Nanofabricated Neural Probe System for Dense 3-D Recordings of Brain Activity

Thesis by

Gustavo Rios

In Partial Fulfillment of the Requirements for the

degree of

Doctor of Philosophy

Caltech

California Institute of Technology

Pasadena, California

2016

(Defended May 12, 2016)

ORCID: 0000-0003-1411-4933

Para mi familia y mi esposa.

Abstract

Computations in brain circuits involve the coordinated activation of large populations of neurons distributed across brain areas. However, monitoring neuronal activity in the brain of intact animals with high temporal and spatial resolution has remained a technological challenge. Here we address this challenge by developing a dense, three-dimensional (3-D) electrode array system for electrophysiology. The front-end of the system is composed of nanofabricated probes with ultrathin shanks that are engineered to minimize tissue damage. The probes are connected via flexible cables to custom PCBs that multiplex the electrophysiological signals. This system architecture decouples the front-end both mechanically and thermally from the PCB which carries all active electronics for signal conditioning and multiplexing. This system was validated *in vivo* with hippocampal recordings from head-fixed mice. The culmination of these efforts was a 3-D array with 1024 sites packed within 0.6 mm^3 of tissue that yielded the densest electrophysiological recordings to date.

Published Content and Contributions

Rios G., Lubenov E.V., Chi D., Roukes M.L., Siapas A.G., "Nanofabricated Neural Probes for Dense 3-D Recordings of Brain Activity" to be submitted, 2016.

Author contributions: system architecture and specification: GR, EVL, MLR, AGS; component design: GR, EVL; component mask layout: GR, EVL; component fabrication prototyping: GR, DC; component mass fabrication: GR, EVL, MLR, AGS; system packaging: GR, EVL; system software development: EVL; system bench-top evaluation: GR, EVL; system in vivo evaluation: GR, EVL, AGS; manuscript preparation and figure creation: GR, EVL with input from AGS, MLR, DC.

Acknowledgements: We thank Denis Renaud, Eric Rouchouze, and Hughes Metras for their help with foundry fabrication at LETI. We thank Jennifer Mok for histological processing of brains from the in vivo experiments.

Project background: GR's participation in this neural nanoprobe effort originated in the research group of Michael Roukes. Prototype fabrication was carried out in the Kavli Nanoscience Institute with Roukes group nanofabrication engineer Derrick Chi. Foundry fabrication was carried out at CEA/LETI, Grenoble under the aegis of the Alliance for Nanosystems VLSI.

Contents

Chapter 1

Introduction

Brain functions such as perception, motor control, learning, and memory arise from the coordinated activation of neuronal assemblies distributed across multiple brain areas. While major progress has been made in understanding the response properties of individual cells, circuit interactions remain poorly understood. One of the fundamental obstacles to understanding these interactions has been the difficulty of observing the activity of large distributed populations of neurons in behaving animals [1]. Electrophysiology has been the gold standard for monitoring the brain because it measures the electrical activity of neurons directly and at a high temporal resolution, sufficient to capture in detail even the fastest neuronal events [2; 3]. The main drawback of electrophysiology has been the invasiveness of the recording electrodes and the associated practical limits on the spatial extent and spatial resolution of the obtained signals.

Research on electrical probes has focused on overcoming these challenges by scaling up the number of recording sites while minimizing their invasiveness. These are inherently competing objectives as smaller probes with mechanical dimensions that minimize tissue displacement offer less surface area and volume for electrode sites, interconnects, and active circuit elements. Furthermore, as the number of electrodes increases, so does the need to bring active signal conditioning and multiplexing components closer to the brain, as the number of passive interconnects exceeds the practical limits of connector and tether cable density. This in turn introduces another dimension to the invasiveness of the recording system: the amount of heat it

dissipates into the brain tissue [4]. Chronic viability of the probes imposes additional constraints on the biocompatibility of all materials that come in direct contact with brain tissue and the flexibility of the probe itself and its coupling to the rest of the system, which must allow for brain movement relative to the skull [5; 6]. Finally, relating the extracellular signals to the underlying circuit elements requires solving an inverse problem in order to obtain a detailed current source density (CSD) estimate. The quality of the CSD estimate critically depends on the density of electrodes and their three-dimensional (3D) arrangement on a grid of known dimensions and relative position to the tissue [7]. While significant progress has been made in solving these problems individually, addressing them simultaneously within a full system has remained a challenge.

1.1   Approach and system architecture

This thesis describes the development of a modular, scalable system for dense 3-D chronic electrophysiology that addresses many of the challenges above. The front end of the system is comprised of passive high-density nanofabricated neural probes (nanoprobes, Figure 1.1)—2-D arrays of minimally invasive shanks with nanoscale interconnects—that are subsequently stacked into a 3-D array of precise geometry with over a thousand recording sites. The front-end of the system is mechanically and thermally decoupled from all active components through high-density flexible cables (Figure 1.1b), which interface the neural probes to the signal conditioning, multiplexing, and digitizing circuitry. The latter is housed on compact, lightweight PCBs, compatible with acute and chronic experimentation (Figure 1.1c).

The design of all individual components is informed by the overall system requirements and guided by the demand for an experimentally practical and configurable solution. The modular architecture adopted here offers a number of advantages. First, it decouples the neural nanoprobe front end from the digitizing back end, thus allowing multiple different probe designs to be utilized and evolved independently of the PCBs. Second, it allows for multiple, possibly different, recording layers to
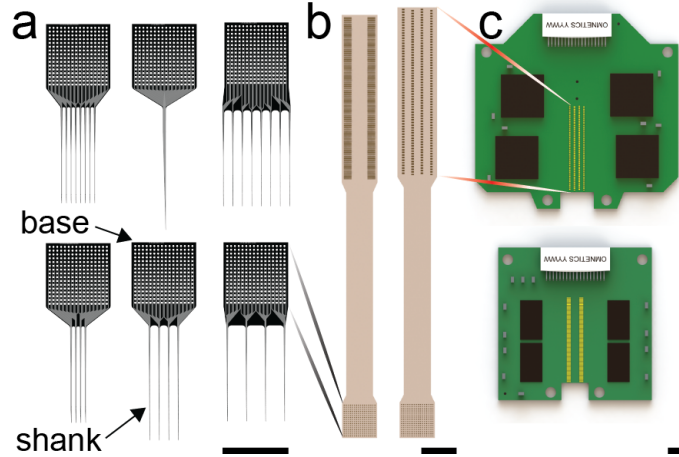
Figure 1.1: **Modular recording layer**. (a) Examples of 6 (of 10) realized neural probe designs. The number of shanks (1-8), inter-shank spacing (250-1000 m), recording site arrangement and pitch (20-65 m) are configurable. All designs support 256 electrodes per layer, connected to a standard 16x16 interconnect matrix with 200m pitch at the probe base. (b) Designs of 2 ultra-flexible cables (fabricated on either 10m thick Parylene C or 15m thick Polyimide) used to interface the neural probes to the signal conditioning PCB. (c) Designs of 2 different signal conditioning PCBs. Each performs analog signal conditioning, multiplexing, and digitization of 256 analog inputs. The top circuit employs 8 Intan RHD2132 QFN packaged chips, weighs 4.1 grams, and requires 8 output LVDS lines (16 wires), while the bottom circuit uses 4 Intan RHD2164 bare dies, weighs 1.2 grams, and requires 4 output LVDS lines (8 wires) All scale bars: 3.5mm.

be combined into 3-D arrays with controllable spacings and offsets, thereby greatly increasing the number of brain targets that can be studied. Third, it mechanically and thermally isolates the neural probes, thereby improving the biocompatibility of the system. Added together, these features endow the system with unprecedented scalability and flexibility. This is demonstrated by assembling and testing in vivo the densest 3-D array described to date.

1.2 Thesis overview

The thesis chapters describe the design, fabrication, and assembly of the system and its performance characteristics. The realized yield and quality of electrophysiological recordings are demonstrated in experiments with awake head-fixed mice.

Chapter 2 describes the analytical framework employed in guiding the design of the neural nanoprobes. It includes the following: (1) mechanical studies of shank dimensions and layer stack thicknesses compatible with straight long narrow shanks capable of withstanding implantation loads, (2) electrostatic simulations to extract lumped circuit parameters from the neural probe materials and geometry, (3) equivalent circuit modeling to infer the range of interconnect geometries compatibly with the desired signal specifications (attenuation and crosstalk) of the system, and (4) thermal modeling to establish limits on the level of heat that can be dissipated through the neural probe and to infer a power budget for the active elements in the back-end PCB.

Chapter 3 outlines the development of procedures for nanoprobe prototype fabrication at the Kavli Nanoscience Institute (KNI) foundry at Caltech and the adaptation of the fabrication process for a commercial CMOS/MEMS foundry. It includes a description of the eight different neural nanoprobe designs that were successfully built.

Chapter 4 describes the development of a modular 1024-channel data acquisition system architecture providing the back-end for nanoprobe arrays, suitable for chronic experimentation in freely behaving animals. It also outlines the design of custom 256 channel DAQ PCB headstages, which provide signal conditioning, digitization, and multiplexing of the nanoprobe outputs. Both an acute and a chronic version of the DAQ PCB headstage was designed and built.

Chapter 5 presents the methods used to package nanoprobes and DAQ PCB headstages into 2-D recording layer assemblies and then combined the 2-D recording layers into 3-D arrays. Custom high-density flexible cables are essential for interfacing the nanoprobes to the DAQ PCBs, so their design and fabrication is outlined. The tools and techniques used in the assembly process, including flex-on-board (FOB) and chip-on-flex (COF), are described and discussed.

Chapter 6 presents results from bench top testing and characterization of the nanoprobe devices and 2-D layer assemblies. The properties of the nanoprobe electrodes are measured and the system crosstalk analyzed. Protocols for establishing

open/short connections are described.

Chapter 7 presents data form acute, *in vivo* experiments with head-fixed mice in a virtual reality setup. High-quality neural recordings from single layers (256 channels) and from 3-D arrays (1024 channels) are demonstrated.

Finally, Chapter 8 concludes with remarks on where the project stands in comparison to prior efforts, and outlines some future directions. The ultimate goal is to create fully integrated devices that can be used reliably across species throughout neuroscience and are easily accessible to the community. The true impact for this system will be measured by the discoveries it can enable.

Chapter 2

NanoProbe Design Framework

2.1   Introduction

Intracranial electrophysiology is intrinsically an invasive technique, as it involves the insertion of an electrode inside the brain tissue. One essential component to reducing the invasiveness of the procedure is to minimize the physical dimensions of the portions of the neural probe that enter and reside inside the brain. While recent work has highlighted some advantages of flexible materials [8–18], silicon was selected as the substrate for the neural nanoprobes to enable high electrode count and to guarantee precise and reproducible electrode arrangements in the assembled 3-D arrays. The choice of a relatively stiff substrate dictates that electrodes should be distributed on beams (shanks) that are long enough to position electrodes within deep brain targets, while having minimal cross-sectional area, in order to minimize tissue displacement.

The lower bound on the physical dimensions of the neural probe shanks are constrained by both mechanical and electrical considerations [19]. Mechanically, the shank cross section must be sufficiently large enough to withstand the load associated with brain insertion during implantation without a buckling failure resulting in fracture. This also requires that the shank is properly stress-compensated, so that it is straight in the absence of load. Electrically, the shank dimensions must allow for the housing of a large number of electrodes and interconnects, while maintaining sufficient spacing for signal isolation between neighboring lines (Figure 2.1).

The following sections describe the analysis methodologies used to quantitatively
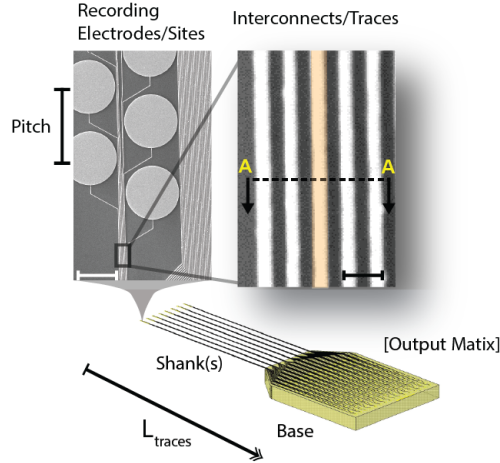
Figure 2.1: **Neural nanoprobe design parameters** *Bottom*: A neural probe consists of a base containing the output pad matrix and one or more shanks housing the electrodes and the interconnects linking them to the output pads at the base. Only the shanks enter the brain and the mechanical parameters that describe them are length, width, and thickness. *Top*: Inset on the left is an SEM of several recording electrodes with circular diameter of 12μm and their interconnects near the shank tip. SEM image on the right further enlarges the nanoscale interconnect lines, which are characterized by their width and spacing. An arbitrary interconnect at the center is highlighted in orange and is treated as the "aggressor" in crosstalk analyses.

evaluate these mechanical and electrical considerations. The buckling load is computed with Euler's formula while the stress-induced shank deformation is evaluated using the finite element method (FEM) in COMSOL. The electrostatic FEM solver in COMSOL is employed to extract the crosstalk capacitance between neighboring lines and the shunt capacitance to ground. These capacitances are then incorporated into an equivalent circuit model solved in SPICE to relate the geometric and material properties of the shank to signal quality measures, such as signal attenuation and crosstalk. The analysis shows that in order to have hundreds of electrodes on shanks that are thin (15 μm) and narrow ($< 65\mu$m) nanoscale interconnects must be used. The chapter concludes with a consideration of the thermal budget of the neural probe: the amount of heat that can be dissipated by the probe before raising the temperature of the neural tissue surrounding the shanks above a safe limit. The analysis indicates that active components, i.e. signal processing and conditioning electronics, should either use very low power (< 1 mW) or must be delocalized away from the neural

probe.

2.2   Mechanical design of the shank

Shanks are thin narrow beams that provide the mechanical substrate for both electrodes and interconnects and constitute the delivery mechanism for inserting electrodes into the brain. Ideally, shanks should be long enough to allow access to the deepest structures in the target brain, while having minimal width and thickness in order to minimize tissue displacement and trauma during implantation. As a reference, the vertical extent of the mouse brain is approximately 6 mm, while that of the rat brain is 10 mm.

During implantation, the tip of the shank is first brought in contact with the brain surface and then is slowly advanced to attempt insertion. As the initial advancement begins, the shank tip presses on the pial surface (assuming that the overlaying dura mater has already been surgically resected) producing a characteristic dimpling deformation of the brain surface. At this point the shank is axially loaded and compressively stressed, while penetration has not yet been achieved. Further advancement either results in penetration of the pia and an associated sharp decrease in the shank stress, as the mechanical resistance of the brain tissue itself is much lower than that of the overlaying membranes, or a continuing increase of the shank load. Once the load reaches a value known as the *buckling load*, the shank no longer compresses axially, but deforms laterally, i.e. buckles. Advancing the probe further quickly increases the lateral deflection and concentrates the maximal stress at the base of the probe. If penetration does not occur before the maximal stress reaches the fracture stress of the material the probe breaks and the implantation fails.

2.2.1   Shank physical dimensions

The force F_{in} required for probe insertion into the brain following dura resection has been experimentally measured to be in the range of 1–10 mN, with the precise value depending on the shape and size of the probe tip and insertion speed, as well

as on the condition of the brain tissue [20; 21]. The physical dimensions of the probe should therefore be selected so that the corresponding buckling load ideally exceeds the insertion force. If buckling is to occur, the probe should be capable of withstanding sufficient lateral deflection, so that fracture is avoided and implantation can still proceed. The maximal lateral deflection is directly related to the mechanical dimensions of the probe and the fracture stress of silicon.

To estimate the buckling load the shank is approximated with a monolithic silicon column of length L with constant rectangular cross section of width w and thickness h. The column is fixed on the base end and allowed to pivot on the tip end. Under these boundary conditions the critical buckling load is expressed by Euler's column formula

$$P = \frac{2EI\pi^2}{L^2} \tag{2.1}$$

where E is the elastic modules of silicon and I is the area moment of inertia of the cross section. For the rectangular cross-sectional area considered here, the moment of inertia with respect to the thickness axis is

$$I = \frac{wh^3}{12} \tag{2.2}$$

Combining the above equations shows that the buckling load scales linearly with shank width, as the cube of the shank thickness, and inversely as the square of the shank length. Figure 2.2a illustrates the buckling load as a function of shank width for a fixed thickness of 15μm and several different shank lengths in the range of 3–20 mm. Notice that shanks longer than 10 mm have buckling load below 1 mN even at a width of 100μm, so they would be difficult or impossible to implant.

For targets inside the mouse brain a shank length of 5–6mm is adequate. The electrical considerations mentioned in the introduction and discussed further below suggest that a shank must reach a width of 60μm in order to house over 100 electrodes and interconnects. Given these values and assuming a minimal implantation force of 1 mN, equation (2.1) shows that the shank must be at least 12–13 μm thick in order to be implantable. The actual physical dimensions of the devices that were built were

as follows: the full stack shank thickness was either 15μm or 18μm, the shank length was either 5 mm or 6 mm, and the width was tapered from about 25μm near the tip to at most 65μm at the most superficial electrode and to 100μm at the probe base in all but one design. The exception was a single shank probe featuring 256 electrodes that had to widen up to 160μm near the probe base, so that it could house all interconnects.

To estimate the maximal lateral deviation that a shank can sustain, the maximal stress experienced by a buckled probe can first be approximated by

$$\sigma_{\mathrm{max}} = \frac{6Ehu}{L^2} \tag{2.3}$$

where u is the maximal level of lateral deflection. When σ_{max} reaches the fracture stress of silicon $(1.7 \times 10^5 \text{ N/cm}^2)$ the shank breaks. Setting σ_{max} to the fracture stress of silicon and solving for u gives the maximal lateral deviation that can be sustained. For an 15μm thick shank that is 5 mm long the critical lateral deviation is 2.78 mm, so that the shank is approximately at a $45°$ angle with respect to the brain surface at the point of fracture.

2.2.2 Shank stress compensation

In the analysis above, shanks are treated as monolithic columns of silicon, but in fact they are composed of several different material layers of varying thicknesses that are sequentially deposited on top of each other during the fabrication process (Fig. 2.2b). As a consequence of the fabrication processes and the material heterogeneity, residual stresses arise and the shank is not guaranteed to remain straight once it is released from the bulk silicon substrate in the final fabrication steps [22; 23]. This represents an important failure mode, since if the shank is not straight in the absence of external mechanical force, its buckling load is effectively zero and implantation may be impossible.

Shank deformation upon release is caused by the stresses that build up in the anchored shank during fabrication. To ensure that shanks remain straight, it is
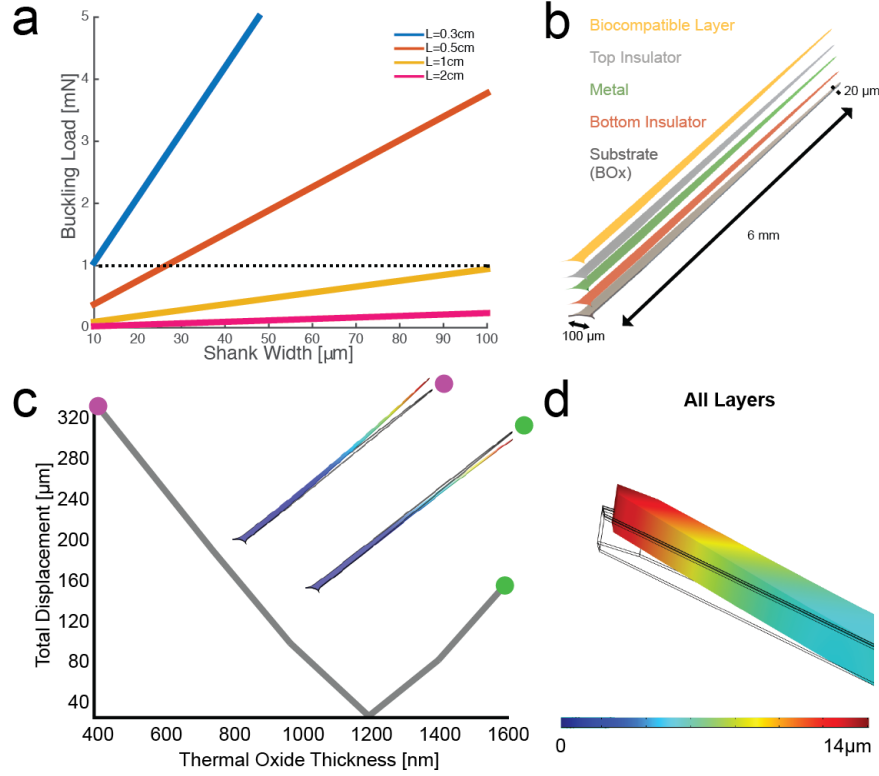
Figure 2.2: **Shank mechanical analysis (a)** Buckling load of $15 - \mu$m-thick shank as a function of shank width. Different curves correspond to different shank lengths as indicated. Interrupted horizontal line marks the minimal force needed for insertion. Shanks with dimensions below this line will buckle during implantation. **(b)** Exploded view of the material layer stack and shank geometry used in the stress compensation analysis. **(c)** Shank tip displacement from the probe base plane as a function of bottom insulator thermal oxide thickness. Straight shanks have 0 displacement. Substrate profile: 17 μm thick silicon, $900 nm$ thick BOx layer. **(d)** Optimal displacement in a simulation using the full layer stack and probe 3-D geometry. Black contour corresponds to perfectly straight shank. Under optimal bottom insulator thickness the tip shank deflection is 14μm upward, so the shank axis is off by less than 0.14°.

essential to consider the impact of each added material layer on the total stress within the structure. These residual stresses are either intrinsic or thermal in nature. Intrinsic stresses arise during the deposition process and are present at the deposition temperature. In contrast, thermal stresses arise because of the different coefficients of thermal expansion (CTE) of the layers in the stack. For example, during fabrication dielectric layers are deposited onto the silicon wafer at elevated temperatures. As the wafer is brought back to room temperature, thermal stress builds up because the

silicon substrate and the newly deposited thin film have different CTEs. This residual thermal stress can be expressed as

$$\sigma = \frac{E}{1 - \nu}(\alpha_{Si} - \alpha)(T - T_d) \tag{2.4}$$

where T is room temperature, T_d is the temperature of deposition, α_{Si} is the CTE of the silicon substrate, and the remaining parameters describe the thin film material, i.e. α is CTE, E is Young's modulus, ν is Poisson's ratio.

The magnitude and direction (tensile, compressive) of intrinsic stresses depend heavily on the stoichiometry of the deposited material, as well as on the nature and the details of the deposition process conditions. For example, the intrinsic stress for silicon nitride (SiNx) deposited using plasma-enhanced chemical vapor deposition (PECVD) can range from being tensile (+500MPa) to being compressive (-1250MPa) depending on the deposition power [24].

Given that the shank is a multilayer structure (Fig. 2.2b), the combination of thermal and intrinsic stresses introduced by each layer has to be properly balanced [25]. Once the shank is fully released from the silicon wafer, the residual stresses should add up to zero and cancel out or the shank will deform. To properly engineer the shank layer stack, the effects of layer thickness and deposition parameters on shank displacement was investigated. This elastostatic analysis enabled the fabrication of devices using layer thicknesses that minimize shank displacement.

The governing equations, in tensor form, of the linear elastostatic boundary value problem are:

$$\nabla \cdot \boldsymbol{\sigma} + \mathbf{F} = 0 \tag{2.5}$$
$$\boldsymbol{\varepsilon} = \frac{1}{2}(\nabla \mathbf{u} + \nabla \mathbf{u}^T) \tag{2.6}$$
$$\boldsymbol{\sigma} = C : \boldsymbol{\varepsilon} \tag{2.7}$$

where $\boldsymbol{\sigma}$ is the stress tensor, $\boldsymbol{\varepsilon}$ is the strain tensor, $\mathbf{u}$ is the displacement vector, $\mathbf{F}$ is the body force, and C is the 4th order stiffness tensor. The last expression is the

constitutive equation for elastic materials (Hooke's law), relating stress and strain via the linear map C, where ":" stands for the double-dot tensor product.

To carry out the analysis, a 3-D geometric model of the shank was build using COMSOL (Fig. 2.2b). Each layer was treated as a homogeneous, isotropic, linear thermoelastic material and could therefore be described in terms of its Young's modulus E, Poisson's ratio ν, and coefficient of thermal expansion α (Table 2.1). Each layer was added iteratively and in order to account for the thermal stresses an extension of Hooke's law was used as the constitutive equation. In particular, the generalized Duhamel-Neumann linear thermoelastic equations was used, given by

$$\boldsymbol{\sigma} - \boldsymbol{\sigma_0} = C : (\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon_0} - \alpha(T - T_d)) \tag{2.8}$$

where $\boldsymbol{\sigma_0}$ and $\boldsymbol{\varepsilon_0}$ are the initial stress and strain, respectively, α is the material COT and T_d is the deposition temperature. It is assumed that thermal equilibrium has been reached and the temperature T is constant throughout. At each iteration the system was solved using boundary conditions on the displacement that fixed all shank walls in space to model the fabrication process where during deposition the probe is physically anchored to the bulk silicon handle wafer and cannot deform. At the end of the iterations, once all material layers were introduced, the system was solved again with different boundary conditions where only the wall of the shank that connects to the probe base was fixed in space and the rest of it was free to deform. The deviation of the shank tip from the neutral plane was used as a scalar measure of deformation (Fig. 2.2d).

The thermally grown oxide layer, which served as bottom insulator (Fig. 2.2b), was chosen to balance the residual stresses in the shank. So the above analysis was repeated for different values of the thermal oxide thickness and the shank deflection was extracted for each thickness value (Fig. 2.2c). This allowed for the selection of 1.2μm thick thermal oxide as fabrication target with predicted deflection of only 14μm (Fig. 2.2d).

Table 2.1: **Mechanical simulation parameters**: layer thickness **t**, deposition temperature T_d, Young's modulus **E**, density ρ, coefficient of thermal expansion (CTE) α.

Material	t	T_d	E	ν	ρ	α
units	nm	$C°$	GPa		kg/m^3	$\times 10^{-6} K^{-1}$
Substrate:						
Silicon	15000	n/a	170	0.28	2329	2.1
BOx	900	n/a	70	0.17	2200	0.5
Bottom Insulator:						
Thermal Oxide	400-1800	1050	70	0.17	2200	0.5
Interconnects:						
Electroplated copper	300	25	120	0.34	8960	16.5
Top Insulator:						
PECVD SiO2	500	400	70	0.17	2200	0.5
Recording sites:						
Electroplated gold	300	100	79	0.44	19000	14

2.3   Electrical design of the shank

The purpose of the shank is to carry as many electrodes as possible inside the target brain area, so that the extracellular potential generated by brain activity can be sensed with high spatial resolution. This is important for isolating the spiking of individual neurons. Indeed, a prerequisite for triangulating a source neuron is that its spikes are sensed with sufficient amplitude on several electrode sites simultaneously [2; 26].

Since the extracellular potential is effectively averaged over the electrode surface, the physical dimensions of the electrode should be minimized [27]. However, the intrinsic thermal noise of the electrode increases with its impedance, which is inversely related to the electrode size. For example, it is estimated that a planar electrode with

$16\mu m^2$ area will have thermal noise of 20μV [28], so this provides a lower bound on the electrode size.

Furthermore, the sensed potential must faithfully reach the amplification stage, so interconnect parameters must be carefully considered to minimize harmful electrical parasitic effects on the signal[28–30]. These include crosstalk coupling between adjacent interconnects, signal attenuation due to shunting to ground, and the addition of thermal noise due to line resistance. Scaling up the recording electrode density also increases the associated interconnect density. As interconnects are brought closer together their coupling capacitance increases allowing signals carried on one line to be sensed on neighboring lines resulting in crosstalk [31].

Crosstalk is reduced by decreasing the coupling capacitance between interconnects. The relationship between interconnect geometry and coupling capacitance is quantitatively studied to inform the interconnect design. The degree to which coupling capacitance translates to crosstalk depends on other system parameters, such as electrode impedance, shunting capacitance, amplifier inputs impedance, etc. These relationships are explored with an equivalent circuit model to guide the selection of probe design parameters from a systems perspective.

2.3.1 Interconnect parasitics

Consider two interconnect lines of width w, thickness t, and length L, running parallel to each other with spacing d over a dielectric layer of thickness t_{btm} (Fig. 2.3a). The dielectric layer separates the lines from a grounded silicon substrate. Using the infinite parallel line approximation, the coupling capacitance between the lines per unit length can be expressed as

$$C_{coupling} \approx \varepsilon_r \frac{hL}{d} \tag{2.9}$$

where ε_r is the permittivity of the dielectric material. Similarly, the shunting capacitance to ground is

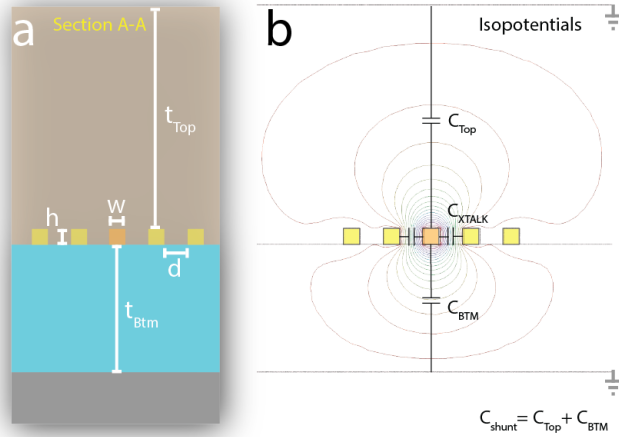$$C_{shunt} \approx \epsilon_r \frac{Lw}{t_{btm}} \tag{2.10}$$

Figure 2.3: **Electrical parasitics and interconnect dimensions**. **a.** Shank cross section reveals 5 interconnects of width w, height h, and spacing d, insulated above and below by dielectric layers of thickness t_{top} and t_{btm}, respectively. The highlighted interconnect in the center is treated as the "aggressor", while the neighboring lines are the "victims". **b.** The aggressor and victim lines form a parasitic coupling capacitance C_{xtalk} (or $C_{coupling}$), while the aggressor and the dielectrics form a shunting capacitance to ground (C_{shunt}). Following [19], the values of these capacitances are first approximated with an analytical expression and then extracted from an electrostatic FEM simulation in COMSOL.

Assume that potential V is applied to one of the two lines, referred to as the "aggressor". The level of crosstalk can be quantified by the potential V_{xtalk} that develops across $C_{coupling}$ on the other line, referred to as the "victim". The potential on the victim line can be evaluated by noticing that the coupling and shunting capacitances form a voltage divider, so the expression is

$$V_{xtalk} = \frac{C_{coupling}}{C_{coupling} + C_{shunt}} V \qquad (2.11)$$

Taking the ratio of the potential on the victim and aggressor lines gives the crosstalk as a fraction

$$\frac{V_{xtalk}}{V} = \frac{1}{1 + C_{shunt}/C_{coupling}} \qquad (2.12)$$

It is clear from the above equation that in order to minimize crosstalk the ratio of shunting to coupling capacitance should be as large as possible. Keeping in mind

that increasing the shunting capacitance leads to signal attenuation the objective is to reduce the coupling capacitance. Substituting eqs. 2.9, 2.10 above expresses the ratio in terms of the interconnect parameters

$$\frac{C_{shunt}}{C_{coupling}} = \frac{dw}{ht_{btm}} \tag{2.13}$$

Notice that the interconnect length L is no longer part of the expression, so the crosstalk can be understood in terms of the cross-sectional dimensions only. Increasing the interconnect width w and decreasing the dielectric thickness t_{btm} reduces crosstalk, but at the expense of increased shunting capacitance (eq. 2.10) and associated signal attenuation. Similarly, lowering the interconnect height h reduces crosstalk, but at the expense of increased thermal noise. This is because the resistance R of the line increases with reduced cross sectional area (wh) according to

$$R = \rho \frac{L}{wh} \tag{2.14}$$

where ρ is the resistivity of the interconnect material, while the thermal noise grows with the resistance

$$\delta V = \sqrt{4k_B R T \Delta F} \tag{2.15}$$

where ΔF is the bandwidth, T is the temperature, and k_B is the Boltzmann constant. Finally, the coupling capacitance can be controlled by the spacing of the interconnects d. Increasing d has no other drawbacks than increasing the real estate taken by the interconnects and thereby limiting the interconnect density.

In order to support many electrodes on shanks of minimal width the interconnect pitch must be small. As the interconnect pitch $(d + w)$ is reduced, coupling capacitance is more and more dependent on fringing effect [32; 33]. Consequently, further reduction of interconnect height h no longer decreases crosstalk and the parallel line approximation (eqs. 2.9, 2.10) is no longer accurate.

To extend the analysis to this regime the coupling and parasitic capacitances must be extracted from the behavior of the electromagnetic field in the multiple conductor

cross section illustrated in Fig. 2.3a. The capacitive behavior of a multi-conductor system can be characterized through a capacitance matrix C that relates the potential of each conductor (with respect to a point at infinity) $V = [V_1 \ldots V_n]^T$ and the total conductor charges $Q = [Q_1 \ldots Q_n]^T$, i.e.

$$CV = Q \tag{2.16}$$

The entries in C are obtained by solving a series of electrostatic problems, which was carried out in COMSOL by a finite element method (FEM) in a 2-D model of the cross section. In order to get the ith column of C, the potential Φ must be obtained throughout the domain with the boundary condition that the potential is 1 Volt on the surface of the ith conductor and 0 on all remaining conductors, i.e. $V = [\cdots 0\ 1\ 0 \cdots]^T$. Because there are no charges present in the dielectric ($\rho = 0$) the potential is given by

$$\nabla \varepsilon \cdot \nabla \Phi = 0 \tag{2.17}$$

where ε is the permittivity of the medium. This is a generalized form of Laplace's equation $\nabla^2 \Phi = 0$ for non-constant permittivity. The contour lines in Fig. 2.3b illustrate one solution of eq. 2.17. Once the potential is available throughout the domain, the charge at each conductor i is obtained using Gauss' law through the surface integral

$$Q_i = \int_{S_i} -\varepsilon \nabla \Phi \cdot dS \tag{2.18}$$

where S_i is a closed surface around conductor i and dS is the outward normal on S. Once the vector Q is available the result follows, since given the boundary conditions the product CV gives the ith column of C.

The results of the electrostatic FEM study are illustrated in Fig. 2.4 summarizing the extracted coupling and shunting capacitance values as a function of interconnect pitch. Notice the large effect fringing has on the coupling capacitance when line spacing decreases below the line width of 100 nm. The effect is especially prominent for thick interconnects with height-to-width ratio > 1. As expected, coupling
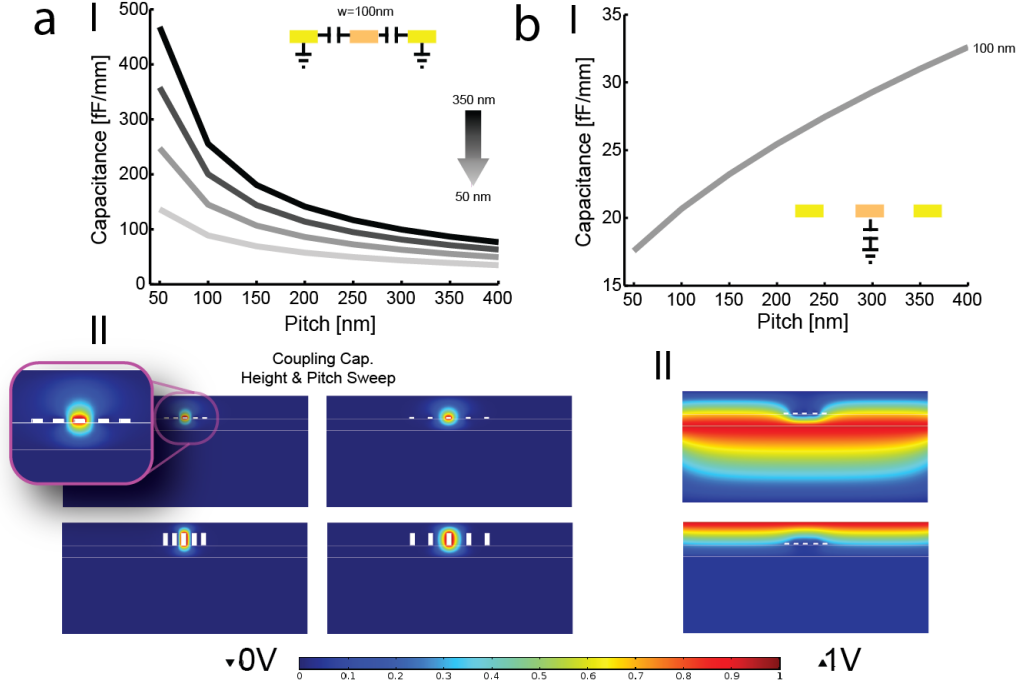
Figure 2.4: **Interconnect parasitics extracted from electrostatic FEM solutions**. **a**. Coupling capacitance as a function of interconnect pitch. Line width is fixed to $100\,\mu$m and line spacing d is varied. Different curves correspond to four interconnect heights h: 50, 150, 250, 350 nm. (*bottom*) Electric field Φ when the aggressor line in the center is held at 1 Volt. Several interconnect configurations are illustrated. **b**. Shunting capacitance between aggressor interconnect ($w = h = 100$ nm) and ground as a function of pitch. (*bottom*) Electric field Φ when either the top or bottom conductor bounding the domain is held at 1 Volt.

capacitance is reduced for interconnects with smaller height. Shunting capacitance decreases with reduced interconnect spacing, as more of the electric field couples to neighboring interconnects instead of the ground planes.

Table 2.2: Dielectric permittivity

	ε	**Value**
Bottom insulator		3.9
Silicon		11.8
Top insulator		3.1

2.3.2   Equivalent circuit model

The analyses in the preceding section establish a relationship between the physical dimensions of interconnects and parasitic capacitance magnitudes. However the acceptability thresholds for these parasitics can only be evaluated in the context of the full system, as other system components interact to influence the overall signal quality. In particular, the impedance properties of the recording electrode and the input impedance of the amplifier are essential in determining the level of crosstalk and signal attenuation. Therefore, an equivalent circuit model of the recording system was constructed and studied.
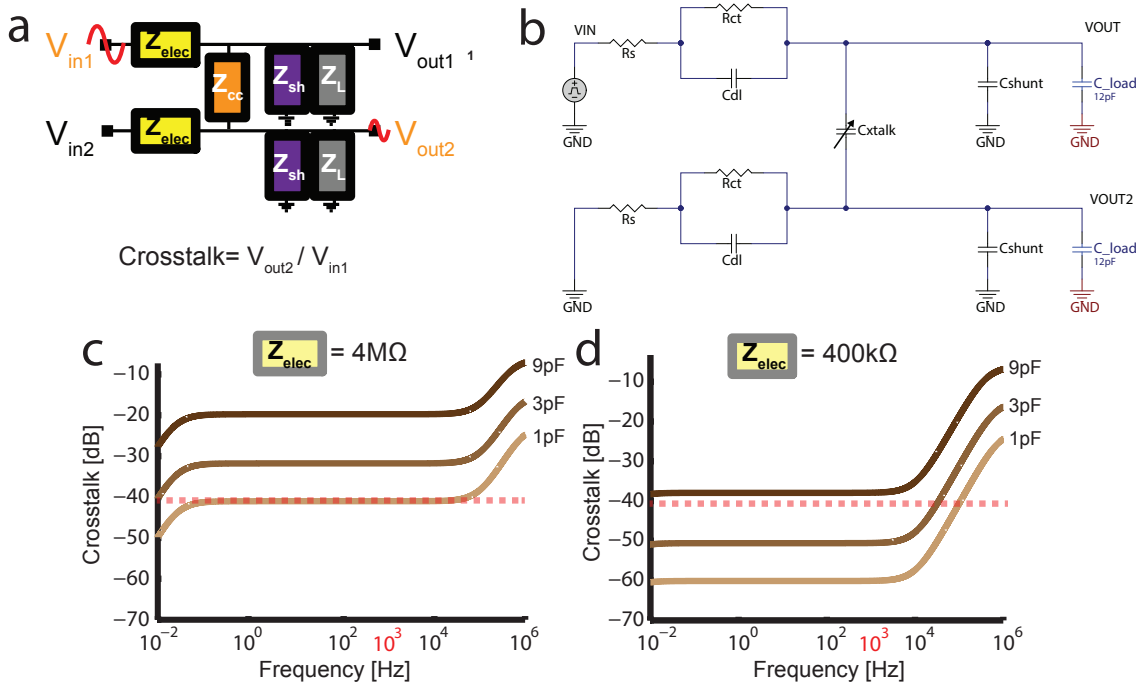

Figure 2.5: **Equivalent circuit study of crosstalk in the recording system.** **a**. Equivalent circuit representation of two recording channels with adjacent interconnects featuring electrode-electrolyte impedance Z_{elec}, parasitic coupling capacitance Z_{cc}, parasitic shunting capacitance to ground Z_{sh}, and amplifier input (load) impedance Z_L. **b**. Actual schematic of the model in (a) used for parameter-sweep simulations in SPICE. The electrode-electrolyte impedance Z_{elec} is represented by Randles circuit containing charge transfer resistance R_{ct}, double layer capacitance C_{dl}, and spreading resistance R_s. **c**. Crosstalk as a function of frequency for an electrode with 4 MΩ impedance at 1 KHz. Different curves correspond to three values of the coupling capacitance (1, 3, and 9 pF). Interrupted horizontal line marks crosstalk of 1% (-40dB). **d** Same as (c), but for an electrode with 400 KΩ impedance at 1 KHz.

The equivalent circuit models two recording channels with adjacent interconnects (Fig. 2.5a,b) and crosstalk is assessed by the ratio of the signal magnitudes at the output of the "victim" line and the input to the "agressor" line, with the input to the "victim" line grounded (V_{out2}/V_{in1} with $V_{in2} = 0$). The interplay between coupling capacitance and electrode impedance can be intuitively understood by considering a version of the circuit with an idealized amplifier (having infinite input impedance) and in the absence of shunting capacitance. Under these conditions Z_{sh} and Z_L are both eliminated from the circuit in figure 2.5a and what remains is a simple voltage divider. Consequently the crosstalk can be expressed as

$$\frac{V_{out2}}{V_{int1}} = \frac{Z_{el}}{2Z_{el} + Z_{cc}} = \frac{1}{2 + Z_{cc}/Z_{el}} \approx \frac{1}{2 + C_{dl}/C_{coupling}} \tag{2.19}$$

where the last approximation treats the electrode simply as a double-layer capacitor. It is now evident that crosstalk depends on the ratio of the electrode double-layer capacitance to the parasitic coupling capacitance, rather than the absolute value of either. Thus systems with lower impedance electrodes (higher C_{dl}) can tolerate proportionately more stray coupling capacitance $C_{coupling}$ than systems with high impedance electrodes.

The same conclusion is recapitulated in the analysis of the full equivalent circuit (Fig. 2.5c,d). The parameter definitions and choice of values used in the SPICE simulations of the model are summarized below:

R_s **solution resistance** In Randles circuit of the electrode-electrolyte interface models the series resistance between the recording and reference electrodes introduced by the electrolyte solution. It is usually determined from the spreading resistance according to the formula $R_s = \rho/(4r)$, where $\rho = 72\,\Omega$cm is the resistivity of physiological saline and $r = 6\,\mu$m is the radius of a circular electrode (geometric area of $113\,\mu$m^2). These numbers give a value of 30 KΩ and in simulations the commonly cited value of 10 KΩ was used [34–36].

R_{ct} **charge-transfer resistance** In Randles circuit it models the ability of the elec-

trode to carry Faradaic currents. The value depends on the electrode surface area and material. It is estimated to be roughly $10^6\,\mathrm{M\Omega \cdot \mu m^2}$ for gold, corresponding to 10 GΩ for the $100\,\mu\mathrm{m}^2$ electrodes considered here. Empirically, values closer to 1 GΩ are commonly cited for microelectrodes.

C_{dl} **double-layer capacitance** In Randles circuit it models the electrode-electrolyte double layer capacitance. The value depends on the electrode surface area and material. It is estimated to be roughly 0.4 pF/$\mu\mathrm{m}^2$ for gold [36], corresponding to 40 pF for an unplated $100\,\mu\mathrm{m}^2$ electrode. Notice that the surface area and double-layer capacitance can be increased by an order of magnitude through electroplating, while keeping the geometric area of the electrode unchanged.

C_{xtalk} **coupling capacitance** Models the interconnect parasitic coupling capacitance $C_{coupling}$. The value depends on the interconnect geometry and based on the FEM analysis in figure 2.4 is in the range of 0.15–0.5 pF/mm for 100 nm interconnects. Assuming 10 mm long shanks gives the range of 0.5–5 pF. Accounting for the full interconnect path, which includes the probe base, flex cable, and PCB, approximately doubles the range to 1–9 pF.

C_{shunt} **shunting capacitance** Models the interconnect parasitic coupling to ground. Based on the FEM analysis in figure 2.4 the value is in the range of 20–30 fF/mm, or 0.2–0.3 pF for 10 mm long lines. However, significant additional parasitics arise at the probe base which features 100 μm diameter output pads and much wider interconnects. These increase the value by an order of magnitude to 2 pF.

C_{load} **amplifier input capacitance** Models the input impedance of the amplifier, which according to the Intan specifications has the value of 12 pF.

Notice that according to the full model SPICE simulations a system with a 4 MΩ electrode (100 μm^2, unplated) cannot tolerate more than 1 pF of parasitic coupling capacitance, whereas a system with 400 KΩ electrode (100 μm^2, plated) can tolerate almost 9 pF (Fig. 2.5c,d). In comparison, according to the analysis in the

previous section, 10 mm of nano-scale interconnects are expected to have a coupling capacitance in the range of 1.5–5 pF, depending on their precise height and spacing (Fig. 2.4). This highlights the importance of post-processing the nanoprobe microelectrodes to lower their impedances, which can be achieved by electrodeposition of gold, platinum, as well as other materials.

2.4 Biocompatibility design

One of the great challenges to establishing and maintaining high quality neuronal recordings is not electrical or mechanical, but involves the biological response to the neural probe implanted in the brain [37–39]. This is especially relevant for chronic applications. By definition biocompatibility is concerned with the biological, chemical, and physical properties of an implant and it can be considered truly biocompatible only if meets a set of requirements [40–43]. First, the implant must be composed of materials that are not toxic or react adversely in the brain. Second, it must not cause an immune reaction. Third, it should not harm or destroy enzymes, cells or tissues. Finally, the implant must be suitable for chronic application. In practice, there is usually evidence for a biological response to the implant, but the tissue should produce only mild foreign body reaction and encapsulation, without evidence of necrosis or implant rejection [40–43].

A common approach to improving the chronic viability of an implant is to coat it with a biocompatible material [23]. Parylene was selected as the outer layer covering the shanks and the reasons behind this choice are summarized below. However, an adverse tissue response may not be triggered by the materials that come in contact with the tissue, but also by the energy the implant dissipates in the tissue as heat. The final section quantifies this energy/thermal balance and discusses its implications for the architecture of the recording system.

2.4.1 Parylene as biocompatibility layer

Parylene has been approved according to USP VI as material for chronic implants and has shown no cytotoxicity according to ISO 10993. Parylene thin-films are very good insulators against both electrical and thermal conduction. Parylene is also very chemically stable and insoluble in all organic solvents. It is resistant to chemical attack at room temperature from most reagents. In addition, it has very low gas permeability.

Another important property of Parylene is its biocompatibility and biostability. Parylene film is formed from a pure molecular precursor (a monomer gas) in modest vacuum. In these deposition conditions there is very low chance for the Parylene coating to be contaminated by any foreign particles or chemical substances. Unlike spin coated polymer films, Parylene films do not suffer from chemical outgassing (e.g solvent, plasticizer, catalyst, or accelerant). In biomedical applications the resulting films have demonstrated very low thrombogenic properties and low potential to trigger immune response. Parylene's high chemical resistance also ensures its stability in corrosive body fluids, electrolytes, proteins, enzymes, and lipids. Thus, Parylene coating works as a very effective barrier in biological settings, not only protecting the substrate from the environment, but also protecting the biological milieu. In the case of implant coating, the passage of potential contaminants from the substrate is blocked [38; 44].

Several sterilization methods can be applied to Parylene enabling its use for medical applications. These methods include steam autoclaving, gamma and e-beam radiation, and sterilization via hydrogen peroxide plasma or ethylene oxide. One of drawbacks of Parylene coatings compared to inorganic coating, however, is its low thermal stability. Parylene has melting points in the range of 300–400 °C and glass transition temperatures of 80–150 °C that are relatively low. Recently, Parylene HT, with improved thermal properties, was commercialized by Specialty Coating Systems and was selected as the coating material for the neural probe shanks.

2.4.2 Heat dissipation limits

Excessive heating in the brain can produce profound negative effects on neuronal function [45](Figure 2.6a). Therefore, it is critical to consider the effects of power dissipation from active electronics on brain temperature. There are benefits to bringing signal amplifiers close to recording electrodes as this minimizes the length of passive interconnects, thereby reducing parasitics, noise susceptibility, and crosstalk. Consequently, the effects of placing an active electronics (ASIC chip) directly to the base of the neural probe were investigated.

In particular, the impact of this implant architecture on the temperature of the brain was modeled in COMSOL. Figure 2.6b shows the geometry used for the simulation. It consists of an ASIC chip, acting as a power source, packaged directly on top of the neural probe base. The shanks of the neural probe are implanted into brain tissue, modeled as a cylindrical volume. The base of the neural probe is surrounded by air. The temperature at the boundary of the cylindrical volume representing the brain was fixed at 37 °C, while the portion of the domain representing the air above the brain was open to exchange heat with the surroundings held at 25 °C.

Pennes bioheat equation [4] was used to describe the heat transfer in the brain domain by considering the effects of blood perfusion (cooling), metabolism (heating), and power dissipative elements (heating):

$$\rho_b C_P \frac{\partial T}{\partial t} = \nabla(k_b \nabla T) + Q_{asic} - Q_{bp} + Q_m \tag{2.20}$$

where the brain tissue parameters are density ρ_b, specific heat C_P, thermal conductivity k_b, T is the temperature and the three heat sources model the ASIC Q_{asic}, blood perfusion Q_{bp}, and metabolism Q_m. Q_{bp} models the cooling effect of blood perfusion in brain tissue. It is significant considering the brain takes in 20% of cardiac output. The amount of heat blood will remove from the brain through convection is expressed by

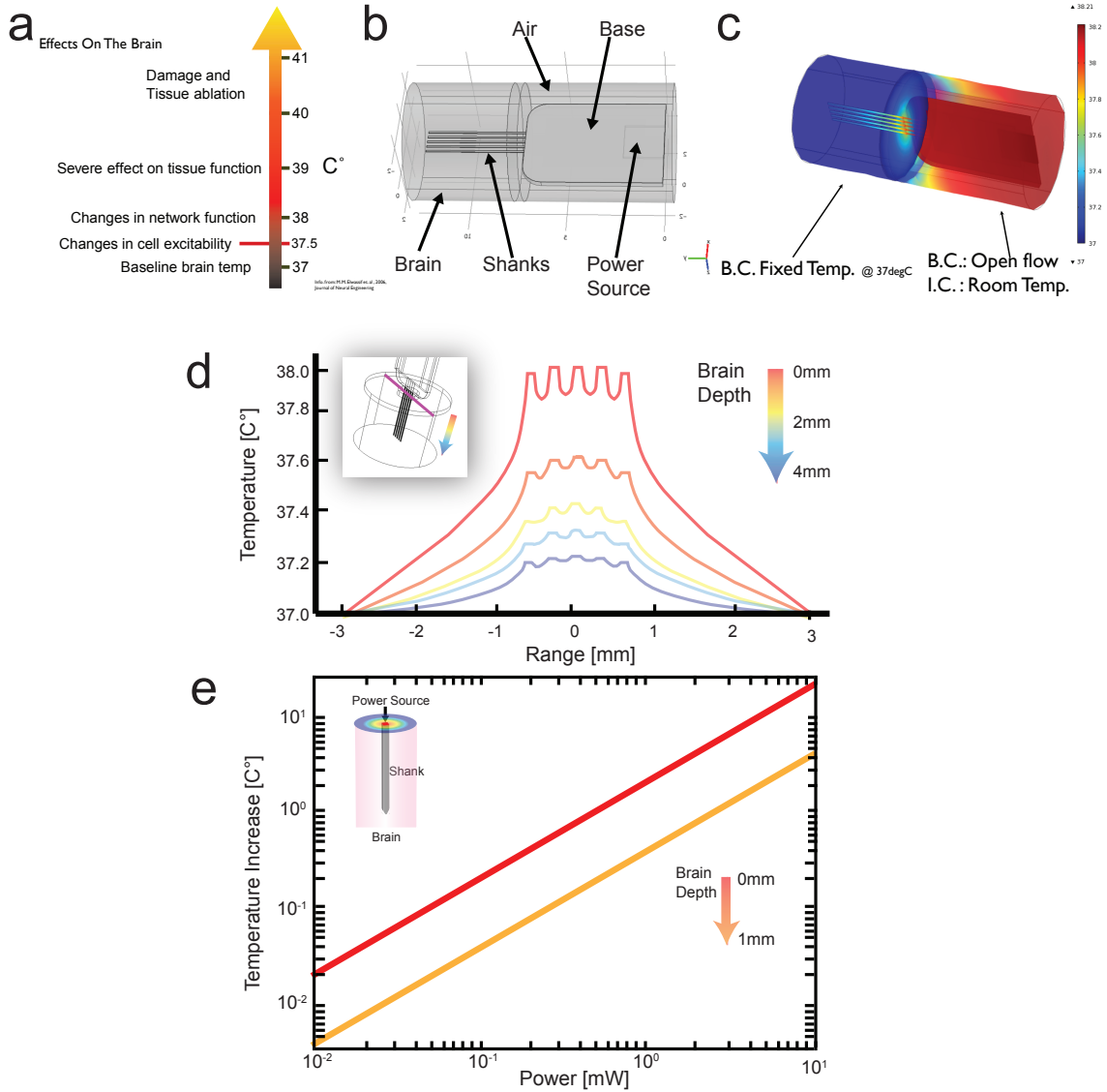$$Q_{bp} = \rho_{bl} \omega_{bl} C_{bl}(T - T_{bl}) \tag{2.21}$$

Figure 2.6: **Bioheat model of power dissipation in brain tissue.** (a) Effects of elevating brain temperature. Adapted from [4]. (b) Domain geometry used in COMSOL simulation. Shanks are 5mm long. (c) Boundary and initial conditions used for simulations. (d) Temperature profiles along line segments adjacent to the shanks at different depths (e) Power sweep assuming active electronics are embedded in the shank. Temperature taken 50 μm away from the shank.

where ρ_{bl} is the blood density, ω_b is the volumetric blood perfusion rate, C_{bl} is the heat capacity of blood, and T_{bl} is the body core temperature. Q_m represents the heat generated by normal metabolic brain activity and was set to 0 in this study. Q_{asic} represents the heat dissipated by the ASIC. Since the power consumption of a

32-channel Intan bioamplifier is 15.6 mW and the die volume is 3x3x0.1 mm^3, the value of 17.33 mW/mm^3 was used.

Figure 2.6d shows the temperature profile across horizontal line segments in a plane adjacent to the shanks. Each line segment is at a different depth (0–4 mm) and passes near each shank, with corresponding peaks in the temperature profiles. Notice that the increase in steady-state temperature at the surface of the brain has reached the 1°C threshold. The cooling effect of blood perfusion is seen in the deeper curves, as evidenced by the decrease in steady-state temperatures. However, even 1 mm deep inside the brain the temperature near the shanks is increased by 0.6°C, which can affect neuronal excitability, and bias the recorded data. Notice that the heat source used to model the ASIC represents the power dissipation for a single 32-channel Intan bioamplifier. Clearly, further scaling of channel count using this chip and packaging method is not advisable.

What is the upper limit on power dissipation per shank? This questions was studied by constructing a simpler model consisting of an implanted shank with a boundary heat source at the base wall (inset in Fig. 2.6e). Again, the brain was modeled as a cylinder ($r = 5$ mm, $h = 9$ mm), and the fully-implanted shank as a rectangular slab ($w = 50\,\mu$m, $h = 5$ mm, $t = 15\,\mu$m). It was further assumed that the top surface of the brain is insulated.

Figure 2.6e summarized the results of this study. The brain initial temperature was set to 37 °C and metabolic heat production was ignored. Power was swept from 10 μW to 10 mW and the associated change in the brain temperature computed at a distance of 50 μm away from the shank both at the surface and 1 mm inside the brain. Figure 2.6e demonstrates that as little as 0.3 mW of power dissipated through the shank can raise brain temperature by 1 °C at the surface and 3 mW are sufficient to do the same 1 mm inside the brain.

Another approach to estimating the power budget is based on heat flux guidelines. For implants located in muscle and lung the upper limit for the heat flux is 40 mW/cm^2 (0.4 mW/mm^2) [46]. While no equivalent guidelines have been established for brain tissue yet, it is likely that the number is lower, given the well recognized

sensitivity of neural tissue to heat. The surface area of a shank that is 5 mm long, 50 μm wide, and 20 μm thick is 0.7 mm^2. Taking the heat flux limit above translates to a power budget upper bound of 0.28 mW per shank, which agrees well with the estimate derived from the model.

Table 2.3: Bio-heat simulation parameters used for the brain domain

Property	Value
k_b	0.5 $[W/m^\circ C]$
ρ_b	1040 $[kg/m^3]$
C_P	3650 $[J/kg^\circ C]$
T_i	37 $[^\circ T]$
ω_{bl}	0.004 $[ml/scm^3]$
ρ_{bl}	1057 $[kg/m^3]$
C_{bl}	3600 $[J/kg^\circ C]$
T_{bl}	37 $[J/kg^\circ C]$

2.5   Conclusion

The analyses presented in this chapter lead to several important conclusions. The mechanical studies show that minimally invasive shanks of 15 μm thickness and 50 μm width are sturdy enough for implantation as long as their length is in the order of 5 mm. Doubling the length to 10 mm translates to a buckling load well below the force needed for implantation.

The interconnect studies indicate that lines as narrow as 100 nm are admissible, as long as the line spacing is at least as large as the line width. Similarly, the interconnect height should not exceed its width. This implies that at least 200 nm of the shank width is required per electrode, so if all interconnects are routed in the same plane a 50 μm wide shank can house at most 250 electrodes. The electrodes themselves should have sufficient surface area, so that their impedance relative to the interconnect parasitics is low and does not result in excessive crosstalk between neighboring signal

lines. Given the equivalent circuit model, unplated electrodes cannot be smaller than $100~\mu m^2$ without leading to crosstalk exceeding 1%. Electroplating the electrodes is essential for providing a tolerance margin for the stray capacitance in the system.

Finally, the thermal modeling gives specific power limits for all electronics directly coupling to the neural probes. Specifically, the limit is estimated at 0.3 mW per shank. For a single shank neural probe with 250 electrodes this translates to a power budget of 1.2 μW per electrode. This is in the order of the power requirements for a state-of-the-art ultra low power analog bioamplifier. Since the active electronics must at a very minimum feature high frequency analog multiplexers, the power budget is quickly exceeded. This argues for delocalized the active electronics away from the neural probe, which is the system architecture adopted here.

Chapter 3

NanoProbe Fabrication

3.1 Introduction

Electrophysiological recordings have been instrumental for the study of neural circuits. Early efforts relied on microelectrodes fabricated in individual laboratories, making reproducibility and precise dimensional control difficult. Kensall Wise and his colleagues pioneered the use of integrated-circuit technology for neural probe fabrication [47–49], opening a path for large-scale production of probes with precise and reproducible dimensions. Silicon based neural probes have since matured into a powerful tool that has been widely used in systems neuroscience research [50–57].

The need to maximize the number of recording sites across the shanks of neural probes has long been appreciated, since brain computations emerge from the coordinated activation of large distributed assemblies [1; 58]. As discussed in chapter 2, scaling up the number of recording sites introduces important design constraints that must be strictly met during probe fabrication. This chapter describes the fabrication approaches that were employed to realize the neural probes.

The fabrication efforts described here have benefited from significant advances in several domains. The CMOS industry has pushed the frontier of small feature size in order to pack in more transistors and more computing power into smaller microchips. The IC fabrication node has shrunk from $10\,\mu$m in 1971 down to 14 nm today. Reliability and yield have improved and costs have decreased. Furthermore, advances in the MEMS industry have provided powerful fabrication techniques to form

very small, high-aspect ratio mechanical structures. These techniques are classified as either bulk or surface micromachining. The former involves using various etches to remove silicon from a substrate to form the desired device. The latter involves depositing material onto a silicon substrate and removing it to form the desired device.
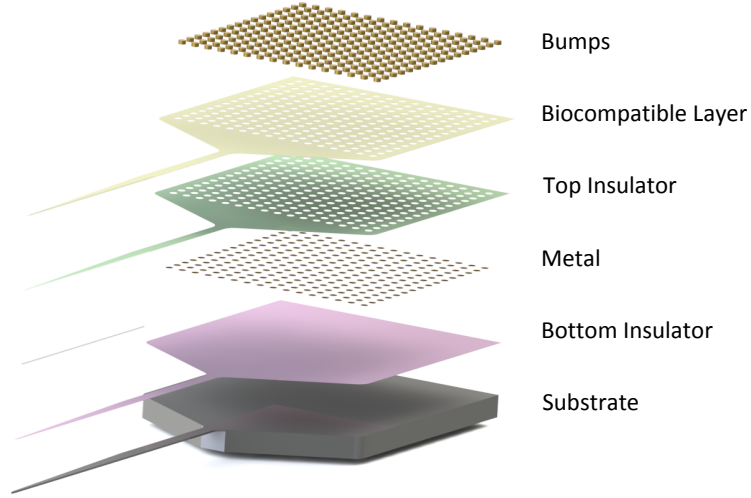

Figure 3.1: **Neural nanoprobe layer stack**. Exploded view of the layers that are stacked on top of each other to form the neural probe. Each layer has a distinct function, material makeup, and is usually created in separate fabrication steps.

To build the neural nanoprobe devices, techniques combining both CMOS and MEMS fabrication were leveraged. Figure 3.1 shows the layer stack makeup of the neural probe. The functionally critical stack component is the metal interconnect layer, which was patterned using integrated circuit (IC) techniques. Surface micromachining was employed to deposit the top insulator and biocompatibility layers, required to isolate the metal interconnects from the brain and to protect the neural tissue from the implant. Bulk micromachining was used to pattern the substrate layers into shanks connected to the probe base. Finally, to allow integration of the device within a functional system, an output matrix of metal bumps was created to electrically interface the neural probe to the signal conditioning electronics via flip-chip bonding. The packaging and system integration strategies are described separately in Chapter 5.

Guided by the design rules derived in the previous chapter, two generations of neural probes were designed and prototypes built at the Kavli Nanoscience Institute (KNI) cleanroom at Caltech (Figure 3.2). The first generation of devices (Gen1, Figure 3.2a,b) was designed and built primarily for fabrication development purposes. The most important of the fabrication procedures developed were the use of electron beam lithography to pattern the nanoscale interconnects, silicon etching techniques to define the shanks, i.e. reactive ion etching (RIE), deep reactive ion etching (DRIE), cryogenic etching, as well as Parylene C processing techniques to coat the top surface of the shanks and to create the electrode openings.
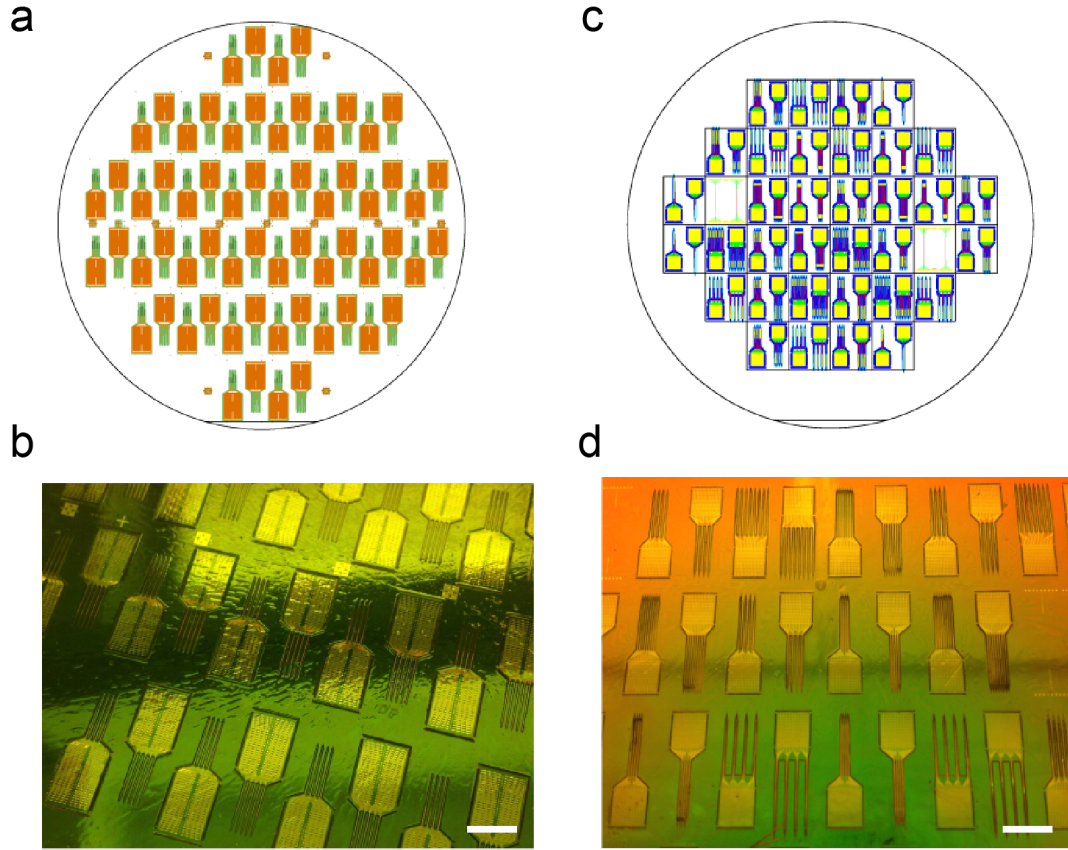
Figure 3.2: **Neural nanoprobes prototyped at the KNI at Caltech.** Wafer layout drawings (a,c) and close-up photographs (b,d) of fully fabricated wafers. Silicon on isolator (SOI) wafers with 100 mm diameter were used for prototyping. The drawings are composed to show all the individual layers used to create the different photomasks. (a,b) Generation 1 wafers required 5 photomasks and 1 electron beam lithography (EBL) step. (c,d) Generation 2 wafers required 4 photomasks and 1 EBL step. Scale bar is 5 mm.

This chapter describes the fabrication development of the second generation of devices (Gen2, Figure 3.2c,d). The fabrication protocols established with Gen1 devices were evolved and extended to create Gen2 device prototypes on 100 mm SOI wafers at the KNI. While fabricating at a university level cleanroom like the KNI was very useful for prototyping, assessing design choices, and learning about device handling techniques, there were inherent challenges to achieving high reliability and yield of the fabricated probes. In order to go beyond proof-of-concept devices and to pursue reliable large scale production, a transition to foundry based fabrication was carried out. The transition and the results of the fabrication efforts at a commercial CMOS/MEMS foundry (LETI, Grenoble, France) are also reported below.

3.2 NanoProbe Models

The overall modular design of the system enables leveraging a variety of nanoprobe front-ends that can target circuits in different species with different spatial resolution and coverage. Therefore a catalog of ten neuroprobe models was created, primarily optimized for targeting structures in the mouse and rat brain (Fig. 3.3). The multiplicity of neuroprobe models combinatorially expands the number and flexibility of 3-D array configurations that can be customized to specific brain areas with convoluted anatomy, such as the rodent hippocampus.

Nanoprobe models ranged in shank count (1, 4, or 8), shank pitch (250, 500, or 750 μm), and shank length (5 or 6 mm). The width of shanks was kept at a minimum in order to reduce mechanical invasiveness through tissue displacement. In all but one design, the maximal shank width in the span containing electrodes was less than 65 μm (50 μm on average), while shanks where much narrower near the tip (24 μm) and only gradually widened to about 100 μm near the probe base.

All nanoprobe models contain 256 recording microelectrodes distributed along one or more shanks in single or double row configuration (Fig. 3.4). The microelectrodes are small in area (117 μm^2) and shaped as ovals elongated parallel to the shank axis (8 $\times$ 16 μm). Single row configuration electrodes are offset from the center axis of the
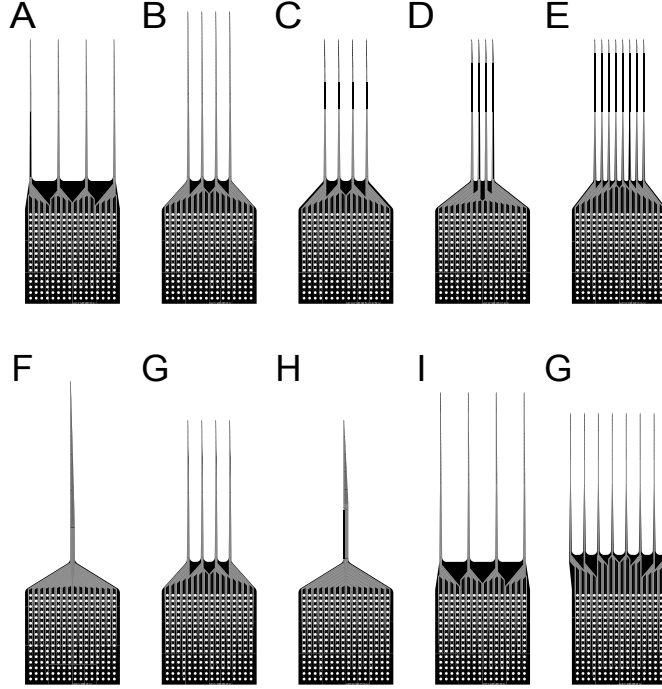
Figure 3.3: **Catalog of neural nanoprobe models.** Model designs are biologically driven and feature diverse electrode configurations. Standardization and modularity is achieved by designing each model to contain 256 recording sites, distributed differently across a variable number of shanks, but connected to a standard 16×16 output pad matrix at the probe base. Shanks are 5 mm long, except for models B, F, I which have 6 mm long shanks. See Fig. 3.4 for the different electrode configurations used.

shank and are positioned only 4 μm away from the shank edge.

The key to keeping the shanks narrow is the use of nanoscale interconnects to passively route the recording sites on the shank to their corresponding output pads at the base. The lines are 280 nm wide as they come out of the recording sites, gradually widen to 680 nm as they reach the probe base and immediately extend to 3 μm at the base. In the prototyping phase of the fabrication, widening the interconnects at the probe base was essential to allow the use of contact photolithography to pattern this portion of the device. Interconnect routing and dimensions adhere to the design rules developed in the previous chapter. They are meant to ensure that electrical parasitics are kept at a minimum as described in Chapter 2.

To compensate for fabrication variations that arise during electron beam lithography and metal lift-off, the interconnect width was set to 280 nm and the edge-to-edge

spacing was 320 nm. In the prototyping phase interconnect height was 100 nm. The next section discusses additional electron beam lithography scattering compensation techniques implemented to guarantee adherence of the fabricated interconnects to design specifications.
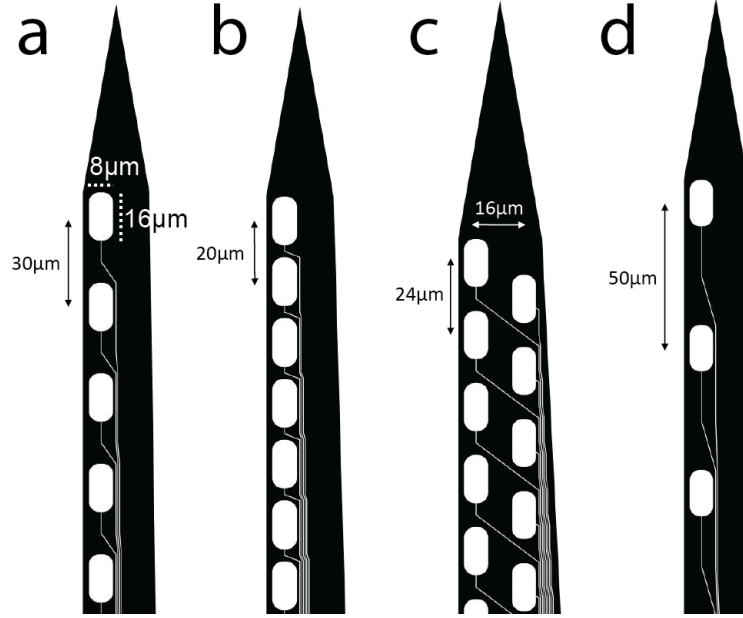

Figure 3.4: **Recording site layout configurations**. Recording sites are vertically distributed in a single column 2–4 μm from one edge of the shank with center-to-center pitch of (a) 30 μm (b) 20 μm (d) 50μm or (c) in a two-column staggered configuration with 16 μm inter-column distance and 24 μm center-to-center electrode pitch.xs

Finally, the base (3.4×3.4 mm) of the nanoprobes houses a 16×16 interface matrix of $100 \, \mu$m circular pads with $100 \, \mu$m edge-to-edge spacing, which constitutes the standardized output interface between the probe and the rest of the system. This design choice requires the use flip-chip bonding techniques to interface the probe to the rest of the system. The wirebonding alternative requires distributing the pads around the perimeter of the base and would have occupied significantly larger base area [31; 59], making the assembly of compact 3-D arrays deployable in a chronic setting impractical or impossible. The interface pad dimensions were chosen to allow the use of gold stud bumps during post processing to prepare the probe for flip-chip bonding (see Chapter 5). Table 3.1 summarizes the parameters of the different

nanoprobe models.

Table 3.1: **Nanoprobe model parameters**: number of shanks **S#**, shank length **SL**, shank pitch **SP**, number of electrode columns **EC#**, electrode vertical pitch **EP**, electrode configuration identifier in Fig. 3.4 **EID**.

Model	S#	SL	SP	EC#	EP	EID
units		mm	μm		μm	
A	4	5	750	1	30	a
B	4	6	500	1	50	d
C	4	5	500	1	50	d
D	4	5	250	2	24	c
E	8	5	250	2	24	c
F	1	6	n/a	1	20	b
G	4	5	500	1	30	a
H	1	5	n/a	2	24	c
I	4	6	750	1	50	d
J	8	5	500	1	30	a

3.3   NanoProbe prototype fabrication

Nanoprobe prototypes were fabricated at the KNI, Caltech's nanofabrication clean-room facility. Fabrication leveraged silicon on insulator (SOI) wafers as its starting point. Unlike a conventional bulk silicon substrate, an SOI wafer is a three layer stack of relatively thin prime quality silicon (device layer), a buried layer of electrically insulating silicon dioxide (BOX layer), and a bulk silicon support wafer (handle layer). The thickness of the nanoprobe shanks is mainly determined by the thickness of the SOI device layer [51; 60]. The top side of the SOI wafer is processed first to build the layer stack on top of the device layer (Fig. 3.1) and involves the sequential

creation of bottom insulator, metal interconnects, top insulator, and biocompatibility coating. This is followed by first etching the top side and then the bottom side of the SOI wafer to define the perimeter of the nanoprobes and release them from the bulk substrate.

Nanoprobe prototype fabrication at the KNI required the use of one electron beam lithography step and four photolithography steps, each defined by a separate photomask. Square 5" photomasks were made with chrome features (3 μm resolution) patterned on a soda lime substrate (Photosciences, Torrance, CA). Mask layout drawings were generated using AutoCAD software (AudoDesk). All mask drawing were parametrically controlled and programatically generated using a custom LISP software extension suite developed by Evgueniy Lubenov at Caltech. Figure 3.2c shows the layout of the wafer. Each nanoprobe model was paired with itself or with a similar model to form a rectangular "die". Devices in each die are layed out facing each other to maximize packing efficiency. There are 34 rectangular dies (68 devices) tiled across the wafer. This layout created grid coordinates defining the corners of every die which was essential for properly aligning the electron beam lithography (EBL) step to the photolithography steps. All alignment marks were placed along the center x-axis of the wafer starting with marks patterned by the EBL step. To overcome uncertainties with visually inspecting the quality of metal deposition and the completeness of dielectric etches, two die slots were allocated on the wafer for test chips (white rectangles in Fig. 3.2c). These test chips contain five interconnects running parallel to each other for 8 mm, connected on both sides to test pads.

Table 3.2: **Prototype fabrication materials and thicknesses.**

Layer	Material	Thickness
Substrate	Silicon / BOX	15 μm / 900 nm
Bottom Insulator	Thermal Oxide	850 nm
Metal	Ti / Au	7 nm / 100 nm
Top Insulator	Silicon Nitride	100 nm
Biocompatible Layer	Parylene C	1.5 μm

3.3.1 Fabrication protocol

The nanoprobe prototype fabrication steps are outlined in Figure 3.5. Fabrication is implemented on 4" (100 mm) SOI wafers, polished on both sides (Ultrasil Corporation). The device layer is 15 μm thick ($\pm 0.5\,\mu$m), the BOX layer is 1 μm ($\pm 5\%$) thick, and the handle layer is 300 μm thick. Given this device layer the thickness of the nanoprobe shank is 18 μm. The handle layer determines the $\sim 320\,\mu$m thickness of the probe base. The BOX layer is important for two reasons. First, it is used as a stress compensation layer to keep the shanks as straight as possible (Chapter 2). Second, it serves as an etch-stop when etching the handle layer form the back side of the wafer.
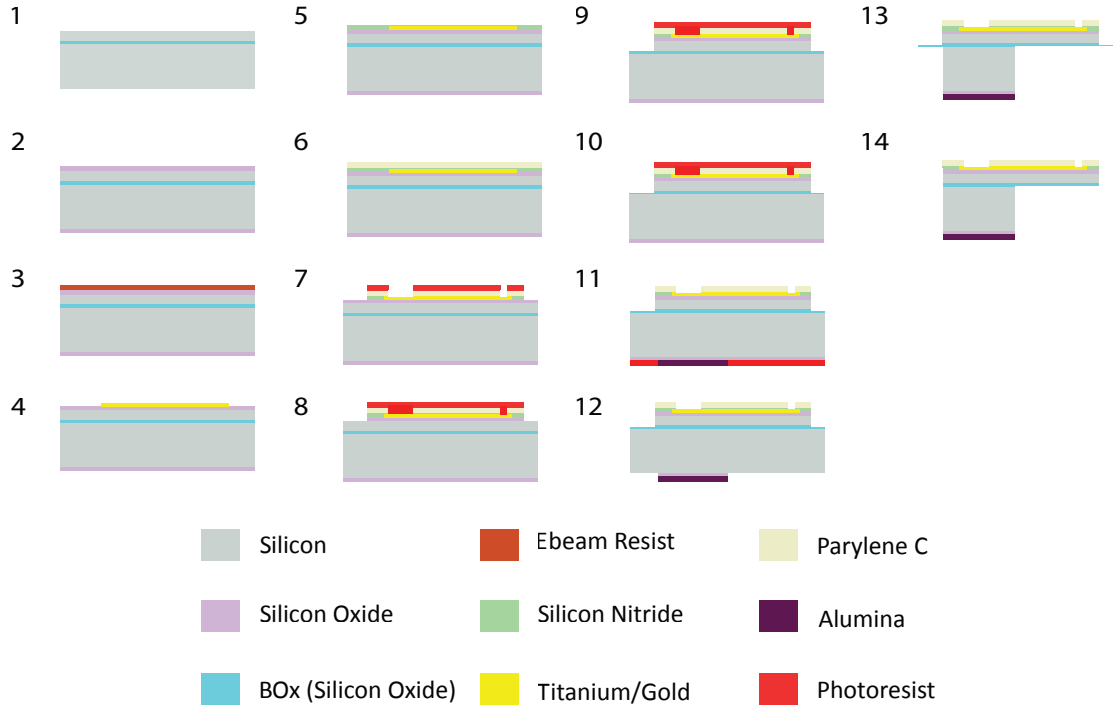

Figure 3.5: **Nanoprobe prototype fabrication steps**. Numbered steps used to create the following targets: bottom insulator (1-2), metal electrodes and interconnects (3-4), top insulator (5), biocompatibility coating (6-7), top side device definition (8-10), back side device definition and release (11-14). See text for details. Color legend defines the materials involved in the process.

The fabrication protocol is as follows with numbers referring to the steps illustrated in Figure 3.5.

Bottom insulator (1-2) SOI wafers (1) are first thoroughly cleaned. To remove oils and organic residues, wafers are rinsed with acetone and immediately after, rinsed with isopropyl alcohol (IPA). To further eliminate organic remnants, Nano-Strip, a stabilized formulation of sulfuric acid and hydrogen peroxide compounds (Cyantek Corp.), is applied at room temperature. To remove any native oxide the wafers are dipped in buffered hydrofluoric acid (HF : NH4F ratio of 12.5 : 87.5, Microchemicals, 30s). (2) Silicon oxide layer is thermally grown onto the SOI wafer (850nm, Tystar Furnace, 1050 °C).

Metal deposition (3-4) Electron beam lithography (EBL, Leica Microsystems EBPG-5000) is used to serially pattern the metal layer features (3). A double layer of Polymethyl Methacrylate (PMMA, MicroChem Corp. 495 A8/950 A2) is used for lift-off optimization. Two beam profiles are employed: high resolution (3.5 nA beam current) to pattern features smaller than 600 nm; low resolution (20 nA beam current) to pattern all other features. Write time for a full wafer (Figure 3.6d) is 22 hours. After patterning, the double layer of PMMA is developed (1 min in 1:3 MIBK:dH20) and gold is deposited using electron beam evaporation (CHA Mark 40, Ti/Au at 7nm/100nm). (4) Lift-off is performed by soaking the wafer in acetone ($>$ 8 hours) and exposing the wafer to gentle ultrasonic agitations (2s intervals, as needed). Figure 3.6 shows features from the resulting metal layer.

Top insulator (5) The metal layer is covered with a 100 nm layer of Si_3N_4 (Oxford PECVD).

Biocompatibility coating (6-7) The wafer is treated with silane (A174, 45 minutes) to increase Parylene C adhesion [61; 62]. (6) This is followed by depositing a 1.5 μm thick layer of parylene C (Specialty Coating Systems, PDS 2010). Parylene deposition on the back side of the wafer is avoided by protecting it with dicing tape. This dielectric stack serves as the top insulator for the metal as well as the biocompatible layer. (7) To expose the recording sites and interface pads, the parylene layer is patterned (AZ9245, 3.6 μm) and
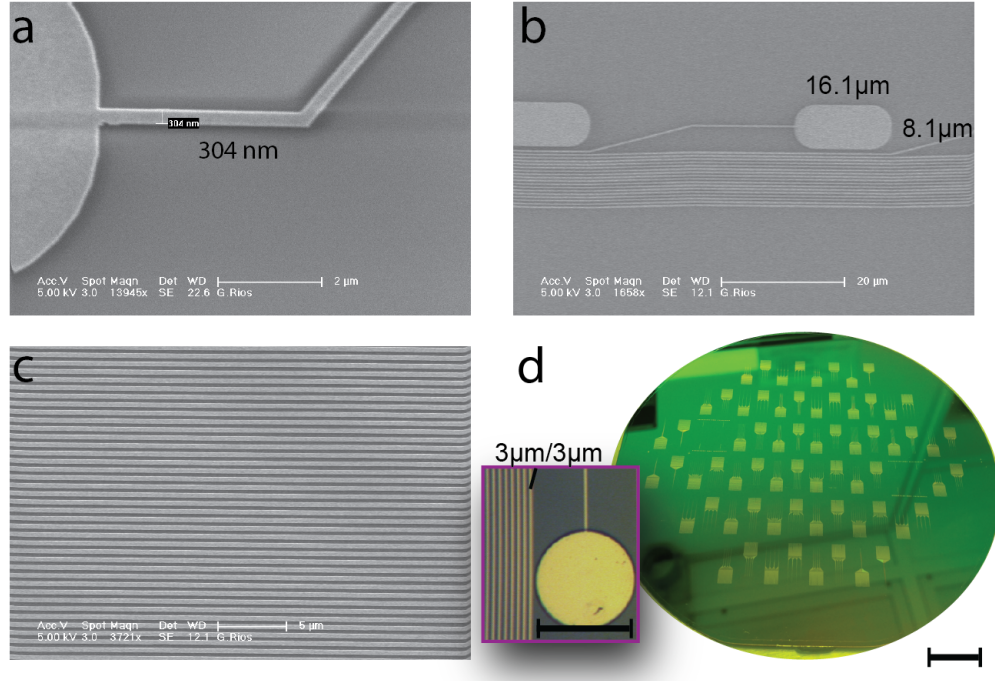
Figure 3.6: **Metal layer following gold lift-off**. All metal features of the probe, i.e. traces,recording sites, and output pads, are patterned using electron beam lithography. (a-c) SEM images of gold features after lift-off. (a) 300 nm interconnect entering a recording site ($2\,\mu$m SEM scale bar). (b) Recording sites have smooth perimeters to facilitate pattern transfer development ($20\,\mu$m SEM scale bar) (c) Interconnects on the shank with relative width:spacing ratio of 1:1.1. Increased spacing minimizes electron beam proximity effects ($5\,\mu$m SEM scale bar) (d) Left image shows an output pad connected to a 3 μm traces ($100\,\mu$m scale bar). Center image shows full wafer after patterning and gold lift-off (12.7 mm scale bar). Electron beam lithography step takes 23 hours and 18 minutes.

etched through (RIE, O2/CF4). Keeping the same patterned photoresist layer the Si_3N_4 layer is further etched (RIE). The former etch is isotropic, so precise pattern alignment to the wafer is essential to avoid exposing the metal. To determine completion of the etch, the test chip pads are tested for continuity across the metal lines. Incomplete etch prevents proper electrical contact to the testing probe tips and the line tests as an open circuit.

Top side device definition (8-10) (8) The nanoprobe outline is patterned (AZ5214, 1.1 μm) and etched down to the BOX layer. The first etch removes the thermal oxide layer around the nanoprobe, exposing the silicon in the device layer

(Oxford III-V RIE, C4F8/O2). (9) The silicon device layer (15 μm) is anisotrop-ically etched using a cryostatic etch (Oxford III-V, -120 °C). (10) Finally, the BOX layer is etched (900 nm out of $1\mu m$, Oxford III-V RIE, C4F8/O2). This completes the processing on the top side of the wafer.
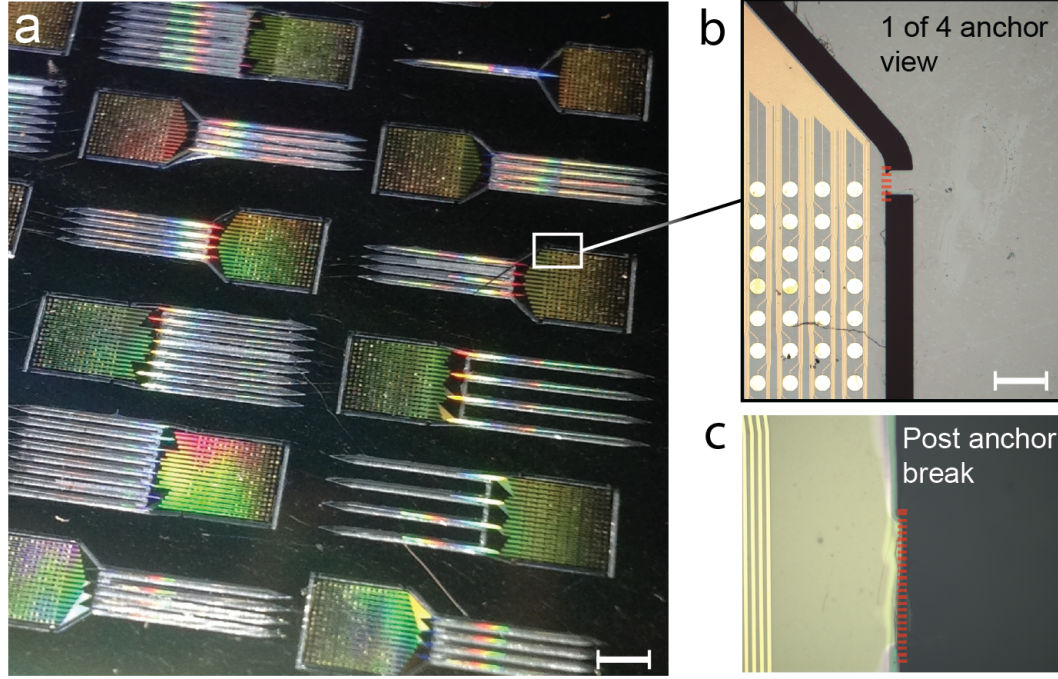

Figure 3.7: **Nanoprobe anchors.** (a) Image of a fully processed wafer showing released, but anchored devices. Scale bar: 3 mm. (b) Close-up image of one corner anchor tab (out of four) keeping the probe in place. Red dashed line indicates where anchors are designed to break. Scale bar: 300 μm. (c) Probe from (b) is safely extracted from the wafer using a vacuum pen, leaving little remnants of the anchor.

Back side device definition (11-13) The remaining processing steps are performed on the back side of the wafer. (11) Alumina is sputtered (TES, 300nm) and de-fined using lift-off (AZ5214, image reversal). This serves as an etch mask for the subsequent etching steps. (12) The thermal oxide layer is etched (RIE, Oxford III-V RIE), exposing the silicon of the handle layer. (13) The handle layer sili-con (300 μm) is anisotropically etched using the Bosch process (Oxford, DRIE). (14) Final release of the nanoprobes is accomplished by carefully etching away the left over BOX layer (100 nm, RIE). Nanoprobes remain attached to the wafer (Figure 3.7a) through several anchor tabs at the base of the nanoprobe.

These anchors are defined on the device layer which makes them only 15 μm thick. Figure 3.7b shows an example of an anchor point. The anchors keep the nanoprobe in place, but can be cleanly and easily broken off at a later point (Figure 3.7c). Device extraction is done by gently tugging at the base of the probe using a vacuum pen.

Figure 3.8 shows fully released nanoprobe prototypes fabricated at the KNI. The measured shank thickness for all models is 18 μm (Figure 3.8f, inset SEM image). Immediate fabrication yield, as defined by a successful release, is typically high (> 90%).
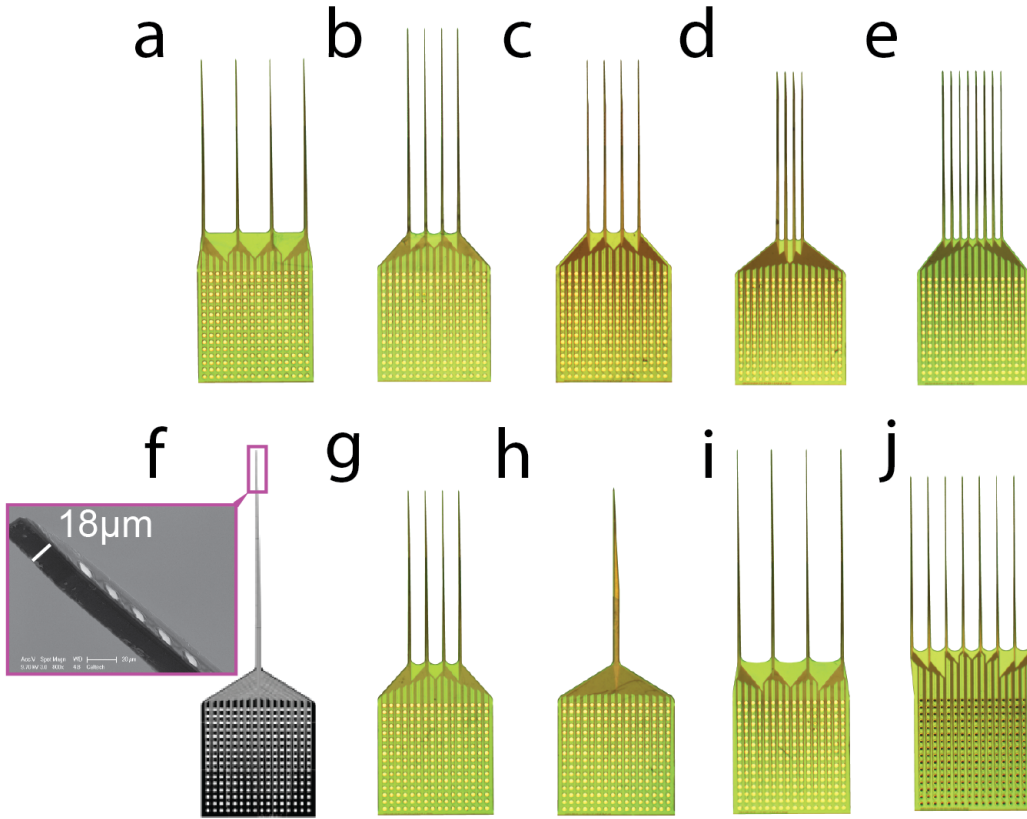

Figure 3.8: **Fabricated nanoprobe prototypes.** (a-e,g-j) Microscope (1x) images of fully released devices. Images are cropped around the perimeter of the device. (f) SEM image of a single-shank design showing the tip of the shank.

3.4 NanoProbe foundry fabrication

Pursuing reliable large scale nanoprobe production, characterized by high yield and reproducibility requires transitioning from prototype fabrication to foundry-based fabrication. Therefore, significant effort was made to adapt the prototype fabrication processes for deployment at LETI, a commercial microelectronics foundry in France. LETI houses both a CMOS and a MEMS foundry and is capable of co-integrating the technologies (200mm). They are experts in standard CMOS/MEMS processes like patterning nanoscale features and etching deep silicon trenches.

The protocols developed during prototype fabrication at the KNI provided a critical starting point for the transition toward foundry-based production at LETI. Equally important were the simulation platforms developed to extract design rules, focused on minimizing electrical parasitics and maintaining the mechanical integrity of the shanks. While the initial fabrication protocols quickly evolved to be more compatible with LETI facilities, the simulation platforms remained instrumental in making essential design parameter choices throughout this effort. The shank stress compensation simulation platform was particularly useful since LETI had stricter rules on dielectric thicknesses and compositions. Table 3.3 summarizes the dielectric materials and thicknesses chosen (see Figure 3.1 for reference).

Table 3.3: **Foundry fabrication materials and thicknesses**

Layer	Material	Thickness
Substrate	Silicon / BOx	17 μm / 900 nm
Bottom Insulator	Thermal Oxide	1.2 μm
Metal-Interconnects	Cu	300 nm
Metal-Electrodes and Pads	Au	300 nm
Top Insulator	SiO2 / SiN	500 nm / 40 nm
Biocompatible Layer	Parylene HT	1.5 μm

Eight out of the ten nanoprobe models prototyped at the KNI were selected for
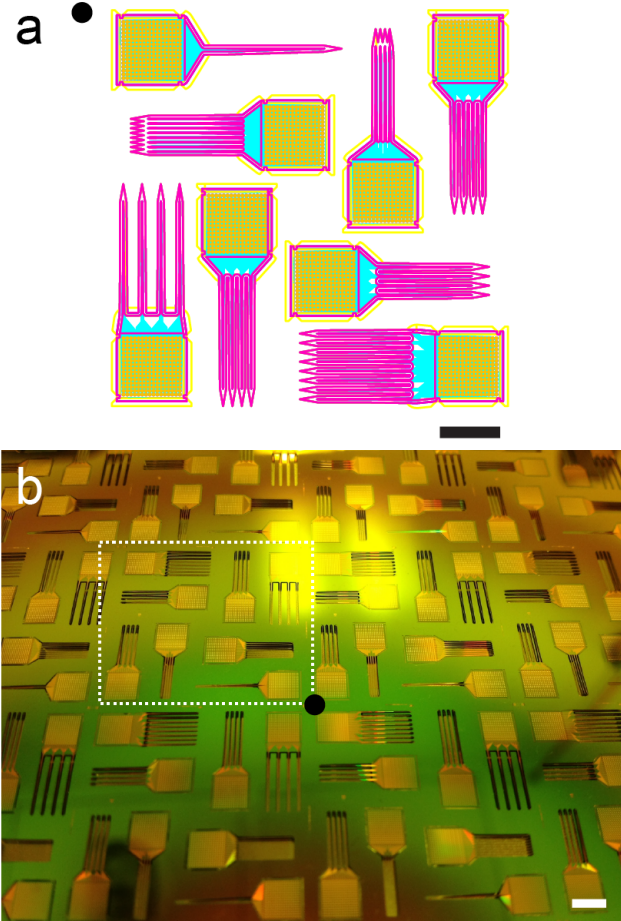
Figure 3.9: **Foundry-based fabrication of nanoprobes.** (a) Stepper field layout drawing used for foundry-based fabrication at LETI. The layout contains 8 (out of 10) Gen2 probe models arranged to maximize stepper field coverage. This field is tiled ("stepped") across a 200 mm SOI wafer. Scale bar: 3.5mm. (b) Close-up image of fully released devices anchored onto the wafer. White dashed rectangle outlines one stepper field. Black circle indicated relative position to (a). Scale bar: 3.5mm.

foundry fabrication at LETI. Figure 3.9a shows the stepper field layout we designed containing the eight chosen nanoprobe models arranged to maximize packing efficiency (designs H and I were excluded, see Figure 3.3 for reference). The following section summarizes the fabrication strategy adapted at LETI, focusing the discussion on novel procedures not available at the KNI.

3.4.1 Fabrication protocol

The nanoprobe foundry-based fabrication steps are outlined in Figure 3.10 . Fabrication was implemented on 8" (200 mm) SOI wafers and consisted of 26 processing steps requiring the use of 7 different stepper field masks. Both the CMOS and MEMS facilities at LETI were required: CMOS for the nanoscale interconnects and MEMS for everything else. As a result, the order of the fabrication procedures was carefully evaluated. One important consideration was the deposition of gold on the recording sites and interface pads. Gold is a serious contaminant if it makes its way inside a CMOS facility, since it diffuses rapidly in silicon which can cause electrical issues at the transistor level. Therefore gold deposition could not occur until after the CMOS facility procedures were completed. In addition to using the CMOS/MEMS facilities, an outsourced step for Parylene deposition was needed (Specialty Coating Systems). Figure 3.10 summarizes the full fabrication procedures in 11 steps.

The fabrication protocol is as follows with numbers referring to the steps illustrated in Figure 3.10.

Bottom insulator (1-2) (1) The first major difference between the fabrication efforts at the KNI and those at LETI was the use of 200 mm SOI wafers. This quadrupled the area available for nanoprobe fabrication. While the handle layer (700 μm) is much thicker than that of the SOI wafers used in the KNI, the device layer (15 μm) and BOX (1 μm) were kept the same. (2) The first process step involves thermally growing a layer of silicon oxide on the wafer. This provides the bottom insulator for the interconnect metal. It also serves as a crucial stress compensation layer, since its thickness is determined by design and not by process requirements, as is the case with subsequent dielectric depositions. Through simulations, it was determined that the thermal oxide thickness required for proper stress compensation is 1.2 μm (see Chapter 2).

Metal interconnects (3) Dielectric layer stack of SiO / SiN (300 nm / 40 nm) is deposited and the interconnect metal lines are patterned using a 248 nm deep ultra violet (DUV) stepper (ASML300). While this tool is obsolete in
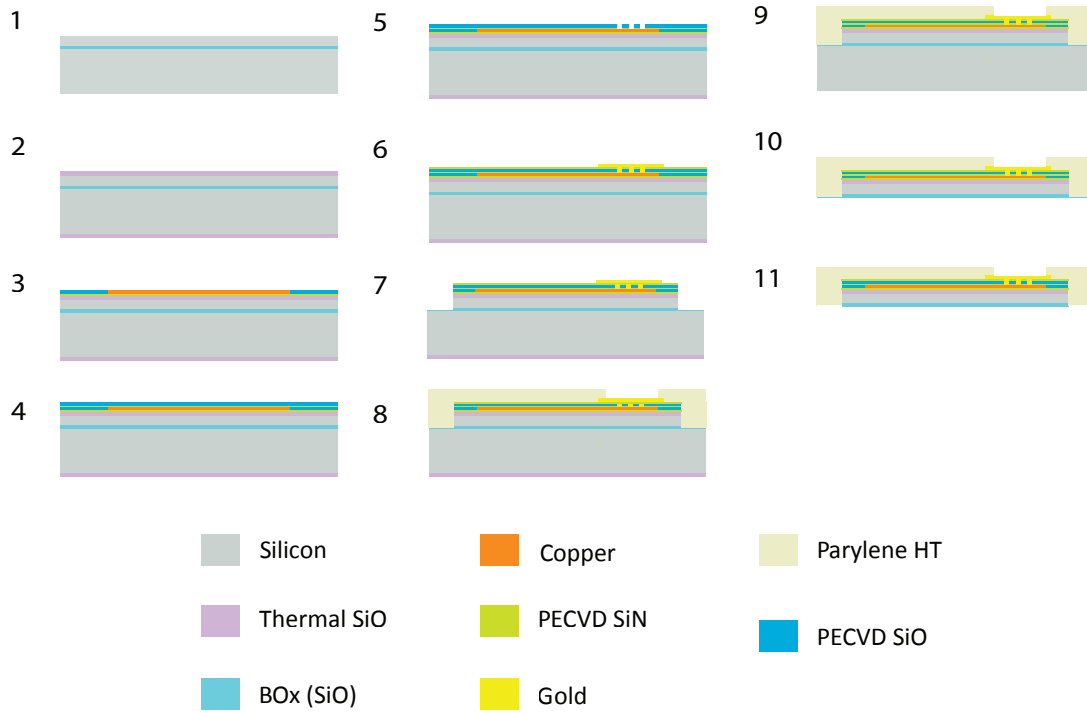
Figure 3.10: **Nanoprobe CMOS/MEMS foundry fabrication steps**. Numbered steps used to create the following targets: bottom insulator (1-2), copper interconnects (3), top insulator (4-5), gold electrodes and pads (6), top side device definition (7), biocompatibility coating (8), back side device definition and release (9-11). See text for details. Color legend defines the materials involved in the process.

the CMOS industry (current node size is 14 nm) it was ideal for patterning 300 nm wide interconnects along with the rest of the metal layer features. In contrast to electron beam lithography, DUV stepper lithography exposes and pattern transfers a field size of 22 × 22 mm at a time (see Figure 3.9a for field design). It then quickly steps through the wafer, 22 mm at a time, exposing the full wafer. This process is significantly faster than electron beam lithography, capable of exposing a 200 mm wafer in minutes (∼10 s/field x 80 fields = 13 minutes). This step would take days using electron beam lithography. Copper is then deposited using the Damascence process. This process was introduced in the 1990s by IBM as a way to replace traditionally used aluminum with lower-resistive copper interconnects [63]. The process steps are as followed:

i Trenches are etched into the SiO / SiN (300 nm / 40 nm) dielectric stack

using the DUV-patterned, metal-layer photoresist as an etch mask.

ii A barrier layer is deposited (TiN, 20 nm) to prevent Copper diffusion into the substrate.

iii A seed layer is deposited (Cu, 200 nm) to serve as the anode for the subsequent electrodeposition step.

iv A thick layer of copper is electroplated into the trenches (ECD, 300 nm), intentionally overfilling them.

v Adhesion bake (400 °C, 30 min)

vi Chemical Mechanical Polishing (CMP) is used to remove the excess copper, leaving behind only the copper filling up the trenches. This becomes the patterned interconnect layer.

The copper Damascene process holds several advantages over gold lift-off. For starters, the measured resistivity for the copper interconnects (1.7 $\mu\Omega cm$) was lower than those made from gold (2.6 $\mu\Omega cm$). Furthermore, open and short circuits are reduced (see Chapter 5). Finally, and relevant to future scaling efforts, the damascene process is compatible with multilayer interconnect structures. This provides a path to further increase the number of recording electrodes without increasing the width of the shanks.

Top insulator (4-5) (4) To insulate the copper interconnects a dielectric stack of SiN / SiO (40 nm / 500 nm) is deposited (PECVD). (5) Vias are etched (RIE) through the top insulator layer at the recording electrode and interface pad locations. A second copper damascene process is used to fill the exposed vias. These vias interface the copper interconnect layer to the subsequent gold metal layer. From this point forward, the wafer is transferred over to the MEMS facility.

Electrode and interface pad definition (6) One major disadvantage of using copper in nanoprobes is that it is not biocompatible. Due to its redox activity, copper can lead to the generation of toxic reactive oxygen species in the brain

[64–66]. Therefore, gold was used for the recording electrodes. Similarly, gold was used to define the interface pads. The gold layer (300 nm) is selectively electroplated on top of the copper vias/top insulator using on a Nickel (700 nm) seed layer. The newly defined recording electrodes seal off the copper vias/interconnects, assuring direct copper contact with the brain is avoided. This step concludes the metalization requirements.

Top side device definition (7) The nanoprobes are then defined by etching through the multiple dielectric layers (RIE), the silicon device layer (DRIE), and 90 percent of the BOX layer (RIE). This is similarly to what is done at KNI.

Biocompatibility coating (8) Parylene HT (1.5 μm) layer is deposited (CVD, Specialty Coating Systems) onto the primed wafer (Silane, A174). The Parylene HT variant was used instead of Parylene C because it is more stable at higher temperatures (short term up to 450 °C). Its relevant material properties (i.e. biocompatibility, pin-hole free, low dielectric constant, chemical resistance) are similar to those of Parylene C [44]. However, it possesses two added advantages apart from the higher temperature stability just mentioned, in that it has a lower coefficient of friction and higher penetrating ability during deposition. The former provides an advantage during device implantation since it decreases the frictional force between the shank and brain tissue. The latter, and more relevant to this process step, assures a higher degree of conformal coverage to cover the exposed sides of the shank (top + 2 side walls). Following deposition, the Paryelene HT is protected with photoresist optimized for step coverage and the recording sites are patterned. Finally, the Parylene is etched (RIE).

Back side device definition (9-11) (9) The back side thermal oxide layer is patterned and etched (RIE) exposing the silicon device layer. (10,11) Final DRIE etch is performed to completely etch away the silicon from the device layer (700 μm) as shown in Figure 3.12a. The left-over BOX layer (100 nm) serves as an etch stop and guarantees a smooth back-side surface required for subsequent packaging steps (Figure 3.12b). Finally, the thin BOX layer is etched (RIE),

fully releasing the nanoprobes. Anchors keep the nanoprobes in place on the wafer until they are manually extracted (see Figure 3.7 for anchor design).
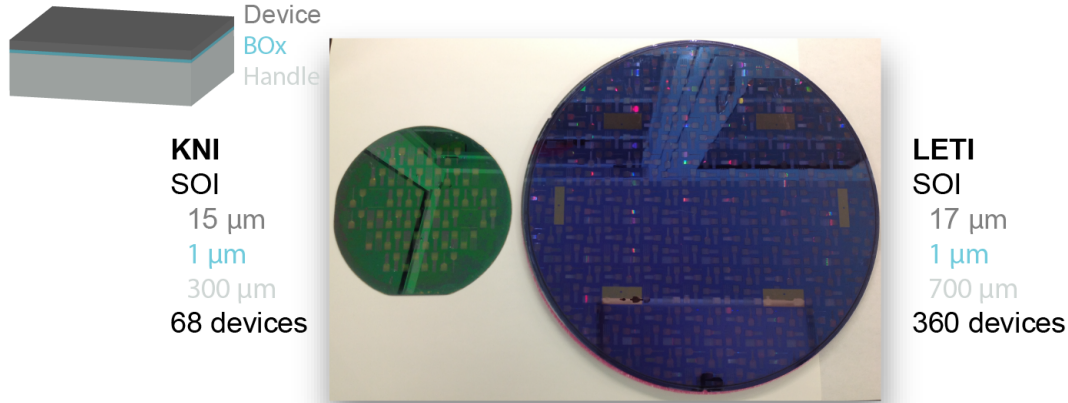
3.5 Conclusion

Figure 3.11: **Fabrication efforts comparison**. On the left is a 100 mm wafer fabricated at a university cleanroom (KNI) and on the right is a 200 mm wafer fabricated at a commercial foundry (LETI). The resulting devices are superficially the same, but are built differently.

The modular system design enables the use of a variety of nanoprobe models, which combinatorially expands the number and flexibility of 3-D array configurations. A catalog of nanoprobe models was created according to the design rules described in Chapter 2. Fabrication methodologies were developed for building nanoprobes both at a university nanofabrication facility and at a commercial CMOS/MEMS foundry. The difference in the production scale of the two approached is illustrated in Figure 3.11. The creation of a fabrication protocol compatible with a commercial CMOS/MEMS foundry represents an important step towards mass-production and dissemination of the neural nanoprobes.
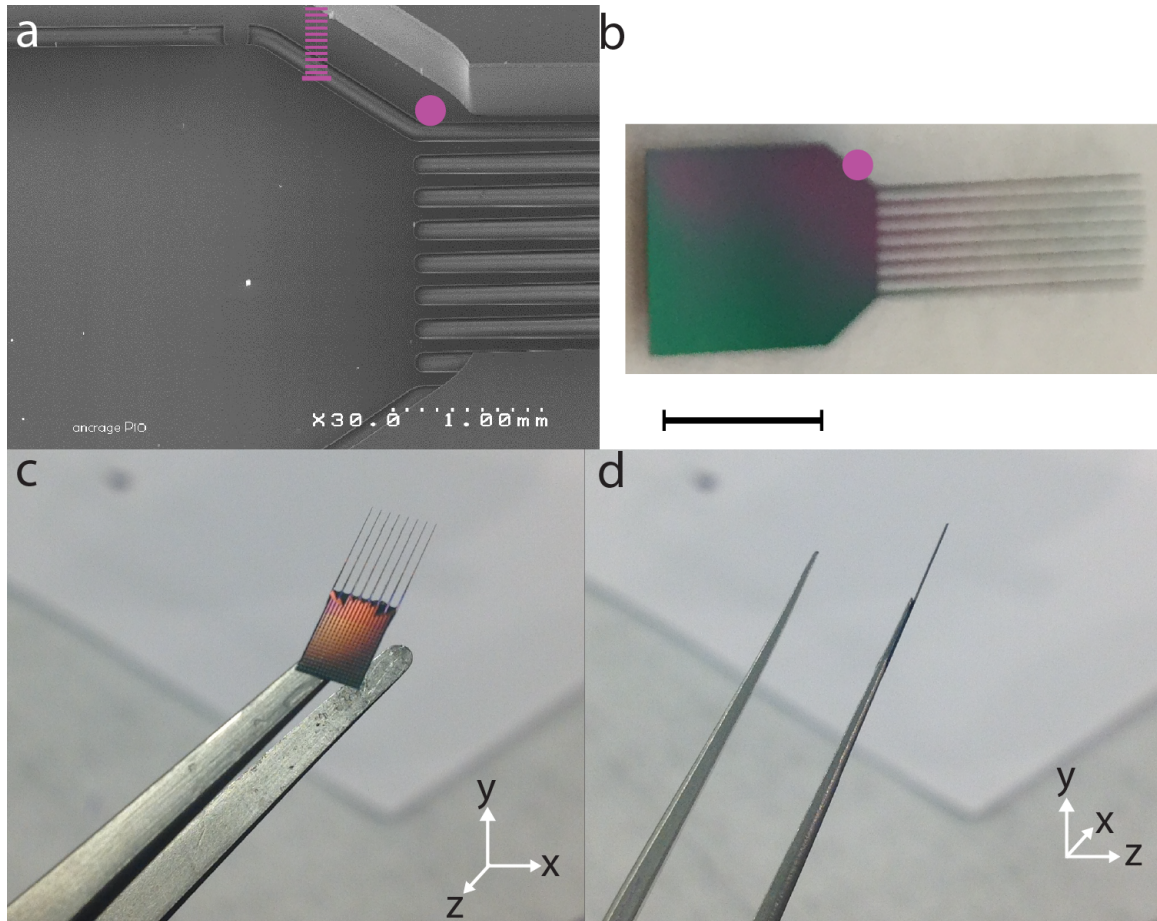
Figure 3.12: **Nanoprobes are etched down to the BOx layer** (a) SEM image of the back side region of an 8-shank probe (SEM taken at LETI). The silicon handle layer (700 μm) has been etched completely exposing the buried oxide (BOx layer), which serves as an etch stop. SEM scale bar: 1mm. (b) Photograph of a device after it is extracted from the wafer. The purple/green colors indicate that the BOx layer forms the back surface of the device. Pink dot used for position references with (a). Scale bar: 3mm. (c) Photo showing the front-side of a probe held onto a pair of tweezers by the surface tension of a water droplet. (d) Side-view of the probe in (c) demonstrating that shank and base are the same thickness (18 μm).

It is important to emphasize that the commercially fabricated devices were less than 20 μm thick throughout, including the probe base (Fig. 3.12). Consequently, special device handling techniques were required. For example, conventional tweezers could no longer be used (Fig. 3.12c,d), so nanoprobes were handled using a vacuum pen and had to be specially stored (Fig. 3.13b). This device thickness also represented a significant challenge during system assembly as described in Chapter 5.
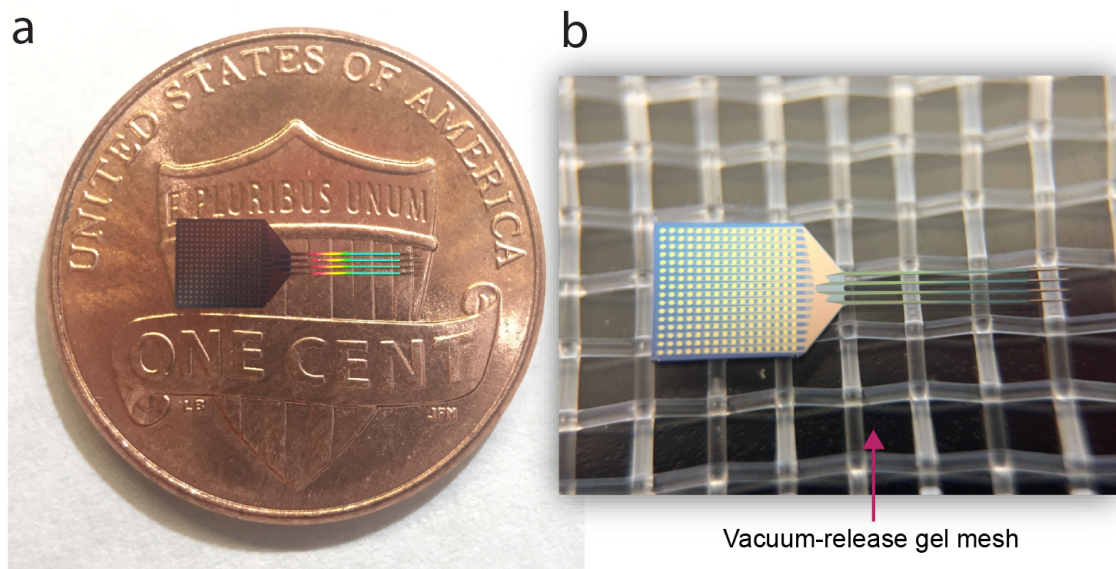
Figure 3.13: **Fully released nanoprobes.** Nanoprobes can be extracted from the wafer without damage but must be carefully handled and stored. (a) Photograph of a probe (U.S. penny for scale) after being cleanly extracted from the wafer. (b) Once detached from the wafer, released probes must be carefully stored. Vacuum releasable gel substrates (pink arrow) are used to store the probes.

Chapter 4

Data Acquisition System

4.1   Introduction

In order to take advantage of nanoprobes with hundreds of recording electrodes, they must be successfully interfaced to signal conditioning electronics and a data acquisition system. Until recently, electrophysiology setups for freely behaving animals employed individual fine wire tethers to carry the buffered raw analog potentials from each electrode in the brain to rack-mounted signal processing hardware. The practical scaling limit to this approach is ∼100 channels, due to physical constraints on the maximal number of connectors, the size of interface boards, and the weight and reliability of the fine wires.

An alternative strategy is to co-integrate active signal conditioning and multiplexing electronics directly with the probe itself using CMOS technology [29; 67]. However, as discussed in Chapter 2, the power dissipation of the active electronics and the limited thermal budget impose a scaling bound on this approach as well. Consequently, the system architecture adopted here delocalizes the active electronics away from the nanoprobe, but within the implant package. This is possible by taking advantage of commercially available integrated circuits (IC) developed for electrophysiological applications (Intan Technologies). The ICs provide amplification, conditioning, digitization, and multiplexing for 32 or 64 channels and have been widely adopted by the neuroscience community [68–72].

This chapter outlines the components of a modular 1024-channel data acquisition

(DAQ) system, capable of supporting 3-D arrays with up to 4 nanoprobes. It describes the design of two different DAQ PCB headstages, i.e. custom, high-density, printed circuit boards (PCB) that interface to an individual nanoprobe and employ multiple Intan chips to support signal conditioning and digitization of 256 channels. One DAQ PCB headstage is intended for acute and the other for chronic applications, but both are designed to be stackable in order to support 3-D arrays.

4.2   Modular 1024-channel DAQ architecture

Each module of the data acquisition (DAQ) system is designed to support 256 channels, i.e. the number of recording electrodes on a nanoprobe. Just as the nanoprobes are stacked to build a 3-D recording array, the DAQ modules are stacked to scale up the back-end data acquisition system. Figure 4.1 shows 4 modules assembled into a 1024-channel system.
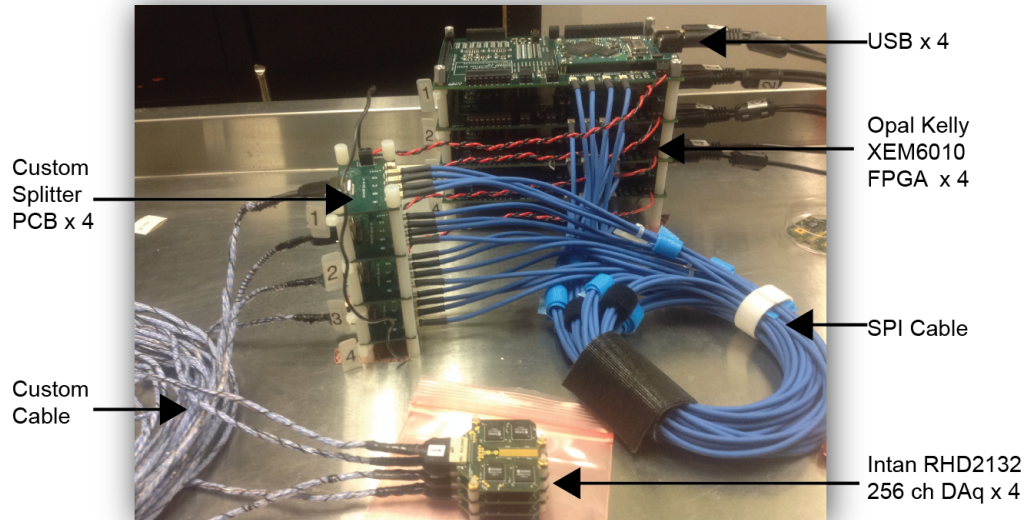

Figure 4.1: **1024-channel data acquisition system.** The 1024 system is a stack of 4 identical 256-channel modules. Each module consists of the following components: DAQ PCB headstage, 14' SPI cable tether, splitter PCB, 4 SPI cables, Intan evaluation board with XEM6010 FPGA Opal Kelly daughter board, USB cable. The first three components are custom designed and the last three are commercially available. See text of details.

Block diagram of the 256-channel data acquisition module is presented in Figure 4.2. The raw analog signals from the recording electrodes feed into a custom headstage (DAQ PCB). The purpose of the DAQ PCB is to provide signal conditioning and multiplexing of nanoprobe signals from within the implant package. Two versions of these boards were designed and built. The *acute* version, depicted in the block diagram, has eight 32-channel Intan chips (RHD2132). The *chronic* version contains four 64-channel Intan bare dies (RHD2164). At the DAQ PCB signals are amplified (2-stage, Gain = 200), filtered (0.1Hz-7.5kHz), multiplexed (256:8), digitized (16-bit), and transmitted out (SPI bus) through a custom cable.
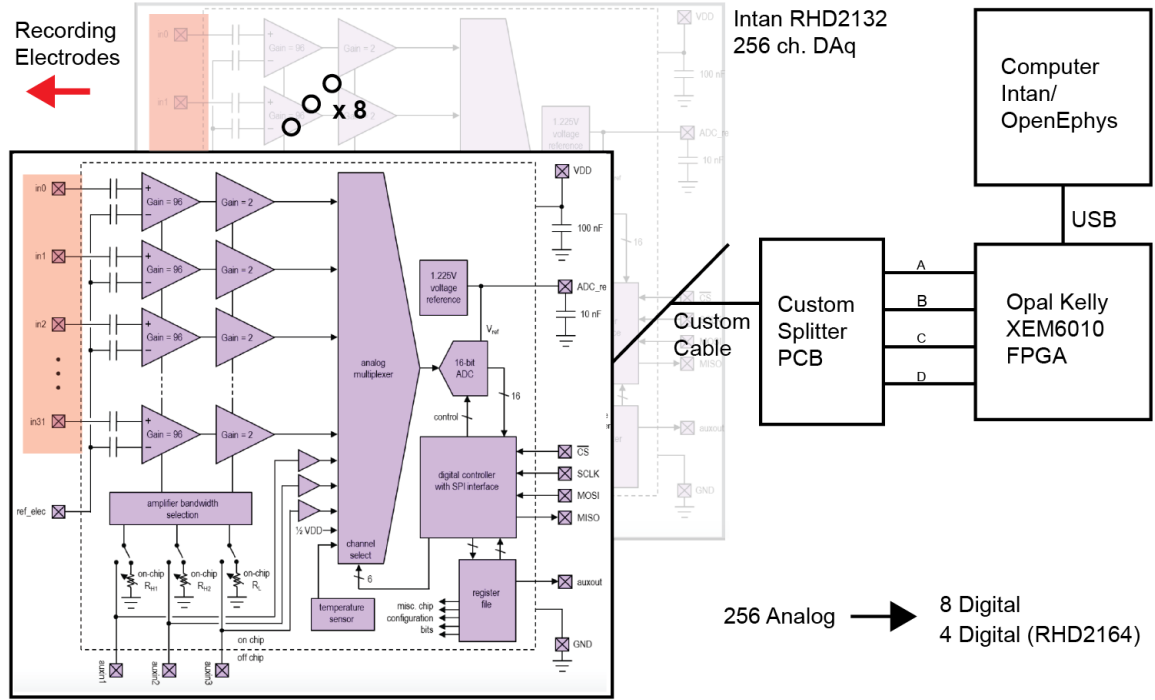
Figure 4.2: **256-channel module block diagram.** A nanoprobe with 256 recording electrodes is interfaced to a DAQ PCB housing eight RHD2132 Intan Technology digital electrophysiology interface chips. Each Intan chip provides 32 low-noise amplifiers, programmable analog and digital filters, a 16-bit analog-to-digital converter, and SPI bus. 8 LVDS lines output multiplexed, digitized data, while 3 LVDS inputs control the behavior of the chips. All lines ultimately connect to an FPGA board (Opal Kelly XEM6010) which controls the hardware. The recorded data are transmitted to the computer over USB 2.0.

The custom SPI cable provides the physical link layer for the SPI bus. It is 14 feet long to allow for tethered recordings in chronically implanted freely behaving

animals. It is composed of a bundle of very thin twisted pair wires (34 AWG) capable of handling LVDS signal transmission. The advantage of using LVDS is that it enables transmission of serial digital data over several feet with high signal integrity. This cable also delivers power and ground to the DAQ PCB using slightly larger single ended wires (32 AWG, to reduce ohmic drop). The composite wire bundle (11 twisted pairs, 3 single ended) is lightweight and highly flexible. It interfaces the DAQ PCB (Omnetics A79027-001) to a custom splitter PCB (Omnetics A29100-037).

The splitter PCB reroutes all custom cable lines (SPI, Power, Ground, test) to allow interfacing directly to Intan's back-end platform using commercial SPI cables. From there signals enter an USB/FPGA interface module (Opal Kelly XEM6010). The FPGA in this module (Xilinix Spartan-6 XC6SLX150-2) is programmed to control the behavior of the DAQ PCB over SPI and to provide both data and command interface to the host computer over USB 2.0 (Rhythm API). Using this platform, 256 channels can be simultaneously acquired and streamed from the DAQ PCB at sample rates up to 30 kSamples/s per channel. Intan's RHD2000 interface software and Open Ephys, both open source, were modified for compatibility with the custom DAQ PCB and enabled real-time data viewing, acquisition, and storage at the host computer.

4.3   Acute DAQ PCB

The initial version of the DAQ PCB was designed using eight QFN-packaged Intan chips (RHD2132, 8mm×8mm). Four of the chips are placed on the top side of the PCB, the remaining four on the back. The PCB takes in 256 analog inputs and outputs 8 digital signals. This version of the DAQ PCB is referred to as *acute*, because the weight (4.1 g) and size (39mm×37mm) of the board (Fig. 4.3aII) makes it more suitable for acute *in vivo* experiments in head-fixed animals. The QFN packaging of the chips, i.e. the plastic encapsulate, is responsible for a significant portion of the size and weight of the board.

The board layout design started by first establishing the placement orientations

for the chips on the PCB. The input/output pads on the Intan chip are separated: one side of the chip contains the 32 analog input pads and the other side contains all digital control (SPI) and power/ground lines. The chip footprint and QFN pins were oriented so that all of the analog input pads faced the center of the PCB, while all control/power pads faced the perimeter. This set up the strategy for signal routing.

In mixed-signal PCB design it is important to segregate analog from digital signals to avoid interference. This becomes critical when the analog signals are low-amplitude. The chip placement and orientation create a natural partition of the PCB: all digital signals are routed along the board perimeter and all input analog signals are routed at the center of the board. The PCB comprises 7 layers, of which 2 are used for analog signals, 2 for low-voltage differential signals (LVDS), 2 for ground planes (GND), while the remaining layer is a power plane (VDD) (see table 4.1 for layer stack). SPI digital outputs (8 MISO), SPI control inputs (MOSI, CS, SCLK), power and ground are interfaced to the PCB using an Omnetics connector (A79024-001). Power and ground lines are connected to power and ground planes.

Table 4.1: **Acute DAQ PCB layer stack**

Layer Name	Layer Type
Intan Top Layer	Signal
GND-1	GND plane
LVDS-1	LVDS
VDD	Power plane
LVDS-2	LVDS
GND-2	GND plane
Intan Bottom Layer	Signal

The top and bottom PCB layers are used to route the raw analog input signals from the nanoprobe. These analog inputs are interfaced to the PCB through a 64×4 matrix of pads. The individual pad size is $500\mu\text{m} \times 160\mu\text{m}$ and the footprint of the input matrix is 16.2mm$\times$3.25mm (Figure 4.3aII). Signals arriving at columns 1 and

4 of the input matrix remain on the top layer, while those at columns 2 and 3 are routed to the bottom layer using filled copper through vias (5 mil), embedded behind the input pads. Analog input trace lengths are kept short and are routed using 3 mil (76.2 μm) wide traces. This trace width is close to what is currently achievable through high-density PCB manufacturing technology (2 mil, 50.8 μm). Ultimately, minimum trace width sets the limit on the minimal size of the input pad-matrix.

SPI lines use low-voltage differential signaling (LVDS) and are routed on their own separate layers. All chips receive the same SPI input command lines (CS, SCLK, MOSI), but each chip generates its own SPI output (MISO). LVDS lines carry digital signals where logic levels are transmitted differentially between two wires/traces. Essentially, the transmitter injects a constant current (± 3.5 mA) into the wires. These wires are terminated with a resistor (100 Ω) at the receiver end, so the current from one wire passes through the resistor, creating a voltage drop (± 350mV), and returns via the other wire. The sign of the current injected by the transmitter establishes the polarity of the voltage drop at the receiver from which the logic level is determined. Because of this differential nature, digital LVDS lines are less prone to contaminating analog signals, but only if there is a tight electric/magnetic coupling between the two lines. Ideally equal and opposite EM fields are created and cancel out. Therefore, when routing LVDS traces the following guidelines were followed: (1) differential traces were routed as closely as possible; (2) discontinuities in the differential impedance was avoided by routing traces as straight as possible and using 45° bevels instead of 90° turns, when necessary; (3) trace electrical lengths were kept approximately the same to avoid skew; and (4) the number of vias was minimized.

The maximum power dissipation for this board is 298 mW (at full recording/sampling capacity). Recall that the power budget per nanoprobe shank is < 1 mW. Although the nanoprobe is thermally isolated from the DAQ PCB by the flexible cable, it is clear that additional heat sinking strategies may be needed to minimize the heat flowing into the nanoprobe. To allow heat sinking at the PCB the following three strategies are implemented. First, vias are placed directly underneath the chips. These vias are connected to both ground planes (copper fill), creating a low resistance thermal
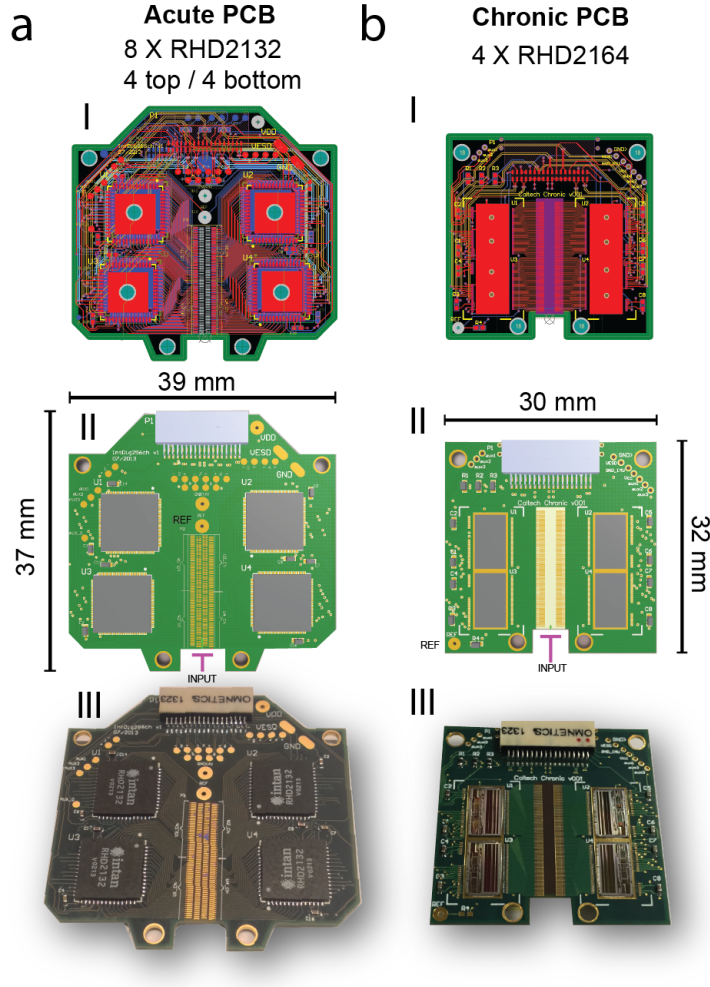
Figure 4.3: **256-channel DAQ PCBs.** Both the (a) acute and (b) chronic PCBs process 256 raw analog inputs from the electrodes of a single nanoprobe. (aI) Acute PCB is routed using 7 layers with the top and bottom layers each containing 4 QFN-packaged Intan chips (RHD2132). (aII) 3D model showing PCB dimensions. Input pads are arranged in a 32×4 matrix. The two middle columns contain filled through hole vias which carry half of the input signals to the bottom layer. (aIII) Image of fully populated acute PCB. (bI) Chronic PCB is routed using 6 layers and contains 4 bare dies on its top layer (RHD2164). (bII) 3D model showing PCB dimensions. Input pads are arranged in a 64×2 matrix. (bIII) Image of fully populated PCB. Both PCBs use the same Omnetics (A79024-001) connector and are compatible with the same custom cables and back-end system. Routing for both PCBs requires high-density interconnect PCB technology (Line width/pitch: 2mil/2mil, Sierra Circuits).

path away from the chips. Second, two mounting holes (80mil/2mm) are placed at opposite corners of the PCB. These mounting holes are also connected to both ground planes and can serve as an exit path for the dissipated heat by, for example, mounting

the PCB onto a thermally conductive, high-surface area radiator. Finally, the ground planes themselves are designed with a gap around the perimeter of the analog/digital boundary. This gap is bridged only at four points (10 mil wide copper pours) to establish electrical connectivity. Since the chips are thermally coupled to the digital side of the ground plane, the gap reduces the thermal conductivity in the direction of the analog input pad matrix and the nanoprobe.

The fully populated board is shown in Figure 4.3aIII. All components are surface mount and are soldered onto the PCB (Screaming Circuits). The input pad matrix is kept free any solder mask or solder paste.

4.4   Chronic DAQ PCB

The second version of the DAQ PCB uses 4 Intan bare dies, i.e. IC chips without packaging (RHD2164, 7.3mm×4.2mm die). All four chips are placed on the top layer of the PCB. This version of the board (Fig. 4.3bII) is smaller (30mm×32mm) and lighter (1.2g with potting), compared to the acute PCB, making it suitable for chronic experiments on freely-behaving animals (see Future Directions). Both PCBs are designed to be compatible with the back-end system and they use the same Omnetics connector and custom cable.

The routing and heat sinking strategy for the chronic DAQ PCB is similar to the one described for the acute version. A key difference involves the analog input interface matrix profile. Here, the matrix is layed out in a 128×2 configuration. The individual pad size is $762\mu m \times 7.6\mu m$ and the matrix footprint is 15.8mm×3mm (Fig. 4.3bII). Input pads are routed using 2 mil (50.8 μm) wide traces keeping their lengths short (< 3.5 mm). All other signal traces, including those for LVDS, are 5 mil wide (127 μm). The power dissipation for this board is 212 mW (at full recording/sampling capacity). Heat sinking precautions similar to the ones used in the design of the acute PCB were used here as well.

The board consists of 6 layers (see table 4.2 for stack details). All chips and all other components are placed on the top layer, keeping the bottom layer unpopulated

Table 4.2: **Chronic DAQ PCB layer stack**

Layer Name	Layer Type
Intan Top Layer	Signal
GND-1	GND plane
LVDS-1	LVDS
VDD	Power plane
GND-2	GND plane
Intan Bottom Layer	Signal

and flat. This helps during the subsequent wirebonding (COB) and packaging process (Chapter 5), since the board can be kept flat on a substrate without the use of a special adapter.
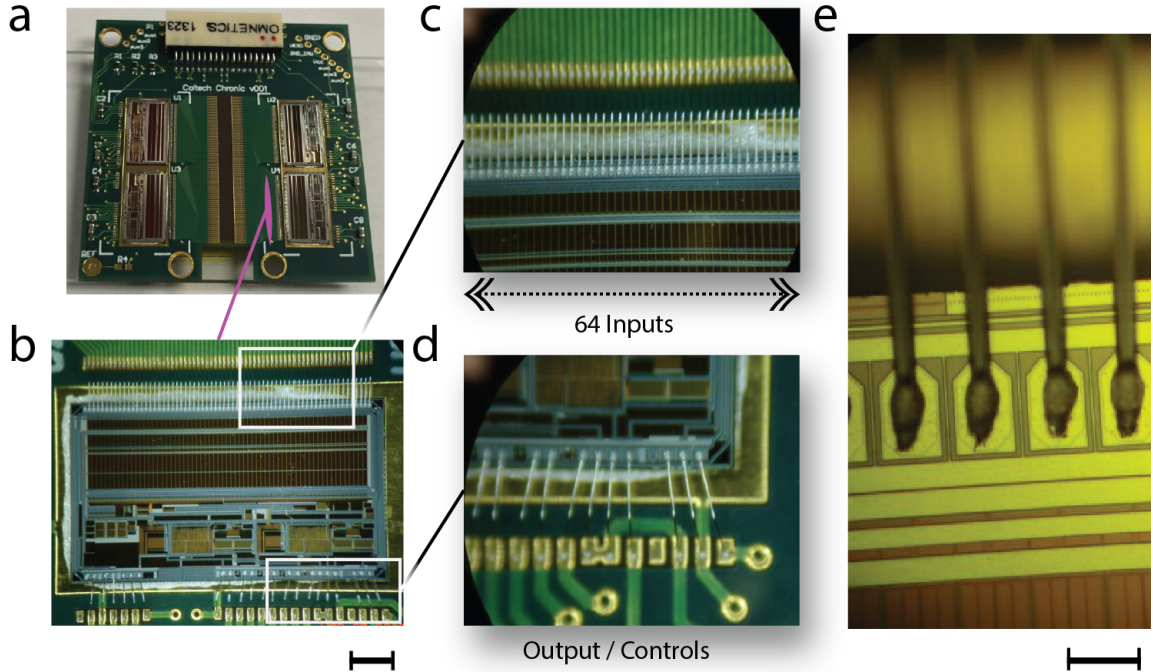

Figure 4.4: **Chronic DAQ PCB chip-on-board (COB) assembly** (a) Fully-assembled chronic DAQ PCB. (b) Micrograph of a wirebonded Intan RHD216 bare die onto the DAQ PCB (c) Wirebond on the input side of the chip require 2 mil wide PCB pads. This is at the manufacturing limit of high-density PCB technology. (d) Output/control pad wirebonds (e) Aluminum 1 mil diamater wedge bonds are used (assembly by IDAX).

The fully populated board is shown in Figure 4.3bIII. All passive components, including the Omnetics connector, are soldered. The chips are electrically connected

to the PCB through a chip-on-board process (COB) wirebonding process (Fig. 4.4). In particular, the chips are first attached onto the PCB using silver-filled conductive epoxy (AbleBond 84-1LMIT). This creates a high-thermally conductive path between the chip and the heat-sink vias. Next, the chips are wirebonded to the PCB soft gold pads using aluminum wedge bonds (1 mil diam wire). The complete assembly requires a total of 348 wirebonds. Finally, the wirebonded chips are potted and protected using epoxy (Hysol fp4650). The height of the potting epoxy is minimized to avoid adding extra weight and increasing the height of the board.

4.5 Conclusion

The high number of electrodes per nanoprobe/3-D stack required the development of high channel count, modular, and scalable data acquisition (DAQ) system. In designing the system an effort was made to leverage widely available hardware components and open-source software, to provide an easy path to system adoption. In particular, standard Intan RHD2000 USB interface boards and SPI cables were used as the system final stage before the host computer. The critical custom component we designed and built are the DAQ PCB headstages, which exploit commercial Intan chips to provide signal conditioning, digitization, and multiplexing of the nanoprobe outputs. The use of custom 14' SPI cables allows for chronic experimentation in tethered freely behaving animals.

The DAQ system has been demonstrated to support 1024 electrodes from a 3-D stack (Chapter 7). When input data are sampled at 30 KHz the resulting data rate is slightly over 120 MB/s. This is twice the nominal data rate (60 MB/s) of USB 2.0 and in fact the DAQ system interfaces to the host computer via 4 separate USB 2.0 ports. While the DAQ hardware can be easily scaled to support more channels, ultimately the limits of the host computer USB 2.0 capabilities provide a practical limit. Transition to USB 3.0 avoids this bottleneck, but requires significant redevelopment of the FPGA/USB interface board and associated software.

Chapter 5

System Packaging

5.1 Introduction

The recording arrays that we developed are composed of three basic modules: (1) a nanofabricated neural probe (nanoprobe), (2) a DAQ PCB for conditioning and multiplexing the electrophysiological data, and (3) a high-density flexible cable connecting the nanoprobe to the DAQ PCB. This chapter describes the methods used to assemble the three components into fully functional recording layers, and the methods for stacking such layers into dense 3-D recording systems.

We begin by describing the design and fabrication of the flexible cables. We then describe the use of a state-of-the-art flip-chip bonder to align and bond the flexible cable to the nanoprobe and DAQ PCB using anisotropic conductive film (ACF). Finally, we describe the successful assembly of full 3-D recording systems with 1024 recording sites.

5.2 Design and fabrication of high-density flexible cables

Mechanical and thermal decoupling between the nanoprobe and the DAq PCB can be done by using polymer-based cables [73]. They are more compliant than PCBs (FR-4) and have a low thermal conductivity. We have designed and fabricated flexible

cables made from Parylene (KNI) and Polyimide (Metrigraphics). The Parylene prototypes were important for evolving the design specifications and enabled fast testing of the cables, while the Polyimide cables provided a higher yield of functional connections and were fabricated with electroplated gold bumps at the interface matrix that facilitated bonding to the nanoprobes. The design and fabrication of both types of cables are described in more detail below.
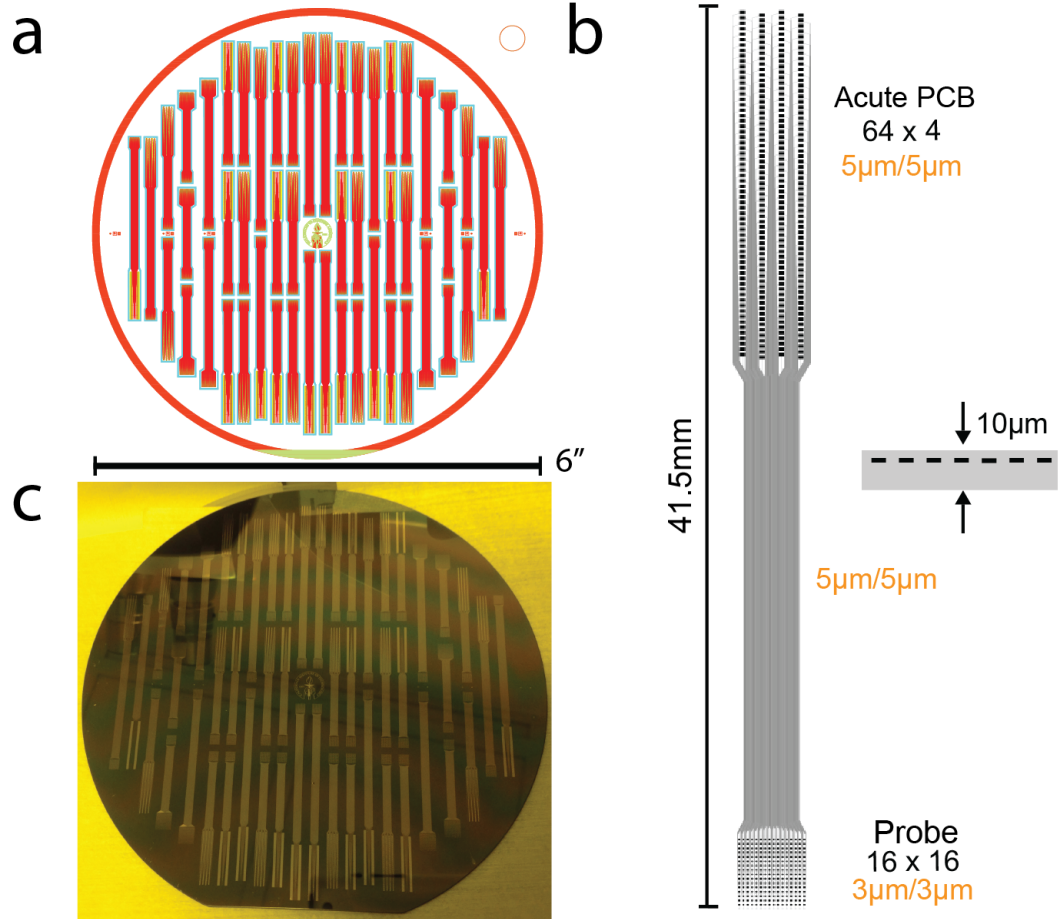

Figure 5.1: **Parylene flexible cables (ParyFlex)**. (a) AutoCAD drawing of wafer layout (v1). (b) Metal layout for the acute DAQ PCB cable. Orange text defines interconnect width/pitch. *Right* Cross-section cartoon of 10 μm thick cable. (c) Fully processed wafer prior to device release.

5.2.1 Parylene flexible cables

Two parylene flexible cable designs were prototyped using microfabrication. The first design is compatible with the Acute DAQ PCB (64x4, Figure 5.1b) and the

second with the Chronic DAQ PCB (128x2). Both designs are compatible with all nanoprobes through the standardized 16x16 interface matrix. Mask layout drawings were generated using AutoCAD software (AudoDesk, Figure 5.1a). All cable drawings were parametrically controlled and programatically generated using a custom LISP software extension suite developed by Evgueniy Lubenov at Caltech.

The flexible cables consist of 256 passive gold interconnects running between two layers of Parylene C. At the probe interface matrix, the gold interconnects have width/spacing of $3\mu m$ and then widen to 5μm as they merge out of this area. Cables are 41.5mm in length and only $10\mu m$ thick. They are the most compliant, 256 channel, flexible cables reported to date.

ParyFlex fabrication steps are outlined in Figure 5.2. Fabrication of the cables was performed at the KNI. Parylene C deposition was performed in the Roukes Lab. The full process requires four 7" photomasks (Figure 5.1a). Our fabrication procedures were adapted from similar previous efforts [74–76].
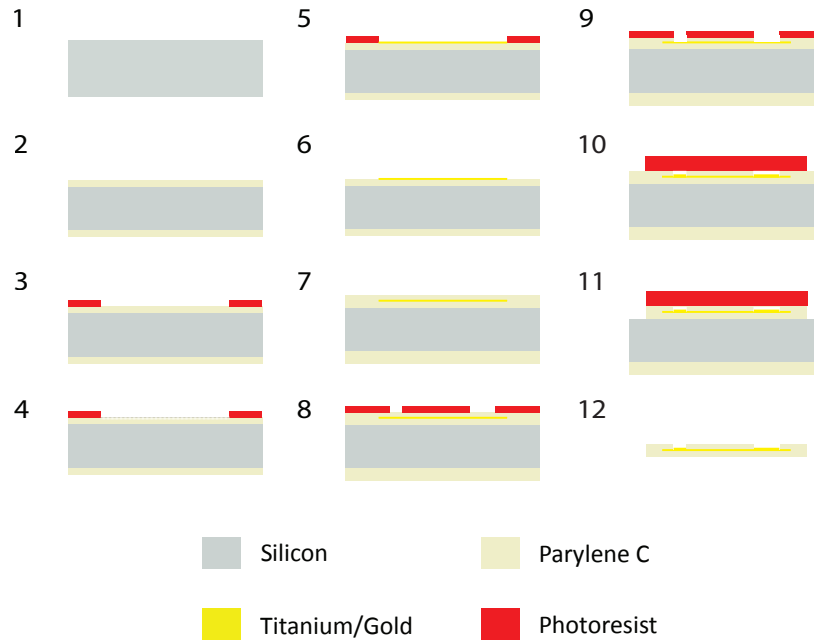

Figure 5.2: **Paryflex fabrication procedures performed at the Kavli Nanoscience Institute (KNI) at Caltech.**

Wafer preparation (1) Fabrication is implemented using 6" silicon wafers. The

wafer is thoroughly cleaned (acetone, IPA, dH2O)–any debris or dust at this stage can present itself later in the form of bumps on the bottom parylene layer, causing open circuits after metal definition.

Bottom Insulator (2) The bottom parylene C layer is deposited onto the silicon wafer (8 μm, SCS PDS 2010). Adhesion promoter is not used since final device release will rely on a weak adhesion between this layer and silicon. Instead, parylene C is allowed to deposit on both sides of a wafer. This creates a parylene shell around the wafer which keeps the parylene in place during further processing. Care must be taken to avoid puncturing this shell since it might lead to peeling during subsequent processing steps.

Metal Layer (3-6) The metal layer is patterned using vacuum-contact lithography (AZ5214 in image-reversal mode - 3s exposure). During exposure, a vacuum is drawn between the mask and the photoresist providing the closest contact possible. As a result, the 3 μm interconnects can be resolved (Figure 5.1b at Probe 16x16). The exposed photoresist is then developed (1:4 MIF400k:dH20, 30s) followed by an oxygen plasma ashing step (μRIE, 2min). (4) This ashing step roughened the exposed parylene substrate, increasing the surface area which helps increase Ti/Au adhesion. (5) Gold is deposited using electron beam evaporation (Lesker Labline, Ti/Au: 10nm/100nm). (6) Lift-off is performed in warm Remover PG (MicroChem, 80 degC, 5 minutes) with short bursts of ultrasonic agitation (2s, as needed). The wafer is then thoroughly cleaned (Acetone, IPA, dH20). Figure 5.1c shows the resulting wafer.

Top Insulator (7) A second layer of parylene C ($2\mu m$) is deposited. Prior to deposition, the wafer is soaked in silane to increase parylene-parylene adhesion (A174:IPA:dH20, 1:100:100, 45min).

Pad Etching (8-9) Pad openings are patterned (AZ4260, 4 μm) and the top parylene layer is etched (RIE, O2/CF4: 30/7sccm). For faster prototyping, the cables can now be released and singulated using scissors (Figure 5.3). To do

this, an incision is made around the perimeter of the wafer, cutting through the parylene layers and exposing the silicon wafer. The wafer is then submerged into a water bath where the parylene layers quickly peel off.

Cable Outline Etching (10-11) Cable outline is patterned (AZ9260, double-layer coat, $24\mu m$) and both parylene layers are etched.

Final Release (12) Finally, individual cables are released by submerging the wafer into a water bath. Table 5.1 summarizes the layers involved.

Table 5.1: **ParyFlex material stack summary.**

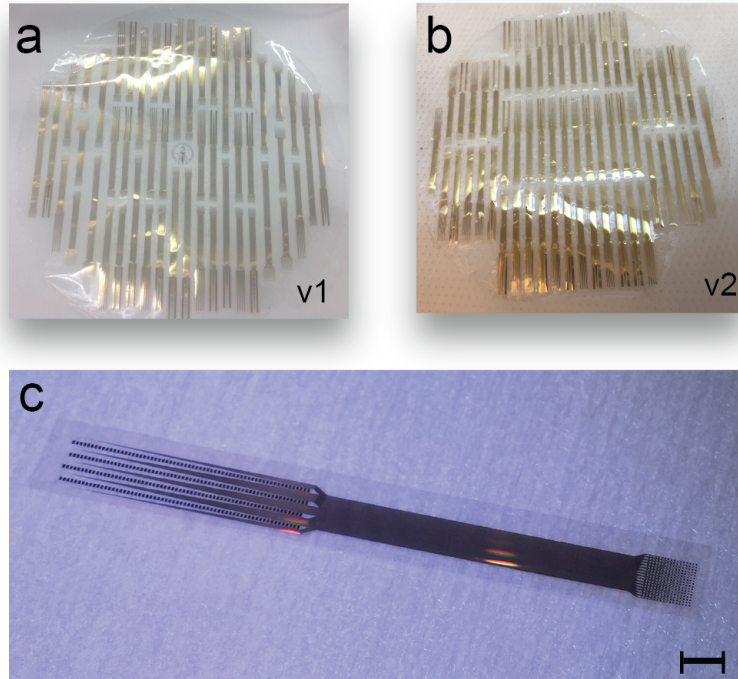| Layer	Material	Thickness $[\mu m]$
Top Insulator	Parylene C	2
Interconnect metal	Gold	0.1
Bottom Insulator	Parylene C	8

Figure 5.3: **Released ParyFlex cables**. (a) wafer layout v1 (b) wafer layout v2. (c) Image of an individual ParyFlex separated with scissors. Scale bar: 3.5mm

5.2.2 Polyimide flexible cables

As described in more detail below, successful bonding of the flexible cables to the nanoprobes is facilitated by the presence of metal bumps at the interface pads. These bumps can be placed at the interconnect matrix of either the nanoprobe or the flexible cable. To fabricate flexible cables with electroplated gold bumps, we worked closely with Metrigraphics, a company specializing in the fabrication of polyimide-based cables. Polyimide is more widely used for flex-circuit applications (Kapton flexPCBs) and is biocompatible and, chemically resistant and has great dielectric properties. Its main advantage over Parylene is that it adheres better to itself, providing a more reliable insulation layer.

Polyimide is popular amongst other neural technology efforts [6; 15; 18; 77–79]. It is used as a dielectric layer or to create similar interface cables. Most notably, it is also used by Neuronexus to manufacture their commercially available interface cables [80].

The fabrication of Polyimide based cables was performed on a 4x4 inch glass substrate. General process details are as follows:

  i Bottom polyimide layer (10 μm) is spun onto the glass substrate and cured.

  ii Interconnect metal layer is patterned using photolithography and metal is deposited using electroplating techniques(Au, 300nm).

  iii Top polyimide layer ($4\mu m$) is spun and cured.

  iv Interconnect pad locations are patterned using photo lithography and revealed.

  v Gold bumps are electroplated (20 μm) both on the nanoprobe and DAQ PCB interface pads.

  vi Outline of the flexible cable is defined using a laser cutter.

  vii Glass substrate is then diced forming glass rectangles around individual cables.

  viii Final release of polyimide cables is performed at Caltech.

Table 5.2: **PolyFlex material stack summary** Starting from top to bottom.

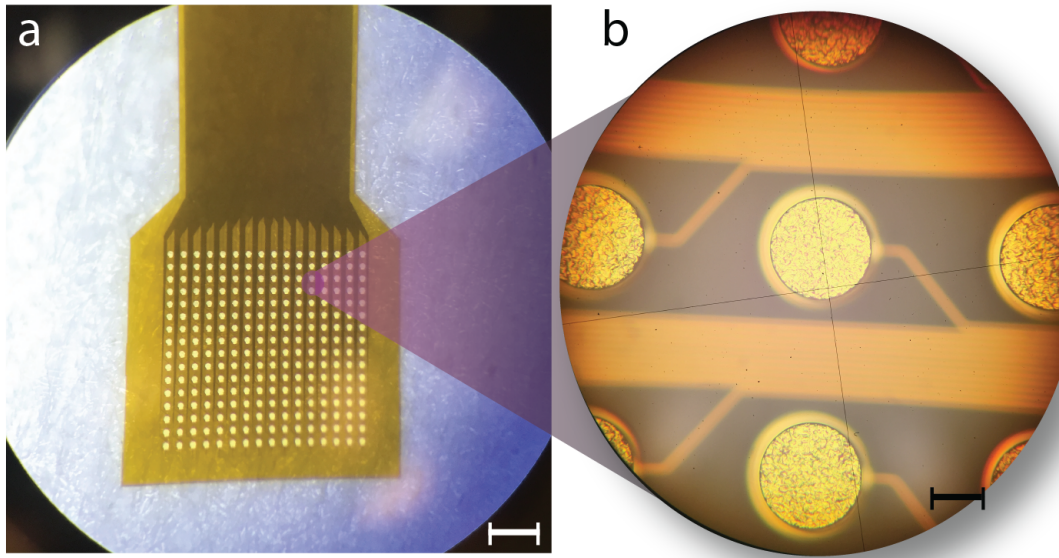| Layer	Material	Thickness [μm]
Interface Pad Bumps	Gold	20
Top Insulator	Polyimide	4
Interconnect metal	Gold	0.3
Bottom Insulator	Polyimide	10

Figure 5.4: **PolyFlex cables with gold bumps** a. Microscope image showing a gold-plated nanoprobe interface pads on a fully-released polyimide cable. Scale bar: 0.5mm. b. Close-up of 20 μm tall gold bump. Scale bar: 50 μm

5.3   System Assembly

This section describes our strategy for bonding the flexible cables to DAQ PCBs and to the nanoprobes.

5.3.1 Flip-chip bonding using Anisotropic Conductive Film

Flip chip bonding describes the method of electrically connecting the device to the package carrier [81; 82]. In contrast, to wire bonding, the interconnection between the device and carrier is made through a conductive bump that is placed directly on the device. The bumped device is then flipped over, aligned so that its pads align with matching pads and placed face down, with the bumps connecting to the mating substrate. Bumps are typically 50-120μm wide and 20-100μm tall. The electrical connection between the bumps and the substrate can be done in many ways. The most established way is using solder bumps and reflowing the solder to establish the connection (Controlled Collapse Chip Connect, C4). Alternatively, this weld can also be done using non-solder bumps. Gold bumps, for example, can be welded onto their mating pads (also gold) using a combination of temperature, pressure or ultrasonic energy (i.e. thermocompression, thermosonic, and ultrasonic bonding). For our application however, the high temperatures ($> 250°C$) and ultrasonic energy required to form these bonds can damage our nanoprobes.

An alternative approach is to use anisotropic conductive film (ACF) technology to form the electrical connection to the nanoprobes [81–84]. This film is a thermosetting resin filled with conductive beads separated from each other far enough such that the film is not conductive itself. As this film is pressed between a bumped die and a substrate, the beads will touch both surfaces and a conductive path will be established in one direction (Figure 5.5a). Heat is applied to cure the resin, keeping the beads trapped between the surfaces. This also establishes a mechanical connection. Since no metal weld is formed, process temperatures are lower and compatible with our nanoprobes ($< 190°C$). One major advantage is that ACF technology is fully capable of handling the electrical connection requirement on both the fine-pitch pads (100 μm) at the nanoprobes and the larger pads on the DAQ PCBs [84]. This means we don't have to alternate between flip chip bonding technologies to fully-package our system.

For the ACF process to work, accurate alignment combined with reproducible

bonding forces and temperatures is needed. For this purpose, a flip chip bonder is used (Figure 5.5b). A flip chip bonder is a machine that enables picking up a device using a bond-head, aligning the device to its corresponding carrier using an overlay vision alignment system (Figure 5.5c-d), and accurately placing the device onto the carrier applying a combination of force and heat (Figure 5.5e).
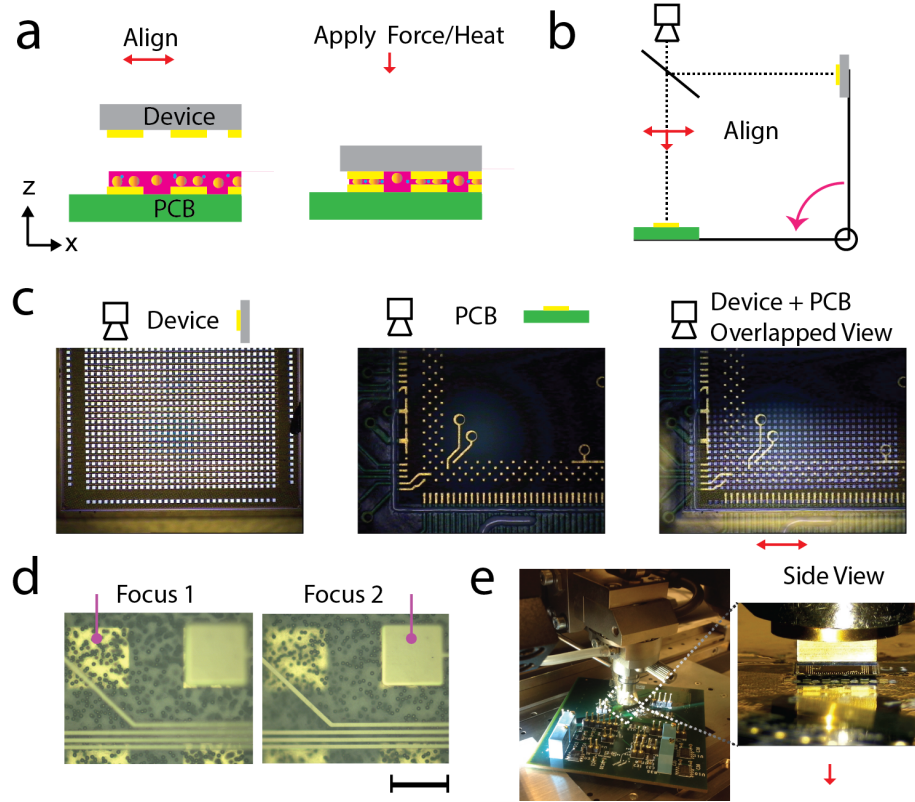

Figure 5.5: **Flip-chip bonding with anisotropic conductive film (ACF).** (a) ACF bonding requires accurate alignment and thermocompression—applying force and heat. Electrical conductivity is only on the z-axis. (b) Alignment is performed with a flip chip bonder capable of aligning with sub-micron resolution. (Finetech lambda).(c) Alignment is performed using an overlapped camera view (*right*) of the device (*left*) and PCB (*center*). (d) Overlapped close-up view of mating pads—one from the PCB and the other form the device—offset from each other, marked with purple circles. To assure alignment, both must be in focus. Scale bar: 100 μm (e) Side-view images showing an example of an ACF bond. *Left:* View of PCB while the bond head is forming the bond. *Right:* Close up view of the bond head placing a bumped device.

5.3.2   Bonding the flexible cable to the DAQ PCB

To create the electrical and mechanical bond between the flexible cable and the DAQ PCB we flip-chip bonded them together using ACF (HnS Hightech, TGP20505N). Pad alignment is performed using a flip chip bonder (Finetech lambda). The flex cable is picked up using a custom bond head while the ACF-tacked PCB is vacuum-held on the substrate of the bonder. We then apply force (150N) and heat (185° C, 10s) to crush and trap the conductive beads and cure the ACF's resin.

The trapped conductive beads (5 μm Ni beads, 1,500 beads/mm^2) provides the electrical connectivity between the interface pads. Contact resistance is directly proportional to the number of entrapped beads. Therefore to assure a low contact resistance ($< 1\Omega$), we had to fine-tune our process parameters (temperature, force, tacking-strategy) and design choices.

The mechanical connection is made by the cured epoxy resin in the ACF. It was important to chose an ACF that has a low coefficient of thermal expansion (CTE, 60ppm/$C°$) to help reduce thermomechanical stress caused by CTE mismatches between our parylene-based cables and FR-4-based PCBs. Otherwise we can experience mechanical damage that could lead to electrical failures.

One major advantage of using ACF at this interface is that the bond can be reworked. This means that we can remove the flex cable and re-use the DAQ PCBs if needed (acetone rinse + heat at 80° C). Secondly, the relatively low process temperature required to form the bond is compatible with our parylene-based cables (melting temperature: 290° C).

5.3.3   Bonding the flexible cable to the nanoprobe

Following a successful bond between the flexible cable and the PCB, completion of the assembly requires bonding the flexible cable to the nanoprobe. As described above, the ACF flip chip bonding works by trapping small conductive beads. However, if the interface pads are not protruding from the surface of the interface, as is the case with the nanoprobes and flexible cables, bead trapping may be difficult [85]. The

probability of successful connections can be significantly improved by creating bumps on the interface matrix. The bumps can be placed either on the interface pads of the probe, or the flexible cable. Section 5.2.2 describes the development of flexible cables with electroplated gold bumps. Here we also describe the alternative strategy of creating bumps on the interface pads of the nanoprobe itself.

5.3.3.1 Bonding flexible cables to stud-bumped nanoprobes

This can be achieved through a process called stud-bumping (Quik-Pak, San Diego,CA). This process is performed using a modified wire bonding machine. A gold wire (30 μm) is fed through a capillary tip where it is melted using a high-voltage electric charge. Because of surface tension, the molten gold forms into a ball and then quickly solidifies once heating subsides. This gold ball is then lowered onto a heated (120° C) pad and welded on using a jolt of ultrasonic energy. The trailing wire is then quickly terminated forming a stud. This is then performed for all 256 pads on the nanoprobe.

Figure 5.6 shows a fully stud-bumped nanoprobe. Notice a small length of the trailing wire is still present on all the studs (Figure 5.6c,d). This remnant is undesirable for ACF flip chip bonding for the following three reasons. First, the trailing wire is sharp and can puncture the parylene-based cable during bonding. Second, a flat bump surface is required to maximize bead contact area. Finally, bumps must have the same height to assure bead are crushed equally throughout the interface pad matrix.
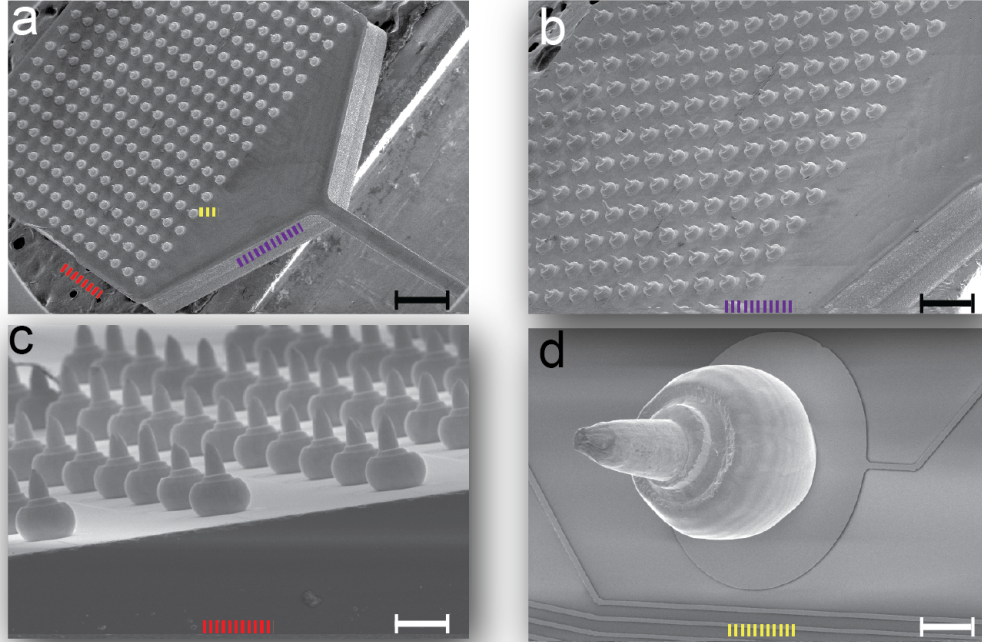
Figure 5.6: **Probes are post-processed by adding gold stud bumps on the interface matrix.** (a-d) SEM images. (a) Base of an one-shank probe device after post-processing. Colored dashed lines serve as view-point reference for (b-d). Scale bar: 1mm. (b) All 256 pads are bumped in a serial process using a modified, ball bond, wire bonding technique. Scale bar: 500 μm. (c) Bump height range is: 100-140 μm. Scale bar:80 μm. (d) Bump diameter is smaller than the output pad to allow lateral expansion during the coining step. Typical diameter range is: 70-85μm. Scale bar: 20 μm.

To eliminate these issues, we flattened (or "coined") the stud-bumps by applying mechanical pressure (400N) through a very smooth surface (polished Si die). Gold is malleable so applying mechanical pressure presses the trailing wire into the ball, flattening the surface. It also makes more uniform bump heights. Figure 5.7 shows a nanoprobe after coining. The initial bump height for this nanoprobe ranged from $100 - 150\mu m$. Through coining, the bumps were compressed down to 20 μm (Figure 5.7c) and the height variation minimized ($\pm2\mu m$). The main advantage for using this method is that it permits easy bumping of individual nanoprobes without the need for wafer-level processing. As long as the bond pad metal is compatible with wire-bonding (Au), gold studs can be placed. Even though it is a serial process, current equipment is capable of placing as many as 12 bumps per second (21 seconds per

nanoprobe), which is adequate for prototyping.

The bumped nanoprobe is flip chip bonded onto the flexible cable using ACF (HnS Hightech, TCF1051GY) through the same process as the flex to PCB bonding. The assembly procedures for a recording layer using the bumped nanoprobe devices is summarized in Figure 5.8.
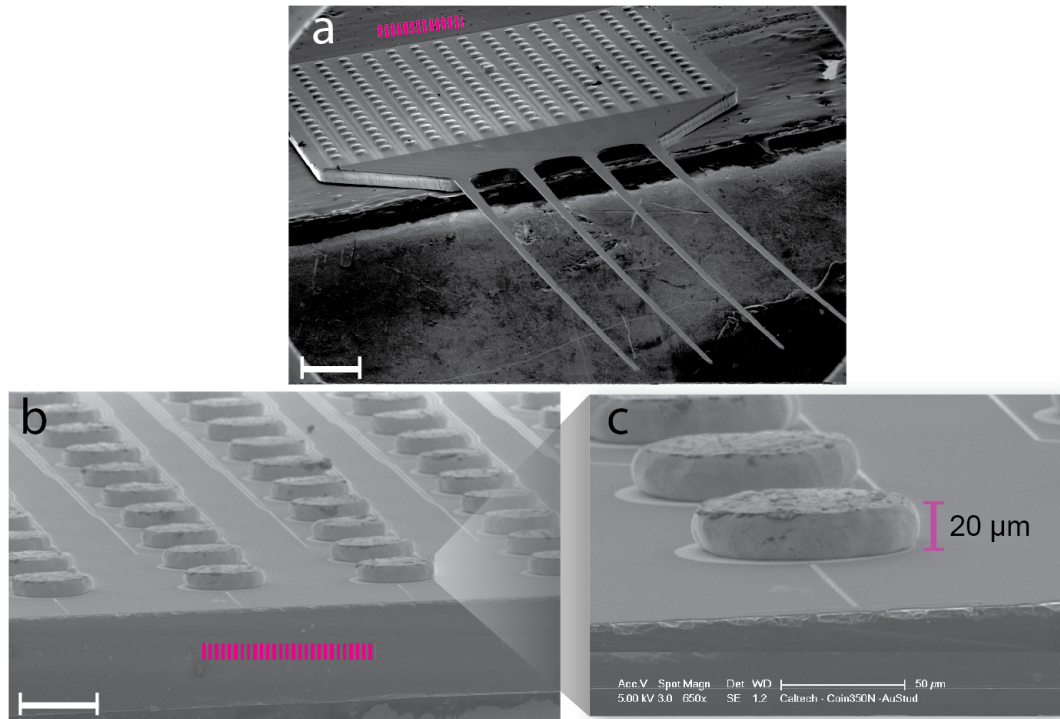

Figure 5.7: **Coined stud-bumps.** (a) SEM image of a 4-shank, Gen2 probe after stud-bump coining. Scale bar: 500 μm. Pink-dashed line indicates SEM view-point for (b). (b,c) Close-up SEM images of stud-bumps after 400N of mechanical pressure is applied. Scale bar for (b): $100\mu m$
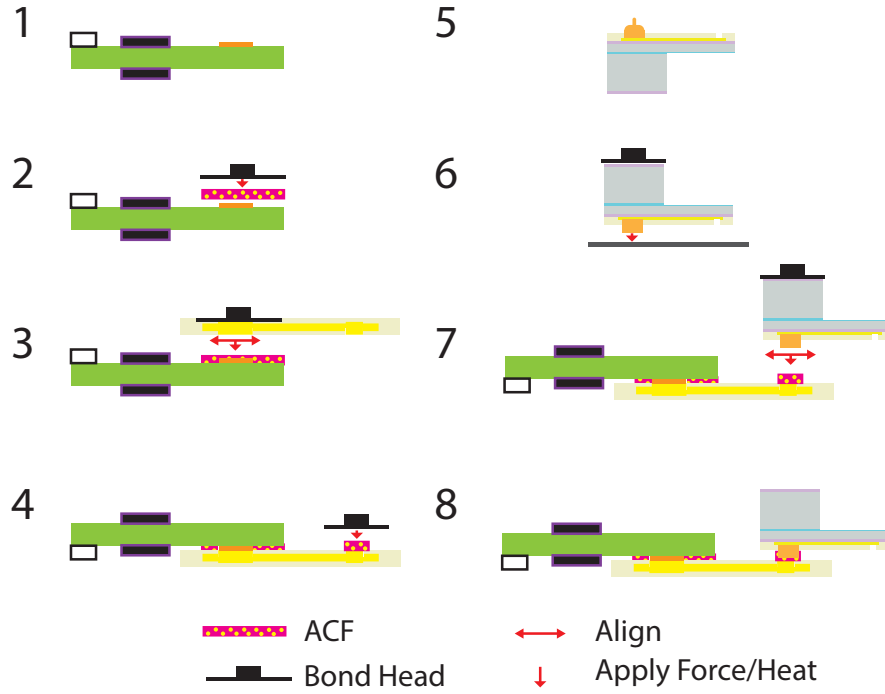
Figure 5.8: **Bonding of recording-layer using stud-bumped nanoprobes.** 1) The pads of the DAQ PCB are cleaned. 2) Flex-on-board (FOB) ACF is tacked. 3) FlexPCB is aligned and bonded. 4) Chip-on-flex(COF) ACF is tacked. 5) Gold-bumped nanoprobe is cleaned. 6) Gold-bumps are coined on a smooth substrate. 7) nanoprobe is aligned and bonded. 8) Fully assembled 2-D recording layer. *Bottom:* Symbol definition.

5.3.3.2   Bonding nanoprobes to gold-bumped flexible cables

A subset of the fabricated nanoprobes have a very thin base ($18\mu m$), making nanoprobe gold-stud bumping impractical. First, the thin base cannot be held in place by the mechanical clamps used during the stud bumping process. Recall that this process uses force, heat, and ultrasonic energy to form the gold stud bump. Therefore, improper coupling to the machine's substrate during processing would cause the process to fail. We tried solving the mechanical coupling issue by mounting the nanoprobe onto a silicon spacer (see Figure 5.12) but the epoxy film used to bond the layers did not hold up during bumping. The increase in temperature caused the epoxy to soften and ultimately to fail. Second, even if we succeeded in stud bumping the interface matrix, the high forces needed to coin the bumps would likely break the nanoprobe.

These ultrathin probes initially motivated the development of flexible cables with electroplated bumps, and can be bonded in two steps. First, the nanoprobe is picked up by a custom bond head and aligned onto the ACF tacked Polyimide flexible cable (PolyFlex). It is gently pressed onto the PolyFlex cable but not with high enough force to create the bond. The vacuum hole in the bond-head holding onto the probe would cause there to be an uneven force distribution on the nanoprobe. This would lead to fracturing. Instead, a double sided polished silicon spacer is picked up by the bond head and used to apply the required bond force and heat to create the COF bond. The assembly process using bumped flexible cables is summarized in Figure 5.8.
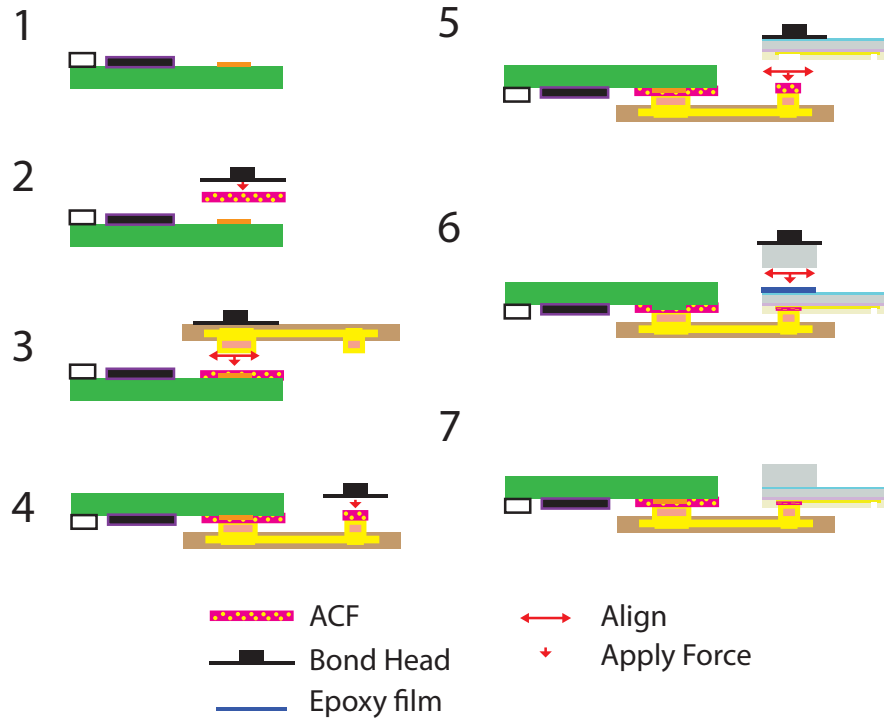

Figure 5.9: **Bonding of recording layer using gold-bumped flexible cables.** 1) The pads of the PCB are cleaned. 2) FOB ACF is tacked. 3) Gold bumped FlexPCB is aligned and bonded onto chronic PCB. 4) COF ACF is tacked. 5) Nanoprobe is aligned and bonded. 2-D layer is ready as is or a spacer can be subsequently mounted. 6) Silicon spacer—of any desired thickness—is aligned and bonded using ultra-thin epoxy film. 7) 2-D recording layer ready for 3-D packaging. *Bottom:* Symbol definition.

5.4   2-D recording layer assembly

In the previous section, we described two packaging approaches used to interface the nanoprobes to the DAQ-PCBs. Both, have led to successfully building fully-functional, 2-D recording layers. Assembly yield—defined as the successful formation of both electrical and mechanical contact between each layer—is high as discussed further in Chapter 6.

Figure 5.10a shows a fully assembled 2-D recording layer packaged together using a stud-bumped nanoprobe (KNI), a parylene-based cable (ParyFlex), and an acute-PCB.

Figure 5.10b shows a fully assembled 2-D recording layer packaged together using an ultrathin nanoprobe (LETI), a gold-bumped polyimide-based cable (PolyFlex), and an acute-PCB.
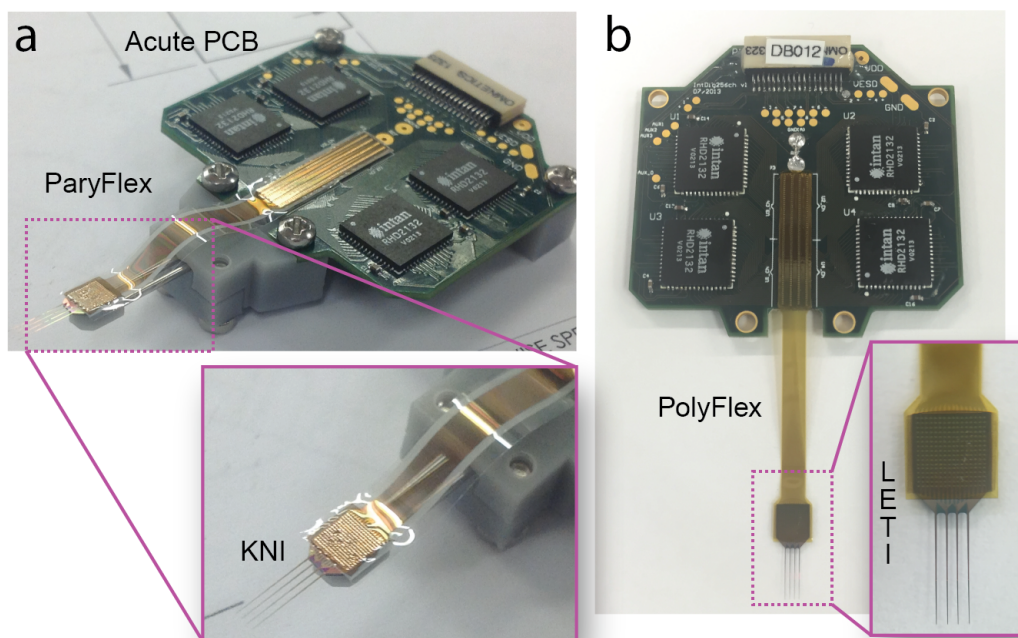

Figure 5.10: **Fully-assembled 2-D recording layers.** Nanoprobes connected to a DAQ PCB via a Parylene (a) or Polyimide (b) flexible cable. The probe in (a) was fabricated at the Caltech KNI, while the probe in (b) was fabricated at LETI.
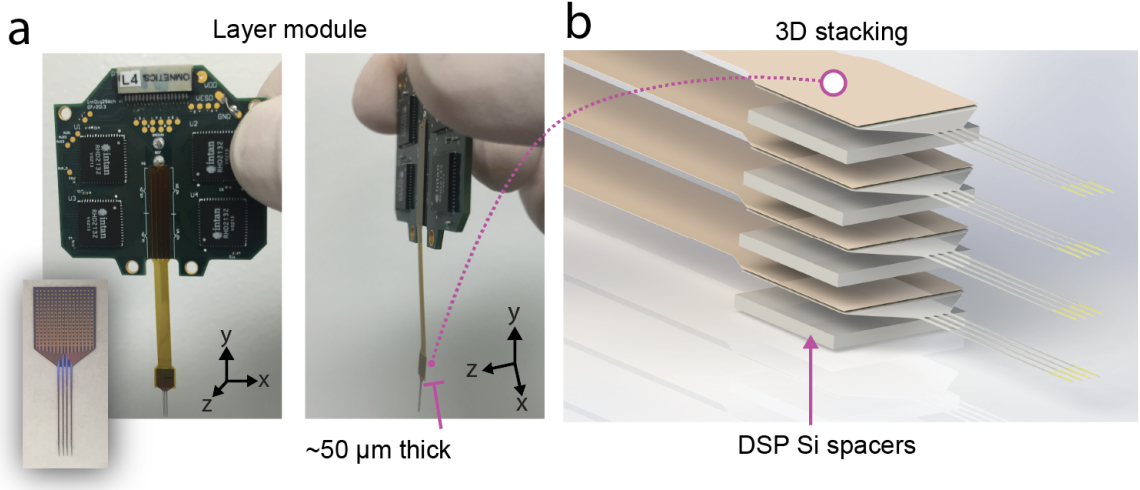
Figure 5.11: **2D recording layers as modules for 3D recording systems.** (a) Module: fully-assembled single recording layer. Side-view image (*right*) shows a thin COF interface of approximately 50 μm thick—which includes the thickness of the probe + gold bumps + cable + epoxy film. *Bottom-left:* Image of the 4-shank, 250 μm shank-pitch, probe used in this layer. (b) Cartoon showing the 3D-packaging concept. DSP silicon (Si) spacers separate each recording layer with a pitch equal to the thickness of the chosen spacer plus 50 μm (for COF thickness). Entities can be temporarily bonded together using PEG or permanently bonded using ultra-thin, die-attach epoxy film (AiT Technology). Stacking alignment is performed using our flip chip bonder which provides sub-micron resolution (0.5nm).

5.5 3-D array assembly

The modular design of our system enables stacking multiple 2-D layers into dense 3-D configurations (Figure 5.11). Each 2-D layer can be assembled using any of the available nanoprobe designs. Separation between layers is achieved with silicon interposers. The minimum separation between layers is $50\mu m$, the total thickness of the nanoprobe, ACF, and polyFlex cable. These interposers are silicon rectangles that are precisely diced (diamond saw) from double-sided polished (DSP) silicon wafers (Ultrasil, Wafer thickness range available: 50-1000 μm).

Figure 5.12 shows that nanoprobes can be properly mounted on silicon interposers (300 μm). The nanoprobe is aligned and placed using the flip-chip bonder. Similarly, we use the flip-chip bonder to align and place each module on top of each other, one layer at time. To do this, we pick up each module using a custom bond-head

integrated with custom 3-D printed clamp. The bond head picks up the nanoprobe region using vacuum while the 3-D printed clamp picks up the PCB. The flexible cable is left floating. Alignment is performed at the shank-tip level between each layer ($< 1\mu m$ resolution). Modules are bonded together at the nanoprobe level, either using thin epoxy film or poly(ethylene glycol) (PEG).
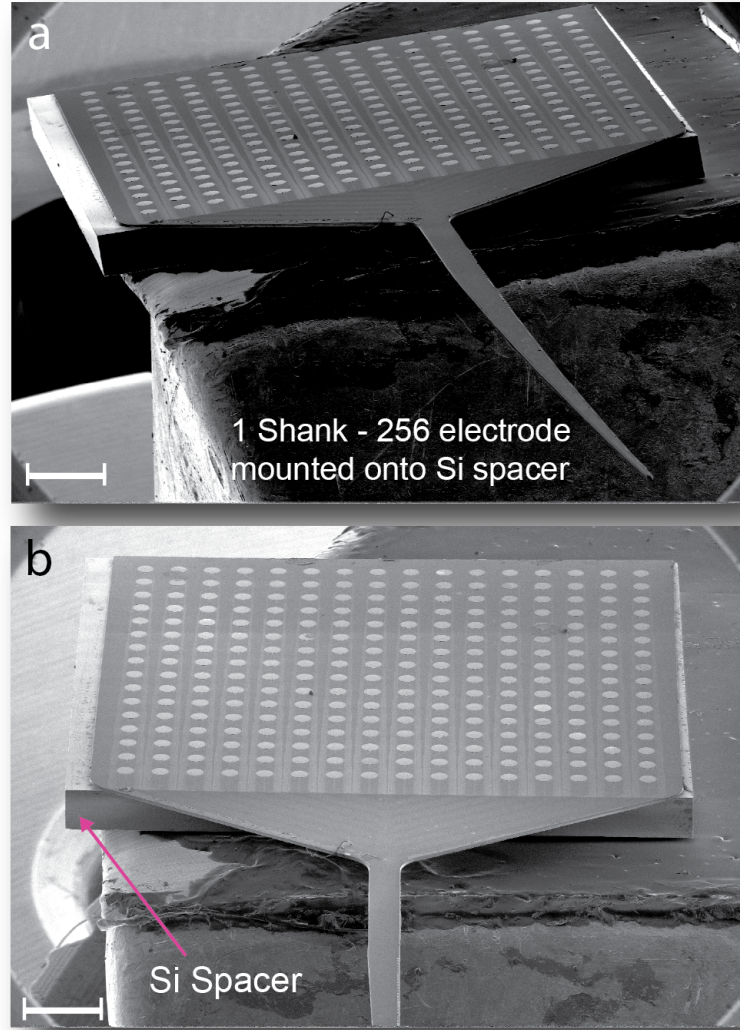

Figure 5.12: **Precision mounting of ultra-thin nanoprobes onto silicon interposers**. Nanoprobes with $18\mu m$ thick base are mounted onto double-polished, diamond-sawed, silicon (Si) interposers using epoxy film. (a,b) SEM images of a single-shank probe mounted onto a 300 μm thick Si spacer. Scale bar: 500 μm.

To demonstrate the power of this stacking approach, we assembled four 2-D layers into a fully-functional 3-D assembly for dense *in vivo* recordings (Figure 5.13). The

DAQ PCBs are separated using brass spacers which provide electrical connectivity between layers to form a common ground, as well as a low-resistance path for heat dissipation (heat-sinking). Separation between PCBs also creates air-flow paths for convective cooling. The nanoprobes are held together using high molecular-weight PEG (Sigma Aldrich). Finally, the full assembly is mounted onto a 3-D printed holder which fits a precision manipulator used for brain insertion (Figure 5.13b). The resulting 3-D electrode grid contains 1024 electrodes within 0.6 mm^3 ($750\mu m \times 1050\mu m \times 756\mu m$), which corresponds to each electrode covering a unit cube with edge length of less than $100\mu m$ (Figure 5.14).
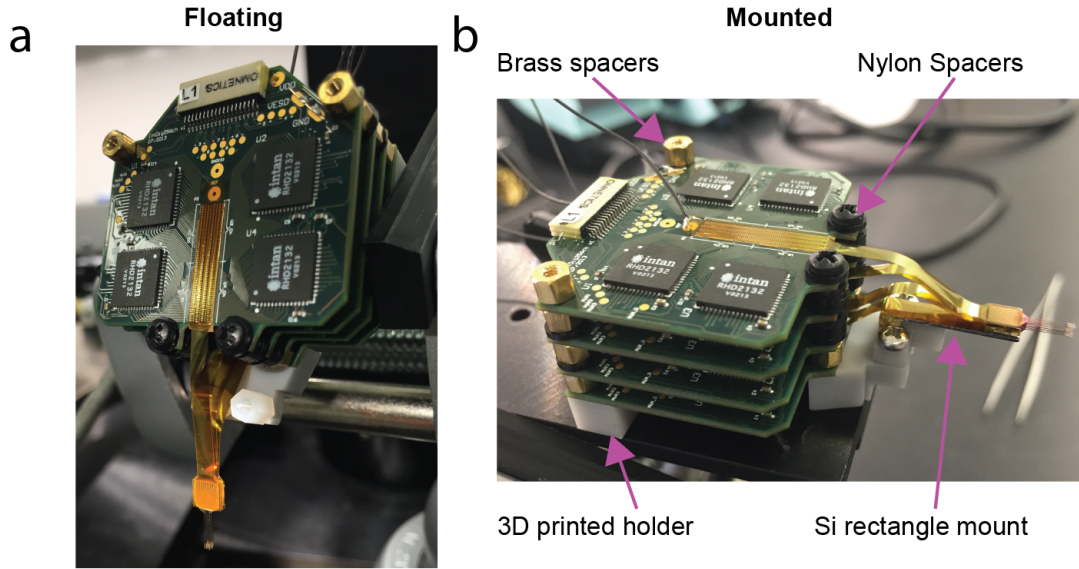

Figure 5.13: **Fully-assembled, 1024-channel, 3-D recording system.** (a) Unmounted or "floating" assembly. Probe-stack is not rigidly held and it is mechanically constrained only by the polyFlex cables. (b) Image taken after the temporarily mounting the probe-stack onto a rectangular silicon piece using PEG (MW: 3000, Sigma Aldrich). Acute PCBs are spaced apart using 3mm tall, brass spacers on one end. On the other end, the PCBs are spaced using nylon spacers with a 12mm long nylon screw passing through them.
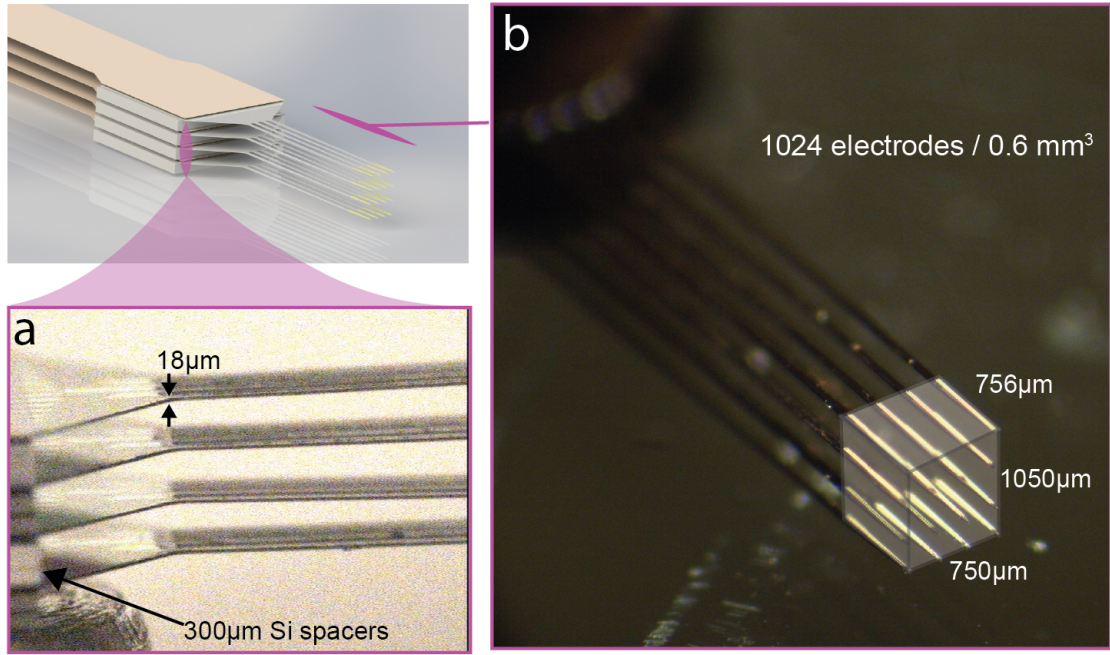
Figure 5.14: **3-D recording electrode grid.** (a) Side-view image of the probe stack showing the 300 μm silicon spacers precisely separating each probe layer. (b) Stereoscope image showing all 16 shanks. Shanks are 5mm long but contain recording electrodes on only $756\mu m$ of their length. The recording volume created by the electrode grid highlighted in white. The resulting recording density is 1024 electrodes in 0.6 mm³, which constitutes the densest 3-D recording density reported to date. *Top-left:* Cartoon used to reference view-point for images.

5.6 Conclusion

We developed a 3-D packaging strategy which is modular and scalable (Figure 5.15). Stacking each layer using a flip-chip bonder provides the following advantages. First, it allows us to stack each layer with any desired alignment offset (Figure 5.15a). This is desirable when targeting curved deep structures in the brain, like the hippocampus. Second, the recording electrode grid is highly configurable through choice of nanoprobe model (Figure 5.15c). Third, multiple layers can be assembled. With our current modules, we can potentially scale our layer count up to eight (Figure 5.15b)—creating 2048-channel, 3-D recording systems. Further scaling of layer count is limited more by the length of the flexible cables and by the acquisition system's data handling capacity rather than by the stacking strategy.

This packaging strategy also enables combining nanoprobes arrays with other types of probes, such as probes with optical elements [86; 87] or microfluidic channels [9; 50; 88–91]. It is also compatible with replacing silicon as the nanoprobe substrate with a more flexible material like parylene or polyimide [8; 11; 16; 18; 92; 93] (Figure 5.15d).
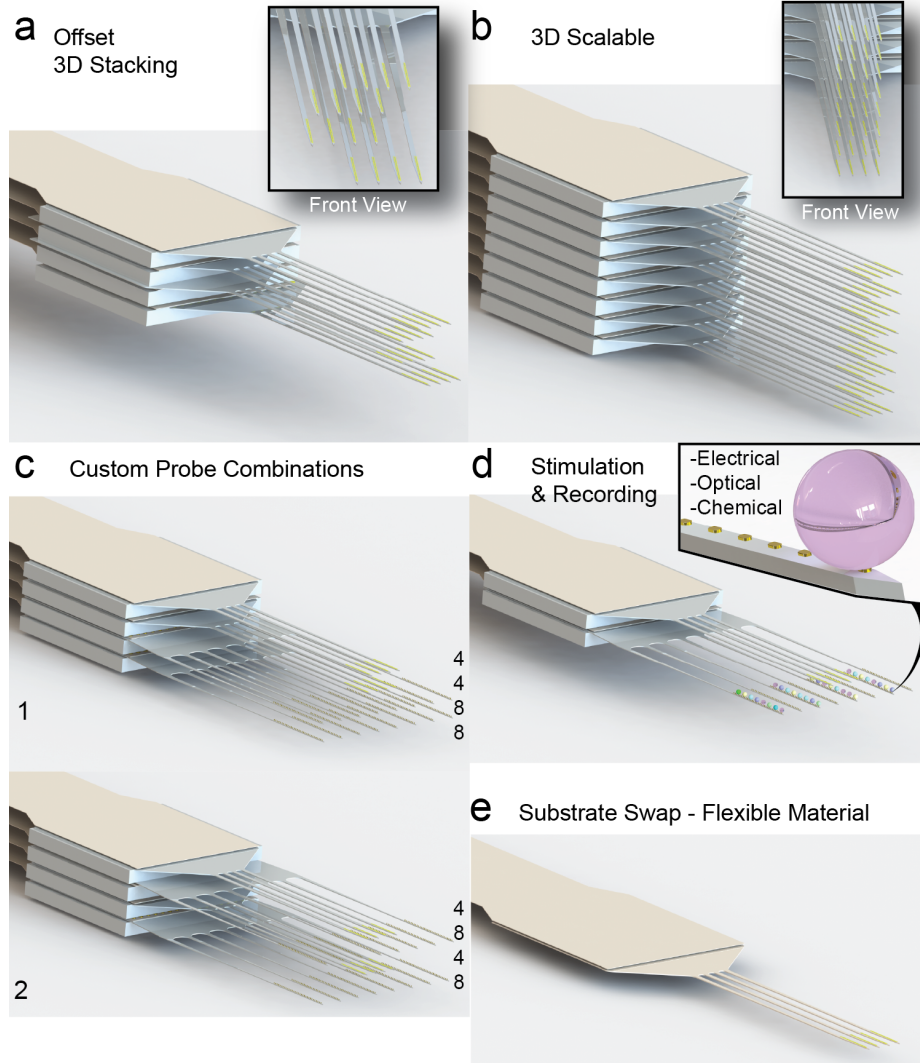
Figure 5.15: **The 3-D stacking strategy is scalable and adaptable** (a) Layers can be stacked with any desired lateral or vertical offset. (b) Maximum layer count is only constrained by flexible cable length and acquisition system capacity. (c) Probe selection and stacking order can vary. Vertically arranged numbers indicate the number of shanks on the probe. (d) Different types of probes can be stacked together. (e) Silicon based probes can be replaced with probes with flexible substrates.

Chapter 6

System Characterization and Benchtop Testing

6.1 Introduction

In this chapter we characterize the nanoprobe structural features that are important for reliable implantation and biocompatibility, and the electrical properties of the fully assembled system.

In particular, we verify that proper stress compensation has been achieved during the nanoprobe fabrication process, resulting in straight shanks with minimal deflection. This is critical for ensuring smooth insertion of the probes into the brain and minimization of tissue damage. Furthermore, we use scanning electrode microscopy (SEM) to visualize a cross section through the nanoprobe shank that reveals the proper size and alignment of fabricated features and enables verifying that shanks are covered with Parylene to enhance biocompatibility.

In addition, we describe tests of connectivity across the fully assembled system, and characterize electroplating of the recording sites. Electroplating increases the surface to volume ratio of the deposited gold, which enhances sensitivity of the recordings while maintaining selectivity in a small volume of tissue [94–98]. Finally, we map the impedance profile of the recording sites and fit an equivalent circuit model that enables cross-talk analysis. Our analysis establishes that the fully assembled system achieves a low-noise and low-crosstalk performance.
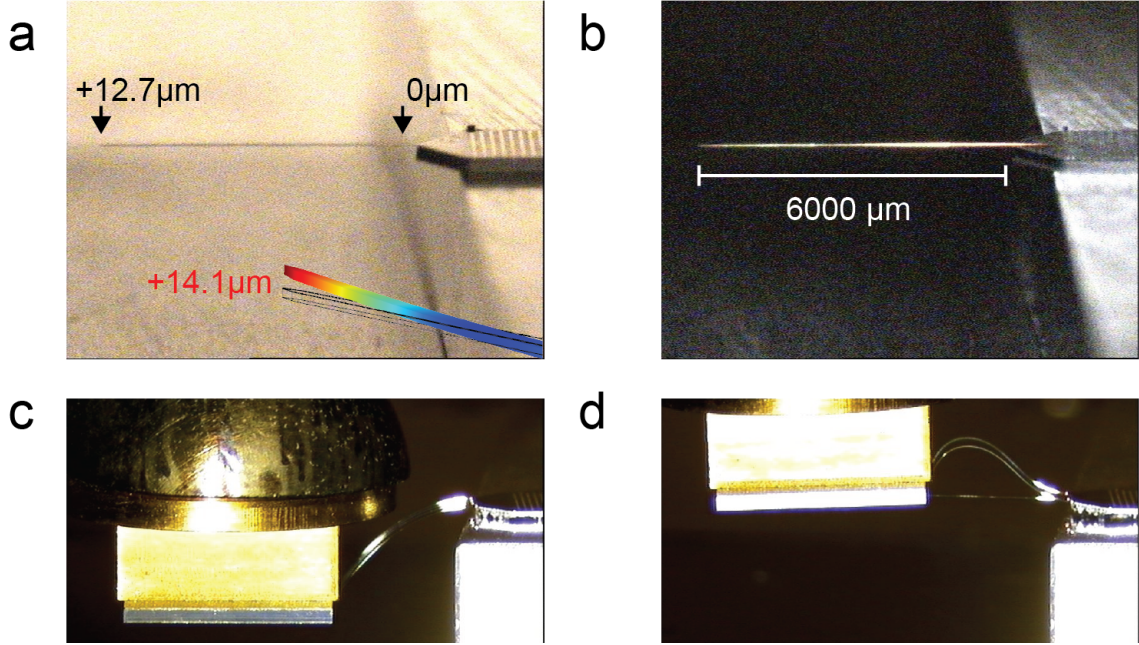
Figure 6.1: **Mechanical profile of nanoprobe shanks.** (a,b) A nanoprobe with a single 6mm long shank shows an upward deflection of 12.7 μm. The base of the shank is defined as having 0 μm deflection. The inset in (a) shows that the shank deflection measurement is in agreement with the expected 14.1 μm deflection derived from COMSOL simulations (see Chapter 2). (c,d) Shanks are flexible and can withstand forced buckling. The flip-chip bonder head is used to apply forces perpendicular to the shanks (c), as well as as head-on (d), revealing remarkable flexibility of the shanks and resistance to breaking.

6.2 Characterization of nanoprobe shanks

6.2.1 Shank deflection

To verify that the shanks have been properly stress compensated, we measured the displacement at the tip of the shank under a microscope with a high resolution ($<$ $1\mu m$) stage. We first brought the base of the shank into focus and noted the position of the microscope stage. We then brought the tip of the shank into focus and measured the difference in stage positions, which corresponds to the deflection of the shank off its neutral axis. Figure 6.1a,b shows the results of this measurement for a 6mm-long nanoprobe shank. The shank tip deflects upward (top of image) by 12.7 μm, which is in close agreement with the expected shank deflection previously determined through simulation ($14.1\mu m$, see Chapter 2 for simulation). We also verified that

shanks are structurally stable, and can withstand significant force without breaking (Figure 6.1c,d).
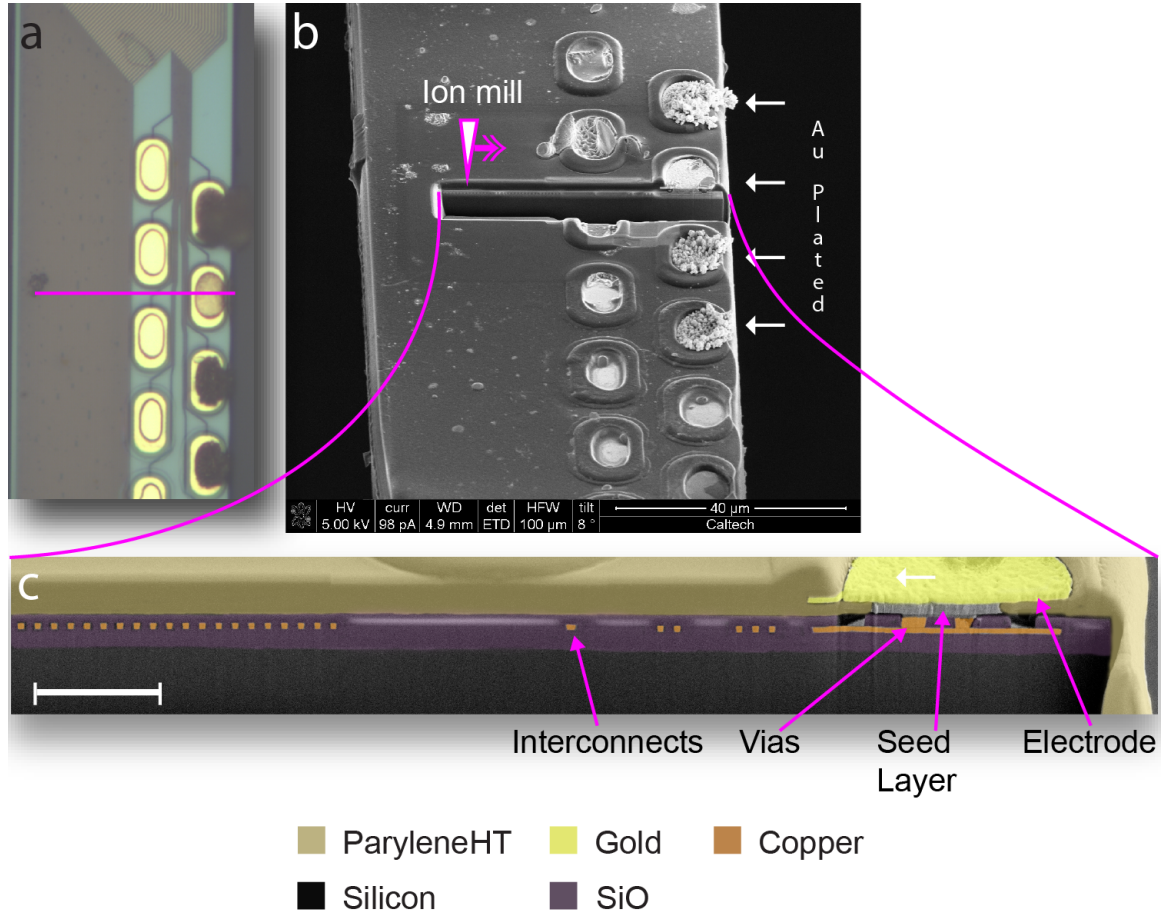

Figure 6.2: Anatomy of a nanoprobe. (a) Microscope image of a nanoprobe with two rows of recording sites. The purple line indicates the location of a cut performed using a focused ion beam machine (Nova 600). (b) SEM image of the probe in (a) taken after ion milling a portion of the shank. Four electrode sites (top-right) have been electroplated with gold. SEM scale bar: 40μm. (c) Pseudo-colored SEM image of the exposed cross-section. Color legend defines the materials involved. Scale bar: 8μm.

6.2.2 Parylene coverage

To confirm the proper fabrication of the nanoprobes we milled a cross-section using a focused ion beam (FIB, NOVA 200) across the edge of a shank and through a recording electrode (Figure 6.2a,b). This cross-section reveals proper dimensions and alignment

of the fabricated features (interconnects, through-vias, electrode and associated seed layer), and enables verifying that Parylene HT wraps around the sides of the shank (Figure 6.2c). This confirms that the nanoprobes are fully covered with biocompatible materials (Parylene on three sides, and silicon oxide (glass) on the fourth).
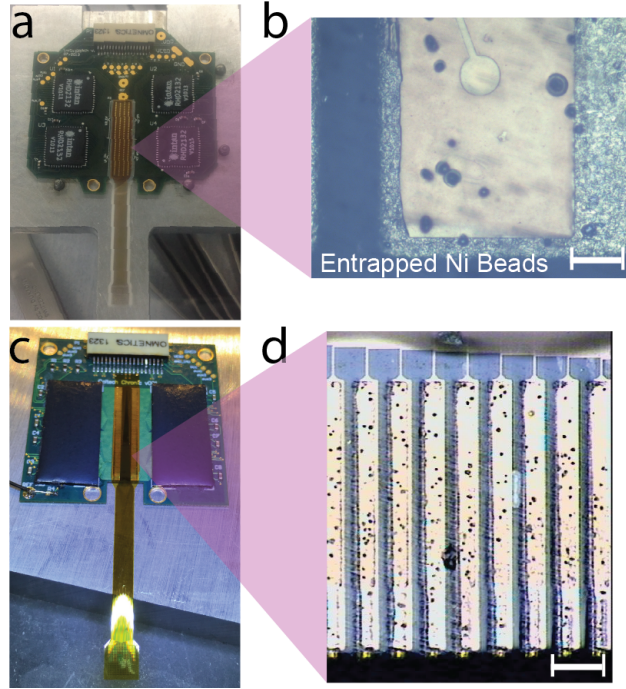

Figure 6.3: **Bond between a flexible cable and acquisition PCB**. (a,b) An acute (a) and chronic (b) PCB with a bonded flexible cable. (b,d) Magnifications of the bonded segments that reveal entrapped and crushed nickel beads (black circles), an indication of a successful electrical bond (Scale bar: $100\mu m$).

6.3   Connectivity tests

The nanoprobe system is composed of several modules (nanoprobe, flexible cable, acquisition PCB) which need to be bonded together. Achieving high yield on the high-density connections across these modules is challenging.

The fastest feedback for debugging and improving these bonds is through visual inspection. Examples of flexible cables bonded to acquisition PCBs and to a nanoprobe are shown in Figure 6.3, 6.4, respectively. As described in Chapter 5, anisotropic conductive film (ACF) containing conductive beads is used for the bonds.

To determine whether a bond was successful we can test whether all interface pads contain entrapped beads. Examples of entrapped beads for a portion of pads are shown in Figure 6.3b,d and Figure 6.4c (black dots). The contact resistance is inversely proportional to the number of entrapped beads. In case of bonding failures, the flexible cable can be removed by heating the assembly to 80° C to soften the ACF and gently pulling on the cable. Cleaning the interface with acetone allows for a new flexible cable to be bonded.
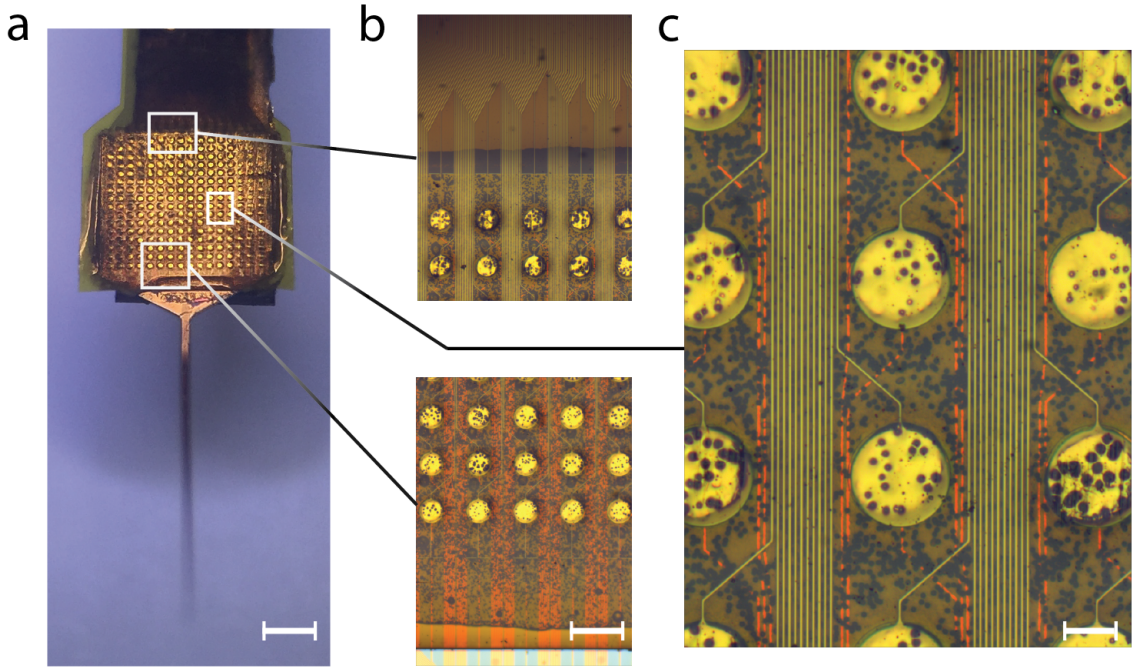

Figure 6.4: **Bond between a flexible cable and nanoprobe base.** (a) Nanoprobe with a flexible cable bonded to its base (Scale bar: 1mm). (b) Anisotropic conductive film (ACF) is used for the bond and remains contained around the perimeter of the input matrix leaving the shank region intact. Scale bar: 300 μm (c) Microscope image showing entrapped and crushed gold and nickel beads, and indication of successful electrical bond (Scale bar: 80 μm).

While visual inspection is useful for providing fast feedback for fine tuning the bonding parameters in different stages of the assembly process, electrical measurements on the fully assembled system are needed to quantitatively assess the quality and yield of the connections. Failures can occur either because (1) a connection is **high impedance** or completely open, or (2) because two connections are **shorted**

together.

We identify high impedance connections by exploiting their higher susceptibility to noise. In particular, we recorded broadband signals (0.1Hz - 10KHz, 30 KHz sampling rate) while electrodes are submerged in a grounded solution. In this configuration, the noise pickup is significantly higher for high impedance connections. Figure 6.5a shows the distribution of power spectral density (PSD) values at 60Hz over 256 channels of a nanoprobe with four shanks. Notice that the distribution is bimodal with most channels exhibiting low levels of 60 Hz noise ($\approx$ -110dB), and only four channels exhibiting high 60 Hz pickup (>-70 dB). To determine the location of these broken channels, we can map the PSD values to the physical location of four different interfaces (electrodes, pads, traces, PDB pads, Figure 6.5b). Contiguous broken channels in one of these maps can point to a specific failure point, and have helped in debugging some of the connection issues in early prototypes.
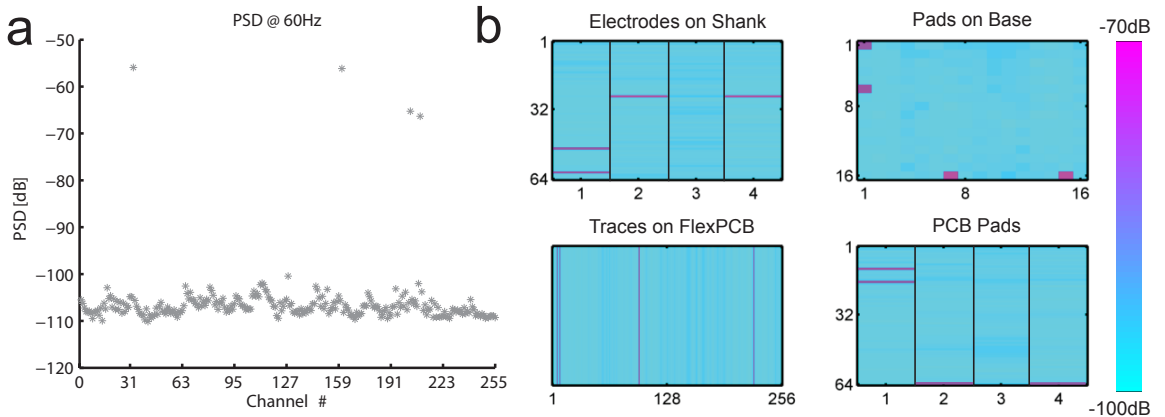
Figure 6.5: **Identifying high impedance connections.** (a) Power spectral density (PSD) at 60Hz of recordings from all 256 channels of a nanoprobe submerged in a grounded solution. In this assembly, 4 out of 256 channels are high impedance connections ($PSD_{60Hz} > -70dB$) (b) PSD at 60 Hz mapped according to physical location across four different interfaces: (1) electrode position on the shanks, (2) location on the connection grid at the base of the nanoprobe, (3) flexible cable trace location, and (4) location on the connection grid of bhe PCB.

To identify shorted channels we exploit the fact that they would exhibit a much higher degree of correlated noise. In particular, we record the broadband signals across all channels in a Faraday cage, a context in which common noise pickup is

minimized. In this configuration, the correlations across shorted channels is close to 1, significantly higher than the average correlation across non-shorted pairs. Figure 6.6a shows an example of the distributions of the highest correlations between each channel and every other channel in the nanoprobe. Notice that only one channel exhibits pairwise correlations close to one with a different channel. The resulting pair represents shorted connections in the nanoprobe assembly. By mapping the location location of the channels across the four different interfaces (electrodes, pads, traces, PDB pads, Figure 6.6b) we can identify the failure point across the assembly. For example, in Figure 6.6b , the correlated channels appear as neighbors at the level of the flexible cable and, hence this is likely the failure point in this specific assembly.
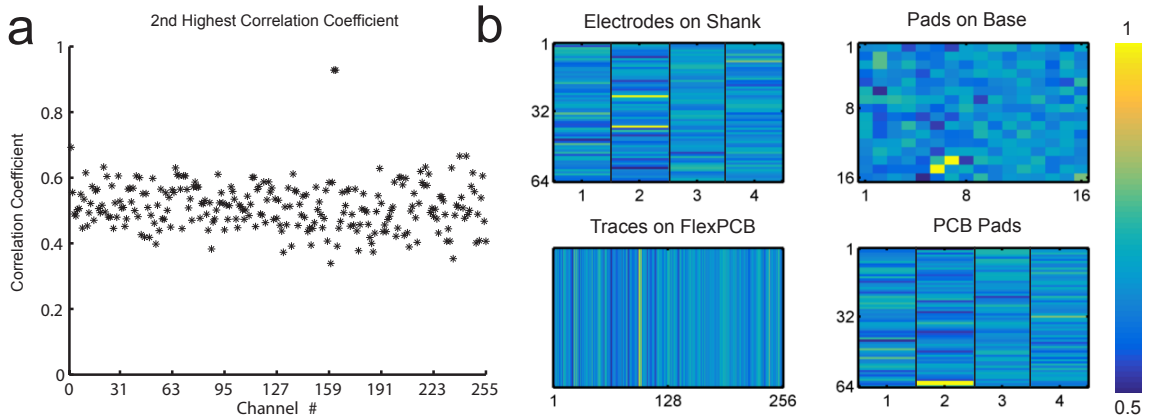

Figure 6.6: **Identifying shorted connections**.(a) Scatter plot showing the highest correlation coefficients between each recording channel and every other channel in the array also analyzed in Figure 6.5. Notice that there is one outlier with a correlation close to 1 with another channel. (b) Maximum correlations for each channel mapped according to physical location across four different interfaces: (1) electrode position on the shanks, (2) location on the connection grid at the base of the nanoprobe, (3) flexible cable trace location, and (4) location on the connection grid of bhe PCB. Notice that the short likely occurred at the level of the flexible cable.

A high yield of successful connections has been achieved in our assemblies. For example, the array analyzed in Figures 6.5, 6.6 shows only 4/256 open channels, and only 2/256 shorted channels, with an overall yield of 97.6 % (250 out of 256 functional channels). The 3-D assembly shown in Figure 5.14 has a ≈ 94% yield (926/1024 functional channels).

6.4 Recording electrode characterization

Electroplating with gold has been used to increase the surface to volume ratio of the electrodes. This decreases the electrode impedance, thus reducing cross-talk and thermal noise. Electroplating protocols can increase the roughness of the gold deposits while keeping the volume occupied low. This results in increased signal to noise ratio, while preserving selectivity within a small volume of tissue. The latter is critical for proper isolation of single unit activity (Chapter 7).
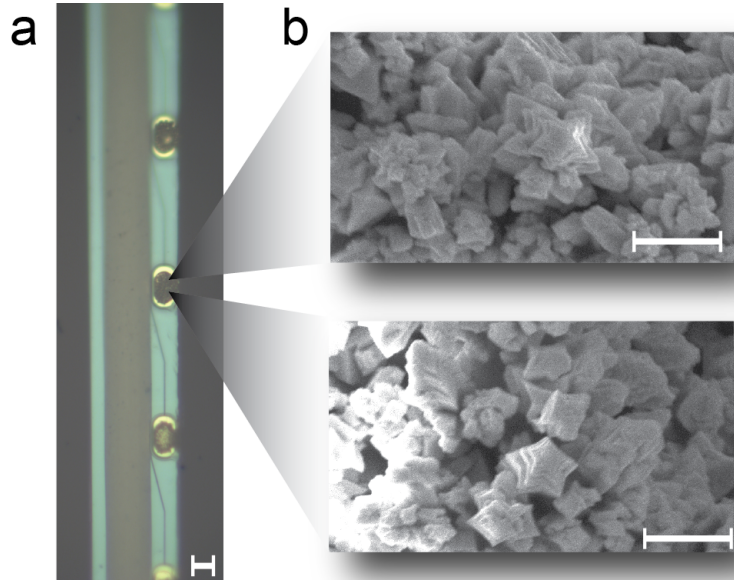

Figure 6.7: **Recording electrodes are electroplated with gold to increase their effective surface area.** (a) Microscope image showing 3 consecutive, gold-electroplated recording electrodes (5355 gold, Sifco). The electrode pitch is $50\mu m$. The lateral growth of the electrode must be kept minimal while increasing the surface area through the roughness of the deposited gold. Scale bar: $8\mu m$. (b) SEM images showing examples of the rough surface of the deposited gold. Scale bar: 500nm.

To increase the effective surface area of the recording electrodes we electroplated them in gold solution (Sifco, 5355 gold) using a Gamry Reference 600 potentiostat. We evaluated different electroplating profiles using a probe station setup prior to electroplating fully assembled nanoprobes. In this setup we created a gold solution droplet around the shanks of a nanoprobe and submerged a probe tip ($10\mu m$, Tungsten electrode) into the droplet. A second probe tip contacts an interface pad at the base of the nanoprobe. Using the potentiostat we experimented with different current

or voltage profiles applied across the two probe tips. Figure 6.7a shows an example
of three recording sites electroplated using this setup by applying a constant current
of -100nA for 10 seconds, five times. Inspecting the resulting deposits through SEM
reveals a high degree of surface roughness in the deposit while keeping the volume
constrained within the recording site (Figure 6.7b).

The same electroplating protocol has been applied across the fully assembled ar-
ray by passing current through the Intan chips. These chips can be used to route the
current applied through the potentiostat to different electrodes in the array. This pro-
cess has been automated for efficient electroplating of the large number of recording
sites.
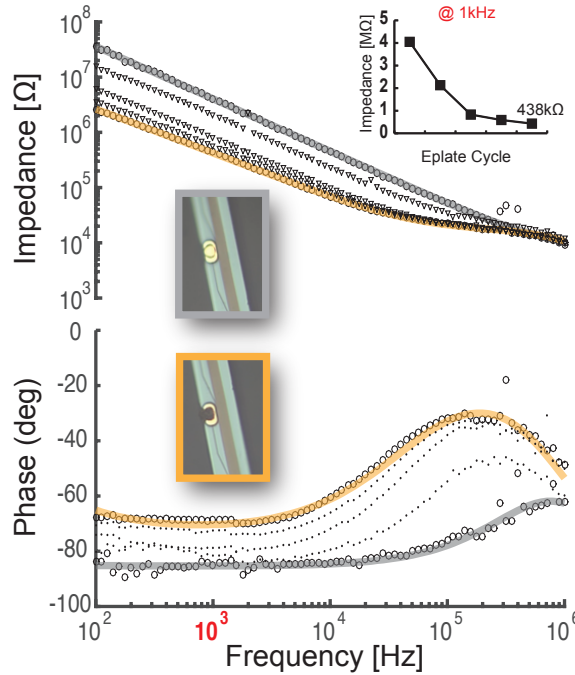
Figure 6.8: **Electroplating decreases recording electrode's impedance.** Plot
shows the evolution of the electrochemical impedance spectra (EIS) as the corre-
sponding recording electrode undergoes 4 cycles of gold electroplating. For each cycle
a 1 μA current is passed for 2 seconds. The inserts show images of the unplated
(top) and plated (bottom) electrode. The plot on the top right shows the electrode
impedance at 1 KHz as a function of the electroplating cycle.

6.4.1 Electrode impedance profile

To quantitatively characterize the electrode-electrolyte interface we measured electrochemical impedance spectra (EIS) for the electrodes and used the data to fit an equivalent circuit model.

To that end we passed sinusoidal current of varying frequencies through the electrode while the array is submerged in PBS within a Faraday cage. The current is passed via the Gamry potentiostat in a 3-cell configuration, with the array electrode as the working electrode, a Ag/AgCl electrode as the reference, and a platinum mesh as the counter electrode. Figure 6.8 shows the evolution of the measured impedance spectrum over four cycles of gold electroplating. These measurements demonstrate the progressive decrease of electrode impedance with electroplating. At 1 KHz the impedance drops from 3.9 MΩ for the bare electrode to 438kΩ after electroplating. This impedance value has been found to be adequate for high-quality single unit isolation (Chapter 7). Further reduction of impedance is possible with continued electroplating cycles but comes at the expense of increased lateral dimensions of the deposits, and an associated reduction of selectivity [94–98].

To fit an equivalent circuit model we captured a high resolution impedance spectrum for an unplated (Figure 6.9, red curves) and plated (Figure 6.9, blue curves) electrode. We fit the data to an equivalent circuit model (Figure 6.9 insert) that uses a constant phase element (CPE) instead of a double-layer capacitor (Randles circuit, Chapter 2), as this better approximates the non-ideal capacitance behavior of the electrode-electrolyte interface. The impedance of the CPE element is described by

$$Z_{CPE} = \frac{1}{(j\omega)^\alpha Y_0}$$

where α is an empirically derived constant ($0 \leq \alpha \leq 1$). Note that $\alpha = 1$ describes an ideal capacitor with capacitance Y_0, while $\alpha = 0$ describes a pure resistor with resistance $1/Y_0$. The equivalent circuit includes a charge-transfer resistance (R_{ct}) in parallel with the CPE and a solution resistance (R_s) in series with the CPE, and the parasitic capacitances from the measurement setup (C_{cell}). The fitted parameters are

summarized in table 6.1.
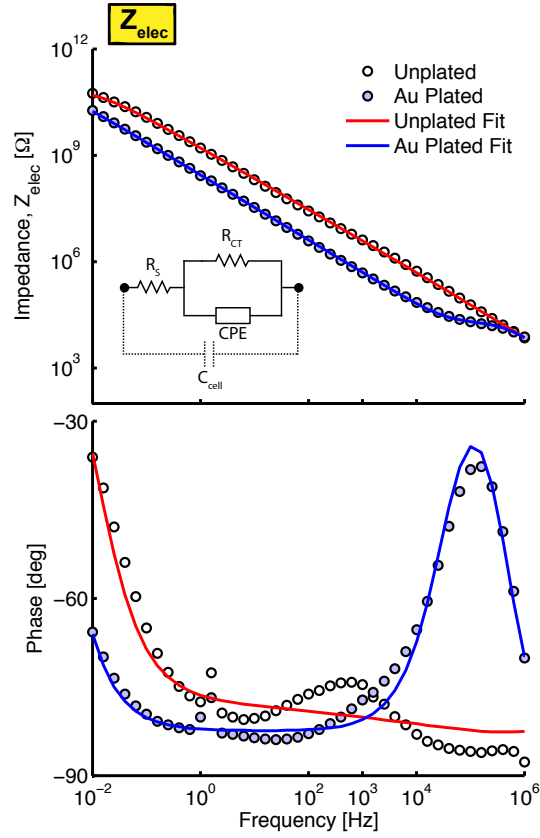
Figure 6.9: **Electrode-electrolyte equivalent circuit model for a recording electrode.** Electrochemical impedance spectrum for a recording electrode before (red) and after (blue) gold electroplating. Notice the order-of-magnitude decrease in impedance after electroplating with gold ($Z_{unplated} = 3.8M\Omega$ to $Z_{plated} = 0.5M\Omega$ evaluated at 1kHz). The data are fitted (red and blue curves) to a constant-phase element (CPE) electrode-electrolyte model (R_s = spreading resistance, R_{CT} = charge-transfer resistance, C_{cell} parasitic capacitances from the measurement setup). The measurement was performed inside a Faraday cage using a 3-cell/potentiostat configuration, with the nanoprobe electrode as the working electrode, a Ag/AgCl electrode as the reference, and a platinum mesh as the counter electrode.

Table 6.1: **EIS data fit parameters for unplated(plated) recording electrodes.** Goodness of fit: 7.6e-3(3.5e-3)

Parameter	Value	Units
R_s	20(15)	$k\Omega$
R_{CT}	55(89)	$G\Omega$
Y_0	60e-12 (750e-12)	$S * s^\alpha$
α	0.88 (0.91)	
C_{CELL}	11	pF

6.5   Crosstalk analysis

The data in Figure 6.9 provides a measurement-verified model for the nanoprobe recording electrodes that can be used to characterize cross-talk in the system, building on the analysis of Chapter 2 (Figures 2.5,6.10).
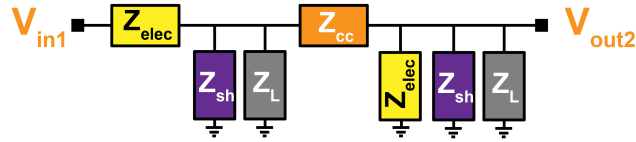
Figure 6.10: **Crosstalk model of two recording electrodes with adjacent interconnects.** This is derived from Figure 2.5.

In this context the cross-talk corresponds to:

$$Crosstalk[dB] = V_{in1}/V_{out2} = -20\log(2 + \frac{Z_{cc}}{Z_{elec}} + \frac{Z_{cc}}{Z_{sh}//Z_L})$$

where Z_{elec} is the electrode's impedance (provided by the equivalent circuit derived from the data in Figure 6.9), Z_{cc} is the coupling capacitance (1.16pF +/- 0.35pF), Z_{sh} is the shunting capacitance (2.5pF), and Z_L is the load impedance(12pF). Figure 6.11 shows the expected crosstalk for five different electrode impedances (Z_{elec}) including the unplated(red) and plated(blue) electrodes measured in Figure 6.9. This analysis

shows that crosstalk is negligible ($< 1\%$) for electrode impedances below 1MΩ at 1kHz.
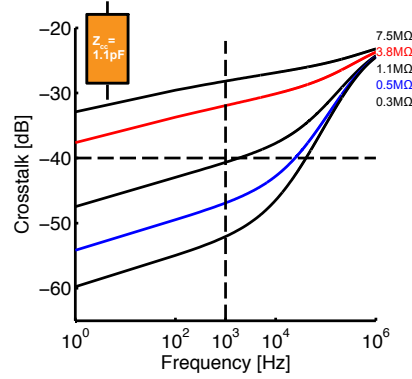
Figure 6.11: **Crosstalk as a function of frequency for five different electrode impedance profiles**. The colored curves correspond to the impedance spectra of the unplated (red) and plated (blue) electrode analyzed in Figure 6.9.

What is the maximum allowable coupling capacitance Z_{cc} for passive intercon-nects in order to keep crosstalk below 1%? Figure 6.12 demonstrates the constraints are severe for high impedance electrodes, with $Z_{cc} < 6$ pF even for low impedance electrode ($Z_{elec} = 0.3M\Omega$ at 1 KHz). This places constraints on the length, size, and distribution of passive interconnects which are satisfied in our system.
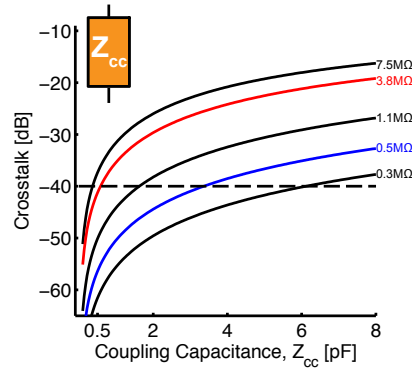

Figure 6.12: **Crosstalk as a function of coupling capacitance.** The colored curves correspond to the impedance spectra of the unplated (red) and plated (blue) electrode analyzed in Figure 6.9.

6.6 System Noise

To characterize the thermal noise contribution from the passive interconnects, we measured the resistance of a fully assembled nanoprobe array. The total resistance is $14 \pm 4 \ k\Omega$ with the contributions from the different components being: (1) nanoprobe: $13 \pm 2 \ k\Omega$, (2) flexPCB: $1 \pm 0.2 \ k\Omega$, (3) DAq-PCB: $< 10 \ \Omega$; (4) flip chip bonds: $< 1 \ \Omega$). The thermal noise from this resistance amounts to only $1.74 \ \mu V_{rms}$ (Bandwidth: 0.1-7.5kHz, room temperature). This is negligible and below the input referred noise of the amplifiers ($2.4 \ \mu V_{rms}$).

Figure 6.13 shows the power spectral density (PSD) across a grounded nanoprobe assembly with 256 channels. The black curve is the median PSD, while the red curve shows the PSD of an Intan chip with grounded inputs. These data show that the overall systems exhibits low noise ($4.8 \ \mu V$ RMS), which lies below the noise floor for brain signals ($> 10 \ \mu V$ RMS).
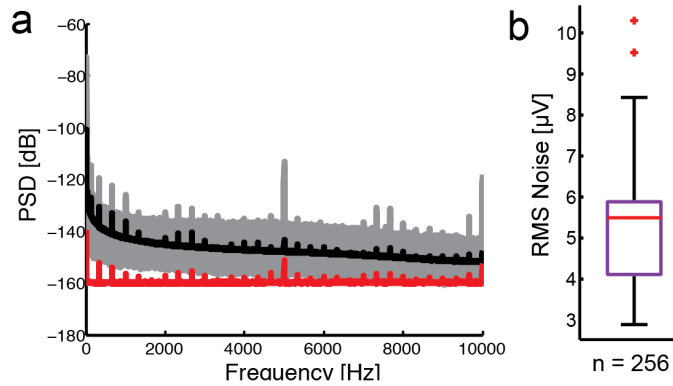
Figure 6.13: **Low-noise system performance of the nanoprobe array system.** (a) Power spectral density (PSD) of 256 recording electrodes (gray curves) after electroplating. The median PSD is shown in black, while the red curve shows the PSD of an Intan amplifier whose inputs are tied to ground. The harmonics are due to Intan's digital circuitry ($< 2\%$ of noise power). (b) Box plot showing the root mean square (RMS) noise distribution for 256 electroplated recording electrodes (all channels are included even if broken or shorted). The noise contributions from the amplifiers is not removed from the data ($2.4 \ \mu V$). Red line denotes the median ($V_{median} = 4.8 \ \mu V$). The edges of the box are the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually. (a,b) Bandwidth: 0.1Hz - 10Hz. Sampling frequency = 30kHz.

6.7 Conclusion

In this chapter we characterize the nanoprobe structural features that are important for reliable implantation and biocompatibility, and the electrical properties of the fully assembled system.

We demonstrated that stress compensation during fabrication has resulted in straight shanks, with minimal deflection consistent with the predictions of the simulations in Chapter 2. We have further confirmed that the nanoprobes are fully covered with biocompatible materials (Parylene on three sides, and silicon oxide (glass) on the fourth).

In addition, we have demonstrated that a high yield of proper connections across the fully assembled system can be routinely obtained ($< 5\%$ high impedance or shorted channels). Finally, using impedance spectroscopy we fitted an equivalent circuit model that enabled cross-talk analysis. Our analysis establishes that the fully assembled system achieves a low-noise and low-crosstalk performance.

Chapter 7

In Vivo Evaluation of Nanoprobe arrays

7.1 Introduction

The ultimate success for any neural probe development effort is determined by the quality of *in vivo* recordings that can be obtained. Reliable and efficient *in vivo* testing requires: (1) Targeting a brain circuit with clear electrophysiological signatures, and (2) using an experimental system that enables precise probe insertion and reliable brain recordings.

To satisfy these requirements, we decided to target the hippocampus, a brain circuit that plays a critical role in memory formation and has several experimental advantages. First, the hippocampus has a simple three-layer structure, with a single well-defined cell layer, and a laminar distribution of inputs on the dendritic fields. Second, this circuit exhibits characteristic local field potential (LFP) oscillations, with well-characterized laminar profiles. Thus the location and expected electrophysiological signatures of nanoprobe recordings can be reliably assessed. Third, the dense packing of cells in the pyramidal cell layer presents a challenge for isolating the activity of individual units. The Siapas lab has extensive experience in hippocampal recordings that can be leveraged to assess the quality of nanoprobe recordings. Furthermore, a validation of the nanoprobes in this circuit provides a powerful spring board for follow up scientific experiments.
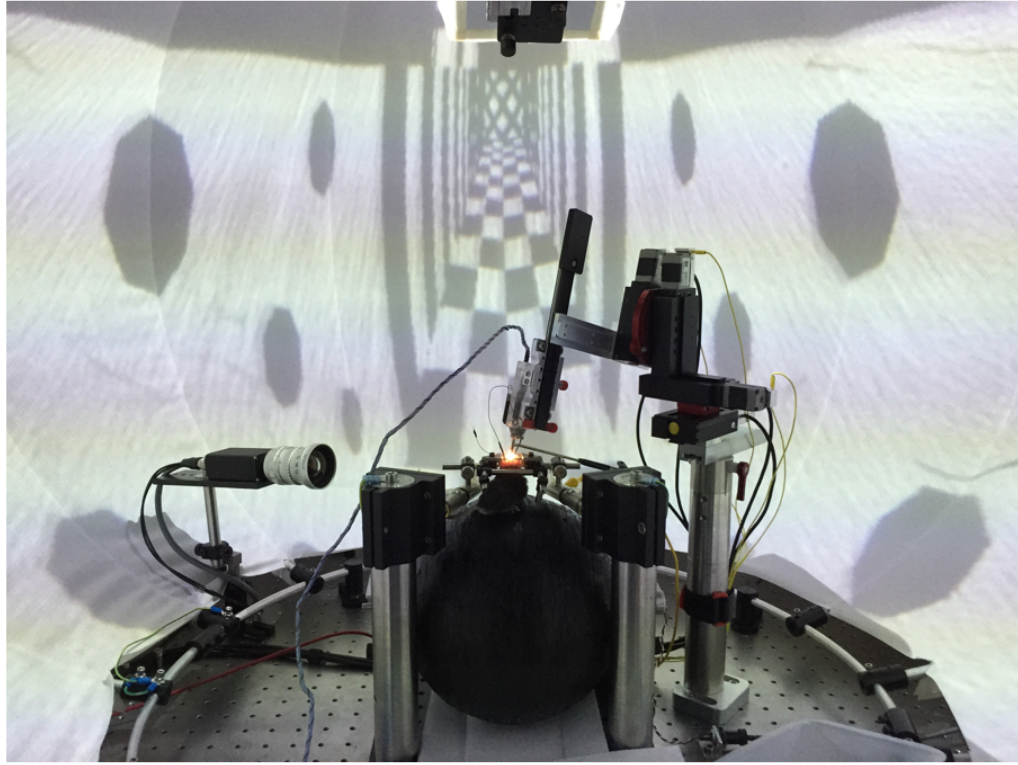
Figure 7.1: **Virtual reality system for head-fixed mice used to evaluate nanoprobe arrays.** A head-fixed mouse is free to walk on a spherical treadmill. Motion of the treadmill drives a virtual reality engine, that projects a virtual environment in a toroidal screen surrounding the mouse. A precision manipulator is used to insert a nanoprobe array in the brain (4-axis Luigs-Neuman manipulator)

7.2 *In vivo* experiments

To facilitate *in vivo* evaluation of the arrays, we built a virtual reality (VR) system for head-fixed mice that enables efficient testing of different nanoprobe designs [99]. In this system, a head-fixed mouse is free to walk on a spherical treadmill (Figure 7.1). The motion of the treadmill drives a virtual reality engine that projects a virtual environment on a toroidal screen surrounding the mouse, presenting a virtual environment that the mouse can navigate. This system has several advantages: (1) recordings can be obtained from awake, behaving mice, to engage the natural activation modes of the hippocampus, (2) experiments can be performed in a single-day (in contrast to

chronic evaluation that often spans weeks or even months), (3) precision manipulators enable inserting probe assemblies that may be too large and heavy to be carried by freely behaving mice, thus facilitating the testing and evolution of prototypes, and (4) insertion of the probes can be achieved under visual guidance, providing the ability to evaluate probe buckling, dimpling of the brain surface, or bleeding due to insertion. These undesirable side effects of probe insertion were confirmed to be minimized due to the small dimensions and straight profile of the nanoprobe shanks.

Theta oscillations are the hallmark of hippocampal activity during awake behavior and REM sleep. These are regular 4-10 Hz oscillations that clock the hippocampal circuit and strongly modulate activity in this network [100; 101]. In contrast, during quiet wakefulness and slow-wave sleep hippocampal activity is characterized by the presence of *ripples*, transient fast oscillations associated with population bursts that are believed to play a critical role in memory consolidation [102–105].

Figure 7.2 shows nanoprobe simultaneous recordings from areas CA3, CA1, and dentate gyrus (DG) of the hippocampus (Figure 7.2a,b). Nanoprobes provide high-quality recordings of ripples (Figure 7.2c) and theta oscillations (Figure 7.2d) with unprecedented resolution and scale. Figure 7.3 shows three ripples recorded on the same shank, illustrating the complex amplitude and phase profiles of individual ripple events. This suggests that dense 3-D recordings with nanoprobes are critical for proper characterization of the extent and timing of these events across the hippocampal circuit.

The modular architecture we have developed enables configuring arrays with different trade-offs between anatomical coverage and spatial resolution. Previous experiments with tetrodes have illustrated the need of closely spaced recording sites for proper single unit isolation. Each spike is picked up on multiple nearby sites with different amplitudes, enabling reliably isolating the activity of individual neurons through a process analogous to triangulation. We next examined whether high amplitude spikes can be obtained with nanoprobe arrays, and whether densely spaced recording sites would enable reliable spike sorting.

To this end we configured a dense 3-D array with 1024 recording sites packed

within 0.6 mm^3 of tissue (Figures 5.13, 5.14). Spikes obtained from this array are shown in Figure 7.4, with full broadband recordings across the array shown in figure 7.5. This is a landmark experiment as it represents the densest electrophysiological recordings obtained to date.

To examine whether the recorded spikes enable reliable isolation we have extracted spikes across the array (Figure 7.6a,b) and demonstrated that spiking persists even as we move the nanoprobes deeper into the tissue (Figure 7.6b). This shows that the damage due to nanoprobe insertion is minimal, as spiking persists as the nanoprobes penetrate deeper into the circuit. Furthermore, plotting the amplitudes of the recorded spikes reveals the existence of clear clusters demonstrating the ability to reliably isolate the activity of individual units (Figure 7.6 c,d).

7.3   Conclusion

Using recordings from behaving head-fixed mice navigating a virtual environment, we demonstrated that the nanoprobe arrays acquire high quality signals at different spatial resolutions, capable of revealing the complexity of LFP events such as ripples and theta oscillations with unprecedented resolution and scale. The culmination of these efforts was the implantation of a 3-D array with 1024 channels within 0.6mm^3 of tissue, which yielded the densest electrophysiological recordings to date.

Finally, we demonstrated that high amplitude spikes can be obtained in the cell layers of the hippocampal subfields. Analysis of spiking across contiguous channels of nanoprobes demonstrates that our recordings enable high-quality single unit isolation.
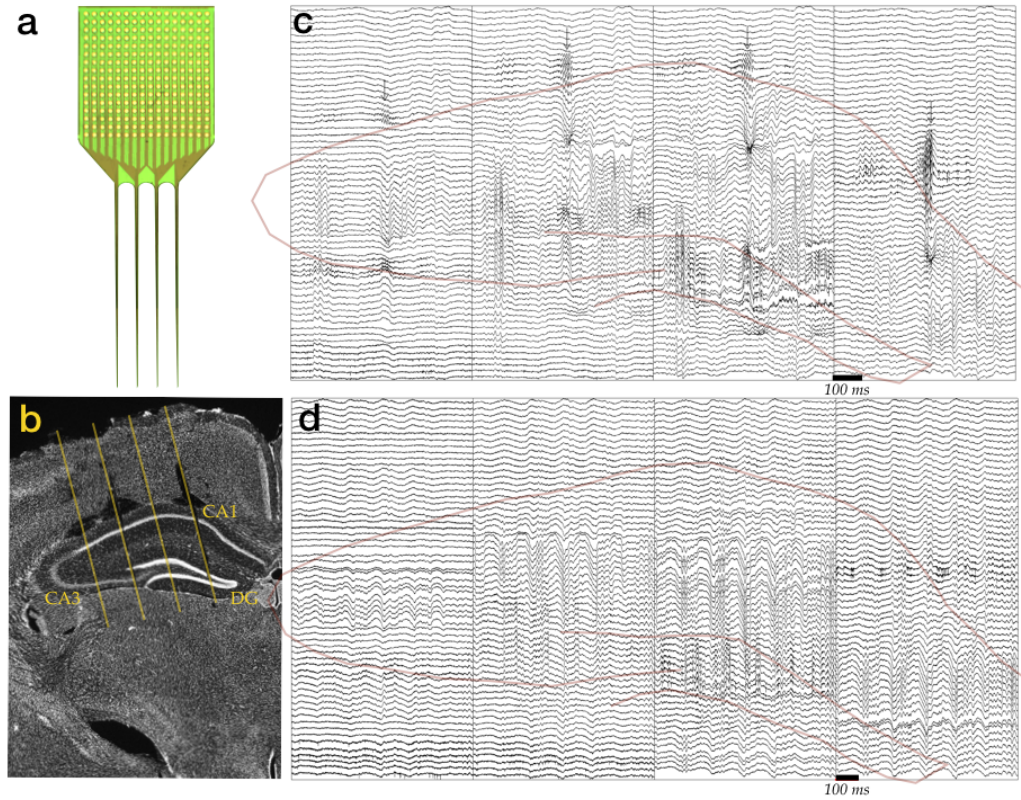
Figure 7.2: **Nanoprobe recordings of hippocampal ripples and theta oscillations.** (a) Picture of the nanoprobe used for the recordings (500 μm shank spacing, 64 recording sites per shank that are spaced at 30 μm). (b) Histological verification of the nanoprobe location in relation to the hippocampal circuit. Areas CA1, CA3, and dentate gyrus (DG) are marked in yellow. (c) Nanoprobe recordings during quiet wakefulness. The red arrows mark the onset of a hippocampal ripple. Notice the complex profile across the different hippocampal subfields, revealing complex modulation of the whole circuit with unprecedented resolution and scale. (d) Recordings from the same probe during active movement in the virtual environment. Notice the robust theta oscillations across the full array, with different amplitude and phase relationships throughout.
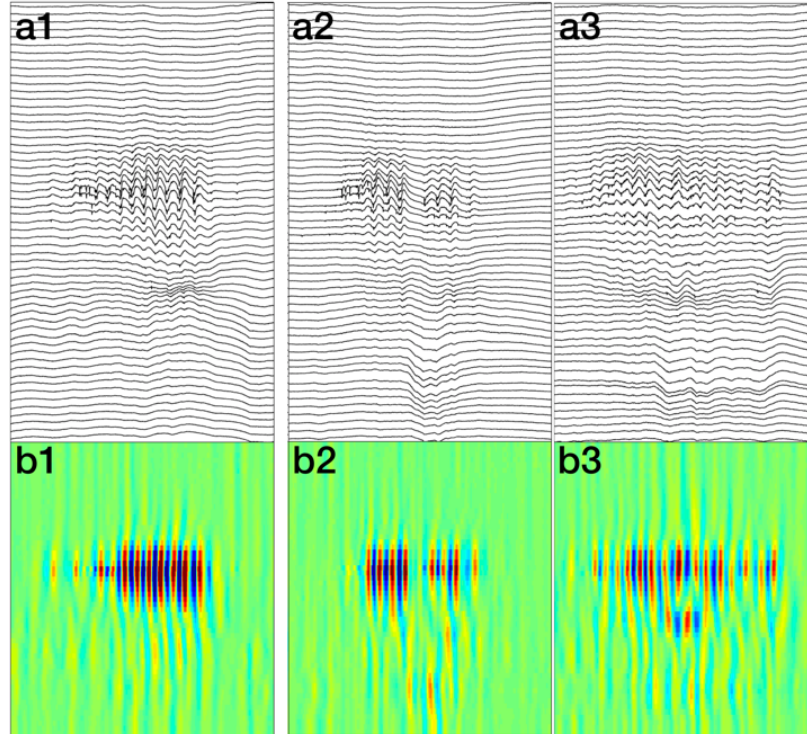
Figure 7.3: **Ripples have diverse depth profiles.** (a1-a3) Recordings of ripples from the same nanoprobe (64 sites spaced at 30 μm). The time window is 150 ms. (b1-b3) The traces from (a1-a3) filtered in the ripple frequency band (80-250 Hz) and displayed as images over the same 150 ms window. Notice the diverse amplitude and phase profiles with depth.
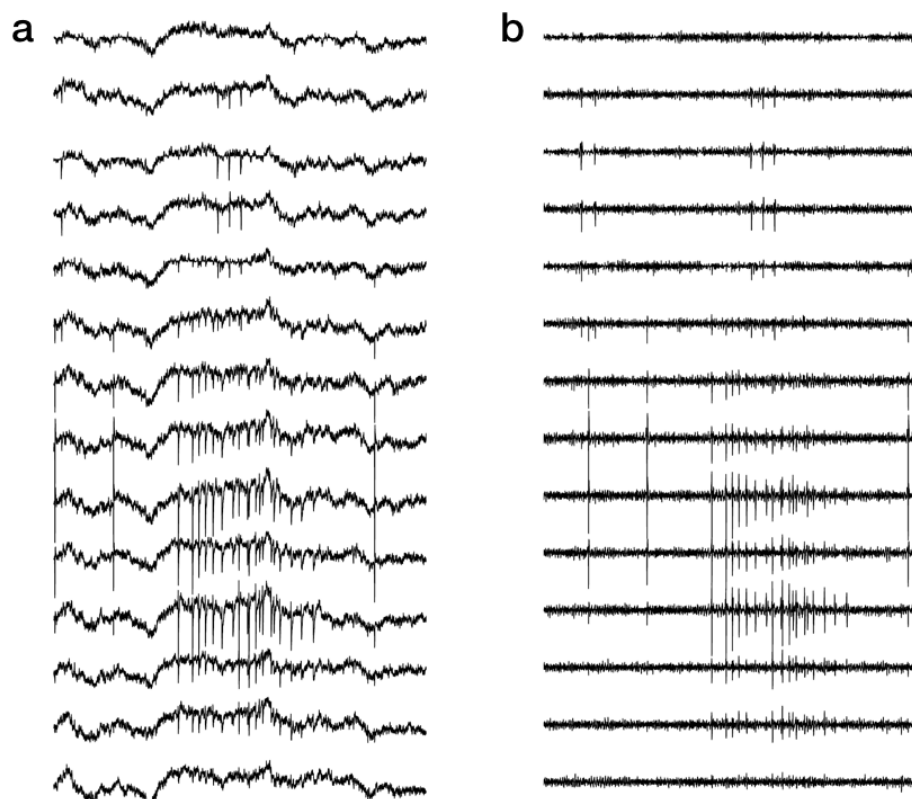
Figure 7.4: **Spikes recorded in consecutive sites of a nanoprobe.** (a) Broadband signal (0.1Hz - 10KHz). (b) The recordings in (a) filtered in the spike frequency range (400Hz-10KHz).
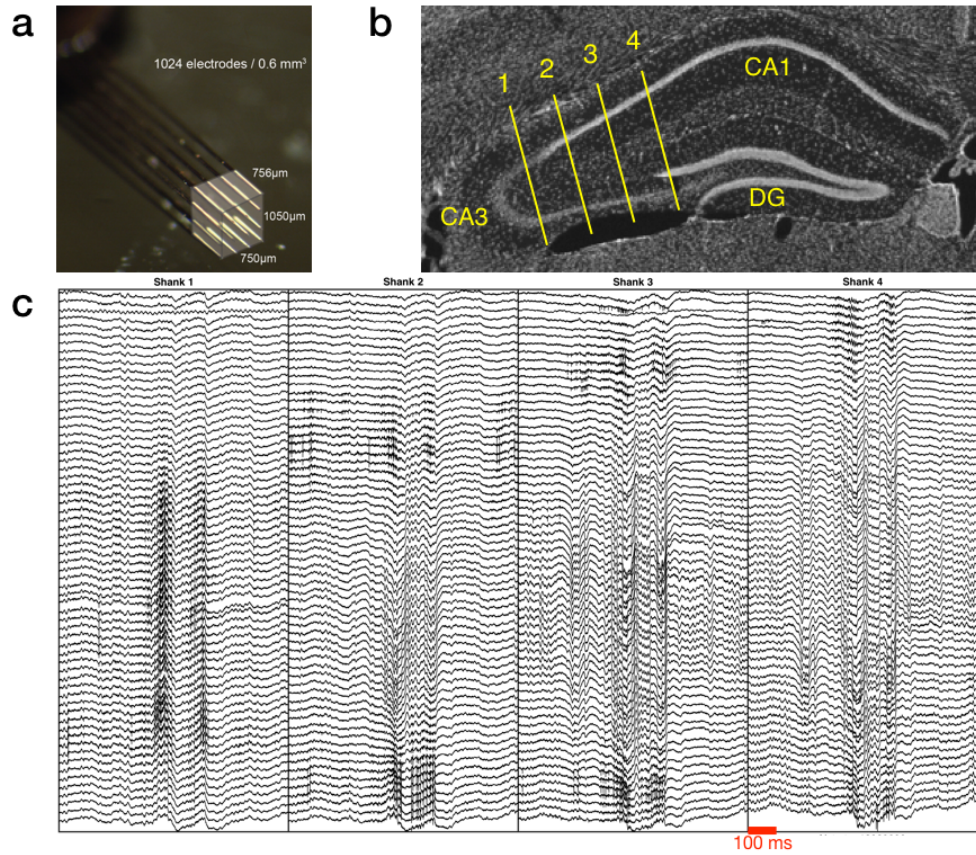
Figure 7.5: **Hippocampal nanoprobe recordings from a 3-D array with 1024 sites.** (a) Nanoprobe 3-D array with 1024 sites in 0.6 mm^3 of tissue (Figures 5.13, 5.14). (b) Histological verification of the location of the recording sites. (c) Nanoprobe recordings of a hippocampal ripple with the array shown in (a). This represents the densest electrophysiological recordings obtained to date.
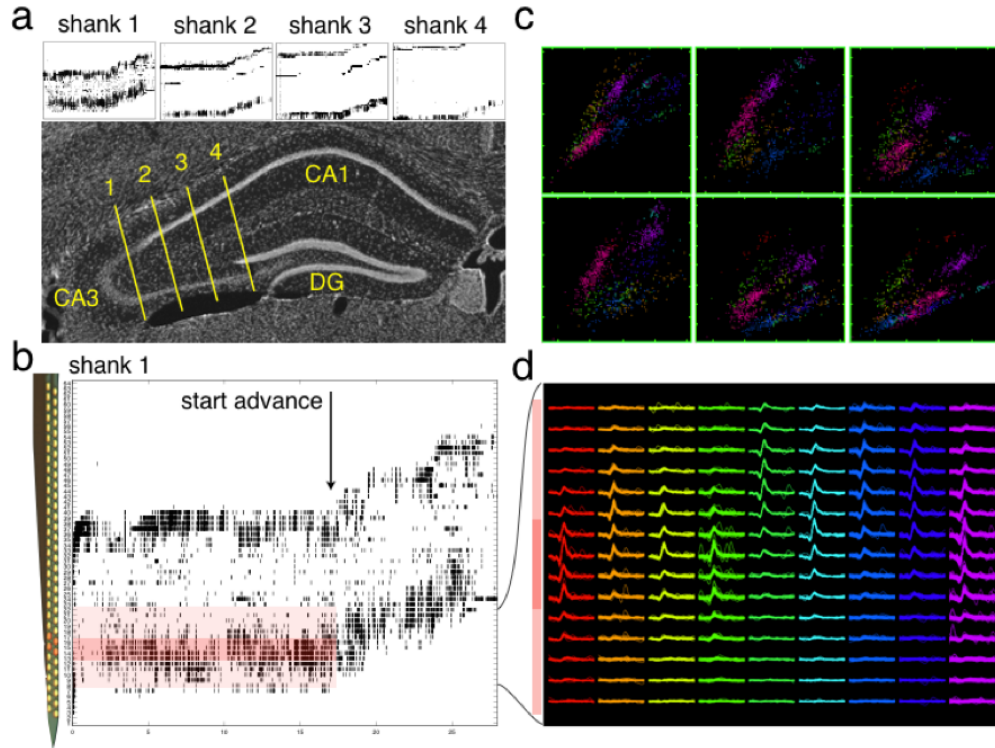
Figure 7.6: **Single unit isolation with nanoprobe arrays.** (a) (bottom) Hoechst stained coronal section indicating shank tracks identified for implantation of the 3-D array shown in Figures 5.13, 5.14. Yellow lines show the recording extent of each shank (756 μm, 64 sites). (top) Spike rasters extracted from each site on a shank are ordered from superficial to deep. Notice that only sites close to cell layers detect spikes and that spiking activity starts shifting up as the nanoprobe is gradually advanced. (b) Electrode site distribution on shank 1 (left) and detailed view of the corresponding spike rasters (right). Notice that as the shank is advanced spiking activity is spared and translated to more superficial sites, owing to the low invasiveness of the probe. (c) Spike amplitude scatter plots of a virtual tetrode formed by 4 neighboring sites (dark red circles/bands in b) before the probe is advanced. Units are automatically isolated and colored accordingly. (d) Multiple waveforms from each isolated unit demonstrate the quality for the clustering. Notice that waveforms extend anatomically beyond the sites used for isolation (light and dark red bands in b), so clustering methods exploiting all sites will perform even better.

Chapter 8

Conclusion

This thesis describes the development of a modular and scalable system that enables 3-D recordings of brain activity from > 1000 electrodes. The front-end of the system is composed of nanofabricated probes with ultrathin shanks that are engineered to minimize tissue damage. The probes are connected via flexible cables to custom PCBs that multiplex the electrophysiological signals. This system architecture decouples the front-end both mechanically and thermally from the PCB which carries all active electronics for signal conditioning and multiplexing. This modular design also enables evolving and debugging the different components efficiently. This system was validated with *in vivo* with hippocampal recordings from head-fixed mice. The culmination of these efforts was a 3-D array with 1024 sites packed within 0.6 mm^3 of tissue that yielded the densest electrophysiological recordings to date.

8.1 Comparison to previous work

The development of multi-electrode recordings has transformed our ability to monitor populations of neurons in behaving animals. An important proof-of-concept, and a benchmark of what can be achieved, is provided by the development and maturation of chronic tetrode recordings. It was previously demonstrated that it is possible to achieve high quality recordings for months, with stability of recordings for weeks [106]. This level of stability is rather remarkable for recordings from freely behaving animals. Several factors contribute to the success of these chronic tetrode recordings,

among which is the small diameter of these probes ($< 50 \ \mu m$) and the close spacing of the recording sites, which enhances the ability to triangulate signals from nearby neurons. However, one of the key disadvantages of tetrode recordings is their limited spatial coverage: one can only record from the very tip of the tetrode.

Integrated circuit technology has provided a successful path for microfabrication of electrode arrays and has enabled both increasing the number of recording sites per shank, and precise dimensional control over the spacing and configuration of recording sites [47]. Since the original pioneering work by Wise and colleagues there has been a sustained effort to further scale up the number of recording sites.

Among challenges to achieving a further increased scale-up of recording sites is managing the increased wiring complexity on the probe itself and, subsequently, creating a compact interface to the data acquisition system. Most probes typically export the recording sites to a printed circuit board (PCB) that houses chips that condition and buffer the electrophysiological signals. Current setups for freely behaving animals use ≈ 100 channels fed through individual fine wires. This is at the practical limit due to the physical constraints on the maximal number of connectors, the size of interface boards, and the weight and reliability of the fine wires.

In principle, increasing the density and number of recording sites can be achieved by co-integrating active signal conditioning/multiplexing electronics directly within the probe itself. Today, active probes are being built using standard CMOS processes that can ultimately permit their large-scale production. Early pioneering efforts were limited to proof-of-concept prototypes [107; 108]. Recent architectures for active probes employ active buffering at each recording site and require selection of only a small subset of the available sites for simultaneous recording. This approach has been introduced as part of the European Integrated Research Project NeuroProbes, and has been termed electronic depth control (EDC) [109; 110]. The highest number of simultaneously recorded channels reported to date is only 52 out of a possible 455 electrodes [111]. This approach severely restricts the number of recording sites available simultaneously and, in turn, greatly limits the scale and complexity of neural circuitry that can be explored. There are significant technological drawbacks

that must also be addressed with the active-probe paradigm: achieving co-integration of CMOS with neural probes is costly; and the large active probe cross-sections (100 μm wide, 50 μm thick) will likely compromise their chronic performance. For example, arrays of such probes spaced at 300 μm would displace $\approx 8\%$ of the tissue, compared to $< 1\%$ of tissue for probes described in this thesis. Finally, the proximity of electronics to the brain poses significant challenges for minimizing heating of the brain tissue, as discussed in Chapter 2.

Recently developed multichannel chips for electrophysiology (Intan technologies) offer a means for increasing the number of channels, while multiplexing to minimize the number of output wires. The highest density commercially available chips enable multiplexing 64 channels to one output line (Intan RHD 2164). We have employed these chips to develop custom PCBs to assemble 256-channel headstages. Two other recent efforts have followed a similar approach [112; 113]. These efforts chose to bond probes directly to large PCBs making it impossible to pack probes in dense 3-D configurations, and to create compact implantable interfaces for chronic implantations in freely behaving animals. In contrast, our modular system has enabled the densest recordings to date (Figure 7.5), and enables coupling to compact, light-weight PCBs that can be chronically implanted. Furthermore, it enables mechanically and thermally decoupling the probes that are inserted in the brain, to optimize biocompatibility and long-term performance.

8.2   Future Directions

A key next step is to design and build chronically implantable interfaces that can be used with freely behaving animals. This critical as many brain functions can only be probed in freely moving animals performing natural behaviors. Furthermore, chronic recordings offer the only viable path for studying the evolution of neural patterns throughout the course of learning, and across behavioral and brain states such as the sleep-wake cycle. The functional role of sleep remains one of the persistent mysteries in neuroscience and long-term chronic recordings is the only path to elucidating this
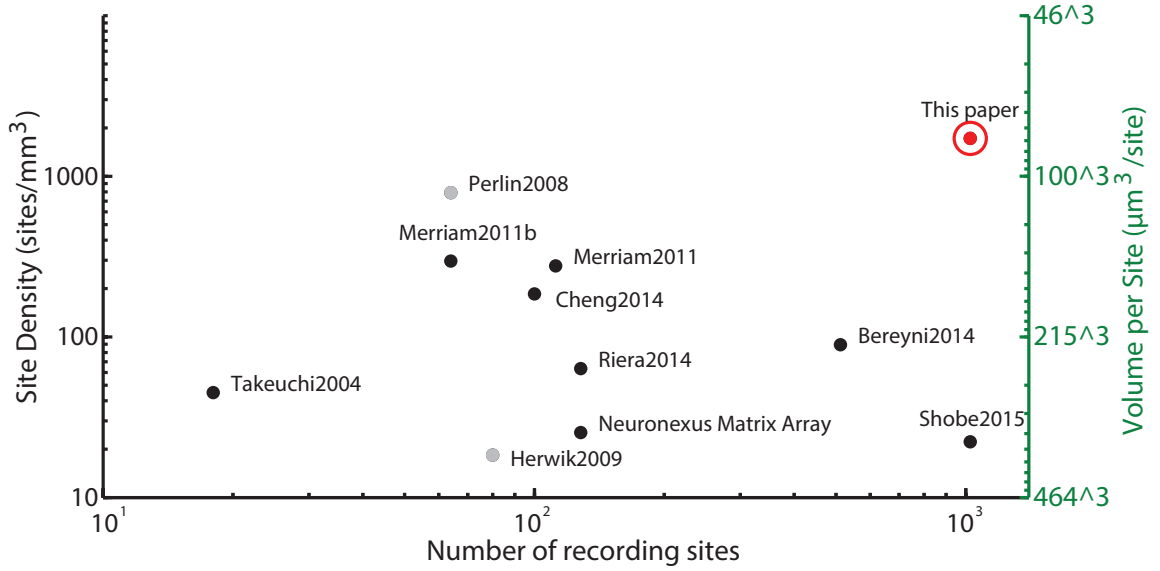
Figure 8.1: **Comparison of the 1024 channel 3-D neural probe recording system to other state of the art systems.** The dense 1024 channel 3-D array (Figures 5.13, 5.14, 7.5) provided the densest electrophysiological recordings to date. The references are: Takeuchi2004 [114], Merriam2011b [115], Merriam2011 [116], Perlin2008 [117], Cheng2014 [118], Riera2014 [119], Herwik2009 [120], Berenyi2014 [121], Shobe2015 [113].

critical open question.

The study of long-term performance of neural implants opens new challenges. It requires a computational framework for quantifying the quality and stability of long-term recordings and characterizing the tissue response to the implant, including gliosis and long-term degradation of the probe insulating layers. Our system is conceived and built to perform chronically, by encapsulating the probes with a biocompatibility layer (Parylene), and enabling compact interfaces that can be chronically implanted, in contrast to other recent efforts [112; 113].

Another avenue for further technical development is the further scale up of the number of electrodes and the integration of recording and electrical stimulation. This requires the development of dense probes that employ multiple layers of nanoscale interconnects, and the development of custom electronics that can multiplex higher number of channels and enable patterned stimulation as well as recordings.

The design philosophy and approach employed in this thesis provides a powerful

platform technology for the development of multifunctional probes that can combine not only electrical recording and stimulation, but also additional functionalities, such as electrochemical measurements and optical detectors and emitters.

Bibliography

[1]   A. P. Alivisatos et al. "Nanotools for neuroscience and brain activity mapping." *Acs Nano* 7:3, 2013, pp. 1850–1866.

[2]   G. Buzsáki et al. "Tools for Probing Local Circuits: High-Density Silicon Probes Combined with Optogenetics." *Neuron* 86:1, 2015, pp. 92–105.

[3]   D Kipke et al. "Advanced neurotechnologies for chronic neural interfaces: new horizons and clinical . . . " *Journal of Neuroscience*, 2008.

[4]   M. M. Elwassif et al. "Bio-heat transfer model of deep brain stimulation-induced temperature changes." *Journal of neural engineering* 3:4, 2006, pp. 306–315.

[5]   J. F. Hetke et al. "Silicon ribbon cables for chronically implantable micro-electrode arrays." *Biomedical Engineering, IEEE Transactions on* 41:4, 1994, pp. 314–321.

[6]   S Kisban et al. "Microprobe Array with Low Impedance Electrodes and Highly Flexible Polyimide Cables for Acute Neural Recording." *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, 2007, pp. 175–178.

[7]   J. A. Freeman and C Nicholson. "Experimental optimization of current source-density technique for anuran cerebellum." *Journal of Neurophysiology* 38:2, 1975, pp. 369–382.

[8]   A. Tooker et al. "Towards a large-scale recording system: Demonstration of polymer-based penetrating array for chronic neural recording." In: *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, pp. 6830–6833.

[9]   B.-y. Lee et al. "Flexible neural probe integrated microchannel and optical fiber for multifle stimulation." In: *Solid-State Sensors, Actuators and Microsystems*. IEEE, 2013, pp. 321–324.

[10] A. A. Fomani and R. R. Mansour. "Flexible neural microelectrode arrays reinforced with embedded metallic micro-needles." *Sensors, 2010 IEEE*, 2010, pp. 1601–1604.

[11] H.-Y. Lai et al. "Design, simulation and experimental validation of a novel flexible neural probe for deep brain stimulation and multichannel recording." *Journal of neural engineering* 9:3, 2012, p. 036001.

[12] T. D. Y. Kozai and D. R. Kipke. "Insertion shuttle with carboxyl terminated self-assembled monolayer coatings for implanting flexible polymer neural probes in the brain." *Journal of Neuroscience Methods* 184:2, 2009, pp. 199–205.

[13] Y Kato et al. "Preliminary study of multichannel flexible neural probes coated with hybrid biodegradable polymer." *2006 International Conference of the IEEE Engineering in Medicine and Biology Society, year = 2008, month = jan.*

[14] S Takeuchi, T Suzuki, and K Mabuchi. "3D flexible multichannel neural probe array." *Journal of Micromechanics and Microengineering*, 2004.

[15] "Partially flexible MEMS neural probe composed of polyimide and sucrose gel for reducing brain damage during and after implantation." *Journal of Micromechanics and Microengineering* 24:2, 2014, pp. 025010–11.

[16] Z. Xiang et al. "Ultra-thin flexible polyimide neural probe embedded in a dissolvable maltose-coated microneedle." *Journal of Micromechanics and Microengineering* 24:6, 2014, p. 065015.

[17] J Seymour and D Kipke. "Fabrication of Polymer Neural Probes with Subcellular Features for Reduced Tissue Encapsulation." *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*, 2006, pp. 4606–4609.

[18] Y.-Y. Chen et al. "Design and fabrication of a polyimide-based microelectrode array: Application in neural recording and repeatable electrolytic lesion in rat brain." *Journal of Neuroscience Methods* 182:1, 2009, pp. 6–16.

[19] K Najafi, J Ji, and K. D. Wise. "Scaling limitations of silicon multichannel recording probes." *Biomedical Engineering, IEEE Transactions on* 37:1, 1990, pp. 1–11.

[20] W Jensen, K Yoshida, and U. G. Hofmann. "In-vivo implant mechanics of flexible, silicon-based ACREO microelectrode arrays in rat cerebral cortex." *Biomedical Engineering, IEEE Transactions on* 53:5, 2006, pp. 934–940.

[21] A. A. Sharp et al. "In vivo penetration mechanics and mechanical properties of mouse brain tissue at micrometer scales." *IEEE transactions on bio-medical engineering* 56:1, 2009, pp. 45–53.

[22] S. J. Oh et al. "A High-Yield Fabrication Process for Silicon Neural Probes." *Biomedical Engineering, IEEE Transactions on* 53:2, 2006, pp. 351–354.

[23] M.-Y. Cheng et al. "A low-profile three-dimensional neural probe array using a silicon lead transfer structure." *Journal of Micromechanics and Microengineering* 23:9, 2013, p. 095013.

[24] A Tarraf et al. "Stress investigation of PECVD dielectric layers for advanced optical MEMS." *Journal of Micromechanics and Microengineering* 14:3, 2004, pp. 317–323.

[25] S. Cho, K Najafi, and K. D. Wise. "Internal stress compensation and scaling in ultrasensitive silicon pressure sensors." *Electron Devices, IEEE Transactions on* 39:4, 1992, pp. 836–842.

[26] C. Rossant et al. "Spike sorting for large, dense electrode arrays." *Nature Neuroscience*, 2016, pp. 1–10.

[27] M Hughes and K Bustamante. "Effects of electrode size on the performance of neural recording microelectrodes." *. . . in Medicine and . . .* 2002.

[28] K. Najafi, J. Ji, and K. D. Wise. "Scaling limitations of silicon multichannel recording probes." *Biomedical Engineering, IEEE Transactions on* 37:1, 1990, pp. 1–11.

[29] "An Implantable 455-Active-Electrode 52-Channel CMOS Neural Probe." *IEEE Journal of Solid-State Circuits* 49:1, 2013, pp. 248–261.

[30] J. Du, M. L. Roukes, and S. C. Masmanidis. "Dual-side and three-dimensional microelectrode arrays fabricated from ultra-thin silicon substrates." *Journal of Micromechanics and Microengineering* 19:7, 2009, p. 075008.

[31] J. Du et al. "Multiplexed, high density electrophysiology with nanofabricated neural probes." *PLoS One* 6:10, 2011, e26204.

[32] N. D. Arora et al. "Modeling and extraction of interconnect capacitances for multilayer VLSI circuits." *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 15:1, 1996, pp. 58–67.

[33] G. Shomalnasab, H. M. Heys, and L. Zhang. "Analytic modeling of interconnect capacitance in submicron and nanometer technologies." *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2013, pp. 2553–2556.

[34] D Cantrell, S Inayat, and A Taflove. "Incorporation of the electrode–electrolyte interface into finite-element models of metal microelectrodes." *Journal of Neural . . .* 2008.

[35] J. B. Bates and Y. T. Chu. "Electrode-electrolyte interface impedance: experiments and model." *Annals of Biomedical Engineering* 20:3, 1992, pp. 349–362.

[36] D. A. Robinson. "The electrical properties of metal microelectrodes." In: *Proceedings of the IEEE.* 1968, pp. 1065–1071.

[37] V Polikov, P Tresco, and W Reichert. "Response of brain tissue to chronically implanted neural electrodes." *Journal of Neuroscience Methods*, 2005.

[38] B. D. Winslow et al. "A comparison of the tissue response to chronically implanted Parylene-C-coated and uncoated planar silicon microelectrode arrays in rat cortex." *Biomaterials* 31:35, 2010, pp. 9163–9172.

[39] M Welkenhuysen et al. "Effect of Insertion Speed on Tissue Response and Insertion Mechanics of a Chronically Implanted Silicon-Based Neural Probe." *IEEE transactions on bio-medical engineering* 58:11, 2011, pp. 3250–3259.

[40] C. Marin. "Biocompatibility of intracortical microelectrodes: current status and future prospects." *Frontier in Neuroengineering* 3, 2010.

[41] D. J. Edell et al. "Factors influencing the biocompatibility of insertable silicon microshafts in cerebral cortex." *Biomedical Engineering, IEEE Transactions on* 39:6, 1992, pp. 635–643.

[42] G Altankov, V Thom, and T Groth. "Modulating the biocompatibility of polymer surfaces with poly (ethylene glycol): effect of fibronectin." *... Research Part A*, 2000.

[43] D. F. Williams. "On the mechanisms of biocompatibility." *Biomaterials* 29:20, 2008, pp. 2941–2953.

[44] G. E. Loeb et al. "Parylene as a chronically stable, reproducible microelectrode insulator." *Biomedical Engineering, IEEE Transactions on* 24:2, 1977, pp. 121–128.

[45] S. Kim et al. "Thermal Impact of an Active 3-D Microelectrode Array Implanted in the Brain." *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* 15:4, 2007, pp. 493–501.

[46] W. M. Reichert and P. D. Wolf. "Thermal Considerations for the Design of an Implanted Cortical Brain–Machine Interface (BMI)," 2008.

[47] K. D. Wise, J. B. Angell, and A. Starr. "An Integrated-Circuit Approach to Extracellular Microelectrodes." *Biomedical Engineering, IEEE Transactions on* BME-17:3, 1970, pp. 238–247.

[48] K. D. Wise and K Najafi. "Microfabrication techniques for integrated sensors and microsystems." *Science (New York, NY)* 254:5036, 1991, pp. 1335–1342.

[49] K Najafi, K. D. Wise, and T Mochizuki. "A high-yield IC-compatible multichannel recording array." *Electron Devices, IEEE Transactions on* 32:7, 1985, pp. 1206–1211.

[50] K. Cheung et al. "Implantable multichannel electrode array based on SOI technology." *Journal of Microelectromechanical Systems* 12:2, 2003, pp. 179–184.

[51] P Norlin, M Kindlundh, and A Mouroux. "A 32-site neural recording probe fabricated by DRIE of SOI substrates." *Journal of ...* 2002.

[52] M HajjHassan, V Chodavarapu, and S Musallam. "NeuroMEMS: neural probe microtechnologies." *Sensors*, 2008.

[53] M Kindlundh, P Norlin, and U. Hofmann. "A neural probe process enabling variable electrode configurations." *Sensors and Actuators B: Chemical* 102:1, 2004, pp. 51–58.

[54] K Wise, A Sodagar, and Y Yao. "Microelectrodes, microelectronics, and implantable neural microsystems." *Proceedings of the ...* 2008.

[55] P Ruther et al. "The neuroprobes project-multifunctional probe arrays for neural recording and stimulation." *Biomed ...* 2008.

[56] M Merriam et al. "A Three-Dimensional 160-Site Microelectrode Array for Cochlear Nucleus Mapping." *Biomedical Engineering, IEEE Transactions on* PP:99, 2010, pp. 1–1.

[57] D. Kipke et al. "Advanced Neurotechnologies for Chronic Neural Interfaces: New Horizons and Clinical Opportunities." *The Journal of . . .* 2008.

[58] G Buzsaki. "Large-scale recording of neuronal ensembles." *Nature Neuroscience* 7:5, 2004, pp. 446–451.

[59] J. Scholvin et al. "Close-packed silicon microelectrodes for scalable spatially oversampled neural recording," 2015.

[60] K Cheung et al. "A new neural probe using SOI wafers with topological interlocking mechanisms." *Microtechnologies in Medicine and Biology, 1st Annual International, Conference On. 2000*, 2002, pp. 507–511.

[61] D. Rotermund et al. "Open Hardware: Towards a Fully-Wireless Sub-Cranial Neuro-Implant for Measuring Electrocorticography Signals." *bioRxiv*, 2016, p. 036855.

[62] J. P. Seymour et al. "Novel multi-sided, microelectrode arrays for implantable neural applications." *Biomedical microdevices* 13:3, 2011, pp. 441–451.

[63] P. C. Andricacos et al. "Damascene copper electroplating for chip interconnections." *IBM Journal of Research and Development* 42:5, 1998, pp. 567–574.

[64] I. F. Scheiber, J. F. B. Mercer, and R. Dringen. "Metabolism and functions of copper in brain." *Progress in neurobiology* 116, 2014, pp. 33–57.

[65] M. C. Cortizo and M. F. L. de Mele. "Cytotoxicity of copper ions released from metal." *Biological trace element research* 102:1-3, 2004, pp. 129–141.

[66] D. S. del Amo et al. "Biocompatible Copper(I) Catalysts for in Vivo Imaging of Glycans." *Journal of the . . .* 132:47, 2010, pp. 16893–16899.

[67] K Seidl et al. "CMOS-Based High-Density Silicon Microprobe Arrays for Electronic Depth Control in Intracortical Neural Recording." *Microelectromechanical Systems, Journal of* 20:6, 2011, pp. 1439–1448.

[68] G. Guitchounts et al. "A carbon-fiber electrode array for long-term neural recording." *Journal of neural engineering* 10:4, 2013, p. 046016.

[69] J. P. Kinney et al. "A direct-to-drive neural data acquisition system." *Frontiers in Neural Circuits* 9, 2015, pp. 1–8.

[70] J. J. Wheeler et al. "An implantable 64-channel neural interface with reconfigurable recording and stimulation." *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 2015, pp. 7837–7840.

[71] D. A. Schwarz et al. "Chronic, wireless recordings of large-scale brain activity in freely moving rhesus monkeys." *Nature methods* 11:6, 2014, pp. 670–676.

[72] J. Du et al. "Multiplexed, High Density Electrophysiology with Nanofabricated Neural Probes." *PLoS One* 6:10, 2011, e26204.

[73] P Ledochowitsch et al. "Nanoflex for neural nanoprobes." In: *2013 Transducers & Eurosensors XXVII: The 17th International Conference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS & EUROSENSORS XXVII)*. IEEE, pp. 1278–1281.

[74] P. Ledochowitsch et al. "Fabrication and testing of a large area, high density, parylene MEMS." *2011 IEEE 24th International Conference on Micro Electro Mechanical Systems*, 2011, pp. 1031–1034.

[75] R. Huang et al. "Parylene coated silicon probes for neural prosthesis." *Nano/Micro Engineered and Molecular Systems, 2008. NEMS 2008. 3rd IEEE International Conference on*, 2008, pp. 947–950.

[76] C Pang, Y Tai, and J Burdick. "Electrolysis-based parylene balloon actuators for movable neural probes." *Nano/Micro Engineered . . .* 2007.

[77] S Metz et al. "Flexible polyimide probes with microelectrodes and embedded microfluidic channels for simultaneous drug delivery and multi-channel monitoring of bioelectric activity." *Biosensors and Bioelectronics*, 2004.

[78] S. Takeuchi et al. "3D flexible multichannel neural probe array." *Journal of Micromechanics and Microengineering* 14:1, 2004, pp. 104–107.

[79] S. Myllymaa et al. "Fabrication and testing of polyimide-based microelectrode arrays for cortical mapping of evoked potentials." *Biosensors and Bioelectronics* 24:10, 2009, pp. 3067–3072.

[80] J. F. Hetke et al. "Silicon ribbon cables for chronically implantable microelectrode arrays." *Biomedical Engineering, IEEE Transactions on* 41:4, 1994, pp. 314–321.

[81] R. S. Pai and K. M. Walsh. "The viability of anisotropic conductive film as a flip chip interconnect technology for MEMS devices." *Journal of Micromechanics and Microengineering* 15:6, 2005, pp. 1131–1139.

[82] M. A. Uddin et al. "Adhesion strength and contact resistance of flip chip on flex packages—effect of curing degree of anisotropic conductive film." *Microelectronics Reliability*, 2004.

[83] T.-O. Chong et al. "Low cost flip chip package design concepts for high density I/O." In: *Electronic Components and Technology Conference, 2001. Proceedings., 51st.* 2001, pp. 1140–1143.

[84] M. J. Yim et al. "Highly reliable flip-chip-on-flex package using multilayered anisotropic conductive film." *Journal of Electronic Materials* 33:1, 2004, pp. 76–82.

[85] K. Y. Chen et al. "Ultra-thin electronic device package." *IEEE Transactions on Advanced Packaging* 23:1, 2000, pp. 22–26.

[86] H. Cao et al. "An Integrated LED Optrode for Optogenetic Stimulation and Electrical Recording." *IEEE transactions on bio-medical engineering* 60:1, pp. 225–229.

[87] F. Wu et al. "Monolithically Integrated μLEDs on Silicon Neural Probes for High-Resolution Optogenetic Studies in Behaving Animals." *Neuron*, 2015, pp. 1–14.

[88] D Moser et al. "Fabrication of microfluidic neural probes with in-channel electrodes." *Sensors, 2012 IEEE*, 2012, pp. 1–4.

[89] G. Petit-Pierre, A. Bertsch, and P. Renaud. "Neural probe combining microelectrodes and a droplet-based microdialysis collection system for high temporal resolution sampling." *Lab Chip*, 2016, pp. 1–9.

[90] H. J. Lee et al. "A multichannel neural probe with embedded microfluidic channels for simultaneous in vivo neural recording and drug delivery." *Lab Chip* 15, 2015, pp. 1590–1597.

[91] T Suzuki, D Ziegler, and K Mabuchi. "Flexible neural probes with micro-fluidic channels for stable interface with the nervous system." *... in Medicine and ...* 2005.

[92] F. Wu et al. "Implantable neural probes for chronic electrical recording and optical stimulation." *Electron Devices Meeting (IEDM), 2013 IEEE International*, 2013, pp. 8–8.6.4.

[93] J. G. McCall et al. "Fabrication and application of flexible, multimodal light-emitting devices for wireless optogenetics." *Nature Protocols* 8:12, 2013, pp. 2413–2428.

[94] S. A. Desai et al. "Improving impedance of implantable microwire multi-electrode arrays by ultrasonic electroplating of durable platinum black." *Frontiers in Neuroengineering* 3, 2010, p. 5.

[95] J. E. Ferguson, C Boldt, and A. D. Redish. "Creating low-impedance tetrodes by electroplating with additives." *Sensors & Actuators: A. Physical* 156:2, 2009, pp. 388–393.

[96] S Venkatraman et al. "In Vitro and In Vivo Evaluation of PEDOT Microelectrodes for Neural Stimulation and Recording." *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* 19:3, 2011, pp. 307–316.

[97] T. Kozai et al. "Chronic in vivo evaluation of PEDOT/CNT for stable neural recordings." *Biomedical Engineering, IEEE Transactions on*, 2015, pp. 1–1.

[98] K. Ludwig, J. Uram, and J Yang. "Chronic neural recordings using silicon microelectrode arrays electrochemically deposited with a poly(3,4-ethylenedioxythiophene) (PEDOT) film." *Journal of Neural ...* 2006.

[99] C. D. Harvey et al. "Intracellular dynamics of hippocampal place cells during virtual navigation." *Nature* 461:7266, 2009, pp. 941–946.

[100] E. V. Lubenov and A. G. Siapas. "Hippocampal theta oscillations are travelling waves." *Nature* 459:7246, 2009, pp. 534–539.

[101] G. Buzsáki. "Theta oscillations in the hippocampus." *Neuron* 33:3, 2002, pp. 325–340.

[102] M. A. Wilson and B. L. McNaughton. "Reactivation of hippocampal ensemble memories during sleep." *Science* 265:5172, 1994, pp. 676–679.

[103] A. K. Lee and M. A. Wilson. "Memory of sequential experience in the hippocampus during slow wave sleep." *Neuron* 36:6, 2002, pp. 1183–1194.

[104] D. J. Foster and M. A. Wilson. "Reverse replay of behavioural sequences in hippocampal place cells during the awake state." *Nature* 440:7084, 2006, pp. 680–683.

[105] K. Diba and G. Buzsáki. "Forward and reverse hippocampal place-cell sequences during ripples." *Nature neuroscience* 10:10, 2007, pp. 1241–1242.

[106] A. S. Tolias et al. "Recording chronically from the same neurons in awake, behaving primates." *Journal of neurophysiology* 98:6, 2007, pp. 3780–3790.

[107] B. K. D. Wise et al. "Microelectrodes, microelectronics, and implantable neural microsystems." *Proceedings of the IEEE* 96:7, 2008, pp. 1184–1202.

[108] K. Najafi and K. D. Wise. "An implantable multielectrode array with on-chip signal processing." *Solid-State Circuits, IEEE Journal of* 21:6, 1986, pp. 1035–1044.

[109] H. P. Neves and P. Ruther. "The neuroprobes project." In: *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*. IEEE. 2007, pp. 6442–6444.

[110] H. P. Neves et al. "The NeuroProbes project: a concept for electronic depth control." In: *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*. IEEE. 2008, pp. 1857–1857.

[111] C. M. Lopez et al. "An implantable 455-active-electrode 52-channel CMOS neural probe." *Solid-State Circuits, IEEE Journal of* 49:1, 2014, pp. 248–261.

[112] J. Scholvin et al. "Close-packed silicon microelectrodes for scalable spatially oversampled neural recording," 2015.

[113] J. L. Shobe et al. "Brain activity mapping at multiple scales with silicon microprobes containing 1,024 electrodes." *Journal of neurophysiology* 114:3, 2015, pp. 2043–2052.

[114] S. Takeuchi et al. "3D flexible multichannel neural probe array." *Journal of micromechanics and microengineering* 14:1, 2003, p. 104.

[115] M. E. Merriam et al. "A three-dimensional 64-site folded electrode array using planar fabrication." *Microelectromechanical Systems, Journal of* 20:3, 2011, pp. 594–600.

[116] M. E. Merriam et al. "A 3-D 160-site microelectrode array for cochlear nucleus mapping." *Biomedical Engineering, IEEE Transactions on* 58:2, 2011, pp. 397–403.

[117] G. E. Perlin and K. D. Wise. "A compact architecture for three-dimensional neural microelectrode arrays." In: *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE.* IEEE. 2008, pp. 5806–5809.

[118] M.-Y. Cheng et al. "3D probe array integrated with a front-end 100-channel neural recording ASIC." *Journal of Micromechanics and Microengineering* 24:12, 2014, p. 125010.

[119] J. J. Riera, T. Goto, and R. Kawashima. "A methodology for fast assessments to the electrical activity of barrel fields in vivo: from population inputs to single unit outputs." *Front. Neural Circuits* 8:4, 2014, pp. 10–3389.

[120] S. Herwik et al. "Fabrication technology for silicon-based microprobe arrays used in acute and sub-chronic neural recording." *Journal of Micromechanics and Microengineering* 19:7, 2009, p. 074008.

[121] A. Berényi et al. "Large-scale, high-density (up to 512 channels) recording of local circuits in behaving animals." *Journal of neurophysiology* 111:5, 2014, pp. 1132–1149.