Differentially synchronized spiking enables multiplexed neural coding

Milad Lankarany, Dheakra Al-Basha, Stéphanie Ratté, and Steven A. Prescott

Multiplexing refers to the simultaneous encoding of two or more signals. Neurons have been shown to multiplex, but different stimuli require different multiplexing strategies. Whereas the frequency and amplitude of periodic stimuli can be encoded by the timing and rate of the same spikes, natural scenes, which comprise areas over which intensity varies gradually and sparse edges where intensity changes abruptly, require a different multiplexing strategy. Recording in vivo from neurons in primary somatosensory cortex during tactile stimulation, we found that stimulus onset and offset (edges) evoked highly synchronized spiking, whereas other spikes in the same neurons occurred asynchronously. Stimulus intensity modulated the rate of asynchronous spiking, but did not affect the timing of synchronous spikes. From this, we hypothesized that spikes driven by high- and low-frequency stimulus features can be distinguished on the basis of their synchronization, and that differentially synchronized spiking can thus be used to form multiplexed representations. Applying a Bayesian decoding method, we verified that information about high- and low-frequency stimulus features can be recovered from an ensemble of model neurons receiving common input. Equally good decoding was achieved by distinguishing synchronous from asynchronous spikes and applying reverse correlation methods separately to each spike type. This result, which we verified with patch clamp recordings in vitro, demonstrates that neurons receiving common input can use the rate of asynchronous spiking to encode the intensity of low-frequency features while using the timing of synchronous spikes to encode the occurrence of high-frequency features. We refer to this strategy as synchrony-division multiplexing.

The nervous system processes phenomenal amounts of information despite biological constraints on the number of neurons and their maximal firing rate, plus other factors that limit signal-to-noise ratio (1). Such constraints necessitate efficient neural coding strategies. In telecommunication systems, efficiency is increased by sending multiple messages simultaneously over a single channel. This so-called multiplexing involves representing different messages in separate frequency bands (frequency-division multiplexing) or temporal epochs (time-division multiplexing), among other strategies (2). The brain can also multiplex (3–16), but different stimuli require different multiplexing strategies, many of which have yet to be elucidated.

Ideas about multiplexed neural coding date back nearly a century (17). In auditory nerve, the phase-locking of spikes to a periodic input enables temporal coding of stimulus frequency, whereas the probability of spiking per stimulus cycle (which is <1 when stimulus frequency exceeds the maximal firing rate) enables rate coding of stimulus intensity (18). In this scenario, the same spikes contribute to both codes (SI Appendix, Fig. S1A), but distinct cochlear nuclei (19) use high- or low-pass filtering to extract either the time- or rate-encoded information. The frequency and intensity of vibrotactile stimuli are similarly encoded by the timing and rate of spikes in somatosensory cortex (3). In these examples, intensity refers to the peak-to-peak amplitude, or envelope, of the periodic stimulus. But not all stimuli are periodic; for instance, natural scenes comprise areas over which intensity varies gradually (low contrast), and sparse edges in which intensity changes abruptly (high contrast) (20). If spikes evoked by low- and high-frequency features are interspersed in a given neuron’s spike train, the spikes evoked by each feature must be disambiguated to decode the feature. How might this occur?

High-contrast features tend to evoke spikes whose timing is more precise than spikes evoked by low-frequency features (21–23), but assessing spike-timing precision requires comparison with the stimulus or some other reference (e.g., other spikes). We hypothesized that precisely timed spikes driven by high-frequency features occur synchronously across neurons receiving common input, whereas other spikes driven by low-frequency features occur asynchronously. If different stimulus features evoke differentially synchronized spiking, information about each feature could be recovered from each spike “type.” In this scenario, unlike for multiplexed coding of periodic stimuli, the representation of each feature is based on separate (synchronous vs. asynchronous) spikes that occur in the same neurons (SI Appendix, Fig. S1B). Other scenarios are also possible; for instance, where synchronous and asynchronous spikes occur in different neurons (SI Appendix, Fig. S1C). The appropriate multiplexing strategy ultimately depends on the stimulus.

Starting with in vivo recordings from somatosensory cortex, we confirmed that the onset and offset of tactile stimuli (high-contrast

Significance

The nervous system processes phenomenal amounts of information. This processing must be conducted efficiently. In telecommunications systems, efficiency is increased by transmitting multiple signals through a single communication channel, or multiplexing. Neurons also multiplex. Here, we demonstrate a strategy for multiplexing different features of aperiodic stimuli: Cortical neurons use the rate of asynchronous spiking to encode stimulus intensity while using the timing of synchronous spikes to encode abrupt changes in stimulus intensity. This is possible because high-frequency features (edges) evoke spikes that transiently synchronize across neurons, whereas low-frequency features evoke sustained asynchronous spiking whose rate is proportional to stimulus intensity. Differentially synchronized spiking evoked in the same neurons by different stimulus features enables the formation of multiplexed representations.
features) evoked transiently synchronized spiking, whereas intervening spiking occurred asynchronously. The rate of asynchronous spiking varied with the intensity of sustained tactile input (a low-contrast feature). In other words, the rate of asynchronous spiking encoded stimulus intensity, whereas the timing of synchronous spikes encoded high-contrast edges. In subsequent simulations and in vitro recordings, we verified the importance of distinguishing between synchronous and asynchronous spikes to recover information about stimulus intensity and contrast.

**Results**

**Rate and Temporal Coding of Tactile Information in Primary Somatosensory (S1) Cortex.** To explore how the different aspects of tactile stimulation are encoded in S1 cortex, we recorded single units in vivo (Fig. 1A) while applying steps of increasing force to the whisker pad of lightly sedated rats. Sedation facilitated reproducible stimulation while avoiding the altered sensory processing caused by anesthesia (24). Firing rate histograms (FRHs) were calculated using a broad or narrow kernel to, respectively, track slow or fast changes in firing rate (25). The high-resolution FRH (in black) exhibited distinctive blips that, when appropriately thresholded, predict stimulus onset and offset with near-perfect sensitivity and specificity (Fig. 1B). These blips coincided with large negativities in the local field potential (Fig. 1C). An expanded view of the rasters (Fig. 1D) reveals that firing rate blips result from a volley of synchronous spikes; some neurons produce a rapid burst of two to four spikes, but most contribute a single spike to each volley. All spikes occurring while the high-resolution FRH exceeded threshold (red shading) were designated as synchronous and colored red; all other spikes were designated as asynchronous and colored blue.

The latency of synchronous spikes was insensitive to stimulus intensity (Fig. 1E), whereas the rate of asynchronous spiking correlated with stimulus intensity (Fig. 1F). Consistent with our hypothesis, these data argue that abrupt changes in stimulus intensity (i.e., high-contrast edges) are encoded by the transient synchronization of spiking, whereas intensity is encoded by the rate of asynchronous spiking. One may have expected synchronous and asynchronous spikes to be segregated to different neurons, as occurs upstream, but our data clearly show synchronous and asynchronous spikes interspersed in the same neurons.

**An Ensemble of Neurons Receiving Common Fast and Slow Signals Can Multiplex.** The ability of cortical neurons to multiplex using differentially synchronized spiking raises the issue of what inputs they receive. In particular, we must ask how stimulus intensity (a first-order feature) and changes in intensity, or contrast (a second-order feature), are represented in primary sensory neurons, as it is the output of these neurons that eventually becomes the input to cortical neurons. Similar to image compression algorithms that decompose scenes according to first- and second-order features (26), retinal ganglion neurons (27), somatosensory...
Slow and fast signals can be demultiplexed from responses to mixed signals. Neurons behaving as low-pass filters (integrators) encode stimulus intensity, whereas others behaving as high-pass filters (coincidence detectors, edge detectors) are sensitive to contrast. Differently tuned sensory neurons are coactivated by natural stimuli, which comprise independent first- and second-order features, yet each neuron type responds to (encodes) different stimulus features, meaning each feature is initially encoded by distinct sets of sensory afferents (Fig. 2A and SI Appendix, Fig. S2). Retinal ganglion cells sensitive to luminance or contrast, respectively, follow the parvocellular and magnocellular pathways to the brain (32), but those pathways at least partially reconverge (33–36). Distinctly tuned somatosensory (37, 38), vestibular (29), and auditory (19) similarly reconverge. Central representations tend to expand rather than contract, and although convergence can lead to representations being deliberately transformed through computations, data in Fig. 1 show that information initially conveyed by segregated pathways comes to be represented using different spike types in common cortical neurons.

Because only multiplexed representations can be demultiplexed, we tested whether multiplexing occurs by testing whether slow and fast signals, reflecting the activity evoked by low- and high-contrast features in integrators and coincidence detectors, respectively, can both be recovered (demultiplexed) from the spiking evoked by a mixed signal (Fig. 2B). Specifically, we compared standard reverse correlation and Bayesian decoding methods applied to simulated responses from a single neuron or from a 30-neuron ensemble (SI Appendix); spikes were not subdivided into different "types" for this initial analysis. The slow signal was recovered from the single-neuron and ensemble responses using either decoding strategy, but the fast signal was recovered only from the ensemble response, and only with the Bayesian method (Fig. 2C). These results show that multiplexing can occur, but that it requires a multineuron representation, consistent with the hypothesized role of synchrony for disambiguating spikes driven by different stimulus features.

Differentially Synchronized Spiking Enables the Formation of Multiplexed Representations. Guided by results in Fig. 1, we reanalyzed data in Fig. 2 after separating synchronous and asynchronous spikes. In Fig. 3, the low-resolution FRH (shown in green) shows rate fluctuations that track the intensity of the slow signal, whereas the high-resolution FRH (shown in black) exhibits blips representing synchronous spikes that coincide with events in the fast signal. The high-resolution FRH was thresholded to separate synchronous spikes (shown in red) from asynchronous spikes (shown in blue). Pairwise cross-correlograms (CCGs) constructed from each spike type are distinct (Fig. 3B), and their superposition explains CCGs with a broad base and narrow peak (Discussion).

If each spike type is driven by a different component of the mixed signal, the spike-triggered average (STA) calculated from the mixed signal should be distinct for each spike type. Consistent with slow and fast signals driving asynchronous and synchronous spikes, respectively, the STA calculated from the slow signal using asynchronous spikes (STA_{slow}^{async}) was broad, whereas the STA calculated from the fast signal using synchronous spikes (STA_{fast}^{sync}) was narrow (Fig. 3C, dark blue and red STAs). Access to component signals is unnecessary, as similar STAs were recovered from the mixed signal, using asynchronous or synchronous spikes (pale-colored STAs). Consistent with the slow signal not driving synchronous spikes and the fast signal not driving asynchronous spikes, STA_{slow}^{async} and STA_{fast}^{sync} were unstructured (gray STAs). Given that STAs are differently shaped depending on which spike type is used for triggering, we reasoned that reverse correlation could be improved by convolving each spike type with its respective STA, rather than convolving all spikes with STA_{uni}. As expected, the fast signal was recovered by convolving synchronous spikes with STA_{fast}^{sync} or STA_{fast}^{async}, and the slow signal was recovered by convolving asynchronous spikes with STA_{slow}^{async} or STA_{slow}^{sync} (Fig. 3C, Bottom). This decoding strategy, which we term synchrony-based demultiplexing, matched the performance of Bayesian decoding (Fig. 3D). Failure of Bayesian decoding to recover the fast signal when the encoding model was prevented from learning the fast signal (Fig. 3D, light green and SI Appendix) confirms that identifying each spike type is necessary for the Bayesian model’s performance. In short, separating synchronous and asynchronous spikes is necessary and sufficient for demultiplexing under the conditions tested.

For synchrony-based demultiplexing to occur in the brain, a biologically implementable decoder must be able to distinguish between synchronous and asynchronous spikes. This is readily achieved through high- or low-pass filtering (Fig. 3E), which can be implemented by cellular (31), synaptic (39), or microcircuit (40) mechanisms. Our data do not demonstrate how or where demultiplexing occurs, but the observation that spike timing and rate both contribute to tactile perception (4) argues that it must occur if the underlying representations in S1 cortex are multiplexed, as shown in Fig. 1.

**Figure 2.** Slow and fast signals can be demultiplexed from responses to mixed signal. (A) Basis for mixed signal. (A, Top) Decomposition of image by sensory neurons behaving as low-pass (LP) or high-pass (HP) filters. (A, Bottom) Gray-scale intensity along cut through LP image corresponds to slow signal representing luminance, a first-order stimulus feature. Same cut through HP image yields a series of discrete events representing edges and other areas of high contrast, a second-order stimulus feature. Convergence of those signals creates a “mixed” signal. See also SI Appendix, Fig. S2. (B) Sample rasters from 10 model neurons (Bottom) receiving a common mixed signal (Top) and independent noise (not illustrated). Spiking evoked by the fast component is not obviously different from spiking evoked by the slow component. (C) Decoding of the mixed signal using standard reverse correlation (orange) or a Bayesian decoding method (green) applied to the response of a single neuron (open bars) or a 30-neuron ensemble (filled bars). (Inset) Original mixed signal (black) overlaid with signal reconstructed from the ensemble response (color; Methods and SI Appendix, Fig. S3). Signal reconstruction was quantified as coding fraction, \( CF = 1 - \frac{\text{original} - \text{reconstructed}}{\text{original}} \), where \( CF = 1 \) represents perfect reconstruction and \( CF \leq 0 \) represents failure to explain any variance.
record pyramidal neurons in a slice preparation of mouse S1 cortex. A noisy, high-conductance state (41) was recreated using dynamic clamp; conductance noise was different for each neuron and each trial, whereas the mixed signal was the same for all neurons and all trials, as in simulations. Neurons were recorded sequentially, but responses were aligned on the basis of the common mixed signal. As expected, synchronous spikes coincided with events in the fast signal, whereas the rate of asynchronous spiking tracked intensity of the slow signal (Fig. 4A). The same decoding strategies applied to simulation data yielded comparable results when applied to these experimental data (Fig. 4B). Additional testing of neurons in a noisy, low-conductance state (Fig. 4C) or without any added noise (Fig. 4D) yielded similar results, indicating that synchrony-division multiplexing is feasible across a broad range of conditions.

**Discussion**

Our results demonstrate that small ensembles of cortical neurons can multiplex by using the rate of asynchronous spikes to encode stimulus intensity and the timing of synchronous spikes to encode high-contrast features such as edges. We refer to this as synchrony-division multiplexing because different stimulus features are represented by spikes that are differentiated by their degree of synchrony, as opposed to being represented in different frequency bands or temporal epochs (as in frequency- or time-division multiplexing). Importantly, synchronous and asynchronous spikes occur in the same cortical neurons, which is unlike the segregated representation of intensity and contrast across dichotomously tuned primary sensory neurons (28).

To the best of our knowledge, this is the first time multiplexing on the basis of differentially synchronized spiking in the same neurons has been described, although the nervous system has been shown to use other multiplexing strategies. For instance, the rate and timing of spikes in auditory afferents have been shown to respectively encode stimulus intensity and frequency (18), and cochlear nuclei are known to demultiplex those representations (19). We have not demonstrated whether or how multiplexed representation in S1 neurons are demultiplexed, but simple high- or low-pass filtering similar to processing carried out in the cochlear nuclei would suffice (Fig. 3E). Demultiplexing is likely to occur, given that tactile perception depends on both the timing and rate of spikes in S1 neurons (4). Indeed, there is extensive work showing that spike timing and rate both carry information about tactile input. Our results do not contradict those findings but, rather, help reconcile them by revealing how different coding schemes can coexist. It is easy to overlook synchrony-division multiplexing by testing with simple stimuli, measuring only certain aspects of a multifaceted response, or neglecting what happens in neighboring neurons, but it is important to recognize that different spikes within the same spike train may encode different information.

Information about the intensity and frequency of a periodic signal such as sound (17, 18) or vibrotactile input (3) is carried by the same spikes, unlike in synchrony-division multiplexing, where...
Pyramidal neurons can multiplex in vitro. (A) Sample rasters from short length of 100-s-long responses. All neurons received the same mixed signal but different conductance noise on each trial. Four neurons were tested with 7 trials each; brackets on left group responses by neuron. Black FRH was thresholded to identify synchronous (red) or asynchronous (blue) spikes. (B) Decoding of the mixed signal from ensemble response (28 trials) illustrated in A, based on neurons tested in the noisy, high-conductance state. Same analysis conducted on neurons tested in a noisy, low-conductance state (31 trials from six neurons) (C) or in neurons without any added noise (29 trials from five neurons) (D). For all different conditions, the four decoding strategies yielded a pattern of CF values very similar to that seen in simulations (Fig. 3D).

Diversity of which reflects the diverse signals the nervous system must efficiently process.

Materials and Methods

All procedures were approved by the Animal Care Committee of The Hospital for Sick Children. Responses of single units to tactile stimulation of the whisker pad were recorded from S1 cortex of five rats lightly sedated with fentanyl. In separate patch clamp recordings, responses to mixed signals applied through the recording pipette were recorded from layer 5 pyramidal neurons in mouse brain slices. Conductance-based neuron models were simulated using Morris-Lecar equations. See SI Appendix for details. Experimental data and code for simulations and analysis are available from prescottlab.ca.

Reverse Correlation. Spikes were passed through a linear filter to reconstruct the signal that evoked them. The filter used here was the STA. The STA kernel was calculated (using training data) by averaging the signal $x$ over a certain time window (kernel width = 50 ms) preceding each spike (SI). The reconstructed signal was obtained by convolving spikes (using test data) with the STA kernel

$$s_{\text{est}} = R_{\text{obs}}(t) * \text{STA}^{(t)}_{\text{obs}} = \int R_{\text{obs}}(q) \text{STA}^{(t)}_{\text{obs}}(t - q) dq,$$

where $s_{\text{est}}$ is the estimated signal and $R_{\text{obs}}$ is the observed response, which includes all spikes for standard reverse correlation but only specific spike types for demultiplexing as described under Synchrony-Based Demultiplexing.

Synchrony-Based Demultiplexing. By using different spike types ($y$) as the trigger and drawing the spike-triggered ensemble from different components of the signal ($x$), we calculated STA variants, denoted STAx. Reverse correlation was carried out as described earlier, but now $R_{\text{obs}}$ comprising only specific spike types was convolved with the respective STA to reconstruct each component signal, which were then summed to reconstruct the mixed signal.

Bayesian Decoding. Given an encoding model (SI Appendix), we applied a Bayesian strategy to estimate the most likely input given an observed output. The rate-based encoding model was fit to the firing rate observed in a conductance-based model neuron or ensemble thereof during stimulation with $I_{\text{noise}}$ and $I_{\text{cover}}$.

Quantification of Signal Reconstruction. Reconstruction of the original signal was quantified as coding fraction (CF) (52) and is equivalent to variation accounted for (16), where
and \(|\|_{2}\) indicates the norm 2. CF lies within \([-1, 1]\), where 1 represents perfect reconstruction. Negative values occur when the SD of the difference between original and reconstructed signals is larger than that of the original signal, but are reported here simply as <0. See SI Appendix for further details.

Supplementary Information for

Differentially synchronized spiking enables multiplexed neural coding

Milad Lankarany, Dhekra Al-Basha, Stéphanie Ratté, and Steven. A. Prescott

Corresponding author: Steven A. Prescott
Email: steve.prescott@sickkids.ca

This PDF file includes:

Supplementary methods
Figs. S1 to S5
References for SI reference citations
SUPPLEMENTARY METHODS

**In vivo recordings**

*Surgery.* Adult male (230-300 g) Sprague Dawley rats (Charles River, Montreal, Quebec) were anesthetized with isoflurane (3% for induction, 2% for maintenance). The upper thoracic region was shaved and anesthetic cream (2.5% lidocaine, 2.5% prilocaine; AstraZeneca, Mississauga, ON) was applied to the shaved skin 10-15 min prior to surgery. To cannulate the jugular vein, a small incision (~10 mm) was made just below the left clavicle. The jugular vein was separated from surrounding tissue and two loose knots were made around it. Immediately after tightening the rostral knot, the vein was incised and saline-primed PE60 tubing (Becton Dickinson, Mississauga, ON) was inserted and fixed in position by tightening the second knot. After ensuring that there was no resistance to flow, saline was delivered at a rate of 0.5 ml/hr using a Legato 101 syringe pump (KD Scientific, Holliston, MA) and the incision was closed. Next, the trachea was exposed by making a midline incision (~ 25 mm) between the lower jaw and upper thorax. The trachea was incised below the larynx and PE240 tubing (Becton Dickinson) already connected to an Inspira ventilator (Harvard Apparatus, Saint-Laurent, QC) was inserted into the trachea and secured, and the incision was closed. The rat was ventilated at 85 breaths/min and tidal volume of 2 cc.

The rat was transferred to a stereotaxic frame (Narishige, Amityville, NY) and its head fixed with ear bars. The head was shaved and anesthetic cream applied before exposing the skull. A 3x3 mm window was made through the skull over the right primary somatosensory cortex (S1), 1-4 mm caudal and 2.5-5.5 mm lateral to bregma. Ear bars were removed and the skull was cemented to a custom-made head plate and whiskers on the left side were trimmed. Heart rate and oxygen level were monitored (Nonin, Plymouth, MN) and body temperature was maintained at 37˚C using a feedback-controlled heating pad (TR-200; Fine Science Tools, Foster City, CA). After giving a loading dose of fentanyl (10 µg/kg) and pancuronium bromide (1.6 mg/kg), both drugs were infused continuously at 10 µg/kg/hr and 1.6 mg/kg/hr, respectively.

*Single unit extracellular recordings:* All recordings were conducted in a dimly lit and quiet room. A multielectrode array comprising 4 shanks each with 4 recording sites (A4 type, NeuroNexus, Ann Arbor, MI) connected to an Omniplex data acquisition system (Plexon, Dallas, TX) was lowered into the S1 cortex. Isoflurane was discontinued while monitoring the animal. Signals were amplified, digitized at 40 kHz, and high-pass filtered at 300 Hz. Receptive fields were identified by gently touching the whisker pad with a paint brush or blunt probe. Recordings were obtained 2.5-3.5 mm caudal and 4-5.5 mm lateral to bregma, 1.4-1.8 mm deep (layer 5).
Once a receptive field (RF) was identified, a Model 300C-I force-feedback mechanical stimulator (Aurora Scientific, Aurora, Ontario) with 1 mm diameter-wide blunt plastic tip was positioned with a micromanipulator. The timing and force of all stimuli were controlled (and the resulting force recorded) using a Power1401 computer interface and Signal v5 software (Cambridge Electronic Design). Triggers sent from Signal were used to sync stimulation with neural recordings. Spikes were sorted using the Plexon offline sorter. Spike times were exported to Matlab for analysis described below.

A total of 44 well isolated single units were recorded from 5 rats. Stimulus sequences (trials) were repeated ≥4 times per animal but, for those units selected (see below), 4 trials were randomly drawn from each so that each unit contributed equally to the final data set. To identify units responsive to whisker pad stimulation, the firing rate evoked by 10, 12.5 and 15 g stimulation was averaged and compared to the spontaneous firing at the start of that trial. Units whose evoked firing rate was significantly greater than the spontaneous firing rate (p < 0.05, paired t-test) were included for further analysis, resulting in a total of 17 units. Units were not selected on the basis of any information about spike timing.

**In vitro recordings**

*Slice preparation.* In vitro recordings were conducted on adult (6-8 week old) mice of either sex derived by crossing Pvalb-2A-Cre-D mice (JAX #012358) with Ai32 mice (JAX #012569). Offspring express channelrhodopsin-2 in parvalbumin-expressing interneurons but experiments targeted pyramidal neurons and did not involve optogenetic stimulation. Mice were anesthetized with 3% isoflurane and decapitated as previously described (1). The brain was rapidly removed to ice-cold carbogenated (95% O₂ and 5% CO₂) sucrose-substituted artificial CSF (ACSF) containing (in mM) 252 sucrose, 2.5 KCl, 2 CaCl₂, 2 MgCl₂, 10 glucose, 26 NaHCO₃, 1.25 NaH₂PO₄, and 5 kynurenic acid. Coronal slices of primary somatosensory cortex (S1) were cut at 400 µm thickness using a VT-1000S microtome (Leica, Concord, Ontario) and were kept in regular ACSF (126 mM NaCl instead of sucrose and without kynurenic acid) at room temperature until recording. Slices were transferred to a recording chamber perfused with ACSF maintained at 31±1°C.

*Patch clamp recordings.* Pyramidal neurons in layer 5 of S1 were recorded in whole-cell mode with >70% series resistance compensation using an Axopatch 200B amplifier (Molecular Devices, Sunnyvale, CA). The pipette solution contained (in mM) 125 KMeSO₄, 5 KCl, 10 HEPES, and 2 MgCl₂, 4 ATP (Sigma), and 0.4 GTP (Sigma); pH was adjusted to 7.2 with KOH. Synaptic transmission was blocked via bath application of (in µM) 10 CNQX, 40 D-AP-5, and 6 gabazine.
The same mixed signal was applied on each trial via current injection with or without added noise, which if present, differed across trials (see below). Responses were low-pass filtered at 2 kHz and digitized at 20 kHz using a Power1401 computer interface and Signal 5 software (Cambridge Electronic Design, Cambridge, UK).

To recreate the noisy, high-conductance state observed in vivo (2), irregularly fluctuating conductances generated by Ornstein-Uhlenbeck (OU) processes \(g_{\text{exc}} = 1 \text{nS}, g_{\text{inh}} = 4 \text{nS}, \sigma_{\text{exc}} = 0.3 \text{nS}, \sigma_{\text{inh}} = 0.75 \text{nS}, \tau_{\text{exc}} = 3 \text{mS} \text{ and } \tau_{\text{inh}} = 10 \text{ms}\) were applied using dynamic clamp. Alternatively, noise was introduced as a fluctuating current modeled as an OU process \(\sigma_{\text{noise}} = 10 \text{pA}, \tau_{\text{noise}} = 5 \text{mS}\). In either case, noisy input caused membrane potential fluctuations of \(\sim 2 \text{mV}\) and spontaneous spiking of \(\sim 5 \text{spikes/s}\). The same mixed signal by different noise was applied for each 100 s-long trial. Other trials were conducted with the same mixed signal but without any added noise. Data were included from all cells in which \(\geq 4\) trials were completed for at least one condition. No more than 7 trials per condition were collected per cell. Data collection was stopped when the data set for each condition neared or exceeded 30 trials.

**Simulations**

Neurons were modeled as previously described (1) using equations adapted from Morris and Lecar (3):

\[
C \frac{dV}{dt} = I_{\text{mixed}}(t) + I_{\text{noise}}(t) - g_{Na}m_\infty(V)(V - E_{Na}) - g_{Kw}(V - E_K) - g_{L}(V - E_{L}) - g_{\text{exc}}(V - E_{\text{exc}}) - g_{\text{inh}}(V - E_{\text{inh}}) \tag{S1}
\]

\[
dw/dt = \frac{w(V) - w}{\tau_w(V)} \tag{S2}
\]

\[
dz/dt = \frac{1}{1 + e^{(\beta - V)/\gamma_z}} - z / \tau_z \tag{S3}
\]

\[
m_\infty(V) = 0.5 \left[1 + \tanh\left(V - \beta_m / \gamma_m\right)\right] \tag{S4}
\]

\[
w_\infty(V) = 0.5 \left[1 + \tanh\left(V - \beta_w / \gamma_w\right)\right] \tag{S5}
\]

\[
\tau_w(V) = 1 / \cosh\left(V - \beta_w / 2\beta_w\right) \tag{S6}
\]
where $g_{\text{Na}} = 20 \text{ mS/cm}^2$, $g_{\text{K}} = 20 \text{ mS/cm}^2$, $g_{\text{L}} = 2 \text{ mS/cm}^2$, $g_{\text{AHP}} = 25 \text{ mS/cm}^2$, $g_{\text{exc}} = 1.2 \text{ mS/cm}^2$, $g_{\text{inh}} = 1.9 \text{ mS/cm}^2$, $E_{\text{Na}} = 50 \text{ mV}$, $E_{\text{K}} = -100 \text{ mV}$, $E_{\text{L}} = -70 \text{ mV}$, $E_{\text{exc}} = 0 \text{ mV}$, $E_{\text{inh}} = -70 \text{ mV}$, $\beta_m = -1.2 \text{ mV}$, $\gamma_m = 18 \text{ mV}$, $\beta_w = -19 \text{ mV}$, $\gamma_w = 10 \text{ mV}$, $\beta_z = 0 \text{ mV}$, $\gamma_z = 2 \text{ mV}$, $\tau_a = 20 \text{ ms}$, $\phi = 0.15$, and $C = 2 \mu\text{F/cm}^2$. These parameter values produce a hybrid operating mode (4). Notably, inclusion of background excitatory and inhibitory synaptic conductance reproduces a “balanced” high-conductance state (2). Surface area was set to 200 $\mu\text{m}^2$ so that $I_{\text{mixed}}$ is reported in pA, like in experiments (see below), rather than as a density.

The mixed signal ($I_{\text{mixed}}$) is the sum of fast ($I_{\text{fast}}$) and slow ($I_{\text{slow}}$) signals. Rather than high- and low-pass filtering a common signal and then combining the outputs (as in Figs. 2A and S1), we created comparable mixed signals as described below. $I_{\text{fast}}$ was generated by convolving a randomly (Poisson) distributed Dirac delta function with a synaptic waveform with a normalized peak amplitude of 1, $\tau_{\text{rise}} = 0.5 \text{ ms}$ and $\tau_{\text{fall}} = 3 \text{ ms}$. Fast events occurred at a rate of $\sim 1 \text{ Hz}$ and were scaled by $a_{\text{fast}} = 170 \text{ pA}$. $I_{\text{slow}}$ was generated using an OU process, $\frac{dI_{\text{slow}}}{dt} = -\frac{I_{\text{slow}}(t) - \mu}{\tau} + \sigma \sqrt{\frac{2}{\tau}} \xi(t)$ \hspace{1cm} (S7)

where $\xi$ is a random number drawn from a Gaussian distribution with 0 average and unit variance, $\tau = 100 \text{ ms}$ to produce a slow-varying random walk with average $\mu = 30 \text{ pA}$ and $\sigma = 120 \text{ pA}$. By generating $I_{\text{fast}}$ and $I_{\text{slow}}$ independently and summing them to form $I_{\text{mixed}}$, the two components of $I_{\text{mixed}}$ are sure to be independent. An OU process with $\tau = 5 \text{ ms}$, $\mu = 0 \text{ pA}$, and $\sigma = 1 \text{ pA}$ was used to create noisy current ($I_{\text{noise}}$). The same instantiation of $I_{\text{mixed}}$ but different instantiations of $I_{\text{noise}}$ were applied to all neuron models within a set.

**Quantification of signal reconstruction**

Reconstruction of the original signal was quantified as coding fraction (CF), as explained in the main text. When synchrony-based demultiplexing was not compared back-to-back with a different decoding strategy (Fig. 3C), CF was calculated by comparing the reconstructed fast and slow signals with the original fast and slow signals. In other cases (Figs. 2C, 3D and 4B-D), the reconstructed mixed signal was separated into fast and slow components by filtering (see below) so that CF could be measured for each component; in other words, because most decoding strategies reconstruct the mixed signal rather than the individual components, the individual components reconstructed through demultiplexing were added and then separated (the same way as other reconstructed signals were) to prevent differences in CF values from arising through technical differences in how original and reconstructed signals are compared (see Fig. S3).
**Bayesian Modeling**

Bayesian decoding (see main text) requires an encoding model. Parameters of the encoding model were fit to a 20 s-long training set by maximizing the log-likelihood function, 

$$
\sum_i R_{obs}(t) \log(\lambda(t)) - \lambda(t),
$$

using the GLM method (5), where $$\lambda(t) = \exp(kx + \mu)$$ represents the conditional intensity (or instantaneous rate) that depends on a linear filter $$k$$, stimulus $$x$$, and the cell’s baseline log-firing rate $$\mu$$, and $$R_{obs}(t)$$ represents the observed spikes (superimposed from all conductance-based neuron models). In the test set, to estimate $$x(t)$$, we drew samples $$s_i$$ ($$i=1:1000$$) from a Gaussian distribution $$P(x)$$ whose standard deviation was $$1 \times$$ or $$0.1 \times$$ the standard deviation of the true input distribution. The lower standard deviation prevented the encoding model from learning to encode fast stimulus fluctuations (see Fig. 3D, light vs. dark green). For each time $$t$$, we computed the conditional probability $$P(R_{obs}|s_i)$$ to measure the likelihood that the observed population response was generated by stimulus $$s_i$$. From Bayes rule, we know that the posterior $$P(s_i|R_{obs})$$ is proportional to $$P(R_{obs}|s_i)$$. Thus, the Bayes least squares estimate is given by

$$
\hat{x}(t) = \frac{\sum_i s_i(t)P(R_{obs}|s_i(t))}{\sum_i P(R_{obs}|s_i(t))},
$$

(S8)
**SUPPLEMENTARY FIGURES**

**Figure S1. Comparison of multiplexing strategies that use spike rate and timing to represent different stimulus features.** Differences boil down to whether rate- and time-based representations comprise the same or different spikes occurring in the same or different neurons. (A) In auditory afferents, the same spikes in the same neurons represent different stimulus features. Spikes in all neurons occur at a preferred phase of the periodic stimulus, and thus occur synchronously at an interval reflecting stimulus frequency. But spikes do not occur on every stimulus cycle; instead, their probability varies with stimulus intensity, thus enabling the firing rate to encode stimulus intensity (6). (B) In primary somatosensory (S1) cortex, different spikes in the same neurons represent different stimulus features. During low-contrast stimulation, spikes occur asynchronously...
at a rate proportional to stimulus intensity. High-contrast features (i.e. abrupt changes in intensity) transiently synchronize spiking across all (most) neurons, enabling synchronous spikes to encode contrast. (C) In the olfactory bulb, different spikes in different neurons represent different stimulus features. A small subset of mitral cells phase-lock to network oscillations that develop during an odor; the synchronized spikes occurring in this minority of phase-locked cells encode odor category whereas odor identity is represented by the rate of asynchronous spiking in mitral cells that do not phase-lock (7).

(B vs C) In S1 cortex, each high-contrast stimulus transiently synchronizes spikes across most neurons but spikes occur asynchronously outside those brief (5-10 ms-long) epochs; in other words, transiently synchronized spiking does not preclude asynchronous spiking in the same neuron (as long as high-contrast features are relatively infrequent). Conversely, network oscillations in the olfactory bulb are protracted. If mitral cells produce mostly synchronous spikes while phase-locked, then nearly all asynchronous spikes occur in mitral cells that do not phase lock (see Supplementary Fig. 1 of ref. 8). Phase-locked neurons could conceivably produce asynchronous spikes (e.g. if they fired at rates much higher than the network oscillation frequency), but that does not appear to be the case. At the very least, different mitral cells contribute differentially to time- and rate-based representations depending on how strongly they phase lock, which is unlike the equal contribution of S1 neurons to each code.

(B vs A) In auditory nerve, the number of spikes per stimulus cycle reflect the intensity (envelope) of the periodic input. In S1 cortex, in the absence of a periodic input, asynchronous spikes encode the intensity of the low-contrast (non-synchronizing) input, not the envelope of a synchronizing input. Temporal coding uses synchronized spikes in both cases but whether the rate-based code involves the same or different (synchronous vs. asynchronous) spikes depends on the nature of the stimulus.

In many sensory systems, different stimulus features can be simultaneously encoded using rate- and time-based codes, but there is no one-size-fits-all multiplexing strategy. Instead, whether multiplexed representations involve the same or different spikes occurring in the same or different neurons depends on the stimulus.
**Figure S2.** Visual analogue of tactile stimulus used for *in vivo* experiments reported in Figure 1: bars of increasing luminance represents steps of increasing force. Decomposition of this image into first- and second-order stimulus features is shown like in Figure 2A to depict the processing conducted by dichotomously tuned upstream sensory neurons. Neurons at the earliest stage of many sensory pathways behave as low- or high-pass filters such that they spike preferentially to first- or second-order stimulus features, respectively. The eventual re-convergence of pathways carrying information about these stimulus features gives a mixed signal to the neuron(s) of interest.
Figure S3. Reconstructing fast and slow components of a mixed signal (A). Left: Schematic shows original signals (left) given to a conductance-based model neuron or ensemble thereof, or to a rate-based model fitted to the output of the conductance-based model neuron(s) (middle), and the signal reconstructed from each model’s output (right). Synchrony-based demultiplexing directly reconstructs the fast and slow components whereas all other decoding strategies reconstruct the mixed signal, which must then be split into fast and slow components for comparison with the original fast and slow signals. To make comparison between decoding strategies fair, the fast and slow signals reconstructed through demultiplexing were summed to give a mixed signal that was then split the same way as for all other decoding strategies (except for CF calculations reported in Figure 3C, since there was no comparison across decoding strategies in that case). Fast and slow signals recovered by splitting the mixed signal are marked (*). Middle: Mixed signal was split by
applying a low-pass filter with a 30 Hz cutoff frequency (designed by Matlab R2014b command “designfilt”) to isolate the slow signal from the fast signal. Right: Traces show examples of the fast and slow components isolated from a mixed signal using this approach. (B) Example of the original mixed signal (black) and its reconstruction from the response of a single neuron using reverse correlation (orange) or Bayesian decoding (green). Both decoding strategies recovered the slow signal but neither recovered the fast signal. See Figure 2C for corresponding CF values. (C) Same as B but with reconstructions based on the response of the 30-neuron ensemble. See Figure 3D for corresponding CF values. (D) To link discriminability of the fast and slow signals with spike type, local maxima in the original mixed signal were plotted against local maxima in the reconstructed signal using an x or o depending on whether coincident spikes were synchronous or asynchronous, respectively. Distributions of x and o were separable in the original signal (for the stimulus parameters used here) but, for reconstructed signals, this was only true for demultiplexing and Bayesian decoding in which fast events were successfully learned (i.e. high standard deviation).
Figure S4. Comparison of the distribution of input (A) and output (B) measures labeled in Figure 3A. The distribution of intervals between synchronous spikes (x_{response}) resembles the distribution of intervals between events in the fast signal (x_{signal}) (red) whereas the distribution of asynchronous spike rate (y_{response}) resembles the distribution of intensity of the slow signal (y_{signal}) (blue). This dichotomy is consistent with synchronous and asynchronous spikes forming distinct representations of different features of the mixed signal.
Figure S5. Mutual information (MI) (A) between all spikes and the mixed signal, (B) between asynchronous spikes and the slow signal, and (C) between synchronous spikes and the fast signal. (D) The sum of MI calculated in B and C is greater than that calculated in A, thus showing that bandwidth is increased by synchrony-division multiplexing. As a final comparison, we computed the coding efficiency (CE) by normalizing MI by the total entropy of the spike train. When CE equals 1 (ideal case), every variation in the output spike trains corresponds to a unique (distinguishable) change in the input signal (9). For asynchronous spikes, CE can be calculated by normalizing the lower bound of information to its upper bound, which yields $CE_{async} = 0.76$. Note that for synchronous spikes, the total entropy of the spike train does not exceed that of the fast events unless the noise level is unrealistically high ($\sigma_{\text{noise}} > 80$ pA, which drives voltage fluctuation with standard deviation $\sigma_{V} > 16$ mV). Hence, the CE of synchronous spikes is
calculated by normalizing the MI by the total entropy of the fast events (7.9 bit/sec), which yields \( CE_{\text{sync}} = 0.97 \), which implies that the coding efficiency of synchronous spikes is nearly ideal. The difference in coding efficiency between synchronous and asynchronous spikes reflects the different coding scheme employed by each spike type.

Methods: For panels A and B, the signal was reconstructed and compared against the original signal to determine the reconstruction error, from which the signal-to-noise ratio (SNR) and lower bound of the information rate were computed (10-12):

\[
\text{SNR}(f) = \frac{S(f)}{\text{Err}(f)}
\]

where \( S(f) \) and \( \text{Err}(f) \) are the Fourier transform of the signal \( S(t) \) and the error between the original and reconstructed signals, i.e., \( \text{Err}(t) = S(t) - S_{\text{est}}(t) \). For A, \( S(t) \) corresponds to the mixed signal. For B, \( S(t) \) corresponds to the slow signal.

The upper bound on mutual information corresponds to trial-to-trial variability in the neural responses to repeated presentations of the same input (13). Given the responses \( R_i, i=1, 2, \ldots, L \), obtained from \( L \) presentations of the same input \( S(t) (L=3) \), one can compute the coherence between these responses as

\[
C_{RR}(f) = \frac{1}{L(L-1)} \sum_{i=2}^{L} \sum_{j=1}^{i-1} \left[ P_{R_i,R_j}(f) \right]^2
\]

where \( C_{RR}(f) \) is the response-response coherence, \( P_{R_iR_j} \) is the cross-spectrum (Fourier transform of the cross-correlation function) between responses \( R_i \) and \( R_j \). The response-response coherence, at each frequency \( f \), represents the strength of correlation between neural responses obtained from repeated presentations of the same input (14). Any trial-to-trial variability in the response to repeated presentations of the same stimulus will decrease the response-response coherence. The upper bound on MI can be estimated from that coherence according to (15-18):

\[
\text{Info}_{\text{UB}}(R, S) = -\int_{-f_{\text{cutoff}}}^{f_{\text{cutoff}}} \log_2 \left[ 1 - \sqrt{C_{RR}(f)} \right] df
\]

To calculate MI between synchronous spikes and the fast signal in C, the stimulus and response were treated as a series of 0s and 1s where 1s represent time bins in which synchronous
spiking or a fast signal event occurred. A bin size of 5 ms was used throughout. For the stimulus, because the timing of fast events was generated from an independent and identical distribution (i.i.d), the probability of having 1s or 0s is independent across bins; because fast events are sparse (i.e. << than the maximal firing rate), that independence also holds true for the response. The independence across bins enables us to estimate the probability of having 1 or 0 at each bin by counting the number of 1s and 0s during the simulation. Moreover, the joint probability of fast events and synchronous events can be estimated in the same manner. Then, the MI rate between fast-events and synchronous events can be calculated as

\[ MI_\Delta (R, S) = \sum_{x \in R_{\text{fast}}} \sum_{y \in S_{\text{fast}}} p(x, y) \log_2 \left( \frac{p(x, y)}{p(x)p(y)} \right) \]  

(S13)

where \( R \) and \( S \) represent synchronous spiking and events in the fast signal, respectively, and \( MI_\Delta \) represents the mutual information between two random binary strings in a bin of \( \Delta = 5 \) ms. Eq.13 can be simplified as

\[ MI_\Delta (R, S) = \sum_{i \in \{0,1\}} \sum_{j \in \{0,1\}} p(r = i, s = j) \log_2 \left( \frac{p(r = i, s = j)}{p(r = i)p(s = j)} \right) \]  

(S14)

where \( p(z = i) = \sum_{j \in \{0,1\}} p(z = i, y = j) \)  

(S15)

and \( z \) represents \( r \) or \( s \). Coding Efficiency (CE) was calculated according to (9)

\[ CE = \frac{MI(R, S)}{H(R)} \]  

(S16)

where \( H(R) \) is the total entropy of the spike train (with the same unit as \( MI \)).
SUPPLEMENTARY REFERENCES