $\mu = 1.3$; *A1* and *A2*: $D = 6.6 \times 10^{-4}$ s; *B1* and *B2*: $D = 3.8 \times 10^{-4}$ s; *A1* and *A2*: $c = 0.53$; *B1* and *B2*: $c = 0.09$.
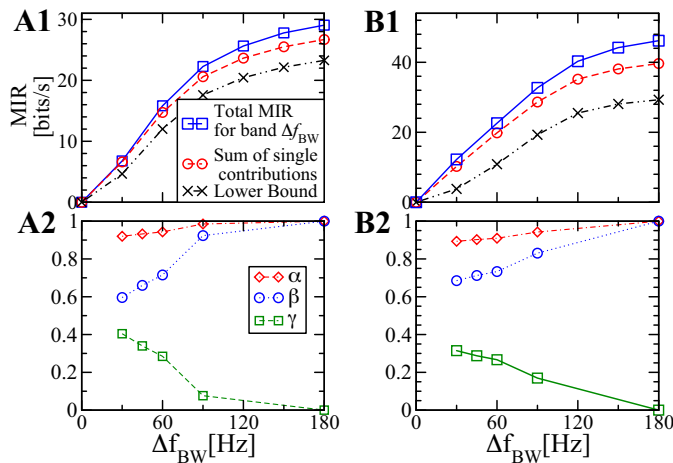
Fig. 10. Synergetic effects for the synchronous spikes. *A1* and *B1*: MIR for a total, containing all partial signals with lower frequency, signal (blue squares) and cumulative sum of partial signals (red circles). *A2* and *B2*: fraction of frequency-resolvable information α, fraction of the difference between full MIR and lower bound $I' - I'_{LB}$, which is explained by the intraband difference β and the synergy between bands γ. Parameters: $\Delta t = 1$ ms, $D = 3.8 \times 10^{-4}$ s, $\tau_{abs} = 2$ ms, $c = 0.53$; *A1* and *A2*: $\mu = 1.3$; *B1* and *B2*: $\mu = 2.0$.

i_k. The intensity of the baseline current also plays a role but in a way that is difficult to interpret. The RR estimate looks similar to the lower bound density and is strongly exceeded by the MIR density in the lower frequency range. To find out if the MIR density converges to the lower bound for decreasing signal strength, simulations for both cases of $\mu = 1.3$ and $\mu = 2.0$ were carried out with a weaker stimulus and the results are shown in Fig. 9, *B1* and *B2*. Figure 9*B1* displays the results for the lower suprathreshold baseline current ($\mu = 1.3$). While the qualitative picture is unchanged, the relative difference between i_k and $i_{LB,k}$ in the lowest frequency band decreases to slightly less than one-half of what is observed in Fig. 8*B1*. A further study of this slow convergence would require a further decrease in the signal strength, which would bring information rates to values too small to be reliably measured in this setup and the question of the convergence to the lower bound remains open. Figure 9*B2* shows the results for the case of a higher baseline current ($\mu = 2.0$ as in Fig. 9*A1*) for the very weak stimulus. The low-pass behavior of the MIR density is still observed, and the relative difference between i_k and $i_{LB,k}$ in the lowest frequency band is even larger than for the stronger signal. This suggests that the MIR density i_k might not converge to the lower bound $i_{LB,k}$ even for smaller and smaller signal intensity. In both cases, $i_{RR,k}$ is close to the lower bound density.

The plots in Fig. 10 summarize the synergetic effects in the two parameters choices of Figs. 8*B1* and 9*A1*. In both cases, synergy between bands causes the MIR for a partial signal and the cumulative sum of the contributions of single bands of bandwidth 30 Hz to fork (Fig. 10*A1* for $\mu = 1.3$ and Fig. 10*B1* for $\mu = 2.0$). The lower bound, plotted with black crosses, performs better for the case of the smaller baseline current (Fig. 10*A1*). The ratio between MIR and the lower bound for the total signal is 1.2 (1.6) in the case of the smaller (higher) baseline current (cf. Fig. 10, *A1* and *B1*). Figure 10, *A2* and *B2*, shows that the fraction of frequency-resolvable MIR is large in both cases (>90% even for the finest frequency resolution) and that the synergy between partial signals is lower than in the

case of a single LIF model for both parameter choices. The plots also indicate that in this system β might not vanish, not even in the limit of very small bandwidth, especially in the case of the high baseline current (Fig. 10*B2*), i.e., that the lower bound approximation would not be exact even in the limit of weak signals.

The results of this subsection for the encoding by synchronous spikes are not easy to decipher, because the information-filtering properties of the system depend on the details. For high values of the baseline current μ, a peak in the MIR density at nonvanishing frequency did not occur, while for a smaller value of μ, the peaks can be observed. On the one hand, it is not completely unexpected that significant deviations from the lower bound are observed, because the operation of sorting out spikes from one spike train by looking at another one is strongly nonlinear. On the other hand, this result is puzzling, because it is not clear what is the difference between two different values of a suprathreshold baseline current. To understand this behavior, further investigations are needed, but extensive parameter scans are not a practicable option because of the computational cost of every single simulation. It is also difficult to cover all parameter combinations because of the practical restrictions imposed by the correlation time of the system: high firing rates and small noise intensities cause the correlation time to grow beyond the values that are tractable in our algorithm.

Frequency-Resolved Information Transfer for the EIF Model

To check if our findings hinge on the particular choice of the LIF model, we performed extensive simulations of the EIF model. In Fig. 11, *A1* and *A2*, we show the results for a single EIF neuron in the fluctuation-driven regime, that is for sub-threshold values of the baseline current. In Fig. 11*A1* (Fig. 11*A2*), the input signal has moderate (strong) intensity. As in the case of the single LIF model, the lower bound approximation is very good and the frequency-resolved MIR exceeds the lower bound approximation at lower frequencies in the case of a strong signal. Results in the tonic firing regime (with baseline
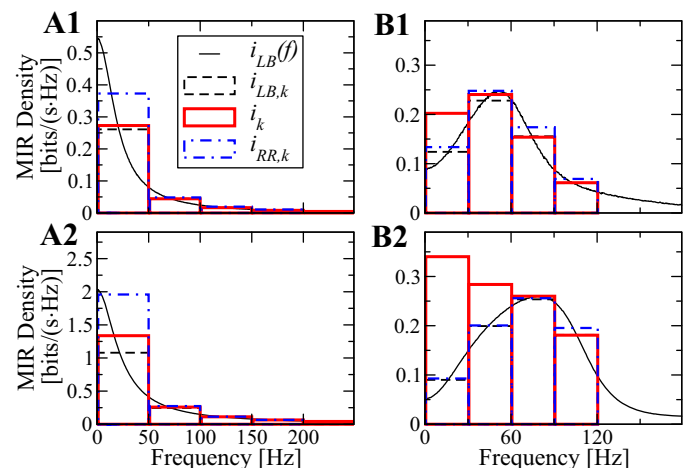

Fig. 11. Simulation results for the exponential integrate-and-fire model (EIF) model are qualitatively very similar to those for the LIF model for both single spike trains (*A1* and *A2*) and synchronous (*B1* and *B2*) spikes. Parameters: $\Delta t = 1$ ms; *A1* and *A2*: $\mu = 0.7$; *B1*: $\mu = 1.6$; *B2*: $\mu = 2.5$; *A1*: $D = 7.6 \times 10^{-4}$ s; *A2*: $D = 3.3 \times 10^{-3}$ s; *B1* and *B2*: $D = 9.5 \times 10^{-4}$ s; *A1* and *A2*: $\tau_{abs} = 0$ ms; *B1* and *B2*: $\tau_{abs} = 2$ ms; *A1*, *B1*, and *B2*: $c = 0.53$; *A2*: $c = 0.98$.

current larger than the spike initiation threshold) are similar to the corresponding results for the LIF model: the information filtering is low-pass and the lower bound is very close to the frequency-resolved MIR (not shown).

In Fig, 11, *B1* and *B2*, we plot the results for the synchronous spikes of two EIF models driven by a common signal. The baseline currents were adjusted to obtain values of the synchronous firing rate ($\approx$27 Hz in Fig. 11*B1*; $\approx$64 Hz in Fig. 11*B2*) and coefficient of variation (CV; $\approx$0.9) similar to the values of Fig. 8*B1* (rate $\approx$28 Hz, CV $\approx$0.8) and Fig. 9*A1* ($\approx$28 Hz, CV $\approx$0.8), respectively. As for the LIF model, the lower frequencies are encoded nonlinearly. In the case of higher firing rate, the excess of information is large enough to make the frequency-resolved MIR look qualitatively different from the lower bound approximation (low pass instead of band pass).

DISCUSSION

In this study we have developed a method to resolve the information transmission of time-dependent Gaussian signals through neural systems with respect to their frequency. Gaussian noise stimuli have statistically independent Fourier components and can thus be regarded as a superposition of independent information sources. More specifically, for a signal with cutoff frequency f_c and a desired frequency resolution Δf_{BW}, we have $f_c/\Delta f_{BW}$ such sources. For each of these sources, we can separately measure the mutual information with the output of the neural system (e.g., a spike train) and ascribe a certain fraction of the full MIR to the specific frequency band. We compared this frequency-resolvable part of the information rate to the predictions of the lower bound, which is based on the spectral coherence function. In particular, we tested whether the information transmission is still maximal where the coherence is maximal, i.e., whether the lower bound measure at least qualitatively captures the information-filtering character of the respective neuron or neural system. Further, because we deal in general with nonlinear systems, only a certain fraction of the total MIR can be actually frequency resolved. We quantified the remaining synergetic part of the total MIR that accounts for the difference between the full MIR and the sum of MIR contributions from single frequency bands. We also compared our results to the predictions based on the square root of the RR coherence, which yields an upper bound to the MIR under restrictive assumptions (Borst and Theunissen 1999). Some of these assumptions are technical (description of the effective noise in the system as Gaussian, additive, and independent of the stimulus) and can be in principle verified for the particular data set under investigation. The most restrictive assumption, however, is the encoding hypothesis, i.e., the premise that the information about the input stimulus is exclusively encoded into the time-dependent firing rate. It is not clear how this can be justified unless the way the system encodes information is known a priori.

Let us first of all emphasize that despite the synergetic (not frequency-resolvable) part of the information transfer, in all cases considered here, the notion of frequency resolution still represents a useful concept. The fraction of frequency-resolvable MIR for weak stimuli does not fall below 90% (see Figs. 6*A2* and 10, *A2* and *B2*) and even in the case of strong stimulation it is rarely below 80% (see Fig. 6*B2*) of the total

MIR. In addition, this fraction also depends on Δf_{BW}, the frequency resolution of the total MIR (the width of the resulting frequency bands). The question arises what are meaningful values for this width of the frequency band for specific sensory systems. As a proxy for a reasonable frequency resolution, we can take the scale over which the coherence displays significant changes. In the auditory systems of the bullfrog (Rieke et al. 1995) and of the cricket (Marsat and Pollack 2004) and in the visual system of the cat (Passaglia and Troy 2004), for instance, it has been found that the coherence varies over a few tens of hertz, suggesting a frequency resolution similar to the one we used in this article. However, there are also systems in which the lower bound estimate varies over only a few hertz (Massot et al. 2011).

The choice of the frequency resolution gains additional importance in the special situation of different frequency bands being associated with distinct sensory stimuli. A popular biological model system in this respect is the electrosensory system of weakly electric fish (Kreiman et al. 2000; Krahe et al. 2008). Behaviorally relevant prey signals have most power below 20 Hz (Fotowat et al. 2013) while the likewise important communication stimuli by other fish are distributed over a frequency range of 20–120 Hz (Stamper et al. 2010). To estimate whether a certain upstream neuron in the electrosensory system encodes preferentially information of prey or communication signals, it would be thus sufficient to look at the MIR with a frequency resolution of 20 Hz.

The Poisson neuron was considered in the first place, because it is the only point process for which limited analytical results are available. The numerical results confirmed both the validity of the procedure and the practical limitations to its applicability (the convergence of the biased estimators is slow and the time scales of the system autocorrelations have to be short-ranged). Although the Poisson neuron can encode signals only in the time-dependent firing rate by definition, it showed a nontrivial interaction between frequencies: despite the fact that the information of the source (the entropy of the Gaussian ensemble) is additive, the information transfer of the total signal is strictly larger than the sum of its components.

Our results for the single white-noise driven LIF neuron have shown for the first time that in the case of a weak signal the lower bound on the MIR is a tight bound. This implies that the coherence function of this popular neuron model gives a fairly complete picture of the information transfer for small stimulus amplitudes. Our results confirm, in particular, that the LIF neuron is a low-pass filter of information, i.e., that most information is encoded about low-frequency components of the stimulus. Stimuli beyond the weak-stimulus paradigm seem to add preferentially information at low frequencies for a single LIF neuron and thus to strengthen the low-pass filtering of information. How to calculate this nonlinear part of the information rate analytically is still an open problem.

Although our results for the single LIF model have confirmed the kind of information filter that was predicted by the lower bound estimate, we have also found counter examples. For the simplest case of a synchrony code, a maximum of the coherence at a nonvanishing frequency f_{max} does not necessarily imply but also does not exclude that in a frequency band around f_{max}, the frequency-resolved MIR is maximal. Varying the baseline current makes the system switch from a frequency-selective encoding to a broader low-pass information filtering.

Because measuring the frequency-resolved MIR is computationally very expensive, extensive parameter scans could not be easily performed and we could neither investigate the exact point at which this transition occurs nor how it varies with other system parameters. However, our results on the information filtering by synchrony suggest that caution must be used when interpreting results that are solely based on the coherence function and the lower bound estimate. One should also be careful in using the square root of the RR coherence as an information transmission measure. While the risk of using this measure as a general upper bound has been already pointed out by means of a very simple model (Rozell and Johnson 2004), our results give an explicit example of how the RR coherence can fail as an upper bound in a more realistic setup. All our findings are robust with respect to the substitution of the LIF model with the EIF model both in the case of the single cell and the spikes fired in synchrony.

Our result on the information filtering by synchrony has only a preliminary character in other respects. First of all, the main application of such synchrony code is certainly the firing of a postsynaptic cell that acts as a coincidence detector, i.e., that fires only if stimulated almost simultaneously by n presynaptic cells. Here we have studied only the simplest yet not most realistic case of $n = 2$. Although we found band-pass information filtering in the lower bound estimate and (for some parameter sets) also in the frequency resolved MIR, the frequency-dependent differences were only modest. In other words, the part of the total information rate that makes up the difference between low-pass filter and high-pass filter is in the considered case rather small. However, this does not neutralize the fundamental importance of verifying whether a maximum in the lower bound MIR density at f_{max} translates into a maximum of the true MIR density. Typically, the number of presynaptic neurons that have to fire in synchrony to make a postsynaptic cell fire is higher than two. For $n > 2$ the question of information filtering and whether it is correctly predicted by the lower bound estimate may find a more unambiguous answer than provided in our article for $n = 2$.

Regarding the differences between lower bound MIR and true MIR density, we can only speculate. Certainly, the procedure of sorting out synchronous spikes is a highly nonlinear operation. The lower bound is essentially based on the linear correlations between input and output (it uses the coherence function, which is essentially a correlation coefficient in the Fourier domain). The coherence measure is blind to nonlinear correlations or higher order correlations. Hence, we could expect that a part of the information that according to the lower bound estimate is suppressed at low frequencies may be encoded in higher moments or higher order correlation functions between input and output. More analytical efforts and in particular tractable examples are needed to gain some intuition in this matter and to clarify how much information can be encoded beyond the time-dependent firing rate.

High-pass filtering of information has also been observed in the vestibular system of primates, both in afferent neurons (Sadeghi et al. 2007) and in higher order vestibular nuclei (Massot et al. 2011). These properties are also reflected by a model (Sadeghi et al. 2007) that comprises an integrate-and-fire neuron with dynamical threshold mimicking an adaptation mechanism and includes a prefilter for sensory stimuli (head velocity). An application of our framework to this adaptive

neuron model could confirm or disprove this form of information filtering.

Turning back to our method, we would like to remark that its application to data from real neurons might look difficult, given the huge amount of data that were generated to eliminate the undersampling bias. However, the approach pursued here was only guided by computational efficiency: the run time of our algorithm grows linearly with the ensemble size, while the computational needs of other more refined bias-reduction techniques grow more rapidly. Therefore, the fastest way to reduce the bias was to generate more data. In experiments the amount of data is limited, but some bias reduction techniques seem to give good results also if the ratio between histogram bins and samples is of order one (Paninski 2003; Nemenman et al. 2004; Panzeri et al. 2007), and this suggests that an application of the method to experimental data may be possible. A first indication about the feasibility of this application could be to investigate whether the same results presented in this work can be reproduced on the basis of a more realistically sized amount of data by using one of the mentioned entropy estimators with better bias reduction performance. Another extension of our method can be made by using alternative approaches for the estimation of entropy rates (Shlens et al. 2007).

Our approach is not limited to a single time-dependent output. Although in this study the output states were chosen to be single spike trains, calculating entropy rates is possible with any choice of output states. This permits, in principle, a straightforward extension to the analysis of network population data. In practice, neural population activity is represented by a vector of spike trains, which makes the dimensional explosion of the output space much more severe than for the case of a single spike train. Nevertheless, a reduced description of the output states may permit the investigation of information filtering by neural populations.

Numerical estimation of entropy rates from neural data is a field in constant development. We believe that our results present a new and theoretically sound framework to use these tools and that it will encourage other researchers to further investigate information filtering in neural systems.

APPENDIX: NUMERICAL CALCULATION OF ENTROPY RATES

In this APPENDIX we report the details of our implementation of the direct method (Strong et al. 1998) and on the procedures we used to calculate all entropy rates.

Implementation of the Direct Method

In the following, we consider the output x to be a binary sequence where, say, a 1 or a 0 stands for the presence or absence of a spike in the corresponding time bin, respectively. Therefore, the time discretization step Δt has to be small enough that at most one spike falls into a single time bin.

Many stimuli are presented to the studied spike generator and spike trains are recorded. A time resolution Δt is set and spike trains are converted to binary sequences whose length depends on the time step: if a time segment of length T is considered, a bit sequence (a "word") of $T/\Delta t$ symbols identifies the system response, and there are $2^{T/\Delta t}$ of such possible words. The probability of each word can be estimated by counting its occurrences, although the exponential growth of the set of possible words requires huge data sets if longer words are considered (i.e., longer time windows T or a finer temporal resolution Δt). Computing $H(x,T,\Delta t)$, the entropy of the words of length T, from
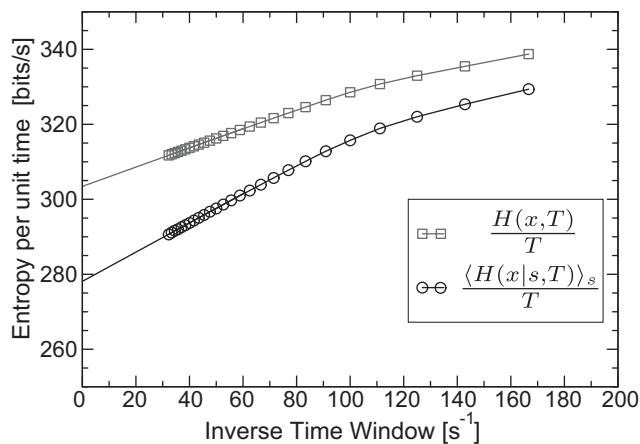
Fig. A1. Example of extrapolation to infinite time window. The data result from the synchronous output of 2 LIF neurons driven by a Gaussian stimulus. The intercept of the 2 regression lines gives the entropy rates. The noise entropy is calculated for a partial signal corresponding to the frequency band 0–75 Hz. Parameters: $\Delta t = 1$ ms, $\mu = 2.0$, $D = 3.8 \times 10^{-4}$ s, $\tau_{abs} = 2$ ms, $c = 0.34$.

the estimated probability distribution is a tricky problem, because plugging the estimated probabilities into the definition of entropy always gives a negatively biased result (Paninski 2003). If the probability distribution is not well sampled, the bias can be very large and the entropy estimate can be completely meaningless. Assume now that a reliable estimate of $H(x,T,\Delta t)$ can be obtained from the estimated probabilities for the words. To compute the entropy rate, the limit of $H(x,T)/T$ has to be computed (because the time resolution is fixed, the dependence on Δt will be dropped in the following). Under the realistic assumption that the range of correlations in the response is finite and extends up to a time T_c, for times larger than T_c the entropy grows linearly:

$$H(x, T) \approx H(x, T_c) + (T - T_c)H'(x)$$
$$= C + H'(x) \cdot T, \qquad (30)$$

where $H(T_c)$ and C are in general unknown constants. Dividing by T yields:

$$\frac{H(x,T)}{T} = H'(x) + \frac{C}{T}. \qquad (31)$$

Because the correlation time is usually unknown, the ratio $H(x,T)/T$ is plotted as a function of $1/T$ and the linear part of the curve, which is correctly described by *Eq. 31*, has to be determined a posteriori from the graph itself (as in the example of Fig. A1, red squares).

Deviations from linear behavior occur at small T because of correlations and may be observed at large T, where a negative bias due to undersampling sets in. The intercept of the linear regression is the full entropy rate $H'(x)$.

To compute the noise entropy, the same stimulus s has to be presented many times. A particular spot t on the stimulus timeline is marked, and only time windows of length T beginning at time t from independent trials contribute to the calculation of $H(x|s,T,t)$. Averaging over t yields the noise entropy of words of length T:

$$H(x|s, T) = \langle H(x|s, T, t) \rangle_t. \qquad (32)$$

The last expression was computed for other fixed stimuli and finally averaged over the stimulus ensemble. Given the ergodic property of Gaussian random functions, this last step would not be necessary if the stimulus were long enough. However, it turned out to be numerically more convenient to generate ~100 signals of 2–3 s than a single long signal. By means of the same extrapolation procedure as for the full entropy, the noise entropy rate can then be extracted (example in Fig.

A1, black circles). Subtracting the noise entropy from the full entropy rate gives the MIR

$$I'(\Delta t) = H'(x, \Delta t) - \langle H'(x|s, \Delta t) \rangle_s \qquad (33)$$

where the dependence on the time resolution has been explicitly indicated again.

An Efficient Scheme to Manage the System Output: Words and Dictionaries

For the extrapolation to infinite time window to be accurate, the system autocorrelation time has to be covered by the binary words, and because the standard time resolution used in the simulations was set to one millisecond (the minimal width of an action potential), words of tens of symbols had to be considered. In the case of 30 bins the number of possible words is $2^{30} \approx 10^9$. The extrapolation procedure also requires to determine this high dimensional probability distribution for different word lengths. The calculation of the noise entropy additionally requires to have one of such sets of probability distributions for each time segment the output is divided in. If the spike train lasts a few seconds, a few hundreds of such sets of probability distributions need to be stored. To effectively manage and quickly fill up such a large number of probability arrays, an efficient system to label word distributions is needed. A convenient way to label words is to consider a binary sequence as a binary number read from left to right. It is easy to convert each binary string into the corresponding integer number in base ten. This can be used as the index for the probability array. For example, in an array storing the words of length five, the word 00000 is assigned the position 0, the word 10000 is assigned the position 1, the word 00101 the position 20, and so on. This is an efficient way of labeling words, and it is very fast because it requires only few very fast operations (addition and multiplication). However, the problem of the exponential explosion of the number of possible words (and of the memory requirements) remains. On the other hand, if the time resolution is on the millisecond scale, spikes are sparse. This means that most of the $2^{T/\Delta t}$ possible words have actually negligible probability and the binary number assignment system has the downside that huge memory resources are allocated just to store an enormous number of zeros and a few interesting numbers. A simple observation is that words containing much more spikes than the word length times the mean rate practically never occur. This suggests a way to find a better assignment system, in which first words with no spikes are considered, then words with

Table A1. *Word indexing*

Binary Words	Key String	Value
0 0 0 0 0	$\rightarrow (5,0)$	$\rightarrow 0$
1 0 0 0 0	$\rightarrow (5,1)$	$\rightarrow 1$
0 1 0 0 0	$\rightarrow (5,2)$	$\rightarrow 2$
0 0 1 0 0	$\rightarrow (5,4)$	$\rightarrow 3$
0 0 0 1 0	$\rightarrow (5,8)$	$\rightarrow 4$
0 0 0 0 1	$\rightarrow (5,16)$	$\rightarrow 5$
1 1 0 0 0	$\rightarrow (5,3)$	$\rightarrow 6$
1 0 1 0 0	$\rightarrow (5,5)$	$\rightarrow 7$
1 0 0 1 0	$\rightarrow (5,9)$	$\rightarrow 8$
1 0 0 0 1	$\rightarrow (5,17)$	$\rightarrow 9$
0 1 1 0 0	$\rightarrow (5,6)$	$\rightarrow 10$
0 1 0 1 0	$\rightarrow (5,10)$	$\rightarrow 11$
0 1 0 0 1	$\rightarrow (5,18)$	$\rightarrow 12$
0 0 1 1 0	$\rightarrow (5,12)$	$\rightarrow 13$
0 0 1 0 1	$\rightarrow (5,20)$	$\rightarrow 14$
0 0 0 1 1	$\rightarrow (5,24)$	$\rightarrow 15$

Left column: all the binary words up to the maximum expected spike count $m(L)$. In this example $m(5) = 2$. *Middle column*: list of the key string for the map. *Right column*: value mapped by the dictionary. Number of entries = 16. Number of possible binary words = 32. Compression ratio = 0.5.

only one spike, then words with two spikes, and so on. An algorithm goes through all words up to $m(L)$ spikes per word of length L and transforms the binary string into a decimal number, which is then assigned a position in the compact indexing (Table A1). This correspondence from the binary to the compact indexing is stored in an associative array (a map). This map forms a "dictionary" that translates the fast but sparse binary indexing into the compact indexing, computed only once for each word at the beginning of the simulation. In the pedagogical example of Table A1, the compression ratio is exactly one-half, but in a realistic scenario for the LIF model ($L \approx 25$ and $m \approx 5$) the compression ratio ranges from about 10^{-3} to 10^{-4}. If a sufficiently large refractory period ($\tau_{abs} \geq 2 \cdot \Delta t$) is included in the model, the words containing spikes in neighboring bins can be skipped. In this case, the compression ratio can be further reduced, typically by an additional factor 1/2.

A nontrivial point is that $m(L)$, the maximum number of spikes per word of length L, which are included in the dictionary, has to be fixed before the simulation starts. If $m(L)$ is too small, the results of the entropy calculation can be affected while large $m(L)$ leads to a great waste of computational resources, because the dictionary would be unnecessarily large. A preliminary simulation determines the count statistics for segments of length L. The smallest value with zero counts should be $m(L)$. However, it is still possible that few words of length L with spike count $k > m(L)$ occur and it is actually desirable to be able to handle these words without having to increase $m(L)$. These "oversize" words are stored all together in the last positions in the dictionary and are distinguished only by the spike count. Because of the translational invariance, a roughly uniform distribution can be assumed for words with a fixed spike count (this assumption is also made by Strong et al. 1998 to justify the use of the Ma-entropy lower bound). As there are $\binom{L}{k}$ possibilities for such words, the probability of a coincidence is $\sim 1/\binom{L}{k}$, and it is very likely that "oversize" words are all different from each other, and so they are treated in the entropy calculation. If a coincidence does happen, the entropy is overestimated by $2/N$, where N is the number of samples (in our simulations $N \geq 10^6$ for the noise entropy, and $N \geq 10^8$ for the full entropy). The expectation value of the overestimation is about $2/[N \cdot \binom{L}{k}]$. If the number of oversize words makes the relative error on $H(T)/T$ larger than 10^{-6}, the corresponding point (typically the largest time window) is not included in the extrapolation.

Entropy Estimators

If the estimated probabilities of a distribution are simply plugged into the definition of entropy [the maximum likelihood estimator (MLE), also called "naive" by Strong et al. 1998], it is a well-known fact that an always negatively biased estimate of the entropy is obtained (see for example Paninski 2003). Let q indicate the number of bins of the histogram of the words and N the number of samples. In the original direct method procedure (Strong et al. 1998) a quadratic fit of the MLE calculated for data subsets of size $N/2$, $N/3$, and $N/4$ is performed and the intercept of the quadratic fit is taken as the value of the true entropy. In our simulations, this extrapolation tended to be unstable (because $N \gg q$, the variance of the points is of the same order as the bias, see Paninski 2003) and this method was not used. Several other methods have been developed to correct for this bias (Paninski 2003; Panzeri et al. 2007). Simpler methods like the Miller-Madow correction (Miller 1955) or the jackknifed (Quenouille 1956) version of the MLE perform well if many samples are available, that is if $N \gg q$. Other methods were recently developed to reduce bias even in the undersampled regime $N \sim q$, like the BUB estimator (Paninski 2003) or the coincidence-based NSB method (Nemenman et al. 2004), which seems to have the best bias reduction performance (Panzeri et al. 2007). These methods are extremely useful if data are limited, as in real experiments. In the asymptotical regime, although, their performance is comparable to the simpler methods described above at a higher computational cost (Paninski 2003; Panzeri et al. 2007). Because in computer simulations a very large amount of data can be generated, we used exclusively the Miller-Madow correction and the jackknifed MLE.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

Author contributions: D.B. and B.L. conception and design of research; D.B. performed experiments; D.B. analyzed data; D.B. and B.L. interpreted results of experiments; D.B. prepared figures; D.B. and B.L. drafted manuscript; D.B. and B.L. edited and revised manuscript; D.B. and B.L. approved final version of manuscript.

REFERENCES

Aldworth AN, Dimitrov AG, Cummins GI, Gedeon T, Miller JP. Temporal encoding in a nervous system. *PLoS Comput Biol* 7: e1002041, 2011.

Badel L, Lefort S, Brette R, Petersen CC, Gerstner W, Richardson MJ. Dynamic I-V curves are reliable predictors of naturalistic pyramidal-neuron voltage traces. *J Neurophysiol* 99: 656–666, 2008.

Bialek W, Deweese M, Rieke F, Warland D. Bits and brains–information flow in the nervous system. *Physica A* 200: 581–593, 1993.

Bialek W, Zee A. Coding and computation with neural spike trains. *J Stat Phys* 59: 103–115, 1990.

Borst A, Theunissen F. Information theory and neural coding. *Nat Neurosci* 2: 947–957, 1999.

Burkitt AN. A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input. *Biol Cybern* 95: 97–112, 2006.

Carr CE. Processing of temporal information in the brain. *Annu Rev Neurosci* 16: 223–243, 1993.

Chacron MJ, Doiron B, Maler L, Longtin A, Bastian J. Non-classical receptive field mediates switch in a sensory neuron's frequency tuning. *Nature* 423: 77–81, 2003.

Chacron MJ, Lindner B, Longtin A. Threshold fatigue and information transfer. *J Comput Neurosci* 23: 301–311, 2007.

Deemyad T, Kroeger J, Chacron MJ. Sub- and suprathreshold adaptation currents have opposite effects on frequency tuning. *J Physiol* 590.19: 4839–4858, 2012.

Droste F, Schwalger T, Lindner B. Interplay of two signals in a neuron with heterogeneous short-term synaptic plasticity. *Front Comp Neurosci* 7: 86, 2013.

Fotowat H, Harrison RR, Krahe R. Statistics of the electrosensory input in the freely swimming weakly electric fish apteronotus leptorhynchus. *J Neurosci* 33: 13758–13772, 2013.

Fourcaud N, Brunel N. Dynamics of the firing probability of noisy integrate-and-fire neurons. *Neural Comput* 14: 2057–2110, 2002.

Fourcaud-Trocmé N, Hansel D, van Vreeswijk C, Brunel N. How spike generation mechanisms determine the neuronal response to fluctuating inputs. *J Neurosci* 23: 11628–11640, 2003.

Gabbiani F. Coding of time-varying signals in spike trains of linear and half-wave rectifying neurons. *Network Comp Neural* 7: 61–85, 1996.

Gerstner W, Kistler WM. *Spiking Neuron Models.* Cambridge, UK: Cambridge Univ. Press, 2002.

Koch C. *Biophysics of Computation–Information Processing in Single Neurons.* Oxford, UK: Oxford Univ. Press, 1999.

König P, Engel AK, Singer W. Integrator or coincidence detector? The role of the cortical neuron revisited. *Trends Neurosci* 19: 130–137, 1996.

Krahe R, Bastian J, Chacron MJ. Temporal processing across multiple topographic maps in the electrosensory system. *J Neurophysiol* 100: 852–867, 2008.

Kreiman G, Krahe R, Metzner W, Koch C, Gabbiani F. Robustness and variability of neuronal coding by amplitude-sensitive afferents in the weakly electric fish eigenmannia. *J Neurophysiol* 84: 189–204, 2000.

Lindner B, Gangloff D, Longtin A, Lewis JE. Broadband coding with dynamic synapses. *J Neurosci* 29: 2076–2088, 2009.

Machens CK, Stemmler MB, Prinz P, Krahe R, Ronacher B, Herz AV. Representation of acoustic communication signals by insect auditory receptor neurons. *J Neurosci* 21: 3215–3227, 2001.

Marsat G, Pollack GS. Differential temporal coding of rhythmically diverse acoustic signals by a single interneuron. *J Neurophysiol* 92: 939–948, 2004.

Massot C, Chacron M, Cullen K. Information transmission and detection thresholds in the vestibular nuclei: single neurons vs. population encoding. *J Neurophysiol* 105: 1798–1814, 2011.

McGillivray P, Vonderschen K, Fortune ES, Chacron MJ. Parallel coding of first- and second-order stimulus attributes by midbrain electrosensory neurons. *J Neurosci* 32: 5510–5524, 2012.

Middleton JW, Longtin A, Benda J, Maler L. The cellular basis for parallel neural transmission of a high-frequency stimulus and its low-frequency envelope. *Proc Natl Acad Sci USA* 103: 14596–14601, 2006.

Middleton JW, Longtin A, Benda J, Maler L. Postsynaptic receptive field size and spike threshold determine encoding of high-frequency information via sensitivity to synchronous presynaptic activity. *J Neurophysiol* 101: 1160–1170, 2009.

Middleton JW, Yu N, Longtin A, Maler L. Routing the flow of sensory signals using plastic responses to bursts and isolated spikes: experiment and theory. *J Neurosci* 31: 2461–2473, 2011.

Miller G. Note on the bias of information estimates. *Information Theory in Psychology II-B.* Glencoe, IL: Free Press, 1995, p. 95–100.

Neiman AB, Russell DF. Sensory coding in oscillatory electroreceptors of paddlefish. *Chaos* 21: 047505, 2011.

Nemenman I, Bialek W, van Steveninck RD. Entropy and information in neural spike trains: progress on the sampling problem. *Phys Rev E Stat Nonlin Soft Matter Phys* 69: 056111, 2004.

Oswald AM, Chacron MJ, Doiron B, Bastian J, Maler L. Parallel processing of sensory input by bursts and isolated spikes. *J Neurosci* 24: 4351–4362, 2004.

Paninski L. Estimation of entropy and mutual information. *Neural Comput* 15: 1191–1253, 2003.

Panzeri S, Senatore R, Montemurro MA, Petersen RS. Correcting for the sampling bias problem in spike train information measures. *J Neurophysiol* 98: 1064–1072, 2007.

Papoulis A, Pillai SU. *Probability, Random Variables and Stochastic Processes.* New York: McGraw-Hill Higher Education, 2002.

Passaglia CL, Troy JB. Information transmission rates of cat retinal ganglion cells. *J Neurophysiol* 91: 1217–1229, 2004.

Press HP, Teukolsky S, Vetterling WT, Flannery BP. *Numerical Recipes (3rd ed.).* Cambridge, UK: Cambridge Univ. Press, 2007.

Quenouille MH. Notes on bias in estimation. *Biometrika* 43: 353–360, 1956.

Reinagel P, Godwin D, Sherman SM, Koch C. Encoding of visual information by LGN bursts. *J Neurophysiol* 81: 2558–2569, 1999.

Rieke F, Bodnar DA, Bialek W. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferent. *Proc Biol Sci* 262: 259–265, 1995.

Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W. *Spikes: Exploring the Neural Code.* Cambridge, MA: MIT Press, 1996.

Roddey JC, Girish B, Miller JP. Assessing the performance of neural encoding models in the presence of noise. *J Comput Neurosci* 8: 95–112, 2000.

Rozell CJ, Johnson DH. Examining methods for estimating mutual information in spiking neural systems. *Neurocomputing* 65–66: 429, 2004.

Sadeghi SG, Chacron MJ, Taylor MC, Cullen KE. Neural variability, detection thresholds, and information transmission in the vestibular system. *J Neurosci* 27: 771–781, 2007.

Shannon R. The mathematical theory of communication. *Bell Syst Tech J* 27: 379–423, 1948.

Sharafi N, Benda J, Lindner B. Information filtering by synchronous spikes in a neural population. *J Comp Neurosci* 34: 285–301, 2013.

Shlens J, Kennel MB, Abarbanel HD, Chichilnisky EJ. Estimating information rates with confidence intervals in neural spike trains. *Neural Comput* 19: 1683–1719, 2007.

Stamper SA, Carrera E, Tan EW, Fugére V, Krahe R, Fortune ES. Species differences in group size and electrosensory interference in weakly electric fishes: implications for electrosensory processing. *Behav Brain Res* 207: 368–376, 2010.

Stein RB, French AS, Holden AV. The frequency response, coherence and information capacity of two neuronal models. *Biophys J* 12: 295–322, 1972.

Stocks NG. Suprathreshold stochastic resonance in multilevel threshold systems. *Phys Rev Lett* 84: 2310–2313, 2000.

Stratonovich RL. *Topics in the Theory of Random Noise.* New York: Gordon and Breach, 1963.

Strong SP, Koberle R, van Steveninck RR, Bialek W. Entropy and information in neural spike trains. *Phys Rev Lett* 80: 197, 1998.

Vilela RD, Lindner B. Are the input parameters of white-noise-driven integrate and fire neurons uniquely determined by rate and CV? *J Theor Biol* 257: 90–99, 2009a.

Vilela RD, Lindner B. A comparative study of three different integrate-and-fire neurons: spontaneous activity, dynamical response, and stimulus-induced correlation. *Phys Rev E Stat Nonlin Soft Matter Phys* 80: 031909, 2009b.

Warland DK, Reinagel P, Meister M. Decoding visual information from a population of retinal ganglion cells. *J Neurophysiol* 78: 2336–2350, 1997.