

# Generalized analog thresholding for spike acquisition at ultralow sampling rates

Bryan D. He,<sup>1,2</sup> Alex Wein,<sup>1,3</sup> Lav R. Varshney,<sup>4</sup> Julius Kusuma,<sup>5</sup> Andrew G. Richardson,<sup>6</sup> and Lakshminarayan Srinivasan<sup>1</sup>

<sup>1</sup>Neural Signal Processing Laboratory, Department of Radiology, Stanford University, Stanford, California; <sup>2</sup>Department of Computer Science, California Institute of Technology, Pasadena, California; <sup>3</sup>Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts; <sup>4</sup>Coordinated Sciences Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois; <sup>5</sup>Department of Mathematics and Modeling, Schlumberger-Doll Research, Cambridge, Massachusetts; and <sup>6</sup>Translational Neuromodulation Laboratory, Department of Neurosurgery, University of Pennsylvania, Philadelphia, Pennsylvania

Submitted 19 August 2014; accepted in final form 15 April 2015

**He BD, Wein A, Varshney LR, Kusuma J, Richardson AG, Srinivasan L.** Generalized analog thresholding for spike acquisition at ultralow sampling rates. *J Neurophysiol* 114: 746–760, 2015. First published April 22, 2015; doi:10.1152/jn.00623.2014.—Efficient spike acquisition techniques are needed to bridge the divide from creating large multielectrode arrays (MEA) to achieving whole-cortex electrophysiology. In this paper, we introduce generalized analog thresholding (gAT), which achieves millisecond temporal resolution with sampling rates as low as 10 Hz. Consider the torrent of data from a single 1,000-channel MEA, which would generate more than 3 GB/min using standard 30-kHz Nyquist sampling. Recent neural signal processing methods based on compressive sensing still require Nyquist sampling as a first step and use iterative methods to reconstruct spikes. Analog thresholding (AT) remains the best existing alternative, where spike waveforms are passed through an analog comparator and sampled at 1 kHz, with instant spike reconstruction. By generalizing AT, the new method reduces sampling rates another order of magnitude, detects more than one spike per interval, and reconstructs spike width. Unlike compressive sensing, the new method reveals a simple closed-form solution to achieve instant (noniterative) spike reconstruction. The base method is already robust to hardware nonidealities, including realistic quantization error and integration noise. Because it achieves these considerable specifications using hardware-friendly components like integrators and comparators, generalized AT could translate large-scale MEAs into implantable devices for scientific investigation and medical technology.

spike acquisition; multielectrode arrays; sub-Nyquist; finite rate of innovation; brain initiative

THE CAPABILITY TO SIMULTANEOUSLY record the individual activity of most neurons in a brain represents a major goal of present-day neurotechnology research, because it would allow neuroscientists and clinicians to bridge the gap between single-cell function and whole-organism behavior (Alivisatos et al. 2013b; Carmena 2013; Lebedev 2014; Marblestone et al. 2013). Although substantial hardware challenges remain, total channel counts in large-scale multielectrode arrays (MEA) fabrication already exceed 1,000 in vivo (Alivisatos et al. 2013a; Schwarz et al. 2014), and 10,000 in vitro (Spira and Hai 2013). While data analysis for spike trains has been a focus of technique development for over a decade (Brown et al. 2004), the

unprecedented torrent of data from large-scale MEAs now requires a special focus on computational efficiency to handle spike acquisition, interpretation, and storage at massive scale. Translation of large-channel-count MEAs to clinical practice will impose even further demands on efficiency (Fig. 1).

This paper focuses on efficiency in spike acquisition, the first step in any data processing chain. Nyquist sampling is the pervasive approach, where amplified electrode potentials are digitized at sampling rates up to 30 kHz. This approach will be challenging to scale. For a 16-bit analog-to-digital converter (ADC), data rates exceed 3 GB/min for a 1,000 channel array. In contrast, analog thresholding (AT) was developed as an efficient alternative to Nyquist sampling, historically by constraints on commercial ADCs in early MEAs (Meister et al. 1994), and more recently by the need for wireless MEAs (Harrison 2008). In AT, the amplified electrode potential is first passed through a comparator with a threshold optimized based on spike amplitudes. The output of the comparator, an analog square wave, is held temporarily with an analog latch, and then sampled at 1 kHz with a 1-bit ADC, resulting in data rates of 7 MB/min for the same 1,000-channel array.

Efficient spike acquisition is also critical to enable the translation of large-scale MEAs into medical devices. Medical power budgets are estimated to be as low as 10  $\mu$ W for a 10-yr operating period (Chen et al. 2012), possibly closer to 230  $\mu$ W based on our estimates and the Medtronic model 37601 Activa PC neurostimulator battery (Fig. 1). A neural signal acquisition method that could significantly decrease sampling rate would also make substantial advances toward the clinical power constraint, because lower sampling rates would decrease power consumption overall, including the ADC, transmitter, and intervening circuits. By reducing the maximum-needed clock speed, all digital circuits driven by the clock as well as the clock itself would save power. Entire circuits could shut off between samples to save energy, using ultraefficient sleep modes collectively described as deep sleep technology. These basic techniques are described in standard pedagogy on low-power design (Rabaey 2009). Researchers recently reported nano-Watt efficiency for an ambient-light-powered intraocular pressure monitor, an impressive demonstration of deep sleep and slow clocks enabled by low sampling rates (Ghaed et al. 2013).

Address for reprint requests and other correspondence: L. Srinivasan, Neural Signal Processing Laboratory, Stanford Univ. School of Medicine, 300 Pasteur Dr., Stanford, CA 94305-5105 (e-mail: ls2@nsplab.org).

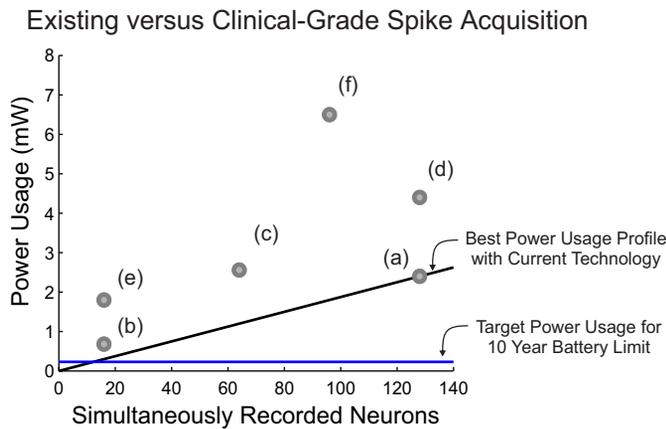


Fig. 1. Power usage for simultaneously recording neurons. The black line represents the best-case existing scenario of power usage vs. number of neurons (0.019 mW/neuron) based on performance achieved in Kusuma et al. (2003). The blue line indicates the maximum total power budget (0.23 mW) available to support 10 yr of uninterrupted operation using the Medtronic model 37601 Activa PC neurostimulator battery. Using current technology, ~12 neurons can be recorded for 10 yr of continuous operation. The data points on this graph correspond to the following references: a, Kusuma et al. 2003; b, Yoo et al. 2012; c, Hegde et al. 2009; d, Charbiwala et al. 2011; e, Bulach et al. 2012; f, Schmale et al. 2013.

While AT represents an order-of-magnitude improvement in sampling rates vs. Nyquist, there are two major barriers that prevent AT from another order-of-magnitude victory. First, temporal resolution in AT is directly proportional to the time between samples (the sampling period). This means that, if we try to increase the sampling period (i.e., decrease the sampling rate), error in spike time grows proportionately. This is true for both event-driven (Meister et al. 1994) and polling (Harrison 2008) implementations of spike-time estimation in AT. Second, more than one spike within a sampling period is discarded. As sampling rates drop, it becomes more likely that more than one spike will occur in the interval, so ignoring extra spikes is less tenable. Dense electrode arrays make matters worse for AT, because they result in spiking from more than one neuron on a single electrode, so refractory periods of individual neurons are no guarantee about the spacing between spikes on a single electrode. Parenthetically, it would be helpful with dense arrays if AT could preserve some descriptor of spike morphology for spike sorting (Gray et al. 1995) or cell type identification (Alivisatos et al. 2013a); the new methodology presented in this paper could possibly be extended to address spike sorting, but that extension requires substantial development beyond the scope of this paper.

Methods other than Nyquist sampling and AT exist, including wavelet compression (CW) (Narasimhan et al. 2007; Oweiss 2006) and compressive sensing (CS) (Bulach et al. 2012; Charbiwala et al. 2011; Hegde et al. 2009; Schmale et al. 2013; Suo et al. 2013), but each of these methods does not simultaneously address the needs for low sampling rates and real-time reconstruction. Both CW and CS require Nyquist-rate sampling as the first step in existing implementations for neural data acquisition. CW additionally requires on-board circuits running at Nyquist clock rates to perform CW (Suo et al. 2013). CS additionally requires an iterative reconstruction procedure to recover spike time and shape, which may not be feasible with dense MEAs that support real-time applications or mission-critical fully-implanted medical devices (Bulach et

al. 2012; Charbiwala et al. 2011; Hegde et al. 2009; Schmale et al. 2013; Suo et al. 2013).

In this paper, we introduce generalized AT (gAT) to achieve the sub-kilo-Hertz sampling rates that eluded AT. Our method incorporates the basic constraint that action potentials are sequences of variable-shape pulses embedded in hash (non-Gaussian spikelike noise), rather than arbitrary, noiseless band-limited signals (Nyquist sampling) or featureless events with timestamps (AT). Where AT uses a comparator followed by a latch, gAT follows the comparator with  $n$ th order integrals. We derive the equations needed to recover spike times from this new sampling method. We compare Nyquist, AT, and gAT using spiking neural data (Fig. 2), previously recorded from motor cortex in two monkeys during center-out reaching movements with a manipulandum (Richardson et al. 2012). Performance is calculated in terms of spike time accuracy and receiver operating characteristics. We then apply these three methods to compare their accuracy in describing spike-local field potential (LFP) phase locking and velocity-related tuning curves, representing applications with varying demands on temporal precision.

**METHODS**

*gAT.* gAT is an extension of AT. In the first step, gAT and AT are nearly identical (Fig. 3). The amplified voltage measured at the electrode passes through a comparator. The output, a square wave, goes high when the electrode voltage exceeds a threshold set by the comparator. In AT, the output stays high until the next sampling period (latched comparator). In gAT, the output returns low immediately when the spike electrode potential falls below threshold (comparator without latching).

In the second step, AT digitizes the latched comparator output and resets the output to zero at the sampling rate. In gAT, a bank of integrators first computes the  $n$ th order integrals of the square wave. Only then are the integrator outputs digitized and returned to zero at the sampling rate. The time of integration is equal to the sampling period, which is equal to one divided by the sampling rate.

In the third step, AT records the spike time using digital logic and a timer. In gAT, a reconstruction procedure calculates spike times from gAT samples. If only one spike typically occurs per interval of duration  $T$ , the spike time  $t_1$  is given in terms of  $y_1$  and  $y_2$ , the first and second integrals of the comparator output:

$$t_1 = T - \frac{y_2}{y_1} \tag{1}$$

The width of the spike measured at the comparator threshold ( $w_1$ ) is given by the first integral of the comparator output:

$$w_1 = y_1 \tag{2}$$

We call this base case gAT-1 to signify that one spike is reconstructed in the interval. Despite the simplicity of these equations, gAT-1 is powerful. As we show in subsequent sections, storing the first and second integrals of the comparator output allows gAT to achieve spike time resolution with one order of magnitude greater efficiency than AT.

In general, the gAT procedure for  $n$  spikes in the interval is denoted gAT- $n$ . When more than one spike occurs in the interval, reconstructing spikes from gAT samples involves the Cauchy formula for repeated integration. We were able to find an exact solution for gAT-2. Higher-order gAT solutions may require iterative procedures. The APPENDIX describes the analytic solutions to gAT-1 and gAT-2 reconstruction, as well as higher-order approaches.

Surgical Site and Sample Data

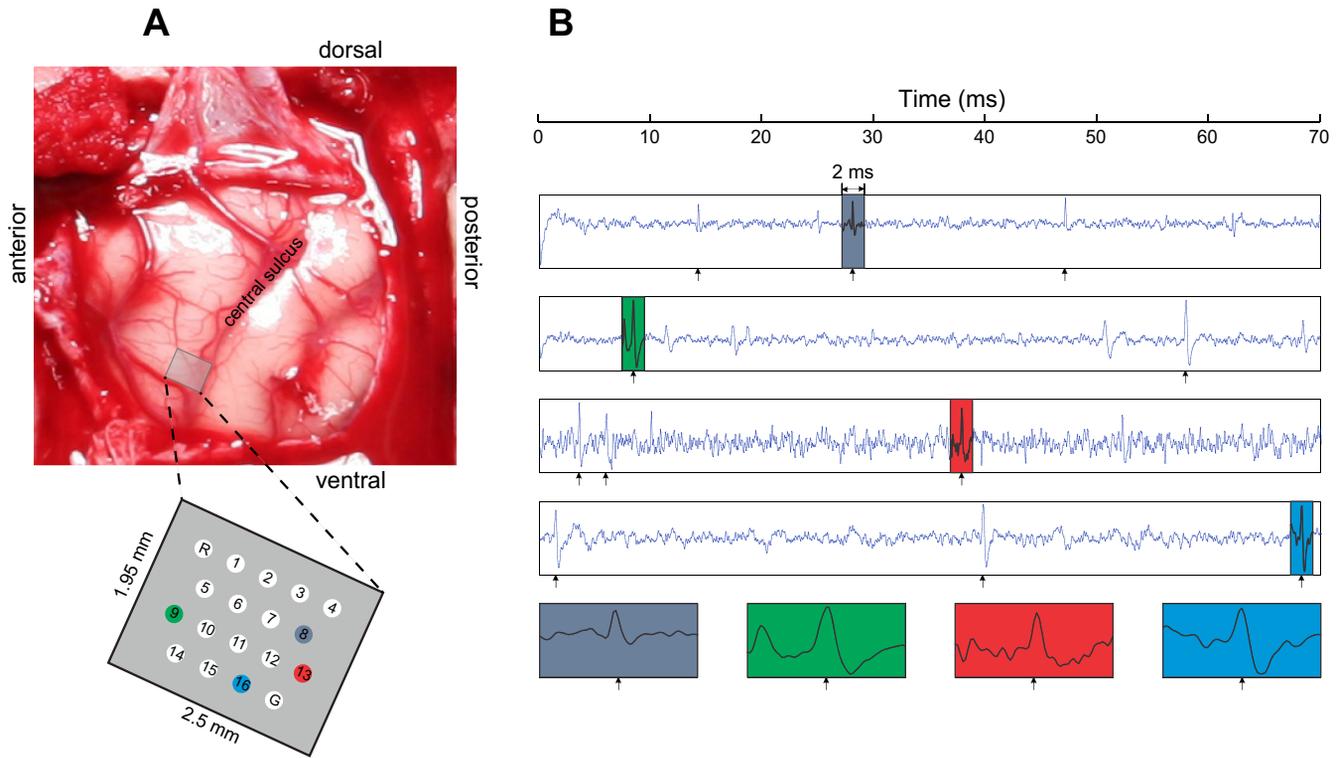


Fig. 2. Electrode positions and sample neuron spikes. *A*: a 16-electrode “floating” microelectrode array and its implantation location in the brain. *B*: example voltage traces on four electrodes. Ground-truth spike times detected from the full 20 kHz recording are indicated by arrows.

*Behavioral task and neural recording.* In each of two *Macaca mulatta*, ensemble spiking activity was previously recorded from primary motor cortex for a separately published study of center-out reaching movements while grasping a two-link robotic manipulandum (Richardson et al. 2012). All animal procedures adhered to National Institutes of Health guidelines on the use of animals and were approved by the MIT Committee for Animal Care. The two-dimensional workspace included 8 possible targets, uniformly spaced on a 10-cm radius circle (22.5°, 67.5°, 112.5°, etc.). Movements were sampled from the robotic manipulandum at 2.5-ms intervals.

A commercial “floating” 16-channel MEA (FMA, MicroProbes for Life Science, Gaithersburg, MD) was used to obtain stable multiday recordings from 10 neurons (28 min, 250,000 total spikes from 10 channels) in one subject and 7 neurons (8.4 min, 18,000 total spikes from 7 channels) in a second subject. Electrode voltages were recorded from each channel at 20 kHz and band-pass filtered at 100–6,000 Hz without downsampling. To determine Nyquist-based spike times, action potentials were detected from the full 20-kHz trace and sorted into single units using threshold crossing and the KlustaKwik technique (Harris et al. 2000). The AT and gAT samples were computed at rates between 10 Hz and 1 kHz derived from the 20-kHz waveforms.

*Quantifying temporal resolution.* Our first performance comparison examined how accurately the various spike acquisition methods determine spike time (Fig. 4). Specifically, we defined spike time error as  $|t_{\text{true}} - t_{\text{reconstructed}}|$ , where  $t_{\text{true}}$  and  $t_{\text{reconstructed}}$  are the true and reconstructed spike times, respectively, in a given block. For each choice of sampling period ( $T$ ), we computed the spike time error by averaging the unsigned error in spike time over all one-spike blocks. The spike time error metric necessarily restricts analysis to blocks where the numbers of reconstructed and true spikes are equal because spike time error is otherwise undefined.

The difficulty of spike acquisition increases with sampling period, in part because a growing fraction of intervals contains more than one spike (Fig. 4A). Larger sampling intervals also make it harder for AT to know when the spike actually occurred. The theoretical spike time error limit for AT (Fig. 4B), is the black line  $y = T/4$  (where  $y$  is the y-axis value and  $T$  is the x-axis value), which was derived as follows. For the block  $[0, T]$ , note that the AT reconstruction spike time is always  $t_{\text{reconstructed}} = T/2$ . Assume that the true spike time  $t_{\text{true}}$  follows a uniform distribution on  $[0, T]$ . The expected value of  $t_{\text{true}} - t_{\text{reconstructed}}$  is  $T/4$ . For large values of  $T$ , the AT curve drops slightly below the theoretical AT curve. To explain this, recall that average spike time error can only be computed using blocks with exactly one true spike and one reconstructed spike. When  $T$  becomes large, the probability of having multiple true spikes in a block becomes significant (Fig. 4A). As a result, given that a block is valid, the true spike is likely to be near the center of the block. To understand this intuition, consider, for example, the extreme case where true spikes are spaced apart by exactly  $T/2 + \epsilon$ , for some small  $\epsilon > 0$ . In this scenario, the only way there can be a valid block is if it is centered on a real spike.

In Fig. 4C, we quantified the fraction of valid sampling intervals, blocks which were eligible for inclusion in the spike time error analysis (Fig. 4B), because spike time error was defined. We define a sampling interval to be active when it contains one or more true spikes. We define a sampling interval to be valid when it is active and the number of reconstructed spikes equals the number of true spikes. We defined the valid fraction as follows:

$$\text{Valid Fraction} = \frac{\text{Total \# Valid Sampling Intervals}}{\text{Total \# Active Sampling Intervals}} \quad (3)$$

The total number of active sampling intervals is the maximum number of blocks that could be used in a spike time error analysis (Fig. 4B). The total number of valid sampling intervals is the number of blocks that were actually included in the analysis. The ratio of these two

Existing vs. Generalized Analog Thresholding (AT)

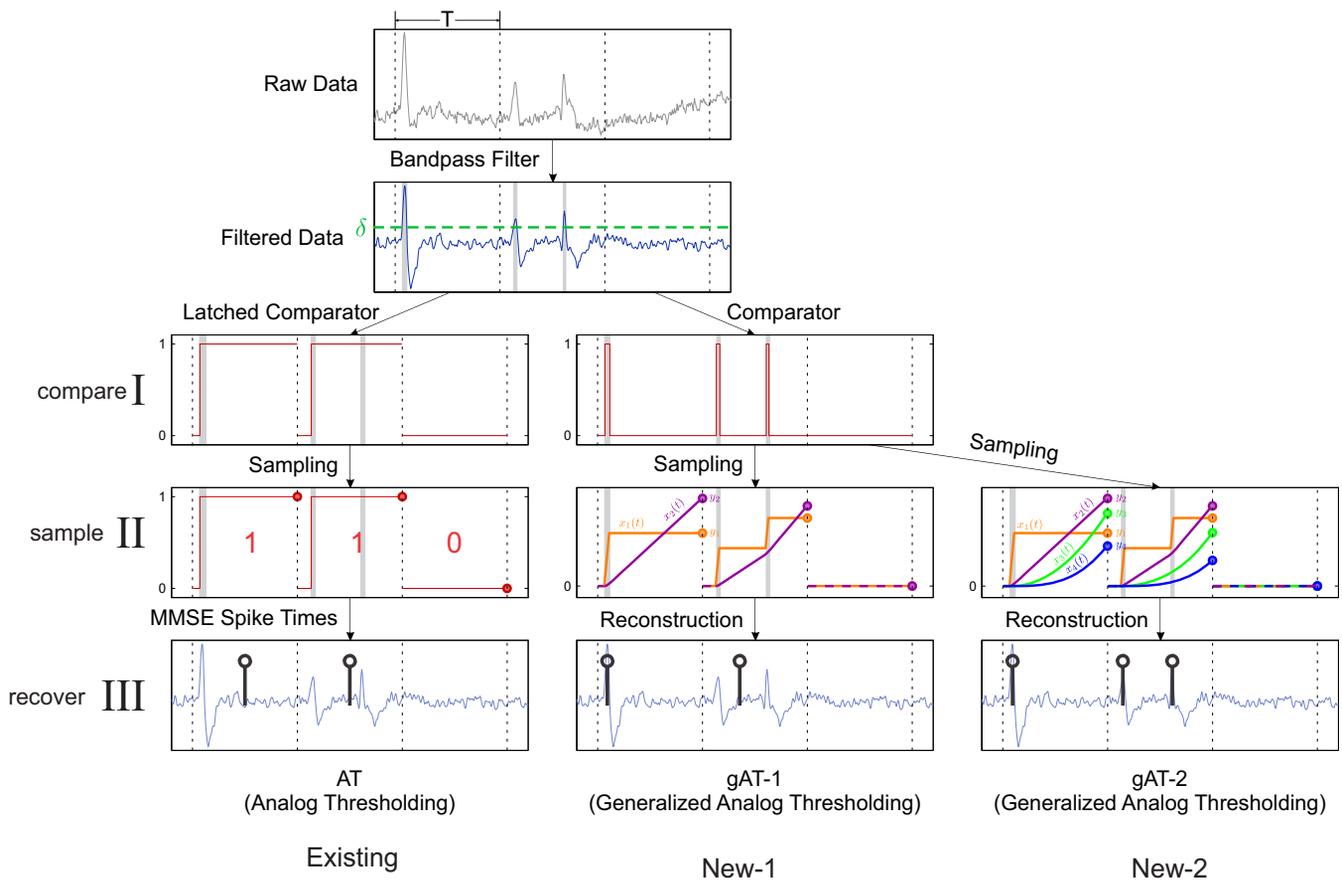


Fig. 3. Analog thresholding (AT) and generalized AT (gAT-1 and gAT-2). Both methods compare the recorded voltage against a threshold. AT uses a latched comparator, which prevents it from detecting multiple spikes per sampling interval or knowing the precise spike time within the sampling interval. In contrast, gAT combines a regular comparator with  $n$ th-order integration. This variation allows gAT to count multiple spikes within a sampling interval and reconstruct spike times to within a few milliseconds of error, despite sampling rates as low as 10 Hz. MMSE, minimum mean squared error.

quantities, the fraction of valid blocks (Fig. 4C), represents the fraction of total data used for each method in the spike time error analysis (Fig. 4B).

**Spike-LFP phase-tuning curves.** The relationship between spike timing and the beta rhythm of motor cortex was estimated in the following manner. First, spike waveforms were removed from the wideband neuronal signal by subtracting the mean spike waveform scaled by a factor that minimized the sum of squared differences between the mean and a  $[-0.5, 2]$  ms window around each spike peak (Zanos et al. 2012). Second, the signal was band-pass filtered in the  $\beta$ -band (20–30 Hz) using a noncausal, zero-phase filter. Third, the Hilbert transform determined instantaneous amplitude and phase of the band-passed signal. Fourth, the phase values at the previously-identified spike times were collected to form a histogram of spike count for 16 phase bins spanning  $0^\circ$  to  $360^\circ$ .

**Velocity-dependent tuning curves.** Velocity-dependent tuning curves were estimated separately for each neuron based on detailed statistical development and validation described previously (Truccolo et al. 2005). This approach uses an inhomogeneous Poisson point process model for spiking with the following conditional intensity function:

$$\lambda(k|v_x, v_y) = \exp[\beta_0 + \beta_1(v_x^2 + v_y^2)^{1/2} \cos(\theta - \theta_p)] \quad (4)$$

Here,  $k$  represents a 2.5-ms time step,  $v_x$  and  $v_y$  represent  $x$  and  $y$  hand speed in the two-dimensional workspace, and  $\theta$  is the direction of hand movement. Under this model, the velocity-dependent tuning

curve for a neuron is characterized by the baseline firing rate parameter  $\beta_0$ , speed modulation parameter  $\beta_1$ , and preferred direction parameter  $\theta_p$ . Parameter fitting for this model is based on maximum likelihood estimation within the generalized linear model framework, fully implemented by the Matlab Statistics Toolbox function `glmfit`.

**Integrator noise and quantization.** The feasibility of hardware implementation was examined by quantifying spike time and width reconstruction error (see Fig. 8A) in the face of integration noise and quantization (see Fig. 8B). Integration noise levels were chosen based on a commercial integrator, the Texas Instruments Precision Switched Integrator Transimpedance Amplifier (IVC102). The IVC102 root mean squared output noise ranges from  $4 \mu\text{V}$  to  $300 \mu\text{V}$  for 1-ms integration time. Quantization error was based on ADCs with 1- to 16-bit resolution, and quantization levels evenly distributed throughout the input voltage range. Performance was examined using the *subject 1* spike times, over the full range of integrator noise and quantization, for gAT-1 and gAT-2 (see Fig. 8, C–H). This analysis on real spiking neural data draws from a recent hardware simulation analysis for square wave reconstruction (He et al. 2015).

RESULTS

We compared AT, gAT-1, and gAT-2 with various measures of performance on neural data recorded from two monkeys using 16-channel multiunit electrode arrays implanted in primary motor cortex (Fig. 2). Results from the first monkey (28

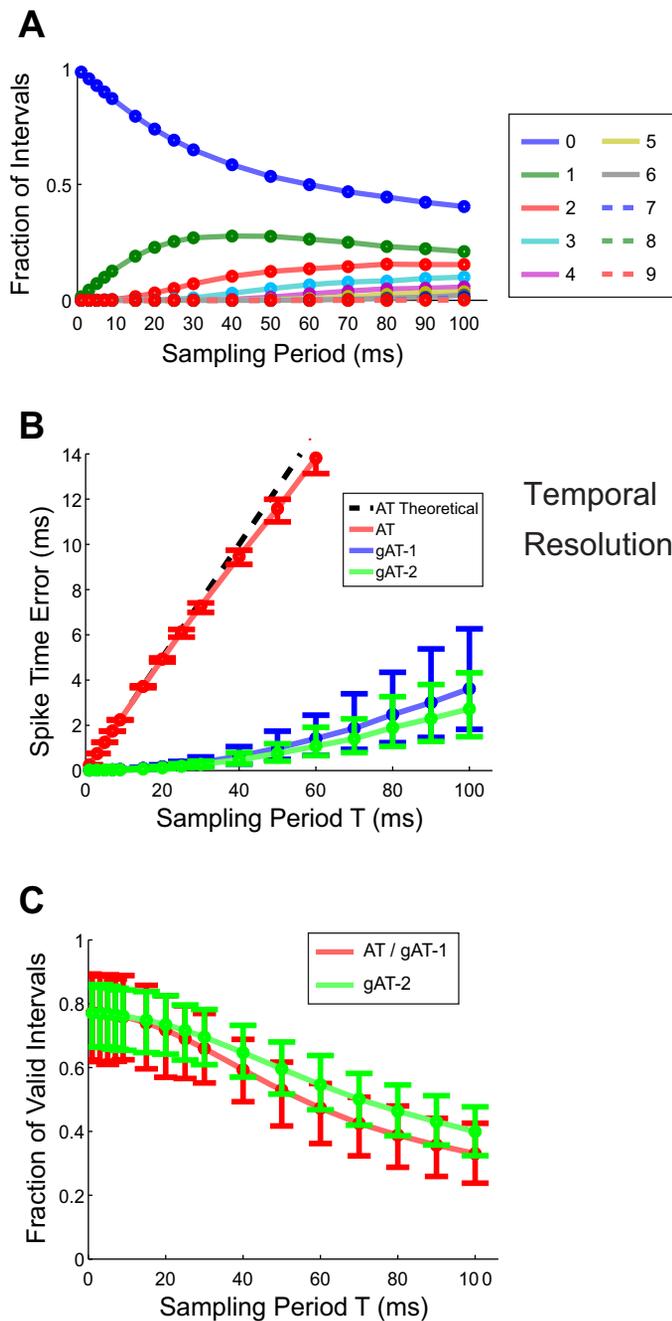


Fig. 4. Temporal resolution at various sampling periods. *A*: the fraction of intervals with a certain number of spikes within the interval. As the sampling period increases, intervals tend to have larger numbers of spikes. *B*: average error in estimated spike time in intervals with exactly one correctly identified spike. AT always predicts that the spike is at the center of the interval, and its behavior closely matches the theoretical curve. gAT-1 and gAT-2 have much lower errors than AT. *C*: the fraction of intervals that correctly identify the number of spikes. AT and gAT-1 always predict the same number of spikes because they are only able to differentiate between intervals with no spikes and intervals with a nonzero number of spikes. gAT-2 is able to differentiate between intervals with one spike and intervals with two spikes, and gAT-2 is able to improve the accuracy of spike count predictions.

min, 250,000 total spikes from 10 channels) are presented in the main text of this paper. Results from the second monkey (8.4 min, 18,000 total spikes from 7 channels) are similar, documented in the APPENDIX. Ground-truth spike times are defined by Nyquist sampling followed by spike detection as

described in METHODS. We compare against this ground truth to compute performance errors under various metrics. As such, these graphs do not display a separate performance trend for Nyquist sampling.

In Fig. 4, we characterized temporal resolution of AT vs. gAT. We first quantified how the difficulty of spike acquisition varies with sampling period for this data set (Fig. 4A). For these primary motor cortical neurons recorded during movement, nearly one-half of all sampling intervals (blocks) contain just one spike, even up to sampling periods of 100 ms, equivalent to sampling rates of 10 Hz. This scenario is favorable to the endeavor of sub-kilo-Hertz spike acquisition, which becomes increasingly difficult with more spikes per interval. We then examined spike time error, showing that gAT is a super-resolution method, meaning that its performance significantly and substantially exceeds the theoretical limit of AT for a given sampling rate (Fig. 4B).

In Fig. 4C, we quantified the fraction of blocks which were eligible for inclusion in the spike time error analysis because spike time error was defined. Both gAT-1 and AT share the same valid fraction (red), because both use the comparator step, both detect at most one spike per interval, and postcomparator processing by gAT-1 never misses spikes that pass the comparator threshold. By comparison, gAT-2 achieves a higher valid fraction because it can also handle sampling periods with two spikes. If only a few blocks were valid, the spike time error characterization (Fig. 4B) would have been misleading. Instead, the fraction of valid sampling periods (Fig. 4C) is generally more than one-half the data for all methods. However, there are still a significant number of blocks that needed to be thrown out because spike time error was not defined for those intervals. Next we consider a different error metric that takes all sampling periods into account to assess overall performance.

In the analysis detailed in Fig. 5, instead of comparing true and reconstructed spikes on a period-by-period basis, we compared a 28-min-long spike train of true spikes with the corresponding spike train of reconstructed spikes. One issue that arises in this setting is the possibility that a true spike intersects the boundary between two sampling periods, triggering threshold crossings in both adjacent sampling periods and creating two reconstructed spikes. To solve this problem, we chose to enforce a mandatory refractory period of 1.1 ms in the reconstructed spike train. Any reconstructed spike whose spike time exceeds that of the previous reconstructed spike by less than the refractory period is expurgated.

After imposing a refractory period, we compared the two spike trains by computing the number of false positives and false negatives. We chose an error tolerance value  $\tau = 5$  ms and paired reconstructed spikes with true spikes that were no more than  $\tau$  ms away. The true spikes that could not be paired up in this manner were considered false negatives, and the reconstructed spikes that could not be paired up were considered false positives. We plotted the proportions of false negatives and false positives relative to the total number of true spikes (Fig. 5). Given a true and reconstructed spike train, there is a tradeoff between false positives and false negatives that is controlled by the spike detection threshold  $\delta$ . Increasing  $\delta$  lowers the number of false positives but increases the number of false negatives. In Fig. 5, we see that gAT gives a more favorable tradeoff than AT. In particular, gAT-1 and gAT-2

ROC Performance

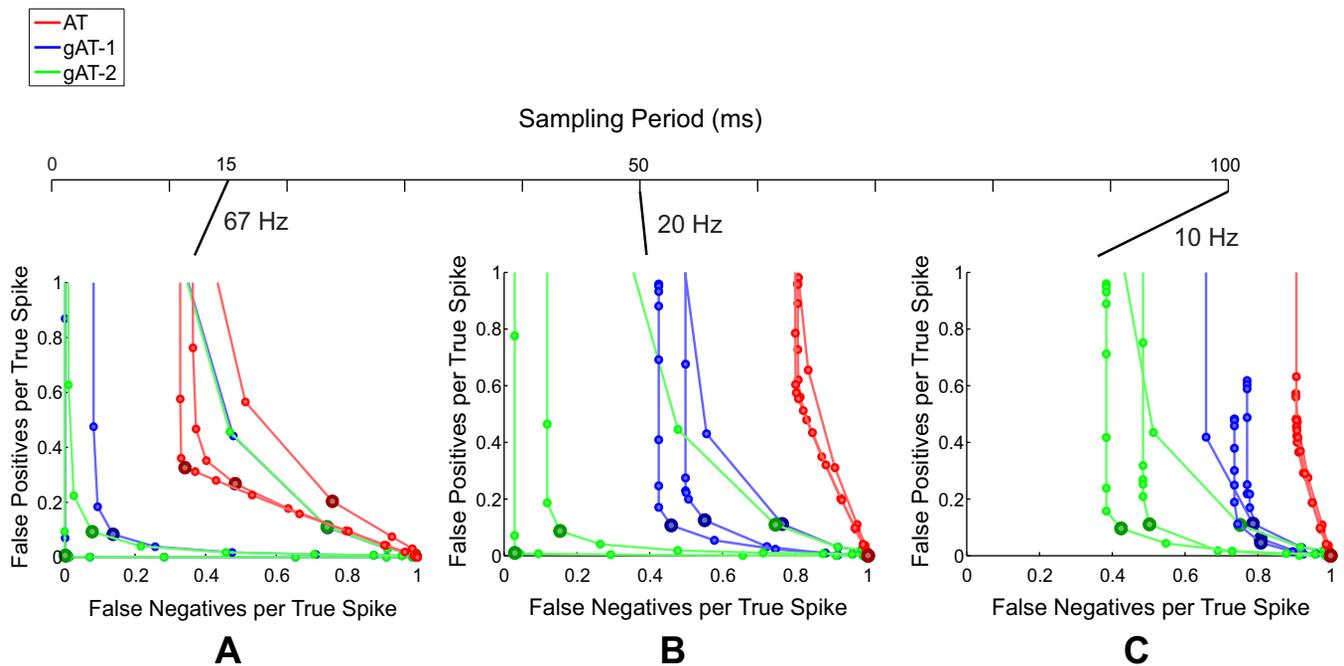


Fig. 5. Spike detection receiver operating characteristics (ROC) at various sampling periods: 67 Hz (A), 20 Hz (B), 10 Hz (C). Three representative curves (worst, median, and best) are shown for each method. False positives are defined as predicted spikes without a corresponding real spike, and false negatives are defined as real spikes without a corresponding predicted spike. An error tolerance of  $\tau = 5$  ms is allowed. Each point corresponds to a different spike detection threshold  $\delta$ . Increasing  $\delta$  lowers the number of false positives but increases the number of false negatives. The large data point on each curve shows the optimal  $\delta$ , which is the point with the fewest possible total errors. Here, gAT-2 has the best performance, and AT has the worst performance. As the sampling period increases, all three methods perform worse.

offer lower error values than AT for longer sampling periods. The three methods have more similar performance when the sampling period is short (high sampling rates), as should be expected.

Next we compared AT and gAT in practical application to data analysis that requires spike acquisition. First we examined the relative abilities of AT and gAT to adequately support spike-LFP phase locking analysis. For some neurons, spikes are most likely to occur during a particular phase of LFP oscillation (Zanos et al. 2012). Given spike times and LFP data, we can perform the following analysis to detect this phase-locking behavior. For each spike, we calculate the phase of the 20- to 30-Hz oscillation in the LFP at the instant in time at which the spike occurred, as described in METHODS. We then plot the distribution of these phase values and calculate the preferred phase. Figure 6A shows an example of this plot for a particular neuron using four methods to acquire spike times: Nyquist (marked True), AT, gAT-1, and gAT-2. Since phase locking is strongest at high-amplitude LFP, spikes that occurred when the LFP amplitude was below some threshold were not included in estimating the phase distribution. The LFP threshold was chosen to be the median of the LFP amplitude recorded at the true spike times, so that approximately one-half of the recordings were not included in the LFP phase analysis.

We then considered average performance in the spike-LFP analysis across all neurons in *subject 1* to determine how performance changes as a function of the sampling period  $T$  using each of the methods, AT, gAT-1, and gAT-2 (Fig. 6,

B–E). Error in preferred phase is defined as the absolute difference between the preferred phase of the true spikes and the preferred phase of the reconstructed spikes (Fig. 6B). Error in phase distribution is defined as the average (over all phase values) of the absolute difference between the phase distribution calculated using the true spikes and the phase distribution calculated using the reconstructed spikes (Fig. 6C). With gAT, the error in preferred phase remains flat as  $T$  increases, whereas with AT, this error increases with  $T$  (Fig. 6D). Similarly, with phase distribution, we see smaller error with gAT than with AT, although error remains essentially flat with respect to  $T$  under this metric (Fig. 6E). As a result of the analysis shown in Fig. 6, we can conclude that gAT outperforms AT in fidelity of phase-locking analysis.

We now investigate an application in which gAT does not provide a significant advantage over AT, namely estimation of preferred direction in velocity-dependent motor cortical tuning curves (Fig. 7). In estimating preferred direction, both AT and gAT provide nearly equivalent performance, in contrast to the spike-LFP application. The reason for this discrepancy relates to differences in time scale between the relevant quantities. In contrast to the phase of 20- to 30-Hz LFP oscillations use in spike-LFP analysis, arm velocity typically changes much more gradually, on the time scale of tens to hundreds of milliseconds, because arm movements are generally smooth. As a result, precise spike time may be critical for determining preferred phase in the spike-LFP analysis, but unnecessary for determining preferred direction in velocity-dependent motor cortical neurons.

Spike-LFP Phase Accuracy

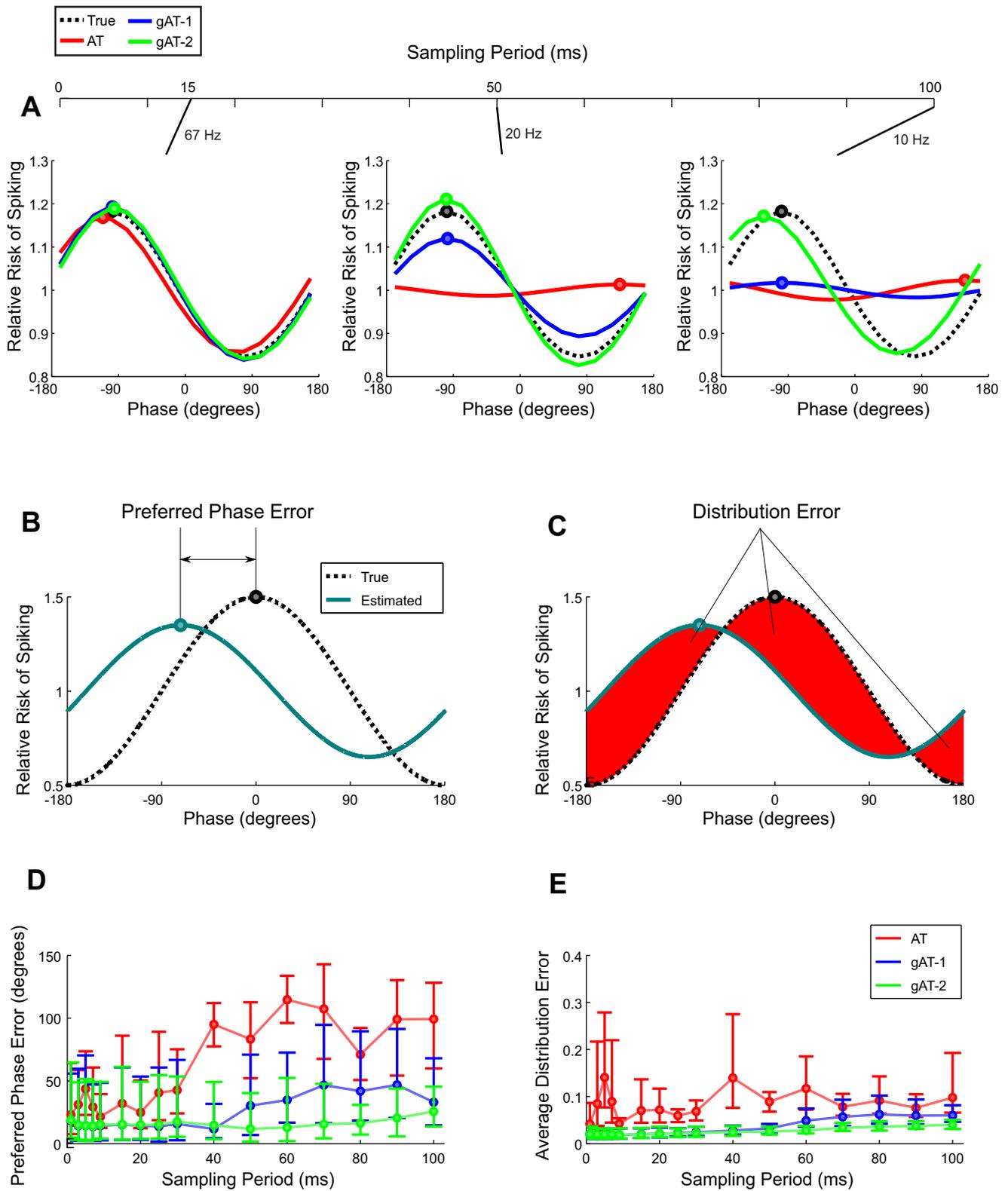


Fig. 6. Robustness of spike-local field potential (LFP) phase-locking estimation to increasing sampling period. *A*: single-neuron illustration of preferred phase estimates (filled circles) at three different sampling periods. *B*: the preferred phase error metric quantifies the difference between the maxima of the two spike-phase curves. *C*: the distribution error metric quantifies the area between true and predicted spike-phase curves. *D*: preferred phase error vs. sampling period, averaged across all 10 neurons from *subject 1*. *E*: average distribution error vs. sampling period for the same data.

Preferred Direction Accuracy

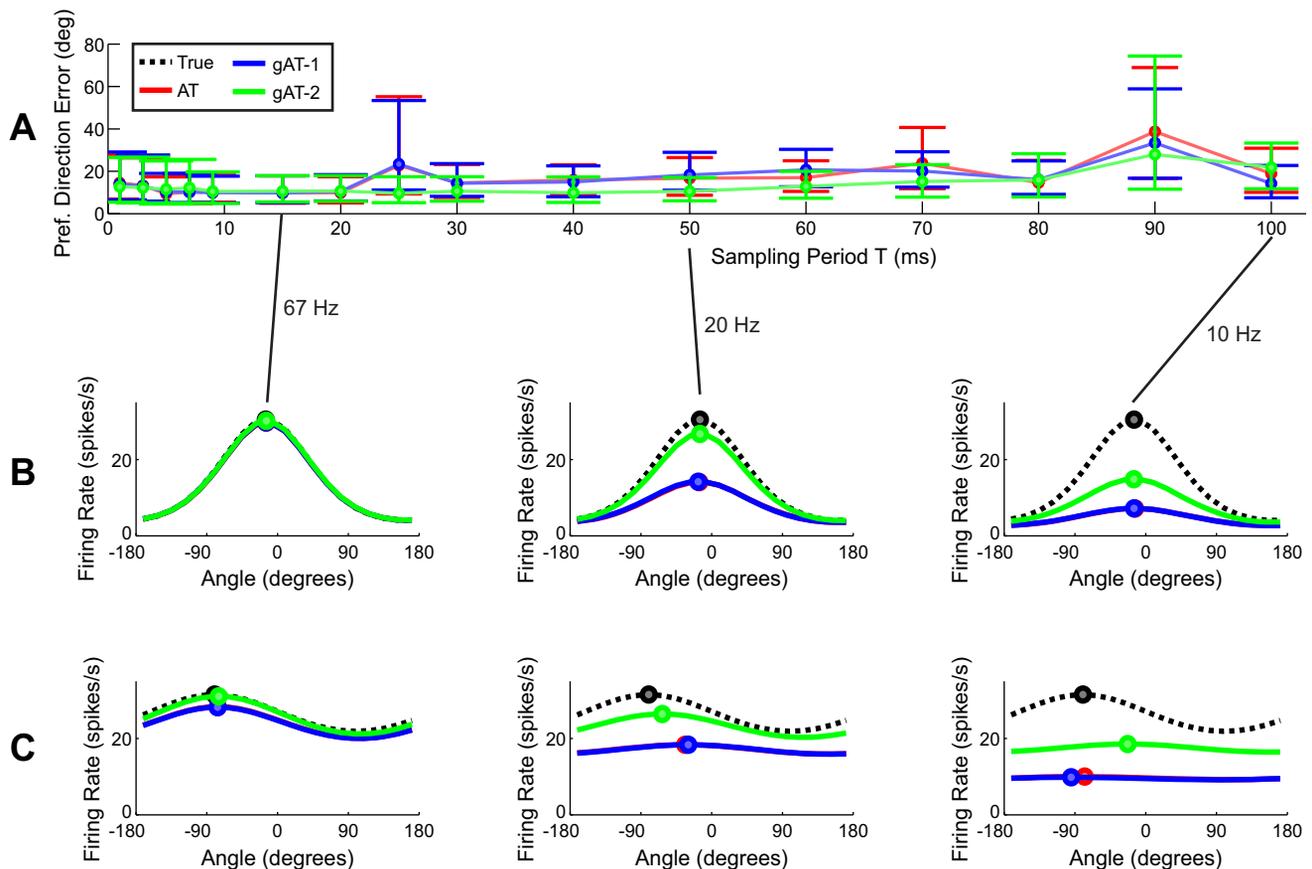


Fig. 7. Robustness of motor cortical preferred direction estimation to increasing sampling period. *A*: error in preferred direction vs. sampling period remains low even for AT. The error for gAT-2 consistently trends to the lowest of the three methods, although confidence intervals heavily overlap for individual sampling periods. *B*: in a single-neuron example, we see that the estimated depth of tuning begins to decline with increasing firing rate, even though the estimated preferred direction remains accurate. *C*: in another neuron with lower depth of modulation and higher baseline firing rate, the estimated preferred direction erodes more significantly. In this neuron, longer sampling periods are more severely inadequate in capturing the difference in spike rates between baseline and maximally modulated directions, because even baseline firing rates are high. We call this performance limitation the saturation effect.

In the preferred direction analysis, we also see a performance characteristic that we call saturation effect (Fig. 7, *B* and *C*). Both AT and gAT variants detect a maximum number of spikes per block. This determines the maximum detectable firing rate. We see from the two different tuning curve examples that tuning curve estimates can be particularly distorted when both the minimum and maximum firing rates approach or exceed the maximum detectable rate, which we call the saturation effect. As long as the minimum firing rate is far from the maximum detectable rate, preferred direction can be estimated accurately.

Finally, we examine the plausibility of hardware implementation. To simulate the major nonidealities of hardware, we corrupt the gAT samples with integrator noise and quantization error based on specifications from commercially available integrated circuits (see METHODS). Reconstruction quality is determined based on ability to determine spike time and width, defined relative to the thresholded experimentally-recorded waveforms (Fig. 8*A*). We simulate effect from a serial integration architecture (Fig. 8*B*). The analysis shows that errors in reconstructed spike time and width deteriorate quickly for gAT-2, but remain robust for gAT-1 across a wide range of realistic noise and quantization characteristics (Fig. 8, *C–H*). For example, gAT-1 achieves millisecond spike time and width

resolution at 10-Hz sampling rates with two gAT samples per interval and 16-bit ADC resolution (320 bits/s). For comparable performance with one electrode, AT requires 1 kHz at 1-bit resolution (1,000 bits/s), and Nyquist sampling requires up to 30 kHz at greater than 8-bit resolution (at least 240,000 bits/s). Both as a path to low sampling rate and low data rate, gAT-1 substantially outperforms these alternatives.

DISCUSSION

Efficient spike acquisition will bridge the divide from large MEA to whole-cortex electrophysiology. In this paper, we present a method for efficient spike acquisition called gAT. This method converts AT into a super-resolution technique for efficient spike acquisition. We show that gAT achieves sampling rates as low as 10 Hz, where equivalent performance would require 1 kHz with AT or 5–30 kHz with Nyquist sampling. We also show that the gAT-1 method is already robust to hardware nonidealities, including integration noise and quantization. Together, these results show that gAT reliably computes timing-sensitive neural statistics at ultralow sampling rates, with a plausible path for translation.

Spike acquisition using gAT offers a number of advantages over existing methods. First, gAT is efficient with sampling.

Effect of Noisy Integration and Quantization

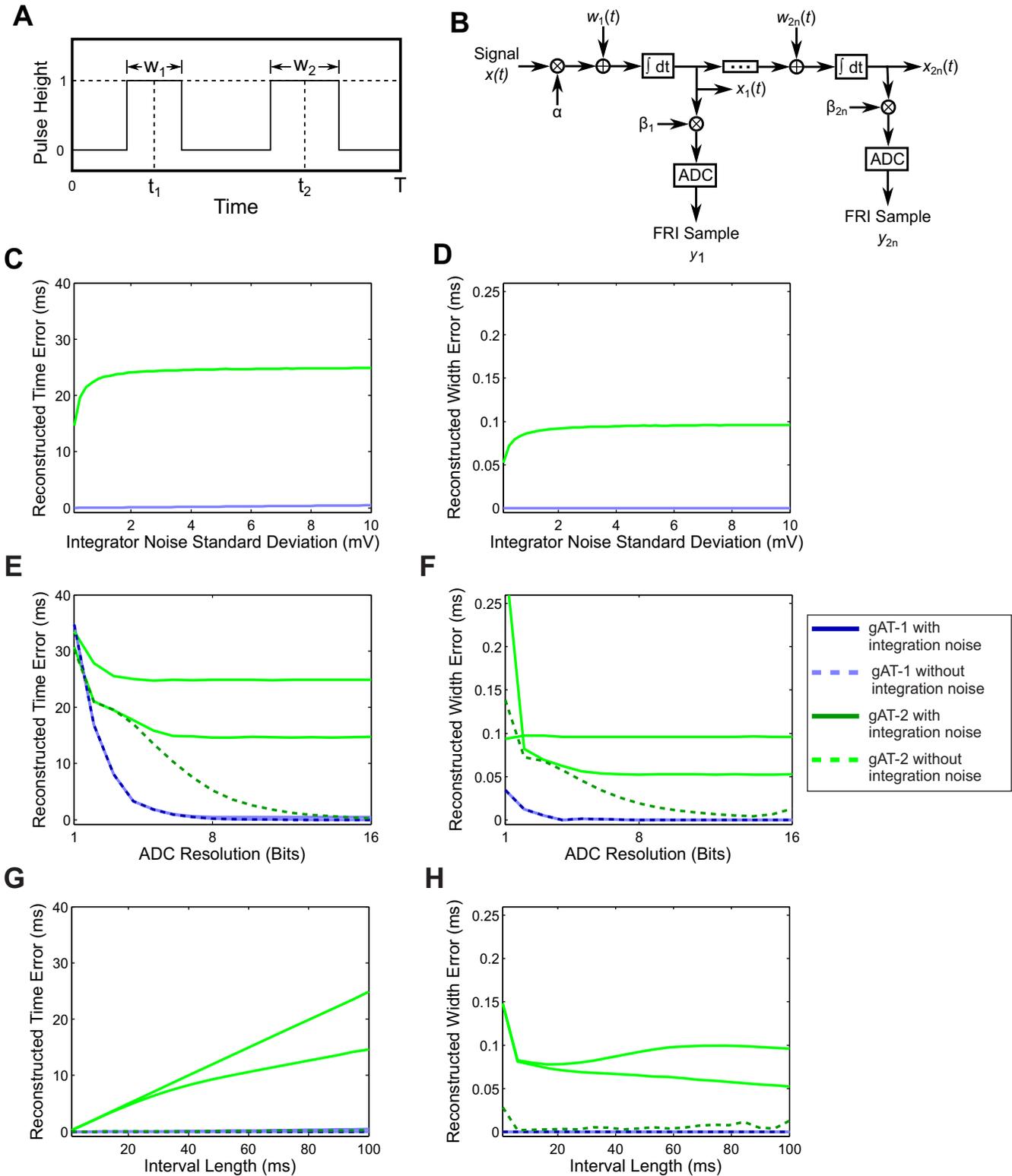


Fig. 8. Performance under hardware nonidealities. *A*: following the comparator step, action potentials in a  $T$ -second analysis interval are defined by width ( $w$ ) and event time ( $t$ ). *B*: major hardware nonidealities include  $w(t)$ , the integrator noise, and quantization performed by the analog-to-digital converter (ADC). Inputs to the ADC are scaled to allow uniform quantization levels over the full range of input signal values. FRI, finite rate of innovation. *C* and *D*: at 10-Hz sampling rates (100-ms sampling interval) and 16-bit ADC, error in reconstructed spike time (*C*) and width (*D*) over a realistic range of integrator noise (defined in terms of standard deviation of the output in mV after integrating zero input for 1 s) are shown. *E* and *F*: at 10-Hz sampling rates, error in reconstructed spike time (*E*) and width (*F*) over a realistic range of quantization (ADC resolution) are shown. *G* and *H*: at a range of sampling rates, error in reconstructed spike time (*G*) and width (*H*) over a realistic range of quantization (ADC resolution) are shown. Results in *E-H* are presented with two levels of integrator noise (solid lines,  $\sigma = 10$  mV and  $\sigma = 0.1$  mV) and without (dashed lines) integrator noise.

This means that application-specific resources (power, bandwidth, memory) can be more effectively utilized. In contrast, existing implementations of spike acquisition based on CW (Narasimhan et al. 2007; Oweiss 2006) or CS (Bulach et al. 2012; Charbiwala et al. 2011; Hegde et al. 2009; Schmale et al. 2013; Suo et al. 2013) require Nyquist sampling as the first step. Consider that a roughly  $1,000\times$  decrease in sampling rate from Nyquist to gAT-1 could be used to sample  $500\times$  more channels with the same bandwidth, given that gAT-1 produces two A/D conversions per sampling interval. Alternatively, the duty cycle (defined as the percentage of time a circuit is active per unit time of operation) could be reduced  $1,000\times$  for the same number of channels. Downstream processing for interpretation and storage becomes faster because the data rate produced per gAT channel is orders of magnitude smaller than Nyquist sampling.

Second, spike reconstruction with gAT is instant. Methods based on CS require iteration to reconstruct the original spike times and shapes over a given sampling interval (Bulach et al. 2012; Charbiwala et al. 2011; Hegde et al. 2009; Schmale et al. 2013; Suo et al. 2013), which is not immediately conducive to real-time applications. In contrast, we have provided simple closed-form solutions for the spike time and width using gAT-1 and gAT-2. This means that recovering the original spike train simply requires entering the gAT samples into the gAT-1 and gAT-2 formulas for spike time and width. As a result, spike reconstruction with gAT can be performed in real time.

Third, gAT provides a path toward elements of spike reconstruction that were prohibited by the classical AT approach. For example, gAT-2 and gAT- $n$  permit reconstruction of more than one spike per interval. Additionally, gAT methods already reconstruct spike width (time spent above threshold), which can be useful in spike sorting with tetrodes and dense MEAs. Based on current trends in theory on signals with finite rate of innovation, further improvements on gAT to reconstruct more detailed parameters of spike shape are on the horizon (Caballero et al. 2012; Kusuma et al. 2003).

Fourth, because of its ultralow sampling rates, gAT enables complementary power-saving opportunities for neural signal acquisition that have been widely used elsewhere (Ghaed et al. 2013; Rabaey 2009). Low sampling rates permit various low-power design strategies, including low clock rates and deep-sleeping entire circuits between samples. Both ADC and transmitter power consumption decrease in proportion to data rates. Where Nyquist sampling requires greater than 10,000 Hz, and AT operates at 1,000 Hz (Harrison 2008; Meister et al. 1994), the gAT method functions below 100 Hz. This affords gAT multiple orders of magnitude improvement in sleep time and data rate.

Fifth, gAT uses hardware-friendly operations like integrators and comparators. These operations are already present within various types of ADCs (Pelgrom 2010). As another example, special-purpose hardware with similar operations has recently been built for efficient sampling of radar pulses (Yoo et al. 2012). As with ADCs in standard Nyquist sampling, precision in these hardware components will be a major determinant of final performance. Nevertheless, existing examples of analog circuits that use these basic operations provide a starting point for the considerable task of translating any algorithm to custom-designed low-power hardware.

The gAT method brings large-scale electrode arrays two to three orders of magnitude closer to the  $10\text{-}\mu\text{W}$  (Chen et al. 2012) or  $230\text{-}\mu\text{W}$  (Fig. 1) total power budget estimates for 10 yr of uninterrupted clinical operation. A recent hardware implementation of CS provides the basis for a first-order projection of power savings at all stages (Chen et al. 2010). The first step, amplification, is normally expensive and does not scale with sampling rates because existing architectures require one amplifier for each electrode. A recent Nyquist-based implementation consumed  $7.6\text{ }\mu\text{W}$  to amplify action potentials from a single electrode (Wattanapanitch et al. 2007). Model-based spike reconstruction allows for noisy amplification that consumes significantly less energy (Chen et al. 2010). The second step, ADC consumption, does scale with sampling rate. These circuits have continued to become more efficient on a per-channel basis relative to the cost of amplification (Murmah 2013). For example, an 8-bit ADC consumes  $0.01\text{ }\mu\text{W}$  for a single channel when sampling at 10 kHz. However, higher bit resolution is still substantially more expensive. A 16-bit ADC consumes  $100\text{ }\mu\text{W}$  for a single channel at the same 10-kHz sampling rate. The gAT-1 method at 10 Hz would consume less than  $1\text{ }\mu\text{W}$  for the 16-bit ADC. Model-based reconstruction has also been shown to use lower bit resolution while still preserving spike timing (Chen et al. 2010). The third step, digital computation, depends on the on-board functions performed, which is greatly simplified in CS and gAT. For example, the CS architecture for signal sampling (but not reconstruction) consumes no more than other data compression schemes (Chen et al. 2010). Power savings from deep sleep are roughly proportional to sampling period, representing  $10\times$  to  $1,000\times$  savings at this step using gAT instead of Nyquist sampling. The final step, wireless transmission, is also a significant power sink that is proportional to data rates. A typical biosignal transmitter (2) consumes  $240\text{ }\mu\text{W}$  for the single channel sampled at 10 kHz with an 8-bit ADC and  $480\text{ }\mu\text{W}$  for the same channel sampled at 10 kHz with a 16-bit ADC. With gAT-1 at 10 Hz, this same transmitter would consume less than  $1\text{ }\mu\text{W}$  with either 8- or 16-bit ADCs.

The performance characterization presented in this paper has limitations that will need to be addressed in future development. First, the extent of power savings with gAT or any other spike acquisition method will depend on specifics related to hardware implementation and application. This makes it difficult to claim in total generality that gAT will offer power savings. Second, we have shown that gAT-1 is robust to realistic hardware nonidealities, but additional work is needed to make gAT-2 perform, despite imperfect integration and quantization error. Third, neural data were limited to motor cortical neurons, where other regimes of spike shape, firing rate, and electrode configuration would reveal different performance profiles. Examples include dense electrode arrays that record duplicate action potentials from multiple neighboring neurons, and neuron types that manifest bursts of spikes or high baseline firing rates. These examples would result in shorter interspike intervals, reducing the sampling rate advantage of gAT over AT. Even in those adverse conditions, gAT would still retain elements of spike shape that are discarded by AT and maintain substantial gains over Nyquist rates. Fourth, sufficiency of performance will ultimately relate to the specific application and target outcomes for which any given spike acquisition method is proposed. As such, gAT should remain a

candidate among a diversity of alternate methods examined for application-specific development. Further validation of gAT could include extensive testing beyond the scope and size limitations of the present paper, such as hippocampal recordings with simultaneous intracellular and extracellular traces (Henze et al. 2000), or many other species, cell types, and electrode configurations.

Apart from electrophysiology, related work on efficient spike acquisition is beginning to appear in the literature on calcium channel imaging. In optical imaging, lower sampling rates translate into larger field of view, similar to functional magnetic resonance imaging. At times, the underlying algorithmic methodology for spike acquisition may be similar between electrophysiology and optical imaging. For example, theory on signals with finite rate of innovation, the basis for our present work, has recently seen application in calcium imaging (Oñativia et al. 2013). Similarly, CS has also been applied to calcium imaging (Pnevmatikakis and Paninski 2013). This is an exciting time for algorithms in neuroscience, where the impending data rate from massive MEAs will require neuroscience methods for acquisition, interpretation, and storage that scale like never before.

APPENDIX

This appendix provides a detailed technical description of gAT. The appendix also reports performance on data from a second subject, demonstrating consistency with trends reported in the main text.

*gAT-1: Derivation and Closed-Form Solution*

First, the simplest variant, gAT-1, is considered. Assume for simplicity that we are dealing with the sampling interval  $[0, T]$ . To analyze this sampling interval, the first and second integrals of the comparator output in the sampling interval are calculated. If the comparator output in the sampling interval is given by the function  $x(t)$  on the domain  $[0, T]$ , the first and second gAT samples  $y_1$  and  $y_2$  are given in terms of the first- and second-order integrals of the comparator output, denoted  $x_1(t)$  and  $x_2(t)$  respectively, and sampled at time  $t = T$ .

$$x_1(t) = \int_0^t x(\tau) d\tau \tag{A1}$$

$$x_2(t) = \int_0^t x_1(\tau) d\tau \tag{A2}$$

$$y_1 = x_1(T) \tag{A3}$$

$$y_2 = x_2(T) \tag{A4}$$

Assume that, in the interval  $[0, T]$  we have either zero or one spike. A spike manifests itself in the comparator output  $x(t)$  as a square-wave centered at  $t_1$  with width  $w_1$  and height 1. We want to detect whether there is a spike, and, if there is one, find its parameters  $t_1$  and  $w_1$ . If there is a spike, the comparator output is given by the signal

$$x(t) = \begin{cases} 1 & \text{if } t_1 - \frac{w_1}{2} \leq t \leq t_1 + \frac{w_1}{2} \\ 0 & \text{otherwise} \end{cases} \tag{A5}$$

Assume the spike is contained completely within the interval, i.e.,  $0 \leq t_1 - w_1/2$  and  $t_1 + w_1/2 \leq T$ . Take the first and second integrals  $x_1(t)$  and  $x_2(t)$  of this signal.

$$x_1(t) \equiv \int_0^t x(\tau) d\tau = \begin{cases} 0 & \text{if } t < t_1 - \frac{w_1}{2} \\ t - \left(t_1 - \frac{w_1}{2}\right) & \text{if } t_1 - \frac{w_1}{2} \leq t \leq t_1 + \frac{w_1}{2} \\ w_1 & \text{if } t > t_1 + \frac{w_1}{2} \end{cases} \tag{A6}$$

$$x_2(t) \equiv \int_0^t x_1(\tau) d\tau = \begin{cases} 0 & \text{if } t < t_1 - \frac{w_1}{2} \\ \frac{1}{2} \left[ t - \left(t_1 - \frac{w_1}{2}\right) \right]^2 & \text{if } t_1 - \frac{w_1}{2} \leq t \leq t_1 + \frac{w_1}{2} \\ w_1(t - t_1) & \text{if } t > t_1 + \frac{w_1}{2} \end{cases} \tag{A7}$$

Sample  $x_1(t)$  and  $x_2(t)$  at  $t = T$ .

$$y_1 \equiv x_1(T) = w_1 \tag{A8}$$

$$y_2 \equiv x_2(T) = w_1(T - t_1) \tag{A9}$$

We can therefore recover the parameters  $t_1$  and  $w_1$  of the spike as follows.

$$w_1 = y_1 \tag{A10}$$

$$t_1 = T - \frac{y_2}{y_1} \tag{A11}$$

Having zero spikes in the interval is equivalent to having one spike with pulse width  $w_1 = 0$ . Therefore, report zero spikes in the interval if  $w_1 = 0$  (i.e.,  $y_1 = 0$ ). Otherwise, report one spike in the interval with parameters  $t_1$  and  $w_1$  as calculated above. Note that this gives gAT superresolution, meaning the ability to pinpoint the time of a spike on scales smaller than the sampling period  $T$ .

*gAT-n*

Next, the general variant, gAT- $n$  is considered, which handles up to  $n$  spikes per sampling interval.

*Derivation of gAT-n.* RELATING GAT SAMPLES TO SPIKE TIMES AND PULSE WIDTHS USING THE CAUCHY FORMULA. Suppose the comparator output in the sampling interval  $[0, T]$  is given by the function  $x(t)$ . Denote the  $k$ th order integral of  $x(t)$  by  $x_k(t)$ . Define the  $k$ th gAT sample  $y_k$  as  $x_k(t)$  evaluated at  $T$ .

$$x_1(t) = \int_0^t x(\tau) d\tau \tag{A12}$$

$$x_{k+1}(t) = \int_0^t x_k(\tau) d\tau \tag{A13}$$

$$y_k = x_k(T) \tag{A14}$$

To reconstruct the spikes that occur in this sampling interval, gAT- $n$  requires us to calculate the first  $k = 2n$  successive integrals of the comparator output in the sampling interval. In other words, the gAT- $n$  procedure requires that we compute  $2n$  gAT samples  $y_1, y_2, \dots, y_{2n}$ .

The general variant gAT- $n$  assumes exactly  $n$  spikes have occurred in the sampling interval  $[0, T]$ , but it can gracefully handle intervals where fewer than  $n$  spikes have actually occurred. For example, gAT-2 assumes exactly two spikes in the sampling interval. If only one spike exists in the interval, gAT-2 reconstructs a second spike with markedly reduced width compared with a real spike, allowing us to detect that only one spike actually occurred. On the other hand, if more than two spikes occur, then gAT-2 will miss the larger number of spikes.

Suppose we have  $s$  spikes in the sampling interval, where  $1 \leq s \leq n$ . Each spike manifests itself in the comparator output  $x(t)$  as a square wave centered at  $t_i$  with pulse width  $w_i$  and height 1. Assume the spikes are contained completely within the interval, i.e.,  $0 \leq t_i - w_i/2$  and  $t_i + w_i/2 \leq T$  for all  $0 \leq i \leq s$ . In addition, the spike times are indexed by increasing order and do not overlap, so  $t_i + w_i/2 < t_{i+1} - w_{i+1}/2$  for  $1 \leq i < s$ .

Using the data provided by gAT during the sampling interval, we want to detect the number of spikes, spike times  $t_i$ , and pulse widths  $w_i$  for each spike. We now derive a formula for these quantities from the gAT samples.

To begin, write the comparator output in terms of spike times and pulse widths:

$$x(t) = \begin{cases} 1 & \text{if } t_1 - \frac{w_1}{2} \leq t \leq t_1 + \frac{w_1}{2} \\ \vdots & \vdots \\ 1 & \text{if } t_s - \frac{w_s}{2} \leq t \leq t_s + \frac{w_s}{2} \\ 0 & \text{otherwise} \end{cases} \quad (A15)$$

Next, write the gAT samples in terms of the comparator output. Recall from Eqs. A12 and A13 that

$$y_i = x_k(T) = \int_0^T \int_0^{\tau_k} \dots \int_0^{\tau_2} x(\tau_1) d\tau_1 \dots d\tau_{k-1} d\tau_k \quad (A16)$$

Next, use the Cauchy formula for repeated integration to create an equation that relates gAT samples  $y_i$  with spike times  $t_i$  and pulse widths  $w_i$ . The Cauchy formula for repeated integration states that

$$\int_0^T \int_0^{\tau_k} \dots \int_0^{\tau_2} x(\tau_1) d\tau_1 \dots d\tau_{k-1} d\tau_k = \frac{1}{(k-1)!} \int_0^T x(\tau) (T-\tau)^{k-1} d\tau \quad (A17)$$

By substituting Eq. A15 into Eq. A17, we obtain an equation for  $y_i$ , where  $1 \leq i \leq k$ , in terms of the  $s$  spike times  $t_1, t_2, \dots, t_s$  and the  $s$  spike widths  $w_1, w_2, \dots, w_s$ :

$$y_i = \frac{1}{(i-1)!} \sum_{j=1}^s \int_{t_j - w_j/2}^{t_j + w_j/2} (T-\tau)^{i-1} d\tau \quad (A18)$$

We need to solve this special case of the Cauchy formula, either analytically or numerically, to recover spike times and pulse widths.

**SOLUTIONS TO THE GAT CAUCHY FORMULA.** We determined the exact solution for  $s = 2$  using Mathematica, presented in *Analytic closed-form solution for gAT-2* below. The resulting equations are provided in the following section. For  $s > 2$  spikes per sampling interval, the Cauchy formula results in a system of equations with higher order polynomials. These may be solved by iterative methods, although specific recommendations are beyond the scope of this work. Our empirical analysis with primate motor cortex (see main text) suggests that even sampling intervals as low as 10 Hz result in more than 70% of those intervals containing no more than two spikes. As such, methods accounting for more than two spikes per interval may be superfluous for many practical applications. One path to higher numbers of spikes per interval could involve approximating the comparator outputs as Dirac delta functions instead of square waves and subsequently applying variants of the annihilating filter (Suo et al. 2013) or the noise-robust IterML method (Wein and Srinivasan 2013), both from theory on finite-rate of innovation signals.

**DETERMINING THE ACTUAL NUMBER OF SPIKES IN THE INTERVAL FROM DATA.** In practice, the actual number of spikes  $s$  in the sampling interval must be determined from the data, where gAT- $n$  entertains the possibility that  $0 \leq s \leq n$ . To do so, starting from  $i = 0$ , the interval is assumed to have  $i$  spikes, and the parameters  $t_1, \dots, t_i$  and  $w_1, \dots, w_i$  are determined based on the first  $2i$  repeated integrals. These

parameters describe the time and width of the  $i$  spikes. The  $(2i + 1)$ -th repeated integral,  $y_{2i+1}$ , is predicted from the parameters.  $i$  is increased until the error between the predicted  $y_{2i+1}$  and the true  $y_{2i+1}$  is sufficiently small, or when  $i$  reaches  $n$ . The value of  $i$  when the algorithm terminates is the predicted number of spikes, and the parameters  $t_1, \dots, t_i$  are the predicted times of the spikes.

*Analytic closed-form solution for gAT-2.* Below, we include the exact solution for spike times  $(t_1, t_2)$  used in gAT-2, as well as pulse widths  $(w_1, w_2)$ .

### A Temporal Resolution (Subject 2)

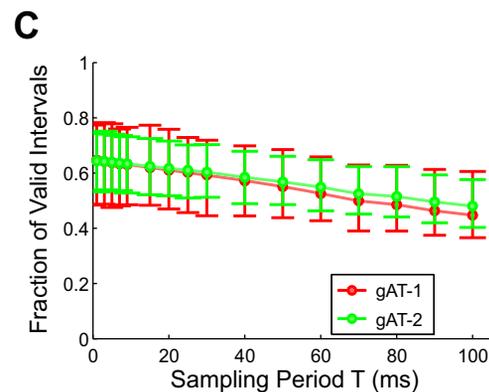
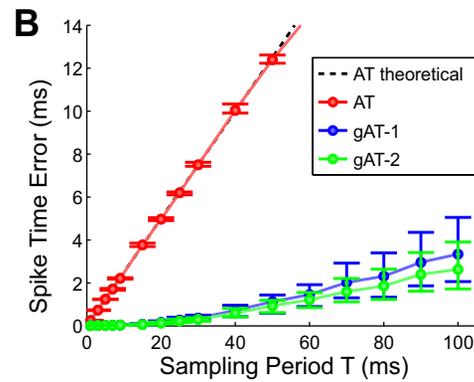
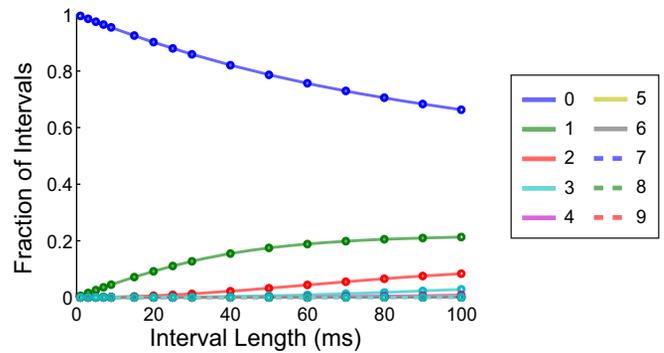


Fig. 9. Temporal resolution at various sampling periods in the second subject. Trends are unchanged compared with *subject 1* in Fig. 4. **A:** the fraction of intervals with a certain number of spikes within the interval. As the sampling period increases, intervals tend to have larger numbers of spikes. **B:** average error in estimated spike time in intervals with exactly one correctly identified spike. AT always predicts that the spike is at the center of the interval, and its behavior closely matches the theoretical curve. gAT-1 and gAT-2 have much lower errors than AT. **C:** the fraction of intervals that correctly identify the number of spikes. AT and gAT-1 always predict the same number of spikes because they are only able to differentiate between intervals with no spikes and intervals with a nonzero number of spikes. gAT-2 is able to differentiate between intervals with one spike and intervals with two spikes, and gAT-2 is able to improve the accuracy of spike count predictions.

ROC Performance (Subject 2)

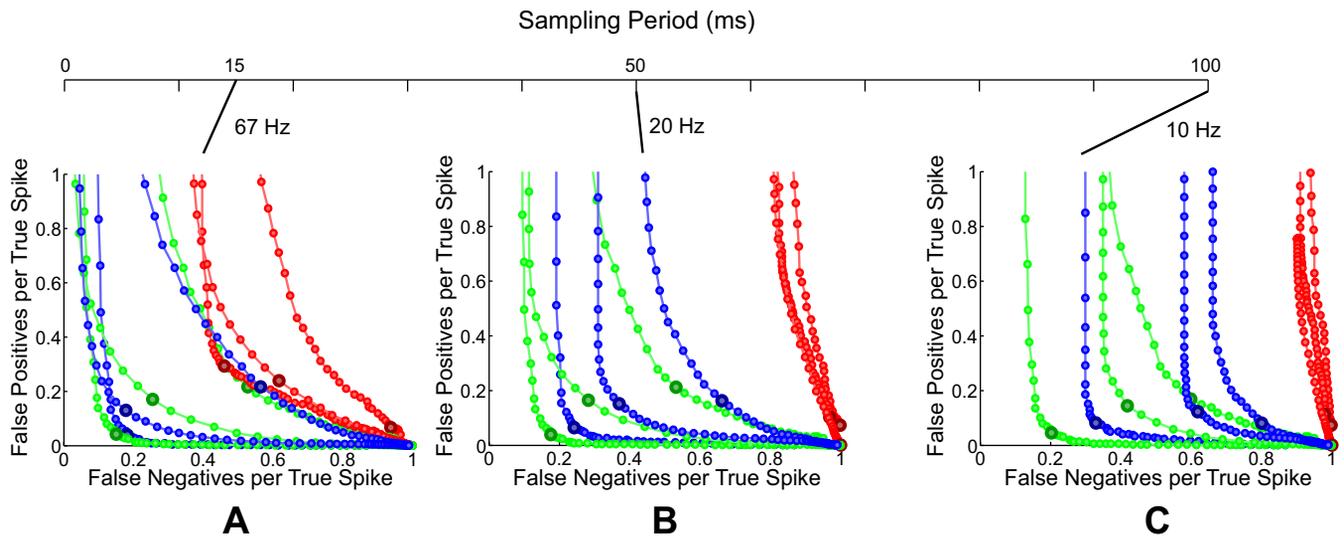


Fig. 10. ROC at various sampling periods in the second subject: 67 Hz (A), 20 Hz (B), 10 Hz (C). Trends are unchanged compared with *subject 1* in Fig. 5. Three representative curves (worst, median, and best) are shown for each method. The large circle on each curve represents the best operating point with fewest total errors. gAT-2 performs the best, followed by gAT-1 and AT. For all three methods, performance decreases with increasing sampling period. Colors are as defined in Fig. 5 legend.

$$\begin{aligned}
 t_1 = & 1/(-y_2^2 + 2y_1y_3) \left[ -2Ty_2^2 + 4Ty_1y_3 + 2y_2y_3 \right. \\
 & - (Ty_2^4)/(-y_2^2 + 2y_1y_3) + (4Ty_1y_2^2y_3)/(-y_2^2 + 2y_1y_3) \\
 & + (y_2^3y_3)/(-y_2^2 + 2y_1y_3) - (4Ty_1^2y_3^2)/(-y_2^2 + 2y_1y_3) - (2y_1y_2y_3^2)/(-y_2^2 + 2y_1y_3) \\
 & - 6y_1y_4 - (3y_1y_2^2y_4)/(-y_2^2 + 2y_1y_3) + (6y_1^2y_3y_4)/(-y_2^2 + 2y_1y_3) \\
 & - (y_2^2\sqrt{-3y_2^2y_3^2 + 8y_1y_3^3 + 6y_2^2y_4 - 18y_1y_2y_3y_4 + 9y_1^2y_4^2})/(-y_2^2 + 2y_1y_3) \\
 & \left. + (2y_1y_3\sqrt{-3y_2^2y_3^2 + 8y_1y_3^3 + 6y_2^2y_4 - 18y_1y_2y_3y_4 + 9y_1^2y_4^2})/(-y_2^2 + 2y_1y_3) \right] \tag{A19}
 \end{aligned}$$

$$t_2 = 1/(-y_2^2 + 2y_1y_3) \times (-Ty_2^2 + 2Ty_1y_3 + y_2y_3 - 3y_1y_4 - \sqrt{-3y_2^2y_3^2 + 8y_1y_3^3 + 6y_2^2y_4 - 18y_1y_2y_3y_4 + 9y_1^2y_4^2}) \tag{A20}$$

$$\begin{aligned}
 w_1 = & \left[ -Ty_2^5 + 5Ty_1y_2^3y_3 + y_2^4y_3 - 6Ty_1^2y_2y_3^2 - 6y_1y_2^2y_3^2 - 8y_1^2y_3^3 - (Ty_2^7)/(-y_2^2 + 2y_1y_3) \right. \\
 & + (7Ty_1y_2^5y_3)/(-y_2^2 + 2y_1y_3) + (y_2^6y_3)/(-y_2^2 + 2y_1y_3) - (16Ty_1^2y_2^3y_3^2)/(-y_2^2 + 2y_1y_3) \\
 & - (5y_1y_2^4y_3^2)/(-y_2^2 + 2y_1y_3) + (12Ty_1^3y_2y_3^3)/(-y_2^2 + 2y_1y_3) \\
 & + (6y_1^2y_2^2y_3^3)/(-y_2^2 + 2y_1y_3) - 3Ty_1^2y_2^2y_4 + 3y_1y_2^3y_4 + 6Ty_1^3y_3y_4 - 6y_1^2y_2y_3y_4 \\
 & - (3Ty_1^2y_2^4y_4)/(-y_2^2 + 2y_1y_3) - (3y_1y_2^5y_4)/(-y_2^2 + 2y_1y_3) + (12Ty_1^3y_2^2y_3y_4)/(-y_2^2 + 2y_1y_3) \\
 & + (18y_1^2y_2^3y_3y_4)/(-y_2^2 + 2y_1y_3) - (12Ty_1^4y_3^2y_4)/(-y_2^2 + 2y_1y_3) - (24y_1^3y_2y_3^2y_4)/(-y_2^2 + 2y_1y_3) \\
 & - (9y_1^3y_2^2y_4^2)/(-y_2^2 + 2y_1y_3) + (18y_1^4y_3y_4^2)/(-y_2^2 + 2y_1y_3) \\
 & - (y_2^5\sqrt{-3y_2^2y_3^2 + 8y_1y_3^3 + 6y_2^2y_4 - 18y_1y_2y_3y_4 + 9y_1^2y_4^2})/(-y_2^2 + 2y_1y_3) \\
 & + (5y_1y_2^3y_3\sqrt{-3y_2^2y_3^2 + 8y_1y_3^3 + 6y_2^2y_4 - 18y_1y_2y_3y_4 + 9y_1^2y_4^2})/(-y_2^2 + 2y_1y_3) \\
 & - (6y_1^2y_2y_3^2\sqrt{-3y_2^2y_3^2 + 8y_1y_3^3 + 6y_2^2y_4 - 18y_1y_2y_3y_4 + 9y_1^2y_4^2})/(-y_2^2 + 2y_1y_3) \\
 & - (3y_1^2y_2^2y_4\sqrt{-3y_2^2y_3^2 + 8y_1y_3^3 + 6y_2^2y_4 - 18y_1y_2y_3y_4 + 9y_1^2y_4^2})/(-y_2^2 + 2y_1y_3) \\
 & \left. + (6y_1^3y_3y_4\sqrt{-3y_2^2y_3^2 + 8y_1y_3^3 + 6y_2^2y_4 - 18y_1y_2y_3y_4 + 9y_1^2y_4^2})/(-y_2^2 + 2y_1y_3) \right] \tag{A21}
 \end{aligned}$$

$$\begin{aligned}
 w_2 = & \left[ -Ty_2^5 + 5Ty_1y_2^3y_3 + y_2^4y_3 - 6Ty_1^2y_2y_3^2 - 8y_1^2y_3^3 - (Ty_2^7)/(-y_2^2 + 2y_1y_3) \right. \\
 & + (7Ty_1y_2^5y_3)/(-y_2^2 + 2y_1y_3) + (y_2^6y_3)/(-y_2^2 + 2y_1y_3) - (16Ty_1^2y_2^3y_3^2)/(-y_2^2 + 2y_1y_3) \\
 & - (5y_1y_2^4y_3^2)/(-y_2^2 + 2y_1y_3) + (12Ty_1^3y_2y_3^3)/(-y_2^2 + 2y_1y_3) + (6y_1^2y_2^2y_3^3)/(-y_2^2 + 2y_1y_3) \\
 & - 3Ty_1^2y_2^2y_4 - 9y_1y_2^3y_4 + 6Ty_1^3y_3y_4 + 30y_1^2y_2y_3y_4 - (3Ty_1^2y_2^4y_4)/(-y_2^2 + 2y_1y_3) \\
 & \left. + (6y_1^3y_3y_4\sqrt{-3y_2^2y_3^2 + 8y_1y_3^3 + 6y_2^2y_4 - 18y_1y_2y_3y_4 + 9y_1^2y_4^2})/(-y_2^2 + 2y_1y_3) \right] \tag{A22}
 \end{aligned}$$

$$\begin{aligned}
 & - (3y_1y_2^5y_4)/(-y_2^2 + 2y_1y_3) + (12Ty_1^3y_2^2y_3y_4)/(-y_2^2 + 2y_1y_3) + (18Ty_1^2y_2^3y_3y_4)/(-y_2^2 + 2y_1y_3) \\
 & + (12Ty_1^4y_2^3y_4)/(-y_2^2 + 2y_1y_3) - (24Ty_1^3y_2y_2^3y_4)/(-y_2^2 + 2y_1y_3) - 18y_1^3y_4^2 \\
 & - (9y_1^3y_2^2y_4^2)/(-y_2^2 + 2y_1y_3) + (18y_1^4y_3y_4^2)/(-y_2^2 + 2y_1y_3) \\
 & - (y_2^5\sqrt{-3y_2^2y_3^2 + 8y_1y_3^3 + 6y_2^3y_4 - 18y_1y_2y_3y_4 + 9y_1^2y_4^2})/(-y_2^2 + 2y_1y_3) \\
 & + (5y_1y_2^3y_3\sqrt{-3y_2^2y_3^2 + 8y_1y_3^3 + 6y_2^3y_4 - 18y_1y_2y_3y_4 + 9y_1^2y_4^2})/(-y_2^2 + 2y_1y_3) \\
 & - (6y_1^2y_2y_3^2\sqrt{-3y_2^2y_3^2 + 8y_1y_3^3 + 6y_2^3y_4 - 18y_1y_2y_3y_4 + 9y_1^2y_4^2})/(-y_2^2 + 2y_1y_3) \\
 & - (3y_1^2y_2^2y_4\sqrt{-3y_2^2y_3^2 + 8y_1y_3^3 + 6y_2^3y_4 - 18y_1y_2y_3y_4 + 9y_1^2y_4^2})/(-y_2^2 + 2y_1y_3) \\
 & + (6y_1^3y_3y_4\sqrt{-3y_2^2y_3^2 + 8y_1y_3^3 + 6y_2^3y_4 - 18y_1y_2y_3y_4 + 9y_1^2y_4^2})/(-y_2^2 + 2y_1y_3) \Big] \\
 & / (6y_2^2y_3^2 - 16y_1y_3^3 - 12y_2^3y_4 + 36y_1y_2y_3y_4 - 18y_1^2y_4^2)
 \end{aligned}$$

Additional Performance Analysis

Algorithm performance in *subject 2* (including 7 motor cortical neurons on separate electrodes over 8.4 min, for a total of 18,000 spikes) confirmed all trends described with *subject 1*. Specifically, gAT again outperformed AT on a number of measures, including temporal resolution (Fig. 9), receiver operating characteristics (Fig. 10), and spike-LFP phase accuracy (Fig. 11). The preferred direction analysis in *subject 2* again showed that temporal resolution in spike time reconstruction is not universally beneficial, such as in velocity-tuning curve estimation where spike times are related to the slowly-varying movement direction (Fig. 12).

GRANTS

B. D. He and A. Wein were supported by the Caltech Summer Undergraduate Research Fellowship. L. Srinivasan was supported by the American Heart Association Scientist Development Grant (11SDG7550015) and the UCLA Radiology Exploratory Development Grant.

DISCLOSURES

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

AUTHOR CONTRIBUTIONS

Author contributions: B.D.H., A.W., L.R.V., J.K., and L.S. conception and design of research; B.D.H., A.W., A.G.R., and L.S. performed experiments; B.D.H., A.W., A.G.R., and L.S. analyzed data; B.D.H., A.W., L.R.V., J.K., A.G.R., and L.S. interpreted results of experiments; B.D.H., A.W., and L.S. prepared figures; B.D.H., A.W., and L.S. drafted manuscript; B.D.H., A.W., L.R.V., J.K., A.G.R., and L.S. edited and revised manuscript; B.D.H., A.W., L.R.V., J.K., A.G.R., and L.S. approved final version of manuscript.

REFERENCES

Alivisatos AP, Andrews AM, Boyden ES, Chun M, Church GM, Deisseroth K, Donoghue JP, Fraser SE, Lippincott-Schwartz J, Looger LL, Masmanidis S, McEuen PL, Nurmikko AV, Park H, Peterka DS, Reid C, Roukes ML, Scherer A, Schnitzer M, Sejnowski TJ, Shepard KL, Tsao D, Turrigiano G, Weiss PS, Xu C, Yuste R, Zhuang X. Nanotools for neuroscience and brain activity mapping. *ACS Nano* 7: 1850–1866, 2013a.

Alivisatos AP, Chun M, Church GM, Deisseroth K, Donoghue JP, Greenspan RJ, McEuen PL, Roukes ML, Sejnowski TJ, Weiss PS, Yuste R. Neuroscience. The brain activity map. *Science* 339: 1284–1285, 2013b.

Brown EN, Kass RE, Mitra PP. Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nat Neurosci* 7: 456–461, 2004.

Spike-LFP Phase Accuracy (Subject 2)

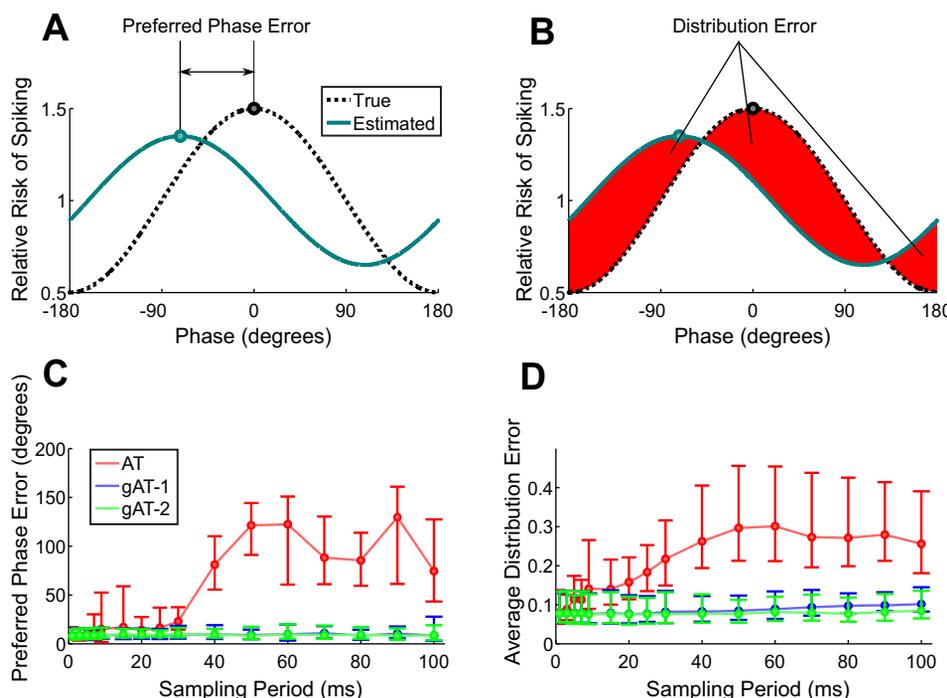


Fig. 11. Robustness of spike-LFP phase locking estimation to increasing sampling period in the second subject. Trends are unchanged compared with subject 1 in Fig. 6. gAT outperforms AT in reconstructing preferred phase, achieving robust performance for sampling rates as low as 10 Hz (sampling period of 100 ms). See main text for details. A: explanation of preferred phase error (difference between the maxima of the two curves). B: explanation of distribution error (area between true curve and predicted curve). C: preferred phase error for different sampling periods. D: distribution error for different sampling periods.

## Preferred Direction Accuracy (Subject 2)

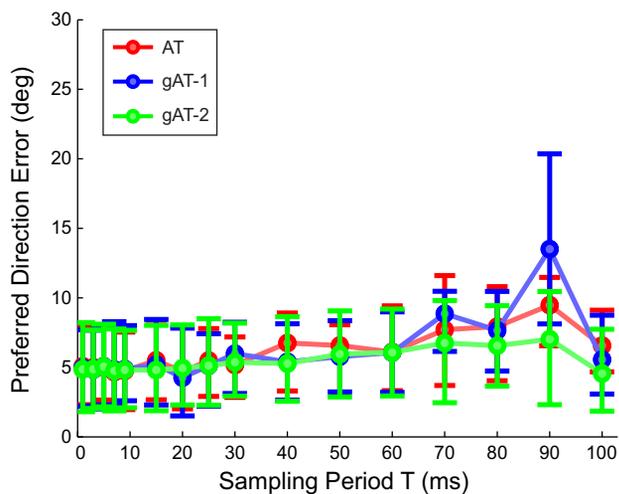


Fig. 12. Robustness of motor cortical preferred direction estimation to increasing sampling period in the second subject. Trends are unchanged compared with *subject 1* in Fig. 7. As with the main text, differences between AT and gAT are negligible in reconstructing preferred direction, and robust across all methods down to 10-Hz sampling rate. In this graph, one neuron out of seven from the second subject was observed to have a very flat tuning curve, which caused large errors in preferred direction. This neuron was removed to avoid skewing the error with this outlier.

**Bulach C, Bihl U, Ortmanns M.** Evaluation study of compressed sensing for neural spike recordings. In: *Proceedings of the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Piscataway, NJ: EMBS, 2012, p. 3507–3510.

**Caballero J, Uriguen JA, Schultz SR, Dragotti PL.** Spike sorting at sub-Nyquist rates. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2012)*. New York: IEEE, 2012, p. 585–588.

**Carmena JM.** Advances in neuroprosthetic learning and control. *PLoS Biol* 11: e1001561, 2013.

**Charbiwala Z, Karkare V, Gibson S, Markovic D, Srivastava MB.** Compressive sensing of neural action potentials using a learned union of supports. In: *Proceedings of the 2011 International Conference on Body Sensor Networks (BSN)*. New York: IEEE, 2011, p. 53–58.

**Chen F, Chandrakasan AP, Stojanovic V.** A signal-agnostic compressed sensing acquisition system for wireless and implantable sensors. In: *Proceedings of the 2010 International Conference on Custom Integrated Circuits (CICC)*. New York: IEEE, 2010, p. 1–4.

**Chen F, Chandrakasan AP, Stojanovic VM.** Design and analysis of a hardware-efficient compressed sensing architecture for data compression in wireless sensors. *IEEE J Solid-State Circuits* 47: 744–756, 2012.

**Ghaed MH, Chen G, Haque RU, Wieckowski M, Kim Y, Kim G, Lee Y, Lee I, Fick D, Kim D.** Circuits for a cubic-millimeter energy-autonomous wireless intraocular pressure monitor. *IEEE Trans Circuits Syst I Regul Pap* 60: 3152–3162, 2013.

**Gray CM, Maldonado PE, Wilson M, McNaughton B.** Tetrodes markedly improve the reliability and yield of multiple single-unit isolation from multi-unit recordings in cat striate cortex. *J Neurosci Methods* 63: 43–54, 1995.

**Harris KD, Henze DA, Csicsvari J, Hirase H, Buzsaki G.** Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *J Neurophysiol* 84: 401–414, 2000.

**Harrison RR.** The design of integrated circuits to observe brain activity. *Proc IEEE Inst Electr Electron Eng* 96: 1203–1216, 2008.

**He BD, Wein A, Srinivasan L.** Feasibility of FRI-based square-wave reconstruction with quantization error and integration noise. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2015)*. New York: IEEE, 2015, In press.

**Hegde C, Duarte MF, Cevher V.** Compressive sensing recovery of spike trains using a structured sparsity model. In: *Proceedings of SPARS'09-Signal Processing with Adaptive Sparse Structured Representations*. Rennes, France: IRISA, 2009.

**Henze DA, Borhegyi Z, Csicsvari J, Mamiya A, Harris KD, Buzsáki G.** Intracellular features predicted by extracellular recordings in the hippocampus in vivo. *J Neurophysiol* 84: 390–400, 2000.

**Kusuma J, Maravic I, Vetterli M.** Sampling with finite rate of innovation: channel and timing estimation for UWB and GPS. In: *Proceedings of the 2003 ICC'03 IEEE International Conference on Communications*. New York: IEEE, 2003, p. 3540–3544.

**Lebedev M.** Brain-machine interfaces: an overview. *Transl Neurosci* 5: 99–110, 2014.

**Marblestone AH, Zamft BM, Maguire YG, Shapiro MG, Cybulski TR, Glaser JI, Amodoi D, Stranges PB, Kalhor R, Dalrymple DA.** Physical principles for scalable neural recording. *Front Comput Neurosci* 7: 137, 2013.

**Meister M, Pine J, Baylor DA.** Multi-neuronal signals from the retina: acquisition and analysis. *J Neurosci Methods* 51: 95–106, 1994.

**Murmann B.** Energy limits in A/D converters. In: *Proceedings of 2013 IEEE Faible Tension Faible Consommation (FTFC)*. New York: IEEE, 2013, p. 1–4.

**Narasimhan S, Tabib-Azar M, Chiel HJ, Bhunia S.** Neural data compression with wavelet transform: a vocabulary based approach. In: *Proceedings of the CNE '07 3rd International IEEE/EMBS Conference on Neural Engineering*. New York: IEEE, 2007, p. 666–669.

**Oñativia J, Schultz SR, Dragotti PL.** A finite rate of innovation algorithm for fast and accurate spike detection from two-photon calcium imaging. *J Neural Eng* 10: 046017, 2013.

**Oweiss KG.** A systems approach for data compression and latency reduction in cortically controlled brain machine interfaces. *IEEE Trans Biomed Eng* 53: 1364–1377, 2006.

**Pelgrom MJM.** *Analog-to-Digital Conversion*. New York: Springer, 2010, p. xx, 455.

**Pnevmatikakis EA, Paninski L.** Sparse nonnegative deconvolution for compressive calcium imaging: algorithms and phase transitions. In: *Adv Neural Inf Process Syst* 2013: 1250–1258, 2013.

**Rabaey J.** *Low Power Design Essentials*. New York: Springer, 2009.

**Richardson AG, Borghi T, Bizzi E.** Activity of the same motor cortex neurons during repeated experience with perturbed movement dynamics. *J Neurophysiol* 107: 3144–3154, 2012.

**Schmale S, Knoop B, Hoeffmann J, Peters-Drolshagen D, Paul S.** Joint compression of neural action potentials and local field potentials. In: *Proceedings of the 2013 Asilomar Conference on Signals, Systems and Computers*. New York: IEEE, 2013, p. 1823–1827.

**Schwarz DA, Lebedev MA, Hanson TL, Dimitrov DF, Lehew G, Meloy J, Rajangam S, Subramanian V, Ifft PJ, Li Z.** Chronic, wireless recordings of large-scale brain activity in freely moving rhesus monkeys. *Nat Methods* 11: 670–676, 2014.

**Spira ME, Hai A.** Multi-electrode array technologies for neuroscience and cardiology. *Nat Nanotechnol* 8: 83–94, 2013.

**Suo Y, Zhang J, Etienne-Cummings R, Tran TD, Chin S.** Energy-efficient two-stage Compressed Sensing method for implantable neural recordings. In: *Proceedings of the IEEE Biomedical Circuits and Systems Conference (BioCAS 2013)*. New York: IEEE, 2013, p. 150–153.

**Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN.** A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J Neurophysiol* 93: 1074–1089, 2005.

**Wattanapanitch W, Fee M, Sarpeshkar R.** An energy-efficient micropower neural recording amplifier. *IEEE Trans Biomed Circuits Syst* 1: 136–147, 2007.

**Wein A, Srinivasan L.** IterML: a fast, robust algorithm for estimating signals with finite rate of innovation. *IEEE Trans Signal Process* 61: 5324–5336, 2013.

**Yoo J, Becker S, Loh M, Monge M, Candes E, Emami-Neyestanak A.** A 100 MHz–2 GHz 12.5 × sub-Nyquist rate receiver in 90 nm CMOS. In: *Proceedings of the 2012 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*. New York: IEEE, 2012, p. 31–34.

**Zanos S, Zanos TP, Marmarelis VZ, Ojemann GA, Fetz EE.** Relationships between spike-free local field potentials and spike timing in human temporal cortex. *J Neurophysiol* 107: 1808–1821, 2012.