



Power-saving design opportunities for wireless intracortical brain–computer interfaces

Nir Even-Chen ^{1,11}✉, Dante G. Muratore ^{1,2,11}, Sergey D. Stavisky ^{1,2,3}, Leigh R. Hochberg ^{4,5,6},
Jaimie M. Henderson ^{2,3}, Boris Murmann ^{1,2,12} and Krishna V. Shenoy ^{1,2,7,8,9,10,12}

The efficacy of wireless intracortical brain–computer interfaces (iBCIs) is limited in part by the number of recording channels, which is constrained by the power budget of the implantable system. Designing wireless iBCIs that provide the high-quality recordings of today’s wired neural interfaces may lead to inadvertent over-design at the expense of power consumption and scalability. Here, we report analyses of neural signals collected from experimental iBCI measurements in rhesus macaques and from a clinical-trial participant with implanted 96-channel Utah multielectrode arrays to understand the trade-offs between signal quality and decoder performance. Moreover, we propose an efficient hardware design for clinically viable iBCIs, and suggest that the circuit design parameters of current recording iBCIs can be relaxed considerably without loss of performance. The proposed design may allow for an order-of-magnitude power savings and lead to clinically viable iBCIs with a higher channel count.

Electrophysiological devices are widely used in basic neuroscience research to measure the activity of a population of neurons (using electroencephalography and electrocorticography) or individual neurons (using intracortical electrodes) to understand the function of the nervous system. Decades of electrophysiological research initially aimed at understanding the brain gave rise to the field of brain–computer interfaces, also referred to as brain–machine interfaces or neural prostheses. Brain–computer interfaces provide a direct communication path between the brain and an external device. These can be used in research to better understand the brain, but they are increasingly intended for clinical applications¹.

Brain–computer interfaces can help restore lost sensory capabilities through stimulation (vision and hearing), or restore lost motor capabilities of people with motor impairments (for example, owing to amyotrophic lateral sclerosis (ALS), brainstem stroke or spinal cord injury). In clinical motor prosthesis applications, brain–computer interfaces estimate the user’s intention from brain activity and use this ‘decoded’ intention to guide the person’s own limb^{2,3} or an assistive device, such as a prosthetic arm^{4–6} or a computer cursor^{7–9} (Fig. 1). Different neural sensors (electroencephalography, electrocorticography and intracortical electrodes) can be used as part of a brain–computer interface for clinical applications. Recent iBCI studies have shown promising results in pilot clinical trials by enabling high-performance computer cursor control, making them prime candidates for contributing to assistive technology for people with paralysis^{3,5,8,10,11}. Although these results are encouraging, for most applications, iBCIs would benefit from further improvements to make them suitable for widespread, standard-of-care human clinical use.

There is considerable industrial and academic interest in continuing to advance all aspects of iBCI design. Two major

requirements for improving iBCIs are: increasing the number of recording electrodes; and implementing wireless transcutaneous implants^{12–14}. Increasing the number of recorded neurons and the variety of recorded cortical areas is predicted to lead to iBCI performance improvements (see refs. ^{13,15}) and might also enable control of more sophisticated prosthetic devices (with higher degrees of freedom)^{14,16,17}. In addition, wireless links will enable the development of transcutaneous systems that minimize the risk of infection and improve aesthetic appearance, user mobility and independence^{12,13}. Existing chronic systems for non-human primates that record from hundreds to thousands of electrodes are power-hungry, and are usually based on wired communication that will not translate well to people^{14,18,19}. Although substantial progress has been made in developing fully implantable wireless technologies that can support approximately 100 electrodes, the field (including future clinical use) will benefit from smaller devices with larger numbers of electrodes (channels)^{20–23}.

Current systems were designed for basic neuroscience research to record and transmit wide-bandwidth signals with high resolution, thus requiring relatively high power consumption. This ‘leave nothing behind’ approach enables the extraction of a variety of signals, including action potentials (spikes) and local field potentials and permits ‘spike sorting’ to attribute action potentials to putative individual neurons²⁴.

As custom signal specifications for iBCI applications become clearer and are diverging from those needed for basic neuroscience, the neuroscience community has investigated what the required signal characteristics for iBCI decoders are^{25–29}. At the same time, the engineering community has focused on designing more efficient neural interfaces. Recent designs, aimed at reducing the quantity of raw data generated by conventional neural interfaces,

¹Department of Electrical Engineering, Stanford University, Stanford, CA, USA. ²Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA, USA.

³Department of Neurosurgery, Stanford University, Stanford, CA, USA. ⁴Department of Veterans Affairs Medical Center, Center for Neurorestoration and Neurotechnology, Providence, RI, USA. ⁵School of Engineering and Carney Institute for Brain Science, Brown University, Providence, RI, USA. ⁶Center for Neurotechnology and Neurorecovery, Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA.

⁷Department of Bioengineering, Stanford University, Stanford, CA, USA. ⁸Department of Neurobiology, Stanford University, Stanford, CA, USA. ⁹Howard Hughes Medical Institute at Stanford University, Stanford, CA, USA. ¹⁰The Bio-X Institute, Stanford University, Stanford, CA, USA. ¹¹These authors contributed equally: Nir Even-Chen, Dante G. Muratore. ¹²These authors jointly supervised this work: Boris Murmann and Krishna V. Shenoy.

✉e-mail: nirec@stanford.edu

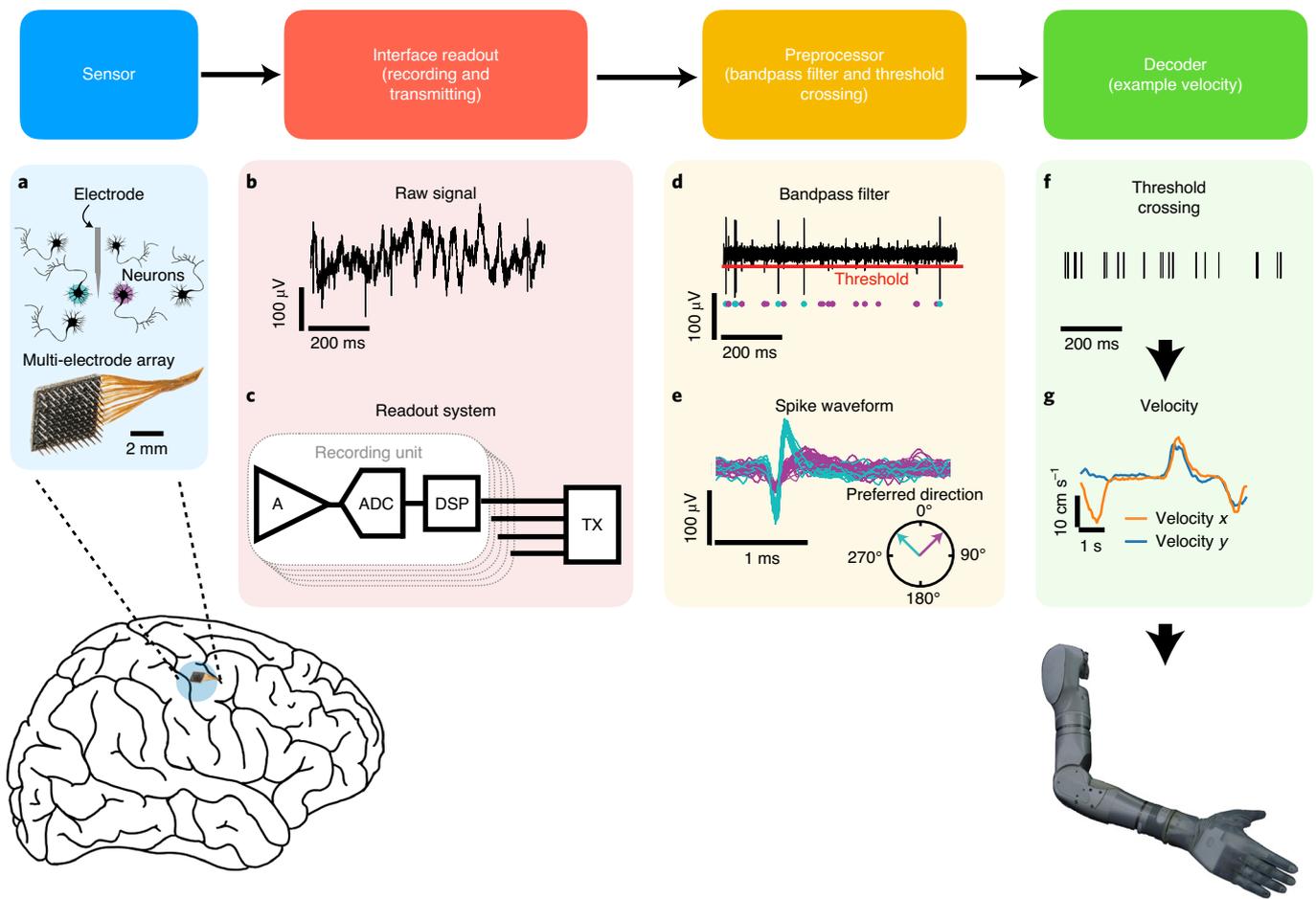


Fig. 1 | iBCI schematic signal flow. Neural activity is recorded using an electrode array (such as the Utah array in **a**), with each electrode measuring the neural activity in its vicinity. In **b**, the raw analogue signal (**c**) is amplified in the frequency band of interest (by a neural amplifier, A), digitized (by an analogue-to-digital converter, ADC) and then transmitted (by a transmitter, TX) to a computer for further processing. In some cases, some preprocessing (such as spike detection) happens before the transmitter by digital signal processing (DSP). Most iBCI neuroprosthetic studies process the signal by applying a further bandpass filter (in addition to that implemented in the neural amplifier) and a simple threshold crossing for spike detection (red line in **d**). The threshold crossings are evaluated for each electrode, although an electrode might record from multiple neurons simultaneously (the cyan and magenta dots in **d** indicate spikes from different neurons), which can mix correlations with behaviour (such as the preferred direction of movement) (**e**). Lastly, a decoder (such as a Kalman filter) estimates the user's intention (for example, the robotic arm's velocity) from the binary threshold-crossing signal (**f**), and sends the control signals (**g**) to the prosthetic controller.

use a wide range of techniques, such as on-chip thresholding^{30–32}, on-chip spike sorting³³, on-chip compression^{34–37} and compressive sensing^{38,39}. Despite these efforts, there is not yet a consensus on the required specifications for iBCI-oriented neural interfaces. As a result, current *in vivo* systems still typically record the neural signal with unnecessarily high fidelity^{20,22,23,40–46}. Here, we will present a holistic analysis validating how neural interface specifications for iBCIs can be relaxed with respect to basic neuroscience systems. Results are obtained using data collected from experimental iBCI measurements in rhesus macaques and a BrainGate2 clinical trial human participant. We suggest a considerably more power-efficient approach designed around the anticipated needs and constraints of clinical use. Our estimates indicate that this could reduce power consumption by an order of magnitude while maintaining high decoding performance. We focus on three main questions: (1) what type of signal should be recorded from the brain, (2) how reliable this signal should be, and (3) how the resulting new specifications will affect the iBCI neural interface design, with particular emphasis on its power consumption.

First, we discuss what type of neural signal is needed for iBCIs and how robust against noise the system must be. Next, we assess

which design specifications can be relaxed without substantially sacrificing iBCI performance. This informs a possible customized recording strategy for implantable clinical iBCIs. Finally, we describe how the new specifications affect the overall power consumption of the system and highlight the potential for substantial power savings.

Neural signal requirements for iBCI decoders

Binary threshold-crossing signals sampled at low rates. Current neural interfaces record and transmit wideband signals (0.1–10,000 Hz) at high resolution (10–16 bits); see Fig. 1b. The low-frequency spectrum contains the local field potential, which reflects a spatial averaging of the neural population activity in the electrode's surroundings^{47–49}. The high-frequency spectrum enables the capture of delicate features in the recorded signal, which can help in investigating spike waveforms, spike sorting and differentiating between neuron classes on the basis of their spike waveform. These properties are of interest for basic neuroscience research, but they are not necessarily essential for estimating the user's neural state or intention for iBCI applications (such as arm movement direction and speed intention)^{27,29,50,51}.

Most high-performing motor iBCIs use only a simple binary signal encoding of threshold-crossing events ('1' if a spike is detected, '0' otherwise) in a short time bin (see Fig. 1f) to decode the user's intention^{8,11,12,25,52–59}. The local field potentials contain less information about the intended movement (such as movement velocity) compared to spikes; they are beneficial for continuous iBCI control mostly when the spike signals have degraded severely^{8,12,26,60,61} or to distinguish a small set of discrete states⁶². A few studies, in addition to threshold-crossing events, also used high-frequency band power (250 Hz up to 3–5 kHz) for iBCIs^{3,9,63}. However, the importance of these additional signals might diminish as the number of electrodes in iBCI systems increases and as neural interfaces improve their ability to record thresholded spikes^{64,65}. Although recording and transmitting only local field potentials could also potentially reduce power consumption and should be further investigated, here we focus only on spike signals because they currently achieve the highest iBCI performance.

Sorting spikes by individual neurons detected on a given electrode (Fig. 1e) can be of interest when investigating single neuron modulation. In some cases, it can potentially improve iBCI performance by distinguishing spikes from neural populations that are tuned differently (for example, they might have different preferred directions) or by separating neurons from high-amplitude noise or high-frequency local field potential. However, this might be unnecessary for decoding purposes when the activity of the entire neuron population is summarized to evaluate the desired user intention (such as by linear combination). Accumulating evidence in non-human primates suggests that the benefit of spike sorting on iBCI performance is minimal compared to simple threshold crossing (a performance difference of about 5%)^{25–29}. Taken together, these developments suggest that iBCIs may benefit most from larger electrode counts, and can compromise on signal fidelity by transmitting only binary threshold-crossing events. In addition, sorting spikes on the device will increase the complexity of the circuit design because it might need to adaptively track moving signals owing to slight array movement and other non-stationary properties of the signal. Even efficient spike sorting techniques^{33,66} increase power consumption substantially when compared to threshold-crossing architectures. Because of the minimal benefits in terms of decoding performance and the high costs of integrating spike sorting into the device, we have decided to focus on neural interface designs that record, transmit and decode threshold-crossing events.

One can imagine transmitting even more abstract signal representations such as binned spike counts (the number of spikes in a defined time window, for example, 20 ms), dimensionality-reduced signals (that is, compressed signals with fewer channels; see ref. ⁶⁷), or just the output control signals for the prosthesis (such as velocity). While it is common in the iBCI field to use threshold crossings for decoding^{8,11,52–54,57,68}, the optimal bin size, dimensionality reduction technique, and the decoding and control algorithms are still under active investigation (see refs. ^{52,53,67,69,70}). There is evidence that small bin sizes (1 ms) can improve performance; however, the effect of bin size on performance is still unclear and probably depends on the decoding algorithm^{8,53,71}. Additional work is required before determining the trade-off between bin size, dimensionality reduction technique, decoder algorithm and system performance. Intensive processing before transmitting the signal may therefore limit the compatibility of the device with future algorithms. Given all this, we have chosen to follow a conservative approach and recommend requirements that are usual in the iBCI field. Thus, at present, we believe that transmitting the presence or absence of unsorted spikes in 1-ms bins is a well balanced level of simplification that fits current iBCI best practices, while still leaving the door open to future research and development.

Tolerance to high recording and transmission error rates. Wired iBCIs used in clinical research, like many other medical and

research devices, are currently designed to have minimal recording and transmission errors. This guarantees minimal spike misclassification. However, this high reliability of recorded and transmitted brain activity comes with a high cost in terms of integrated circuit power consumption. The empirical observation that there is a high degree of redundancy in the neural population spiking activity suggests that high signal fidelity might not be required for iBCIs. The aim in motor iBCI applications is to estimate the user's intention from the ensemble of all recorded neurons. Correlations across electrodes and the temporal smoothness of the underlying intention (such as hand velocity) mitigate the effect of noise on each electrode. Therefore, it may be possible to reduce the recording and transmission reliability of each electrode as long as the final decoding accuracy does not degrade. To demonstrate this, we investigated the tolerance of iBCI decoders to noise and quantified how much we could distort the spiking activity while keeping comparable decoding quality.

Specifically, we tested an iBCI system's robustness to spike errors. Digital communication fidelity is traditionally described by the bit error rate (BER), which is the rate of flipped bits from 0 to 1 and vice versa. For our purposes, to measure the accuracy of the recording and transmission of spikes, which are binary threshold-crossing events, we introduce the notion of the spike error rate (SER). SER is the rate at which a spike event was falsely transmitted or missed, that is, a bit in the binary spike train was flipped. The iBCI decoder is agnostic to the source of the error. Spike errors (that is, the SER) can result from the noise added to the analogue signal or DSP throughout the recording and transmitting system. For example, electrode or amplifier noise can be incorrectly detected as a spike. Also, the low resolution of the ADC can lead to erroneously detected spikes. Moreover, bit errors in the transmission system (BER) can also translate to spike errors, depending on the communication protocol (as will be discussed later). Hence, we distinguish between BER and SER because SER represents the error rate of the entire neural interface system, both in recording and in transmission.

To evaluate the robustness of the iBCI decoder to the SER, we measured non-human primate and human movement behaviour predictability post hoc, using an intentionally distorted neural signal. That is, we randomly flipped bits in the binary threshold spikes (Fig. 1f) and then tried to decode this distorted signal. We note that this noise injection is different from the common practice in the field: we flip spikes in 1-ms resolution and we do not add Gaussian noise to the spike rate in a longer time window (20–100 ms; refs. ^{63,69,72}). The original neural signals were recorded with Blackrock Microsystems recording systems⁷³ from able-bodied monkeys and a person with paralysis while they performed a computer cursor movement task. The monkeys moved the cursor to cued targets using their hands, whereas BrainGate2 clinical trial participant 'T5' controlled it with an iBCI system (see Methods section 'Neural and hand position recording'). As a decoding performance metric, we use the coefficient of determination (R^2), which describes how much of the variability of the cursor velocity can be predicted from the neural activity. First, we generated synthetically distorted neural signals by injecting errors in the recorded data at different SERs (10^{-6} to $10^{-0.5}$). This was done by flipping the binary spiking signal bits with an independent Bernoulli distribution with the probability of the SER (see Methods). Then, we estimated how well the distorted neural signal could predict the cursor velocity using a Kalman filter, a decoder widely used for iBCIs^{8,11,57} (Fig. 2a; see Methods).

Surprisingly, across the three monkeys and the human participant, the SER had only a statistically significant effect on performance at a rate higher than 10^{-3} (see horizontal bars Fig. 2b in the right top corner, two-sided Wilcoxon rank-sum test, $P < 0.05$), when compared to an undistorted signal (R_{org}^2). In other words, the neural interface recording and transmission system can tolerate a SER of up to 10^{-3} while maintaining comparable decoding

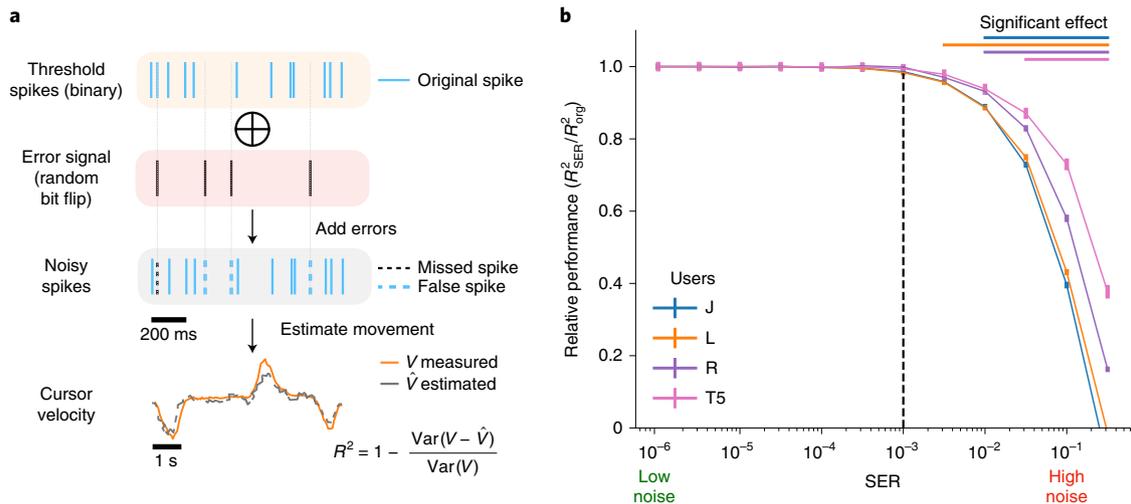


Fig. 2 | iBCI robustness to spike error. Data were recorded during centre-out-and-back cursor trials of the three monkeys (J, L and R) and the human participant (T5). **a**, The spike errors per second (SER) simulation process. First, binary threshold neural signal bits ('1' for spike, '0' otherwise) were flipped at different rates (10^{-6} – $10^{-0.5}$). Second, cursor velocity was predicted from the noisy signal, and the coefficient of determination R^2 was computed. Var, variance. **b**, Decoder performance (velocity coefficient of determination, R^2) as a function of added SER (ranges from 0–1). Values are normalized to performance when decoding an undistorted signal (that is, $R_{SER}/R_{original}$). Vertical bars along the lines represent the standard error of tenfold cross validation across ten days (a total of 100 R^2 estimates). Horizontal bars in the right top corner indicate significant change in performance compared to the undistorted signal (two-sided Wilcoxon rank-sum test, $P < 0.05$). Vertical dashed line indicates the SER in which performance starts to degrade. Bar colours correspond to the user (see key).

performance. Intuitively, if the SER is much smaller than the rate of spikes, it should not have a significant effect on the decoder; but if the SER is of the order of magnitude of the action potential emission rate (firing rate), then the movement intention information will become highly corrupted. Indeed, our results on SER robustness correspond to the recorded firing rate (about 10 spikes s^{-1} , see Supplementary Fig. S5) at the sample rate of 1,000 samples per second (1 kSps). We extended the same analysis to three additional methods: a linear regression decoder to predict cursor velocity (as done with a Kalman filter) and two types of discrete iBCI classifiers using a naive Bayes method and a support vector machine to predict the location of the cued target. Similar to the findings for the Kalman filter, the performance of these decoders started to degrade when the SER approached 10^{-3} (see Supplementary Methods section 'Offline decoders' and Supplementary Fig. S3). Below, we will focus our analysis on the Kalman filter decoder, given its ubiquity in iBCI systems and the comparable results obtained by the other methods.

The proposed metric for spike distortion (that is, SER) and the upper limit for the allowed error should thus guide iBCI neural interface designs, both in terms of recording and transmitting architecture, as well as specifications for the individual blocks in the system. Our results suggest that a neural interface (from electrode to transmitter in Fig. 2a,c) can distort the transmitted spike signal by a rate of up to 10^{-3} , which is several orders of magnitude higher than the BER traditionally targeted by telecommunication systems^{74,75}. Here, we propose a neural interface that transmits a binary threshold-crossing spike signal at 1 kSps with an error rate of up to 10^{-3} . This allows for trade-offs that could lower the device power requirements, and give rise to neural interface designs that are custom-optimized for iBCIs.

Custom neural interfaces for iBCIs

Recording and transmitting binary threshold-crossing signals while tolerating high spike error rates opens an avenue for new recording system architectures that are customized for the needs of clinical iBCIs. These devices can be more power efficient than current neural interfaces (Fig. 1c). To estimate the potential power savings

of this approach, we investigated the benefits of using conventional architectures with distinct parameter choices informed by anticipated clinical iBCI needs. We intend these conventional architecture estimates to be a starting point and an upper bound for very different future designs and architectures. In what follows we describe the key parameters of conventional systems, and perform analyses similar to those above, to examine how these parameters can be relaxed while still maintaining sufficient decoding accuracy.

Neural interface circuit design parameters. Figure 1c shows a conventional wireless implantable neural interface system similar to what is currently used in animal studies^{14,20} and is being developed for human use^{22,76}. The system consists of a sensor (such as a penetrating electrode array) connected to multiple recording units and a wireless transmitter (TX). Each recording unit contains a neural amplifier (A), an analogue-to-digital converter (ADC), and digital signal processing (DSP, such as a spike detector). The amplifier amplifies the neural signal in the frequency band of interest (f_0 , f_1 ; 0.3–7.5 kHz for action potentials). The noise introduced by this stage is defined as the input-referred noise level (\bar{v}_{in}^2) of the neural amplifier (see Supplementary Methods section 'Neural interface parameter simulation'). The output signal is then converted into a digital signal by the ADC. The ADC is defined by the sampling frequency (f_s) and the resolution (B , number of bits). Owing to circuit nonidealities (thermal noise and nonlinearities), the achieved effective resolution is usually 0.5–1.5 less than the number of bits, and it is defined by the signal-to-noise-ratio-based number of bits (SNR bits) when nonlinearities are not taken into account⁷⁷. Threshold-crossing detection is implemented in post-processing and adds negligible power consumption, as discussed in ref. ⁷⁸. Hence, it will not be part of our analysis here. The transmitter data rate is usually defined as $R = f_s B N_c$, where N_c is the number of channels, if no DSP is applied and raw data are transmitted directly. The main factors that guarantee the fidelity of this process are a wide filter frequency band, low noise, high quantizer sampling frequency, high quantizer resolution, and low data link BER. Unfortunately,

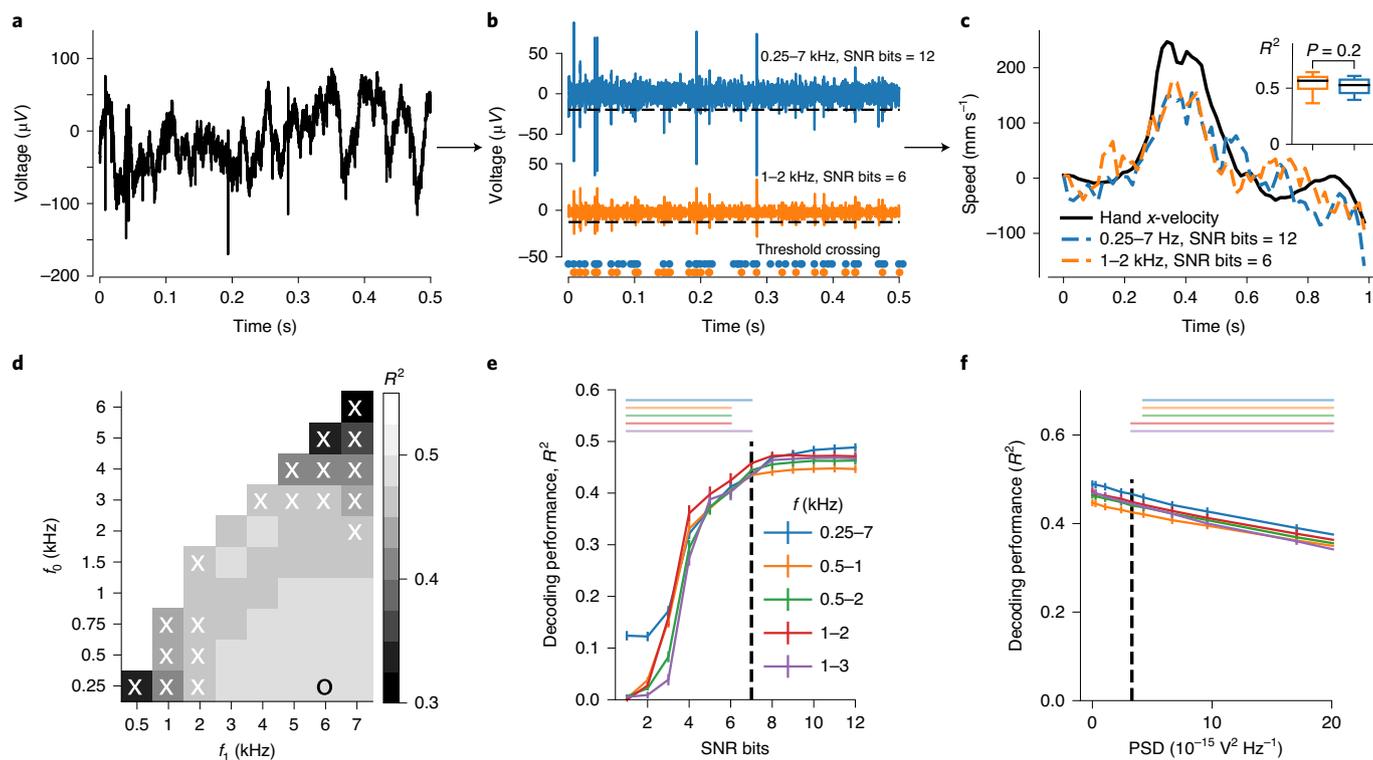


Fig. 3 | Study of iBCI performance as a function of the neural interface parameters (monkey J). **a–c.** Parameter simulation pipeline and examples for the simulation done in **d–f**. **a.** Raw data were filtered, re-quantized (with SNR bits) or corrupted with noise, and then thresholded (**b**). The simulation pipeline corresponds to the raw signal pipeline of a conventional system, where filtering occurs only in the analogue domain: $A \rightarrow \text{ADC} \rightarrow \text{DSP}$ (for spike detection). For more details see Methods section ‘Neural interface parameter simulation’. The threshold was set proportionally to each electrode’s RMS voltage. Orange and blue dots are the corresponding binary spike signals after thresholding for distorted and undistorted signals, respectively. **c.** Hand velocity was then predicted with a Kalman filter (tenfold cross-validated across 10 days with 100 trials in each day) from the preprocessed spike signal. Decoding performance was evaluated with velocity coefficient of determination ($n = 100$ R^2 estimates; see inset). Significance was tested with two-sided Wilcoxon rank-sum test ($P < 0.05$). **d.** R^2 as a function of f_0 and f_1 (the cut-off frequencies for the bandpass filter). The threshold was optimized for each frequency band separately. The nought symbol marks the frequency band with maximum performance, and crosses mark frequency bands that resulted in significantly lower performance than the best frequency band. **e, f.** R^2 (mean \pm standard error) as a function of SNR bits (**e**) and PSD of added noise (**f**). Horizontal bars above the plots show significant differences compared to best R^2 (two-sided Wilcoxon rank-sum test, $P < 0.05$); the bar colours correspond to the respective frequency bands (see key in **e**). Vertical dashed lines indicate the minimum SNR bits and maximum PSD values for which performance starts to degrade. For the other monkeys and the human participant, see Supplementary Fig. 6.

these capabilities contribute to the power budget and total size of the system. Next, we focus on the recording unit’s design parameters, which also affect the transmitter power consumption. The transmitter-specific design parameters will be discussed later.

Custom circuit parameters for a movement iBCI. To investigate the range of circuit parameters that do not compromise performance in the iBCI application domain, we estimated the velocity prediction quality from a neural signal degraded by relaxing the specifications of several specific circuit components (see Fig. 3a–c and Methods section ‘Neural interface parameter simulation’).

First, we varied the observation frequency band, f_0 and f_1 , by sweeping the lower and upper cut-off frequencies of a second-order Butterworth bandpass filter used to filter the raw signal (see Methods). To relax the filter roll-off requirements and avoid aliasing issues, the sampling rate was set to $f_s = 3f_1$, allocating a slight oversampling above the Nyquist limit. Figure 3d shows that decoding performance is kept consistent in a wide range of frequency bands, and that competitive decoding accuracy is also achieved at lower and narrower frequency bands, 0.5 kHz to 3 kHz. These results are consistent with previous studies examining finger position decoding³¹.

Second, we investigated the effect of the quantizer resolution on performance. Figure 3e shows that the decoder performance is fairly independent of ADC resolution for SNR bits ≥ 7 (see horizontal bars in the top left corner). This minimum resolution is limited by the precision requirements on the threshold values used for spike detection (see Methods), rather than by the signal resolution. Thresholding in the analogue domain would allow for 1-bit quantization, but Gibson and colleagues^{78,79} have showed that low-resolution digital spike detection is more efficient than its analogue counterpart.

Third, we investigated the effect of raw signal noise on decoding performance. In a neural interface, there are four main sources of noise: tissue thermal noise; electrode impedance thermal noise; noise from the interface electronics; and neuronal background electrical activity that might carry information from the surrounding neurons, but which for the purpose of spike detection we also consider as noise (Fig. 1a). Very low-noise interfaces (such as $\sqrt{V_{\text{RMS}}^2} = 1\text{--}5 \mu\text{V}_{\text{RMS}}$ ^{20,23,31} over a >5 kHz band of observation) are usually designed to avoid being the main source of noise and to provide measurements limited only by the biological system⁸⁰. However, Fig. 3f shows that decoder performance is robust (across all users) to a larger added Gaussian noise with a power spectral density

Table 1 | Neural interface specifications for a conventional commercialized⁷³ system, academic prototypes^{20,22,23,41–45} and the suggested minimalistic design

Parameters	Blackrock ⁷³	Yin et al. ²³	NeuroPixel ⁴¹	Borton et al. ²²	HermesE ²⁰	Lopez et al. ⁴²	SiNAPS ⁴⁵	De Dorigo et al. ⁴³	WIneRS-8 ⁴⁴	Minimalistic design
\bar{v}_{in}^2 (10^{-12} V ²) ^a	9.0	7.8	25.0	74.0	4.8	41.0	56.3	110.3	9.0	–
PSD(10^{-15} V ² Hz ⁻¹) ^b	0.8 ^c	0.6 ^c	1.6 ^c	6.0 ^c	0.3 ^c	2.7 ^c	5.0 ^c	7.2 ^c	0.4 ^c	4.0^d
f_0 (Hz)	0.3	1	300	0.1	280	300	300	300	400	250
f_1 (kHz)	7.5	7.8	10	7.8	10	10	7.5	10	15	3
f_s (kS s ⁻¹)	30	20	30	20	31.25	30	25	20	50	9
B (bits)	16	12	10	12	10	10	12	10	10	8^e
R_{ADC} (kbits s ⁻¹) ^f	480	240	300	240	312.5	300	300	200	500	72
R_{TX} (kbit s ⁻¹) ^g	480	240	300	240	312.5	300	300	200	500	1^g

^a Reported total integrated noise referred to the input of the interface. ^b Thermal PSD referred to the input of the interface. ^c PSD was calculated from the reported total integrated noise, assuming a first-order filter roll-off and a thermal spectrum. $PSD = \bar{v}_{in}^2 / (\Delta f \pi / 2)$, where $\Delta f = f_1 - f_0$. ^d Total PSD is the original PSD in our recording system⁷³ plus the added PSD in our analysis. ^e SNR bits = 7 ^f Data rate is calculated per channel. ^g Transmitter data rate assumes on-chip thresholding for the minimalistic design. The parameters are the total integrated noise power (\bar{v}_{in}^2), the power spectral density (PSD), the lower and upper frequency corners of the bandpass analogue filter (f_0 and f_1), the sampling frequency (f_s), the number of bits of the ADC (B), the data rate of the ADC (R_{ADC}), and the data rate of the transmitter (R_{TX}). We propose in the column 'Minimalistic design' that parameters can be relaxed up to 2–4× with respect to the state-of-the-art (the R_{TX} can be further reduced thanks to on-chip thresholding).

($PSD = \bar{v}_{in}^2 / \Delta f$) of up to about 3.5×10^{-15} V² Hz⁻¹. Surprisingly, decoder performance does not depend strongly on the observation frequency $\Delta f = f_1 - f_0$, suggesting that the critical factor is to maintain low noise power in the action potential band (that is, around 1 kHz). The PSD of an amplifier scales inversely with its power consumption. Hence, the extra noise budget can be used to reduce the power consumption of the amplifier. Alternatively, larger noise contributions from variations in the electrode impedance in chronic implants can be accommodated. Other noise sources, like motion artefacts and electromagnetic interference, usually appear to the recording system as common-mode signals. Hence, the differential structure of the system is able to reject them quite efficiently.

These results suggest that neural interface designs can be relaxed substantially and that there is a wide space of parameters (f_0 , f_1 , B and \bar{v}_{in}^2) in which comparable iBCI performances can be achieved. Design decisions should be based on the neural interface architecture and the trade-offs between power consumption and complexity. For example, in a conventional design, the ADC power consumption is negligible compared to that of the amplifier. Thus, the amplifier parameters should be optimized first ($f = 0.5$ – 3 kHz and $PSD = 3.3 \times 10^{-15}$ V² Hz⁻¹) and the minimal number of bits that maintain performance with such an amplifier (SNR bits = 7) should then be determined. An ongoing research effort to reduce power consumption already exists, with groups trying to create custom neural interfaces (academic prototype systems)^{20,22,23,40–46,81,82} for a wide range of applications. Table 1 summarizes the system specifications for a conventional system (commercialized⁷³), selected academic prototypes^{20,22,23,41–45}, and a minimalistic iBCI-focused design based on our results. Here we compare against systems designed for in vivo (fully implanted) neural interfaces like the one we characterized in this study^{20,22,23,41–45,73}. It can be noted from Table 1 how neural interfaces have been evolving towards more efficient designs with respect to commercially available systems (that is, amplification and quantization are optimized for maximizing the SNR at the output of the interface without compromising power efficiency). However, if specifications were derived based on iBCI decoding performance, further optimization could be achieved. This presents an opportunity to design custom neural interfaces for iBCI applications based on existing architectures that are power efficient and still support high channel count and wireless communication. Next, we will suggest the power savings that could be achieved by such

specification relaxation as compared to designs intended for neuroscientific measurements.

Neural interface power consumption

The previous results present a wide range of acceptable neural interface designs for use as part of an iBCI. In what follows, we analyse how much power could be saved in each system component by allowing more noise over a reduced bandwidth of interest. The following analysis is based on theoretical limits and/or extrapolations from published work. An important next step is to build a hardware proof-of-concept of the proposed system.

Neural amplifier. To estimate the amplifier power consumption we used the power efficiency factor (PEF) metric (the Muller⁸³ extension to the Steyaert⁸⁴ noise efficiency factor NEF):

$$PEF = NEF^2 V_{DD} = \frac{\bar{v}_{in}^2}{\Delta f} \frac{I_{tot}}{2\pi U_T kT} V_{DD} \quad (1)$$

where V_{DD} is the supply voltage, \bar{v}_{in}^2 is the total integrated noise contributed by the recording electronics (referred to the input of the amplifier), $\Delta f = (f_1 - f_0)$ is the bandpass filter bandwidth, I_{tot} is the total bias current, $U_T = kt/q$ is the thermal voltage, k is the Boltzmann constant, T is the temperature and q is the charge on the electron. The NEF describes how many times higher the noise of an amplifier is than the ideal case of a bipolar junction transistor, operating with the same bias current, and the PEF is used to compare solutions working at different supply voltages. State-of-the-art neural amplifiers^{30,81,85–87} usually have a power budget ($P_A = I_{tot} V_{DD}$) in the range 0.5–10 μ W per channel and a PEF in the range 1–30 V. For more details on the amplifier power consumption, see Methods.

Figure 4a shows the power consumption of an amplifier as a function of the input-referred PSD. $PEF = 1.12$ V from ref. ³⁰ was used here to calculate the power consumption. Given a required PSD, circuit topologies and supply voltage are the only degrees of freedom to reduce power consumption, because the bias current is set by the noise requirements.

ADC. The conversion power for ADCs scales linearly with the sampling rate in designs that are not limited by the transit frequency of the technology. Estimating the conversion energy as a function of SNR bits, however, is a more complex task. A model

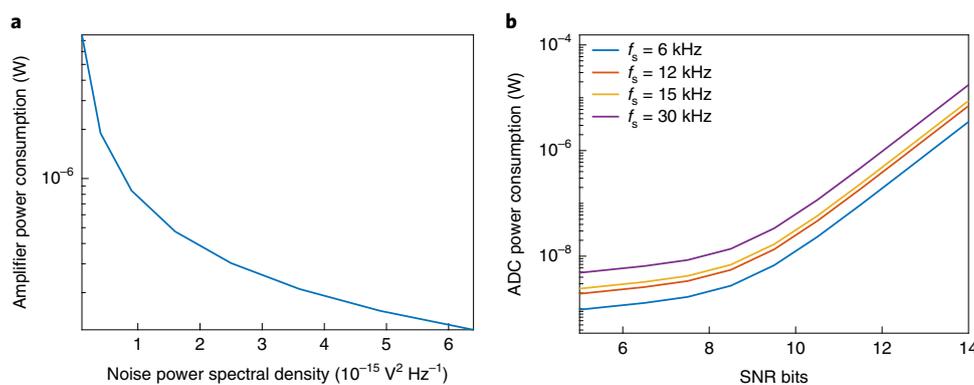


Fig. 4 | Power consumption trends in neural amplifiers and ADCs. a, Power consumption as a function of input-referred noise PSD for a neural amplifier. See equation (1) and PEF = 1.12 V used for this plot³⁰. **b**, Power consumption as a function of SNR bits for an ADC for different sampling frequencies f_s . See Methods for the model used for this plot.

for the conversion energy is used here and described in Methods. This model considers the minimum power consumption of a proxy successive approximation register ADC⁸⁸, and it provides a realistic estimate of the power savings as a function of SNR bits. The model does not take nonlinearities into account, because they do not noticeably affect the performance of the spike detector. Figure 4b shows the power consumption of an ADC as a function of the SNR bits, for different f_s . Reducing the number of SNR bits gives a large benefit in power consumption for medium to high resolutions (SNR bits > 8). However, for low-resolution ADCs, this benefit becomes less dominant and power consumption is dominated by secondary effects (see Methods).

Wireless transmitter. The design suggested here would take advantage of the reduced requirements, both in terms of single channel data rate and BER, to allow for a larger number of channels transmitting simultaneously. Current systems implement simple modulation schemes such as ‘on-off keying’ or ‘frequency-shift keying’ to reduce the complexity at the transmitter^{89,90}, while still achieving a BER lower than 10⁻⁴. The efficiency of the transmitter, E_b , is defined in terms of energy per bit and accounts for both the dynamic and static power consumption needed to transmit a single bit of information. At high data rates, the static power consumption is negligible compared to the dynamic power consumption and the overall efficiency ranges from a few picojoules per bit to a few tens of picojoules per bit^{75,91}. For our analysis, we used the achieved 8.5 pJ per bit value from ref. ⁹¹. Given the data rate R , the total power consumption becomes:

$$P_{\text{TX}} = E_b R \quad (2)$$

Transmitting threshold-crossing spike events instead of wide-band high-resolution signals will reduce the power consumption of the transmitter as long as the transmission efficiency is kept constant. The power saved can be allocated to integrating more channels (electrodes) into the device and maintaining the same data rate, which would yield the same efficiency (given the same design parameters such as coil size, transmission protocol, receiver distance, output power and so on). Compared to a conventional system for basic neuroscience (such as 16 bits at 30 kS s⁻¹ per channel) or academic prototype solutions (such as 10 bits at 20 kS s⁻¹ per channel), the proposed solution tailored for iBCIs transmitting only binary events (for example, 1 kbit s⁻¹ per channel) could enable an increase in the number of channels (electrodes) by 480× and 200×, respectively. Designing for a lower BER (10⁻³ and higher) could further improve the power consumption.

Low total power consumption for an iBCI. Figure 5 summarizes the results. Power consumption is plotted for each component (A, ADC and TX) of the systems in Table 1. The minimalistic design suggested here is analysed for two cases: if raw data are transmitted (‘Minimalistic design’) and if only threshold-crossing events are transmitted (‘Minimalistic design with thresholding, TH’). The transmitter assumes an efficiency $E_b = 8.5 \text{ pJ b}^{-1}$, and the amplifier assumes a PEF = 1.12 V. We note that the results presented here are obtained under the assumptions described above for power efficiency in each component and could vary greatly for different implementations and assumptions. The objective of Fig. 5 is to illustrate the relationship between system specifications and power consumption under reasonable assumptions.

Figure 5 shows how more recent academic systems^{20,22,23,41–45} have lowered power consumption because they use more relaxed quantization strategies than commercialized systems⁷³ and optimize noise specifications for the neural amplifier⁸⁰. Importantly, however, further power savings could be achieved when adopting the relaxed specifications we are proposing (‘Minimalistic design’). If on-chip thresholding is performed (‘Minimalistic design with thresholding’) and only the presence or absence of a threshold-crossing event is transmitted, the power consumption of the transmitter (which is the dominant power draw in most systems) can be drastically lowered with negligible extra cost in power and complexity. Such a system could reduce the total power consumption by two orders of magnitude compared to a commercialized system designed for basic neuroscience, or one order of magnitude compared to emerging research-tailored academic prototypes. These power savings come from loosening the parameters of all three components together. For example, transmitting spike events detected on chip with wide bandwidth (as in ref. ³²) would have reduced the power by only about twofold. In contrast to the conventional system, whose power consumption would be limited by the transmitter (50%–93%), the proposed design’s power consumption is limited by the neural amplifier (92%).

Discussion

Our results suggest that iBCIs with dedicated circuits designed for clinical use could consume an order of magnitude less power than an iBCI built with basic neuroscience-motivated specifications. The suggested minimalistic specifications are substantially relaxed without compromising decoding accuracy. This could give rise to high-electrode-count wireless iBCIs by reducing power and space requirements so that these circuits can support thousands of channels. Next, we identify areas where future research efforts in circuit-level and system-level solutions can further push

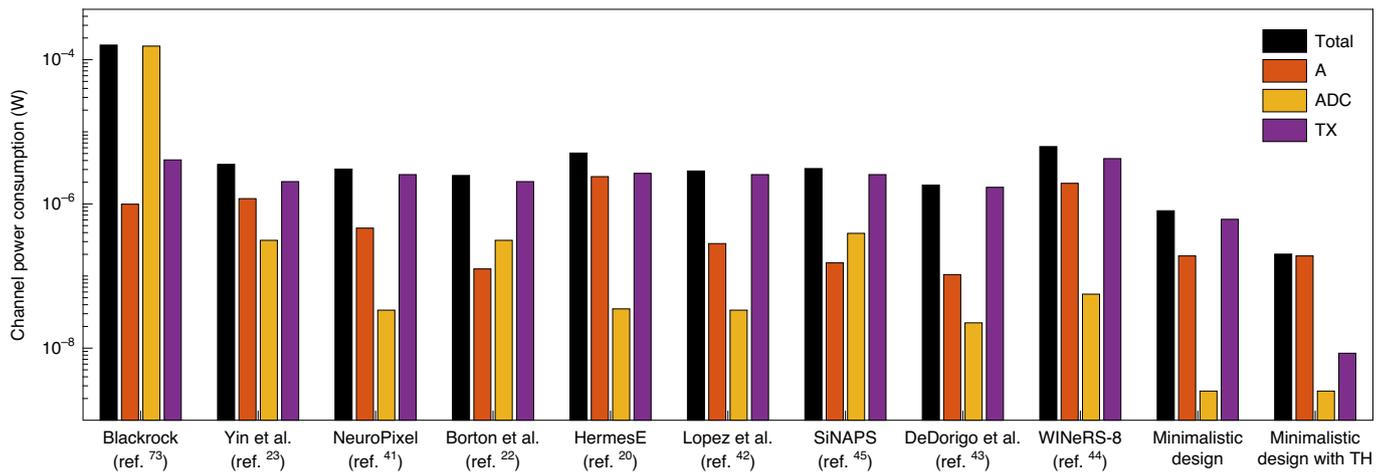


Fig. 5 | Power consumption estimates per channel (log scale) for systems described in Table 1^{20,22,23,41–45,73}. A minimalistic design implementing on-chip thresholding (TH) could reduce the total power consumption by at least one order of magnitude.

an exponential increase in the number of wirelessly transmitted recording channels.

Circuit-level opportunities. *Recording unit and transmitter.* Transmitting only threshold-crossing events does more than just relax the system specifications. It also opens an avenue for new recording system architectures for clinical iBCIs. For example, different spike detectors can be implemented, such as the nonlinear energy operator, which looks at the energy of the signal instead of the absolute value of the voltage trace^{92,93}. Such detectors might provide more robust protection against thermal noise and could further relax the amplifier and ADC specifications.

In our approach, the neural amplifier is the main power consumer. It therefore should be the focus of future studies of circuit topologies and tests of the robustness of iBCIs to noise. While we found that an iBCI decoder can be robust to added Gaussian noise of up to $3.5 \times 10^{-15} \text{ V}^2 \text{ Hz}^{-1}$, we were not able to isolate the different noise sources and calculate the amplifier input referred noise. This is because noise from different sources (biology, electrodes, electronics) is indivisible in the recordings. Future work should aim to isolate these sources and better characterize the iBCI decoder robustness specifically to noise from the neural amplifier. Also, continuous improvement in electrode design and manufacturing is likely to lead to better SNR and larger signals. This trend is likely to increase the robustness to added electronic noise in the system.

The total SER is related to the transmitter BER through the communication protocol. If the raw output of the threshold detector is transmitted ('1' for a spike and '0' otherwise, transmitted per channel at the bin rate), then the BER is equal to the SER, and an error in the transmitted data will result in either a false spike or a missed spike. More elaborate protocols that take into consideration the sparsity of the signal can reduce the transmitter data rate and obtain a different relationship between SER and the transmitter BER. For example, since the average rate at which action potentials arise is typically on the order of ten action potentials per second, only 1% of the bits will be 1 at 1 kS^{-1} (for example, $N/100$, where N is the number of electrodes). Transmitting only the index of the electrodes that have an action potential, which requires $\log(N)$ bits per index, will result in an average data rate of $\log(N) \times N \times 1\%$. In this protocol, a bit error will create an erroneous spike at an incorrect electrode index number, and a missed spike at the actual electrode index number, thereby doubling the SER. On the other hand, error detection and correction techniques at the receiver end could

help to alleviate the requirements on the BER, at the cost of minimal increase in the data rate requirements.

Device area. Another limited resource for an implantable neural interface is the chip area, but quantitative estimates for this area are difficult to obtain. Area depends on many factors, such as technology and architecture. However, reducing the noise requirements of an integrated circuit usually results in a smaller footprint. Smaller transconductance results in smaller active devices, and smaller sampling capacitors result in smaller passive devices. Increasing the high-pass pole of the filter might also reduce the neural amplifier total area. Future work should investigate the limitations of area consumption and find trade-offs that balance reduced area with acceptable performance.

System-level opportunities. *Dimensionality reduction on-chip.* As we mentioned earlier, dimensionality reduction techniques are still under active investigation and consensus on a single reduction strategy has not been reached. Nevertheless, to further reduce the system's data rate, one could implement dimensionality reduction directly on-chip. A common approach is to implement principal component analysis (PCA) before the decoder. However, it is not immediately clear how on-chip PCA would help reduce the overall system power consumption. The data rate reduction is not very large and the power overhead of the hardware implementation for PCA might actually result in an increase in power consumption. For example, for 1,000 channels the original data rate would be, after the threshold detector, $f_{\text{TX},1} = 1 \text{ Mbps}$ (assuming 1 bit at 1 kHz output). If 20 principal components with 8 bit resolution are transmitted after PCA analysis, the new data rate would be $f_{\text{TX},1} = 0.16 \text{ Mbps}$. The gains in reducing the data rate from 1 Mbps to 0.16 Mbps might not be enough to justify the computational cost of PCA and the storage cost of a projection matrix. This conclusion differs from previous work done on on-chip compression of neural signals⁸⁹, since the raw data here is a binary threshold-crossing signal and not a high-resolution signal used for spike sorting.

Currently, PCA is computed using floating-point resolution. To properly estimate the benefit of on-chip PCA, future work should study the resolution requirements of the principal components for iBCI decoders, that is, how many components to use and at what bit resolution to represent each component. This will allow a numerical analysis of the computation and storage cost of on-chip PCA, as well as the data reduction factor.

Real-time iBCI. Here, we estimated the effects of system parameters on iBCI performance using offline decoding analyses based on real movement and iBCI data. However, during real-time iBCI control, the user has continuous feedback about the decoder's performance (for example, by seeing how the cursor or robotic arm is moving). This allows the user to compensate for errors^{60,69,94}. Thus, the robustness to SER in real time is probably higher than in the offline analysis we presented here. This can be verified by future closed-loop studies in which various types of signal processing alterations (such as spike errors or different filtering or bit resolution) are made to the neural data during real-time iBCI control.

Number of required recording channels. A natural question is how many channels are needed for an iBCI? Until high electrode count devices exist ($\gg 1,000$ channels), it will be difficult to determine how many electrodes suffice for each iBCI application. There have been a few attempts to extrapolate the performance of iBCIs with increasing numbers of electrodes^{14,15,95}. However, extrapolations are challenging because they do not reliably predict how neural activity will change in more complex tasks⁹⁶ and how more advanced decoders might make use of this data. Schwarz and colleagues¹⁴ postulated that recording 5,000 to 10,000 neurons is necessary for an iBCI to restore limb movement, and that 100,000 neurons will be required to control whole-body movement. Once a high-electrode-count device exists, it might turn out to be advantageous to have more electrodes even at the expense of the reliability and accuracy of each electrode signal; this would provide opportunities to reduce per-channel power even more by further relaxing the specifications.

Implications beyond movement iBCIs. *Other types of brain-computer interface.* Here we have evaluated circuit specifications for neural interfaces for decoding movement intentions. Our results—that design specifications can be dramatically relaxed—may well apply to other types of neural interfaces, such as those used in retinal prostheses, peripheral nervous system interfaces, and so on. Our prediction is that neural interfaces that rely mainly on decoding spike activity will have similar recording system requirements (f_o , f_i and B), although the SER might change based on the redundancy of the signal and the robustness of the decoders. Applications where spike sorting is required (such as for artificial retinas) could also benefit from a design that uses a holistic approach to reduce power consumption at the interface, as in ref. ³⁴. Similar power calculation and parameter analyses might also be beneficial for wireless miniature microscopes to enable longer recording times (see, for example, ref. ⁹⁷).

Implications for basic neuroscience research. Although this work focuses on neural interface design for a clinical iBCI application, basic neuroscience research might also benefit from low-resolution, high-channel-count devices. Research-based studies traditionally require accurate single neuron recording, but in some scenarios they can benefit from trading off recording from more neurons for better per-neuron signal fidelity. For example, this may be the case when looking for population-level phenomena for which spike sorting is not necessary²⁹, or in studies where wireless recording is essential^{14,23,98}. In particular, this approach could enable longer-duration and wireless recording from model organisms that are too small to carry the bulky electronics needed for high-bandwidth recording and data transmission or storage.

Outlook

When developing a new device, an iterative process of design and user testing is essential. In multidisciplinary research, such as neuroscience, this process may take years or even decades because the design and the testing are sometimes done by separate entities (different research laboratories). A large body of neuroscience research gave rise to the iBCI field, which since its inception has used methods

and tools similar to those of neuroscience. Decades of iBCI research with monkeys and recent clinical trials with human participants have brought the field to a level of maturity and confidence about its neural interface requirements. Future iterations (further on-chip processing) are inevitable, but this study can be used as feedback on current neural interface designs, intended to guide dedicated designs for the next generation of iBCIs. We believe that this study, which arose from a collaborative effort between electrical engineers, neural engineers, clinicians and neuroscientists, indicates the path towards the next generation of clinically viable iBCIs.

Methods

Monkeys hand-movement data. All monkeys' procedures and experiments were approved by the Stanford University Institutional Animal Care and Use Committee. Three male rhesus macaques (monkeys J, R and L) were trained to perform point-to-point movements of a 6-mm-radius virtual cursor in a two-dimensional plane, while their other arm was gently restrained. The monkey performed centre-out-and-back task to 8 targets uniformly distributed on a 8-cm-radius circle (Supplementary Fig. S1). Two mirrors, set up as a Wheatstone stereograph, visually fused the monitors into a single three-dimensional percept for the monkeys, although all task-relevant motion was limited to two dimensions⁹⁹. In this work, about 100 continuous successful trials (about 2 min) from ten experiment session days (about 1,000 total trials) were recorded from each monkey and analysed.

Monkeys were implanted with two (monkeys J and R) or one (monkey L) 96-electrode Utah arrays (Blackrock Microsystems, Inc.), using standard neurosurgical techniques 95 (J), 75 (R) and 91 (L) months prior to the recorded sessions. The arrays contained a 10×10 grid of 1 mm microelectrodes with 400- μ m centre-to-centre spacing between adjacent electrodes. The arrays of J and R were implanted into the left cortical hemisphere; one array went into the primary motor cortex (M1) and the other into the dorsal premotor cortex (PMd). In this study we used only the PMd array of monkey R, because his M1 array was severely degraded and recorded almost no large waveform action potentials. Ls single array was implanted into the right-hemisphere boundary between M1 and PMd.

Human participant iBCI cursor movement data. Permission for these studies was granted by the US Food and Drug Administration (Investigational Device Exemption) and Institutional Review Boards of Stanford University (protocol 20804), Partners Healthcare / Massachusetts General Hospital (2011P001036), Providence VA Medical Center (2011-009), and Brown University (0809992560). The participant in this study, 'T5', was enrolled in a pilot clinical trial of the BrainGate2 Neural Interface System (<http://www.clinicaltrials.gov/ct2/show/NCT00912041>). Informed consent, including consent to publish, was obtained from the participant prior to his enrollment in the study.

Participant T5 is a right-handed man, 63 years old at the time of the study, whose iBCI cursor control research sessions were previously described in refs. ^{8,62}. T5 was diagnosed with a C4 AIS-C spinal cord injury approximately nine years prior to study enrollment. In August 2016, participant T5 had two 96-channel intracortical silicon microelectrode arrays (1.5 mm electrode length, Blackrock Microsystems) implanted in the arm-hand area of the dominant (left) motor cortex.

T5 also performed ten research sessions from which we analysed 2-min durations of a cursor movement task. In his task, a grid spanning $1,000 \times 1,000$ pixels on the computer monitor was divided into a 6×6 or 9×9 grid of equally sized grey squares. Each square was a selectable target, and on each trial, one square would randomly be prompted as the correct target by changing its colour to green. The participant had to select the correct target (which resulted in a trial success) while avoiding selecting any of the other, incorrect targets (which would result in a trial failure).

T5 controlled the computer cursor using an iBCI. In his sessions, neural control and task cuing were controlled by custom software running on the Simulink/xPC real-time platform (The Mathworks), enabling millisecond-timing precision for all computations. Neural data were collected by the NeuroPort System (Blackrock Microsystems) and available to the real-time system with 5-ms latency.

Two-dimensional continuous control of the cursor was enabled by the ReFIT Kalman Filter detailed in refs. ^{8,11}. T5 could select a target by dwelling on it for 1 s or by a discrete 'click' signal. Discrete selection (a click) was achieved using a Hidden Markov Model-based state classifier. The user commanded a click by attempting to squeeze his left hand (that is, the hand ipsilateral to the arrays). For both the continuous cursor-positioning ReFIT Kalman filter decoder and the discrete click-state Hidden Markov Model decoder, spiking activity was binned every 15 ms and sent through the decoders. Since the executed kinematics were an output of a Kalman-filter-based decoder, they were more temporally structured than the monkeys' hand kinematics.

Neural and hand-position recording. We used Blackrock Microsystems neural acquisition systems during both monkey (Cerebus system) and human (NeuroPort

system) research sessions. Both data-acquisition systems achieve $3 \mu\text{V}_{\text{RMS}}$ of input-referred noise over a bandwidth of $[0.3\text{--}7,500]$ Hz, and sample each electrode with 16 bits at 30 kSps. We refer to the system output signal as our raw signal. Nonactive electrodes with zero firing rates were removed from the analysis. During the session, the monkey's contralateral hand position was measured for decoder training and hand kinematics were analysed using an infrared reflective bead tracking system (Polaris, Northern Digital) polling at 60 Hz. Hand velocity was computed from the recorded position of the bead, which was taped to the monkey's reaching hand.

Offline decoders. We used a Kalman filter to estimate the two-dimensional hand velocity ($v_t \in R^2$) from the spike events ($y_t \in \{0, 1\}^N$, where y is the number of spike events, t is time and N is number of electrodes), $v_t = f(y_t)$. In all analyses, decoders were tenfold cross-validated on each day (total of 100 decoders per user) and their quality was measured with R^2 compared to the true hand (monkeys) or cursor (human) velocity. With the monkey datasets, we estimated the velocity of their native hand while they conducted a centre-out reaching task. As T5 could not move his hands, we offline estimated the cursor velocity. During T5's sessions, the cursor was controlled by the iBCI system, which used a ReFIT Kalman filter algorithm to decode his intention in real time. We note that for both the monkey and human movements, the key question was the same: did decoding the same neural data, subject to varying additional spike errors, output similar decoded kinematics?

The same analysis performed for the Kalman filter decoder was extended here to a linear regression decoder to estimate cursor velocity, and its performance was similarly measured using R^2 . Also, two different discrete iBCI decoders were analysed using a naive Bayes and a support vector machine classifier. Each iBCI discrete decoder's task was to detect the intended target of the user on the basis of the neural activity at the beginning of each trial. Similar to the optimal parameters found in ref.¹⁰⁰, we used 64 ms of neural activity starting from 160 ms after target onset. We used classification accuracy to measure the classifiers' performance. For all decoders and classifiers, we used tenfold cross-validation for each day, as described above. Supplementary Fig. S3 summarizes the results for different decoders. Conclusions based on discrete decoding regarding SER robustness are consistent with those drawn for the Kalman filter decoder.

SER simulation. To extract neural spiking activity using the Blackrock system, a 250-Hz high-pass filter was applied to the raw signal. Then, a spike was detected whenever the voltage crossed below a threshold set at $-4.5 \times \text{RMS}$ voltage. This threshold value was updated every session. The spike detector used a window of 1 ms, and multiple spikes were accumulated in non-overlapping 60-Hz bins in order to align to hand-velocity recordings. Spike errors were simulated by independently, randomly flipping the binary signal samples y_t with a Bernoulli distribution with probability equal to the tested SER (Fig. 2a). For example, if the error rate was 10^{-2} , each sample (bits) of the signal was independently, randomly flipped with a Bernoulli distribution with a probability of $P = 10^{-2}$.

Neural interface parameter simulation. To simulate a neural interface with a set of new parameters, we degraded the raw signals (with a recording system output signal of 16 bits at 30 kSps) with a series of manipulations as described in the main text and elaborated below. Then, we detected the spikes from this manipulated raw signal and estimated the hand velocity using the decoders as described earlier (Supplementary Fig. S2). We note that the spikes were detected with a threshold (see Methods subsection 'Threshold-crossing detection') and no spike sorting was performed. Future work could explore the effect of spike sorting on whatever region of the parameter space would still result in comparable performance. Here, the raw signal emulates a continuous-time signal to be processed by our recording system. This is equivalent to a real continuous-time analysis, since the sampling rate of our recording system is well below the sampling rate of the raw signal.

Input-referred noise. The input-referred noise represents the total noise introduced by the circuit via a fictitious input source that captures all circuit-internal noise sources. Noise is usually referred (scaled) to the input so it can be readily compared to the input signal level. To simulate higher input-referred noise ($\bar{v}_{\text{in,RMS}}$), we added Gaussian noise with variance of σ_{noise}^2 to the raw signal.

Bandpass. To simulate the amplifier filtering function, we filtered the signal between f_0 and f_1 (the cutoff frequencies of the filter) with a second-order Butterworth filter.

Sampling. The output of the bandpass filter was sampled at $f_s = 3f_1$, to allocate some oversampling above the Nyquist limit and to relax the bandpass filter performances. This choice is commonly adopted to relax the filter roll-off requirements and avoid aliasing issues. Optimizing the ratio f_0/f_1 can be investigated in future work.

Quantization. The sampled signal was re-quantized at B bits.

Threshold-crossing detection. Threshold detection was applied every millisecond (in a causal 30-sample window) to detect the presence of a putative neural spike. The threshold was set to be proportional to the estimated raw signal RMS for each electrode (e)

$$V_{\text{threshold}}^e = n \times V_{\text{RMS}}^e$$

The number of RMS (n) was optimized by sweeping the range of -6 to -1 using increments of 0.5 for all of the electrodes, for each set of parameters; see

Supplementary Fig. S7 for best n distribution across parameter sets. The Blackrock system's built-in function was used for calculating the RMS voltage of the noise, which is slightly different from the standard RMS calculation²⁷. Specifically, the BlackRock algorithm calculates a biased estimate of the RMS with an aim to exclude spikes and artefacts that inflate the RMS. First, the algorithm computes mean squares of each of 100 non-overlapping bins (x_j) of 600 continuous samples (20 ms) of the raw data (s_i), with a total of 60,000 samples (2 s):

$$x_j = \frac{1}{600} \sum_{i=1}^{600} s_i^2, \quad 1 \leq j \leq 100$$

Then, the RMS is calculated by averaging the 6th until the 25th lowest x_j values (20 out of 100 values):

$$\text{RMS} = \sqrt{\frac{1}{20} \sum_{i=6}^{25} \min_i x}$$

where $\min_i x$ is the i th minimum value of x .

Spike binning. The resulting binary signal was binned in 60-Hz non-overlapping bins aligned to velocity recordings.

Statistical testing. When comparing two different distributions of R^2 , we used a two-sided Wilcoxon rank-sum test with a confidence level of $P = 0.05$ unless stated otherwise.

Neural amplifier noise and power. To gain insight into the amplifier power consumption and its lower limits, let us consider the input-referred thermal noise contribution of a single MOS transistor:

$$\bar{v}_{\text{in}}^2 = \frac{4kT\gamma}{g_m} \quad (3)$$

where γ is a technology-dependent noise factor here approximated to 0.8 (refs.^{101,102}) and g_m is the device transconductance. g_m can be linked to the power consumption through the device transconductance efficiency, g_m/I_D , where I_D is the transistor bias current. For a device working in the sub-threshold region, g_m/I_D can be as large as 30 S/A (the upper limit is $q/kT \approx 38$ S/A for bipolar transistors (where S/A is siemens per ampere)). In these conditions, the power consumed (P_D) to obtain an input referred noise of $5 \mu\text{V}_{\text{RMS}}$ over a bandwidth $\Delta f = 10$ kHz (assuming a supply voltage of 1 V) is:

$$P_D = \frac{1}{(5 \mu\text{V}_{\text{RMS}})^2} \frac{4kT\gamma}{g_m/I_D} \Delta f V_{\text{DD}} = 177 \text{ nW} \quad (4)$$

Although this result might look promising, this refers to a lower limit for biasing a single device that ensures enough thermal noise margin for a complete system. In reality, a differential readout is usually implemented to deal with common-mode noise injections, chopping or correlated double sampling is used to attenuate flicker noise, and multiple gain stages are used for better conditioning of the signal before the ADC. As a result, practical implementations will consume more power than described in equation (4).

Steyaert et al.⁸⁶ proposed a metric for comparing the noise performance of amplifiers, the NEF:

$$\text{NEF} = v_{\text{in,RMS}} \sqrt{\frac{2I_{\text{tot}}}{\Delta f \pi U_T 4kT}} \quad (5)$$

where I_{tot} is the total bias current, $U_T = kT/q$ is the thermal voltage, k is the Boltzmann constant, T is the temperature and q is the charge on the electron. The NEF describes how many times higher the noise of an amplifier is than in the ideal case of a bipolar junction transistor, operating with the same bias current (the NEF is defined for a first-order filter with an effective noise bandwidth of $\Delta f(\pi/2)$, hence the scaling factor of $\pi/2$ in equation (5)). However, the NEF lacks the ability to compare solutions operating at different supply voltages. To overcome this problem, Muller⁸³ introduced the power efficiency factor ($\text{PEF} = \text{NEF}^2 V_{\text{DD}}$), which takes into account both the operating current and the supply voltage.

$$\text{PEF} = \bar{v}_{\text{in}}^2 \frac{2P_A}{\Delta f \pi U_T 4kT} \quad (6)$$

where P_A is the total power consumption of the amplifier. State-of-the-art neural amplifiers^{81,85–87} usually result in a power budget ($P_A = I_{\text{tot}} V_{\text{DD}}$) in the range 0.5–10 μW and a PEF value in the 1–30 V range.

ADC power model. To study the effect of resolution on the ADC energy, we considered a model of a successive approximation register ADC⁸⁸, as shown in Supplementary Fig. S4. The model assumes that the three main sources of power consumption are the capacitive DAC, the comparator and the logic. Also, it assumes that the comparator, the sampling capacitor and the quantization process each

contribute a third of the total noise. From these assumptions, we can derive the minimum energy required to respect the SNR specifications for each component.

Capacitive DAC. The input is sampled by the capacitive DAC, hence the minimum capacitance must satisfy:

$$\text{SNR} = \frac{\frac{1}{2} \left(\frac{V_{\text{in,pp}}^2}{2} \right)^2}{3 \frac{kT}{C_{\text{DAC}}}} \quad (7)$$

where $V_{\text{in,pp}}$ is the peak-to-peak input voltage, k is the Boltzmann constant, T is the temperature, $C_{\text{DAC}} = 2^B C_U$ is the total DAC capacitance, and C_U is the unit capacitance. As a result, the minimum unit capacitance becomes:

$$C_U = -\frac{24kT \text{SNR}}{2^B V_{\text{in,pp}}^2} + C_{U,\text{min}} \quad (8)$$

where $C_{U,\text{min}}$ is the minimum realizable capacitance allowed by the technology (usually 0.1–1 fF). The total energy depends on the switching activity of the DAC. The solution in ref. ¹⁰³ grants:

$$E_{\text{DAC}} = \sum_{i=1}^{B-1} 2^{B-3-2i} (2^i - 1) C_U V_{\text{REF}}^2 \quad (9)$$

where B is the number of bits of the ADC and V_{REF} is the reference voltage.

Comparator. Here, a simple latch model is used for the comparator and the noise is simplified to $\frac{kT}{C_C}$, where C_C is the load capacitance of the latch. For a more complete analysis, see ref. ¹⁰⁴. Similar to the capacitive DAC, the minimum load capacitance for the latch can be derived as:

$$C_C = -\frac{24kT \text{SNR}}{V_{\text{in,pp}}^2} + C_{C,\text{min}} \quad (10)$$

where $C_{C,\text{min}}$ is the minimum load capacitance available (usually [1–10] fF). The total energy then becomes

$$E_{\text{comp}} = (C_C V_{\text{DD}}^2) B \quad (11)$$

where V_{DD} is the supply voltage of the comparator.

Logic. For simplicity, we assume that the logic complexity depends linearly on the number of bits. The total energy then becomes:

$$E_{\text{logic}} = N_B E_G B \quad (12)$$

where N_B is the number of gates required per bit, and E_G is the energy per gate.

The total energy, and contributions from each block are plotted in Supplementary Fig. S4 as a function of SNR. Reference points from the literature are also plotted in figure, ^{20,82,105,106}. For low SNR, the conversion energy is dominated by the logic. For high SNR, the conversion energy is dominated by the comparator and increases 4× per bit (ref. ¹⁰⁷).

Here, we assume that power is a linear function of f_s , which is a realistic assumption for sampling frequencies well below the transit frequency of the technology. Hence:

$$P_{\text{ADC}} = f_s (E_{\text{DAC}} + E_{\text{comp}} + E_{\text{logic}}) \quad (13)$$

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The main data supporting the results in this study are available within the paper and its Supplementary Information. The raw human neural data are available on request to K.V.S. or J.M.H., yet owing to the potential sensitivity of the data and to respect the participant's expectation of privacy, an agreement between the researcher's institution and the BrainGate consortium is required to facilitate the sharing of these datasets. Processed data are available at <https://shenoy.people.stanford.edu/data>.

Code availability

The custom code used to produce the figures is available at <https://shenoy.people.stanford.edu/data>.

Received: 2 November 2018; Accepted: 30 June 2020;

Published online: 03 August 2020

References

- Slutzky, M. W. Brain-machine interfaces: powerful tools for clinical treatment and neuroscientific investigations. *Neuroscientist* **25**, 139–154 (2019).
- Bouton, C. E. et al. Restoring cortical control of functional movement in a human with quadriplegia. *Nature* **533**, 247–250 (2016).
- Ajiboye, A. B. et al. Restoration of reaching and grasping movements through brain-controlled muscle stimulation in a person with tetraplegia: a proof-of-concept demonstration. *Lancet* **389**, 1821–1830 (2017).
- Hochberg, L. R. et al. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature* **485**, 372–375 (2012).
- Collinger, J. L. et al. Functional priorities, assistive technology, and brain-computer interfaces after spinal cord injury. *J. Rehabil. Res. Dev.* **50**, 145–160 (2013).
- Downey, J. E. et al. Blending of brain-machine interface and vision-guided autonomous robotics improves neuroprosthetic arm performance during grasping. *J. Neuroeng. Rehabil.* **13**, 28 (2016).
- Hochberg, L. R. et al. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature* **442**, 164–171 (2006).
- Pandarinath, C. et al. High performance communication by people with paralysis using an intracortical brain-computer interface. *eLife* **6**, 18554 (2017).
- Jarosiewicz, B. et al. Virtual typing by people with tetraplegia using a self-calibrating intracortical brain-computer interface. *Sci. Transl. Med.* **7**, 313ra179 (2015).
- Blabe, C. H. et al. Assessment of brain-machine interfaces from the perspective of people with paralysis. *J. Neural Eng.* **12**, 043002 (2015).
- Gilja, V. et al. Clinical translation of a high-performance neural prosthesis. *Nat. Med.* **21**, 1142–1145 (2015).
- Homer, M. L., Nurmikko, A. V., Donoghue, J. P. & Hochberg, L. R. Sensors and decoding for intracortical brain computer interfaces. *Annu. Rev. Biomed. Eng.* **15**, 383–405 (2013).
- Lebedev, M. A. & Nicolelis, M. A. L. Brain-machine interfaces: from basic science to neuroprostheses and neurorehabilitation. *Physiol. Rev.* **97**, 767–837 (2017).
- Schwarz, D. A. et al. Chronic, wireless recordings of large-scale brain activity in freely moving rhesus monkeys. *Nat. Methods* **11**, 670–676 (2014).
- Carmena, J. M. et al. Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biol.* **1**, e42 (2003).
- Gao, P. & Ganguli, S. On simplicity and complexity in the brave new world of large-scale neuroscience. *Curr. Opin. Neurobiol.* **32**, 148–155 (2015).
- Wodlinger, B. et al. Ten-dimensional anthropomorphic arm control in a human brain-machine interface: difficulties, solutions, and limitations. *J. Neural Eng.* **12**, 016011 (2015).
- Mitz, A. R. et al. High channel count single-unit recordings from nonhuman primate frontal cortex. *J. Neurosci. Methods* **289**, 39–47 (2017).
- Chen, X. et al. 3D printing and modelling of customized implants and surgical guides for non-human primates. *J. Neurosci. Methods* **286**, 38–55 (2017).
- Gao, H. et al. HermesE: a 96-channel full data rate direct neural interface in 0.13 μm CMOS. *IEEE J. Solid-State Circuits* **47**, 1043–1055 (2012).
- Miranda, H., Gilja, V., Chestek, C. A., Shenoy, K. V. & Meng, T. H. HermesD: a high-rate long-range wireless transmission system for simultaneous multichannel neural recording applications. *IEEE Trans. Biomed. Circuits Syst.* **4**, 181–191 (2010).
- Borton, D. A., Yin, M., Aceros, J. & Nurmikko, A. An implantable wireless neural interface for recording cortical circuit dynamics in moving primates. *J. Neural Eng.* **10**, 026010 (2013).
- Yin, M. et al. Wireless neurosensor for full-spectrum electrophysiology recordings during free behavior. *Neuron* **84**, 1170–1182 (2014).
- Kao, J. C., Stavisky, S. D., Sussillo, D., Nuyujukian, P. & Shenoy, K. V. Information systems opportunities in brain-machine interface decoders. *Proc. IEEE* **102**, 666–682 (2014).
- Fraser, G. W., Chase, S. M., Whitford, A. & Schwartz, A. B. Control of a brain-computer interface without spike sorting. *J. Neural Eng.* **6**, 055004 (2009).
- Perel, S. et al. Single-unit activity, threshold crossings, and local field potentials in motor cortex differentially encode reach kinematics. *J. Neurophysiol.* **114**, 1500–1512 (2015).
- Christie, B. P. et al. Comparison of spike sorting and thresholding of voltage waveforms for intracortical brain-machine interface performance. *J. Neural Eng.* **12**, 016009 (2015).
- Li, J. & Li, Z. Sums of spike waveform features for motor decoding. *Front. Neurosci.* **11**, 406 (2017).
- Trautmann, E. M. et al. Accurate estimation of neural population dynamics without spike sorting. *Neuron* **103**, 292–308 (2019).
- Han, D., Zheng, Y., Rajkumar, R., Dawe, G. & Je, M. A 0.45 v 100-channel neural-recording IC with sub-μW/channel consumption in 0.18 μm CMOS. In *IEEE Int. Solid-State Circuits Conf.* 291–292 (IEEE, 2013).
- Irwin, Z. T. et al. Enabling low-power, multi-modal neural interfaces through a common, low-bandwidth feature space. *IEEE Trans. Neural Syst. Rehabil. Eng.* **24**, 521–531 (2016).
- Sodagar, A. M., Wise, K. D. & Najafi, K. A fully integrated mixed-signal neural processor for implantable multichannel cortical recording. *IEEE Trans. Biomed. Eng.* **54**, 1075–1088 (2007).

33. Karkare, V., Gibson, S. & Marković, D. A 75- μ w, 16-channel neural spike-sorting processor with unsupervised clustering. *IEEE J. Solid-State Circuits* **48**, 2230–2238 (2013).
34. Muratore, D. G. et al. A data-compressive wired-OR readout for massively parallel neural recording. *IEEE Trans. Biomed. Circuits Syst.* **13**, 1128–1140 (2019).
35. Aprile, C. et al. Adaptive learning-based compressive sampling for low-power wireless implants. *IEEE Trans. Circuits Syst. I* **65**, 3929–3941 (2018).
36. Pagin, M. & Ortmanns, M. A neural data lossless compression scheme based on spatial and temporal prediction. In *IEEE Biomedical Circuits and Systems Conf.* 1–4 (IEEE, 2017).
37. Wu, T., Zhao, W., Keefer, E. & Yang, Z. Deep compressive autoencoder for action potential compression in large-scale neural recording. *J. Neural Eng.* **15**, 066019 (2018).
38. Okazawa, T. & Akita, I. A time-domain analog spatial compressed sensing encoder for multi-channel neural recording. *Sensors* **18**, 184 (2018).
39. Shoaran, M., Lopez, M. M., Pasupureddi, V. S. R., Leblebici, Y. & Schmid, A. A low-power area-efficient compressive sensing approach for multi-channel neural recording. In *IEEE Int. Symp. on Circuits and Systems* 2191–2194 (IEEE, 2013).
40. Musk, E. Neuralink, an integrated brain-machine interface platform with thousands of channels. *J. Med. Internet. Res.* **21**, e16194 (2019).
41. Jun, J. J. et al. Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551**, 232–236 (2017).
42. Lopez, C. M. et al. A neural probe with up to 966 electrodes and up to 384 configurable channels in 0.13 μ m SOI CMOS. *IEEE Trans. Biomed. Circuits Syst.* **11**, 510–522 (2017).
43. De Dorigo, D. et al. Fully immersible subcortical neural probes with modular architecture and a Delta-Sigma ADC integrated under each electrode for parallel readout of 144 recording sites. *IEEE J. Solid-State Circuits* **53**, 3111–3125 (2018).
44. Lee, B. et al. An inductively-powered wireless neural recording and stimulation system for freely-behaving animals. *IEEE Trans. Biomed. Circuits Syst.* **13**, 413–424 (2019).
45. Angotzi, G. N. et al. SiNAPS: an implantable active pixel sensor CMOS-probe for simultaneous large-scale neural recordings. *Biosens. Bioelectron.* **126**, 355–364 (2019).
46. Fiàth, R. et al. A silicon-based neural probe with densely-packed low-impedance titanium nitride microelectrodes for ultrahigh-resolution in vivo recordings. *Biosens. Bioelectron.* **106**, 86–92 (2018).
47. Einevoll, G. T., Kayser, C., Logothetis, N. K. & Panzeri, S. Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nat. Rev. Neurosci.* **14**, 770–785 (2013).
48. Belitski, A. et al. Low-frequency local field potentials and spikes in primary visual cortex convey independent visual information. *J. Neurosci.* **28**, 5696–5709 (2008).
49. Buzsáki, G., Anastassiou, C. A. & Koch, C. The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nat. Rev. Neurosci.* **13**, 407–420 (2012).
50. Chestek, C. A. et al. Long-term stability of neural prosthetic control signals from silicon cortical arrays in rhesus macaque motor cortex. *J. Neural Eng.* **8**, 045005 (2011).
51. Todorova, S., Sadtler, P., Batista, A., Chase, S. & Ventura, V. To sort or not to sort: the impact of spike-sorting on neural decoding performance. *J. Neural Eng.* **11**, 056005 (2014).
52. Kao, J. C., Nuyujukian, P., Ryu, S. I. & Shenoy, K. V. A high-performance neural prosthesis incorporating discrete state selection with hidden Markov models. *IEEE Trans. Biomed. Eng.* **64**, 935–945 (2017).
53. Shanechi, M. M. et al. Rapid control and feedback rates enhance neuroprosthetic control. *Nat. Commun.* **8**, 13825 (2017).
54. Even-Chen, N., Stavisky, S. D., Kao, J. C., Ryu, S. I. & Shenoy, K. V. Augmenting intracortical brain-machine interface with neurally driven error detectors. *J. Neural Eng.* **14**, 066007 (2017).
55. Collinger, J. L. et al. High-performance neuroprosthetic control by an individual with tetraplegia. *Lancet* **381**, 557–564 (2013).
56. Mueller, K. et al. Autonomy infused teleoperation with application to BCI manipulation. *Auton. Robots* **41**, 1401–1422 (2017).
57. Gilja, V. et al. A high-performance neural prosthesis enabled by control algorithm design. *Nat. Neurosci.* **15**, 1752–1757 (2012).
58. Katyal, K. D. et al. A collaborative BCI approach to autonomous control of a prosthetic limb system. In *2014 IEEE Int. Conf. on Systems, Man, and Cybernetics* 1479–1482 (IEEE, 2014).
59. Oby, E. R. et al. Extracellular voltage threshold settings can be tuned for optimal encoding of movement and stimulus parameters. *J. Neural Eng.* **13**, 036009 (2016).
60. Stavisky, S. D., Kao, J. C., Nuyujukian, P., Ryu, S. I. & Shenoy, K. V. A high performing brain-machine interface driven by low-frequency local field potentials alone and together with spikes. *J. Neural Eng.* **12**, 036009 (2015).
61. Flint, R. D., Lindberg, E. W., Jordan, L. R., Miller, L. E. & Slutzky, M. W. Accurate decoding of reaching movements from field potentials in the absence of spikes. *J. Neural Eng.* **9**, 046006 (2012).
62. Even-Chen, N. et al. Feasibility of automatic error detect-and-undo system in human intracortical brain-computer interfaces. *IEEE Trans. Biomed. Eng.* **65**, 1771–1784 (2018).
63. Brandman, D. M. et al. Robust closed-loop control of a cursor in a person with tetraplegia using Gaussian process regression. *Neural Comput.* **30**, 2986–3008 (2018).
64. Fernández, E. & Botella, P. Biotolerability of intracortical microelectrodes. *Adv. Biosyst.* **2**, 1700115 (2018).
65. Zhai, S., Hunter, M. & Smith, B. A. Performance optimization of virtual keyboards. *Hum. Comput. Interact.* **17**, 229–269 (2002).
66. Zumsteg, Z. S. et al. Power feasibility of implantable digital spike sorting circuits for neural prosthetic systems. *IEEE Trans. Neural Syst. Rehabil. Eng.* **13**, 272–279 (2005).
67. Kao, J. C. et al. Single-trial dynamics of motor cortex and their applications to brain-machine interfaces. *Nat. Commun.* **6**, 7759 (2015).
68. Shanechi, M. M. Brain-machine interface control algorithms. *IEEE Trans. Neural Syst. Rehabil. Eng.* **25**, 1725–1734 (2017).
69. Sussillo, D., Stavisky, S. D., Kao, J. C., Ryu, S. I. & Shenoy, K. V. Making brain-machine interfaces robust to future neural variability. *Nat. Commun.* **7**, 13749 (2016).
70. Glaser, J. I., Chowdhury, R. H., Perich, M. G., Miller, L. E. & Kording, K. P. Machine learning for neural decoding. Preprint at <https://arxiv.org/abs/1708.00909> (2017).
71. Cunningham, J. P., Gilja, V., Ryu, S. I. & Shenoy, K. V. Methods for estimating neural firing rates, and their application to brain-machine interfaces. *Neural Netw.* **22**, 1235–1246 (2009).
72. Perge, J. A. et al. Intra-day signal instabilities affect decoding performance in an intracortical neural interface system. *J. Neural Eng.* **10**, 036004 (2013).
73. *Cerebus: Instructions for Use* (Blackrock Microsystems, 2020); <https://www.blackrockmicro.com/wp-content/ifu/LB-0028-15.00-Cerebus-Instructions-for-Use.pdf>
74. Bahrami, H., Mirbozorgi, S. A., Rusch, L. A. & Gosselin, B. BER performance of implant-to-air high-speed UWB data communications for neural recording systems. *IEEE Proc. Eng. Med. Biol. Soc. Conf.* **2014**, 3961–3964 (2014).
75. Ebraze, A. & Mohseni, P. 30 pJ/b, 67 Mbps, centimeter-to-meter range data telemetry with an IR-UWB wireless link. *IEEE Trans. Biomed. Circuits Syst.* **9**, 362–369 (2015).
76. Harrison, R. R. et al. Wireless neural recording with single low-power integrated circuit. *IEEE Trans. Neural Syst. Rehabil. Eng.* **17**, 322–329 (2009).
77. Walden, R. H. Analog-to-digital converter survey and analysis. *IEEE J. Sel. Areas Commun.* **17**, 539–550 (1999).
78. Gibson, S., Chandler, R., Karkare, V., Markovic, D. & Judy, J. W. An efficiency comparison of analog and digital spike detection. In *2009 4th Int. IEEE/EMBS Conf. on Neural Engineering* 423–428 (IEEE, 2009).
79. Gibson, S., Judy, J. W. & Marković, D. Technology-aware algorithm design for neural spike detection, feature extraction, and dimensionality reduction. *IEEE Trans. Neural Syst. Rehabil. Eng.* **18**, 469–478 (2010).
80. Yang, Z., Zhao, Q., Keefer, E. & Liu, W. Noise characterization, modeling, and reduction for in vivo neural recording. *Adv. Neural Inf. Process. Syst.* **22**, 2160–2168 (2009).
81. Chandrakumar, H. & Marković, D. An 80-mVpp linear-input range, 1.6-G Ω input impedance, low-power chopper amplifier for closed-loop neural recording that is tolerant to 650-mVpp common-mode interference. *IEEE J. Solid-State Circuits* **52**, 2811–2828 (2017).
82. Mendrela, A. E. et al. A bidirectional neural interface circuit with active stimulation artifact cancellation and cross-channel common-mode noise suppression. *IEEE J. Solid-State Circuits* **51**, 955–965 (2016).
83. Muller, R., Gambini, S. & Rabaey, J. M. A 0.013 mm², 5 μ W, DC-coupled neural signal acquisition IC with 0.5 V supply. *IEEE J. Solid-State Circuits* **47**, 232–243 (2012).
84. Steyaert, M. S. & Sansen, W. M. A micropower low-noise monolithic instrumentation amplifier for medical purposes. *IEEE J. Solid-State Circuits* **22**, 1163–1168 (1987).
85. Kim, S.-J. et al. A 0.5-V sub- μ W/channel neural recording IC with delta-modulation-based spike detection. In *IEEE Asian Solid-State Circuits Conference* 189–192 (IEEE, 2014).
86. Dong, H., Yuanjin, Z., Rajkumar, R., Dawe, G. & Minkyu, J. 0.45 V 100-channel neural-recording IC with sub-mW/channel consumption in 0.18 mm CMOS. In *IEEE Int. Solid-State Circuits Conf.* 17–21 (IEEE, 2013).
87. Muller, R. *Low power, scalable platforms for implantable neural recording*. PhD dissertation, Univ. California Berkeley (2015); <http://www2.eecs.berkeley.edu/Pubs/TechRpts/2015/EECS-2015-19.html>

88. McCreary, J. L. & Gray, P. R. All-MOS charge redistribution analog-to-digital conversion techniques. *IEEE J. Solid-State Circuits* **10**, 371–379 (1975).
89. Karkare, V., Chandrakumar, H., Rozgić, D. & Marković, D. Robust, reconfigurable, and power-efficient biosignal recording systems. In *IEEE Custom Integrated Circuits Conf.* 1–8 (IEEE, 2014).
90. Goldsmith, A. *Wireless Communications* (Cambridge University Press, 2005).
91. Miranda, H. & Meng, T. H. A programmable pulse UWB transmitter with 34 energy efficiency for multichannel neuro-recording systems. In *IEEE Custom Integrated Circuits Conf.* 1–4 (IEEE, 2010).
92. Obeid, I. & Wolf, P. D. Evaluation of spike-detection algorithms for a brain-machine interface application. *IEEE Trans. Biomed. Eng.* **51**, 905–911 (2004).
93. Kaiser, J. F. On a simple algorithm to calculate the energy of a signal. *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* **1**, 381–384 (1990).
94. Chase, S. M., Schwartz, A. B. & Kass, R. E. Bias, optimal linear estimation, and the differences between open-loop simulation and closed-loop performance of spiking-based brain-computer interface algorithms. *Neural Netw.* **22**, 1203–1213 (2009).
95. Willett, F. R. et al. Signal-independent noise in intracortical brain-computer interfaces causes movement time properties inconsistent with Fitts' law. *J. Neural Eng.* **14**, 026010 (2017).
96. Gao, P. et al. A theory of multineuronal dimensionality, dynamics and measurement. Preprint at <https://www.biorxiv.org/content/10.1101/214262v2> (2017).
97. Liberti, W. A., Perkins, L. N., Leman, D. P. & Gardner, T. J. An open source, wireless capable miniature microscope system. *J. Neural Eng.* **14**, 045001 (2017).
98. Foster, J. D. et al. A freely-moving monkey treadmill model. *J. Neural Eng.* **11**, 046020 (2014).
99. Cunningham, J. P. et al. A closed-loop human simulator for investigating the role of feedback control in brain-machine interfaces. *J. Neurophysiol.* **105**, 1932–1949 (2011).
100. Santhanam, G., Ryu, S. I., Yu, B. M., Afshar, A. & Shenoy, K. V. A high-performance brain-computer interface. *Nature* **442**, 195–198 (2006).
101. Jindal, R. Compact noise models for MOSFETs. *IEEE Trans. Electron Devices* **53**, 2051–2061 (2006).
102. Scholten, A. et al. Noise modeling for RF CMOS circuit simulation. *IEEE Trans. Electron Devices* **50**, 618–632 (2003).
103. Hariprasath, V., Guerber, J., Lee, S.-H. & Moon, U.-K. Merged capacitor switching based SAR ADC with highest switching energy-efficiency. *Electron. Lett.* **46**, 620–621 (2010).
104. Razavi, B. The strongarm latch [a circuit for all seasons]. *IEEE Solid-State Circuits Mag.* **7**, 12–17 (2015).
105. Harpe, P., Gao, H., van Dommel, R., Cantatore, E. & van Roermund, A. H. A 0.20 mm² 3 nW signal acquisition IC for miniature sensor nodes in 65 nm CMOS. *IEEE J. Solid-State Circuits* **51**, 240–248 (2016).
106. Chandrakumar, H. & Markovic, D. A 15.2-ENOB continuous-time $\Sigma\Delta$ ADC for a 200mV pp-linear-input-range neural recording front-end. In *IEEE International Solid-State Circuits Conf.* 232–234 (IEEE, 2018).
107. Murmann, B. The race for the extra decibel: a brief review of current ADC performance trajectories. *IEEE Solid State Circuits Mag.* **7**, 58–66 (2015).

Acknowledgements

We thank S. I. Ryu for electrode array implantation surgical assistance, M. Risch, M. Wechsler and A. Craig for expert surgical assistance and veterinary care. We thank B. Davis, G. Bell and N. Lam for administrative assistance. We thank W. L. Gore Inc. for donating Preclude artificial dura, used as part of the chronic non-human-primate electrode array implantation procedure. This work was supported by a Stanford Bio-X Institute fellowship (to N.E.-C.), Wu Tsai Neurosciences Institute fellowships (to D.G.M. and S.D.S.), an ALS Association Milton Safenowitz Postdoctoral Fellowship (to S.D.S.), an A. P. Giannini Foundation Postdoctoral Research Fellowship in California (to S.D.S.), a Career Award at the Scientific Interface from the Burroughs Wellcome Fund (to S.D.S.), the Rehabilitation R&D Service, the Department of Veterans Affairs (grants N2864C, A2295R, N9288C and B6453R to L.R.H.), grant NINDS-UH2NS095548 (to L.R.H., J.M.H. and K.V.S.), grant NIDCD-R01DC009899 (to L.R.H., J.M.H. and K.V.S.), grant NIDCD RO1DC014034 (to J.M.H., K.V.S. and L.R.H.), grant NINDS-U01NS098968 (to L.R.H., J.M.H. and K.V.S.), the DARPA 'NESD' (to B.M. and K.V.S.), the DARPA 'NeuroFast' (grant W911NF-14-2-0013 to K.V.S.), NIH Director's Pioneer Award 8DP1HD075623 (to K.V.S.), a Simons Foundation grant 'Simons Collaboration on the Global Brain (SCGB)' 543045 (to K.V.S.), the Office of Naval Research grant W911NF-14-2-0013 (to K.V.S.), and the Howard Hughes Medical Institute (to K.V.S.).

Author contributions

N.E.-C. and D.G.M. designed the study and analysis, and wrote the manuscript with input from all other authors. N.E.-C. and S.D.S. were responsible for data collection. J.M.H. planned and performed T5's array placement surgery. L.R.H. is the sponsor investigator of the multi-site pilot clinical trial. J.M.H., B.M. and K.V.S. were involved in all aspects of the study.

Competing interests

The MGH Translational Research Center has clinical research support agreements with Neuralink Inc., Paradromics Inc. and Synchron Medical, for which L.R.H. provides consultative input. K.V.S. and J.M.H. are consultants to Neuralink Inc. K.V.S. is on the Scientific Advisory Boards of CTRL-Labs Inc., Mind-X Inc., Inscopix Inc. and Heal Inc. These entities did not support this work.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41551-020-0595-9>.

Correspondence and requests for materials should be addressed to N.E.-C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Neural control and task cuing were controlled by custom software running on the Simulink/xPC real-time platform (The Mathworks, Natick, MA). Neural data were collected by the NeuroPort System (Blackrock Microsystems, Salt Lake City, UT).

Data analysis

Data analysis was done using MATLAB (The Mathworks, Natick, MA).
The custom code used in this study to produce the figures is available at <https://shenoy.people.stanford.edu/data>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The main data supporting the results in this study are available within the paper and its Supplementary Information. The raw human neural data is available on request to K.V.S. or J.M.H., yet owing to the potential sensitivity of the data and to respect the participant's expectation of privacy, an agreement between the researcher's institution and the BrainGate consortium is required to facilitate the sharing of these datasets. Processed data is available at <https://shenoy.people.stanford.edu/data>.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

- Sample size
- Data exclusions
- Replication
- Randomization
- Blinding

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | | |
|-------------------------------------|---|
| n/a | Involvement in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Animals and other organisms |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |

Methods

- | | |
|-------------------------------------|---|
| n/a | Involvement in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

- Laboratory animals
- Wild animals
- Field-collected samples
- Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about [studies involving human research participants](#)

- Population characteristics
- Recruitment

Ethics oversight

Permission for the study was granted by the US Food and Drug Administration (Investigational Device Exemption) and the Institutional Review Boards of Stanford University (protocol 20804), Partners Healthcare / Massachusetts General Hospital (2011P001036), Providence VA Medical Center (2011-009), and Brown University (0809992560).

Note that full information on the approval of the study protocol must also be provided in the manuscript.