nature portfolio

https://doi.org/10.1038/s41928-021-00631-8

Supplementary information

Neural recording and stimulation using wireless networks of microimplants

SUPPLEMENTARY INFORMATION

Neural recording and stimulation using wireless networks of microimplants

Jihun Lee[1], Vincent Leung[2], Ah-Hyoung Lee[1,3], Jiannan Huang[4], Peter Asbeck[4], Patrick P. Mercier[4], Stephen Shellhammer[5], Lawrence Larson[1], Farah Laiwalla[1], and Arto Nurmikko[1,6]

[1]School of Engineering, Brown University, Providence, RI, USA
[2]Electrical and Computer Engineering, Baylor University, Waco, TX, USA
[3]Graduate School of Convergence Science and Technology, Seoul National University, Seoul, South Korea
[4]Electrical and Computer Engineering, University of California San Diego, San Diego, CA, USA
[5]Qualcomm Inc, San Diego, CA, USA
[6]Carney Institute for Brain Science, Brown University, Providence, RI, USA

Comparison with other contemporary approaches

Table S1 compares the neurograin approach with sampling of representative state-of-the-art wireless microimplants in terms of wireless powering, data communication, recording and stimulation. A key difference is the first row, namely the number of networked devices. The Neurograin system is shown to have a considerably higher data rate to provide sufficient bandwidth for multi-node communication. Note that with the SoC design which integrates all circuit components as well as a coil on-chip, the present volume of Neurograin is approximately 0.1 mm^3, considerably smaller than other microdevices to date. We have elsewhere demonstrated that this volume can be reduced to 0.01 mm^3 after mechanical thinning and encapsulation with conformal hermetic packaging based on atomic-layer deposition (ALD) [1]. In addition to their unique capability for wireless networking, the miniaturized chips achieve recording and stimulation capability comparable to other microimplants. Designed for ultralow-power operation, the neurograin power budget is less than 30 µW for each chip. The main circuit elements of the neurograin ASICs are summarized in Fig. S1.

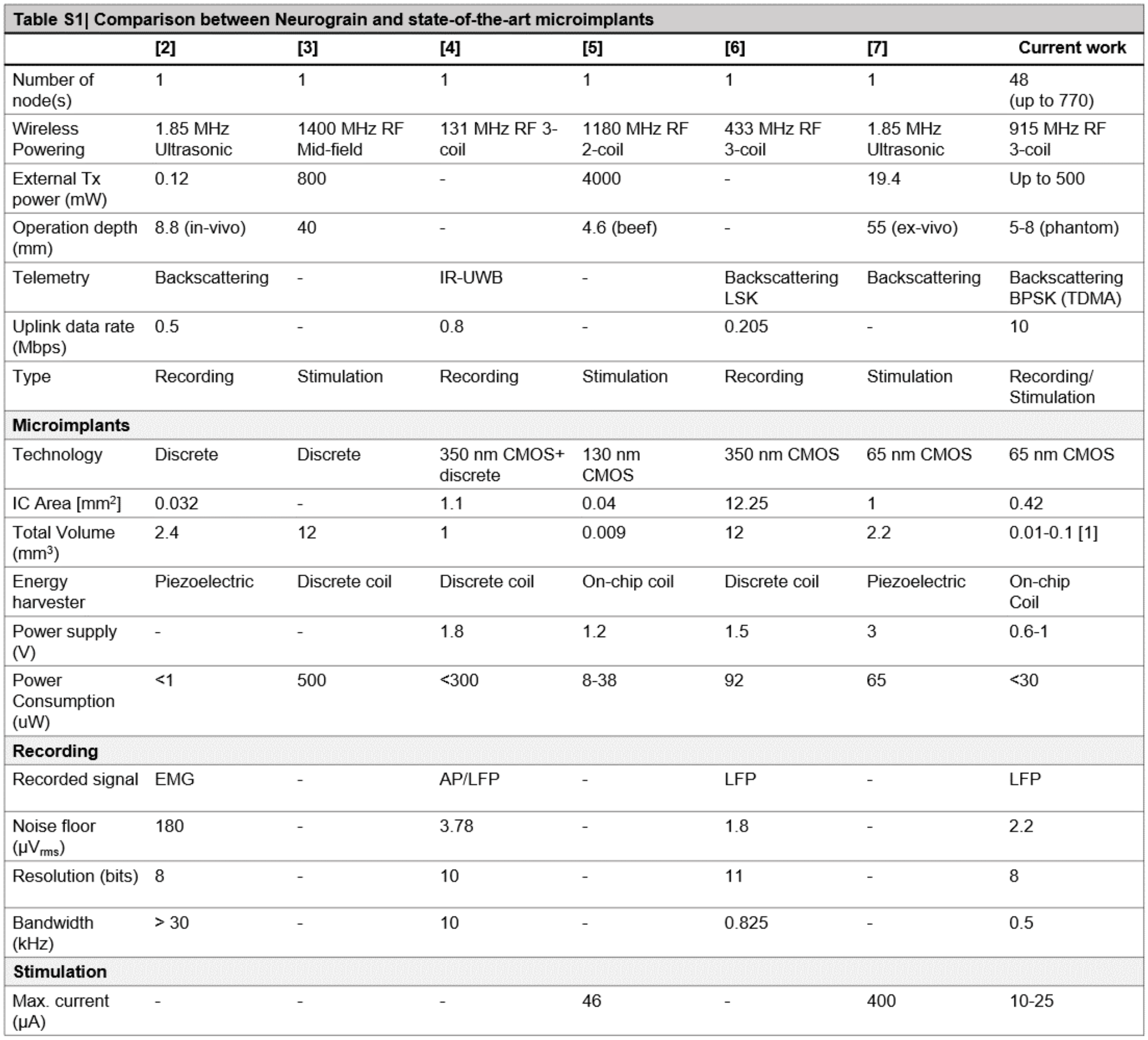


Table S1| Comparison between Neurograin and state-of-the-art microimplants

	[2]	[3]	[4]	[5]	[6]	[7]	Current work
Number of node(s)	1	1	1	1	1	1	48 (up to 770)
Wireless Powering	1.85 MHz Ultrasonic	1400 MHz RF Mid-field	131 MHz RF 3-coil	1180 MHz RF 2-coil	433 MHz RF 3-coil	1.85 MHz Ultrasonic	915 MHz RF 3-coil
External Tx power (mW)	0.12	800	-	4000	-	19.4	Up to 500
Operation depth (mm)	8.8 (in-vivo)	40	-	4.6 (beef)	-	55 (ex-vivo)	5-8 (phantom)
Telemetry	Backscattering	-	IR-UWB	-	Backscattering LSK	Backscattering	Backscattering BPSK (TDMA)
Uplink data rate (Mbps)	0.5	-	0.8	-	0.205	-	10
Type	Recording	Stimulation	Recording	Stimulation	Recording	Stimulation	Recording/ Stimulation
Microimplants							
Technology	Discrete	Discrete	350 nm CMOS+ discrete	130 nm CMOS	350 nm CMOS	65 nm CMOS	65 nm CMOS
IC Area [mm^2]	0.032	-	1.1	0.04	12.25	1	0.42
Total Volume (mm^3)	2.4	12	1	0.009	12	2.2	0.01-0.1 [1]
Energy harvester	Piezoelectric	Discrete coil	Discrete coil	On-chip coil	Discrete coil	Piezoelectric	On-chip Coil
Power supply (V)	-	-	1.8	1.2	1.5	3	0.6-1
Power Consumption (uW)	<1	500	<300	8-38	92	65	<30
Recording							
Recorded signal	EMG	-	AP/LFP	-	LFP	-	LFP
Noise floor (µV$_{rms}$)	180	-	3.78	-	1.8	-	2.2
Resolution (bits)	8	-	10	-	11	-	8
Bandwidth (kHz)	> 30	-	10	-	0.825	-	0.5
Stimulation							
Max. current (µA)	-	-	-	46	-	400	10-25

Comparison of possible TDMA, FDMA and CDMA approaches for Neurograins

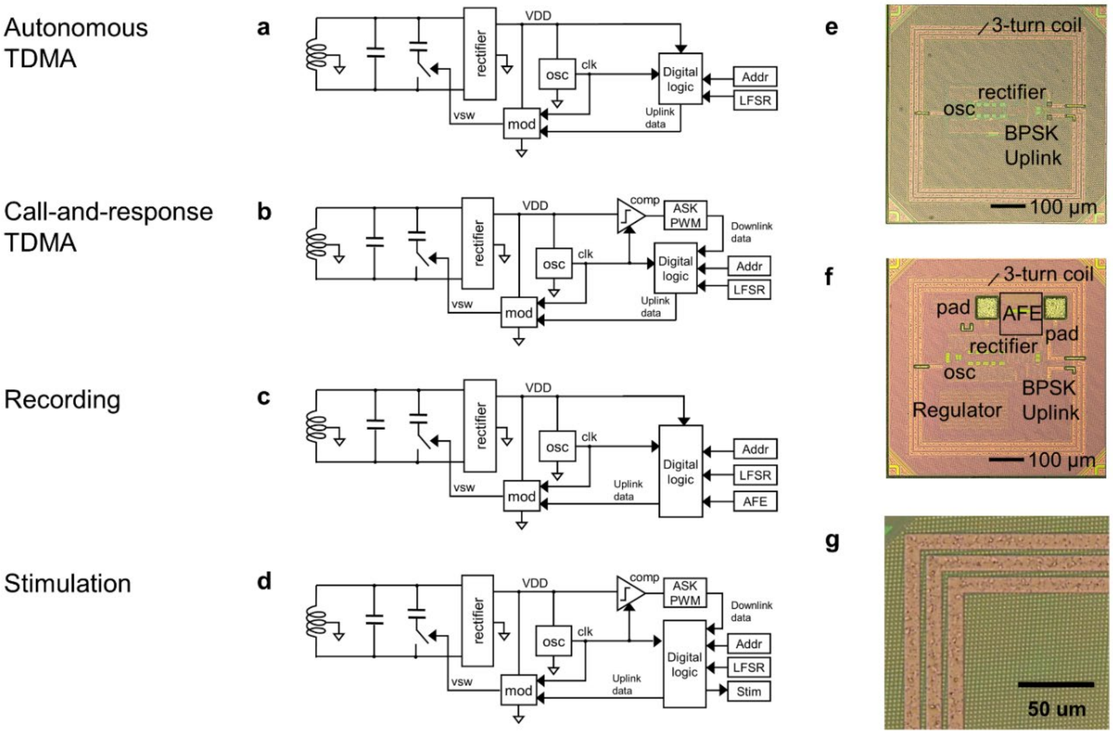
TDMA (Time-division multiple access), FDMA (Frequency-division multiple access) and CDMA (Code-division multiple access) can all be employed for one-to-many communication of wireless neural microdevices. In FDMA, each RF devices use a different frequency band which allows data sampling without delay. For the current Neurograin application, each device can have a minimum bandwidth of 8 kHz while channel bandwidth needs to be increased for overhead to accommodate clock variances. However, as the RF harvesting microimplant approach here incorporates a 3-coil system, the intermediate (unloaded) resonant coil has high-Q and narrow RF bandwidth which limits the range of the group bandwidth of the system and subsequently the channel size. Additionally, when assigning a frequency band for an individual chip, each chip needs to be post-processed after CMOS fabrication to tune its oscillator frequency with high accuracy – rather challenging in the case of a large number of nodes.

For synchronous CDMA application, the synchronization of the network plays a key role and the clock frequency of all nodes need to be precisely tuned as well as in phase. Under the clock variance, synchronous CDMA cannot operate as expected and imposing mutually orthogonal digital signals (which are used as bits) for the individual chip will be difficult, though not impossible. Asynchronous CDMA is not applicable here as in the case of the number of nodes over hundreds, the aggregate noise level will be too high. On the other hand, designing nodes with fixed clock frequency can provide a solution for FDMA and CDMA though this is a significant challenge under the microchip power and area constraint.

A practically useful TDMA protocol requires fast data transmission during a short period of time and has lower spectral efficiency. However, due to the usage of PUF (physical unclonable function)- based device ID, all chips can have the same fixed design and do not require individual post-processing for channel separation. Also, when combined with ASK-PWM downlink as in call-and-response TDMA chips, the clock variance does not significantly affect the number of accessible channels. Particularly under the perspective of the neural interface, call-and-response TDMA allows channel selection for nodes with important neural information and more frequent checking them so as to reduce the latency.

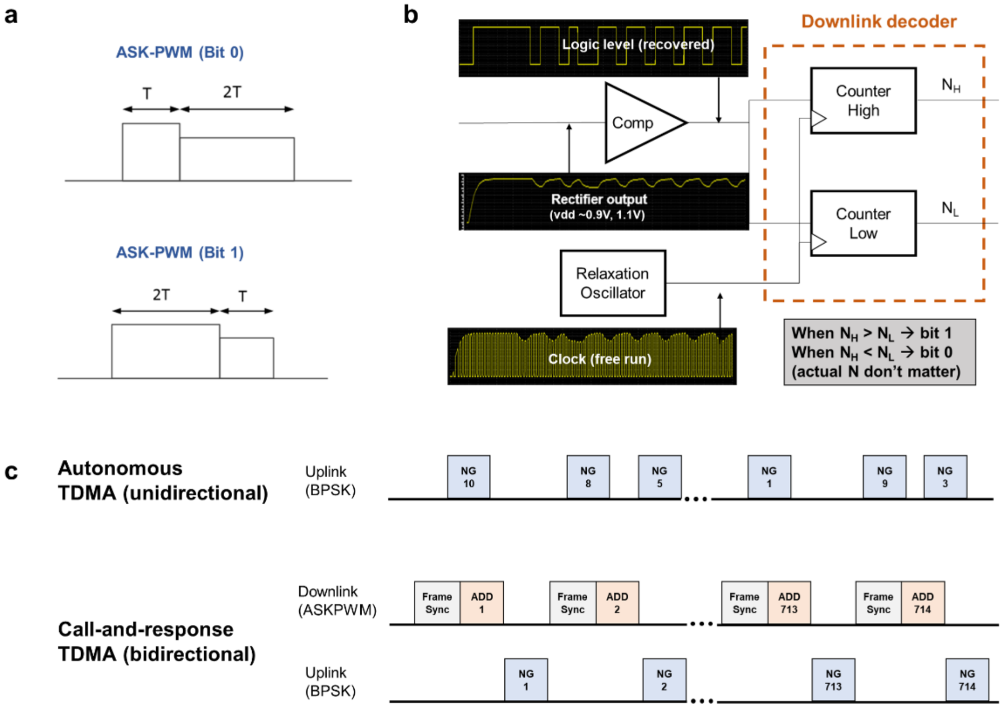
The rationale for the 1 GHz range frequency selection

The 1 GHz baseband RF range was chosen by us as optimal between several competing frequency dependent factors. As the RF frequency is lowered, the achievable quality factor on the microchip Rx coil is limited while the efficiency between Tx to relay coil can improve. At higher RF frequencies, tissue absorption losses increase and the quality factors of the Tx and relay coils can decrease due to self-resonance effects [8, 9]. An HFSS simulation suggests an optimal frequency of the coil system as 377 MHz assuming 1 kOhm circuit load impedance. However, in addition to the tight trade-off among the coil couplings, following RF circuit design can play a critical role in this type of wireless system. In a low-frequency band, a matching capacitor and BSPK modulator need to be scaled much larger in the footprint, which already occupies 23 % of the total at 915 MHz. Lumping the multiple factors together, we chose 915 MHz as the baseband frequency which could also benefit from the availability of mature wireless technologies. The near 1 GHz frequency region also allows the increase of available telecommunication bandwidth, which is important in scaling the system whether to, say, 770 neurograins (case example in this paper) or beyond. We have not considered more complex geometries for the Rx coil in this work.

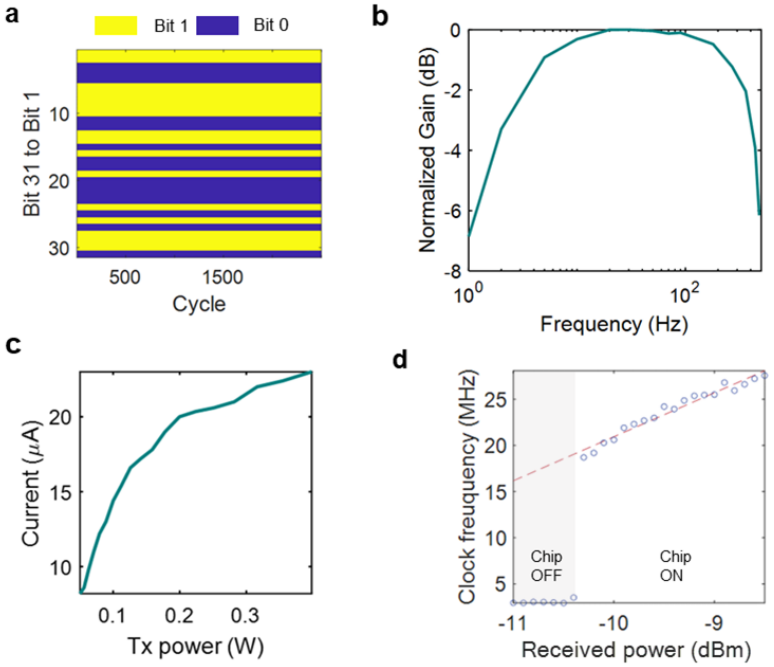


Supplementary Figure 1| Circuit block diagrams for the Neurograin ASICs. (a) Autonomous TDMA networking chip (with PUF address). Using a BPSK modulator, each neurograin backscatters every 100 ms while its data packet length is 100 us. The data packets contain about 1000 predefined bits from a linear-feedback shift register (LFSR) for BER performance evaluation. (b) Call-and-response TDMA protocol-based chip, which shares RF energy harvesting circuits as in the autonomous version but adds an amplitude shift keying, pulse width modulation (ASK-PWM) demodulator for bidirectional communication. (c) Neural recording ASIC with a PUF address and a capacitor-less analog-front-end (AFE). The recording chips have a voltage regulator to ensure stable recording. The chip backscatters in autonomous TDMA mode every 100 ms with the length of each data packet being 82 us. A packet contains predefined 32 bits from the LFSR, 20 bits dedicated to the PUF address, and 800 bits for the ADC output (8 bits/100 samples) (d) Stimulating Neurograin with an ASK-PWM demodulator and a biphasic current source stimulator. The chip generates current output for injection to neural tissue when the address code in the incoming downlink matches a 10-bit laser-written fuse address on-chip. Microphotographs of: (e) The networking test chip, (f) A recording neurograin, (g) Spatial distribution of the top metal fill in the TSMC 65 nm process. Diced chips have a nominal size of 650 μm × 650 μm including a 75 μm gap between the chip edge and the coil where the seal ring is placed. Within a 500 μm × 500 μm footprint, the 3-turn coil occupies 30 % of this area for RF harvesting while the rectifier and uplink circuits employ around 23 % of the total area. For the recording chip, the digital circuit, which handles data input/output and storage, regulator, and other circuits such as AFE and PUF, takes up most of the rest of chip real estate.

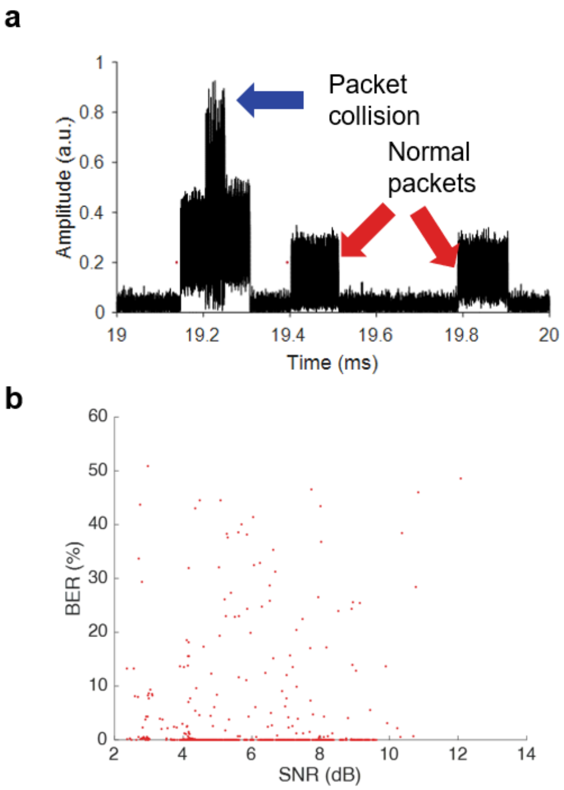
Abbreviations in the circuit diagrams. LFSR: linear-feedback shift register, Addr: address, mod: modulator, osc: oscillator, vsw: switch voltage, comp: comparator, AFE: analog-front-end, stim: stimulator



Supplementary Figure 2| The ASK-PWM scheme and networking protocol. (a) sample ASK-PWM bits for data and frame synchronization; a "short high" represents bit 0 while a "long high" is equivalent to bit 1. (b) The logic decoder for relative pulse width comparison between high and low. In the ASK-PWM protocol, the lower state (RF) energy total is smaller than that of the high state which decreases the power level fluctuation during the downlink onto the chip while the standard PWM would generate larger energy fluctuations. (c) Schematic to illustrate the timing diagram for autonomous and call-and-response TDMA cases, respectively. Each contains a 100 ms frame and 82 us data packet length.



Supplementary Figure 3| Results for ASIC test on the analog-front end and data communication. (a) Results from a test chip of continuous BPSK demodulation with 31 repeating LFSR bits over 2500 cycles showing zero bit error. (b) Recording neurograin: normalized gain spectrum; the low cutoff frequency is determined by the decoupling input capacitors in the test bench which mimics the high DC impedance in the electrode-electrolyte interface. (c) Stimulating neurograin: injected current as a function of the external Tx power in a 2-coil test setup. (d) The relationship between the received power at the Rx coil and its impact on the clock frequency of the on-chip relaxation oscillator.

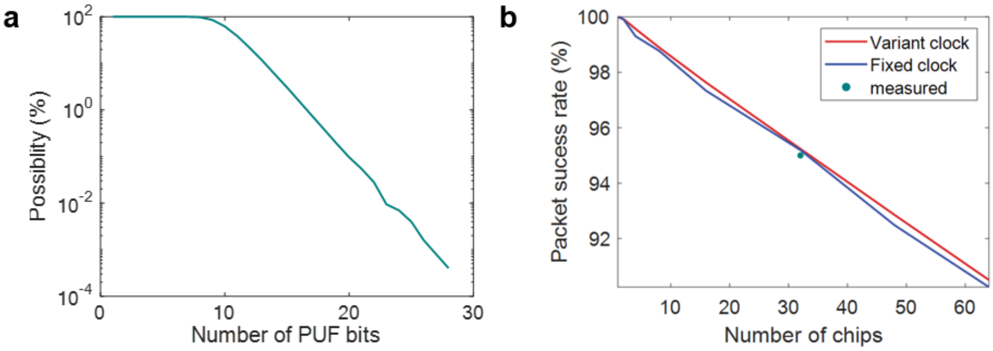


Supplementary Figure 4| Demonstration of the Autonomous TDMA. (a) The transient waveform of received data showing uplink data packet collision from two neurograins in the autonomous TDMA mode. Collisions are caused by unscheduled backscattering and chip clock frequency variations. (b) Signal-to-noise ratio (SNR) and Bit-error-rate (BER) of 1160 packets measured over 2 seconds from an ensemble of 64 autonomous TDMA chips. Partial packet collision was observed from 128 packets, which complies with the packet success rate simulation in supplementary Fig. S5b. When including collided packets, we recovered data with an average BER of 1.82 %.

Scalability of the neurograin TDMA network

Here we study the communication between a high number of potential nodes (on the scale of 1000) and a single external RF hub, adopting the TDMA approach. The multi-node TDMA telecommunication between the external hub and an ensemble of spatially distributed implants is intrinsically subject to overall latency. Our current networking protocol is designed for a 100 ms maximum latency which serves as a constraint to the design of neurograin networking protocol. Each recording neurograin simultaneously collects neural signals at 8 kbps (with ADCs sampling at 8-bit resolution and 1 kHz rate); 800 bits of data are thus accumulated over 100 ms per chip.

To test whether a physically unclonable function (PUF) can provide 1000 unique device IDs, we analyzed the number of PUF bits and the probability of device ID overlap while assuming PUF bits are randomly determined and have an equal chance to be either 0 and 1. Here the target number of nodes was 1000 and, as shown in Fig. S5a, the possibility of device ID overlap decreases for the higher number of PUF bit content. For the 16-bits and 20-bits PUF adopted in our system, we calculate a less than 1.5 % and 0.1 % chance of address overlap, respectively. This simulation shows that PUF-based device ID can provide all 1000 nodes unique addresses without any post-processing labeling after the CMOS fabrication.



Supplementary Figure 5| Analysis of PUF address uniqueness and the packet success rate in the autonomous TDMA. (a) The simulated possibility that any two PUF address among 1000 of them have the same bit sequence, given the assumption that PUF bits are decided randomly. (b) Simulation of the packet success rate assuming a randomized time slot for each chip. A higher number of chips shows a lower packet success rate. For this simulation, 2 seconds length of transient data was synthesized over 1000 repeats and the averaged collision rate was calculated on that. Here, any partial collision between packets has been counted as a packet failure.

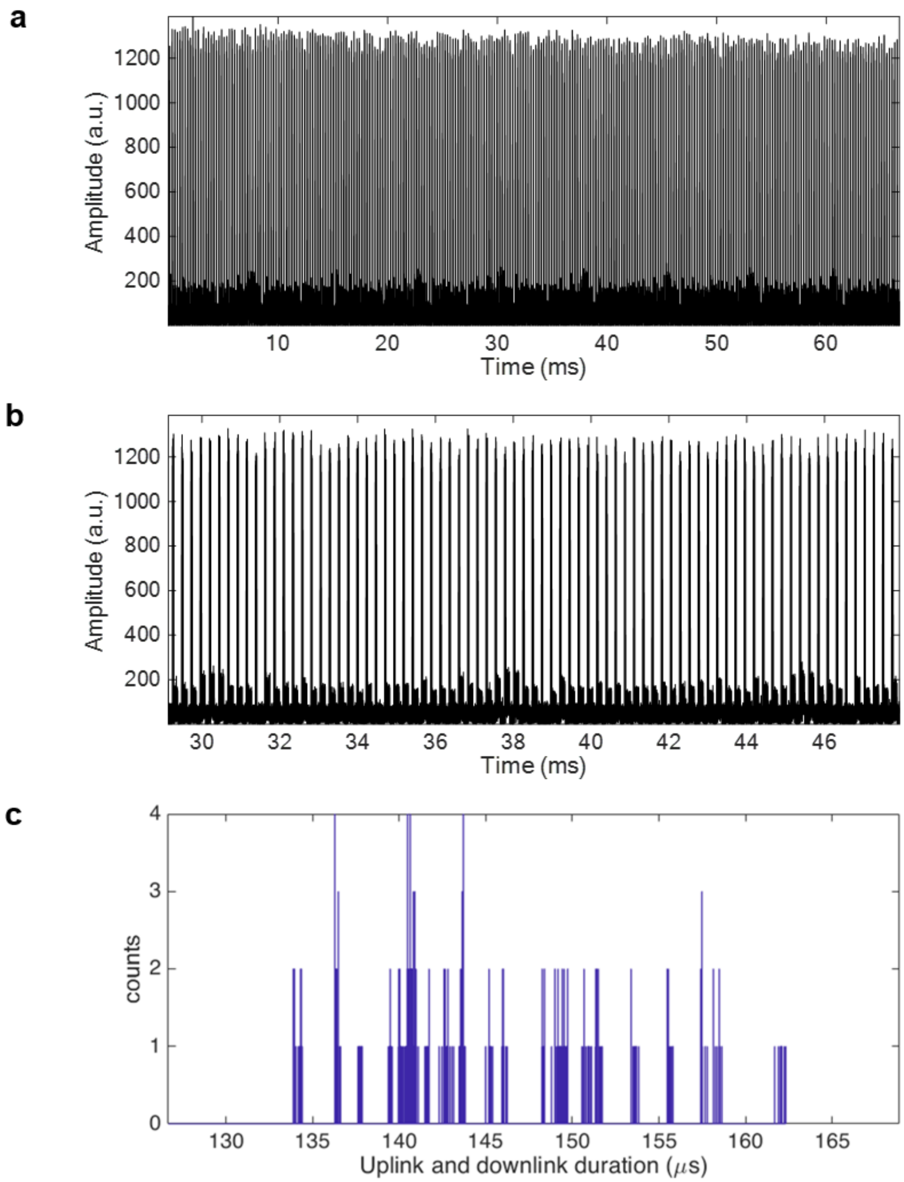
1) Autonomous TDMA

Under autonomous TDMA, all chips backscatter in random time slots every 100 ms using their unique PUF address. This 20-bit PUF address needs to be transmitted along with uplink neural data (total 820 bits) for channel tracking. Due to randomized time slots of autonomous TDMA, data packets from more than two chips may be transmitted at the same time resulting in a “packet collision”. As the nominal backscattering time for 820 bits is 82 µs (10 Mbps), the chance of this data packet collision can be calculated as in Fig. S5b. Any partial collision between packets was considered a packet failure. The simulation shows that the packet success rate decreases as a function of the number of chips. In the case of 32 and 64 chips operating in autonomous TDMA, we obtain a 95.28 % and 90.46 % packet success rate on average, respectively, when the clock variants among chips are included. While autonomous TDMA circuits provide a compact and low-power approach for early ensemble testing in rodents, where available real estate for the implant is limited e.g. by a subjects' anatomy, it has limited scalability if one aspires for a higher number of nodes. For large scale neurograin ensembles, we propose the "Call-and-Response" TDMA as a more sophisticated alternative for the communication protocol as follows:

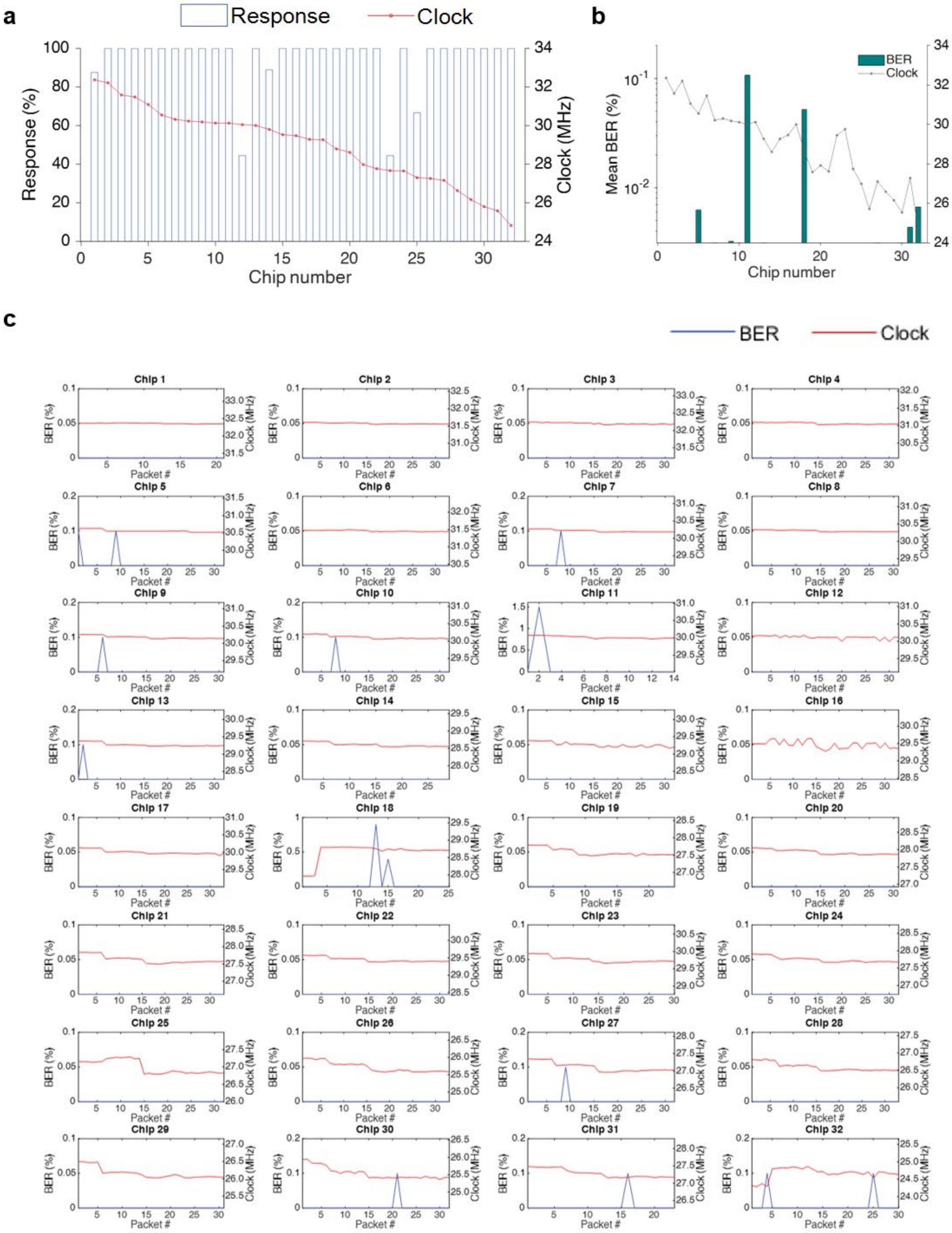
2) Call-and-Response TDMA

In Call-and-Response TDMA, the chip-ID is embedded into the downlink trigger command, causing only the implant with the matching ID to report its data content through the uplink, to prevent packet collision. For the current exploratory version of call-and-response TDMA neurograin, with 16-bit PUF addresses, a downlink rate of 1 Mbps, and Manchester-coding to ensure a data stream, each downlink trigger frame takes up 40 us. Fig. 3e showed an expanded view of the downlink and the uplink to/from three chips. Fig. S6a shows the transient RF amplitude on 32 chips call-and-response TDMA over 66 ms. To simulate a larger number of chips, the 32 sequences of the downlink were transmitted in a loop and the downlink assigned 281 time slots during 66 ms. Due to the finite yield in the CMOS post-process fabrication and dicing processes, 266 uplinks were observed with the BER and clock frequencies are shown in Fig. S7b. This result shows that the exploratory call-and-response chip can communicate with up to 425 Neurograin within 100 ms even allowing for the present 230 μs gap between the downlink sequences. Fig. S6c shows the histogram for the packet duration of one cycle of the uplink and the downlink whereby that the downlink can be transmitted every 170 μs so as to communicate with up to 588 chips without any circuit modification. However, in the present version, one cycle of the downlink and uplink have a 145 μs duration, on average, given that the current version of ASIC transmits additional bit sequences (1024 bits) for the communication link test (such as repeating 31 LFSR sequence and zero paddings). For more efficient data communication, one might choose to transmit the 20-bit PUF address through the downlink (20 μs under the 1 Mbps) to enable 800 bits transmitted through uplink during 80 μs (10 Mbps). Even with an additional time gap between the downlink and the clock variance, each time slot can be reduced to 130 us allowing communication with 770 chips in this Call-and-Response scheme.

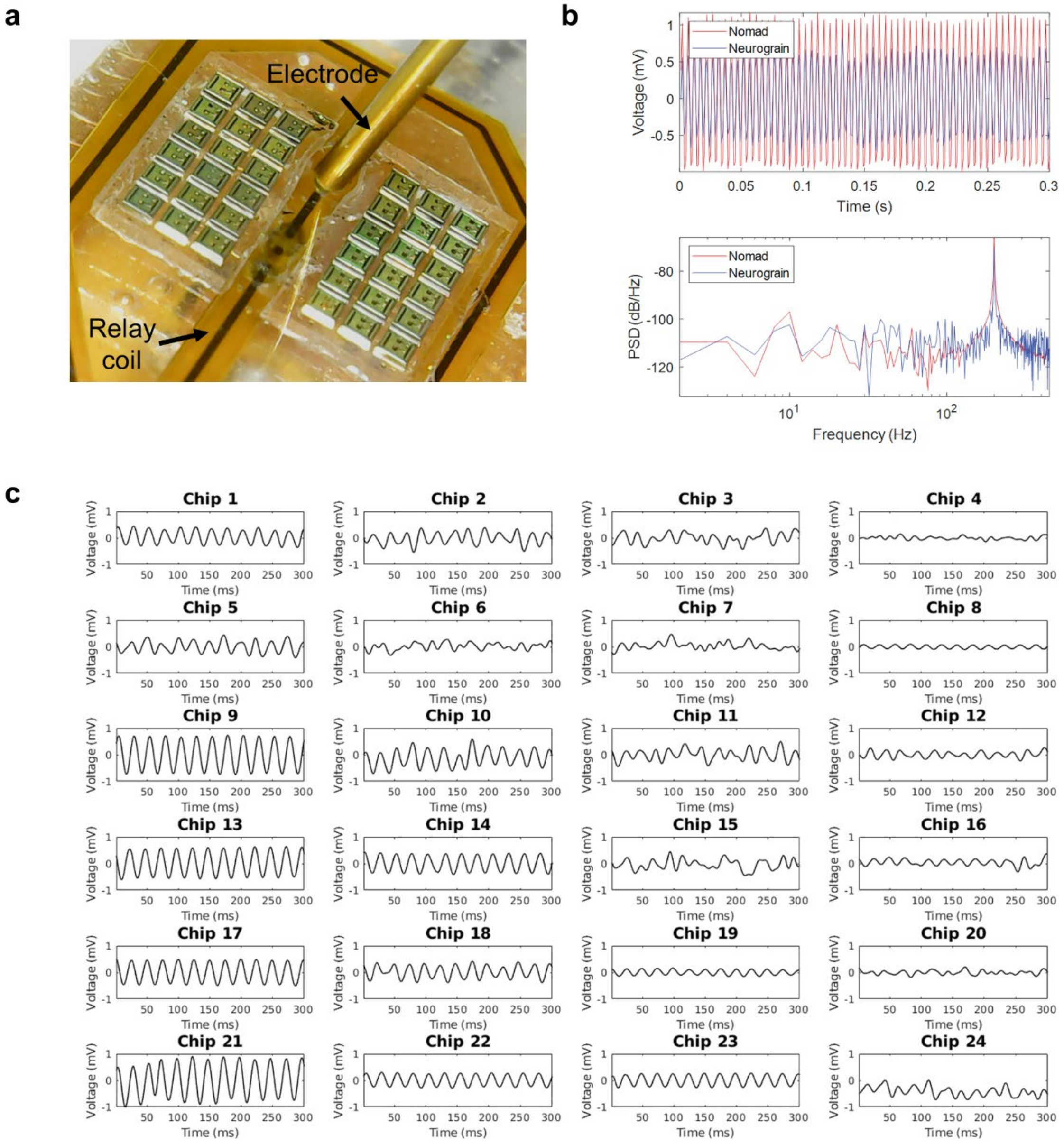
Note that the Call-and-Response approach provides for a way to down-select chip nodes which may contain the most meaningful neural information, paving a way to use this type of bidirectional communication approach for future 'adaptive sensing', to reduce the burden on data processing and to further reduce the system latency.



Supplementary Figure 6| Call-and-response TDMA demonstration. (a) & (b) Transient waveform in a receiving channel with 32 neurograin showing the large amplitude leakage of downlink and the relatively weak uplink data packet. Within 66 ms, 281 cycles of the downlink and uplink was achieved allowing up to 425 neurograins to be multiplexed in 100 ms. (c) The duration of single downlink command and following uplink packet, taking up to 145 µs on average.



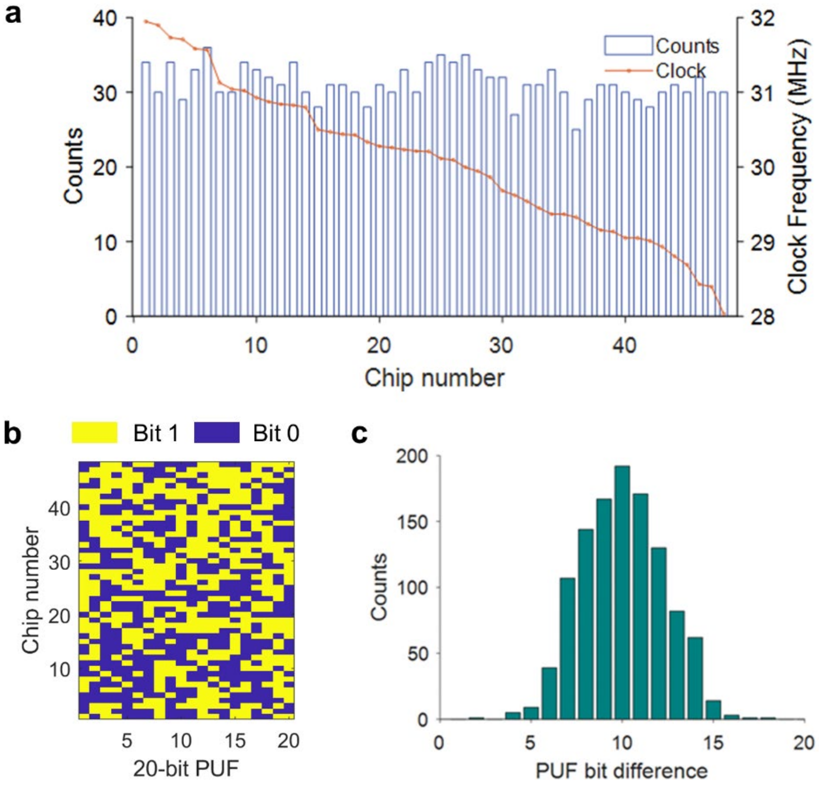
Supplementary Figure 7| Measurements to test the Call-and-Response TDMA. (a) ASK-PWM downlink response rate and the averaged clock frequencies of 32 call-and-respond TDMA chips. (b) Averaged BER and the clock frequency of 32 chips. (c) Variations of BER and clock frequency within received packets over the 32-chip network.



Supplementary Figure 8| Neurograin recording compared to the commercial device and multichannel neurograin recording in the saline to capture localized 40 Hz sinusoidal signal from a concentric electrode. (a) The photograph of the neurograin array in a saline testbench. The Tx and relay coil are underneath the array for wireless powering and data communication. (b) Recording of a 200 Hz signal injected in the saline by neurograin and a wired commercial device (Nomad, Ripple Inc) using gold planar electrodes. Both the transient waveforms (top) and the nonparametric power spectral density (PSD) estimates (bottom) are shown. Compared to the commercial wired device, the recording quality of neurograin is limited due to constraints on the circuit area and power, which may need further improvement. But, the current version of recording neurograin can still capture low and high frequency features of local field potentials. (c) 40 Hz sinusoidal wave recordings by the distributed neurograin network in the saline (only 24 locations are shown). The location of the subplot shows the relative location of the chip inside the saline. Here, a concentric electrode placed between chip 9 and chip 13 generated a localized 40 Hz signal, which has a higher magnitude around chip 9 and 13 and a lower magnitude around chip 4 and 8 based on the electrical field distribution.

In-vivo data analysis: Role of the PUF address

As shown in Fig. S9, in the case of an ensemble of 48 recording neurograins in-vivo, 1697 data packets were detected during 3.3 seconds. Each packet contains a 20-bit PUF address, and exactly 48 distinct addresses were identified as in Fig. S9b. Using this information, we can uniquely associate each packet to the chip which sent it. Fig. S9a shows the counts of received packets per chip. Data packet collisions and clock speed differences contribute to the variation of packet counts. Fig. S9c shows the histogram of the number of different PUF bits between any two chips. One sees a normal distribution with an average of 10 bits (that is, on average, half of the 20 PUF bits between two chips are different). Based on such observations, we conclude that the 20-bit PUF is random across the ensemble of chips, yet stable over time, making it a suitable technique for chip addressing purposes for free-floating, ultra-small micro-implants. Note that through this analysis, we did identify that all 48 chips are activated and transmitted their uplink data in-vivo.



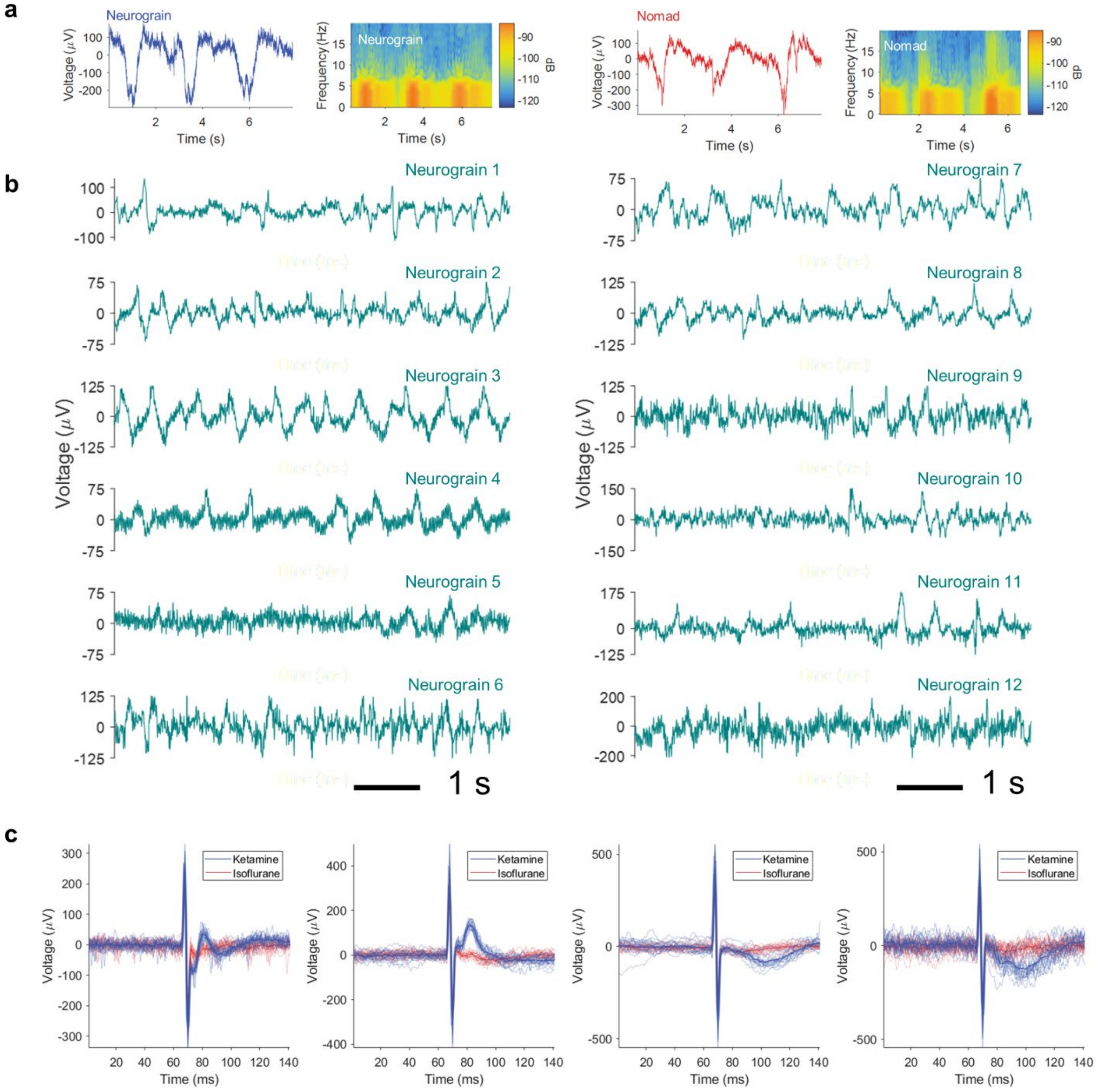
Supplementary Figure 9| PUF address analysis on 48 recording neurograin chips in-vivo. (a) Packet counts and clock frequencies, sorted using distinctive PUF addresses, of 48 chips measured over 3.3s in the rat cortex. (b) The PUF addresses, each 20-bit long, of the 48 chips. (c) The histogram on the count of bit differences (between two PUF addresses) for the ensemble of 48 chips.

In-vivo experiments setup for neurograin

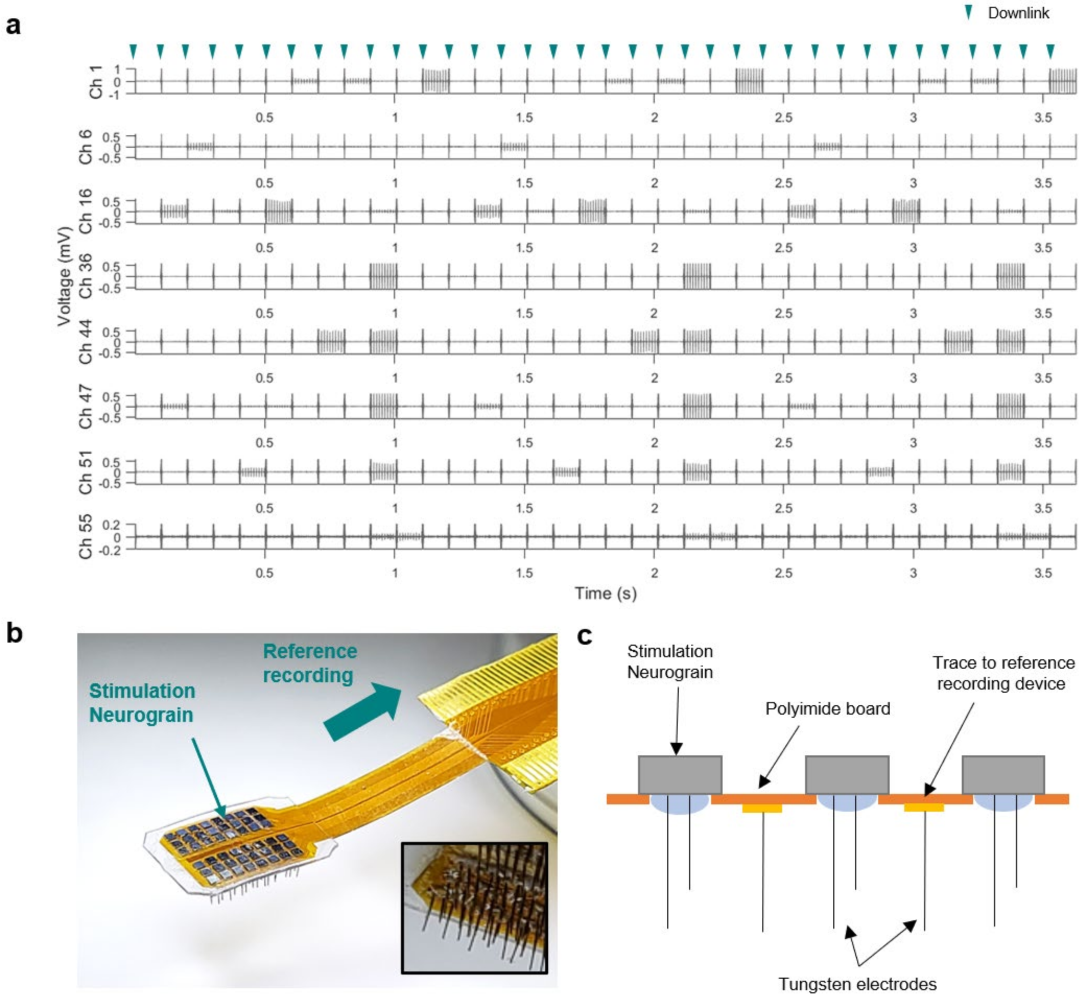
Using the rat model, the validation of recording and stimulation neurograin has been performed in separate experiments. As a reference to calibrate the neurograin system performance we used a commercial multichannel wired neural signal processing system (Nomad, Ripple Inc) recording data from wired electrodes implanted near the wireless neurograins. Note that in contrast with the rack-mounted commercial system, the wireless neurograins are tightly constrained in their area and power. For packaging, the neurograins and the relay coil were encapsulated with PDMS for in vivo use (shown in supplementary Fig. S8a). The chips were placed on the rat cortex to record epicortical signal while the Nomad system was configured also to record the corresponding ECoG signals as shown in Fig. S10a,b. After recording spontaneous cortical activity monitoring, we used the Nomad system to stimulate the cortex at 5 Hz through intracortical (microwire) electrodes with the neurograins capturing evoked responses as shown in Fig. 4e and Fig. S10c. During the experiments, Arduino board (Arduino IDE) and data acquisition unit (USB-6009, National Instrument) were used to start the neurograin recording or stimulation session.

In the next phase of the in vivo work, we deployed the intracortical stimulation neurograins, for which a separate specially fabricated assembly was configured as shown in Fig. S11b. The monolithic template consisted of stimulation neurograins and a number of wired recording electrodes connected to the Nomad as shown in Fig. S11c. The template was assembled on a polyimide substrate. By recording signals from several sites in proximity to the stimulating neurograins, we were able to monitor the overall stimulation patterns such as in Fig. S11a in saline as well as in the animal (Fig. S12). Note that in this case experiment, the commercial Nomad system was operated in the recording mode.

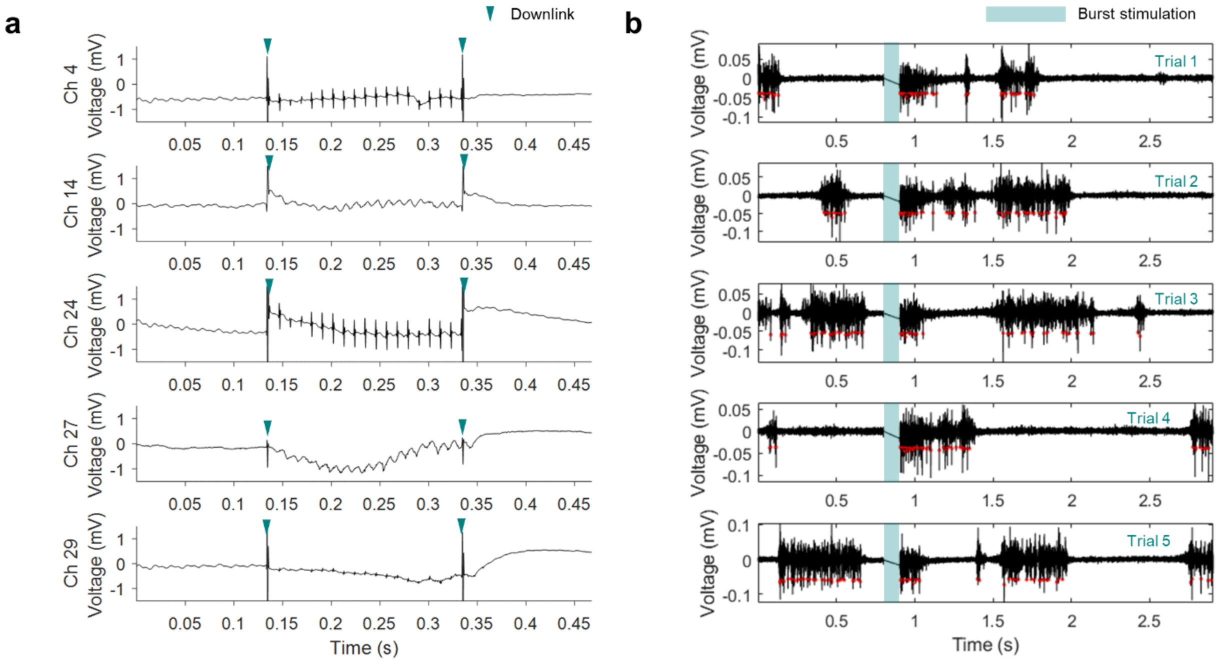
Due to the relatively large size of the construct (Fig. S11b) which required a large size of craniotomy to accommodate a significant number of neurograins, the skin and the skull were not replaced during the surgery. However, according to our RF simulations shown in Fig. S15d, attenuation less than 2.5 dB is expected in the fully closed tissue condition (2.5 mm skin plus 2.5 mm skull).



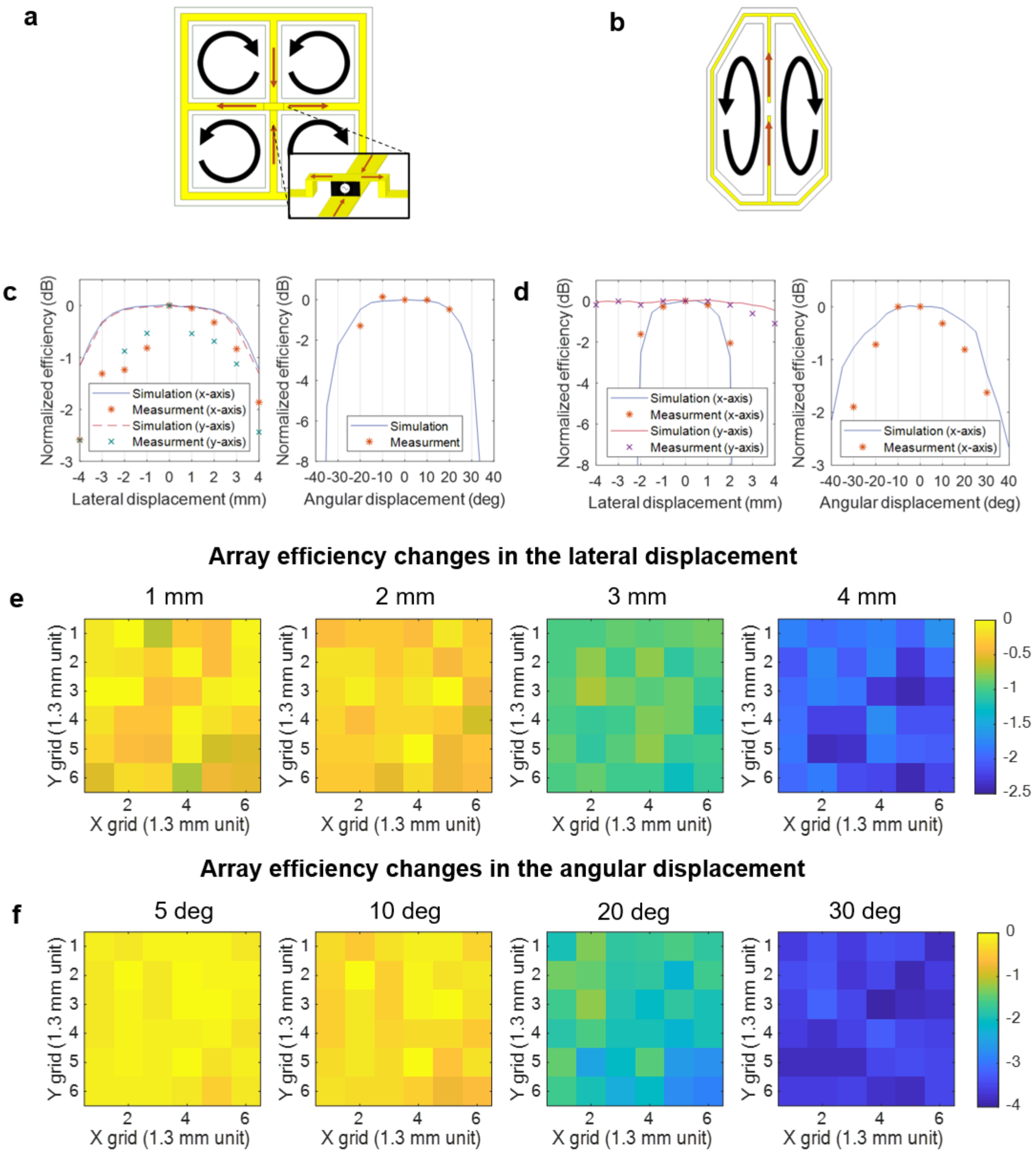
Supplementary Figure 10| Demonstration of Neurograin recordings in vivo. (a) Juxtaposing recordings of a spontaneous low-frequency oscillatory wave by the Neurograin and a commercial multichannel wired neural signal processing system (Nomad system by Ripple Inc) from an anesthetized rat with corresponding multi-taper spectrograms [10]. (b) Multiple neurograin recordings of spontaneous ECoG under ketamine featuring low-frequency cortical oscillations. In these proof-of-concept experiments, we would typically record low-noise and evident brain activity from a fraction of the 48 channels, shown here for 12 channels. Other channels displayed higher noise levels due to multiple possible factors such as poor contact with cortical tissue, i.e. imperfect electrode metal/tissue interface, areas of lower cortical activity coupled with practical limitations in placing the ensemble of microchips within the size craniotomy. (c) Superimposed raw ECoG signals over 20 stimulation trials showing evoked responses with 20 to 70 milliseconds duration. In this experiment, the stimulation was achieved using intracortical wire electrodes while neurograin recorded ECoG signals from the epicortical surface. The solid line shows averaged signal and each subplot shows results from different individual neurograins. The electrical stimulation was applied at an amplitude of 50 uA and pulse width per phase of 1 ms. Note that the evoked response by electrical stimulation has a higher amplitude under ketamine compared to that under isoflurane.



Supplementary Figure 11| Experiments demonstrating multichannel microstimulation by an ensemble of 12 stimulating neurograins in saline. A device carrier was fabricated from polyimide where the neurograins with intracortical microwires were intercalated with an array of tungsten wires which were used to record the patterned stimulation. (a) A sequential activation pattern of 12 Neurograins recorded by 8 channels of a wired commercial device in saline. Each downlink triggered a different stimulation neurograin to inject the current at 100 Hz. As a result, different patterns of stimulation were observed due to varying distances between neurograin to recording electrodes. The data was generated by wireless programming (indicated by downlink arrows) of stimulation patterns over 3 cycles. (b) & (c) Close-up view of the stimulation-recording test device including the neurograin array with post-processed penetrating electrodes and wired tungsten electrodes for recording. Within this test assembly, all neurograins operate in a fully wireless manner and while the proximal wired tungsten recording electrodes read out the stimulation patterns and local neural activities to quantitively assess the effects of neurograin stimulation on the cortex. Due to the manual electrode attachment process, tungsten electrodes are not strictly perpendicular to the device plane as evidenced in the zoomed-in close-up.



Supplementary Figure 12| Cortical responses evoked by microstimulation from a neurograin. (a) 100 Hz Neurograin stimulation and evoked LFP responses over 5 recording channels. (b) Spontaneous burst firing activities under ketamine artificially induced by a 400 Hz Neurograin stimulation for 100 ms and recorded over 5 trials.



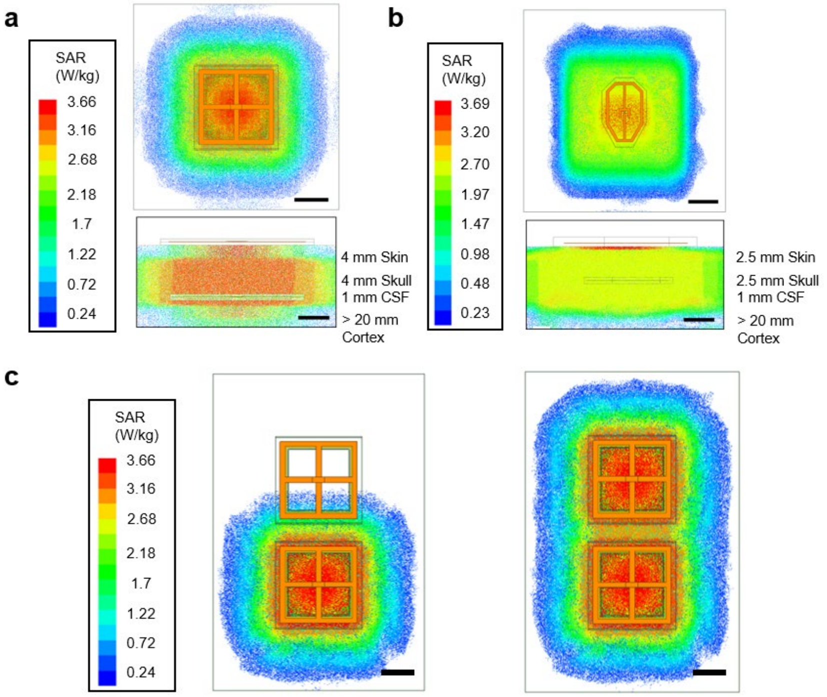
Supplementary Figure 13|Further details of the 3-Coil wireless power transfer system design and measurement. (a) & (b) Four or two sub-coils are combined to realize specific current flow for the (proposed) primate and the (implemented) rodent models, respectively, in order to expand the power harvesting area for scalability. (c) & (d) WPT efficiency as a function of the lateral or angular displacement of Tx coil for the primate and the rodent models, respectively. Here a neurograin was located in the center of the subunit coil. (e) & (f) Simulated efficiency changes across 36 neurograin array depending on the lateral and angular displacement of the coil in the primate coil model. Up to -2.38 dB and -3.86 dB loss is observed for 4 mm lateral and 30 deg angular displacement, respectively. The associated variance of efficiency on neurograin WPT was not significant as the displacement mainly affects the coupling between the Tx and the relay coil.

Dual transmitter wireless power transfer

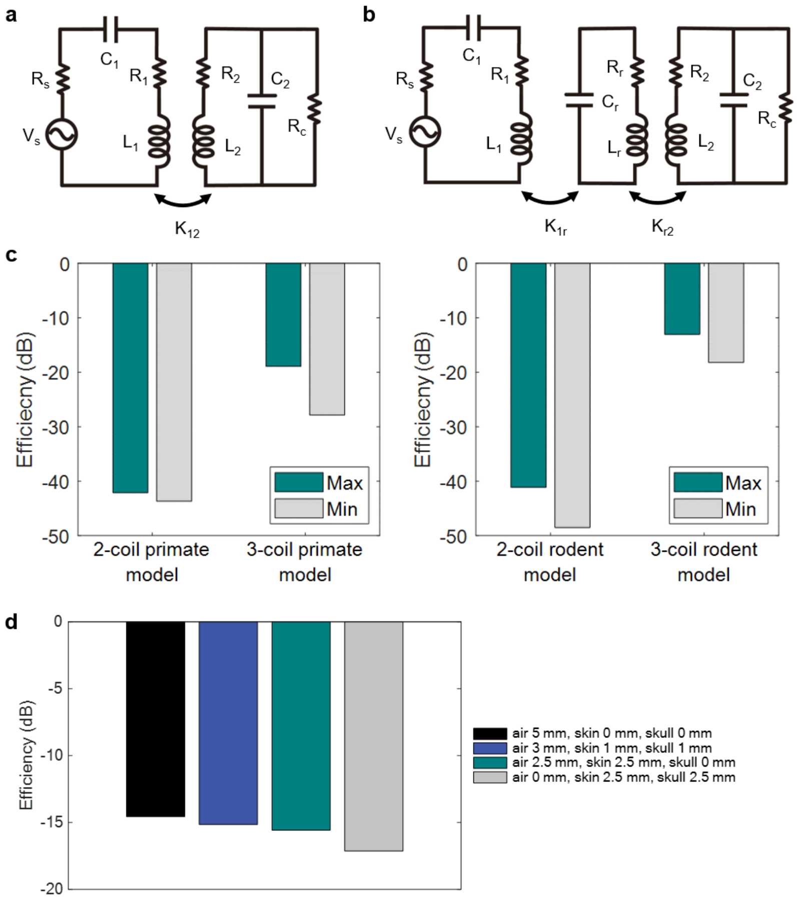
The emphasis in this paper is on a wireless powering method using a single Tx coil and single relay coil as in previous studies which deploy a 3-coil system [4, 11]. However, using multiple pairs of Tx/relay coils can provide an access to a wide-cortical area as well as multiple cortical areas. As such the magnetic field by two adjacent coils can interact destructively, which can significantly impair the wireless harvesting/communication. One remedy is to fine-tune the phase of coils to have an exactly 180-degree phase difference thereby generating a magnetic field in opposite direction. In fact, in our wireless system, the single Tx and relay coils already contain subunits to extend the lateral wireless powering area. Due to such magnetic field distribution, we can prevent destructive interference of adjacent coils if they are in-phase as shown in Fig. 5h (left). Unlike applying phase-shifting at 180-degree, which involves additional hardware, we have ensured the in-phase relationship by simply sharing the RF source. To simulate the situation when the exact phase match deviates from an ideal case, we tested a two-coil configuration driven by two independent sources and evaluated the bit-error-rate in communication in Fig. 5h (right). Fig. 5i shows that the BER is much higher when two coils are out-of-phase emphasizing the importance of phase tuning.

Table S2| Dimension and simulation parameters of WPT coil system at 915 MHz

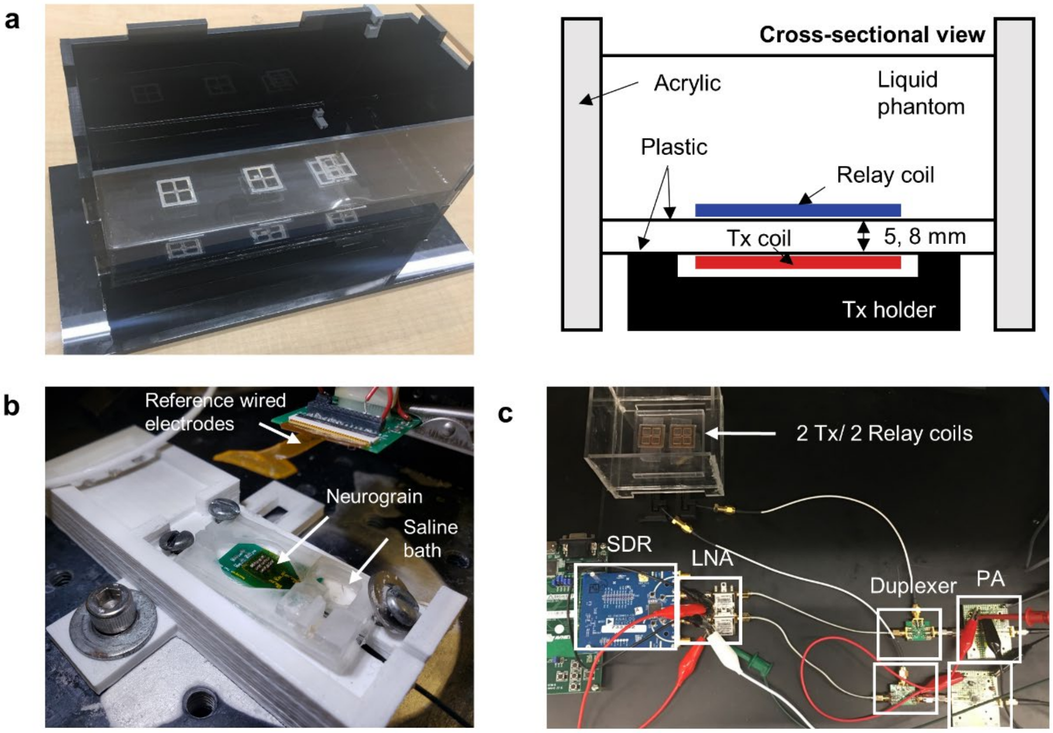
Type	**Primate model coil**	**Rodent model coil**
L_1 (nH)	12.2	20.9
L_R (nH)	11	16
L_2 (nH)	10.3	
Q_1	26.9	50
Q_R	30.8	109.8
Q_2	13.4	
k_{1R}	0.051	0.02
k_{R2}	0.0035-0.0099	0.008-0.015
k_{12}	0.0005-0.0006	0.0003-0.0007
Size Tx (mm)	22.5 x 22.5	20 x 12.3
Size Relay (mm)	20.4 x 20.4	14.2 x 8
Size Rx (mm)	0.5 x 0.5 (3 turns)	
Tissue	4 mm skin, 4 mm skull, >20 mm cortex	2.5 mm skin, 2.5 mm skull, >20 mm cortex



Supplementary Figure 14| Specific absorption rate (SAR) simulation averaged over 10-g tissue. (a) SAR field pattern of the primate WPT system with 0.1 W Tx emission. (b) SAR field of the rodent model coil with 0.072 W Tx emission. (c) Comparison of SAR fields with one Tx channel enabled (left) and two adjacent Tx channels enabled (right) with 0.1 W emission for each. The composition of the tissue layer in (c) was the same in (a). The scale bar is 10 mm. Averaging of SAR over 10-g tissue was done by following the procedure in IEC/IEEE 62704-4 [12].



Supplementary Figure 15| Comparison of WPT between the 2-coil vs. 3-coil configurations and dependence on the composition of tissue layers. (a) Lumped circuit model for the 2-coil wireless powering system. (b) The 3-coil model. (c) Efficiency comparison of the 2-coil vs 3-coil configurations for the primate and rodent systems, using Equation 1 and 2 with coil parameters from Table S2. Vs: source voltage, Rs: source impedance, Rc: circuit impedance. (d) Variation in efficiency for the 3-coil system (rodent model) depending on the composition of tissue layers. For example, up to 2.5 dB additional loss occurs with 5 mm tissue thickness.



Supplementary Figure 16| Neurograin benchtop RF setup. (a) Acrylic and plastic-based benchtop setup to minimize RF interference including a liquid phantom to mimic dielectric properties of brain tissue [12]. (b) Saline bath for testing recording/stimulation neurograins along with standard microwired (tungsten) electrodes. (c) The RF-component set up for dual channel power and data link using two adjacent pairs of Tx/ relay coils to double the channel capacity.

References in Supplementary Information

[1] Jeong, J. *et al.* Conformal hermetic sealing of wireless microelectronic implantable chiplets by multilayered atomic layer deposition (ALD). *Advanced Functional Materials* **29**, 1806440 (2019).

[2] Seo, D., Carmena, J. M., Rabaey, J. M., Alon, E. & Maharbiz, M. M. Neural dust: An ultrasonic, low power solution for chronic brain-machine interfaces. *arXiv preprint arXiv:1307.2196* (2013).

[3] Agrawal, D. R. *et al.* Conformal phased surfaces for wireless powering of bioelectronic microdevices. *Nature biomedical engineering* **1**, 0043 (2017).

[4] Yeon, P., Bakir, M. S. & Ghovanloo, M. Towards a 1.1 mm^2 free-floating wireless implantable neural recording SoC. In *2018 IEEE Custom Integrated Circuits Conference (CICC)*, 1–4 (IEEE, 2018).

[5] Khalifa, A.et al. The microbead: A 0.009 mm^3 implantable wireless neural stimulator. *IEEE transactions on biomedical circuits and systems* **13**, 971–985 (2019).

[6] Ahmadi, N. *et al.* Towards a distributed, chronically-implantable neural interface. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, 719–724 (IEEE, 2019).

[7] Piech, D. K. *et al.* A wireless millimeter-scale implantable neural stimulator with ultrasonically powered bidirectional communication. *Nature Biomedical Engineering* **4**, 207–222 (2020).

[8] Kim, S., Ho, J. S., Chen, L. Y. & Poon, A. S. Wireless power transfer toa cardiac implant. *Applied Physics Letters* **101**, 073701 (2012).

[9] Cheng, Y., Wang, G. & Ghovanloo, M. Analytical modeling and optimization of small solenoid coils for millimeter-sized biomedical implants. *IEEE Transactions on Microwave Theory and Techniques* **65**, 1024–1035 (2016).

[10] Bokil, H., Andrews, P., Kulkarni, J. E., Mehta, S. & Mitra, P. P. Chronux: a platform for analyzing neural signals. *Journal of neuroscience methods* **192**, 146–151 (2010).

[11] Kiani, M., Jow, U.-M. & Ghovanloo, M. Design and optimization of a 3-coilinductive link for efficient wireless power transmission. *IEEE transactionson biomedical circuits and systems* **5**, 579–591 (2011).

[12] Determining the peak spatial-average specific absorption rate (SAR) in the human body from wireless communications devices, 30 MHz to 6 GHz - part 3: Specific requirements for using the finite difference time domain (FDTD) method for SAR calculations of mobile phones. *IEC/IEEE 62704-3:2017* 1–76 (2017).