Technological and Computational Approaches for Large Count High-Density Neural Probes

by

Behnoush Rostami

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical and Computer Engineering)
in the University of Michigan
2024

Doctoral Committee:

Professor Khalil Najafi, Chair
Professor Omar Ahmed
Professor Yogesh Gianchandani
Professor Emeritus Kensall D. Wise

Behnoush Rostami

behnoosh@umich.edu

ORCID iD: 0000-0002-4494-5802

© Behnoush Rostami 2024

Dedication

To my beloved family, whose constant love and endless support have been my guiding light through every challenge and success.

Acknowledgements

I extend my deepest gratitude to my advisor, Professor Khalil Najafi, for his tremendous influence on my academic and personal growth. His unwavering support and the opportunities provided during my PhD studies have been invaluable. I am profoundly grateful for his support and for entrusting me with the opportunity to contribute to remarkable research projects.

I also want to extend my gratitude to my thesis committee members, Professor Kensall Wise, Professor Yogesh Gianchandani, and Professor Omar Ahmed. Their time and invaluable feedback on my thesis have been instrumental in refining my work. I am honored and privileged to have had them on my committee.

A special thanks goes out to the staff of the Lurie Nanofabrication Facility (LNF), especially Sandrine Martin, Pilar Herrera-Fierro, Katharine Beach, Terre Briggs, Kevin Owen, Shawn Wright, Brian Armstrong, Jorge Barreda, and Robert Nidetz who have greatly assisted me during my time in the cleanroom.

The EECS staff members have also assisted me in many ways, and I want to take this opportunity to express my gratitude. Special thanks to Kristen Thornton, Trasa Burkhardt, and Robert Gordenker.

I also express gratitude to all my former and present friends and colleagues in EECS2001 and OmarLab, from whom I have learned many technical and non-technical lessons. Their support and friendship have made this journey more delightful. Special thanks to Vaughn Hetrick, Alex

Johnson, Stacey Yemin Tang, Kimberly Beers, Jae Yoong Cho, Sajal Singh, Jong Kwan Woo, Amin Sandoughsaz, and Tal Nagourney.

Furthermore, I wish to thank my dear friends in Ann Arbor and the US, who have made me feel at home. Their presence throughout the extremes and peaks of this journey has brought me immense joy and support.

I am deeply grateful to my parents for their unconditional love and support throughout the various stages of my life. Their sacrifices for my education and career have been invaluable, and they continue to be a constant source of inspiration for me. I also want to give a big shout out to my incredibly supportive siblings, Behrouz, Behzad, and Behnaz, who have always been there for me in every step of my life. They are not just my siblings but also my best friends, and I am grateful for their support and companionship throughout the years.

Finally, I want to express my heartfelt gratitude to Milad, who has been my rock through the highs and lows of this journey, bringing immense happiness into my life. I feel incredibly blessed to have him by my side, and I know that without his love, support, and patience, completing this journey would not have been possible.

Table of Contents

Dedication ii

Acknowledgements iii

List of Figures vii

Abstract xiv

Chapter 1 Introduction 1

1.1 Penetrating Neural Probes 1

1.2 High-Density and Large-Count Neural Interface 4

1.3 Considerations to Minimize Tissue Damage 11

1.4 Tip Metallization Techniques in Large Count 3D Arrays 16

1.5 Packaging and Integration 20

1.6 Challenges in Spike Sorting for Large-Count High-Density Neural Interfaces 21

1.7 Research Objectives and Thesis Contributions 23

1.8 Organization of the Thesis 25

Chapter 2 High-Density Multi-Shank 2D Planar Silicon Probes for Precise Recording with Reduced Tissue Damage 27

2.1 Introduction 27

2.2 Optimization of Mechanical Properties for Reduced Tissue Damage 30

2.3 Design and Microfabrication Process 37

2.4 Engineered Mechanical and Electrical Properties 51

2.5 Conclusion 70

Chapter 3 High-Density, Large-Count 3D Non-Planar Neural Probes 71

3.1 Introduction 71
3.2 Microfabrication Process 83
3.2.1 Etching and Refilling Steps 83
3.2.2 Integrated Flexible Cable 91
3.2.3 Tip Metallization and Releasing Shanks 97
3.2.4 Tip Metallization Technique for Varying Shank Lengths Neural Probe Arrays ... 112
3.3 Mechanical and Electrical Characterizations 118
3.4 Conclusion 124
Chapter 4 Computational Approaches for Large-Count High-Density Neural Interfaces 125
4.1 Introduction 125
4.1.1 Methods: Manual and Automated 126
4.1.2 Advancements in Automated Spike Sorting for High-Density Neural Interfaces . 130
4.1.3 Curation after Spike Sorting 131
4.2 Dynamic L-ratio 140
4.3 Spike Sorting using Dynamic L-ratio 148
4.4 Conclusion 158
Chapter 5 Conclusion and Future Works 160
Bibliography 165

List of Figures

Figure 1.1: A comparison of the limitations and applications of four main categories of brain study technology: fMRI, EEG, ECoG, and Implantable Microelectrodes. 2

Figure 1.2: Planar and out-of-plane configurations of micromachining-based neural probes. 3

Figure 1.3: The neuroscience version of Moore's Law: scaling up of neural recordings [25, 26]. 6

Figure 1.4: Neural probes for high-density and large-scale recordings, (a) Neuropixels [37, 39], (b) NeuroSeeker [38], (c) High-density microwire bundle connected to a CMOS amplifier array [33], (d) Nanofabricated neural probes [32], (e) Michigan's Sea of Electrode Arrays [23]. 8

Figure 1.5: Neuron-like electronics (NeuE) neural probe, featuring a biomimetic design inspired by the structure and function of neurons [50, 51] 13

Figure 1.6: Nanoelectronics thread (NET) probe, featuring a SU-8 substrate material [52]........ 14

Figure 1.7: Evolution of methods for de-insulating and metallizing needle tips in 3D arrays over the past two decades [56-65]. 18

Figure 1.8: Summary of main challenges and proposed solutions in this thesis. 25

Figure 2.1: Principal concept of crafting T-bar shaped and π-bar shaped cross sections for planar shanks to achieve an optimal balance between their volume and stiffness. 31

Figure 2.2: FEM modeling to show the influence of stiffeners, shank thickness, and cross-sectional modifications on probe mechanical properties. 33

Figure 2.3: A publicly accessible dataset of three-dimensional microvascular networks in the rat brain was employed to simulate the invasion of blood vessels caused by the insertion of different geometries. 34

Figure 2.4: (a) A visual representation of the 3D modeling of the implanted shank and its interaction with the blood vessels, (b) thick shank, (c) π-bar shaped shank, (d) thin shank, (e) T-bar shank. 36

Figure 2.5: Fabrication process of large count high density planar neural probes with embedded stiffeners and integrated flexible interconnect cables. 40

Figure 2.6: SEM images of different shanks after patterning the n-doped polysilicon layer to create conductive sites and traces. 41

Figure 2.7: Cross-sectional view of the probe, showing the total thickness of the shanks and the location of the stiffeners........ 42

Figure 2.8: Schematic comparison between flat and stair shape openings in the insulator layer, highlighting the contact between the metal film and the underlying conductive layer. 43

Figure 2.9: The von Mises Equivalent Elastic Strain rainbow-colored maps, (a) on the surface of the shanks, (b) in the vicinity of the contact area between the metal site and the n-poly silicon........ 44

Figure 2.10: Comparison of von Mises equivalent stress on the surface of shanks between stair shape and flat shape. 45

Figure 2.11: Comparison of von Mises equivalent stress at the contact area between the metal layer and the underlying n-poly silicon for stair shape and flat shape configurations. 46

Figure 2.12: SEM images of the stair shaped openings at the recording sites which are strategically positioned where the metal layer contacts the conductive layer. This design aims to improve the ohmic connection between the metal layer and the conductive layer, while also lowering mechanical strain on the metal layer at the contact region with the underlying conductive layer. 47

Figure 2.13: SEM images showcasing probes post-patterning of the initial Parylene-C layer on backends and cables. The first metal layer on recording sites is visible, along with the removal of Parylene-C from shanks. Additionally, openings on probe backends, representing the connection region between the flexible cable and the probes, are highlighted. The shanks and recording sites are shown to be clean after etching the initial Parylene-C layer. 48

Figure 2.14: SEM images of the metal traces post lift-off, prior to the deposition of the second Parylene-C layer on probes' backends and cables. The image highlights step coverage along the edge of the Parylene-C layer, with visible metal layers on the recording sites. 49

Figure 2.15: SEM images of device components showcasing customizability with diverse shank sizes, tip shapes, stiffener variations (0 to 3), and flexible cables connected to backends. Notably, innovative 3D recording sites enhance effective surface area for more efficient recordings. Illustration depicts devices post-plasma etching on a Santovac 5 layer, before release. 50

Figure 2.16: Exploring shank design flexibility: a collection of SEM images that illustrate various shank shapes, sizes, and angles incorporated in our innovative designs. 53

Figure 2.17: High-density recording sites, different tip shapes, two-step metallization, and compact n-poly silicon conductive traces on the shanks utilized in our designs. 54

Figure 2.18: SEM images of different aspects and different designs of the presented planar probes. 55

Figure 2.19: Optical images from different aspects of the presented planar probes........ 56

Figure 2.20: (a) Comprehensive overview of device components, (b) SEM image of flexible cable, (c) Optical microscopic images of released multi-shank probes with compact, high-density recording sites integrated with flexible cables. 57

Figure 2.21: Wire bonding and packaging process for preparing probes. (a) Overall view of a 4-shank probe packaged and wire-bonded to the PCB. (b) Layout of 64 pads on the backside of the integrated flexible cable and the PCB for wire bonding. (c, d) Zoomed-in views of the 0.7 mil Al wire bonds. (e) Use of a plastic cover to protect the delicate shanks during the wire bonding process. 58

Figure 2.22: The impedance for selected recording sites, (a) Magnitude, (b) Phase.................... 60

Figure 2.23: Impact of stiffeners on shanks - Comparative visualization of thin, T-bar, and π-bar designs in test samples, revealing the influence on post-release orientation for precise implantation. 61

Figure 2.24: Demonstrating the exceptional compliance of probe shanks, entwining upon removal from a solution and effortlessly straightening upon re-immersion. A 4-shank probe is initially removed from a low-concentration agarose gel, revealing its inherent flexibility: (1) straightened shanks in the Agarose gel, (2-5) removing shanks from the Agarose gel, (6) entwined shanks outside of the Agarose gel, (7-9) reinserting the shanks in the Agarose gel, (10) re-straightened shanks inside the Agarose gel. 62

Figure 2.25: Mechanical test: implanting probes into tofu, a brain tissue-mimicking material, with saline covering to facilitate the process, (1-3) implanting the probe into tofu, (4-9) removing intact shanks from tofu. 63

Figure 2.26: Continuous monitoring of recording site impedance performance on a probe immersed in a PBS solution over an 11-month period. 64

Figure 2.27: SEM images of the studied recording sites after over 11 months of immersion in PBS. 66

Figure 2.28: In-vitro assessment of probe recording capabilities. (a) The generated neural action potential-like signal, with a 40 Hz frequency, applied to a reference electrode in PBS solution, demonstrated responses across 14 studied sites. (b-c) A simple lumped circuit model, comprising resistors and capacitors, was derived using the transient responses obtained from the experimental setup. The simulated data closely matched the actual measurements.............. 67

Figure 2.29: Assessment of stimulation capability in an in-vitro test. (a) Color-coded representation of 14 tested recording sites. (b) Stimulated signal marked by red asterisks (*) and recorded signals from 14 neighboring sites, shown in color, demonstrating the probes' responsiveness. (c-d) Equivalent lumped circuit comprising resistors and capacitors, was also derived using the transient responses obtained from this experimental setup. The simulated data closely matched the actual measurements, validating the functionality of the probes in stimulating and recording signals in an in-vitro setting... 69

Figure 3.1: Illustration of the key features of out-of-plane neural probes, including customizable needle density, high-count array, smaller cross-sectional area, customizable length, and tip sharpness. These features are crucial for enhancing the functionality and effectiveness of neural probes in studying complex neural processes and developing advanced brain-machine interfaces [23]. 74

Figure 3.2: Fabrication steps of SEA before metallizing the tip, (a) DRIE of a hole on a Si substrate, (b) filling the hole with multiple insulator and conductor layers, (c) releasing the needle, (d) a 3×3 SEA, (e) Close-up of a tip........ 77

Figure 3.3: Some of the SEA array fabrication limitations: (a) Wafer bonding misalignment (left) results in undesired shifts along the electrode shanks (right), (b) Formation of necking along the electrode shank at the bonded wafers junction [23]........ 78

Figure 3.4: Major limitations and challenges of the SEA array fabrication technique [23]......... 80

Figure 3.5: Illustration of the key features of the proposed technology for fabricating high-count, high-density non-planar probes with minimal invasiveness. 82

Figure 3.6: SEM images of etched holes with a 20µm diameter, exclusively etched from the top side of the wafer using a silicon dioxide and photoresist mask. 84

Figure 3.7: Smoothing the edges by depositing thermal silicon dioxide on the etched wafer and then removing it. This process helps with resolving the necking issue where holes with different diameters meet in the substrate. 85

Figure 3.8: Illustration of the steps involved in the fabrication process of the DSE version of high-count, high-density non-planar neural probe arrays, including depositing different layers and refilling the holes........ 86

Figure 3.9: SEM images revealing the intricate cross-section of refilled holes post-insulator layer deposition, featuring a 45-degree tilted side view from the top side of the substrate. The sacrificial layer is indicated by white arrows, while the insulating layer is represented by black arrows........ 87

Figure 3.10: SEM images of the refilled holes on the topside of the wafer, shown from a 45-degree tilted side view and a top view. 88

Figure 3.11: SEM Images of LPCVD Refilled Holes on the Wafer's Bottom, illustrating near-complete refilling, subtle indentations, and negligible nanometer-scale openings at the base of each. 89

Figure 3.12: SEM images showing the inside of LPCVD refilled holes, illustrating detailed layers within shanks at midpoints and the top points of the shanks—all captured in a 45-degree tilted side view. 90

Figure 3.13: Optical microscopic images showing the etched ONONO layer and the separated conductive layers of individual shanks in each array. 90

Figure 3.14: SEM images depicting the etched insulator layer and the separated conductive layers of individual shanks in each array. 91

Figure 3.15: SEM images of metal lines and contact pads at the base of arrays, highlighting the varying channel connections (16 to 128) and array sizes (10 × 10 to 100 × 100) with small pitches (50µm or 100µm). 92

Figure 3.16: Optical images showcasing key components, including metalized contact pads on array bases, metal lines on the patterned first PI 2611 layer, and their connection to the backend leading to flexible cables. 93

Figure 3.17: Optical images showcasing insulated metal lines, metalized contact pads on array bases, and metal connections to backends after the patterning of the second PI 2611 layer. 94

Figure 3.18: Overview of devices and components fabricated on the top side of the Si substrate in DSE version. 95

Figure 3.19: Illustration of the steps involved in the fabrication process of the SSE version of high-count, high-density non-planar neural probe arrays, including depositing different layers and refilling the holes 96

Figure 3.20: Tip metallization process in DSE followed by releasing the shanks. 98

Figure 3.21: Schematic illustration of the electroplating and tip sharpening process in DSE version. The top image shows a cross-sectional view, while the bottom image shows a top view. The process involves a series of steps, including metallization, masking, electroplating, and etching, to create sharp crescent-wedge-shaped tips with a high effective surface area. 99

Figure 3.22: SEM images of electroplated tips in DSE version, (a) before gold plating, (b) after gold plating, showcasing their sharp and wedge-shaped profiles. The porosity of the electroplated gold, due to isotropic etching, enhances the effective surface area, leading to reduced recording site impedances. 101

Figure 3.23: A 15µm thick photoresist mask (AZ 12 XT) is deposited on the tip locations and cable backends, providing protection during the fast DRIE process. 102

Figure 3.24: Tip metallization process in SSE followed by releasing the shanks. 103

Figure 3.25: Emerged tips of arrays in SSE version. Showing arrays with various sizes and pitches after maskless dry etching of the substrate. The tips are coated with an insulator layer. 104

Figure 3.26: SEM images of the SSE version tip metallization process. (a) Emerged tips after sputtering Cr/Au. (b) Photoresist mask (AZ 12XT, 10µm) used for etching the Cr/Au through an isotropic etching process. (c) 10 x 10 array after isotropic etching of the Cr/Au layer. (d) Metallized tips after stripping the photoresist layer. (e) and (f) SEM images of one emerged tip before and after etching the insulating layer on top, respectively. (g) Zoomed-in image of the same tip after metallization with the SSE version method. 105

Figure 3.27: Process flow diagram illustrating the fabrication steps for both DSE and SSE types of non-planar neural probes. Steps 1-9 detail the process with the left side representing DSE and the right-side representing SSE. From steps 10-19, the wafer is flipped, showcasing the reverse perspective with the right side now illustrating DSE and the left side portraying SSE. 107

Figure 3.28: SEM images of high-count high-density non-planar arrays, integrated cables, after dry etching the substrate, showing variations in sizes and densities among the samples. 108

Figure 3.29: Illustration of the wafer containing samples before and after release from the carrier wafer, showing the polyimide bridges and frames surrounding the devices. 109

Figure 3.30: Comparison of released samples with a finger and a penny for size reference. The Omnetics connector is attached to one of the backends, illustrating the use of the flexible cable's backend as a PCB for easy and fast connector soldering post-release. 110

Figure 3.31: SEM images of needle components in the DSE version, showing ramped junctions and no misalignment between the top and bottom holes. This design improvement addresses breaking points seen in SEA arrays. The arrays feature high-count, high-density, sharp gold-plated tips, and other customizable features. 111

Figure 3.32: EP tip metallizing process, (a) The COMSOL simulation of 3D-P electroplating on a 2 × 2 SEA array. The rainbow-colored legend shows the thickness of EP on the surface of the array, (b-1) a 3D neural probe (b-2) electroplating 3D-P, (b-3) exposing the EP, (b-4) DRIE the insulator layer, (b-5) depositing metal layer, (b-6) lift-off process to create the metallized tips. 114

Figure 3.33: Implementing the EP metallization technique on SEAs, (a) SEM of 2×2 and 3×3 SEA arrays before electroplating photoresist, (b) After electroplating 3D-P, (c) Needles after exposure and developing the EP, (d) Close-up of a de-insulated and metallized tip. 116

Figure 3.34: The impedance for 32 recording sites, (a) Magnitude, (b) Phase. 119

Figure 3.35: Lumped circuit model derived from the impedance measurements, the close alignment between the simulated and measured impedance data indicates the accuracy of the model in predicting the electrical behavior of the system. These results closely match the actual impedance measurements shown in Figure 3.34. 120

Figure 3.36: Assessment of stimulation capability in an in-vitro test. Stimulated signals marked by ‘1’ and recorded signals from 15 neighboring sites demonstrate the probes' responsiveness. 121

Figure 3.37: A lumped circuit model, comprising resistors and capacitors, was also derived using the transient responses obtained from this experimental setup. The simulated data closely matched the actual measurements. 122

Figure 3.38: Implantation of high-density electrode arrays into tofu, showing the shanks robustness and the flexibility of the meandered cables. 123

Figure 3.39: A 40 × 40 array after implantation in tofu, demonstrating the robustness of the electrodes with all shanks remaining intact. 123

Figure 4.1: Comprehensive overview of steps involved in spike sorting, (a-b) Recording changes in extracellular electrical potential resulting from nearby neurons' activity. (c) Preprocessing data (d) Spike detection (e) Detected sikes are aligned based on a common feature, such as the position of the voltage peak. (f) Feature extraction techniques, including principal component analysis, (g-h) The critical steps of clustering to group spikes based on their features, with each cluster ideally representing spikes from a single neuron.................... 128

Figure 4.2: Confusion matrix for spike sorting models. 129

Figure 4.3: Phy GUI, (a) cluster scatter view, (b)cluster view, (c) similarity view, (d) waveform view, (e)amplitude view, (f) correlogram view, (g) ISI view, (h) firing rate view, and (i) probe view. 134

Figure 4.4: Illustration of the dynamic L-ratio calculation process. (a) Cluster locations over time, with the time span divided into small time divisions (Δt). (b) Scatter plot of spike features for each time division. (c) Spike waveforms for six example units recorded from four channels. (d) Distribution of the Mahalanobis distance from the fitted cluster center to the observed spikes, with dashed lines indicating theoretical distributions for Gaussian distributed spikes. . 143

Figure 4.5: A graphical representation of how the input data is transformed or processed by the toolbox to produce the output data.. 146

Figure 4.6: Dynamic L-ratio Analysis of a KiloSort Cluster Over Time Windows of Varying Durations (30, 60, 120, and 180 seconds) in a Mouse Hippocampus Dataset provided by Professor Ahmed’s lab, University of Michigan, Ann Arbor. 148

Figure 4.7: Illustration of the supervised learning process for the spike sorting algorithm, showcasing the sequential steps from raw data sorting to the generation of dynamic L-ratio data and dynamic firing rates, ultimately leading to the classification of clusters within new datasets. 150

Figure 4.8: Visualization of the “loss” landscape travelled by the Adam stochastic optimization algorithm. Adam seeks to minimize loss by traversing the network and adjusting parameters. Reproduced from [141]. 154

Figure 4.9: Illustration of the learning algorithm applied to classify clusters into 'Good' and 'Bad' classes.. 155

Figure 4.10: Performance comparison of the prediction results when training the same learning algorithm using dynamic data from dynamic L-ratio and dynamic firing rate vs. when it is only trained using classic L-ratio data. 157

Abstract

Implantable neural probes with various shapes, designs, and materials are extensively used to study the brain by recording the electrical and chemical responses of neural structures and circuits. This thesis addresses four critical challenges in the development of advanced neural interfaces aimed at mapping large collections of neurons.

The first challenge is to develop innovative technological methods for microfabricating high-count, high-density probes with user-defined features such as density, size, shape, and distribution for specific applications. A new class of silicon-based two-dimensional (2D) planar neural probes is developed incorporating at least four shanks, each featuring over 16 recording sites with high density (320 electrodes/mm^2) and narrow vertical (6μm) and horizontal (10.5μm) separation between sites. Each shank is as narrow as 43μm and as long as 5mm. Furthermore, a new class of high-count, high-density three-dimensional (3D) non-planar silicon-based neural interfaces has been developed, allowing over 10,000 slender shanks each supporting a recording/stimulation site at its tip, and providing design flexibility in array size, density (400 electrodes/mm^2), and distribution.

The second challenge is the engineering of the electrical and mechanical features of individual probe shanks to make them minimally invasive and suitable for long-term measurements. This involves optimizing the probe shank design to be more mechanically compliant to reduce foreign body response while ensuring it is sufficiently stiff to be implanted and reach the targeted region with minimal buckling and without the need for large mechanical insertion shuttles. The proposed planar 2D probes have T-shaped and π-shaped cross-sections, with a top side of 43μm wide and 3μm thick, and vertical stiffeners that are 4μm wide and 10μm thick. These new geometries reduce the shank cross-section and volume by ~3 times compared to conventional thick rectangular cross-section shanks while providing similar mechanical stiffness. The proposed 3D non-planar silicon-based needle arrays have been manufactured with needles ranging in length from 0.5-1.5mm, diameters from 10-15μm, and with sharp tips < 2μm wide. Three advanced technologies are

proposed to create well-defined recording and stimulating sites at the tips of these needles. Electrode robustness, insertion and recording functionality both acute and long-term have been demonstrated by mechanical and electrical in vitro tests.

The third challenge involves packaging and integration. Flexible, miniaturized, and robust Parylene-C and Polyimide cables, which are more than 20mm long, ~ 1mm wide, 5-15µm thick, and carry hundreds of interconnect lines, are monolithically integrated with the 2D and 3D probes. This integration significantly improves the ease of handling and reduces mechanical tethering on probes during measurements. Especially-designed Polyimide cables are integrated with large-count 3D arrays. They are solder attached directly to external connectors or readout electronics, thus avoiding the need for traditional wire bonding. This innovative approach significantly reduces the complexity and labor associated with external connections for large-count electrode arrays, thus making the packaging process faster, more efficient, and more reliable.

Finally, high-count neural recording presents computational challenges, particularly in spike-sorting. A new Python Toolbox that calculates a dynamic version of the L-ratio as a quality metric for assessing cluster isolation as well as providing additional temporal information is developed. By leveraging this new information, a neural network learning algorithm to automatically curate the spike-sorting results has been trained. This approach achieved 50-100% higher F1-score on average compared to when we trained the same algorithm with the classic L-ratio.

Chapter 1 Introduction

1.1 Penetrating Neural Probes

The average human brain is composed of approximately 80 to 100 billion neurons that continuously communicate with one another through electrochemical signals. Since the late 1700s [1], researchers have been intrigued by the study of brain functionality through the exploration of neuronal electrical excitability. In 1791, Italian physician Luigi Galvani made a significant discovery, revealing that both muscle cells and neurons exhibit electrical excitability and responsiveness. Since then, the field of neuroscience has relied on neuronal recording techniques as essential tools for research and practical applications [2].

Over the past two centuries, dedicated efforts by neuroscientists have led to a partial understanding of the mechanisms underlying intricate brain activities, thanks to advancements in technologies to monitor and image activities of neural structures. These technologies can be broadly classified into four main categories: neuroimaging techniques (such as fMRI), scalp electroencephalography (EEG), electrocorticogram (ECoG), and neural interfacing using implantable microelectrodes. These tools enable researchers to both modulate and detect local field potentials (LFP) and single unit spikes, facilitating the investigation of subtle brain phenomena. Through the continuous development of these techniques, our knowledge of the brain's function continues to expand.

Currently, penetrating neural probes stand as the predominant and exclusive technology employed for both micro-stimulation and recording of electrical and chemical activities at the level

of individual neurons within the brain [3-5]. Beyond their research applications, these implantable devices offer significant clinical advantages, such as deep brain stimulation for treating conditions like Parkinson's disease and epilepsy [4, 6]

Moreover, a profound understanding of neurological signals and neurochemical processes holds the potential to restore certain brain functions after severe brain injuries [7]. An illustrative example of such clinical progress lies within the field of neuro-prosthetics, where damaged motor, sensory, or cognitive modalities can be effectively replaced by neural prostheses [8]. Cochlear implants serve as a widely recognized instance of such devices [9]. Figure 1.1 compares the limitations and applications of the four main categories of brain study technologies: fMRI, EEG, ECoG, and Implantable Microelectrodes.

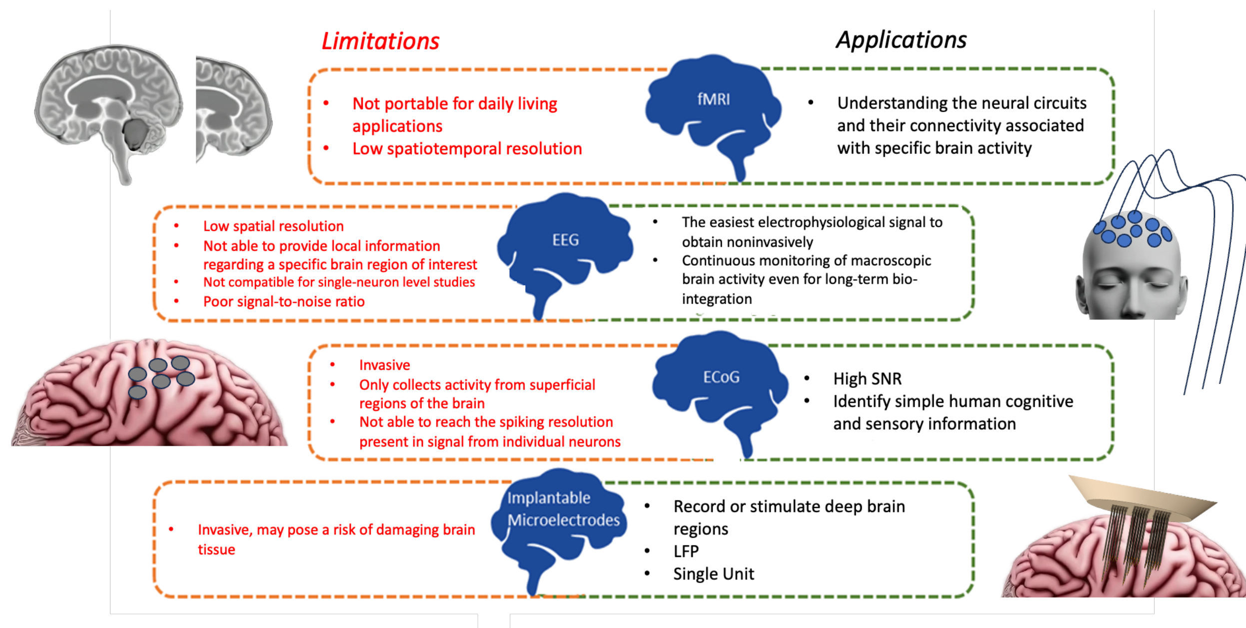

Figure 1.1: A comparison of the limitations and applications of four main categories of brain study technology: fMRI, EEG, ECoG, and Implantable Microelectrodes.

The pursuit to develop intracortical neural probes has a rich history of research, with its origins tracing back to early studies employing metal wire electrodes in the early twentieth century

[10] and subsequent utilization of insulated wires, either singular or bundled, featuring recording sites at their tips for extracellular electrical activity monitoring [11, 12]. In 1970, Kensall Wise introduced pioneering work on silicon-based microelectrodes, leveraging advancements in microfabrication techniques [13]. Over time, the progressive evolution of microelectromechanical systems (MEMS) and silicon micromachining techniques unlocked a multitude of possibilities in developing miniaturized neural probes capable of precise recording and stimulation of neural circuitry [6, 13-20].

Micromachining-based implantable penetrating probes can be classified into two main categories: planar and out-of-plane configurations [15, 21]. In planar probes, such as the Michigan probe, the recording and stimulation sites are positioned along a two-dimensional (2D) shank (Figure 1.2). The inherent 2D nature of planar probes presents challenges when it comes to gathering data from spatially dispersed locations. To address this limitation, individual 2D shanks can be assembled to form three-dimensional (3D) arrays [22]. There are other types of 3D arrays that do not require the assembly of 2D electrodes, such as the Utah Electrode Array (UEA) [21] and Michigan's Sea of Electrode Array (SEA) [23, 24], as shown in Figure 1.2.

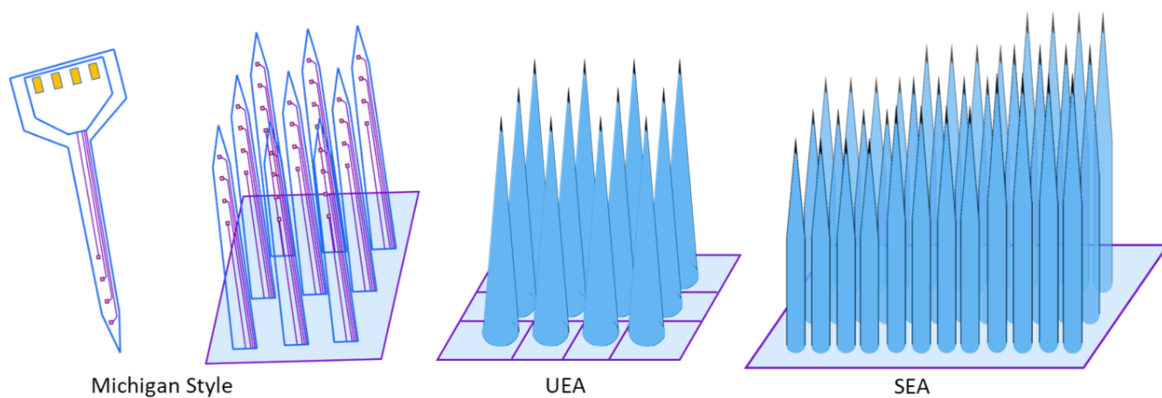

Figure 1.2: Planar and out-of-plane configurations of micromachining-based neural probes.

Dense UEA arrays are created by cutting a highly p-doped silicon substrate into pillars, which are subsequently subjected to a wet etching process to form sharp needles. The needles are then metallized and coated with Parylene for insulation [21]. Michigan's SEA arrays are fabricated using an ultra-high aspect-ratio deep reactive ion etching (DRIE) technique, which enables the creation of arbitrary hole arrangements on a silicon substrate. The holes are then filled with insulator and conductor layers to form a high density of needles [23]. The tips of needles in both of these arrays have to be processed to remove any dielectrics and coated with a suitable recording/stimulation metal.

Over the past few years, there have been remarkable advancements in the application of planar and out-of-plane neural probes, reflecting significant progress made in recent decades. Nevertheless, researchers are currently dedicated to the pursuit of an ideal neural probe—one that not only provides customizable mechanical and electrical features but also minimizes tissue damage and maintains long-term functionality for chronic studies. This dedication to innovation in developing new penetrating neural probe technologies has the potential to greatly enhance our understanding of how the brain works and create exciting possibilities for groundbreaking therapeutic treatments.

1.2 High-Density and Large-Count Neural Interface

In order to enhance our understanding of the coordinated processes involved in brain computations, it is essential to improve our capacity to simultaneously monitor a greater number of individual neurons. Meeting this demand necessitates the use of high-density and large-count arrays of recording sites facilitating the simultaneous interface with a larger population of neurons. The advent of these large-count and high-density probes has revolutionized research methodology, moving away from the traditional emphasis on selecting specific neurons and brain areas. Instead,

neuroscientists can now conduct unbiased surveys of neural responses and cell types. This approach greatly enhances the chances of encountering neurons with rare responses and enables the capture of subtle signals that may be distributed across multiple cells or associated with infrequent cell types.

Moreover, through the simultaneous recording of multiple neurons, we can acquire comprehensive data sets without depending on a large cohort of animals. This ethical consideration promotes the refinement and reduction of experimental procedures.

In 2011, Stevenson and Kording [25] introduced a neural recording version of "Moore's Law," predicting a doubling of concurrently recorded neurons approximately every 7 years, as depicted in Figure 1.3. The figure demonstrates the number of simultaneously recorded neurons using electrophysiology, indicated by dark blue dots (with squares representing the original data used for the fit by Stevenson and Kording, 2011), and optical calcium imaging techniques, represented by red dots [26].

Interestingly, calcium imaging, represented by red dots in the figure, has exhibited even faster advancements than initially anticipated. Calcium imaging is a technique that enables researchers to monitor neuronal activity by detecting changes in intracellular calcium levels. It relies on fluorescent calcium indicators that emit fluorescence signals upon binding with calcium ions. The advancements in fluorescent indicators and optical imaging techniques have facilitated the rapid mapping of large populations of neurons and the observation of neural activity at a microscopic level.

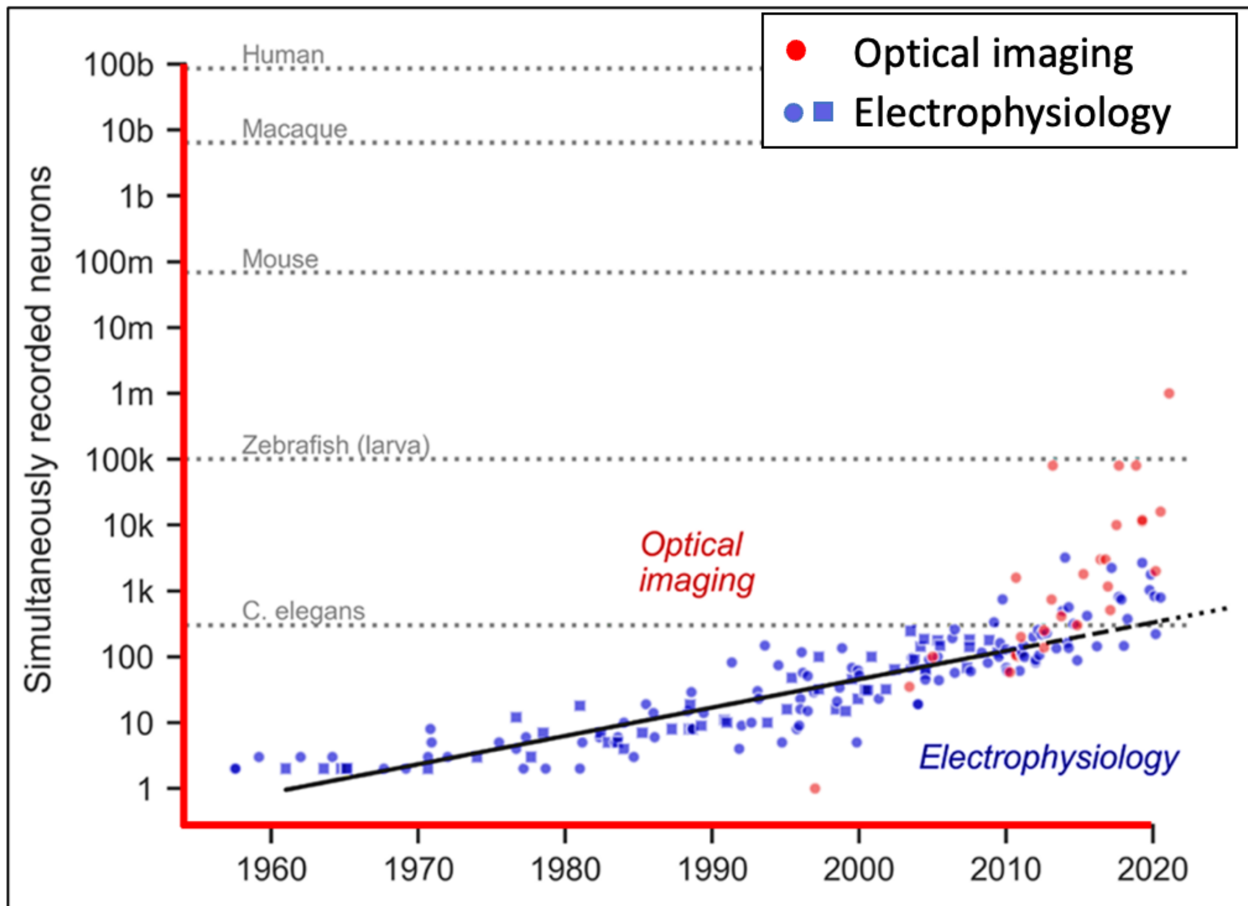

Figure 1.3: The neuroscience version of Moore's Law: scaling up of neural recordings [25, 26].

Calcium imaging, although a valuable technique for studying neuronal activity, has several limitations when compared to extracellular electrophysiological recordings. It indirectly senses neuronal spiking activity through changes in calcium levels, leading to slower kinetics and poorer temporal resolution than electrophysiological methods. Today's calcium indicators have rise and decay rates of tens to hundreds of milliseconds, however, the temporal resolution of electrophysiological recordings can be in the range of microseconds to milliseconds, allowing for the detection of rapid changes in neural activity [27].

Calcium imaging requires the genetic expression of indicators in target cells, which is not necessary for extracellular recordings. Additionally, calcium imaging often involves invasive

procedures to visualize imaged neurons, such as skull removal, whereas electrode implantation is less invasive after the initial implantation and surgery. Consequently, extracellular signals offer more flexibility for extracting relevant information from large populations of neurons, with higher temporal resolution and without the need for genetic manipulation or invasive interventions.

Remarkable progress has been made in simultaneously measuring the activity of individual neurons throughout the entire brain in small animals such as C. elegans [28], larval zebrafish [29], hydra [30], and Drosophila [31, 32] using a combination of electrophysiological and calcium imaging methods. However, capturing electrophysiological recordings from all cortical neurons in mammals remains a formidable challenge, despite its theoretical feasibility [33, 34]. Based on the extrapolation of current trends, it is estimated that the realization of whole-brain, single-neuron recordings in mice could potentially take anywhere from two decades to a century [26].

Planar probes, such as the Michigan-type silicon probe, possess the capability to incorporate multiple recording sites within a single shank. This integration can be expanded in terms of quantity and density by utilizing microfabrication techniques. However, maintaining a small shank width while increasing the number of recording sites poses a challenge.

The incorporation of complementary metal-oxide-semiconductor (CMOS) technology has enabled the integration of recording electrodes with multiplexing circuits, resulting in compact input/output connections [35, 36]. This has facilitated the design and fabrication of large-scale neural recording electrodes such as Neuropixels [37], and NeuroSeeker [38] (See Figure 1.4).

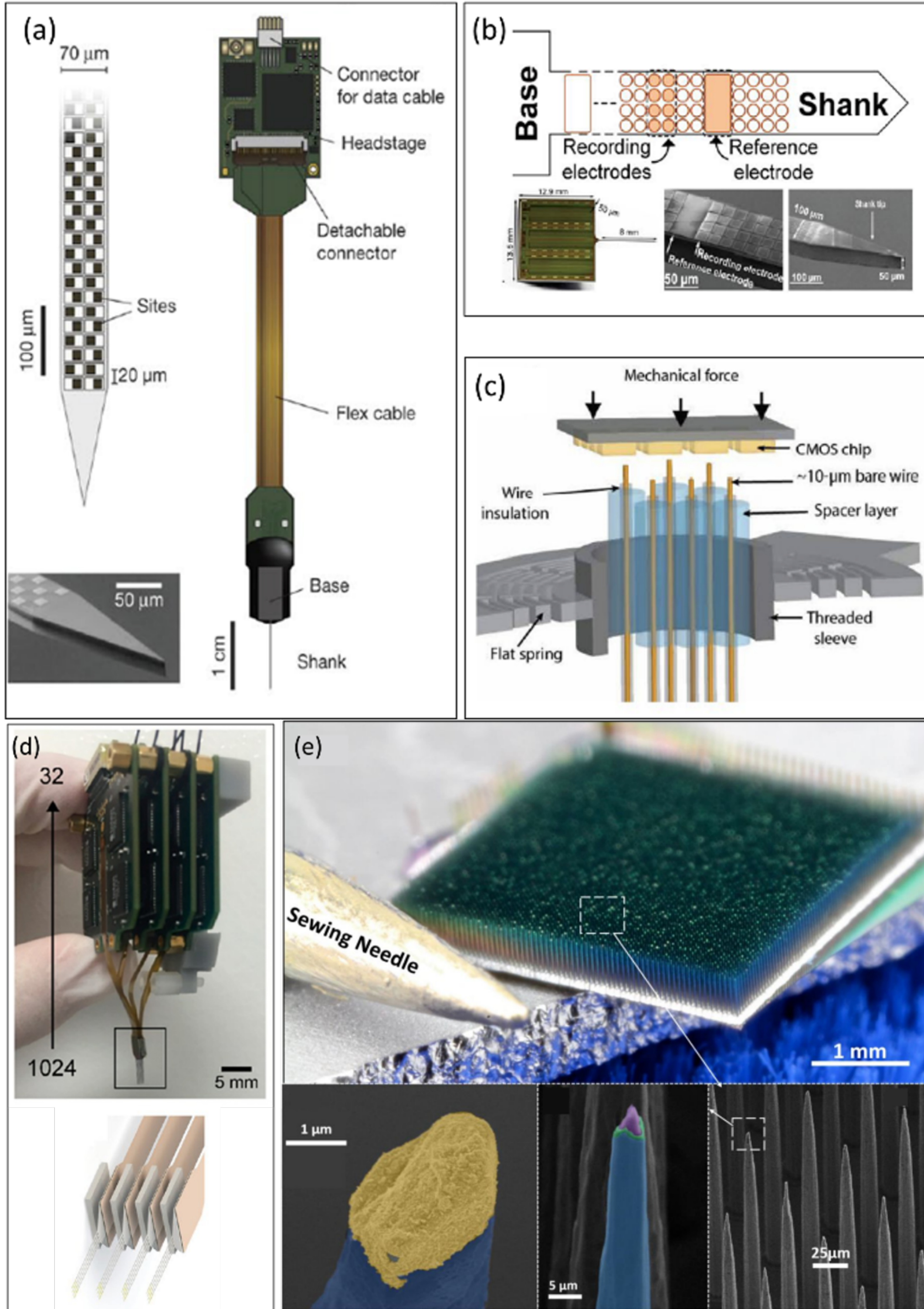

Figure 1.4: Neural probes for high-density and large-scale recordings, (a) Neuropixels [37, 39], (b) NeuroSeeker [38], (c) High-density microwire bundle connected to a CMOS amplifier array [33], (d) Nanofabricated neural probes [32], (e) Michigan's Sea of Electrode Arrays [23].

The Neuropixels 1.0 and Neuropixels 2.0 probes are notable examples of high-count planar probes, featuring 960 and 1280 recording sites respectively [37, 39]. These probes are fabricated using a custom 130 nm CMOS process on either single or multiple shanks. Each recording site measures 12 µm × 12 µm and is spaced at a center-to-center distance of 25 µm in Neuropixels 1.0 and 15 µm in Neuropixels 2.0 (Figure 1.4 (a)).

The shanks have dimensions of 10mm in length, 70 µm in width, and a thickness of 24µm. To facilitate their use, the probes are integrated with a flexible cable measuring 43.5mm in length, 4.0mm in width, and 80 µm in thickness. The multi-shank Neuropixels 2.0 probes are equipped with over 5000 recording microelectrodes.

These probes have been specifically designed to be miniaturized, enabling efficient recording of neuronal activity on a large scale across various regions of the rodent brain. NeuroSeeker probes [38], similar to Neuropixels probes, employ CMOS and MEMS technology and provide simultaneous readouts from 1344 electrodes.

They have the capability to record from over 1000 neurons across diverse brain regions in mice (Figure 1.4 (b)). Data analysis validation and probe localization present significant challenges when dealing with large-scale probes such as Neuropixels and Neuroseeker probes. Additionally, the physical size of Neuropixels probes may limit their application in studies involving smaller animal models or brain regions with limited accessibility.

In a study conducted by Schaefer et al., they devised a novel approach to connect a high-density microwire bundle directly to a CMOS amplifier array (Figure 1.4 (c)) [33]. The microwire bundle consisted of polymer-insulated metal microwires with exposed recording sites at their distal ends and interconnecting tips at their proximal ends. Electrical contacts were established by mechanically pressing the proximal interconnecting tips onto a CMOS amplifier array.

Additionally, the size of the metal microwires was reduced to a diameter range of 5–25 μm to minimize potential damage to brain tissues. This configuration allowed for the simultaneous recording of neural activity from hundreds of neurons in the motor cortex of mice in an in vivo setting. Nevertheless, a significant challenge arises due to the unpredictable arrangement of microwires within each CMOS-bundle device. This poses difficulties in the development of spike sorting algorithms that rely on positional information to accurately distinguish and classify neural activity.

In another study, Rios, et. al. proposed a high-density and large-count out-of-plane probe by assembling 16 2D planar probes, each featuring 64 channels and fabricated using standard CMOS technology (Figure 1.4 (d)) [32]. These planar probes were grouped together to create a 3D array with dimensions of 4 probes × 4 probes × 64 electrodes. However, it is important to note that this method requires manual assembly of the individual components to create the 3D array.

Michigan's Sea of Electrode Array (SEA) is another notable example of high-density and large-count 3D arrays [23]. This innovative technology employs deep reactive ion etching (DRIE) to create ultra-high aspect-ratio holes in a silicon substrate. These holes are then refilled with multiple thin films resulting in the formation of numerous individual needles when released from the substrate. The needles vary in length from approximately 0.5 to 1.2 millimeters, have widths at the base of 20μm, and sharp tips of <1μm (Figure 1.4 (e) [23]).

While significant progress has been made in the development of high-density and high-count neural probes, the research in this field is still ongoing, emphasizing the need for further advancements in both planar and out-of-plane probe designs and fabrication technologies. It is crucial to continue exploring and developing new techniques to create customizable and engineered probes that can cater to specific applications and requirements.

A primary focus of this thesis is the development of techniques aimed at enhancing the customizability of 2D and 3D neural probes for specific applications, particularly for hard-to-access regions within the brain. These techniques are proposed to improve the functionality of both planar and highly non-planar probes by engineering their mechanical and electrical properties, enabling more precise and effective exploration of targeted brain regions.

1.3 Considerations to Minimize Tissue Damage

Conventional neural probes are typically rigid, with large cross-sections and elastic moduli in the range of hundreds of GPa. Brain tissues are soft, with an elastic modulus in the range of kPa. Following the implantation of a neural probe into the brain, the insertion site undergoes continuous damage due to the brain's subtle movements, referred to as micromotion. These movements are induced by variations in blood pressure resulting from cardiac pulsation and breathing.

It has been shown that the density difference between the tissue and the neural probe influences the extent of glial scarring [40]. Notably, hollow neural probes exhibited a significantly lower presence of astrocytes and microglia in their vicinity.

Moreover, a different research group developed a finite element model to investigate the microenvironment between the electrode and tissue, simulating the impact of micromotion on interfacial strains [41]. This model revealed elevated strains at both the edges and the tip of the neural probe. Given that the Young's modulus of the brain is in the order of kilopascals, these findings suggest the potential use of soft materials for neural probes to minimize the mechanical impedance mismatch [41].

Another study showcased a substantial reduction in the long-term chronic neuroinflammatory response through the use of a mechanically compliant implant [42]. Long-term neural signal measurements using rigid neural probes with large volumes often suffer from signal

instability and degradation [43]. To tackle this challenge, reducing the bending stiffness of the implantable probes is considered an effective approach. This reduction enhances their mechanical compliance with neural tissues, leading to improved signal quality over time. The bending stiffness of a cylindrical penetrating neural probe can be estimated using the following equation:

$$k = \frac{3\pi d^4}{64 l^3} E_s$$

Where E_s is the elastic modulus, l is the length, and d is the diameter of the probe. To reduce the bending stiffness of the probe, several strategies can be pursued. One approach is to minimize the cross section of the probe while keeping the length constant (d). This can be achieved by using thinner or more compact designs that occupy less space. Additionally, employing materials with lower elastic modulus (E_s) can effectively decrease the bending stiffness of the probe.

Recently, many research groups have been actively exploring the use of materials with lower elastic modulus, particularly polymer-based materials, for the development of neural probes [44-47]. Polymers offer several advantages in terms of their mechanical properties and biocompatibility. However, there can be challenges associated with their implantation in specific regions of the brain. Precise insertion and accurate positioning of the implantable probes are essential for specific interactions with the targeted cells.

A probe's mechanical stiffness influences the handling and precise deep-brain implantation process [48]. Because of the fragility and lack of mechanical stiffness, ultra-flexible probes always tend to bend, buckle, and deflect during insertion into brain [25]. Therefore, the use of more flexible shanks or probes can indeed make them more challenging to target accurately [49]. In a study by Lieber et al. they introduced a neural probe implantation technique using an injection

syringe [50, 51]. They developed a biomimetic neural probe known as neuron-like electronics (NeuE), featuring a bioinspired structure design (Figure 1.5).

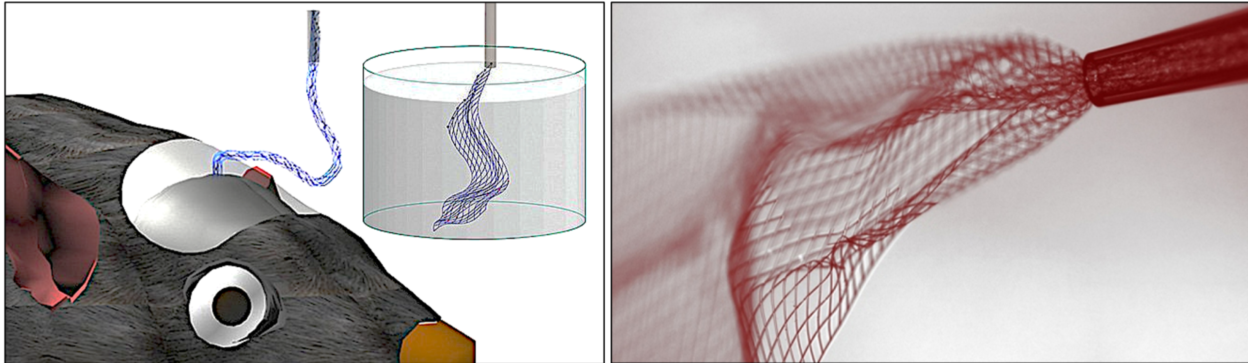

Figure 1.5: Neuron-like electronics (NeuE) neural probe, featuring a biomimetic design inspired by the structure and function of neurons [50, 51]

However, it is important to consider that these mesh-like neural probes, due to their distributed nature, present challenges in terms of removability. As they cover a substantial area within the brain tissue, extracting the entire mesh structure without causing substantial damage or disruption to the surrounding neural tissue becomes a complex task.

In another study Luan et al. introduced an ultra-flexible nanoelectronics thread (NET) probe, utilizing SU-8 as the substrate material (Figure 1.6) [52]. NET probes necessitate implantation into the brain using a rigid and sizable shuttle device. This facilitates accurate placement and positioning of the probes, ensuring precision placement. However, it is important to note that this implantation method can also result in tissue displacement and damage within a considerable area surrounding the implantation site due to the need to use a stiff shuttle.

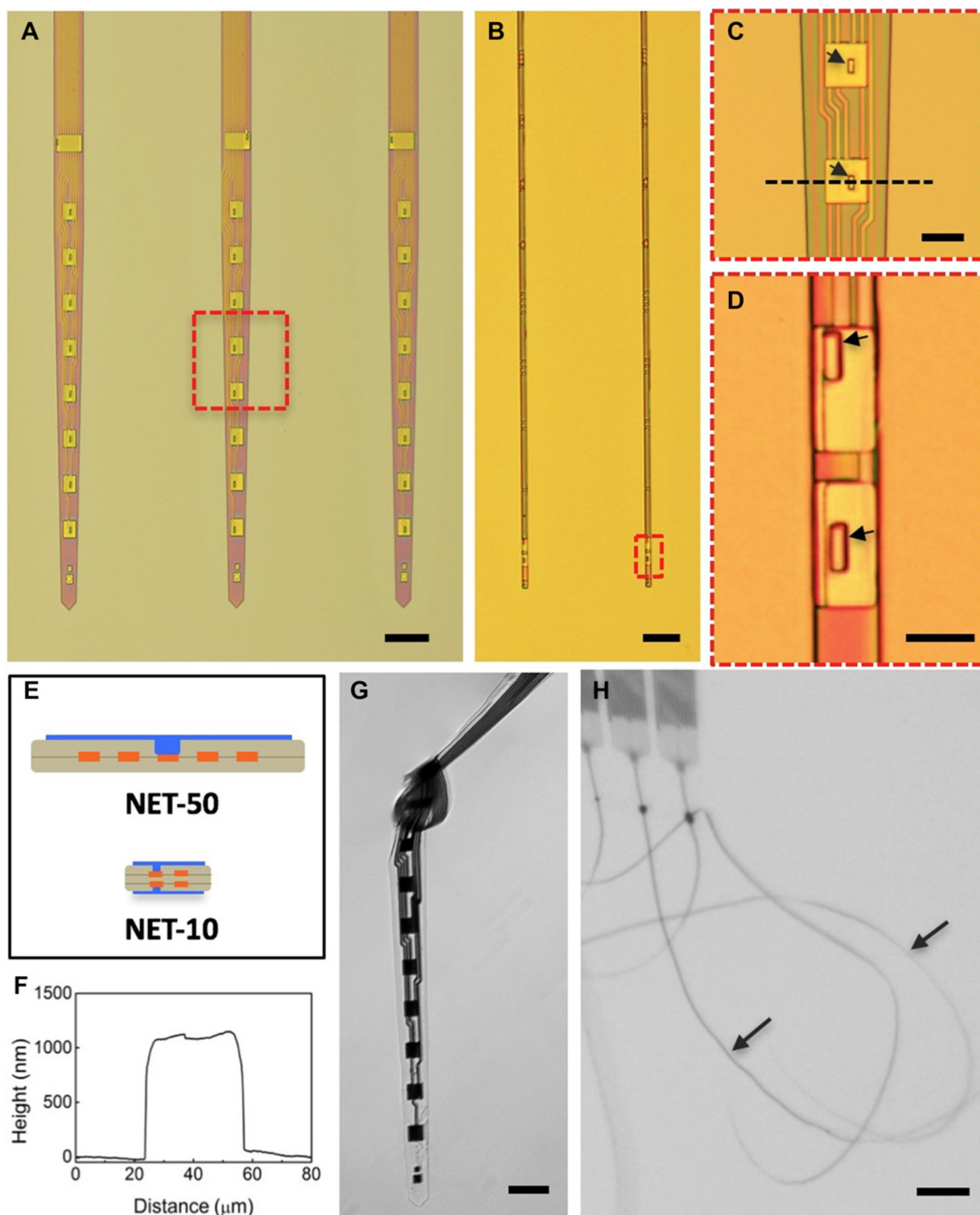

Figure 1.6: Nanoelectronics thread (NET) probe, featuring a SU-8 substrate material [52].

Generally, flexible neural probes require additional guide tools or support structures to aid in their precise implantation. These guide tools are often rigid and can cause larger tissue displacement when inserted into the brain tissue, potentially leading to tissue damage, or disrupting the intended target region. The increased flexibility of the probes can make them less easily controllable during the implantation process, potentially affecting the accuracy of their placement [5, 49].

In some attempts to address this issue, degradable or dissolvable materials have been explored as coating layers for ultra-flexible probes. These materials can provide temporary stiffness during implantation and then gradually degrade or dissolve, allowing the probe to return to a more compliant state within the hydrated neuronal tissue [5]. However, despite these attempts, it is still a challenge to achieve sufficient strengthening of the implantable probes without increasing the thickness of the coating layer. Thicker coatings can indeed lead to a higher risk of tissue displacements and potential damage during implantation [5].

The geometry and structure design of neural probes can significantly impact the development of chronic inflammation. Implanting shanks with larger widths or thickness increases tissue stress, which in turn promotes glial encapsulation and reduces the long-term stability of the implant. In addition to its dimensions, the structural design of the probe plays a crucial role in its compliance with neural tissues [53]. Studies have shown that open-architecture probes promote better integration and interaction with surrounding tissue while minimizing immunogenicity. However, it is important to note that extracting these probes after long-term implantation can be challenging and may potentially cause tissue damage [54, 55].

When designing neural probes, it is crucial to strike the right balance between mechanical stiffness and compliance. If the probe is too compliant, it cannot be easily inserted unless an

insertion tool is utilized, and if the probe is too stiff it could cause long-term tissue damage. This stiffness-compliance tradeoff might also be different for different tissues and applications. Therefore, the "optimal" stiffness/compliance ensures precise insertion and minimizes overall probe volume, which reduces tissue displacement and interference with blood vessels.

The stiffness of the shanks must be sufficient for accurate targeting and placement, and small to reduce the risk of unintended harm to surrounding tissue. Simultaneously, the optimization of the probe's geometry and structure is essential to enhance mechanical compliance with the soft brain tissue. By minimizing the shank volume and cross section, the extent of tissue displacement caused by the probe can be reduced.

Another main contribution of this thesis is developing a probe structure and fabrication technology that allow the engineering of the mechanical properties of neural probes to establish a balance between stiffness/compliance and volume/cross-section. The aim is to enhance targetability while minimizing tissue damage. This is accomplished by optimizing the geometric features of planar 2D multichannel and 3D non-planar silicon-based neural probes.

1.4 Tip Metallization Techniques in Large Count 3D Arrays

In the fabrication of 3D neural probes, the formation of recording/stimulation sites requires selective removal of insulation layers covering the conductors connecting to these electrodes is a vital process. These insulation layers are typically applied to protect the electrodes and prevent unwanted electrical interference. However, to establish direct contact with the neural tissue, it is necessary to remove these insulation layers selectively.

In large-scale and high-density neural probes, where numerous electrodes are closely packed together, the selective removal of insulation layers becomes even more critical. Precise and controlled de-insulation allows for the creation of active electrode sites at specific locations,

enhancing spatial resolution and enabling simultaneous recording from multiple neurons. This capability is invaluable in advancing our understanding of neural networks and complex brain functions.

Methods for de-insulating and metallizing needle tips in 3D arrays have undergone multiple improvements over the past two decades (Figure 1.7). Early UEA de-insulation methods used heat to melt the insulation, or high voltage discharges to ablate it [56]. The heating method caused significant damage to the insulation layer adjacent to the exposed tip.

The high voltage discharge method degraded the adhesion of the Parylene near the tip, resulting in tiny fractures along the needle. As a result, electrode impedance decreased, resulting in reduced encapsulation lifetime, increased crosstalk between channels, and reduced selectivity. Another de-insulation technique uses diffusion-limited deposition (DLD) of Parylene to form unencapsulated electrodes [57].

During this process, the needle-shaped microelectrodes are placed in a specially-designed container with perforated lids, forming a gradient of precursor monomers between the interior and exterior of the container as Parylene is deposited. Since this method requires specific containers for different shapes and arrangements of electrodes, it is not suitable for tip de-insulation in 3D arrays to obtain highly uniform impedance values and a controllable de-insulated region.

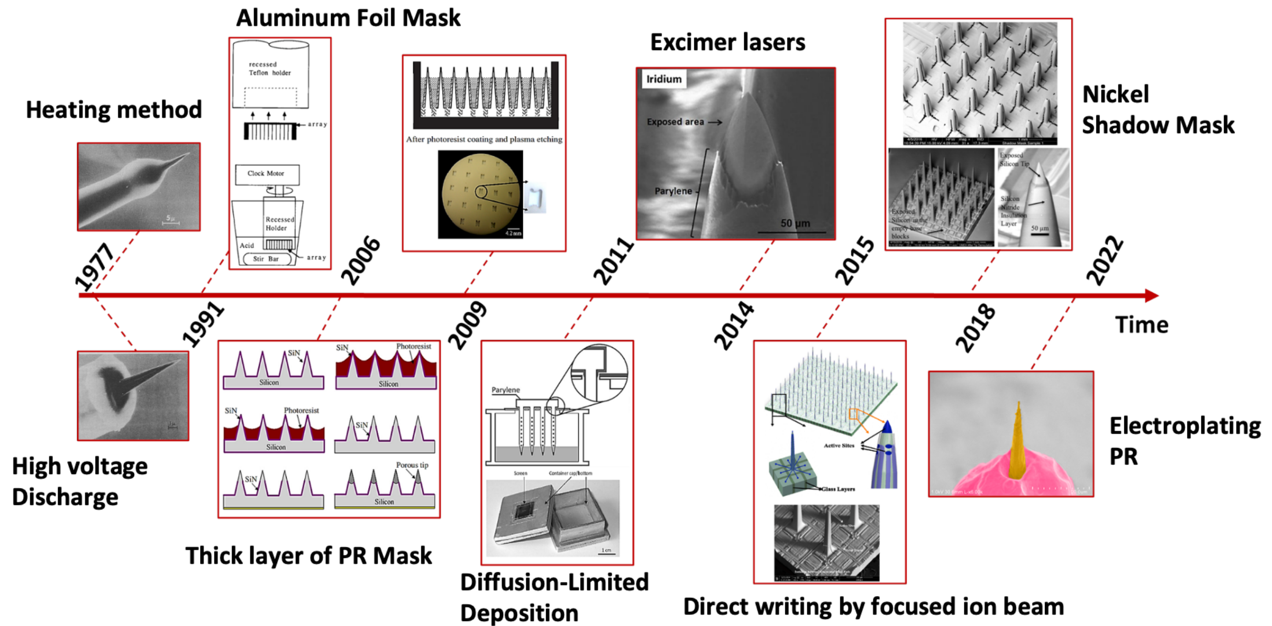

Figure 1.7: Evolution of methods for de-insulating and metallizing needle tips in 3D arrays over the past two decades [56-65].

Oxygen plasma and wet etching have also been used for UEA tip exposure with photoresist and aluminum foil masks [58-60]. The de-insulation of UEA tips through photoresist masks requires fabricating Teflon holders to hold individual electrode arrays within recesses. After the electrode arrays are placed in the holder, thick layers of a photoresist are spun onto them, then vacuumed and baked for a long period of time. However, the long-term baking and acetone removal of thick layers of the photoresist mask degrades the properties of Parylene-C. In addition, Teflon holders must be customized for different array sizes [60].

Meanwhile, aluminum foil masking and poking the electrode tips at the desired exposure length to de-insulate the Parylene layer has become a regular masking technique in UEA technology. Poking the tips through aluminum foil is a tedious procedure as well as challenging when it comes to controlling the tip size, which results in large variations in impedance values due to nonuniformity. Furthermore, foil masks are not suitable for electrode geometries with variable

heights [58]. Excimer lasers were later used to remove Parylene from UEA tips, but this method is relatively slow and expensive because each tip must be etched individually [61]. Direct writing by focused ion beam (FIB) patterning and shadow mask processing [63] have been used recently to create metallic patterns on individual needles in UEAs to expose multiple recording sites on each shaft. However, FIB patterning, coupled with high-resolution scanning electron microscopy (SEM), requires writing each active site on each needle individually, which is both time-consuming and expensive [62].

Shadow masks address some of the FIB challenges, but they must be made from nickel by electroplating and laser ablation on sacrificial UEAs. To fabricate nickel shadow masks, however, extra steps must be taken beyond conventional UEA fabrication. It is additionally time-consuming to make special shadow masks for batches with different electrode array shapes [63].

To overcome the limitations of the existing methods for de-insulating out-of-plane electrode arrays, we have developed a novel alternative method using electroplated photoresist (EP) as a thin conformal mask layer [64, 65] to expose the tips of needles in a 3D probe. In this technique, the time required for baking and stripping the photoresist is relatively short, reducing the risk of damage to the insulator layer. Furthermore, there is no need to fabricate shadow masks or holders, which makes it suitable for electrode arrays of any length, density, or distribution. This approach is effective for arrays with shanks of uniform size, as well as arrays that incorporate shanks of different sizes within the same array [66].

Furthermore, we have introduced two innovative methods for microfabricating large-count and high-density highly non-planar probes, featuring techniques for forming and metallizing the tips. Further exploration of these methods will be conducted in the upcoming chapters.

1.5 Packaging and Integration

Packaging neural probes is of utmost importance for recording and stimulating neural activity. The packaging designs are focused on seamlessly integrating the neural probe with electronic components. To achieve this, flexible cables, integrated circuits, amplifiers, and transmitters are incorporated within the packaging to efficiently process and transmit neural signals. Various approaches can be employed to incorporate these electronic parts, such as directly integrating them onto the interface device [67] or utilizing head-mounted systems [68, 69].

For many years, integrated flexible cables have played a significant role in facilitating access to the recording sites on the neural probe and simplifying its handling during in-vitro and in-vivo tests [70-72]. These cables are constructed using various flexible and biocompatible polymers, such as Polyimide and Parylene-C.

As the number of electrodes in neural probes increases, the need for innovative packaging methods arises to connect to each individual electrode. For instance, relying on wire bonding for each electrode becomes impractical when the number of recording sites increases. Furthermore, an increase in the number of electrodes requires the acquisition of data and stimulation commands through a small number of wires or even wireless channels.

Additionally, active multiplexing is becoming increasingly important as the number of electrodes in neural interfaces grows. This is because relying solely on passive wiring presents significant limitations in terms of reliability, lifespan, and size.

Presently, most neural interfaces employed in chronic experiments operate in a passive manner and depend on extensive wire bundles, resulting in a limited number of channels, typically to a few hundred [73, 74]. To advance the field, it is critical to prioritize the development of platform technologies that seamlessly integrate active electronics with neural arrays. This

advancement is crucial for ensuring high reliability while simultaneously minimizing the size and complexity of the systems. This endeavor holds significant importance, especially in accomplishing the desired objectives of achieving higher channel counts and densities [6, 62, 67, 75].

In this research, we have proposed new ideas for packaging and integration of large-count and high-density neural probes. Our primary goal is to minimize the overall footprint of highly non-planar arrays by eliminating the need for wire bonding, which becomes impractical when dealing with a large number of electrodes. Furthermore, we investigate the integration of a Polyimide-based cables and platforms to facilitate the convenient and cost-efficient addition of connectors or active electronics to the backend of the probes.

1.6 Challenges in Spike Sorting for Large-Count High-Density Neural Interfaces

The advancement of neural devices has enabled simultaneous recordings from a large number of electrodes, presenting new challenges in the analysis of high-density extracellular recordings [76]. Spike sorting, a critical process in extracting neuronal activity from these recordings, becomes increasingly complex due to the increased channel density. The higher likelihood of encountering overlapping spikes in both space and time makes it difficult to resolve spike collisions and complicates the spike sorting process [77].

Furthermore, the exceptional spatial resolution of high-density probes amplifies existing challenges in spike sorting, particularly concerning the issue of drifts, which arise from tissue movement relative to the recording devices. As a result, the development of fully automated and reproducible spike sorting solutions becomes imperative to handle the growing volume of data generated by high-density devices.

Scalability is a significant challenge in spike sorting, as the number of electrodes in neural devices continues to rise. Spike sorting algorithms must adapt to accommodate the increasing number of electrodes. To illustrate, a 1-hour recording from a 384-electrode device can produce over 80 GB of raw data when sampled at 30 kHz, while a 1-hour recording from a 20,000-electrode device can generate a 1 TB file when sampled at approximately 10 kHz [78].

Additionally, achieving reproducibility in spike sorting poses its own set of challenges. Despite the initial use of automatic tools for sorting, manual curation is often necessary to refine the output. Various tools have been developed to assist users in tasks such as removing, merging, and splitting units [78-81]. However, this manual curation process is subjective and can lead to inconsistent outcomes. Consequently, the lack of consistency in manual curation degrades reproducibility and subsequent analyses that rely on spike sorting [82, 83].

To evaluate the reliability of sorted units, unsupervised quality metrics are employed, considering both known biophysical properties and data-driven measurements, such as the unit's isolation in the feature space [84, 85]. Although several metrics have been proposed, there is currently no consensus within the scientific community regarding which metrics are informative and should be utilized to automatically refine spike sorting outputs. Nonetheless, the availability of ground truth and curated datasets emphasizes the importance of exploring and validating these metrics for spike sorting developers and end-users [86].

This research has also investigated computational approaches to enhance the reliability of quality metrics such as L-ratio and presents a dynamic model to improve the accuracy and precision of this metric in assessing spike sorting outputs specifically tailored for large count and high-density neural interfaces.

1.7 Research Objectives and Thesis Contributions

The main objective of this research is to develop technologies for fabricating advanced 2D and 3D multichannel, high-density, and large count penetrating neural interfaces. These interfaces can be specifically designed for targeted areas of the brain, enabling more advanced neuroscience research and precise measurements. The proposed structural designs and fabrication technologies will provide the end user with a much greater range of mechanical and structural features that will potentially improve neural recording quality and longevity of the probes.

This work offers significant specific contributions in the following areas:

1. Development of high-density and large-count neural interfaces in both planar 2D and non-planar 3D configurations with customizable features:
    a. A new class of large count (> 64 recording sites), high-density (6µm vertical separation and 10.5µm horizontal separation between recording sites), multiple shanks (at least 4-shanks), silicon-based 2D planar neural interfaces.
    b. A new class of 3D non-planar (500µm height) silicon-based neural interfaces capable of accommodating a large number of recording sites (100×100) with design versatility enabling customized array size, density, and distribution.
2. Incorporation of mechanically and electrically engineered features to ensure minimal invasiveness and more precise implantation:
    a. Engineered T-bar and π-bar shaped cross sections of the silicon-based 2D planar probes to achieve sufficient mechanical stiffness and yet reduce the volume and cross-sectional area of each shank.

   b. Microfabricated 3D non-planar silicon-based needles with a diameter as small as 10-15µm and sharp tips.
   c. Technologies for forming precisely defined recording/stimulating sites on extremely high-density and large count (100×100) 3D non-planar probes using electroplated photoresist and reliable 1-sided and 2-sided techniques.
3. Packaging and integration innovations to minimize the footprint of the package/connector and eliminate wire bonding, facilitating the integration of active or passive electronics with neural devices:
   a. Innovative integrated packaging and assembly techniques for flexible and robust connections to 3D non-planar neural interfaces with a large number of recording sites and high density.
   b. Parylene-C based integrated flexible cables for easier handling of the 2D probes.
4. Computational techniques for processing vast data from high count high-density neural interfaces:
   a. A new Python-based toolbox to compute a dynamic version of the L-ratio, which is a quality metric in spike-sorting, tailored for large count and high-density neural interfaces, providing much improved reliability compared to existing techniques.

A summary of the main challenges we aimed to address and their proposed solutions in this thesis is illustrated in Figure 1.8.

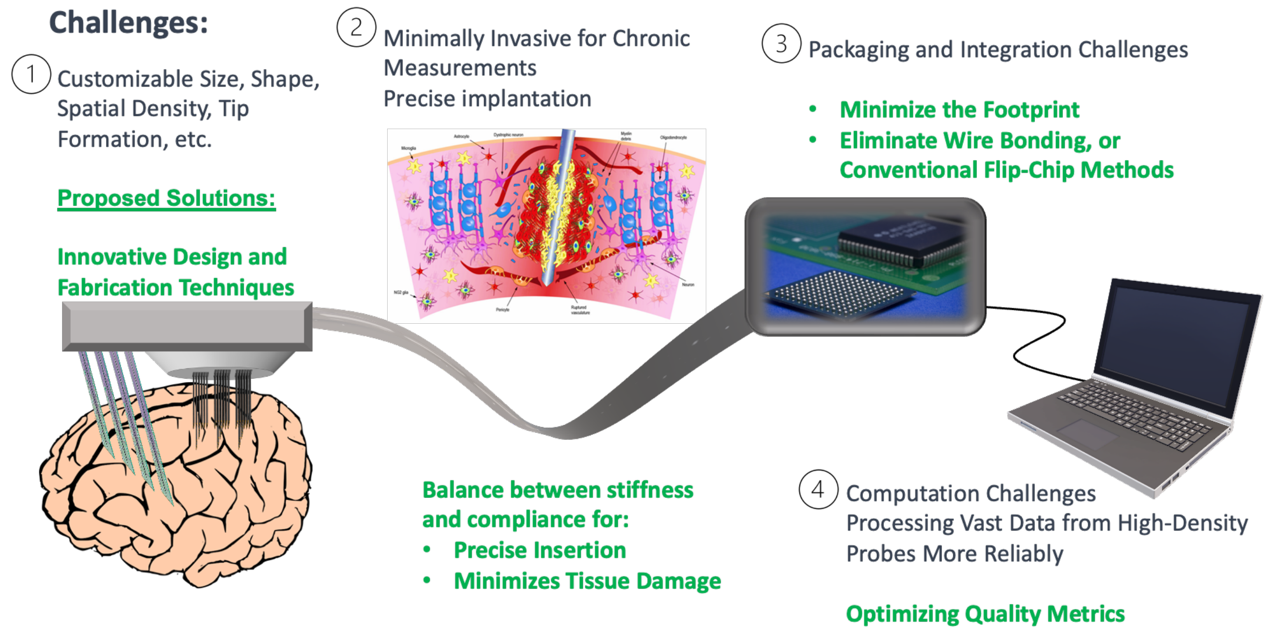

Figure 1.8: Summary of main challenges and proposed solutions in this thesis.

Moreover, comprehensive electrical and mechanical characterizations, along with preliminary in vitro tests of electrode impedances for stimulation and recording, have been successfully conducted on the fabricated probes and validated by the modeled equivalent circuits. Additionally, chronic (for almost a year) impedance measurement tests have been performed in vitro on planar probes to assess the durability of the recording sites.

1.8 Organization of the Thesis

Chapter 1 provides a literature review, outlines the motivation behind the thesis, research objectives, and the contributions it makes to the field.

Chapter 2 introduces a new class of large-count, high-density, multi-shank 2D planar silicon probes engineered to be more compatible for long-term measurements. These probes are fabricated with integrated flexible cables for easier handling and feature customizable characteristics such as size, tip shape, number of electrodes, and number of shanks. Comprehensive

electrical and mechanical characterizations, along with in-vitro tests, have been successfully conducted on the fabricated probes and validated by the modeled equivalent circuits. Additionally, chronic impedance measurement tests have been performed in vitro on planar probes for nearly a year to assess the durability of the recording sites.

Chapter 3 presents new techniques for fabricating ultra-high-count, high-density non-planar 3D silicon probes with technological solutions for the packaging and tip metallization of out-of-plane neural interfaces. Comprehensive electrical and mechanical characterizations, along with in-vitro tests, have been successfully conducted on the fabricated probes and validated by the modeled equivalent circuits.

Chapter 4 introduces a Python Toolbox designed to calculate the dynamic L-ratio; an evaluation metric that provides more temporal information compared to the classic L-ratio. The new information provided by dynamic L-ratio is used to train a neural network learning algorithm in a supervised learning process. Subsequently, the trained algorithm is employed to automatically curate the sorted clusters.

Chapter 5 concludes the proposed work and proposes future steps in this research.

Chapter 2 High-Density Multi-Shank 2D Planar Silicon Probes for Precise Recording with Reduced Tissue Damage

2.1 Introduction

Neural probe technology plays a pivotal role in understanding the complexities of the brain and developing therapeutic interventions for neurological disorders. Planar neural probes, such as the Michigan-type silicon probe, have garnered significant attention in the field of neuroscience due to their ability to incorporate multiple recording sites within a single probe [37]. However, a key challenge arises when attempting to maintain a small shank width while increasing the number of recording sites.

The incorporation of complementary metal-oxide-semiconductor (CMOS) technology has been instrumental in overcoming this challenge by enabling the integration of recording electrodes with multiplexing circuits, resulting in compact input/output connections. This technological advancement has facilitated the design and fabrication of large-scale neural recording electrodes such as Neuropixels [37, 39] and NeuroSeeker [38].

Neuropixels 1.0 and Neuropixels 2.0 probes are notable examples of high-count planar probes, featuring 960 and 1280 recording sites, respectively [37, 39]. These probes are fabricated using a custom 130 nm CMOS process on either single or multiple shanks. Each recording site measures 12μm × 12μm and is spaced at a center-to-center distance of 25μm in Neuropixels 1.0 and 15 μm in Neuropixels 2.0.

The shanks have dimensions of 10mm in length, 70μm in width, and a thickness of 24μm. To facilitate their use, the probes are integrated with a flexible cable measuring 43.5mm in length,

4.0mm in width, and 80μm in thickness. The multi-shank Neuropixels 2.0 probes are equipped with over 5000 recording microelectrodes, making them ideal for efficiently recording neuronal activity on a large scale across various regions of the rodent brain [37, 39].

NeuroSeeker probes, similar to Neuropixels probes, employ CMOS and MEMS technology and provide simultaneous readouts from 1344 electrodes [38]. They have the remarkable capability to record from over 1000 neurons across diverse brain regions in mice. However, the use of large-scale probes such as Neuropixels and NeuroSeeker presents significant challenges in terms of data analysis validation and probe localization. Additionally, the physical size of Neuropixels probes may limit their application in studies involving smaller animal models or brain regions with limited accessibility [46].

This chapter highlights advancements in the field of planar or two-dimensional (2D) neural probe technology, specifically focusing on a new class of high-density and large-count multi-shank planar silicon probes with engineered mechanical and electrical properties.

The first aspect of this innovation involves the development of fabrication technologies that allow for the creation of probes with easily engineered and user-defined mechanical and electrical properties. This adaptability ensures optimal performance of the probes in different tissue environments. Furthermore, these novel probes offer a significant advantage by optimizing their electrical and mechanical properties without requiring ultra-thick shanks or sophisticated process steps. This reduces the complexity and cost of manufacturing while allowing for customization based on specific experimental or clinical requirements. Additionally, the probes are designed for minimal volume and size, thereby reducing potential tissue damage during implantation, and enhancing long-term stability.

The key advancements include microfabrication of super-fine shanks with ultra-high aspect ratios, incorporation of embedded small stiffeners, achieving a high stiffness-to-volume ratio with T and π-shaped cross sections thus producing remarkable straightness upon releasing the shanks, and the integration of flexible interconnect cables. These innovations aim to minimize tissue displacement and reduce the number of invaded blood vessels, surpassing the effects observed with thicker shanks.

Our approach enables the achievement of remarkably small cross sections, approximately ~170 μm^2, allowing customization of the probes for specific applications while minimizing tissue damage and displacement caused by the probe. The design of these probes is intended to enhance the precision of implantation in the neural tissue while also being flexible enough to reduce damage to tissue and vascular networks while being implanted.

We have successfully fabricated 64-electrode 4-shank probes using this approach, which exhibit excellent mechanical and electrical properties. The functionality of these probes is assessed through in-vitro experiments in the laboratory. This involves introducing artificial electrical signals, similar to neural spikes, into a solution of Phosphate Buffered Saline (PBS), and then recording the responses from the probes' recording sites. Furthermore, the durability of the recording sites on these probes is assessed by immersing them in a PBS solution for nearly a year (11 months), during which the site impedances are measured over time.

Overall, the development of this new class of high-density and large-count multi-shank planar silicon probes presents a significant advancement in neural probe technology. The combination of engineered properties, user customization, and optimized electrical and mechanical characteristics provides researchers and clinicians with powerful tools for studying and interacting

with neural circuits in various tissue contexts, potentially leading to further advancements in understanding the brain and developing innovative treatments for neurological disorders.

2.2 Optimization of Mechanical Properties for Reduced Tissue Damage

In this section, we will delve into the crucial design considerations for optimizing the mechanical properties and probe size to reduce tissue damage during implantation. One key factor to focus on is the cross-sectional area of the shank, as it directly influences its stiffness. When comparing shanks with the same width, thicker shanks are stiffer and less susceptible to buckling during insertion.

While this increased stiffness may initially appear advantageous, it is important to consider the potential drawbacks. Thicker shanks have a larger volume, which leads to increased tissue displacement and damage during insertion. Hence, achieving the right balance between shank thickness and volume is crucial to ensure that the probe is sufficiently stiff for easy insertion while being small enough to minimize the risk of tissue damage. Figure 2.1 depicts the principal concept of crafting planar shanks to achieve an optimal equilibrium between their volume and stiffness.

The figure illustrates four different planar shank styles. The thick shank has the largest cross-sectional area, causing significant tissue volume displacement and high tissue damage. In contrast, the thin shank has the smallest cross-sectional area, resulting in reduced damage to the tissue and vascular network when implanted in brain tissue. However, for a thin shank probe with a high enough aspect ratio, the stiffness of the shank may not be sufficient for insertion into the targeted region, leading to buckling or breaking.

To strike a balance between these two extremes, we propose the T-shaped and π -shaped cross-sectional shanks and compare them in the following sections. The T-bar shank features a T-shaped cross-section, with a much lower volume compared to the thick shank. However, the

addition of one stiffener provides enough stiffness for insertion into the tissue without buckling, while still minimizing tissue damage and volume displacement.

The π -bar shank also has a π -shaped cross-section, similar to the T-bar shank but with two stiffeners instead of one. The additional stiffeners increase the stiffness of this shank, making it more robust, yet the volume remains much lower than that of the thick shank.

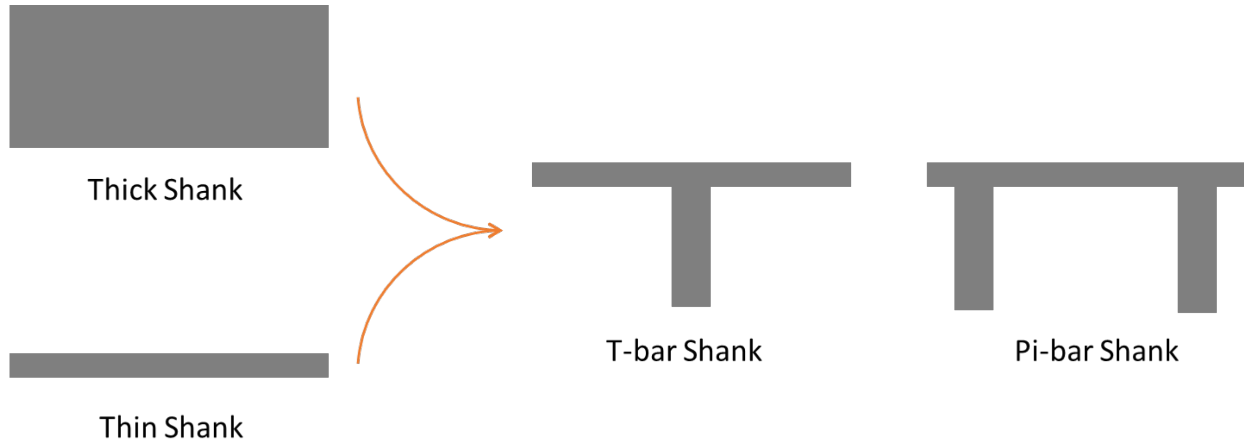

Figure 2.1: Principal concept of crafting T-bar shaped and π-bar shaped cross sections for planar shanks to achieve an optimal balance between their volume and stiffness.

In our probe design study, we utilized COMSOL Multiphysics software to simulate and analyze the effects of different parameters, such as the number of stiffeners and shank thickness, on the mechanical properties, such as spring constant. The primary aim of our investigation was to determine the optimal design configuration that maximizes the stiffness to volume ratio while minimizing damage to blood vessels during and following insertion.

Figure 2.2 summarizes the outcome of a series of simulations conducted on several different shank structures. We conducted a simulated experiment applying a uniform downward force on the tip of shanks with different cross-sections: thin, T-bar, π-bar, and thick. Notably, each shank exhibited distinct displacement characteristics. This observation prompted us to investigate the interplay between the shank cross-section, the number of stiffeners, and mechanical properties.

The accompanying table in Figure 2.2 provides details on the size and volume of each probe (first and fourth columns, respectively) and the corresponding mechanical spring constant obtained through FEM simulation (second column).

Our simulation findings demonstrated the impact of incorporating stiffeners and increasing shank thickness on probe mechanical properties. For instance, the addition of a small stiffener to the thin bar, resulting in a T-shaped cross-section, led to a 17.5-fold increase in stiffness with a minimal volume increase of 1.31 times. Similarly, the introduction of another stiffener and transitioning from a thin shank to a π-shaped cross-section yielded a 27.5-fold increase in stiffness while only increasing the volume by a factor of 1.62.

Furthermore, we investigated the effect of transitioning from a thin to a thick cross-section. As expected, our simulation results confirmed a substantial 75-fold increase in stiffness. However, it should be noted that this improvement came at the expense of a significant volume increase by a factor of 4.34.

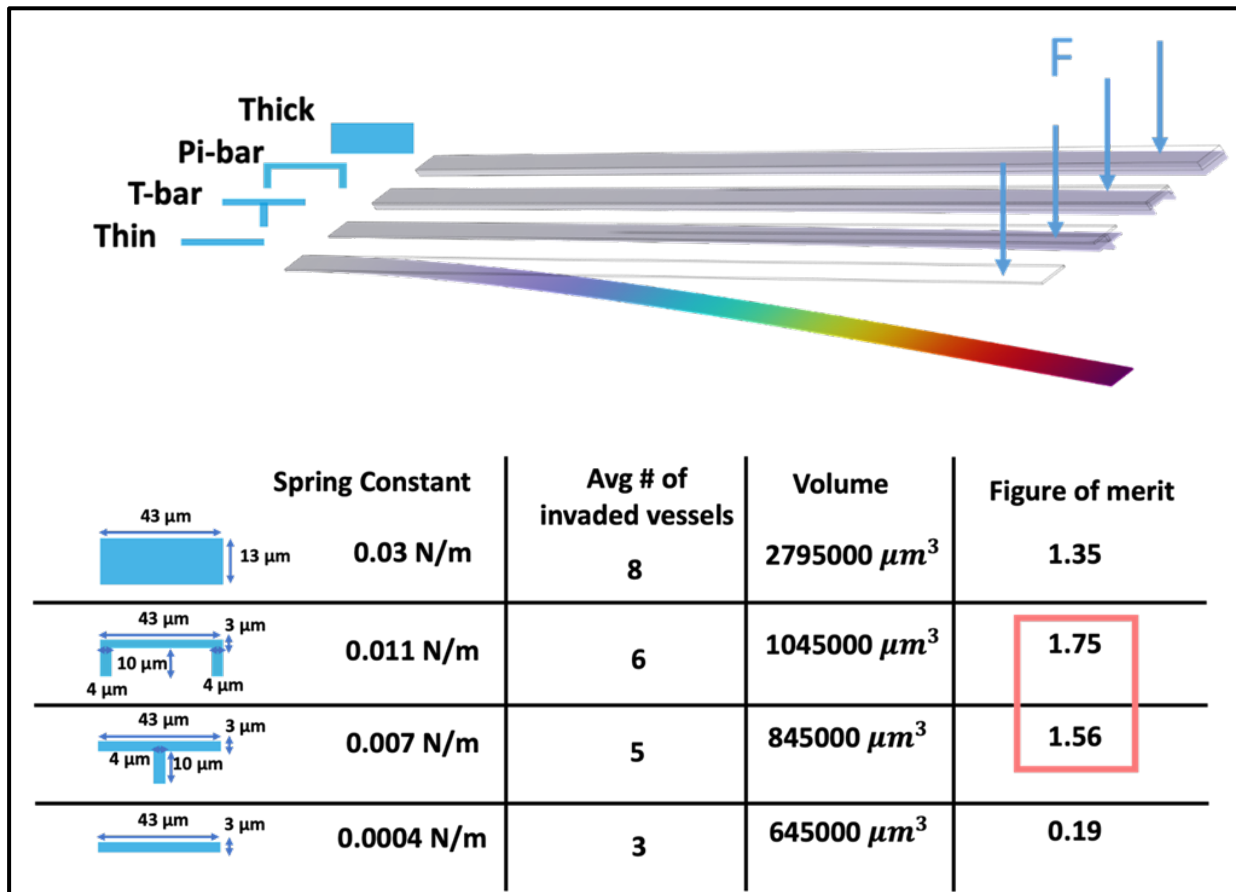

Figure 2.2: FEM modeling to show the influence of stiffeners, shank thickness, and cross-sectional modifications on probe mechanical properties.

To conduct a comprehensive comparison of the different probe designs, we took into account not only their mechanical properties but also their potential impact on the vascular network within the brain. To achieve this, we utilized a MATLAB model representing a rat's brain vascular network and placed each probe design into multiple arbitrary locations.

To ensure accuracy, we employed open-source three-dimensional micro-vascular network data specific to rat brains, allowing us to simulate the invasion of blood vessels resulting from the implantation of various probe geometries as it is shown in Figure 2.3. The model operates under the assumption that the device contour damages any vessel when it traverses the entire width of a

segment. This method allows for the evaluation of the different geometry's performance through a quantitative analysis of blood vessel invasion.

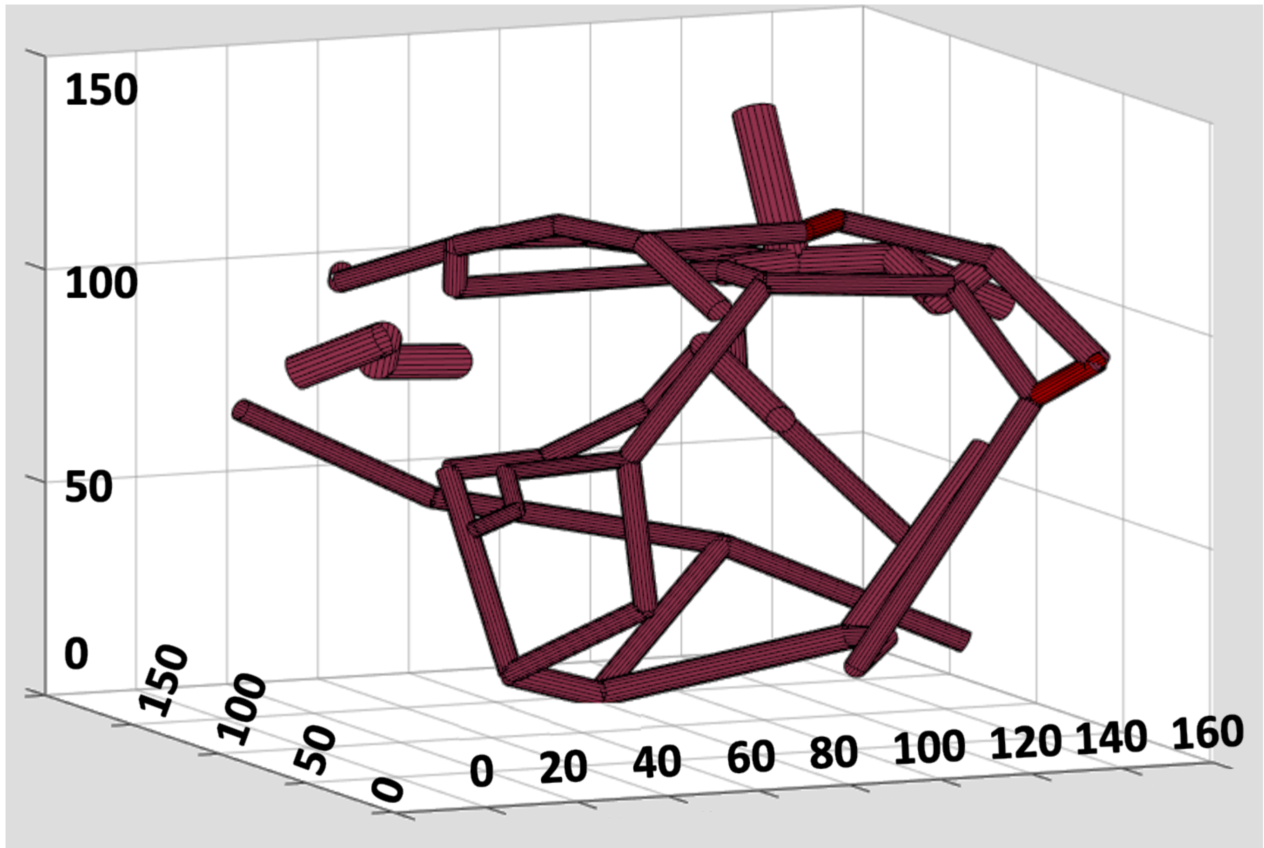

Figure 2.3: A publicly accessible dataset of three-dimensional microvascular networks in the rat brain was employed to simulate the invasion of blood vessels caused by the insertion of different geometries.

To evaluate the extent of blood vessel invasion, we counted the number of blood vessels that came into contact with and were invaded by each implanted probe. As the blood vessels were represented as cylinders in our model, we counted those vessels that intersected with the implanted shank, which was modeled in three dimensions. Figure 2.4 provides a visual representation of the 3D modeling of the implanted shank and its interaction with the blood vessels.

To ensure the accuracy and reliability of our results, the number of microvascular segments intersected by the shanks' footprints was counted over 1000 insertion iterations, with randomized

angles and locations for each geometry, and this information was inserted in the third column of the table within Figure 2.2.

Our findings show that, while increasing the volume and thickness of the probe design significantly improves its stiffness, it also leads to a larger number of invaded vessels. In contrast, the T-bar and π-bar designs exhibit a smaller number of invaded vessels while maintaining a reasonable stiffness-to-volume ratio. Therefore, a careful balance between mechanical properties and the potential impact on the brain's vascular network should be considered when selecting an optimal probe design.

To compare these various probe designs, we established a figure of merit (FOM) defined as the ratio of the shank spring constant to the product of the average number of invaded blood vessels and the shank volume, as shown below:

$$\text{FOM} = \frac{\text{Shank spring constant}}{\text{Avg \# of invaded vessels } \times \text{ Shank volume}}$$

This metric allowed us to assess the trade-off between stiffness and invasiveness for each design. The results of our analysis are presented in the fifth column of the table shown in Figure 2.2.

Our analysis revealed that the π-bar and T-bar designs had the first and second highest figure of merit, respectively. This finding suggests that these designs may be optimal for the targeted application, as they can effectively achieve the desired mechanical properties to reach the targeted region in the brain precisely without causing significant harm to the surrounding tissue. By considering not only mechanical properties but also invasiveness, our approach offers a more comprehensive evaluation of the performance of different designs, leading to a more informed and effective selection of the optimal design for the targeted application.

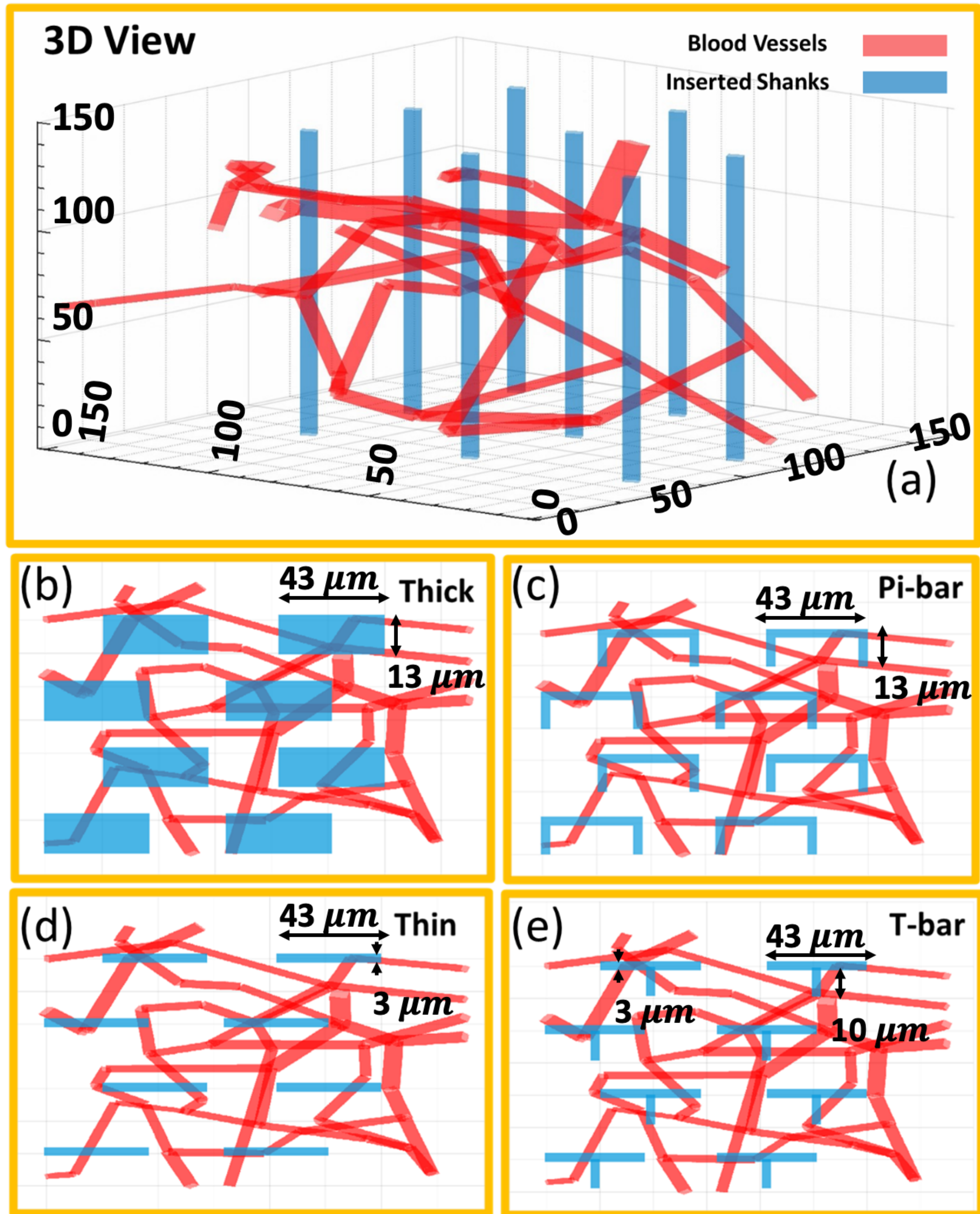

Figure 2.4: (a) A visual representation of the 3D modeling of the implanted shank and its interaction with the blood vessels, (b) thick shank, (c) π-bar shaped shank, (d) thin shank, (e) T-bar shank.

2.3 Design and Microfabrication Process

This section focus will be on exploring the design principles, microfabrication techniques, and distinctive features of these planar probes. By delving into these aspects, we gain valuable insights into the exciting possibilities they offer for advancing our understanding of the brain and pushing the boundaries of cutting-edge neuro-technologies.

These probes possess two features that contribute to their superior performance. First, they employ super fine shanks with ultra-high aspect ratios, allowing for precise and delicate operations. Second, they include small, embedded stiffeners, which boost their structural integrity and functionality, features that can be easily modified by the designer without changing the fabrication process.

Shanks with one stiffener have a T-shaped cross-section, while those with two stiffeners have a π-shaped cross-section. Other designs with more stiffeners are also possible. Overall, the design and fabrication of these probes are optimized to achieve high precision and accuracy in implantation while minimizing tissue damage and displacement.

Overall, the proposed T-bar and π-bar shape of these probes exhibits a higher figure of merit for precise insertion and lower tissue damage as discussed previously. The shanks have a thickness of 2.5 μm where conductive traces are not present, 3.5 μm where conductive traces are present, and 13 μm where the stiffeners are located. The width of the shanks is only 43 μm, allowing for 16 electrodes per shank, and the length of the shanks is 5mm.

These probes are fabricated using planar micromachining techniques. Previous silicon probes used shanks that were predominantly machined out of the host silicon substrate. For example, the Michigan probes used deep boron diffusion and the dissolved wafer process to precisely form the probe shanks to support thin film conductors and dielectrics that formed the

recording/stimulation probes [15, 22]. Other probe technologies, such as Neuropixels and many others [37, 39], utilized silicon thinning and etching using DRIE to form shanks that varied in thickness from 20-100µm.

The fabrication process presented here uses the silicon substrate only for supporting films deposited on top of it to form the probe. The probe shank does not need to keep any of the original substrate once it is released. As such, the 2D and 3D probes presented in this thesis are completely fabricated using deposited thin films. This is a significant feature because it removes many hurdles associated with other fabrication processes, such as having to dope the substrate, and it allows these probes to be fabricated completely using standard CMOS processes.

The probe microfabrication process is shown in Figure 2.5 and begins by etching trenches that will later be refilled to create stiffeners. The resulting multi-shank probes consist of four two-dimensional shanks, each of which is equipped with one or two 4µm wide and 10µm deep stiffeners.

These stiffeners, which are approximately the same length as the shanks, provide increased mechanical stiffness while maintaining a low overall volume. After etching trenches using DRIE for the stiffeners in a (100) p-doped Si substrate, a stack of LPCVD silicon dioxide, silicon nitride, silicon dioxide, silicon nitride, and silicon dioxide (ONONO) with thicknesses of 125 nm, 62 nm, 126 nm, 62 nm, and 125 nm, respectively, is deposited resulting in a total insulator layer thickness of 500 nm. These thicknesses are chosen to cancel out the residual stress of these layers on top of each other.

The subsequent layer, also deposited using LPCVD, involves a mechanically robust polysilicon layer with a thickness of 1µm. This layer serves the purpose of refilling the stiffener

trenches and enhancing the mechanical stiffness of the shanks. Following deposition, the layer is precisely patterned using RIE to selectively cover only the shank area.

Continuing with the layering process, the third layer comprises a combination of silicon dioxide, silicon nitride, silicon dioxide, silicon nitride, and silicon dioxide, forming a stack with specific thicknesses of 125 nm, 62 nm, 126 nm, 62 nm, and 125 nm, respectively.

Through LPCVD, this layer, with a total thickness of 500 nm, is meticulously deposited to completely fill the stiffener trench, ensuring structural integrity and uniformity throughout the design. Subsequently, an LPCVD process is employed to deposit a 1μm thick layer of n-doped polysilicon, specifically intended to serve as a conductive layer. Using RIE the conductive n-doped polysilicon is patterned to create conductive traces and recording sites along the shanks.

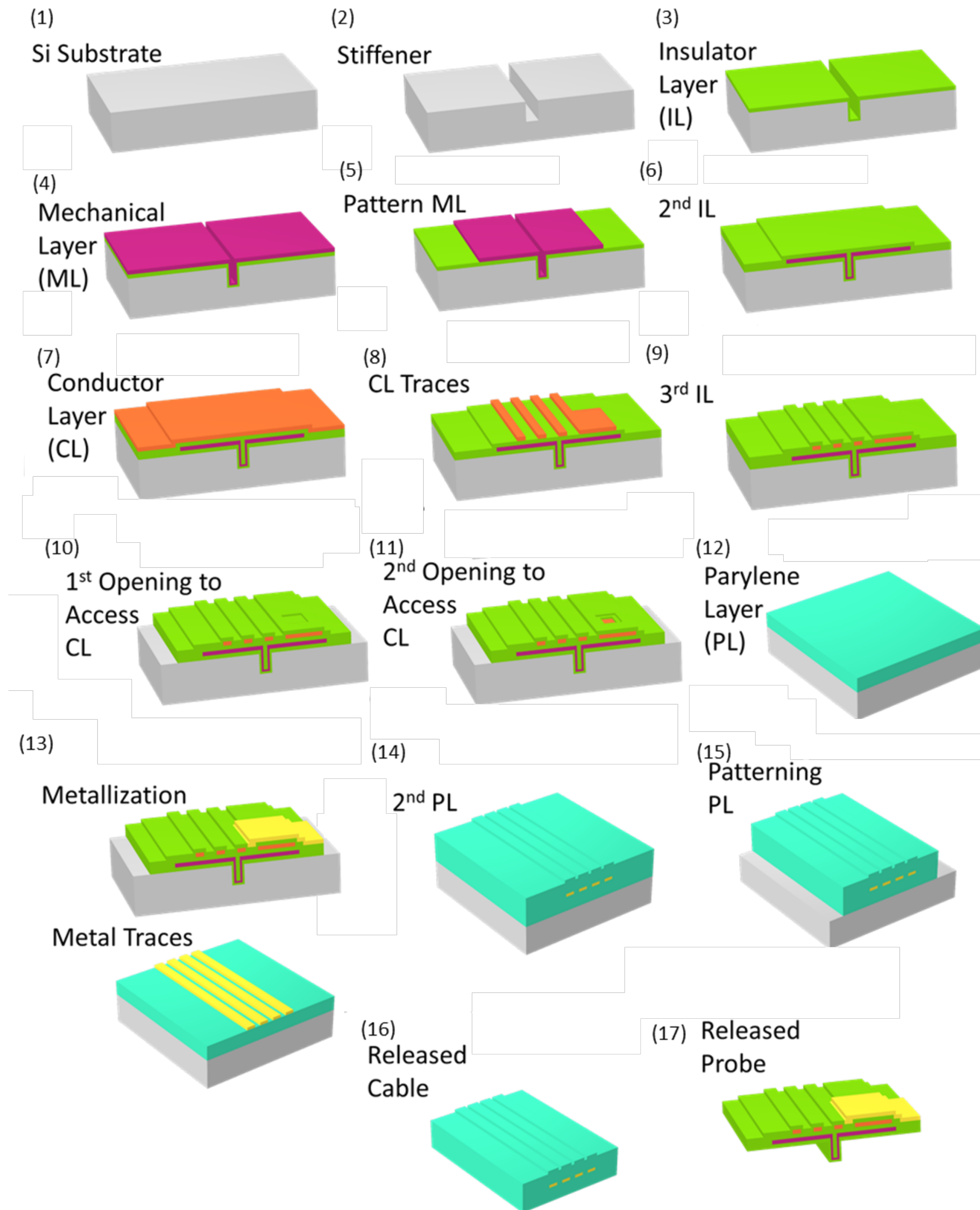

Figure 2.5: Fabrication process of large count high density planar neural probes with embedded stiffeners and integrated flexible interconnect cables.

Each individual shank features a total of 16 recording sites, positioned in close proximity to one another. Notably, one recording site is strategically located at the tip of each shank, serving as a lead site. The purpose of the lead site is to act as a reference point, ensuring successful implantation into the intended region.

The remaining 15 recording sites are positioned linearly alongside the shank, with a fixed distance of 100µm between the lead site and the first of these additional sites. Vertically arranged on both sides of the shanks, the rest of the recording sites are placed at regular intervals of 26µm, forming a well-organized and high-density configuration. Figure 2.6 presents a series of SEM images capturing distinct shanks following the patterning of the n-doped polysilicon layer to generate conductive sites and traces.

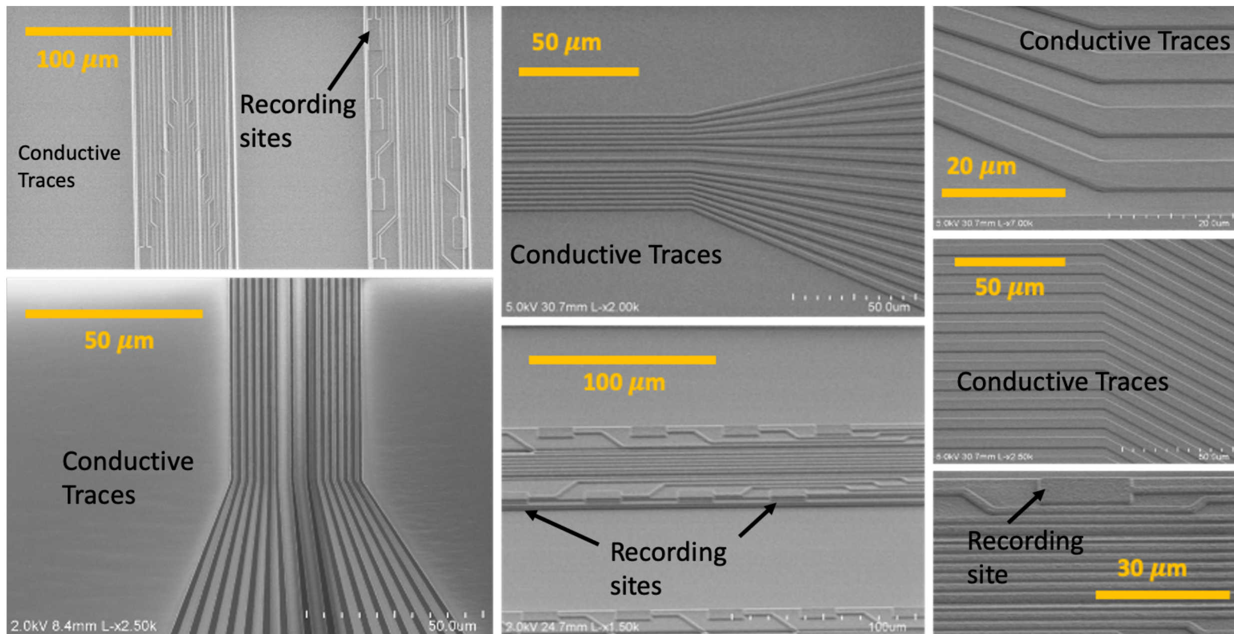

Figure 2.6: SEM images of different shanks after patterning the n-doped polysilicon layer to create conductive sites and traces.

To isolate the conductive layer, a third insulator layer is deposited. This layer consists of a stack of five layers of silicon dioxide, silicon nitride, silicon dioxide, silicon nitride, and silicon

dioxide with thicknesses of 125 nm, 62 nm, 126 nm, 62 nm, and 125 nm respectively. The total thickness of this insulator layer is 500 nm.

Adding this final insulator layer results in a total thickness of the shanks between 2.5µm (where conductive traces are not present) and 3.5µm (where conductive traces are present). The total thickness of the shanks where the stiffeners are located is around 13µm as shown in Figure 2.7.

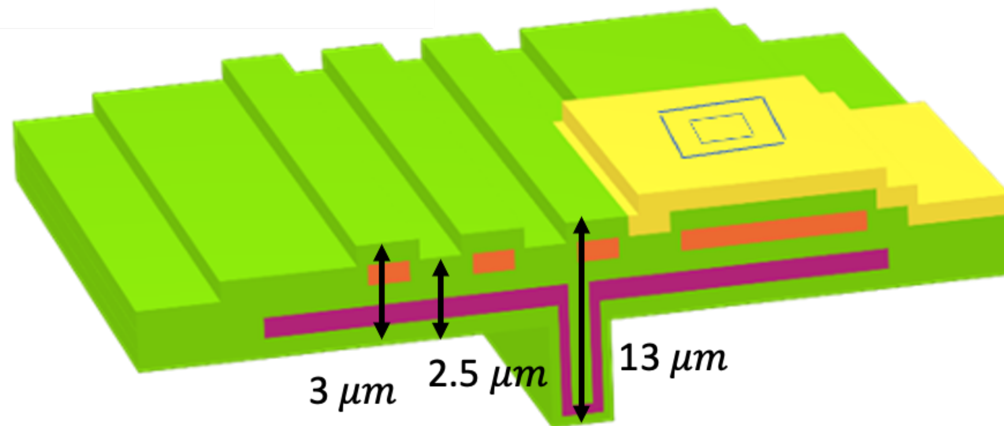

Figure 2.7: Cross-sectional view of the probe, showing the total thickness of the shanks and the location of the stiffeners.

Following this, insulator layers are patterned using plasma etching to define the shape of the four probe shanks. The next step involves accessing the conductive layer at the recording sites. This is done by etching the insulator layer in two steps to ensure that the metal layer completely covers the exposed conductive layer at the sites.

Then, the first layer of metal (300Å Cr/3000Å Au) is deposited using electron beam evaporation and selectively patterned to cover only the recording sites using the lift-off technique. This stair shape opening process ensures a solid contact area between the metal layer and the conductive layer at the recording sites. Figure 2.8 illustrates the difference between a flat shape

opening in the insulator layer and a stair shape opening, which facilitates contact between the metal film and the underlying conductive layer.

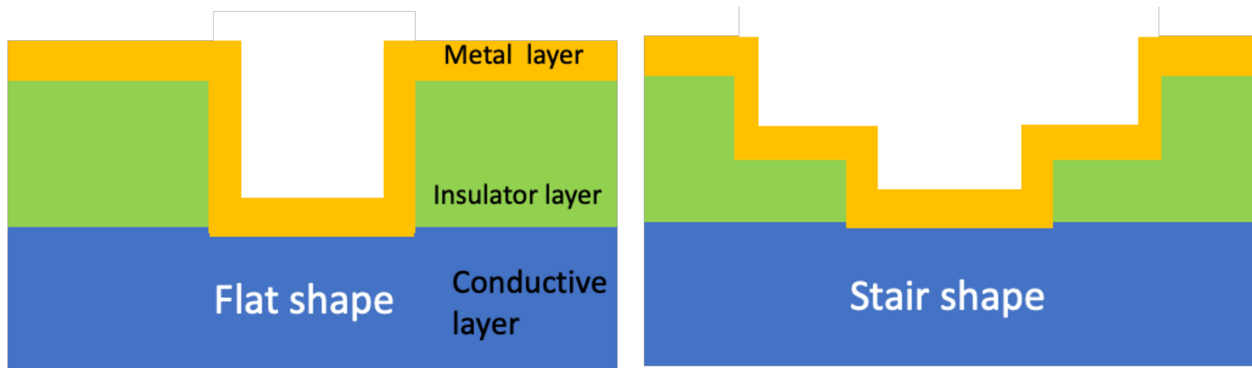

Figure 2.8: Schematic comparison between flat and stair shape openings in the insulator layer, highlighting the contact between the metal film and the underlying conductive layer.

In this process, we employed a two-step mask opening approach to etch the insulator layer. The first mask, featuring window openings of $3\ \mu m \times 10\ \mu m$, was utilized to etch the initial 250 nm of the insulator layer. Subsequently, we employed the second mask with window openings measuring $1\ \mu m \times 5\ \mu m$ to etch the remaining insulator, enabling access to the underlying conductive layer.

This stepped structure offers two notable advantages. First, it ensures a reliable ohmic connection between the conductor layer and the metal layer, which will be deposited atop it. By covering two shallow steps instead of one large step, we achieve improved coverage and connectivity. Second, this approach minimizes mechanical stress at the contact region with the underlying conductive layer.

This reduced stress is highly desirable for long-term mechanical and electrical durability, particularly in chronic measurements. To illustrate the efficacy of this two-step site formation technique, we performed FEM modeling using COMSOL, comparing it to the conventional one-step connection.

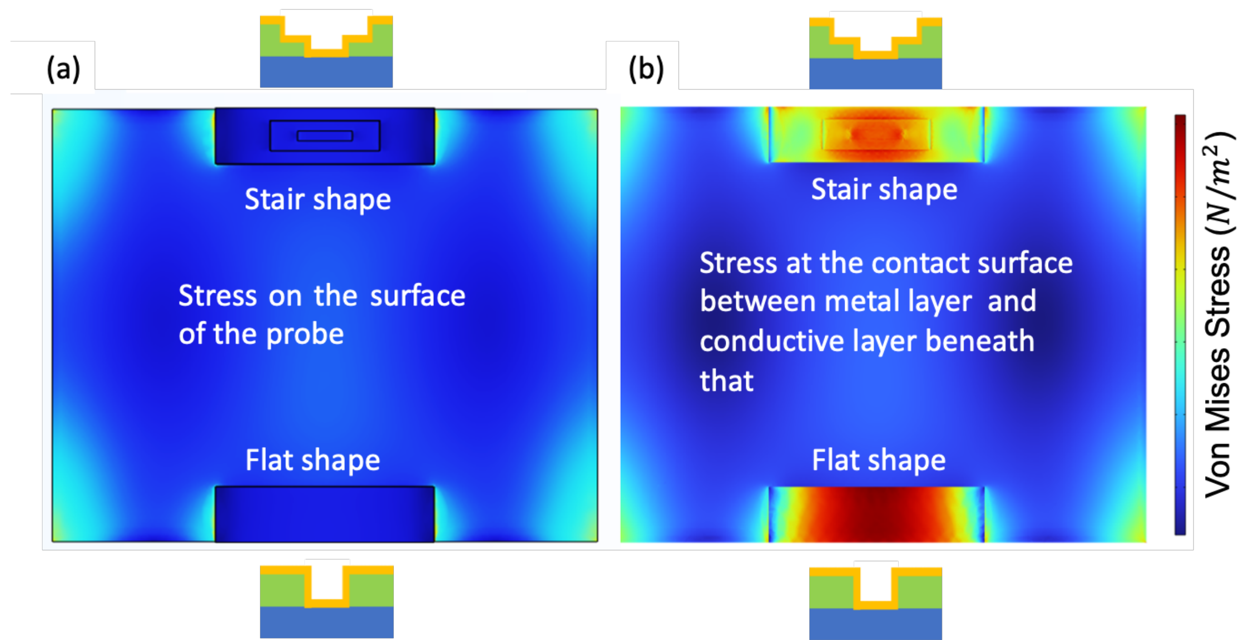

Figure 2.9: The von Mises Equivalent Elastic Strain rainbow-colored maps, (a) on the surface of the shanks, (b) in the vicinity of the contact area between the metal site and the n-poly silicon.

This simulation involved applying a 1µm displacement perpendicular to the flat surface of the shank (i.e., the thickness direction). The objective was to investigate how these probe micromotions influence the mechanical strain of the recording sites. The results of this simulation, presented in Figure 2.9, depicted the von Mises Equivalent Elastic Strain using a rainbow-colored map.

Remarkably, increased strain was detected near the boundaries of the recording sites, which aligns with the findings of T. Kozai et al. [87] who reported that motion-induced mechanical strain in a planar microelectrode array is mainly localized around the electrode site. To gain further insight, the strain within the recording sites was examined, not only on the probe surfaces (Figure 2.10) but also in the vicinity of the contact area between the metal layer and the underlying n-poly silicon (Figure 2.11). The analysis revealed an increased strain in the middle of the recording sites, indicating a higher level of mechanical mismatch with the underlying layer.

Importantly, the analysis demonstrated that the conventional method for metallization of the recording sites was more susceptible to detachment and delamination of the metal pads due to this increased stress. In contrast, based on the simulation results, the two-step metal coverage approach appears to provide enhanced reliability and reduced risk of detachment. These findings highlight the advantages of the two-step method in terms of improving the long-term mechanical durability of the recording sites in planar neural probes.

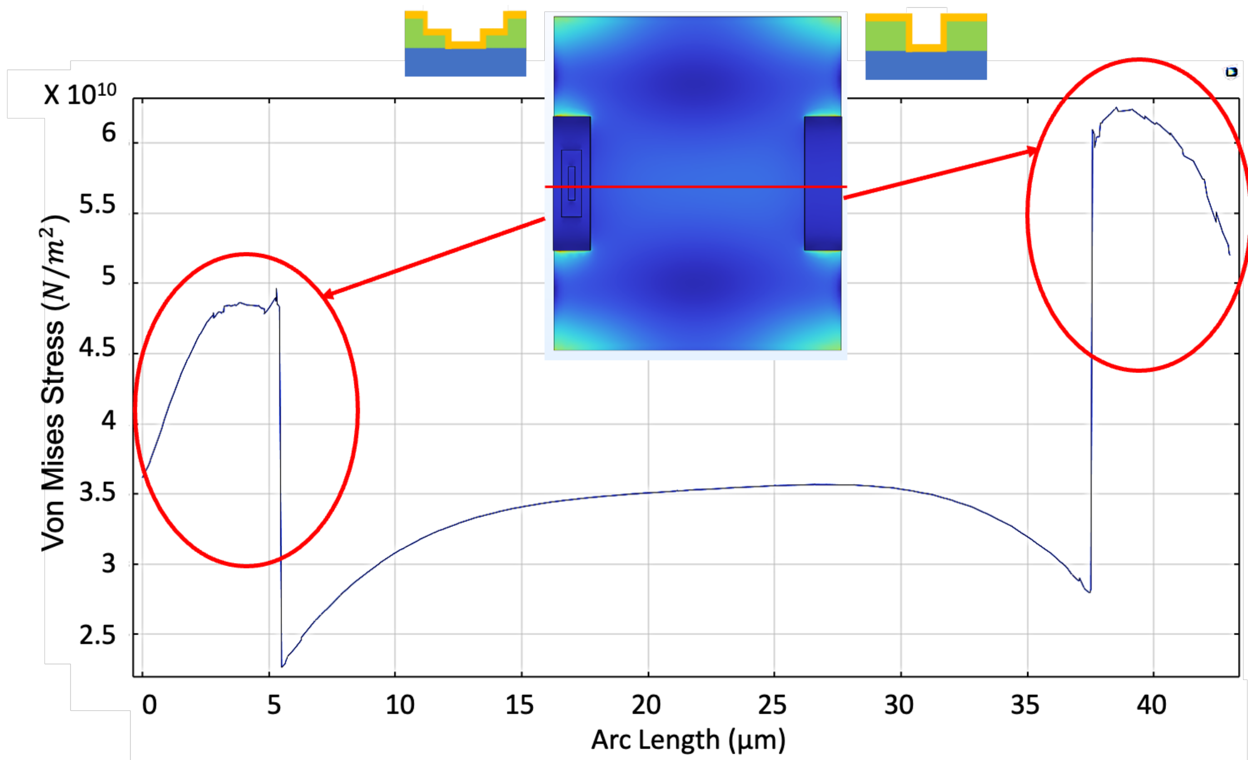

Figure 2.10: Comparison of von Mises equivalent stress on the surface of shanks between stair shape and flat shape.

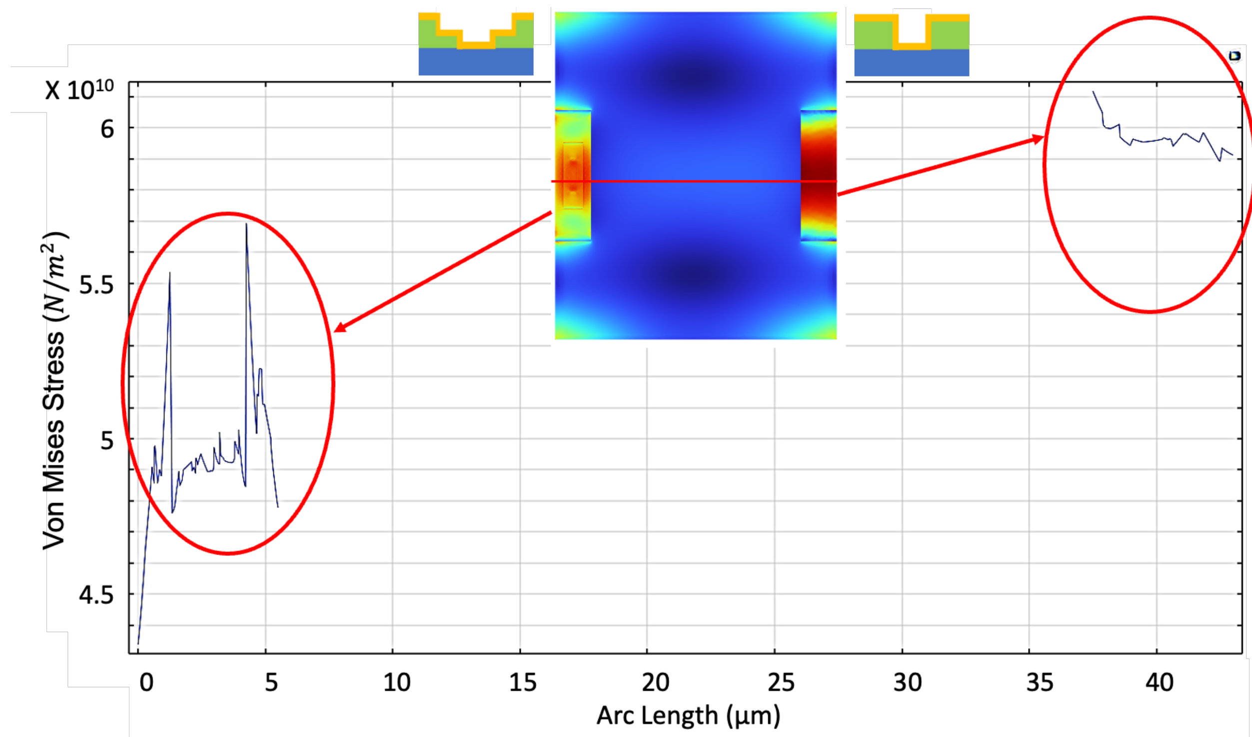

Figure 2.11: Comparison of von Mises equivalent stress at the contact area between the metal layer and the underlying n-poly silicon for stair shape and flat shape configurations.

Figure 2.12 shows SEM images of the two-step openings at the recording sites and the probes' backends which are strategically positioned where the metal layer contacts the conductive layer. The purpose is to enhance the electrical connection (ohmic connection) between the metal and conductive layers, while also reducing strain on the metal layer at the contact region with the underlying conductive layer. The images provide a clear visual representation of this design, emphasizing its goal to improve both electrical performance and mechanical durability.

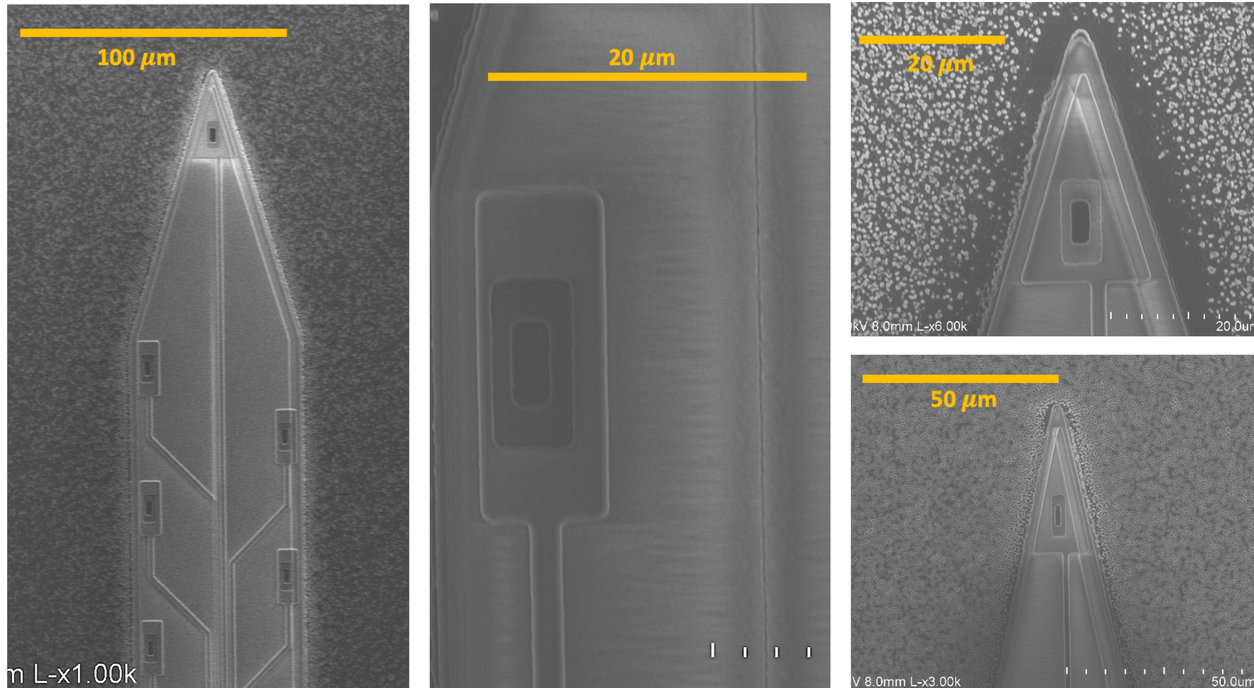

Figure 2.12: SEM images of the stair shaped openings at the recording sites which are strategically positioned where the metal layer contacts the conductive layer. This design aims to improve the ohmic connection between the metal layer and the conductive layer, while also lowering mechanical strain on the metal layer at the contact region with the underlying conductive layer.

The probe fabrication includes the integration of flexible cables, designed to facilitate easier handling during implantation and minimize tethering effects post-implantation. These flexible cables, measuring 20mm in length and 5µm in thickness, are composed of Parylene-C.

The process begins by depositing the first layer of Parylene-C with a thickness of 2.5µm using PVD technique. Subsequently, plasma etching is employed to shape the flexible cables through patterned Parylene-C (Fig. 2.5). Figure 2.13 displays SEM images of the probes after patterning the first Parylene-C layer on the probes' backends and cables.

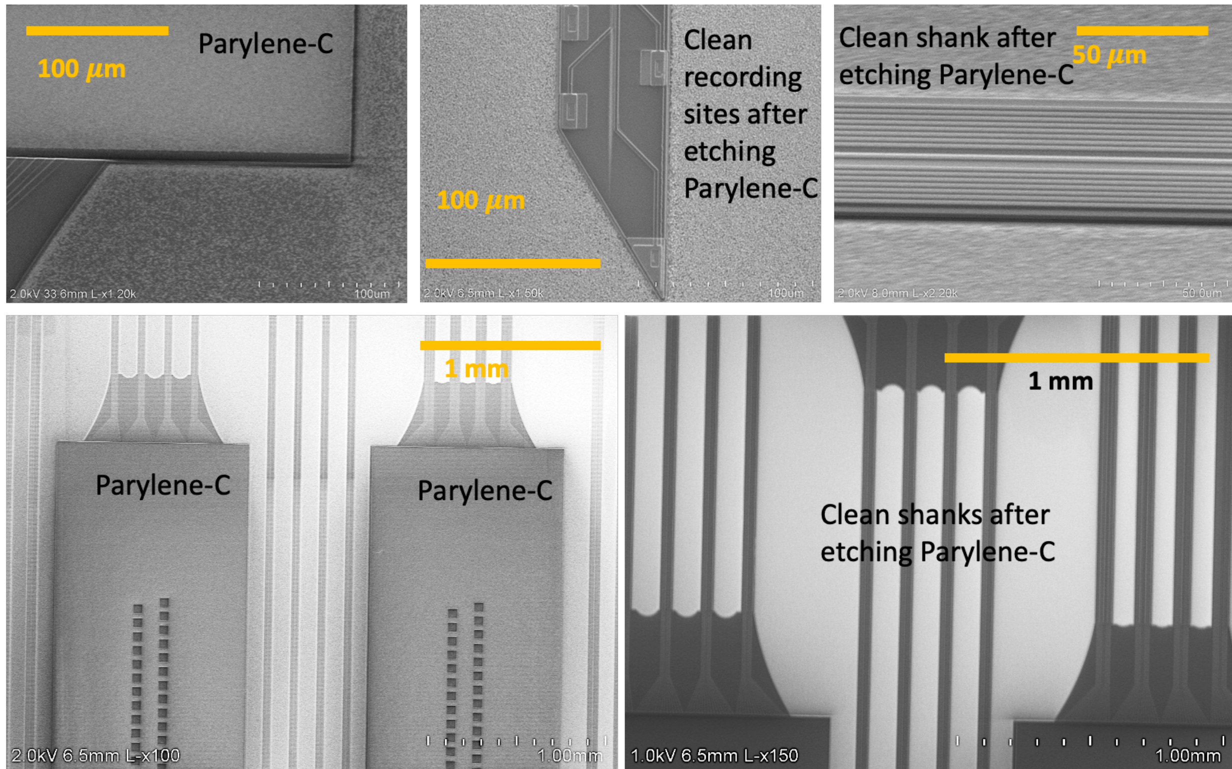

Figure 2.13: SEM images showcasing probes post-patterning of the initial Parylene-C layer on backends and cables. The first metal layer on recording sites is visible, along with the removal of Parylene-C from shanks. Additionally, openings on probe backends, representing the connection region between the flexible cable and the probes, are highlighted. The shanks and recording sites are shown to be clean after etching the initial Parylene-C layer.

Next, a second metal layer composed of Cr/Au/Cr is evaporated onto the patterned Parylene-C using an E-beam. The respective thicknesses of this metal layer are 300/3000/300Å. After the metal layer deposition, lift-off patterning is employed to create metal traces with a width of 10µm and a gap of 6µm between them. The recording sites themselves are sized at 15µm × 20µm, resulting in an area of 300µm^2.

Moreover, an additional layer of Cr/Au with respective thicknesses of 300/3000Å is selectively deposited exclusively at the recording sites. This additional metal layer ensures complete coverage of the recording sites and enhances impedance characterization. Figure 2.14 depicts the metal traces after the lift-off process before depositing the second Parylene-C layer on

the probes' backends and cables. The step coverage on the edge of the first Parylene-C layer is also illustrated, and the metal layer on the recording sites are visible.

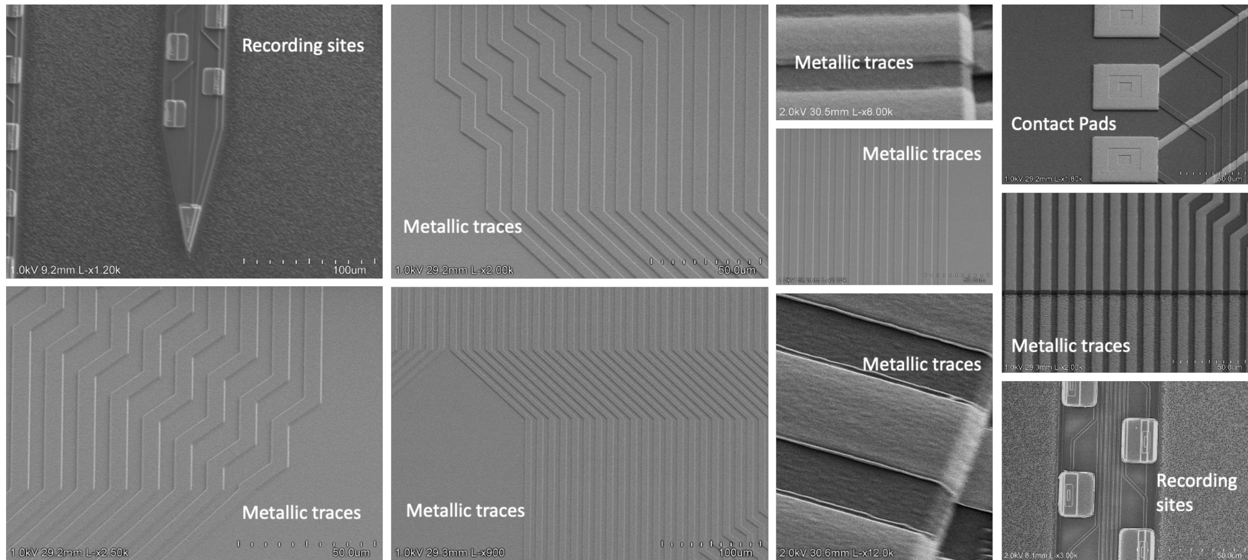

Figure 2.14: SEM images of the metal traces post lift-off, prior to the deposition of the second Parylene-C layer on probes' backends and cables. The image highlights step coverage along the edge of the Parylene-C layer, with visible metal layers on the recording sites.

The subsequent step in the process involves depositing a final 2.5µm Parylene-C layer. This layer serves the purpose of isolating the metal lines on the flexible cables, leaving only the contact pads exposed on the cable backend. This enables wire bonding to the readout printed circuit boards, facilitating data acquisition from the implanted probes.

Finally, the probes and flexible cables are released by DRIE etching from the backside of the Si substrate. Initially, the processed wafer is mounted on a dummy silicon substrate. Through deep reactive ion etching (DRIE) of the backend, the devices are released while still mounted on the carrier wafer. To separate the delicate samples from the substrate, it is essential to immerse them in a suitable solution that dissolves the mounting material. Here we used Santovac 5 as the mounting material.

Figure 2.15 presents the various components of the devices, highlighting the ability to create probes with different shank sizes and tip shapes, as well as varying numbers of stiffeners (ranging from zero to 3) and flexible cables connected to the backends.

In a few samples, the metal recording sites extend outward from the edge, creating 3D recording sites to increase the effective surface and decrease the effective impedance, leading to more efficient recordings. The illustration captures the devices post-plasma etching, situated on a layer of Santovac 5, before the removal of Santovac 5 and the subsequent release of the devices.

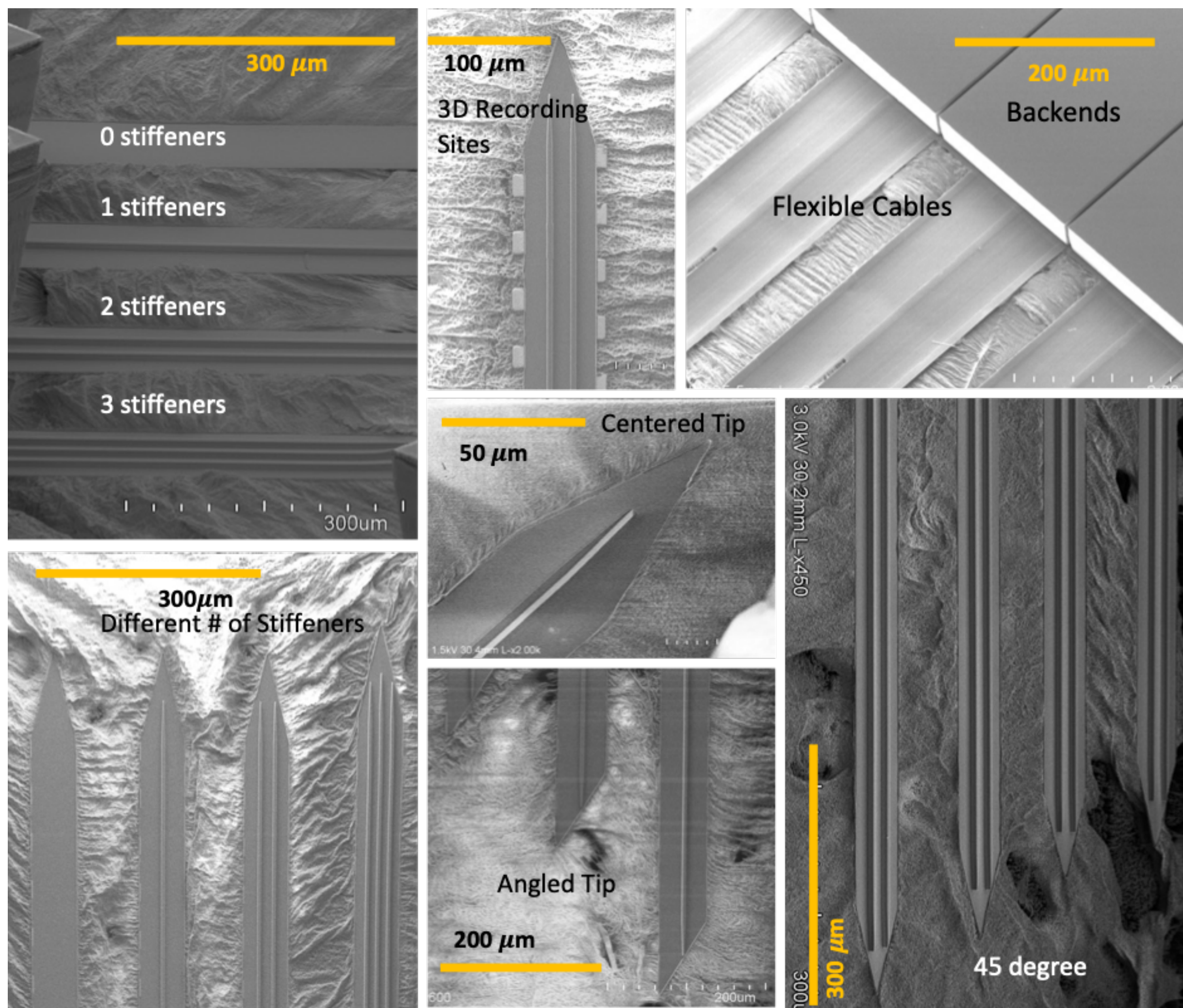

Figure 2.15: SEM images of device components showcasing customizability with diverse shank sizes, tip shapes, stiffener variations (0 to 3), and flexible cables connected to backends. Notably, innovative 3D recording sites enhance effective surface area for more efficient recordings. Illustration depicts devices post-plasma etching on a Santovac 5 layer, before release.

Releasing these small devices with their large backends presented a challenge, but it was successfully achieved. A stainless-steel mesh was utilized to hold the devices in place on the carrier wafer while they were soaked in acetone. This ensured that all components of the devices remained on the carrier wafer without floating into the solution before the final release.

Despite initially facing challenges in completely removing the Santovac 5 residue from the shanks, the devices were ultimately released and successfully separated from the carrier wafer with great effectiveness. Figure 2.5 illustrates the entire fabrication process of planar shanks with the embedded stiffeners at the backend of these probes.

2.4 Engineered Mechanical and Electrical Properties

This section highlights the engineered mechanical and electrical properties of these high-density multi-shank probes. The development of such super-fine shanks enables more precise neural recordings and stimulation while minimizing tissue damage. Additionally, the implementation of silicon-based shanks with ultra-high aspect ratios provides enhanced flexibility and adaptability to different brain region structures.

In this design, the residual stress of the layers is effectively mitigated in order to obtain straight shanks upon their release from the substrate. This straightness serves as a crucial factor for precise positioning and implantation that enables reliable neural recordings. The integration of flexible cables stands as a key element in these advancements.

By implementing a central opening within the cables, their flexibility is significantly enhanced, facilitating easier insertion into tissue structures. The incorporation of super fine and flexible Parylene-C cables, with a thickness of 5 μm and a length of 20mm, not only ensures

reliable signal transmission but also minimizes the physical footprint of the probe, further optimizing its performance.

The probes are designed with widths ranging from 43μm to 61.5μm and feature one or more stiffeners. The length of the probe is tailored to the target animal, with dimensions of 3.8mm to 4mm for mice brains and 4.5mm to 5mm for rat brains. The shank tip angle, which can be defined, plays a crucial role that varies based on the application. For mice brains, the shanks are arranged at a 25-degree angle, while for rat brains, the angle is set at 45 degrees.

Figure 2.16 shows SEM images of our innovative shank designs, displaying their diverse shapes, sizes, and angles. These detailed images offer insights into the structural characteristics and geometric variations of the shanks. They provide a valuable resource for understanding the design flexibility and customization options of our advanced neural interface probes. In addition, our designs feature two distinct tip shapes to accommodate various implantation methods. Some probes are equipped with 45-degree beveled tips for angled insertion ease, while others have centric tips for vertical insertion.

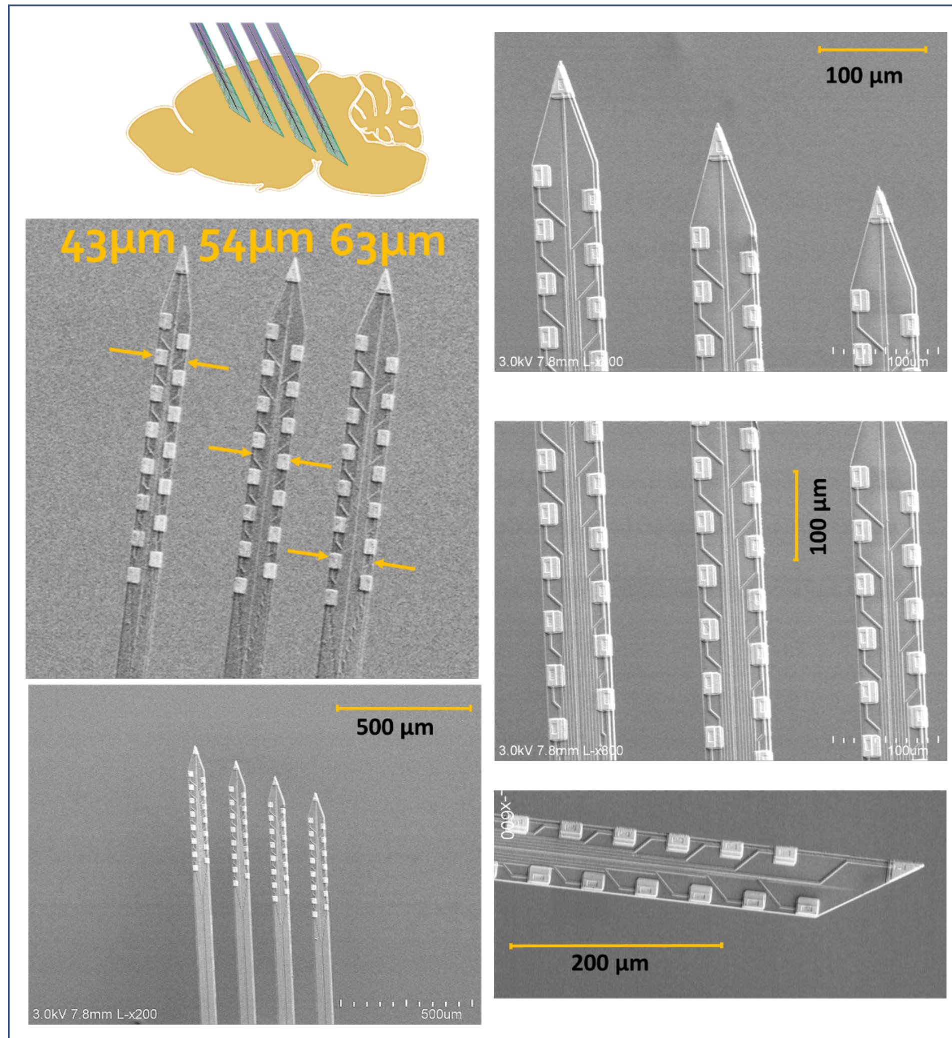

Figure 2.16: Exploring shank design flexibility: a collection of SEM images that illustrate various shank shapes, sizes, and angles incorporated in our innovative designs.

Figure 2.17 shows SEM images that provide a closer look at the different tip shapes, recording sites, two-step metallization, and conductive traces on the shanks utilized in our designs.

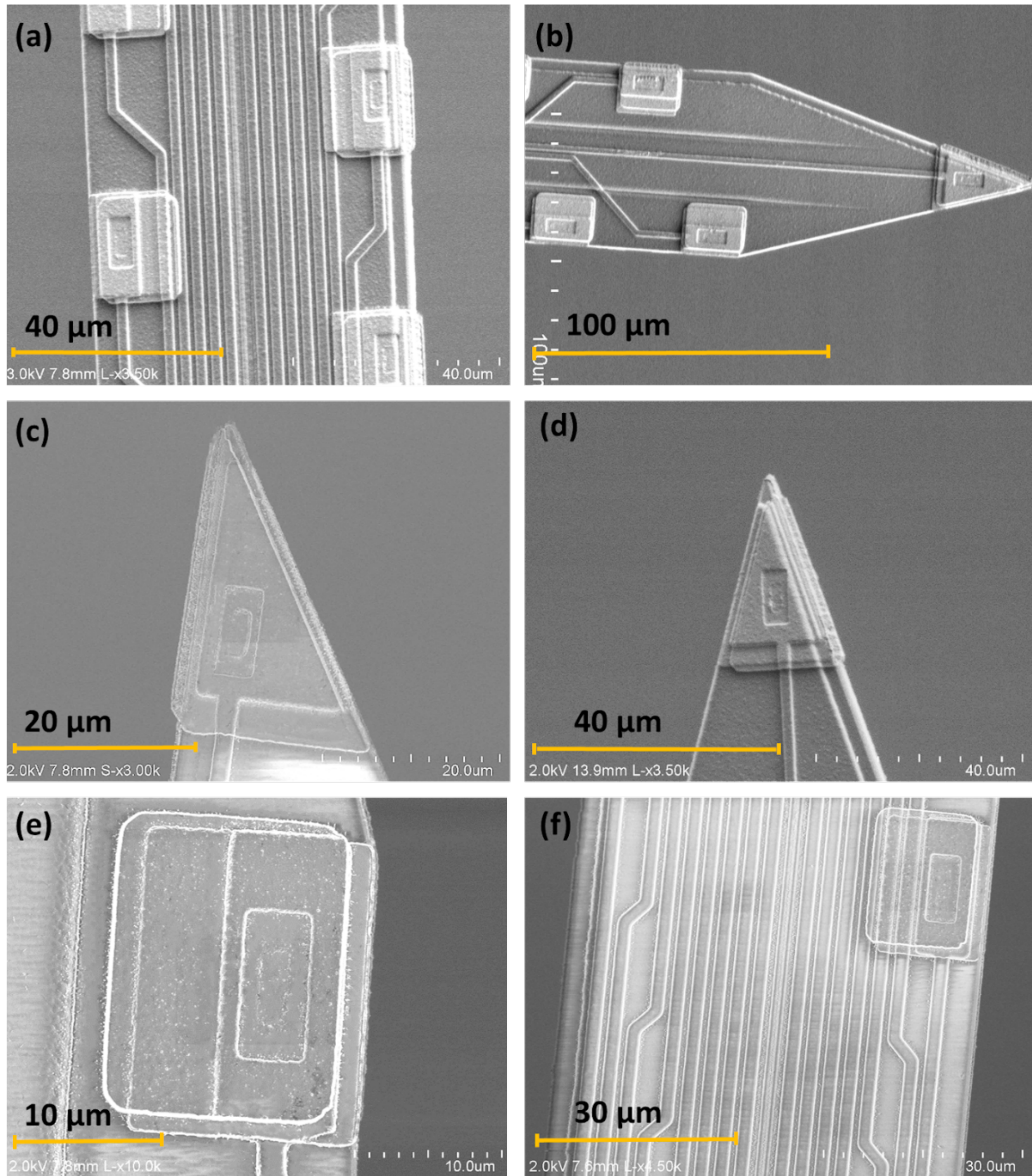

Figure 2.17: High-density recording sites, different tip shapes, two-step metallization, and compact n-poly silicon conductive traces on the shanks utilized in our designs.

Figure 2.18 and Figure 2.19 show more SEMs and optical images from different aspects of the presented planar probes, respectively.

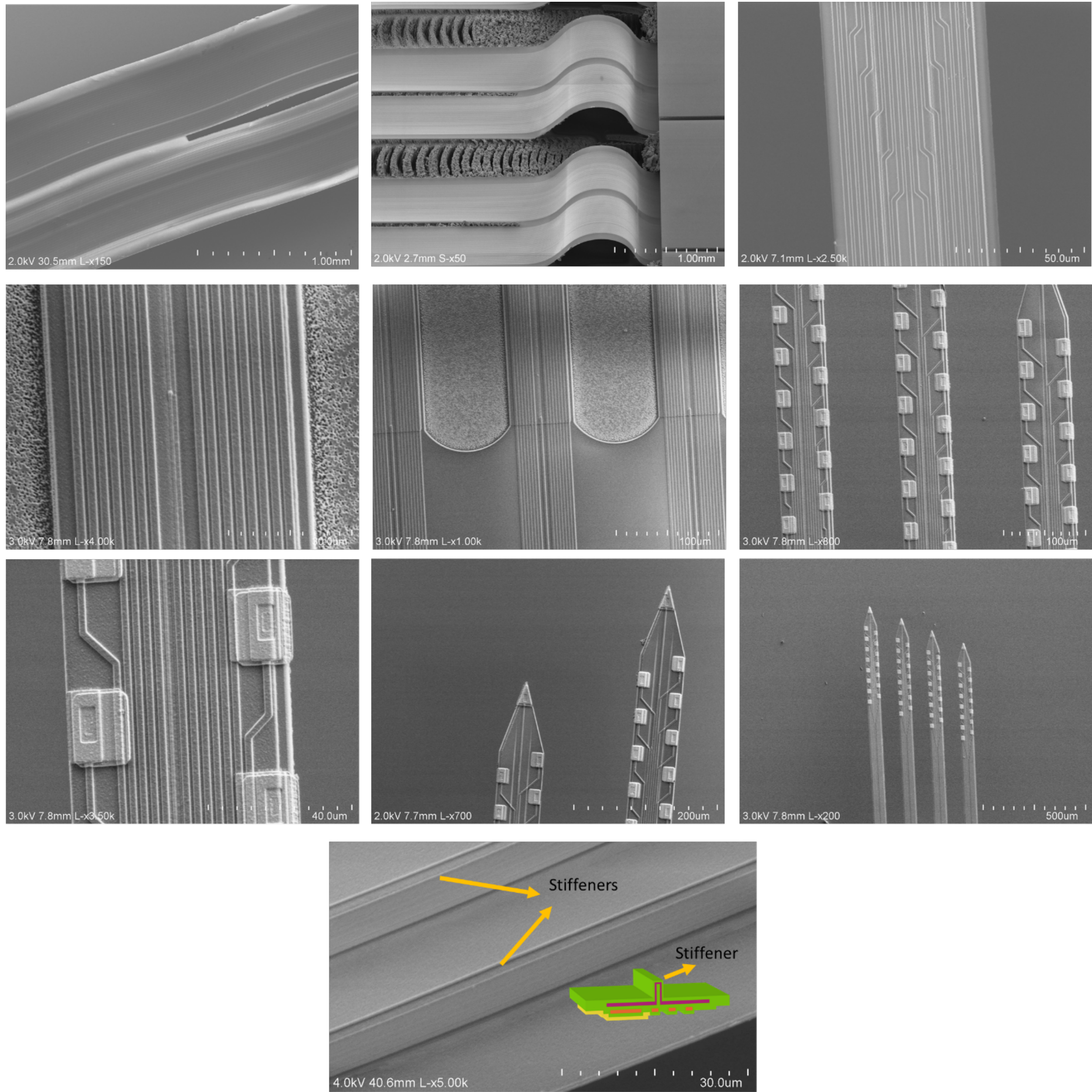

Figure 2.18: SEM images of different aspects and different designs of the presented planar probes.

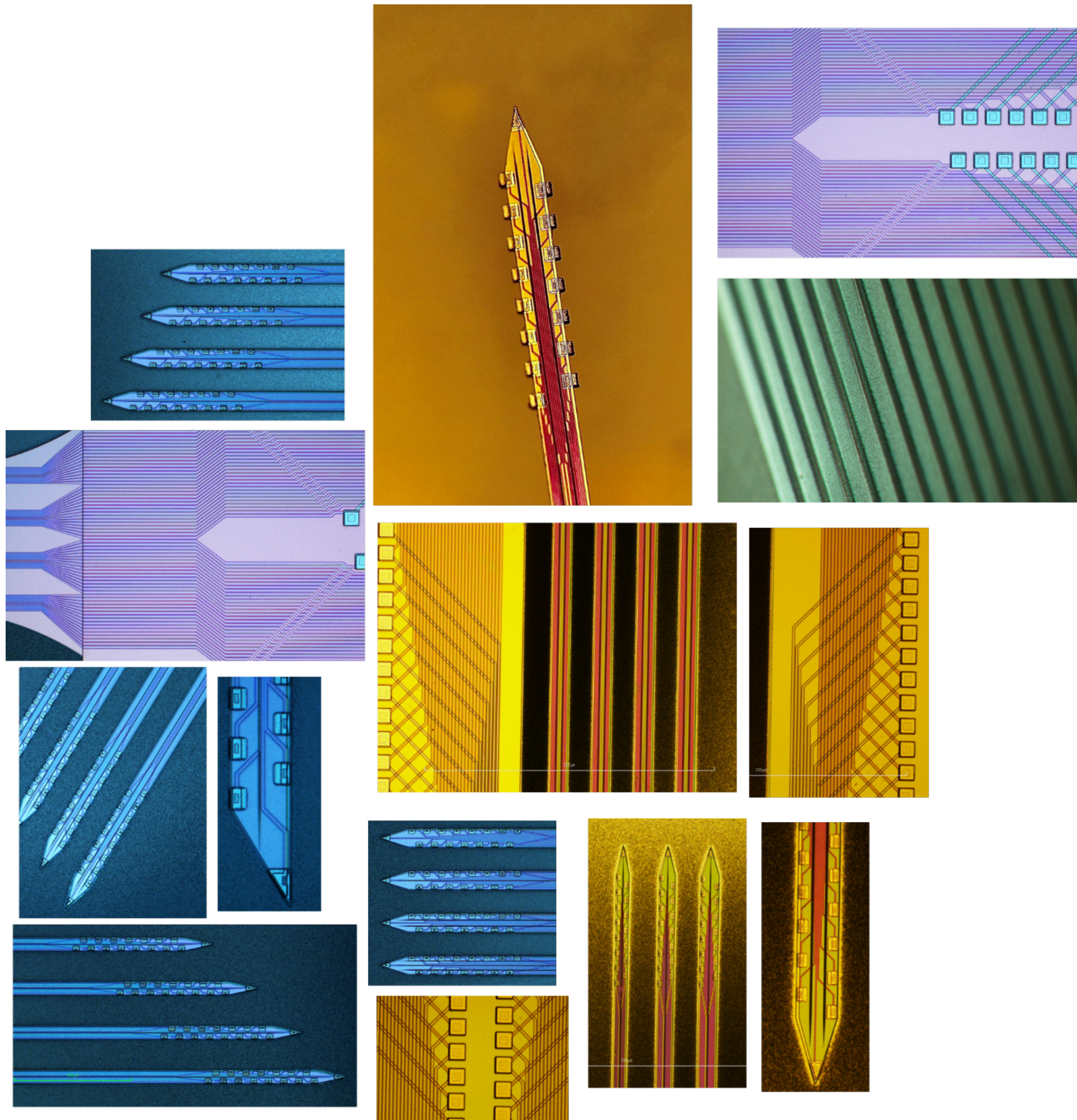

Figure 2.19: Optical images from different aspects of the presented planar probes.

The precise layering process used in fabricating these probes enables the production of durable devices that provide precise and dependable neural measurements. Figure 2.20 (a) illustrates the structure of a complete device, including the flexible cables, backend wire bonding, printed readout circuit board, implanted 4-shank probes, and their base within the rat's brain tissue.

Additionally, Figure 2.20 (b) displays a closer view of one of the flexible cables connected to the backend of the cable, while Figure 2.20 (c) displays optical microscopic images of the released multi-shank probes and their integrated compact and high-density recording sites with the flexible cables.

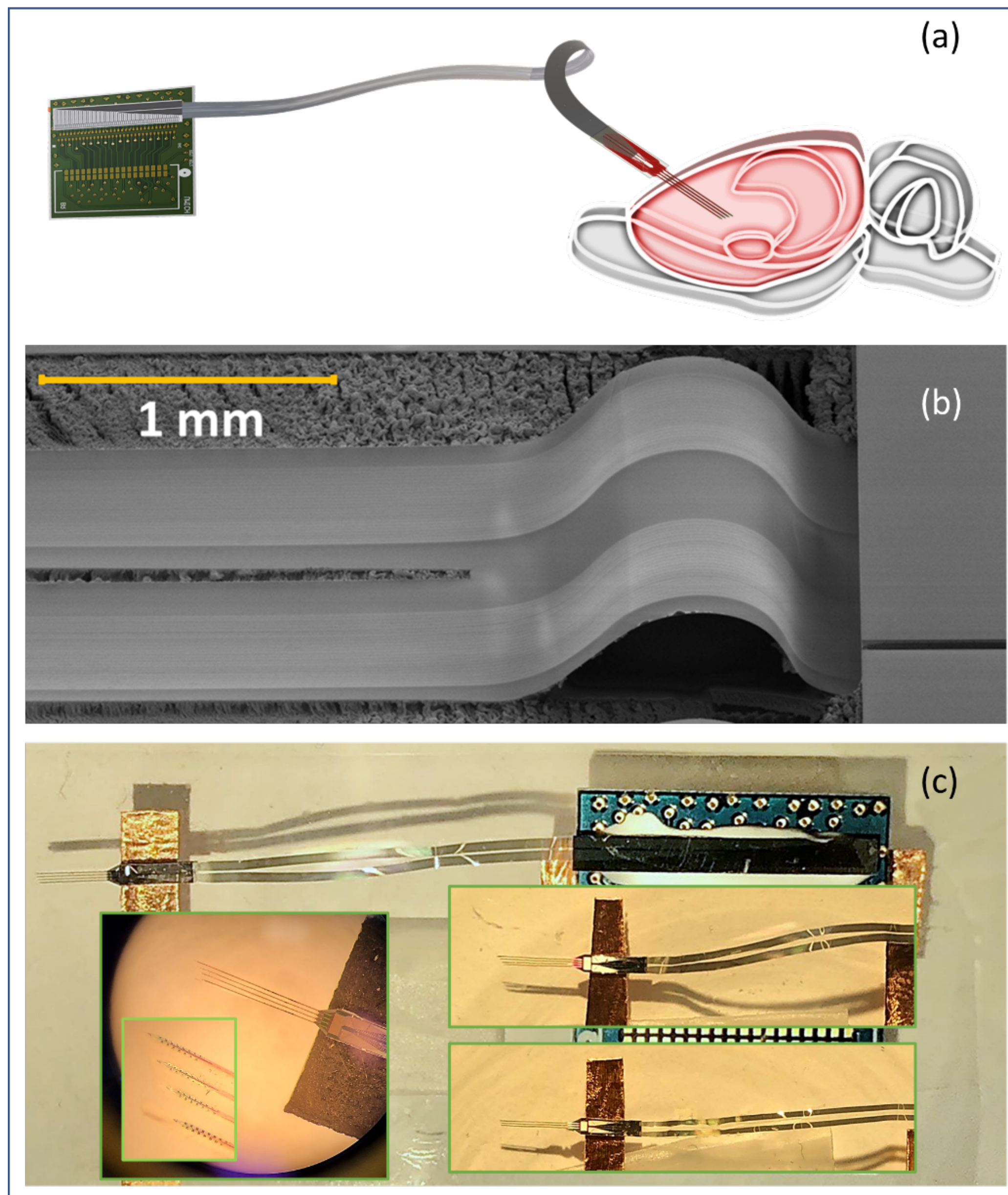

Figure 2.20: (a) Comprehensive overview of device components, (b) SEM image of flexible cable, (c) Optical microscopic images of released multi-shank probes with compact, high-density recording sites integrated with flexible cables.

Following the design and microfabrication phases, the next step involves wire bonding and packaging the prepared probes. This critical stage prepares the devices for in-vitro and in-vivo testing. To connect the small contact pads of the backend to the PCB, we utilized a wedge bonder with 0.7 mil Aluminum wires, as depicted in Figure 2.21.

Figure 2.21 (b) illustrates the layout of pads on the backside of the integrated flexible cable and the PCB for wire bonding. To protect the delicate shanks during the wire bonding process, we mounted the probe and the PCB on the same glass slide and used a plastic cover to shield the shanks, as shown in Figure 2.21 (e).

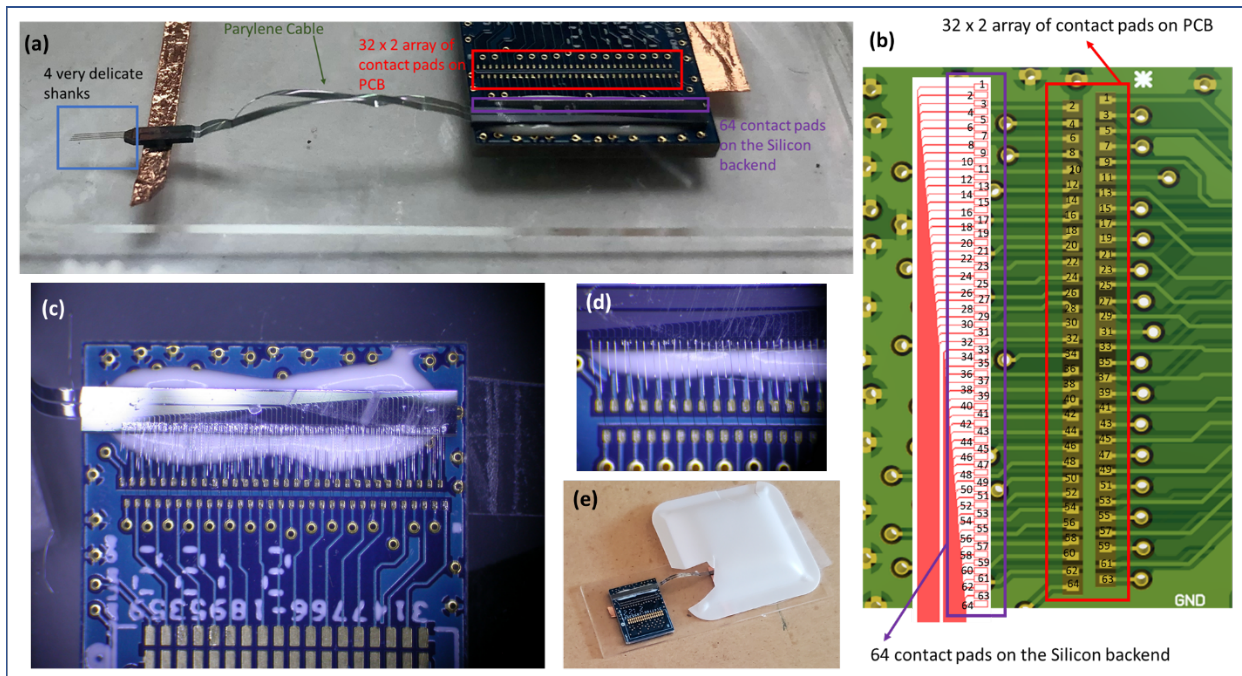

Figure 2.21: Wire bonding and packaging process for preparing probes. (a) Overall view of a 4-shank probe packaged and wire-bonded to the PCB. (b) Layout of 64 pads on the backside of the integrated flexible cable and the PCB for wire bonding. (c, d) Zoomed-in views of the 0.7 mil Al wire bonds. (e) Use of a plastic cover to protect the delicate shanks during the wire bonding process.

We proceeded with soldering connectors onto the PCB using solder paste and a hot air gun. The PCB, measuring 15mm × 18mm, featured 32-position Omnetics connectors on both the front and back, granting access to 64 recording sites. Once the packaging and wire-bonding step was

completed, the subsequent crucial phase involved preparing the probe for impedance characterizations.

These characterizations are vital for ensuring accurate in-vivo measurements. To accomplish this, we employed both the nanoZ impedance tester and Agilent vector network analyzer to measure the impedance of our devices. The obtained impedance measurements yielded promising results, revealing the phase and magnitude of impedance for some of the sites, as depicted in Figure 2.22.

At 1kHz, the magnitude of impedance for the sites varied between 2.2 MΩ and 3 MΩ, while the phase of the impedance varied between -80 and -63 degrees. These impedance values are within the acceptable range for neural recording applications, indicating that our device's performance is satisfactory. To further improve the device's performance, we electroplated gold on the recording sites using nanoZ immediately before implantation, which significantly decreased the impedance magnitude of each site.

The resulting impedance magnitude ranged between 300 kΩ and 400 kΩ, which is much lower than the initial values. This decrease in impedance magnitude is attributed to the increase in the effective electrode area caused by gold electroplating. Moreover, the decrease in impedance magnitude indicates a better signal-to-noise ratio and improved sensitivity, which is crucial for accurate neural recordings. Overall, the in-vitro impedance characterizations of our device have demonstrated its potential for high-quality in-vivo neural recordings.

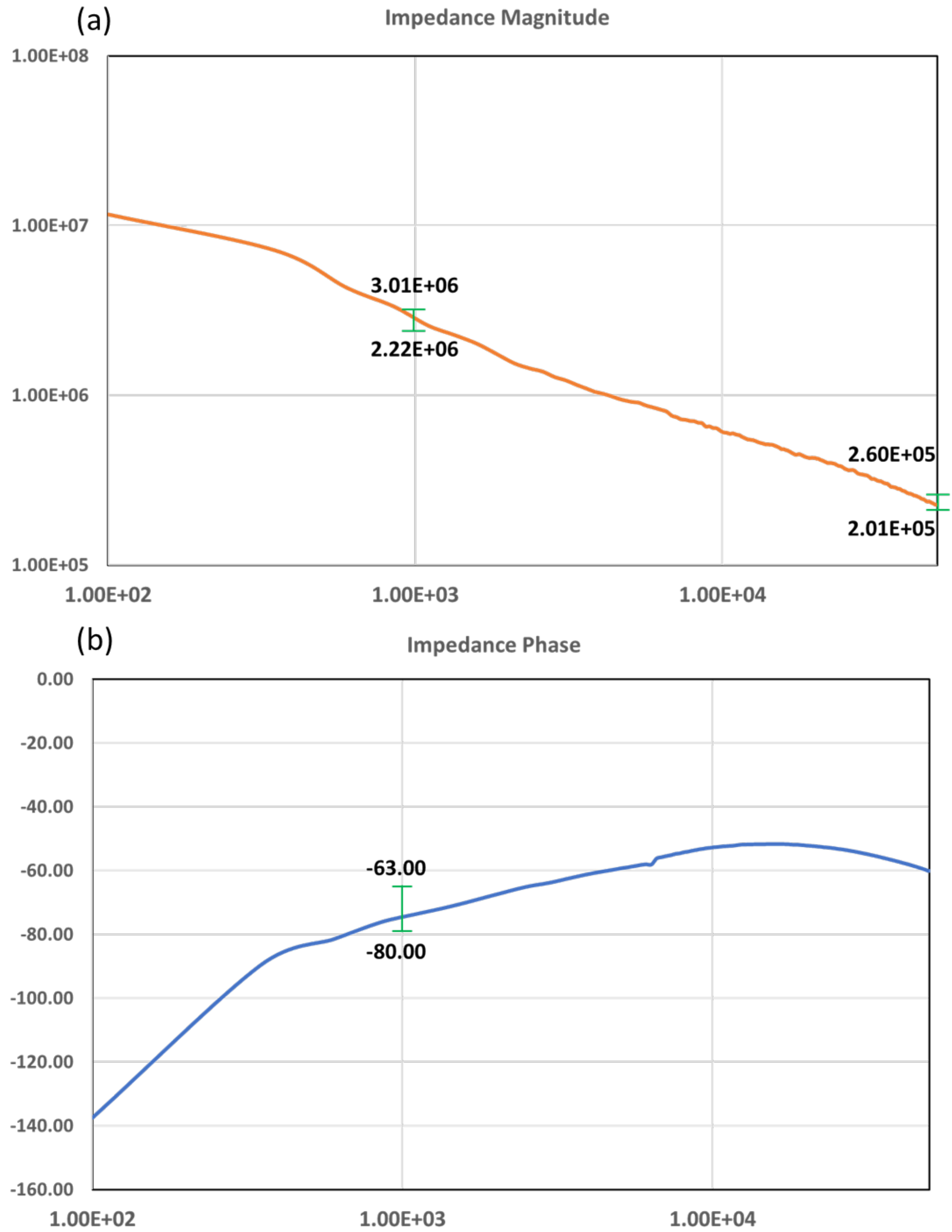

Figure 2.22: The impedance for selected recording sites, (a) Magnitude, (b) Phase.

To explore the impact of stiffeners, we designed and microfabricated a test sample comprising three shanks, all with the same thickness: one without a stiffener (Thin), one with a single stiffener (T-bar), and the last one with two stiffeners (π-bar), as depicted in Figure 2.23.

It is evident that the shanks with stiffeners maintain a perfectly straight orientation after release. In contrast, the thin shank without any stiffeners curls upward, indicating its limited robustness for precise implantation. While it should ideally remain almost flat, its increased flexibility causes it to curl up due to the layers not being completely symmetric and the stress within the composite layers not completely cancelled.

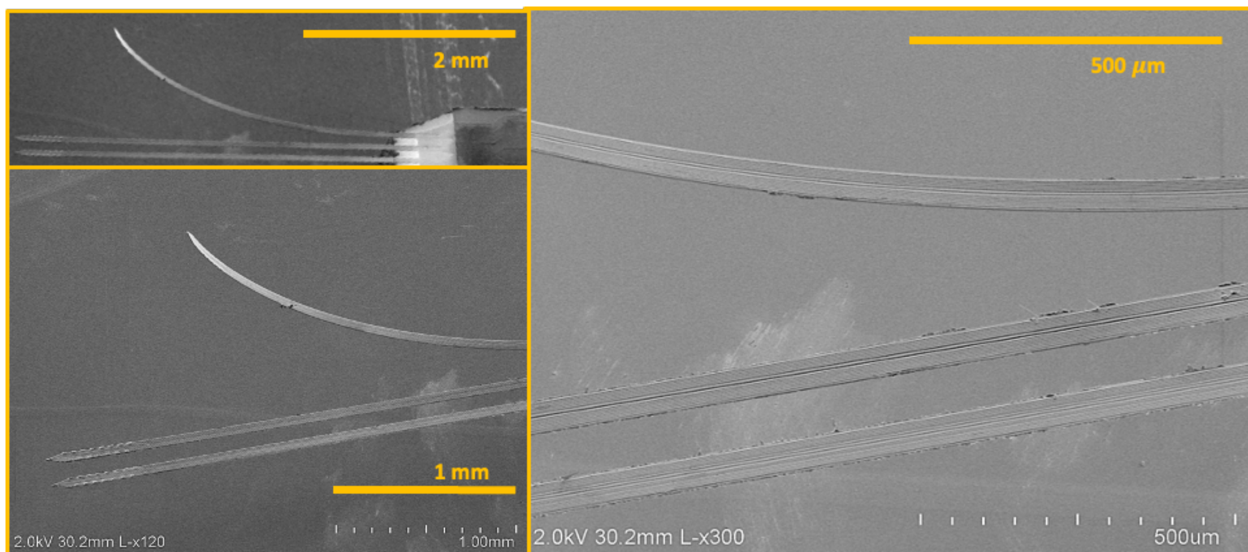

Figure 2.23: Impact of stiffeners on shanks - Comparative visualization of thin, T-bar, and π-bar designs in test samples, revealing the influence on post-release orientation for precise implantation.

The shanks of these probes exhibit exceptional compliance and flexibility, despite not being constructed from flexible materials. This inherent flexibility, combined with their compact size and close spacing, makes them prone to entwining together upon removal from the solution due to surface tension forces pulling the shanks together. However, once re-immersed, they effortlessly straighten back.

This behavior is clearly demonstrated in Figure 2.24, where a 4-shank probe is initially extracted from a low-concentration agarose gel and then re-implanted, showcasing the shanks returning to their straightened state. To evaluate the mechanical characteristics of these probes, we utilized tofu as a brain tissue-mimicking material. In Figure 2.25 the implantation and extraction of these probes into tofu is depicted, with saline covering the tofu to facilitate detangling during implantation.

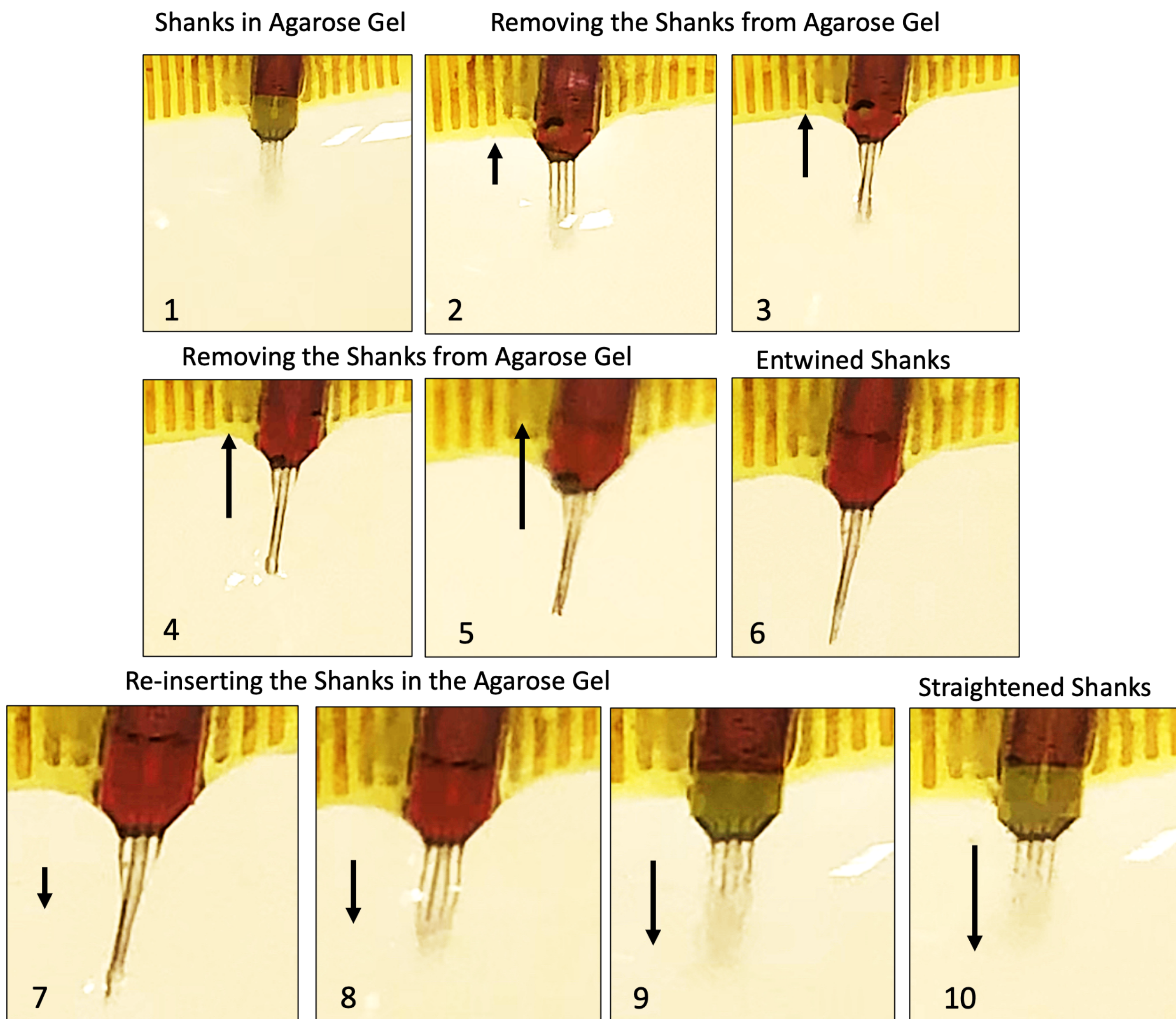

Figure 2.24: Demonstrating the exceptional compliance of probe shanks, entwining upon removal from a solution and effortlessly straightening upon re-immersion. A 4-shank probe is initially removed from a low-concentration agarose gel, revealing its inherent flexibility: (1) straightened shanks in the Agarose gel, (2-5) removing shanks from

the Agarose gel, (6) entwined shanks outside of the Agarose gel, (7-9) reinserting the shanks in the Agarose gel, (10) re-straightened shanks inside the Agarose gel.

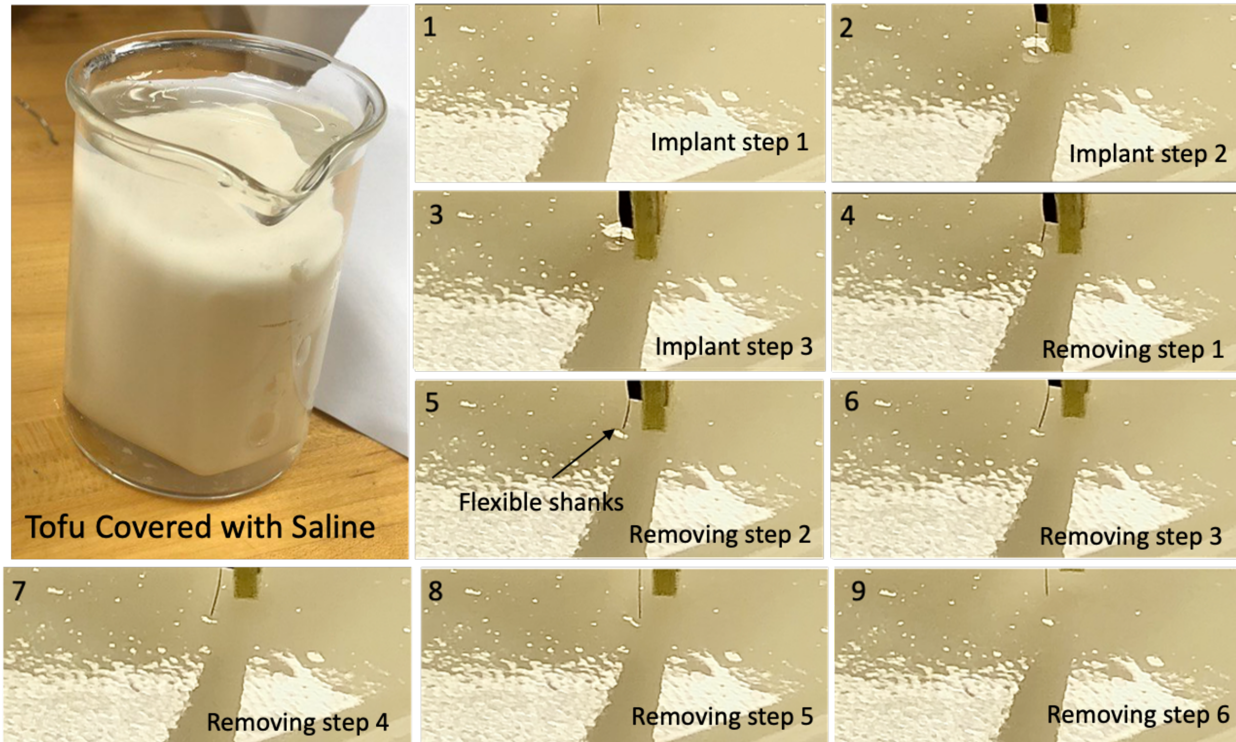

Figure 2.25: Mechanical test: implanting probes into tofu, a brain tissue-mimicking material, with saline covering to facilitate the process, (1-3) implanting the probe into tofu, (4-9) removing intact shanks from tofu.

During this experiment, we observed that none of the shanks broke, which was a positive outcome. The tofu implantation and removal experiments were conducted 2-3 times for each probe, and consistent results were observed each time. However, in a few samples after a few rounds of removal and reinsertion, one or two shanks broke. Nonetheless, we still had a good yield.

To evaluate the long-term durability of the two-step coverage site metallization method, the impedance of recording sites in a sample was continuously measured over an 11-month period, during which the probes were immersed in a PBS solution. Additionally, SEM imaging of the

recording sites was conducted after 11 months of continuous immersion. Figure 2.26 illustrates monthly recordings of the impedance for the 10 recording sites on a probe.

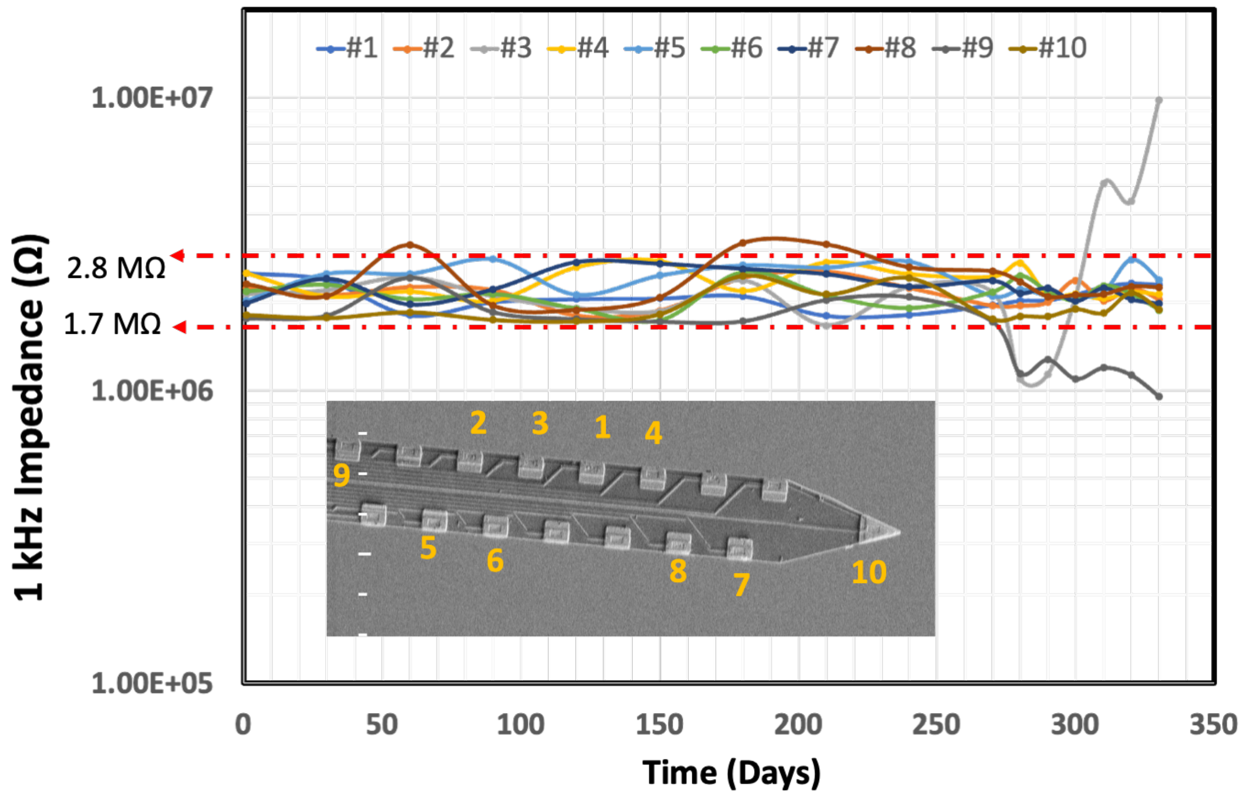

Figure 2.26: Continuous monitoring of recording site impedance performance on a probe immersed in a PBS solution over an 11-month period.

Among the ten recording sites analyzed, seven exhibited consistent performance over time. However, recording site number 3 displayed a decline in impedance after 9 months, suggesting possible insulation breakdown. Subsequently, the impedance sharply increased, indicating a disconnection in the electrode, which was confirmed through SEM analysis as shown in Figure 2.27, revealing material degradation in the electrical trace.

Recording site number 9 also showed a decrease in impedance after 9 months, likely due to insulation breakdown. Similarly, recording site number 8 exhibited significant impedance fluctuations around 1300 kΩ, with SEM analysis revealing a broken recording site, leading to

increased instability over time. Conversely, the remaining recording sites, upon SEM inspection, presented an almost intact and stable appearance, with no observable issues such as delamination or breakdown around the metalized recording sites.

In conclusion, the majority of recording sites demonstrated robust and stable performance throughout the over 11-month immersion period. However, material degradation observed in a few cases emphasizes the importance of further investigation into potential insulation breakdown points for enhanced electrode durability.

The results of this study, in comparison to those of a similar study conducted by Kozai, et al. [87], demonstrate significantly improved durability in the recording sites. In the previous study, over 90 percent of the measured sites failed in less than 190 days. In contrast, in this study, all of the recorded sites remained functional for up to 9 months, with 70 percent of them remaining functional for almost one year. This indicates a substantial enhancement in the longevity and reliability of the recording sites, which is a promising development for the field.

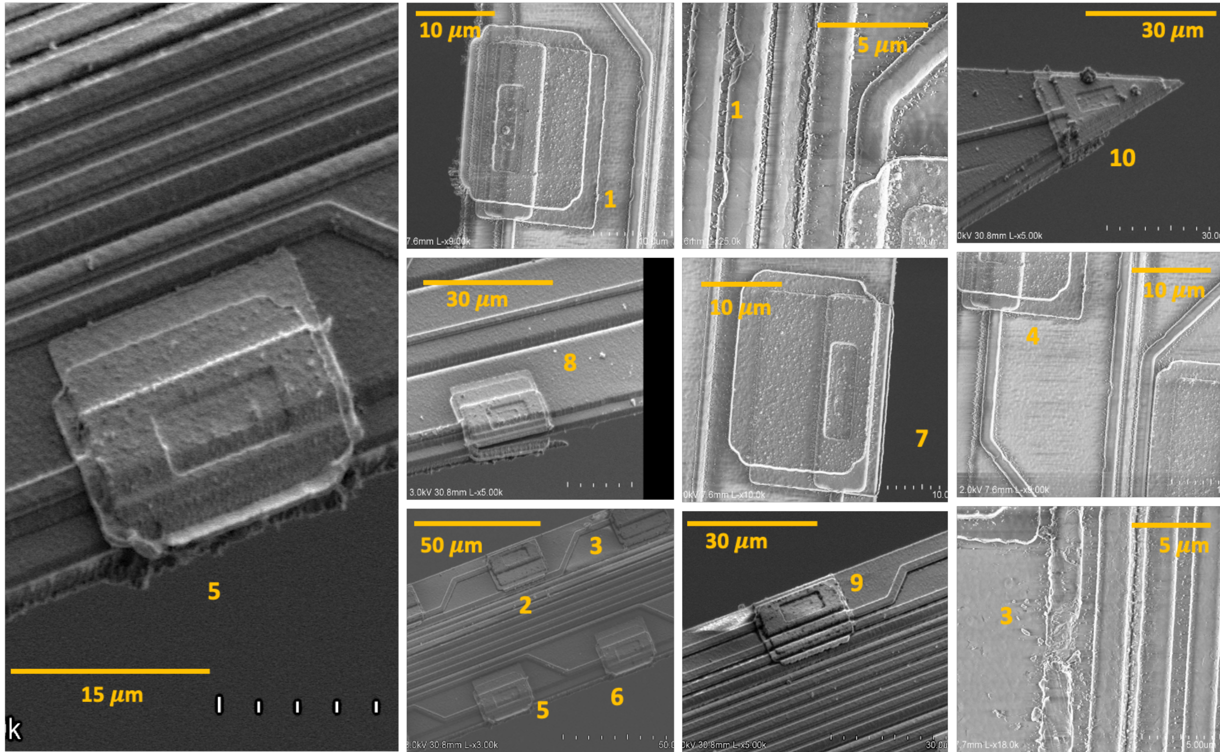

Figure 2.27: SEM images of the studied recording sites after over 11 months of immersion in PBS.

In-vitro tests were conducted to evaluate the recording and stimulation capabilities of these probes. In the initial test, a spike train mimicking a neural action potential was generated at a frequency of 40 Hz. This signal, produced by passing a spike train through an RC lowpass filter, had a peak-to-peak value of approximately 100μV. The signal was then applied to a reference electrode in the PBS solution, the same medium in which the probes were immersed.

Out of the 64 available sites, 14 were studied, though not simultaneously. This accounts for the lack of synchronization in the recorded signals within the time domain. For a deeper analysis of the results, the modeled impedance of the recording sites in the PBS solution was used to simulate the equivalent circuit of the test.

In AC conditions, the electrical double layer at the electrode/electrolyte interface functions as a capacitor, denoted as Cp, and polarization resistance is represented as Rp, which is a very

large parallel resistance and can usually be neglected. For a specific electrolyte, Rs is determined by the resistance between the reference and working electrodes [88, 89]. In practical terms, this ohmic resistance also includes the ohmic resistance of the connection cables and the working electrode.

A simple lumped circuit model, comprising resistors and capacitors, was derived using the transient responses obtained from the experimental setup. The parameter values in the model are extracted from the measured traces, and these values match those obtained through impedance measurements, as illustrated in Figure 2.28 (b).

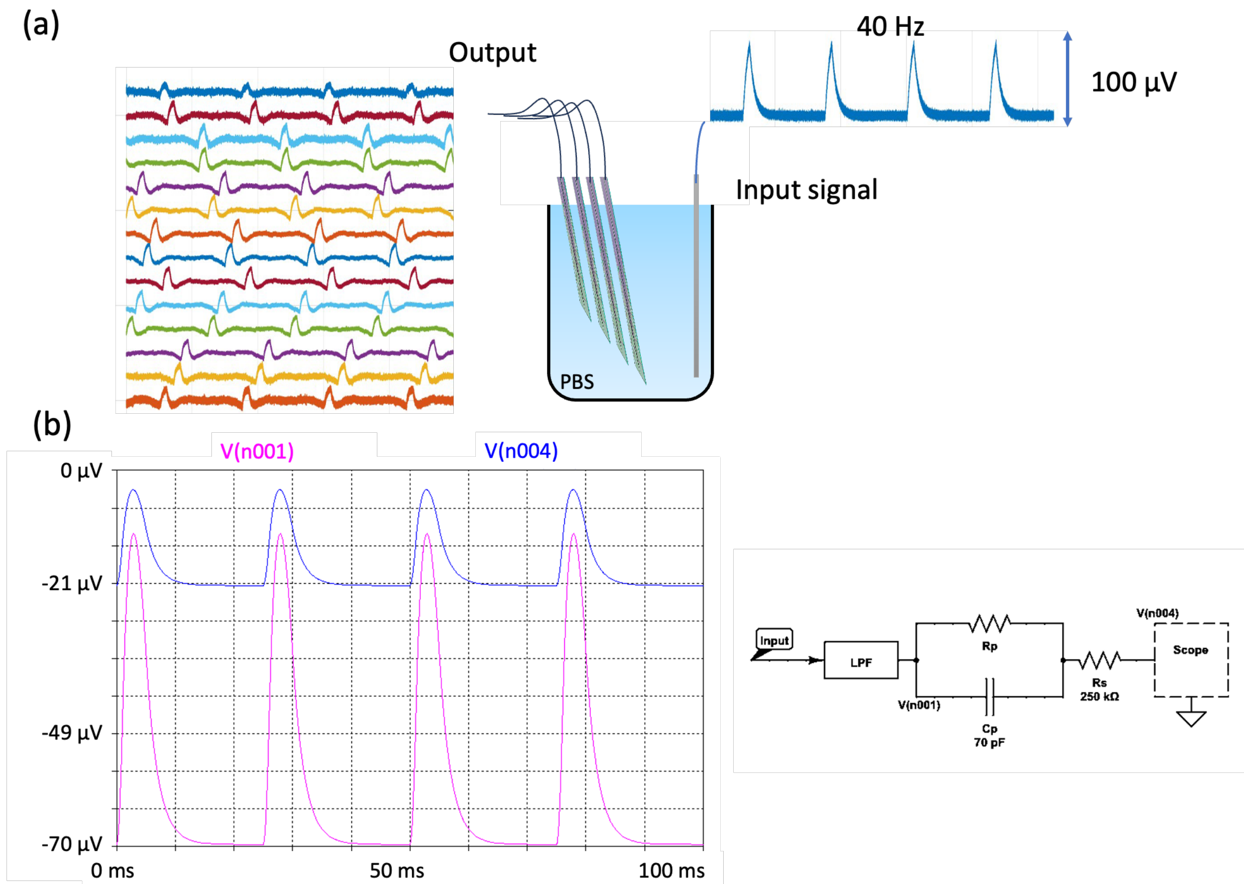

Figure 2.28: In-vitro assessment of probe recording capabilities. (a) The generated neural action potential-like signal, with a 40 Hz frequency, applied to a reference electrode in PBS solution, demonstrated responses across 14 studied sites. (b-c) A simple lumped circuit model, comprising resistors and capacitors, was derived using the transient responses obtained from the experimental setup. The simulated data closely matched the actual measurements.

In the subsequent in-vitro test, we assessed the stimulation capability of the recording sites. In this test, the same spike train was randomly injected into one of the recording sites, and the resulting signal sensed by neighboring recording sites in the same PBS solution was recorded. Figure 2.29 displays the stimulated signals marked by a red asterisk (*) and the recorded signals from the other 14 adjacent recording sites.

A lumped circuit model, comprising resistors and capacitors, was also derived using the transient responses obtained from this experimental setup. The parameter values in the model are extracted from the measured traces, and these values closely match those obtained through impedance measurements, as illustrated in Figure 2.29 (b).

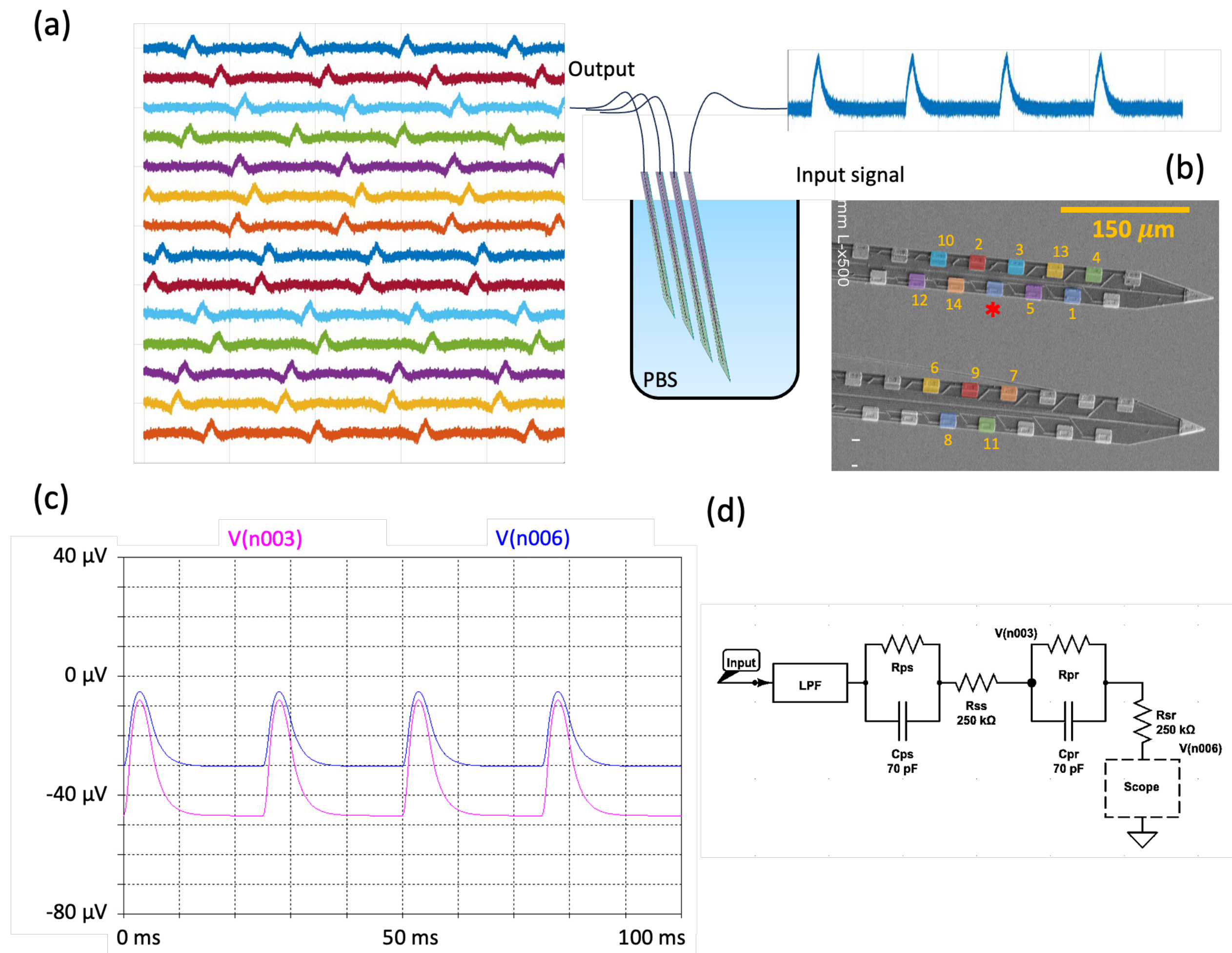

Figure 2.29: Assessment of stimulation capability in an in-vitro test. (a) Color-coded representation of 14 tested recording sites. (b) Stimulated signal marked by red asterisks (*) and recorded signals from 14 neighboring sites, shown in color, demonstrating the probes' responsiveness. (c-d) Equivalent lumped circuit comprising resistors and capacitors, was also derived using the transient responses obtained from this experimental setup. The simulated data closely matched the actual measurements, validating the functionality of the probes in stimulating and recording signals in an in-vitro setting.

The promising results serve as encouragement for the upcoming phases of this study, specifically in-vivo measurements to evaluate the device's performance under realistic conditions in the brains of mice and rats. Due to time constraints, in-vivo results will be deferred to a later stage in the study.

2.5 Conclusion

This chapter presented the development of a new class of high-density multi-shank planar silicon probes. These probes have been designed to provide high density, large count, user customization, and optimized electrical and mechanical characteristics. This innovative design aims to facilitate precise neural recording and stimulation while minimizing tissue damage during the implantation process.

The incorporation of fine shanks, small stiffeners, and unique cross-sectional designs contributes to an improvement in mechanical performance without the need for thicker shanks. These probes exhibit remarkable straightness, compact cross-sectional areas, and flexibility, mitigating potential tissue damage. The successful fabrication of 64-electrode probes with exceptional properties across various types has been realized.

Comprehensive electrical and mechanical characterizations, along with in-vitro tests for stimulation and recording, have been successfully conducted and validated by the modeled equivalent circuits. Additionally, chronic measurement tests have been performed on planar probes (for almost a year) to assess the durability of the recording sites.

This achievement lays the foundation for advancing our understanding and application of high-density multi-shank planar silicon probes in neuroscientific research.

Chapter 3 High-Density, Large-Count 3D Non-Planar Neural Probes

3.1 Introduction

Implantable electrodes with different shapes, designs, and materials have been widely used to study the brain by recording the electrical and chemical responses of neural structures and circuits. Neural recording and stimulation using micro-electrode arrays (MEA) have been an essential part of decoding neural networks. The ever-increasing need to map larger collections of neurons has pushed the scientific and engineering community to develop new technologies capable of allowing this, but with limited success and progress [27].

In the previous chapter, we discussed planar probes, which have recording sites located alongside a planar shank. While these probes are useful for many applications, they have limitations when it comes to studying three-dimensional (3D) brain volumes. Non-planar or out-of-plane probes, on the other hand, are composed of out-of-plane needles or shanks positioned next to each other. This allows them to record or stimulate neurons with the recording sites at their tips or alongside their shanks [90].

Non-planar probes are particularly useful for recording from hard-to-reach regions of the brain or regions with sophisticated geometries. Multiple sites are distributed along the length of several shanks, which are themselves distributed along an area of tissue. They can be used to study complex neural networks in 3D, providing valuable insights into brain function and connectivity.

When fabricating out-of-plane neural probes, three important features are desirable. First, it is crucial to have customizable needle density. This means being able to adjust the density of the

neural probes when using them in different regions. For instance, in areas with highly populated or densely packed neurons, using high-density probes is preferred. High-density microneedle electrodes, with their high-channel counts, can better discern complex neural signals from densely populated neurons or nerve fibers. This increased selectivity enables more sophisticated functions of brain-machine interfaces [27, 90, 91]. Additionally, having arbitrary needle density on the same array provides neuroscientists with more flexibility in effectively reaching their desired regions.

A second desirable feature is a high-count array of neural probes. As highlighted by Stevenson & Kording [25], the number of simultaneously recorded neurons has approximately doubled every 7 years since the 1950s. The number of available sites in neural probes for simultaneous recording has been doubled almost at the same rate since the 1930s. This demonstrates that our understanding of neural circuits has been limited by a lack of suitable multi-channel probes. This limitation is partly driven by available fabrication technologies and materials.

Great progress has been made during the past two decades, especially with the advent of new fabrication capabilities. With a high-count neural probe array, researchers gain access to a larger population of neurons simultaneously, allowing for the collection of larger datasets during recording. This provides more information in a shorter amount of time, which is crucial for studying complex neural processes.

Additionally, the redundancy provided by having more probes ensures that even if some probes fail or become damaged, the overall functionality of the neural probe array is maintained. Overall, the ability to customize the size of the neural probe array, whether large or small, is more desirable for neuroscientists as it allows for more flexibility in experimental design and data collection [92].

A third important feature is a smaller cross-sectional area when implanting the probes into brain tissue. This is crucial for minimizing injury to the brain tissue and reducing the foreign body response, which is particularly important for chronic or long-term studies [93]. A smaller cross-sectional area reduces the physical footprint of the probe, which in turn reduces the amount of tissue displacement and damage during implantation. This not only helps to preserve the integrity of the brain tissue but also minimizes the inflammatory response and the formation of scar tissue around the probe [93-95]. As a result, the probe is more likely to maintain its functionality over extended periods, allowing for more accurate and reliable long-term recordings of neural activity [95-97].

Other important features are customizable length and tip sharpness. Microneedles of different lengths are useful for studying neurons at various depths in complex three-dimensional nervous systems, like the laminated cortex and peripheral nerves [98]. For electrical stimulation, microelectrode arrays with complex shapes are needed for specific applications, such as stimulating nerve fibers or the retina [99, 100]. This is because the central and peripheral nervous systems have multiple layers, each with unique functions [101-107]. Microneedles of different lengths can stimulate these layers individually or together. Additionally, the shape of microneedles is important for delivering drugs or biomolecules to specific areas within the nervous system. A sharp tip is preferred for neuro-prosthetic devices to minimize damage during insertion into neural tissues [107-110] and to reduce the foreign-body response [40, 42]. All these key features of out-of-plane neural probes are depicted in Figure 3.1.

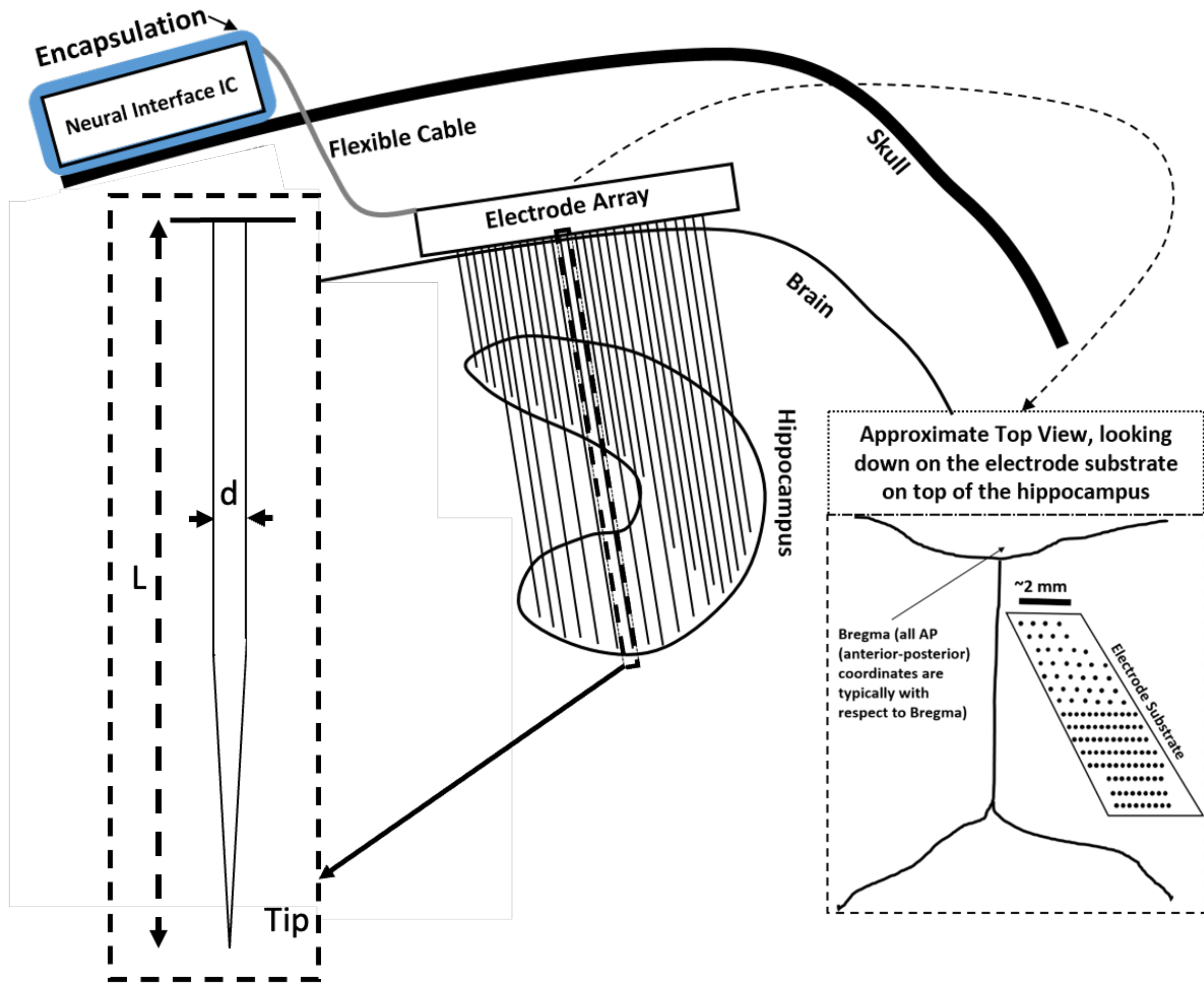

Figure 3.1: Illustration of the key features of out-of-plane neural probes, including customizable needle density, high-count array, smaller cross-sectional area, customizable length, and tip sharpness. These features are crucial for enhancing the functionality and effectiveness of neural probes in studying complex neural processes and developing advanced brain-machine interfaces [23].

The development of non-planar probes has seen significant progress in recent years, with various fabrication methods being employed. One conventional method for creating out-of-plane probes is the Utah electrode array, which involves a series of steps, including dicing of silicon (Si) wafers and wet etching of resulting Si pillars [58, 99]. The dicing kerf lines (~ 100 μm) play a crucial role in determining both the dimension and density of microneedle arrays, making it

challenging to obtain a high-density array with a small feature size. As of now, the Utah arrays have achieved a record of 25 electrodes per mm^2 [111].

Another technique for creating out-of-plane arrays involves fabricating Michigan-style probes and assembling them next to each other to form 3D arrays. The Michigan group demonstrated one of the earliest 3D probes of this kind for both recording and stimulation [112]. This 3D system consisted of multi-shank silicon probes fabricated using planar microfabrication techniques and then assembled onto a platform with precisely arranged features to accommodate and align the planar probes into the final 3D structure. Rios [32] has also demonstrated a large-count 3D probe by stacking several individual planar multi-shank and multi-site silicon probes, instead of assembling them through a platform as the Michigan group did. The Michigan system also included electronics on individual probes or on the supporting platform for signal amplification and multiplexing to enable recording from all sites simultaneously. These 2 systems separated in time by almost 20 years represent true 3D recording probes. The physical structure of both of these probes is similar to the Utah array, that is the probes consist of many needles (shanks), with the difference being that the Michigan/Caltech needles can each support a linear array of recording sites, whereas the Utah needles could each support only one recording site at the tip of each needle [32, 111].

Another method is to make a bundle of microwires [33]. The microwire bundle was composed of polymer-insulated metal microwires, each with exposed recording sites at their distal ends and interconnecting tips at their proximal ends. Electrical contacts were established by pressing the proximal interconnecting tips onto a CMOS amplifier array. To minimize potential damage to brain tissues, the size of the metal microwires was reduced to a diameter range of 5–25μm. This configuration enabled the simultaneous recording of neural activity from hundreds of

neurons in the motor cortex of mice in an in vivo setting. However, a significant challenge arises from the unpredictable arrangement of microwires within each CMOS-bundle device. This poses difficulties in developing spike sorting algorithms that rely on positional information to accurately distinguish and classify neural activity [33].

Another technique involves assembling carbon fibers to create an array of neural probes, with each probe having a recording site at its tip [113]. However, this method requires a significant amount of labor and is very time consuming due to the intricate process of assembling the tiny carbon fibers.

Another approach utilizes etching ultra-deep holes in a silicon substrate and refilling them with multiple insulating and conductive layers, and was employed in the development of Sea of Electrode Arrays (SEA) probes [23]. This technique enables the creation of arrays of non-planar (out-of-plane) needles with high density and a large count, allowing for customization in array size, density, and distribution. Figure 3.2 (a-c) shows a cross-sectional view of the fabrication process for the SEA. The process starts with etching deep and narrow holes in a silicon substrate. The holes can be as deep as 500μm, have sharp tips, and act as molds that are refilled with materials that eventually form the arrays of the needles (Figure 3.2 (d)).

To form deeper holes, and thus longer needles (i.e., >500μm), it is necessary to first etch multiple wafers separately, then align and bond the through-etched wafers. The holes are then filled sequentially by insulating layers of silicon dioxide, silicon nitride, conductive polysilicon, and silicon dioxide and silicon nitride. The silicon substrate is then etched away, leaving behind the insulated probe needles (Figure 3.2 (e)).

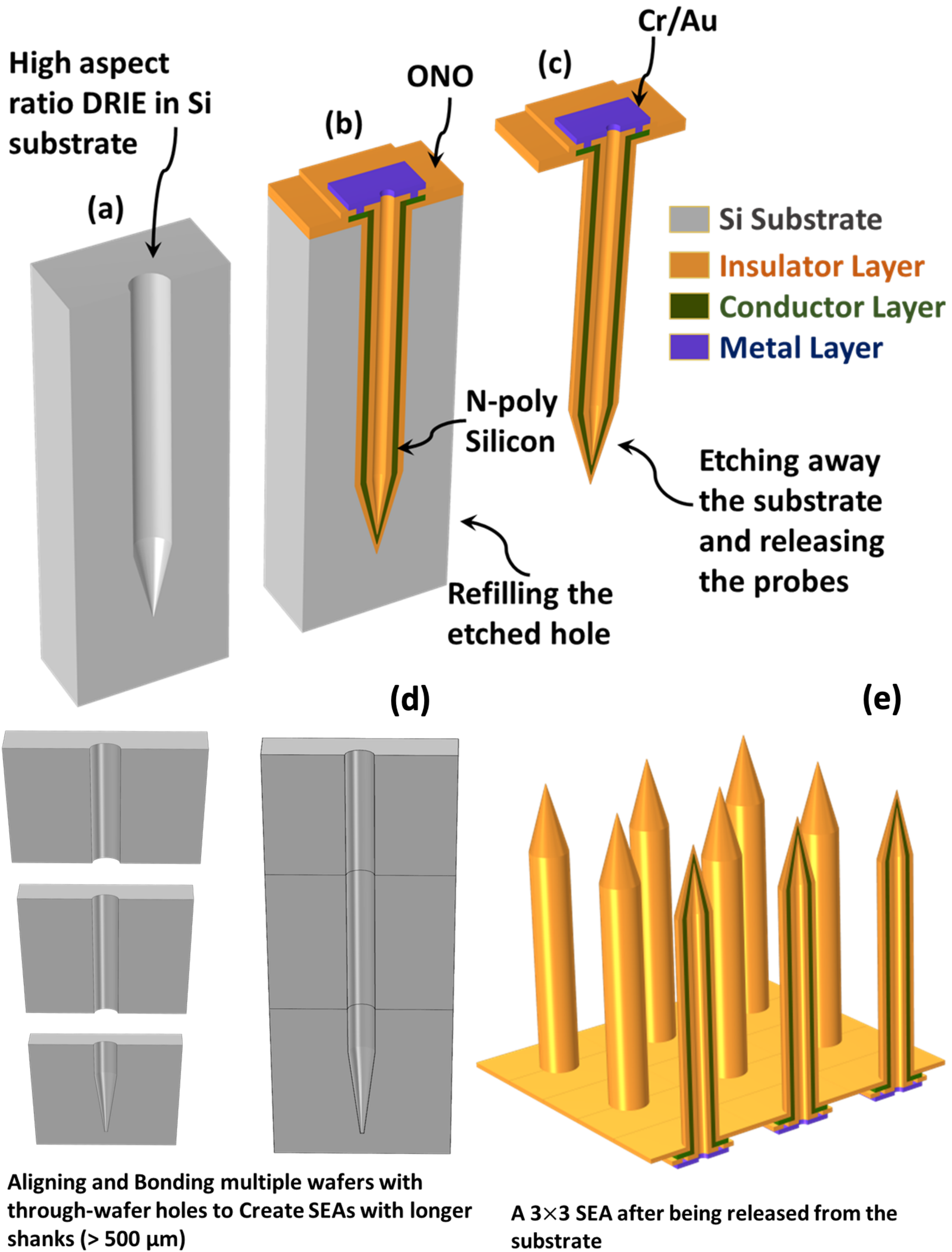

Figure 3.2: Fabrication steps of SEA before metallizing the tip, (a) DRIE of a hole on a Si substrate, (b) filling the hole with multiple insulator and conductor layers, (c) releasing the needle, (d) a 3×3 SEA, (e) Close-up of a tip.

The SEA arrays provide a great capability to create customizable arrays for desired applications. However, some limitations arose with this approach, which we address in this chapter. In Figure 3.2 (d), the bottom wafer, which will create the tips after releasing the shanks (tip wafer), contained smaller holes to create sharper tips.

However, when bonding these wafers, there were a few challenges. At the bonding point between the tip wafer and the rest of the wafers, a shoulder is shaped (Figure 3.3(a)). Additionally, at the bonding junction of other wafers, we also have inevitable misalignments due to the limitations of the alignment tools, potentially causing breaking points on the shank and discontinuity in the deposited conductive layer within the holes (Figure 3.3 (a)).

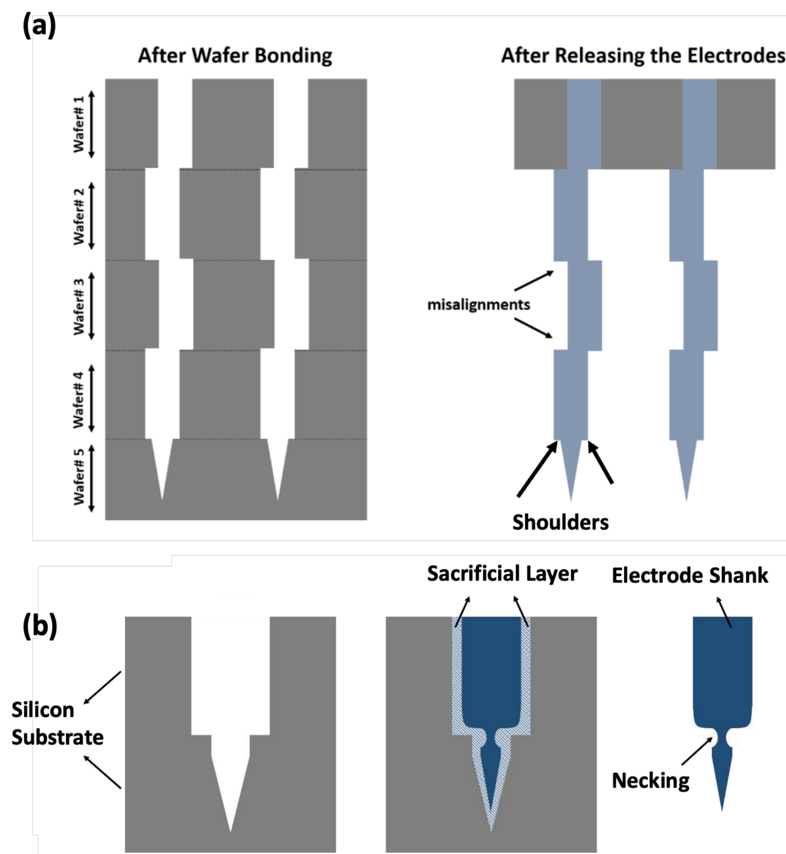

Figure 3.3: Some of the SEA array fabrication limitations: (a) Wafer bonding misalignment (left) results in undesired shifts along the electrode shanks (right), (b) Formation of necking along the electrode shank at the bonded wafers junction [23].

Another limitation is called necking. A neck is formed at the junction of the conical tip part and the cylindrical shank base. This is due to the degraded LPCVD deposition conformality, as shown schematically in Figure 3.3(b). In other words, the sacrificial polysilicon layer is thicker around the step corners at the junction where two wafers are bonded. This results in a reduced diameter of the conical tip part at the junction where it is connected to the shank cylindrical base. The formed neck introduces a structural weak point that can potentially degrade the mechanical robustness of the electrodes during implantation and post-implantation.

Another concern in this method is tip metallization. In the SEA fabrication method, after etching the silicon substrate, the insulation at the tip of the needles should be selectively removed to expose the conductor layer and deposit the electrode site material, as shown in Figure 3.2 (e). In the SEA fabrication method, positioning the recording sites at the tips of the shanks is challenging as etching the insulator layer from the tips of the shanks with control on the size of the tips was challenging.

It was easier to etch the insulator layer from the shoulders of the shanks that were exposed by maskless directional dry etching. Therefore, in SEA arrays, the recording sites were formed on the shoulders of each shank instead of the tips of shanks. The size of the shoulders limits the size and control over this size of the recording sites.

Another challenge pertains to the packaging of SEA arrays, where the backends of the non-planar arrays require wire-bonding through a complex process with lengthy wires to a printed circuit board (PCB) for tests. Integrating a flexible cable with these arrays during fabrication makes it easier to handle them during measurements. Figure 3.4 shows all the discussed limitations and challenges with SEA array fabrication technology that we intend to address in this chapter.

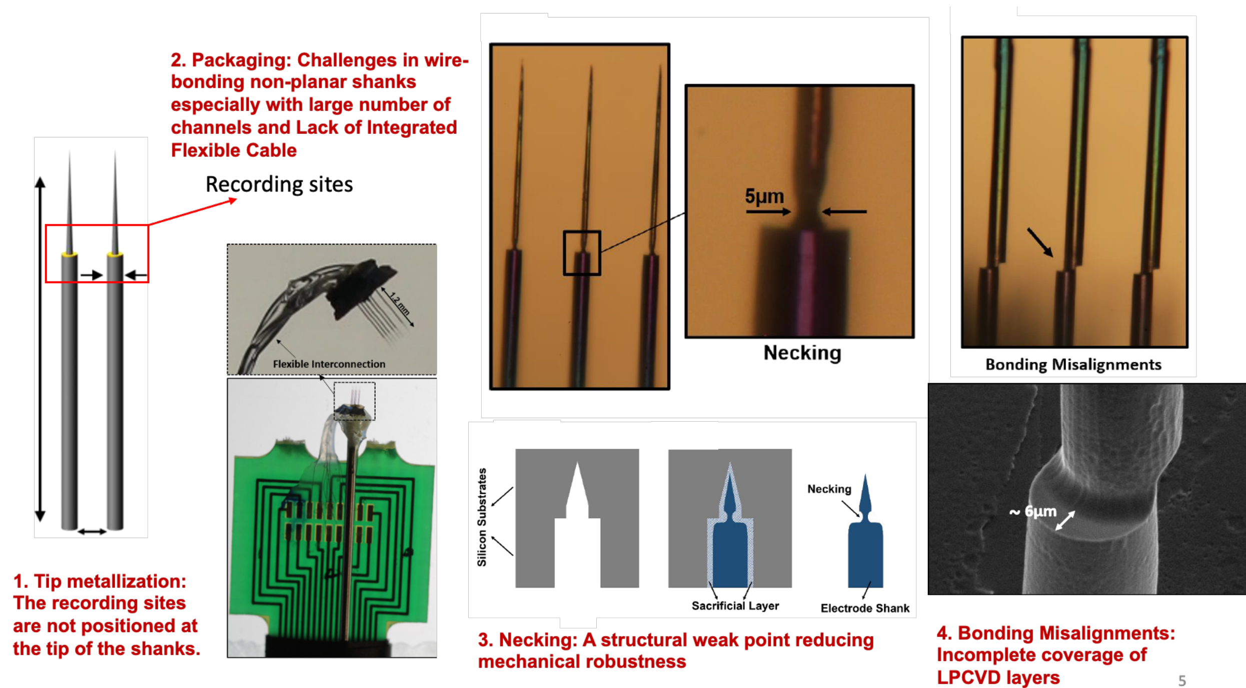

Figure 3.4: Major limitations and challenges of the SEA array fabrication technique [23].

In this chapter, we present a new method for creating non-planar neural probes with high electrode count and density. This technique, inspired by the SEA array fabrication method, overcomes several limitations of traditional methods while maintaining the ability to customize array size, density, and distribution. We introduce innovative techniques for creating high-count and high-density non-planar probes, including precise tip metallization for recording sites.

Our microfabrication technology produces ultra-small needles with diameters of 10 to 15µm and sharp tips, enabling precise recording and stimulating sites. We also propose packaging techniques to simplify array handling, integrating flexible cables and offering more efficient connections to connectors for testing and measurements.

The compact size of these probes allows for a higher density of recording and stimulating sites, enhancing spatial resolution. With shank gaps as small as 30µm and varying up to 80µm,

researchers can obtain detailed data from numerous neural cells or regions, facilitating precise brain function mapping.

The probes also offer customization options for shank separation, shank size, and shank distribution, tailoring them to specific experimental needs. Additionally, their small size makes them suitable for long-term implantation and chronic recordings, ensuring stable and reliable measurements for extended studies. The proposed technology for fabricating high-count, high-density non-planar probes with minimal footprints to be minimally invasive has several key features:

- Adjustable shank density: The shank density is adjustable, and we have demonstrated high-density arrays with separation as small as 30µm (400 electrodes mm^{-2}). This density is 16 times higher than Utah array densities, enabling a higher density of recording and stimulating sites within a small area, leading to enhanced spatial resolution.
- High-count arrays: The arrays are high-count, ranging from 10 x 10 arrays up to 100 x 100 arrays, and this technique is capable of creating higher count arrays as well.
- Recording sites at the tips: The recording sites are located at the sharp tips of the shanks, and this chapter proposes three novel techniques for their formation. This allows for precise control over the size of the recording sites, enabling researchers to tailor the probes to specific experimental needs.
- Integrated flexible cable: The flexible cable is integrated with other array components on the same wafer, eliminating the need for wire bonding on the

backends of the arrays to the PCB or non-integrated flexible cable. This simplifies the packaging process and reduces the risk of damage to the probe during assembly.

- Customizable length and distribution: Arrays with varying shank lengths can be achieved by creating holes with different diameters on the substrate as well as bonding multiple wafers. This allows for the customization of the array size and distribution.

All these key features are illustrated in Figure 3.5.

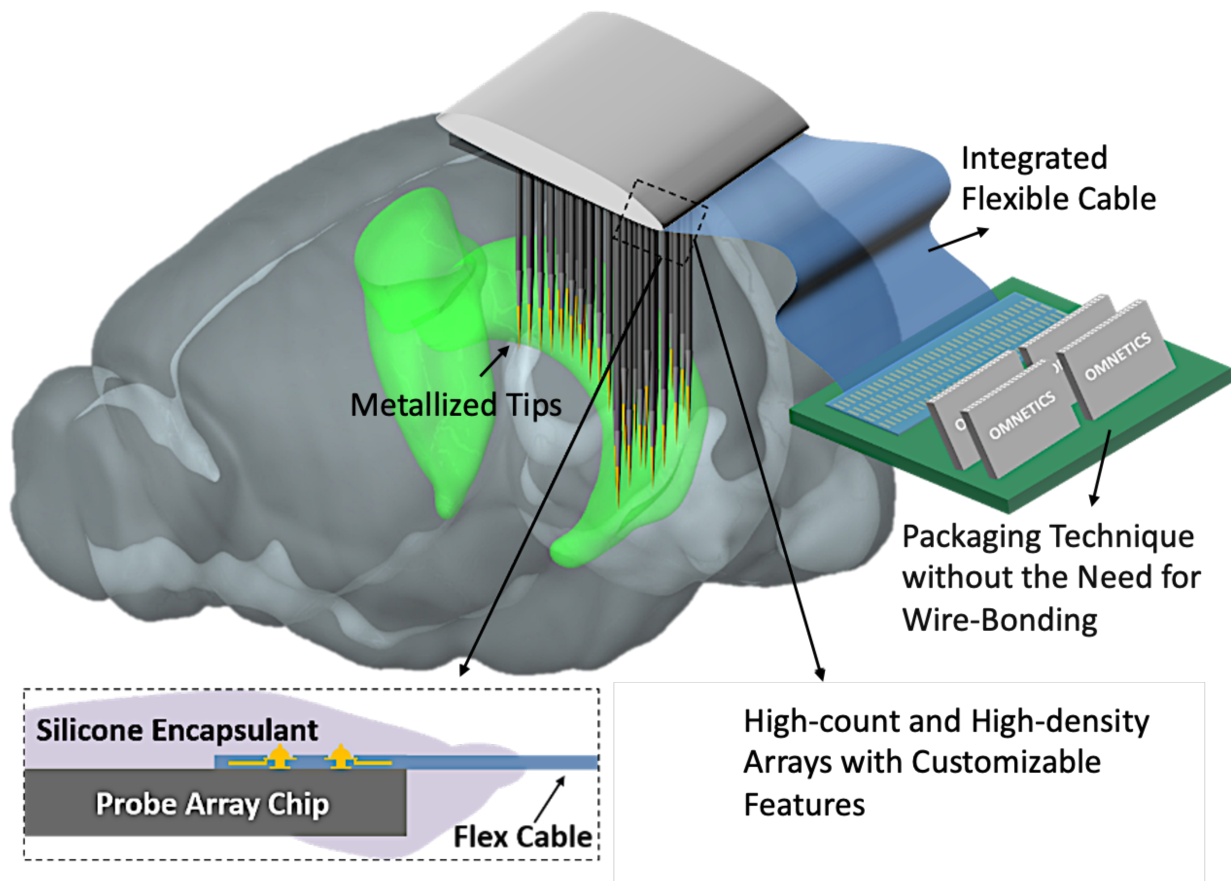

Figure 3.5: Illustration of the key features of the proposed technology for fabricating high-count, high-density non-planar probes with minimal invasiveness.

3.2 Microfabrication Process

The fabrication process begins by coating 4-inch double-sided polished silicon wafers with a 1μm silicon dioxide layer. This silicon dioxide film, along with 10μm SPR 220 photoresist, serves as a mask layer for etching ultra-high aspect ratio through-wafer holes using DRIE. These holes can either be created through-wafer etched from both sides of a silicon wafer (Double-Sided Etch (DSE) version) or non-through wafer, etched only from one side of the wafer (Single-Sided Etch (SSE) version). For the DSE version top-side holes measuring 20μm in diameter, while the bottom-side holes are 15μm. For the SSE version top-side holes measuring 20μm in diameter. The etching is done using ramped parameter DRIE process for both versions [114].

The hole sizes are customizable, and in the DSE version, we intentionally chose different sizes on each side for two specific reasons. First, this provides flexibility to accommodate potential misalignments during the alignment of the top and bottom masks. Second, the shank width is reduced closer to the tips, aiming to facilitate easier implantation.

Furthermore, the ultimate width of the shanks can be fine-tuned by depositing the preferred thickness of the sacrificial polysilicon layer in the subsequent steps. This sacrificial polysilicon layer will reduce the size of the hole so the final size of the shank could be smaller than the hole diameter.

3.2.1 Etching and Refilling Steps

In the DSE version, the holes are initially etched from the top side to achieve a depth of 360μm, while the holes from the backside are etched to a depth of 140μm (Figure 3.8). This process results in through-wafer holes with a 20μm diameter extending to a depth of 360μm, connected to 15μm diameter holes with a depth of 140μm. As mentioned before, this difference in

size between the top and bottom holes aims to resolve misalignment issues that could lead to weak points prone to breakage, as well as reduce the size of the shanks closer to the tips.

The contact aligner we employed for aligning the masks on both the top and bottom sides of the wafer had a minimum feature size resolution of 2µm. The difference in diameter between the holes on the top and bottom sides of the wafer in DSE version is 5µm, providing us with a 2.5µm tolerance for misalignment. In the SSE version, the first step involves etching ultra-deep holes from only one side with a diameter of 20µm and depth of approximately 400µm.

Figure 3.6 illustrates the SEM images of the etched holes with a 20µm diameter, exclusively etched from the top side of the wafer using a silicon dioxide and photoresist mask.

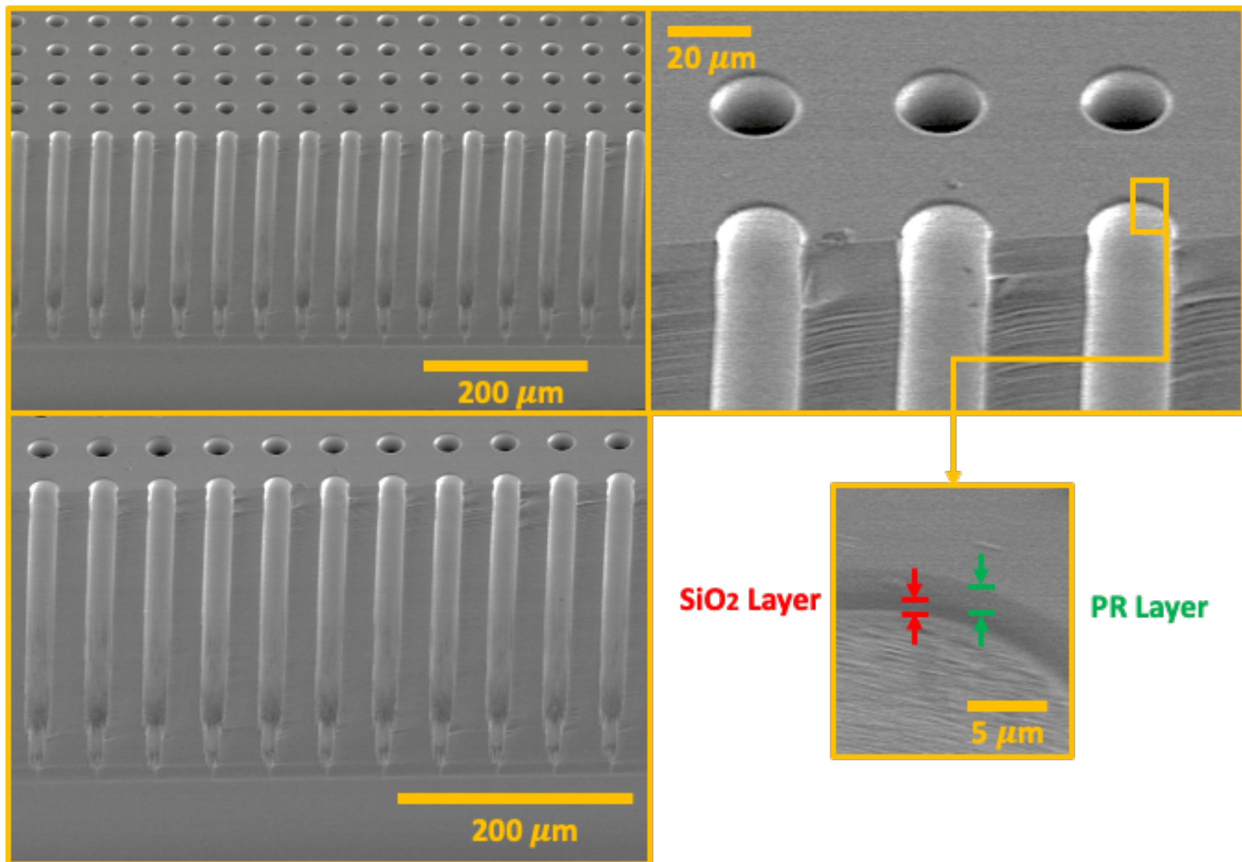

Figure 3.6: SEM images of etched holes with a 20µm diameter, exclusively etched from the top side of the wafer using a silicon dioxide and photoresist mask.

The next step after completing the high aspect ratio etching from both sides is to deposit a 300 nm thermal silicon dioxide on the etched wafer and then remove it to smooth any sharp edges (Figure 3.7). This step aids in tapering the sharp edge at the connection of the top and bottom holes, reducing the necking effect that could lead to breaking points—an issue encountered in the SEA method [23].

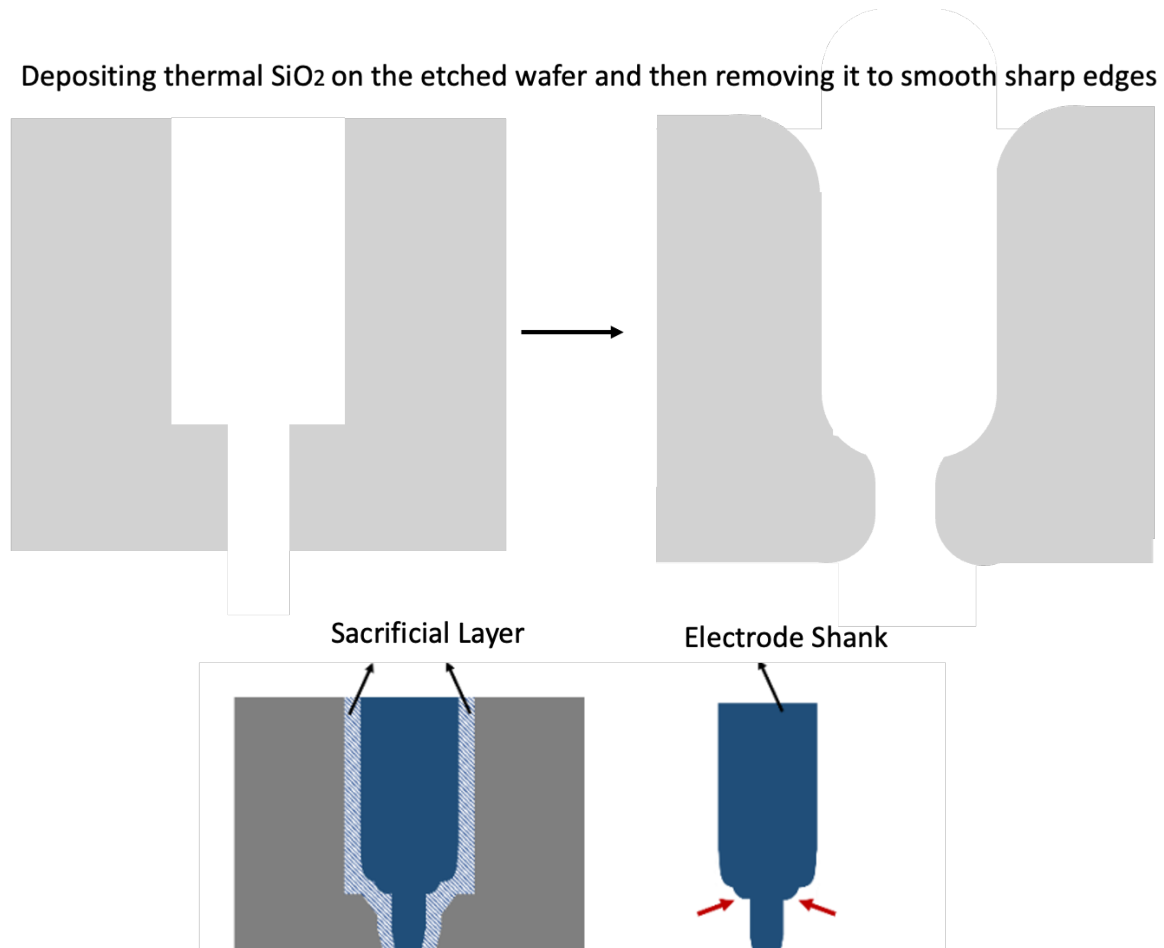

Figure 3.7: Smoothing the edges by depositing thermal silicon dioxide on the etched wafer and then removing it. This process helps with resolving the necking issue where holes with different diameters meet in the substrate.

The next step involves refilling the holes with different layers. The schematic of these steps for DSE version are shown in Figure 3.8. The initial step in refilling the holes is with an approximately 2µm thick polysilicon layer using LPCVD. This layer is referred to as the

sacrificial-poly layer because it will be etched away in the final step, which is the release of the shanks. It is used to decrease the diameter of the final shanks and finely adjust their size (Figure 3.8, step #2).

Subsequently, the second refilling layer is approximately 2µm thick, consisting of a stack of 5 layers: silicon dioxide/silicon nitride/silicon dioxide/silicon nitride/silicon dioxide (ONONO). This layer serves as the insulator layer and will be the outer layer of the shanks after being released (Figure 3.8, step #3). The thickness of these ONONO layers is 5600/1600/5600/1600/5600Å. Figure 3.9 shows SEM images illustrating the cross-section of the refilled holes following the deposition of the insulator layer, providing a 45-degree tilted side view of the top side of the substrate. As shown, the sharp edges of the holes are slightly smoothed, while the sacrificial layer is observed to be deposited with increased thickness in proximity to the hole openings.

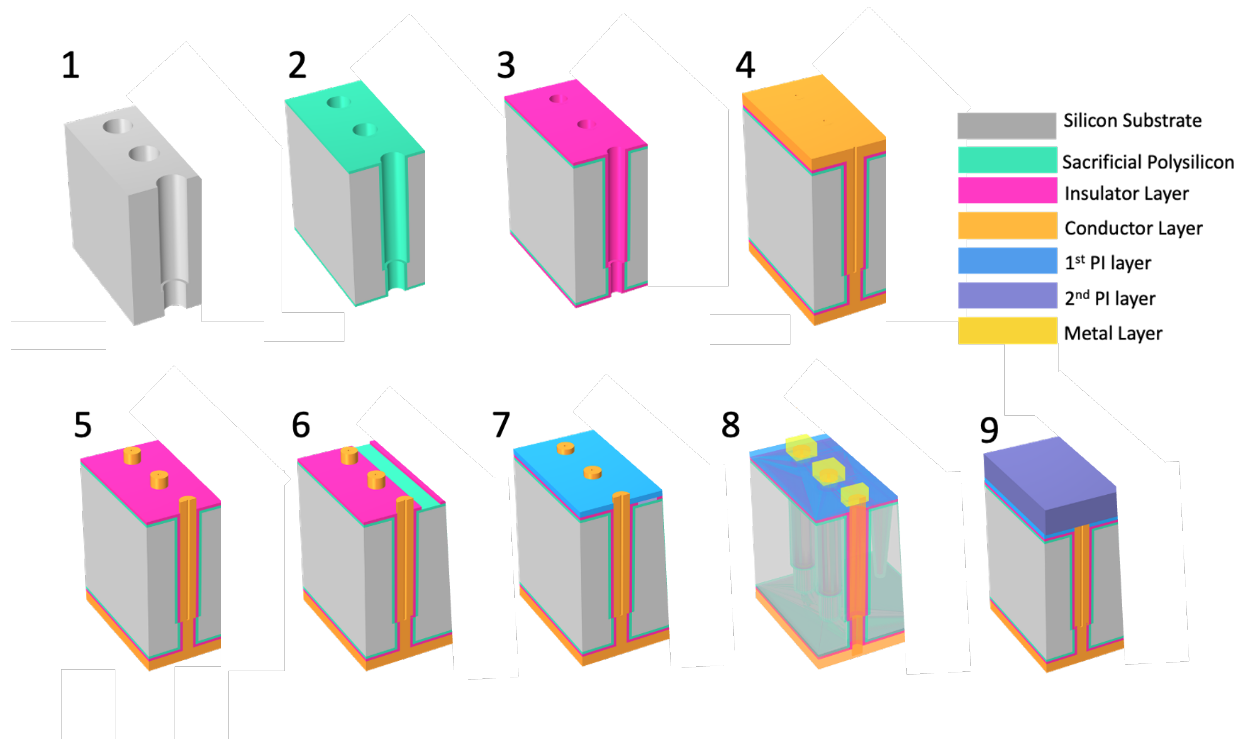

Figure 3.8: Illustration of the steps involved in the fabrication process of the DSE version of high-count, high-density non-planar neural probe arrays, including depositing different layers and refilling the holes.

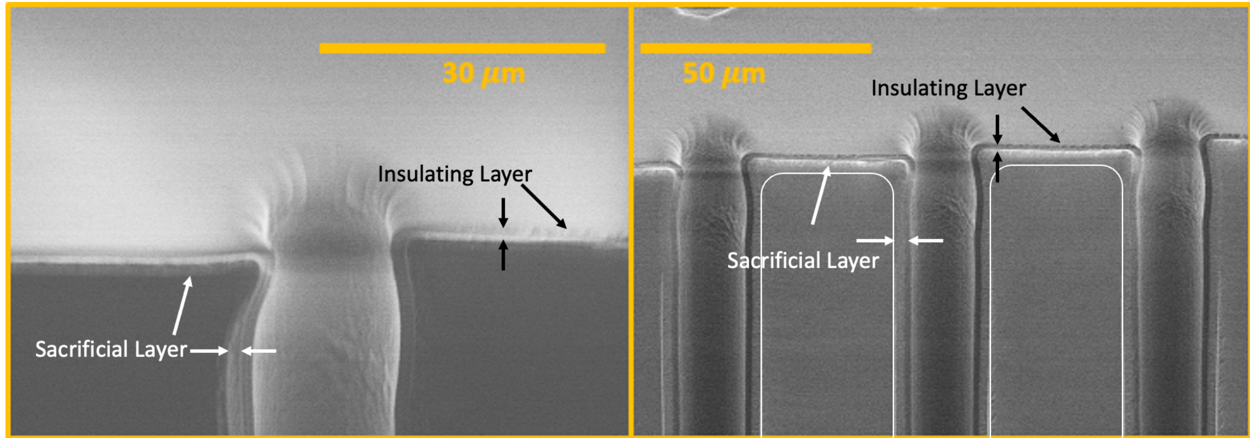

Figure 3.9: SEM images revealing the intricate cross-section of refilled holes post-insulator layer deposition, featuring a 45-degree tilted side view from the top side of the substrate. The sacrificial layer is indicated by white arrows, while the insulating layer is represented by black arrows.

The next step is to deposit a conductive layer, specifically n-doped polysilicon with a sheet resistivity of 40 Ω/sq, using LPCVD. This layer is deposited thick enough to completely refill or close the holes on the bottom side of the wafer, which initially have a diameter of 15µm (Figure 3.8, step #4). This side of the device is considered the tip side of the shanks after being released. It is crucial that the holes at the tip side of the shanks are closed to enable the tip metallization step for the DSE method in the future steps.

Therefore, a thick layer of 5µm n-doped polysilicon was deposited using the LPCVD method (Figure 3.8, step #4). Figure 3.10 shows the SEM images of the refilled holes on the topside of the wafers, shown from both 45-degree tilted side view and top view. As it is shown the holes are not completely refilled but they are closed on the top.

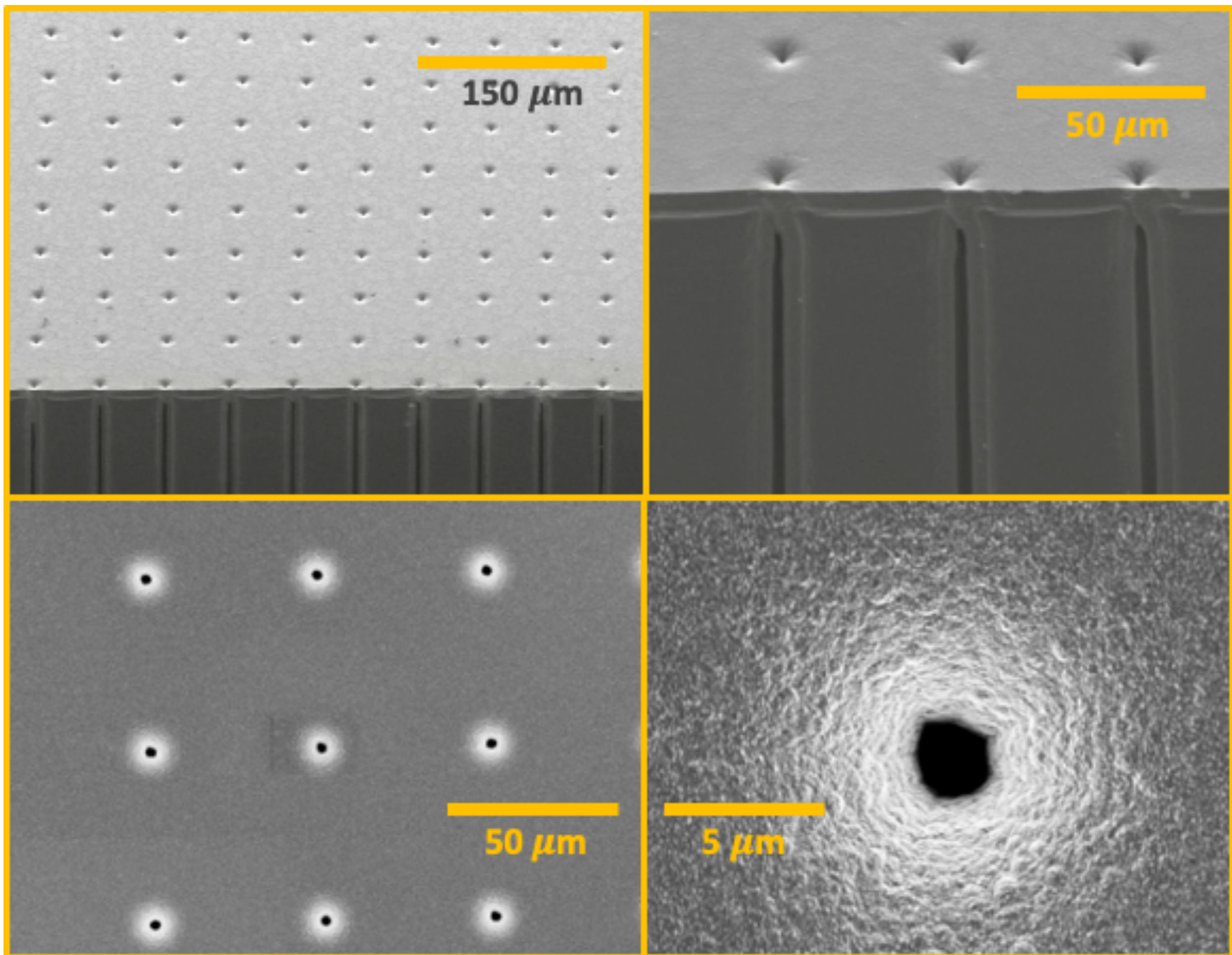

Figure 3.10: SEM images of the refilled holes on the topside of the wafer, shown from a 45-degree tilted side view and a top view.

The holes on the bottom side of the wafers, intended to be closed to finally shape the tip side of the shanks, are almost completely closed with multiple LPCVD layers, as shown in Figure 3.11. While they are nearly refilled, very small indentations are formed at the location of each hole, leaving an opening of < 80 nm at the bottom of each indentation, which is negligible.

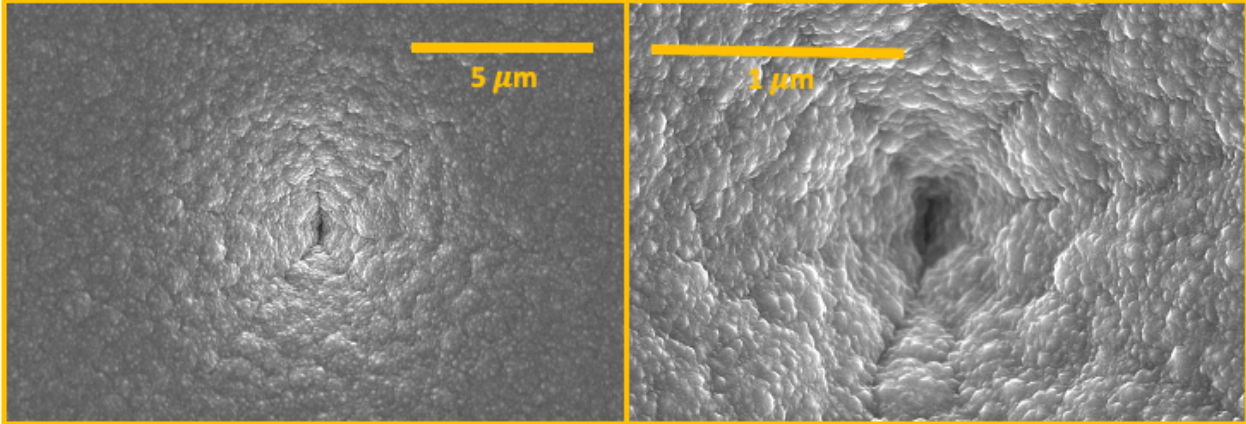

Figure 3.11: SEM Images of LPCVD Refilled Holes on the Wafer's Bottom, illustrating near-complete refilling, subtle indentations, and negligible nanometer-scale openings at the base of each.

Figure 3.12 shows a cross-sectional view of a refilled hole, revealing multiple LPCVD layers. The top images depict cross-sectional views of the shank layers at the midpoints, while the two bottom images show the upper side of the shanks. All images are presented in a 45-degree tilted side view.

After completing the refilling process, the subsequent phase consists of two key steps. Initially, the conductive segments of the shanks undergo separation. This is achieved by using lithography and plasma etching on the n-doped polysilicon layer, thereby isolating the backends of the shanks (Figure 3.8, step #5).

Following this, the array dies are separated from each other by patterning the insulator layer beneath the conductive layer (Figure 3.8, step #6) using RIE. Figure 3.13 presents optical microscopic images, while Figure 3.14 shows SEM images, providing a detailed view of the precise patterning of the conductive polysilicon layer and the subsequent patterning of the insulator layer.

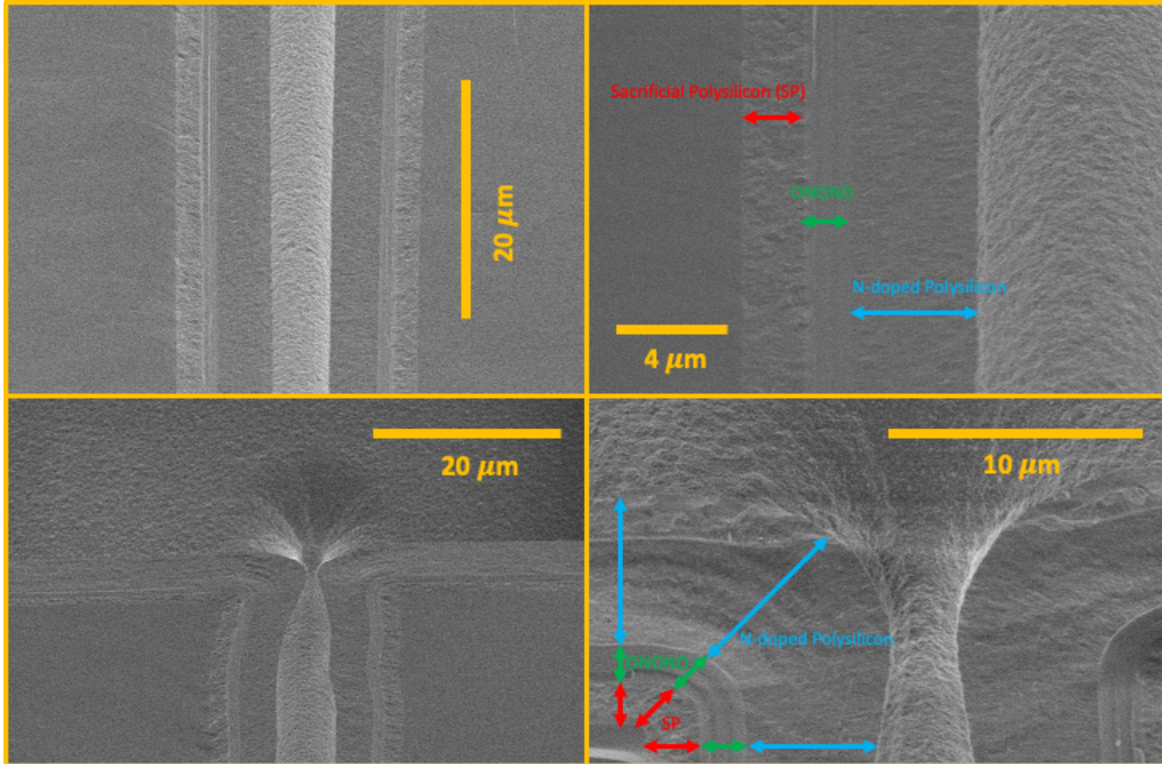

Figure 3.12: SEM images showing the inside of LPCVD refilled holes, illustrating detailed layers within shanks at midpoints and the top points of the shanks—all captured in a 45-degree tilted side view.

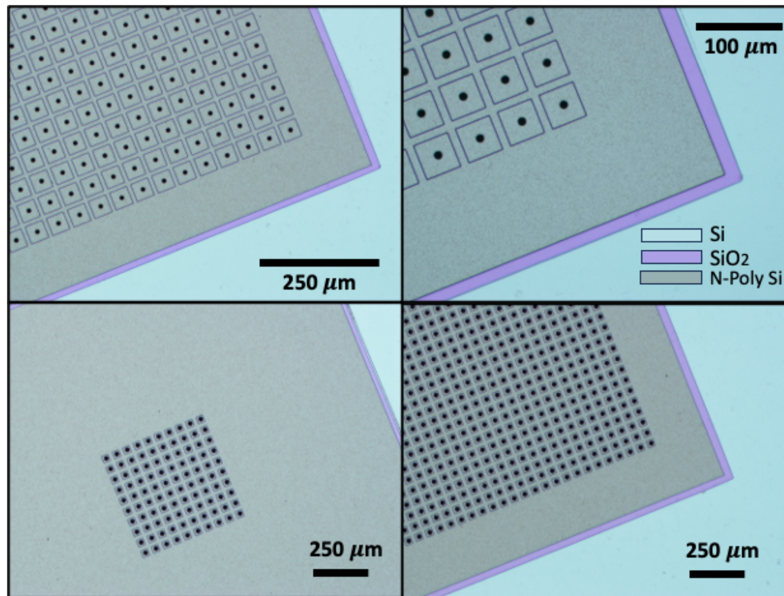

Figure 3.13: Optical microscopic images showing the etched ONONO layer and the separated conductive layers of individual shanks in each array.

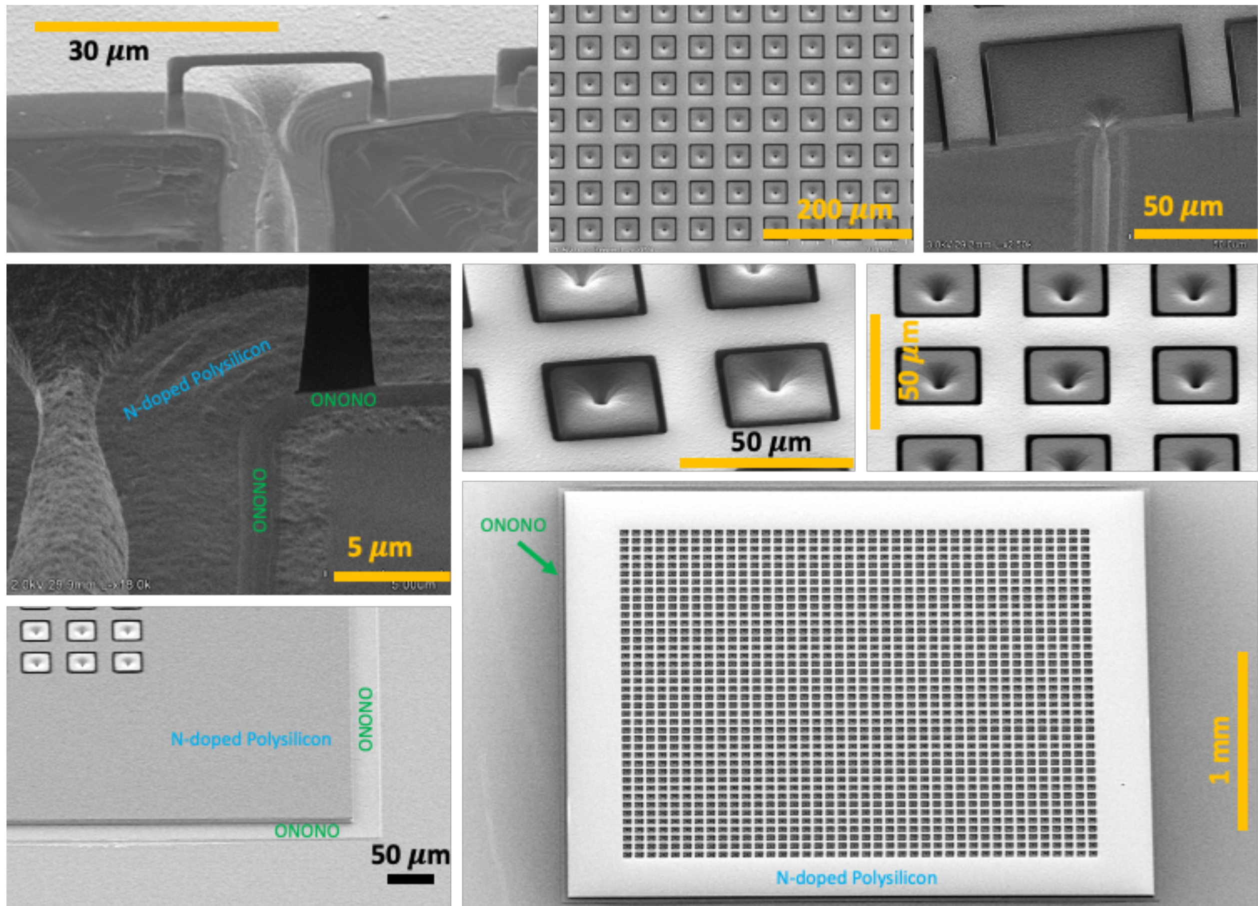

Figure 3.14: SEM images depicting the etched insulator layer and the separated conductive layers of individual shanks in each array.

3.2.2 Integrated Flexible Cable

In the subsequent step, a 3µm layer of Polyimide (PI 2611) is deposited and annealed. This serves a dual purpose: it isolates the conductive segments of the shanks from each other and shapes the integrated flexible cable and backends of the probes. Continuing, the process includes the patterning of the first Polyimide (PI 2611) layer.

This step is essential for creating the cables and backends. Additionally, openings are strategically crafted on the base of the arrays to expose the conductive layer of the shanks (Figure 3.8, step #7). Next, the metallization process takes place. This critical step enables access to the

conductive layer of each individual shank, facilitating the creation of metal lines on the flexible cables and establishing contact pads on their backends (Figure 3.8, step #8).

The metallization is done using a lift-off process, utilizing a stack of Cr/Au with a thickness of 300/9000Å. SEM images of the metal lines and contact pads at the base of the arrays are illustrated in Figure 3.15. As depicted, the number of lines connected to each array ranges from 16 to 128 channels, connecting to arbitrary channels in different locations within each array.

The array sizes vary from 10×10 to 100×100, featuring very small pitches ranging between 50µm or 100µm. Extracting signals from all channels using an integrated flexible cable necessitates much larger cables. To accommodate diverse arrays with their cables on a 4” wafer, we opted to connect up to 128 channels, thus showcasing the feasibility of this approach to create high-count and high-density non-planar arrays.

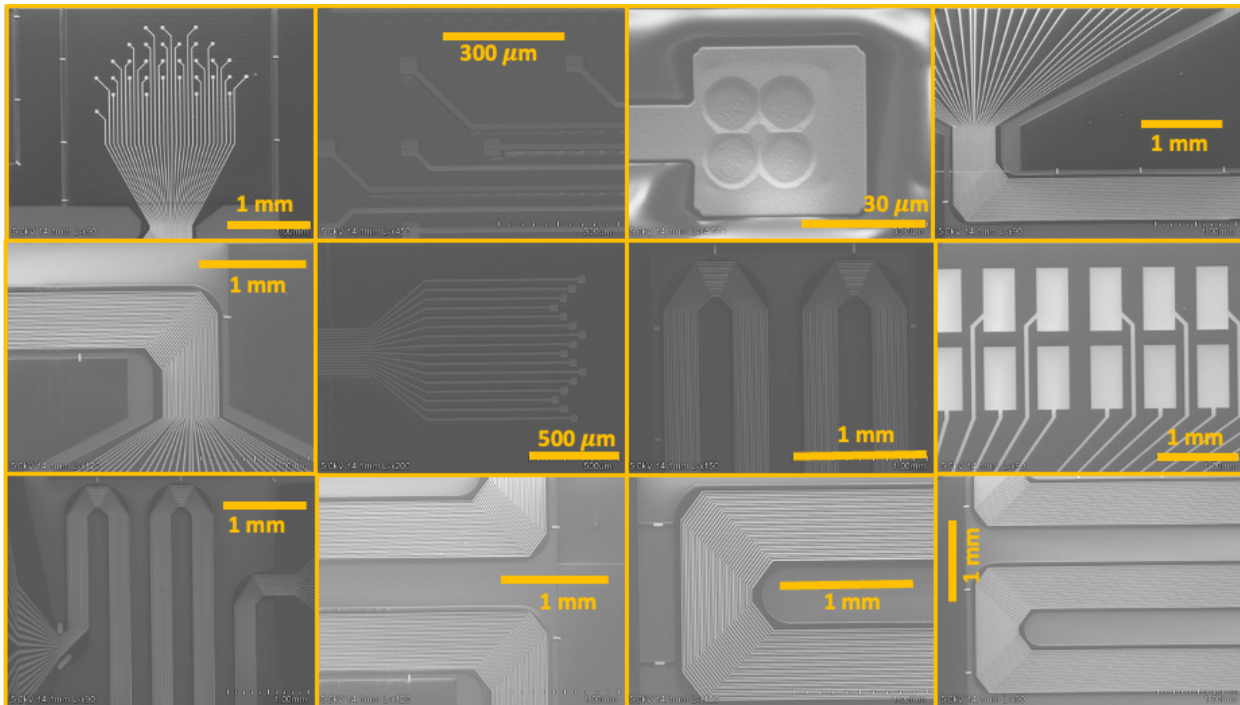

Figure 3.15: SEM images of metal lines and contact pads at the base of arrays, highlighting the varying channel connections (16 to 128) and array sizes (10×10 to 100×100) with small pitches (50µm or 100µm).

Optical images capturing different components of the devices after metallization are presented in Figure 3.16. This includes metalized contact pads on the array bases, metal lines on the patterned first PI 2611 layer, and the connected backend to the flexible cables. The metal lines have a width of 12µm and are separated by an 8µm.

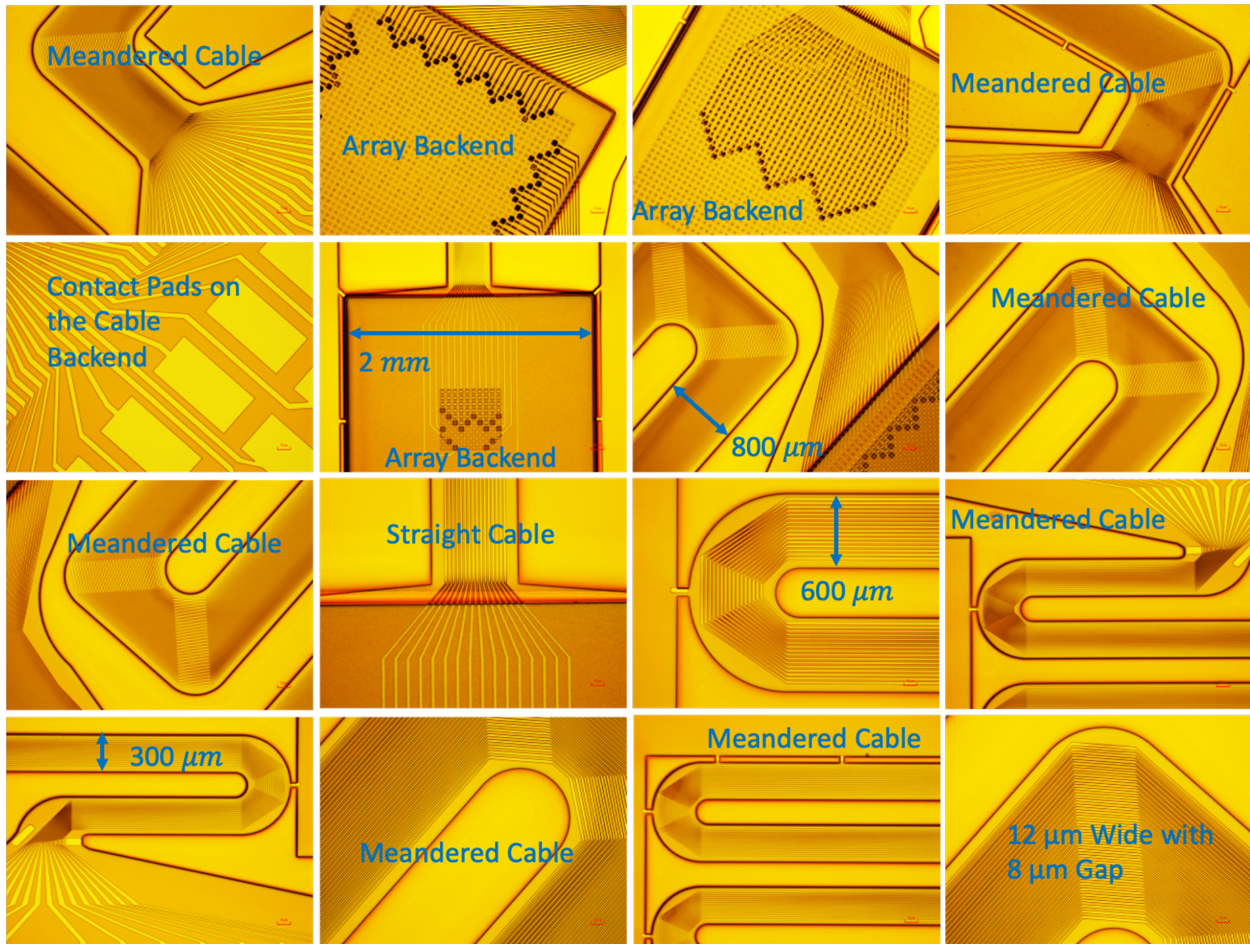

Figure 3.16: Optical images showcasing key components, including metalized contact pads on array bases, metal lines on the patterned first PI 2611 layer, and their connection to the backend leading to flexible cables.

After metallization, a 12 µm-thick layer of Polyimide (PI 2611) was deposited, annealed, and patterned. This layer serves to isolate the metal traces, creating a sturdy base for the arrays, and enhancing the durability of the integrated flexible cables and their backends (Figure 3.8, step #9).

Figure 3.17 presents optical images of key components following the deposition and patterning of the second PI 2611 layer. This includes isolated metal lines and contact pads on array bases, as well as metal lines on the flexible cables. The images also illustrate the openings of contact pads on the backends of the cables, designed for subsequent connection to connectors using solder.

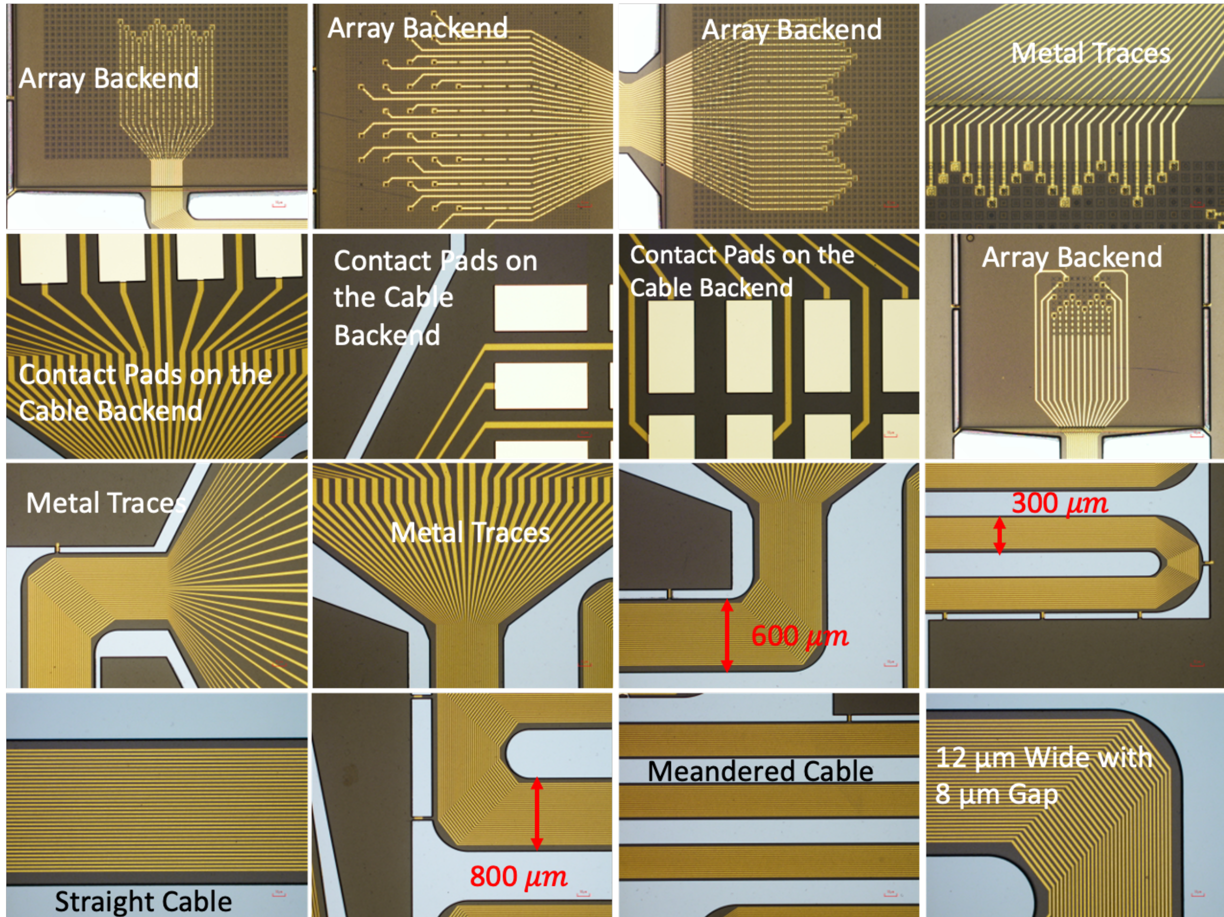

Figure 3.17: Optical images showcasing insulated metal lines, metalized contact pads on array bases, and metal connections to backends after the patterning of the second PI 2611 layer.

Figure 3.18 provides an overall view of all the devices and their components processed on the top side of the substrate. There are a total of 90 arrays on each wafer, varying in sizes such as $10 \times 10, 40 \times 40$, and 100×100, featuring different pitches of 50 μm or 100 μm. To optimize

space utilization on the wafer and accommodate more samples, the connected flexible cables are meandered. Additionally, two samples on each wafer feature straight flexible cables. Note that a significant amount of area on the wafer seems to be blank. The unused areas are where Omnetics connectors will be eventually placed. In future systems, Omnetics connectors could be replaced by IC chips, or by much smaller high-density connectors, thus saving valuable real estate on the wafer.

The cable width ranges from 300 μm to 600 μm in the meandered version and is 800 μm in the straight version. At the center of each wafer, there is a sample with 128 channels connected to four cables, extracting signals from four sides of the array. Some test arrays are not connected to flexible cables; however, they are equipped with contact pads on their bases for future wire bonding, soldering, or mechanical connection to connectors.

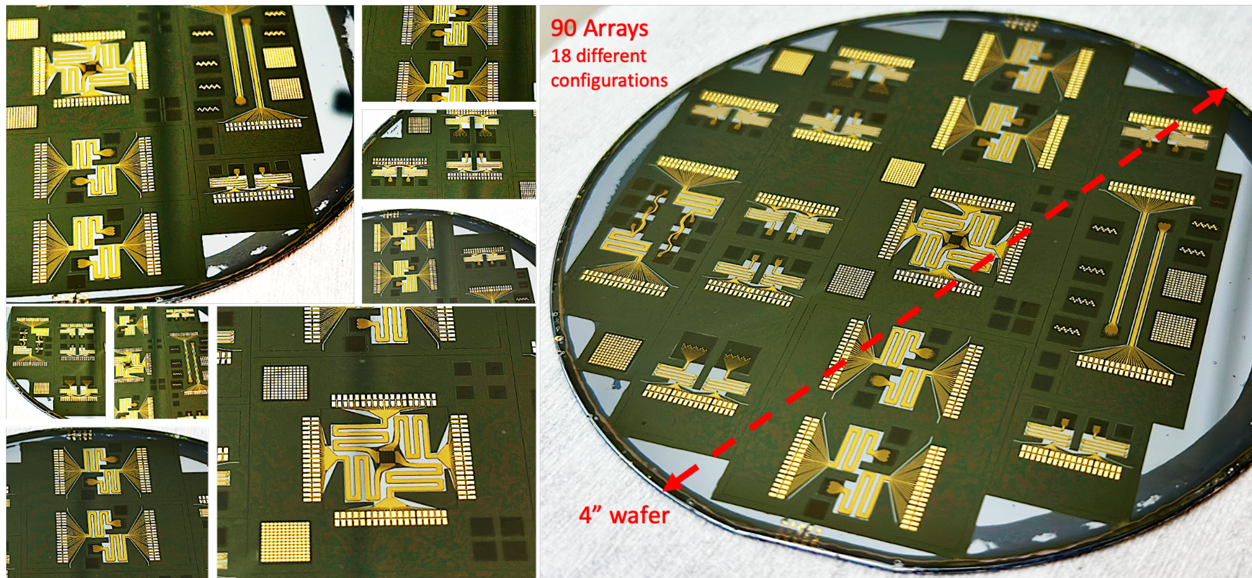

Figure 3.18: Overview of devices and components fabricated on the top side of the Si substrate in DSE version.

The fabrication process of the SSE version before the tip metallization is quite similar to the DSE approach. The first step involves etching ultra-deep holes from only one side with a diameter of 20µm and depth of approximately 400µm (see Figure 3.19, step #1).

Subsequently, these holes are refilled with a sacrificial polysilicon layer of a desired thickness, facilitating fine adjustment of the final size of the shanks in the array (Figure 3.19, step #2). Here, the thickness of the sacrificial polysilicon layer is approximately 2µm.

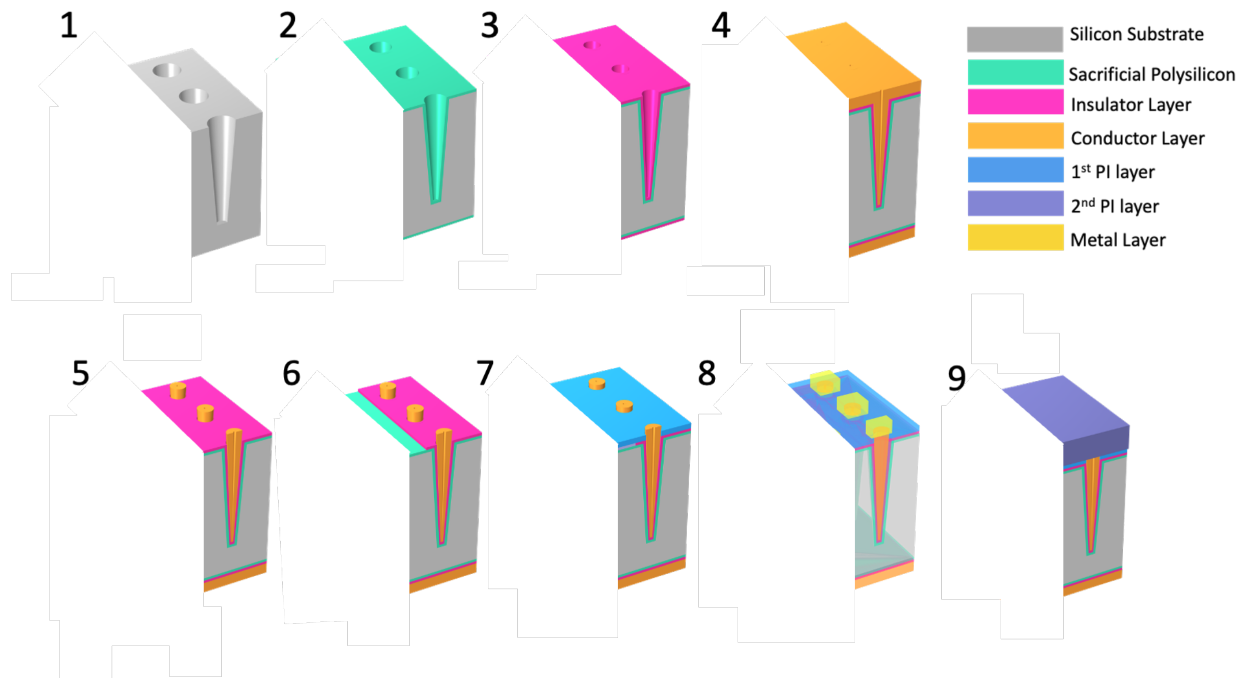

Figure 3.19: Illustration of the steps involved in the fabrication process of the SSE version of high-count, high-density non-planar neural probe arrays, including depositing different layers and refilling the holes.

The next step involves refilling the holes with insulator layer, serving as the outer coating of the shanks (Figure 3.19, step #3). The insulator layer is approximately 2µm thick, consisting of a stack of 5 layers: silicon dioxide/silicon nitride/silicon dioxide/silicon nitride/silicon dioxide (ONONO). The thickness of these layers is 5600/1600/5600/1600/5600Å.

Following this, a thick layer of 5µm n-doped polysilicon is deposited to nearly refill the holes, functioning as a conductive layer (Figure 3.19, step #4). All depositions, including sacrificial polysilicon, the ONONO stack, and n-doped polysilicon, are executed using the LPCVD method.

Afterward, the conductive layer undergoes a patterning process to ensure it remains only at the back end of each shank, effectively isolating the channels by separating the conductive parts of the shanks from each other (Figure 3.19, step #5). Following this, the insulator layer is patterned to isolate the arrays from each other and to eliminate it from the locations of the flexible cables (Figure 3.19, step #6). Next, the first polyimide layer (PI 2611) is deposited with a thickness of 3µm, annealed, and then patterned to create flexible cables. This layer also extends over the base of the arrays, featuring strategically placed openings to expose the patterned conductive layer for the subsequent metallization step (see Figure 3.19, step #7).

The subsequent step involves depositing a Cr/Au layer with a thickness of 300/9000Å and patterning it using the lift-off method to metallize the contact pads at the base of the arrays. Simultaneously, this step creates metal lines on the integrated flexible cables and their connected backends (see Figure 3.19, step #8). Following this, the second polyimide layer (PI 2611) is deposited with a thickness of 12µm, annealed, and patterned to form flexible cables. This layer also extends over the base of the arrays (see Figure 3.19, step #9).

3.2.3 Tip Metallization and Releasing Shanks

The next step, following the patterning of the second PI 2611 layer, involves the tip metallization process. In the DSE approach, initially, we need to pattern the n-doped polysilicon on the bottom side of the wafer, selectively removing it from all areas except the tip locations of the shanks. For this step, the substrate was flipped and patterned from the backside.

We used a mask with holes of 5 µm diameter at the tip locations (Figure 3.20, step #10), and following this step, we dry etched the underlying ONONO layer, keeping it only around the tip locations (Figure 3.20, step #11).

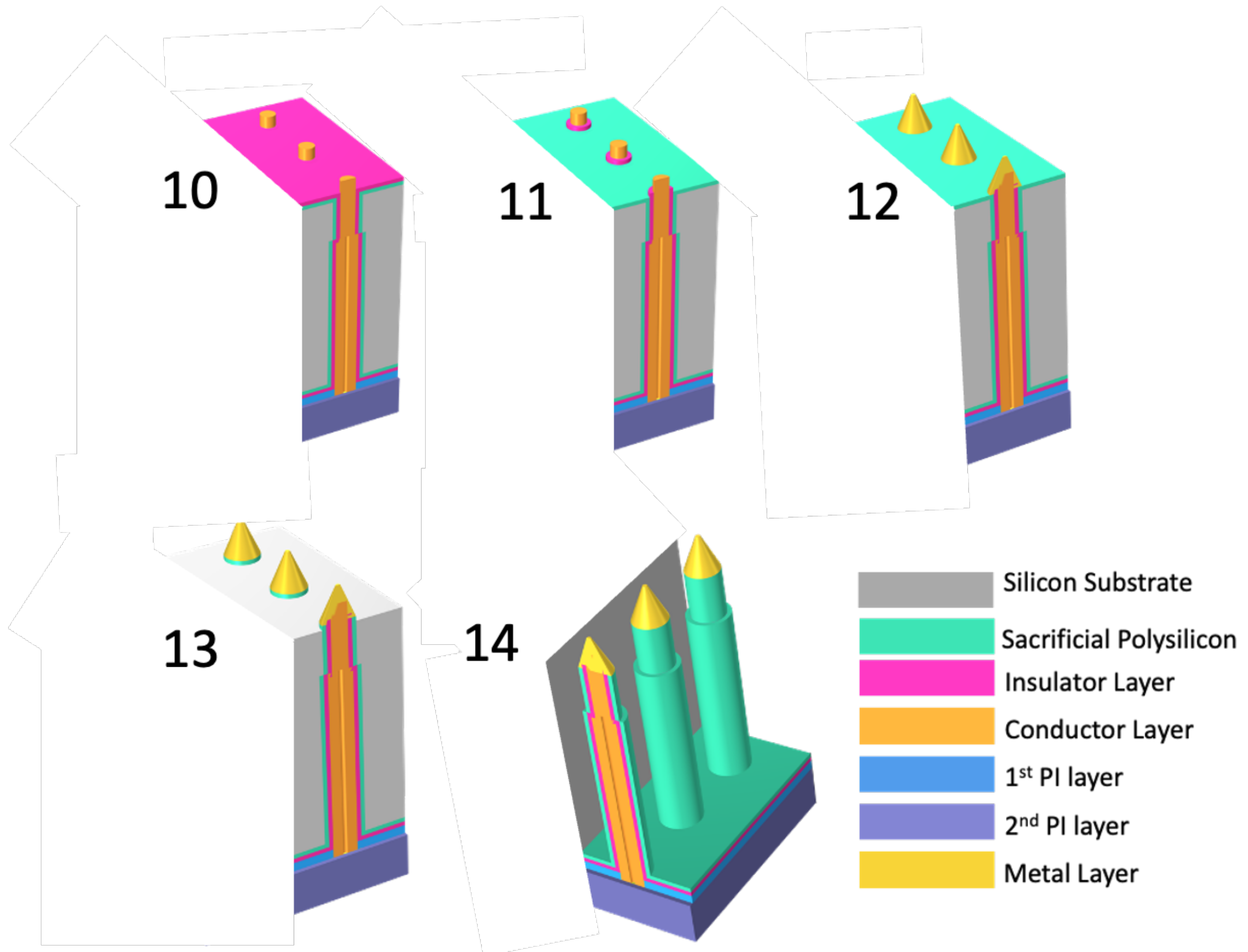

Figure 3.20: Tip metallization process in DSE followed by releasing the shanks.

Up to this point, all alignments and exposures have been conducted using a contact aligner. However, after this step, we must use proximity mode exposure with a spacer to prevent contact with the protruding tips, avoiding potential breakage and damage. The metallization of the tips is achieved through sputtering 300/3000Å Ti/Pt, followed by a lift-off process, wherein the image-reversed SPR 220 with negative slopes serves as a mask for better results.

To make sharper tips, we electroplate gold on the tip locations and then perform isotropic etching with a mask on top of the gold-plated tips to create sharp metallic tips (Figure 3.20, step #12).

Figure 3.21 provides a detailed illustration of the electroplating and tip sharpening process. The top image shows a cross-sectional view, while the bottom image shows a top view. After the lift-off process, a 100/1000Å layer of Cr/Au is sputtered on top to cover all surfaces, serving as the seed layer for electroplating. Next, a 10µm AZ 12XT photoresist is applied as a mask layer, leaving only openings above the tips for electroplating. In the second step, the photoresist layer is removed, and a new 10µm AZ 12XT photoresist layer is applied and patterned.

The new mask is intentionally shifted for 2µm, allowing for the creation of sharp crescent-wedge-shaped tips with a high effective surface area. The gold is then isotropically wet-etched in step 4, which also etches the seed layer. Finally, in step 5, the photoresist layer is stripped, leaving behind the electroplated sharp tips.

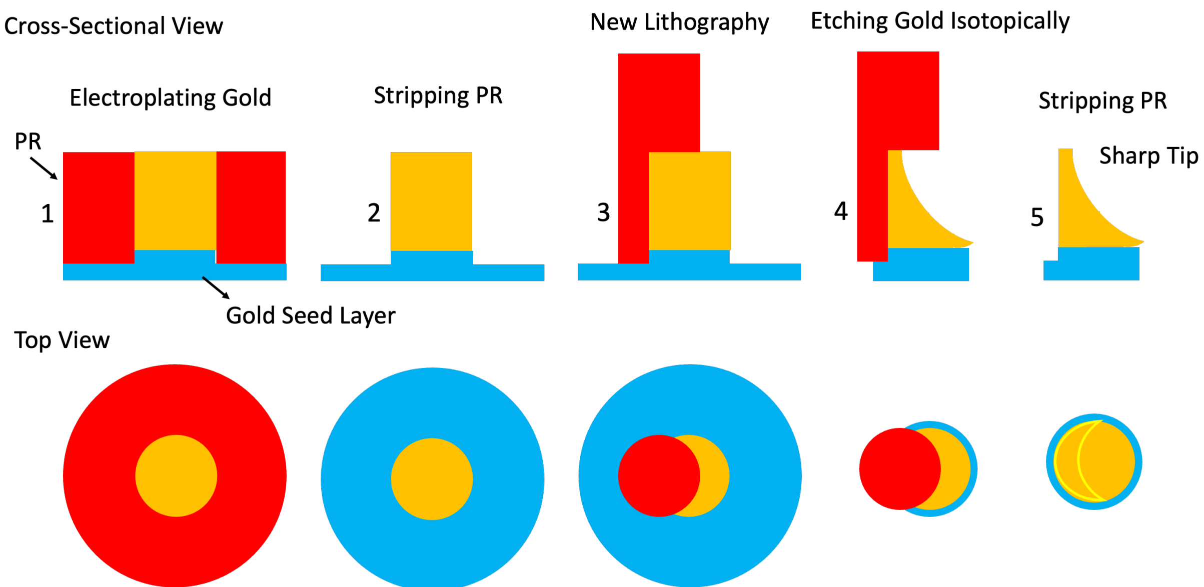

Figure 3.21: Schematic illustration of the electroplating and tip sharpening process in DSE version. The top image shows a cross-sectional view, while the bottom image shows a top view. The process involves a series of steps, including metallization, masking, electroplating, and etching, to create sharp crescent-wedge-shaped tips with a high effective surface area.

Figure 3.22 illustrates SEM images of the tips in the DSE version before (a) and after (b) gold plating and isotropic etch. While the tip shapes vary, each one exhibits sharp and wedge-shaped features. The porosity of the electroplated gold, resulting from isotropic etching, enhances the effective surface area, leading to reduced recording site impedances.

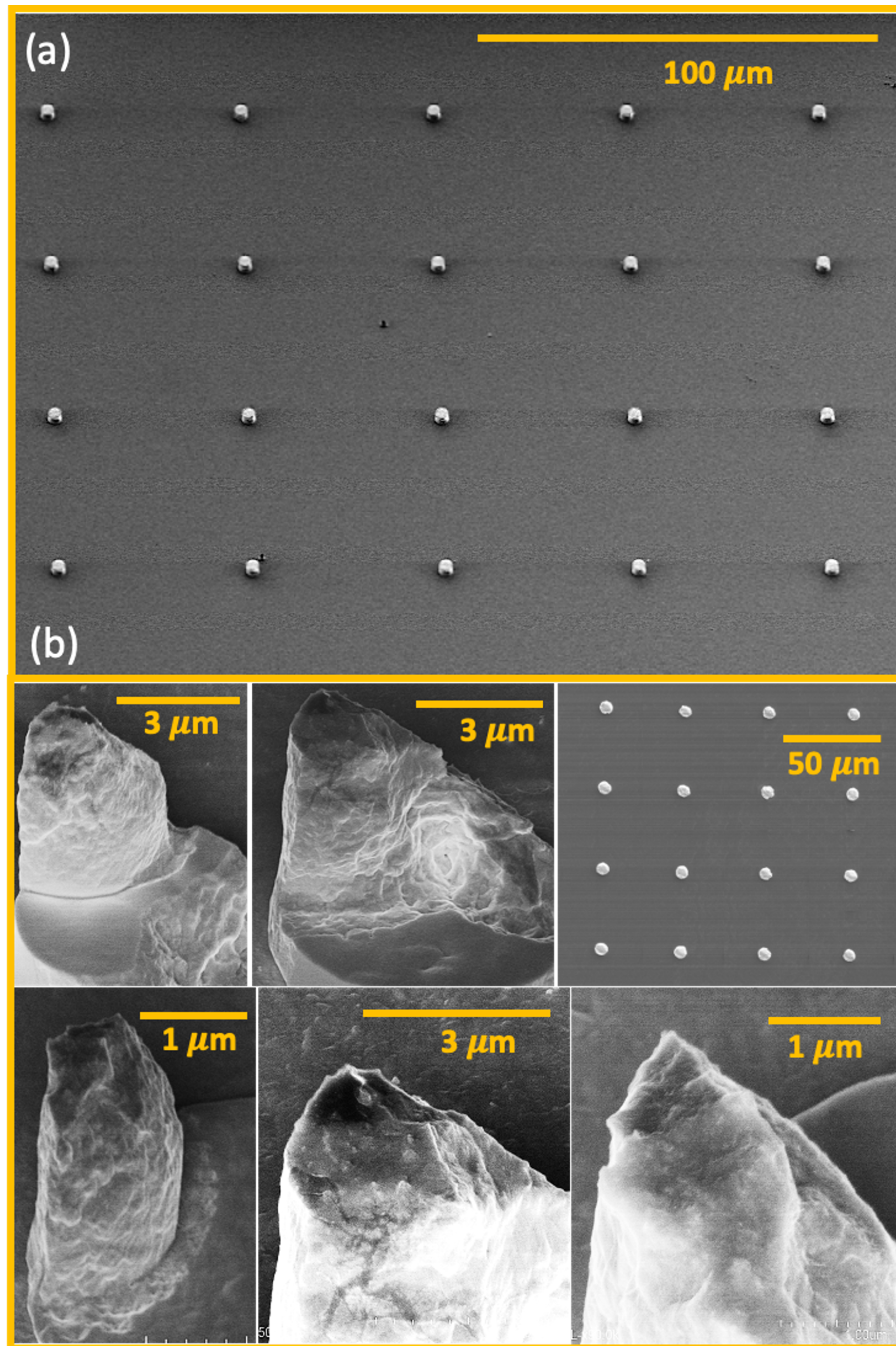

Figure 3.22: SEM images of electroplated tips in DSE version, (a) before gold plating, (b) after gold plating, showcasing their sharp and wedge-shaped profiles. The porosity of the electroplated gold, due to isotropic etching, enhances the effective surface area, leading to reduced recording site impedances.

In the next phase, we proceed by etching the sacrificial polysilicon layer from the backside of the wafer (see Figure 3.20, step #13), followed by substrate etching using a fast DRIE process.

In this process, the back side of the wafer is etched utilizing a mask with holes of 30 μm diameters at the tip locations to cover the tips.

Moreover, during the fast DRIE, the backends of the flexible cables are also protected (Figure 3.20, step #14). Figure 3.23 shows the photoresist mask (AZ 12 XT) with a thickness of 15 μm, deposited on top of the tip locations and the cable backends. This mask serves to protect them during the fast DRIE process. As it is shown, this photoresist can be patterned very well, even on non-uniform surfaces of the arrays that include the tips.

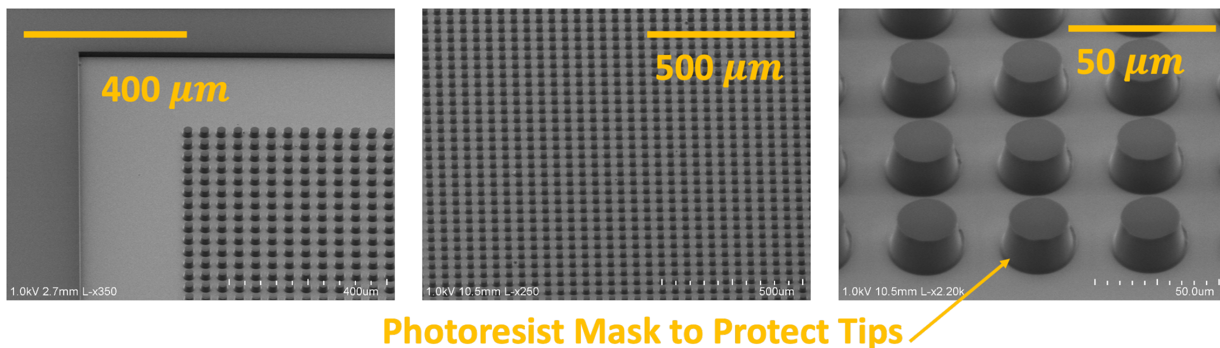

Figure 3.23: A 15μm thick photoresist mask (AZ 12 XT) is deposited on the tip locations and cable backends, providing protection during the fast DRIE process.

In the SSE version, to create metallic tips as the recording sites of the shanks, we flip the wafer and dry etch the backside layers in a maskless process. The first layer to be etched is the insulator layer (Figure 3.24, step #10 and #11). Then, the sacrificial polysilicon is dry etched through a maskless step (Figure 3.24, step #12).

We continue dry etching the silicon substrate through a maskless process until the tips of the shanks emerge (Figure 3.24, step #13). The size of the recording sites in this process is finely adjustable, as we can change the amount of etching of the substrate at this step (Figure 3.2, step #13).

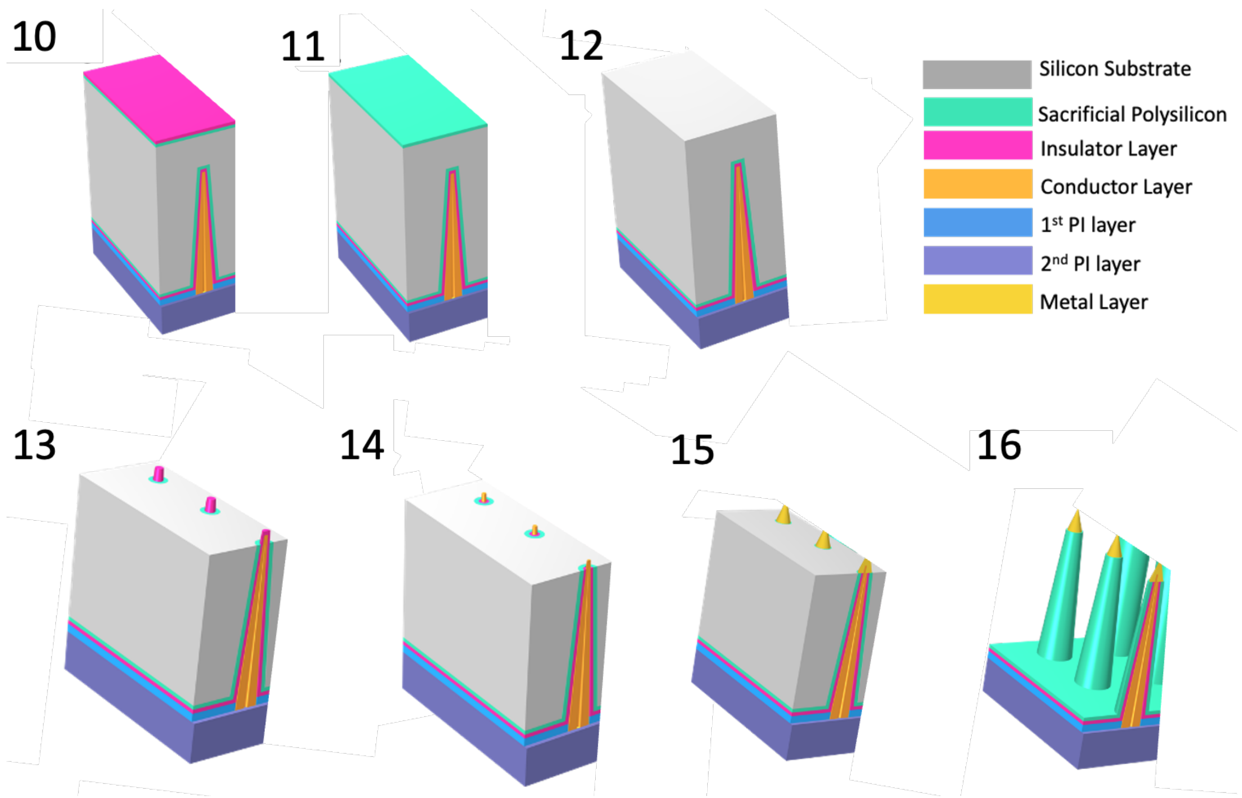

Figure 3.24: Tip metallization process in SSE followed by releasing the shanks.

Figure 3.25 shows the emerged tips of arrays with various sizes and pitches after maskless dry etching of the substrate. These emerged tips are coated with an insulator layer. The tips are nearly tapered, and the diameter at the very top of the tips before etching the insulator layer is approximately 6.5µm.

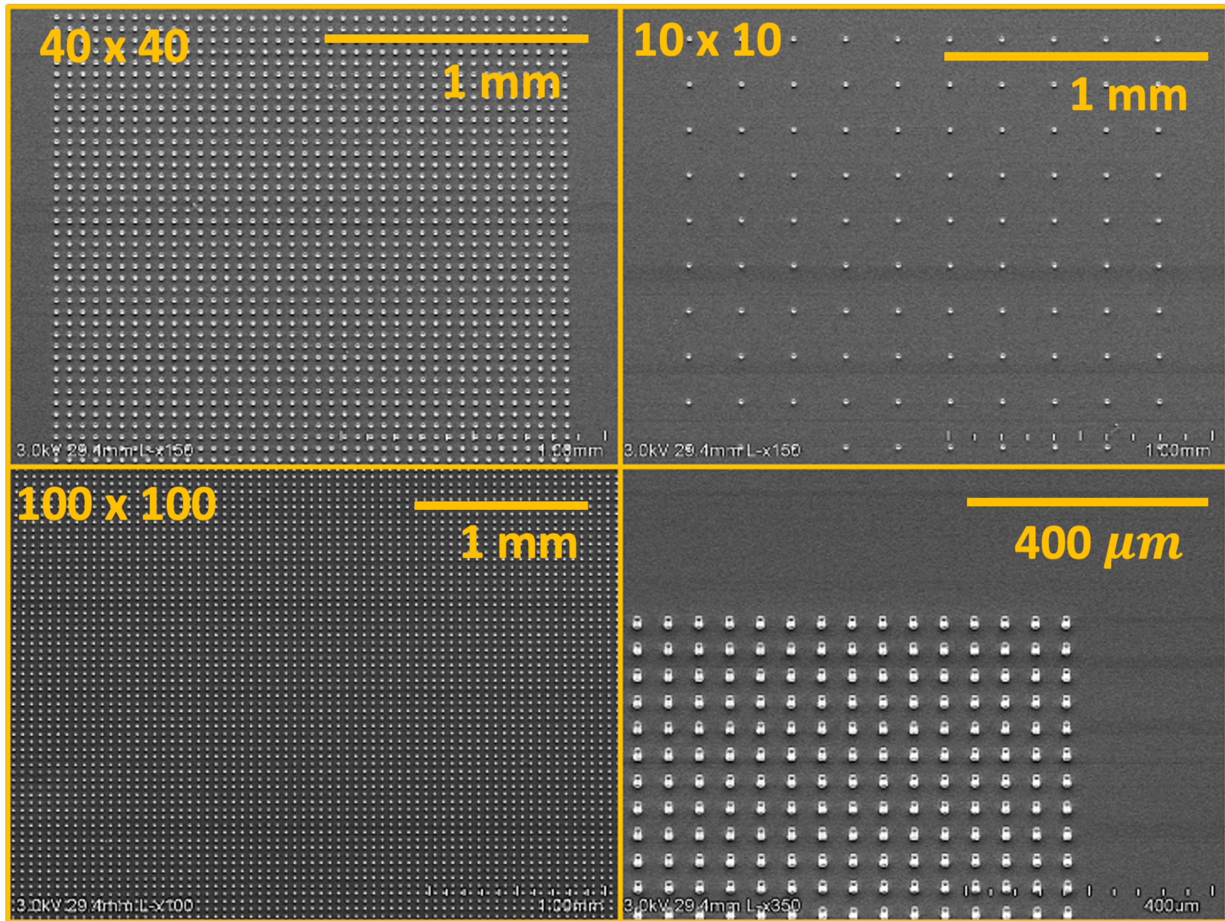

Figure 3.25: Emerged tips of arrays in SSE version. Showing arrays with various sizes and pitches after maskless dry etching of the substrate. The tips are coated with an insulator layer.

As was mentioned, the emerged tips are coated by an insulator layer. To create recording sites, we need to de-insulate the tips first. In the SSE method, tip de-insulation can be easily done in a maskless plasma etch process. By etching the outer insulator layer, the conductive layer at the tips is exposed (Figure 3.24, step #14), and we can proceed with the tip metallization at this step.

We sputter Cr/Au 300/3000Å on top and pattern it with photoresist and etch it through an isotropic wet etching process (Figure 3.24, step #15). Figure 3.26 presents SEM images of the tip metallized using the SSE version tip metallization process.

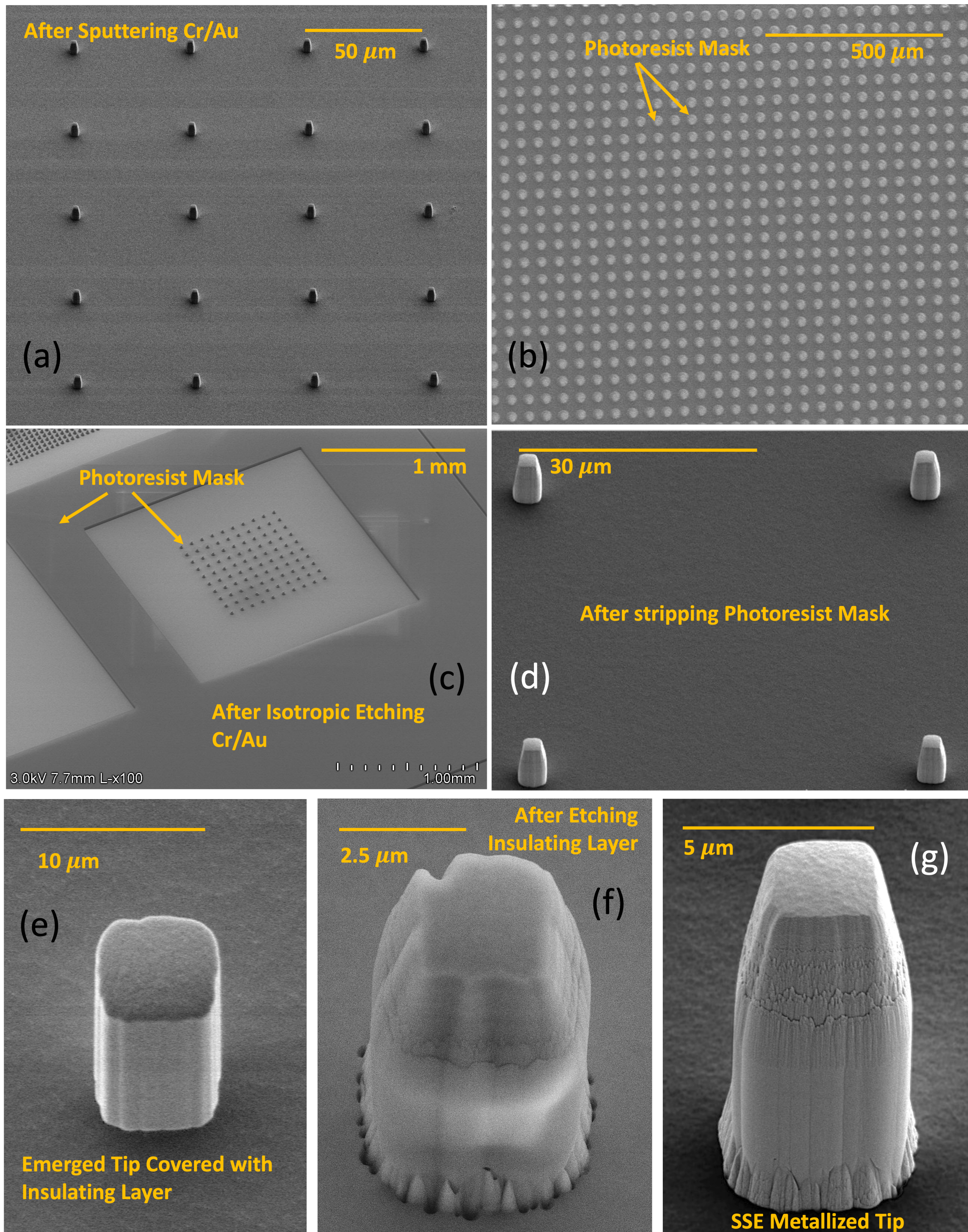

Figure 3.26: SEM images of the SSE version tip metallization process. (a) Emerged tips after sputtering Cr/Au. (b) Photoresist mask (AZ 12XT, 10µm) used for etching the Cr/Au through an isotropic etching process. (c) 10 x 10 array after isotropic etching of the Cr/Au layer. (d) Metallized tips after stripping the photoresist layer. (e) and (f) SEM images of one emerged tip before and after etching the insulating layer on top, respectively. (g) Zoomed-in image of the same tip after metallization with the SSE version method.

In Figure 3.26 (a), emerged tips are shown after sputtering Cr/Au, and (b) displays the photoresist mask (AZ 12XT, 10μm) used for etching the Cr/Au through an isotropic etching process. The photoresist mask coats the metal layer on the tips, with a diameter of 10μm covering each tip. By over etching, we ensure the metal layer only covers the tips. Figure 3.26 (c) shows a 10×10 array after isotropic etching of the Cr/Au layer, with the photoresist mask almost intact. Figure 3.26 (d) shows the metallized tips after stripping the photoresist layer. Figures 3.26 (e) and (f) show SEM images of one emerged tip before and after etching the insulating layer on top, respectively, revealing the conductive n-doped polysilicon in (f). Finally, Figure 3.26 (g) provides a zoomed-in image of the same tip after metallization with the SSE version method.

The tips at the highest point have a width as small as 2.5μm, which is quite sharp but not as sharp as the DSE tip metallization (< 1μm). However, the metallized tips are all of the identical shape, ensuring consistent and uniform measurements. Furthermore, in the SSE method, it is significantly easy to control the size of the emerged tips to be metallized through a maskless dry etching process.

Similar to the DSE version, in the SSE process, after metallization is complete, we cover the tips with a thick layer of photoresist (AZ 12XT, 15μm) to protect them during the remaining substrate plasma etching. In this process, the backside of the wafer is etched using a mask with holes of 30 μm diameter at the tip locations to cover the tips. Here, we first use a fast DRIE process to etch the bulk of the substrate while protecting the tips, the frames around the probes, and the backend of the flexible cables (Figure 3.24, step #16). In both the DSE and SSE processes, after fast etching the silicon substrate, the remaining silicon and sacrificial poly-silicon layer around each shank are removed using a slow isotropic dry etch (thinning process + XeF2 etch), releasing the highly non-planar shanks, flexible cables, and backends (Figure 3.27, step #19).

Figure 3.27 presents a summary the process flow for fabricating both DSE and SSE types of non-planar probes on the same substrate.

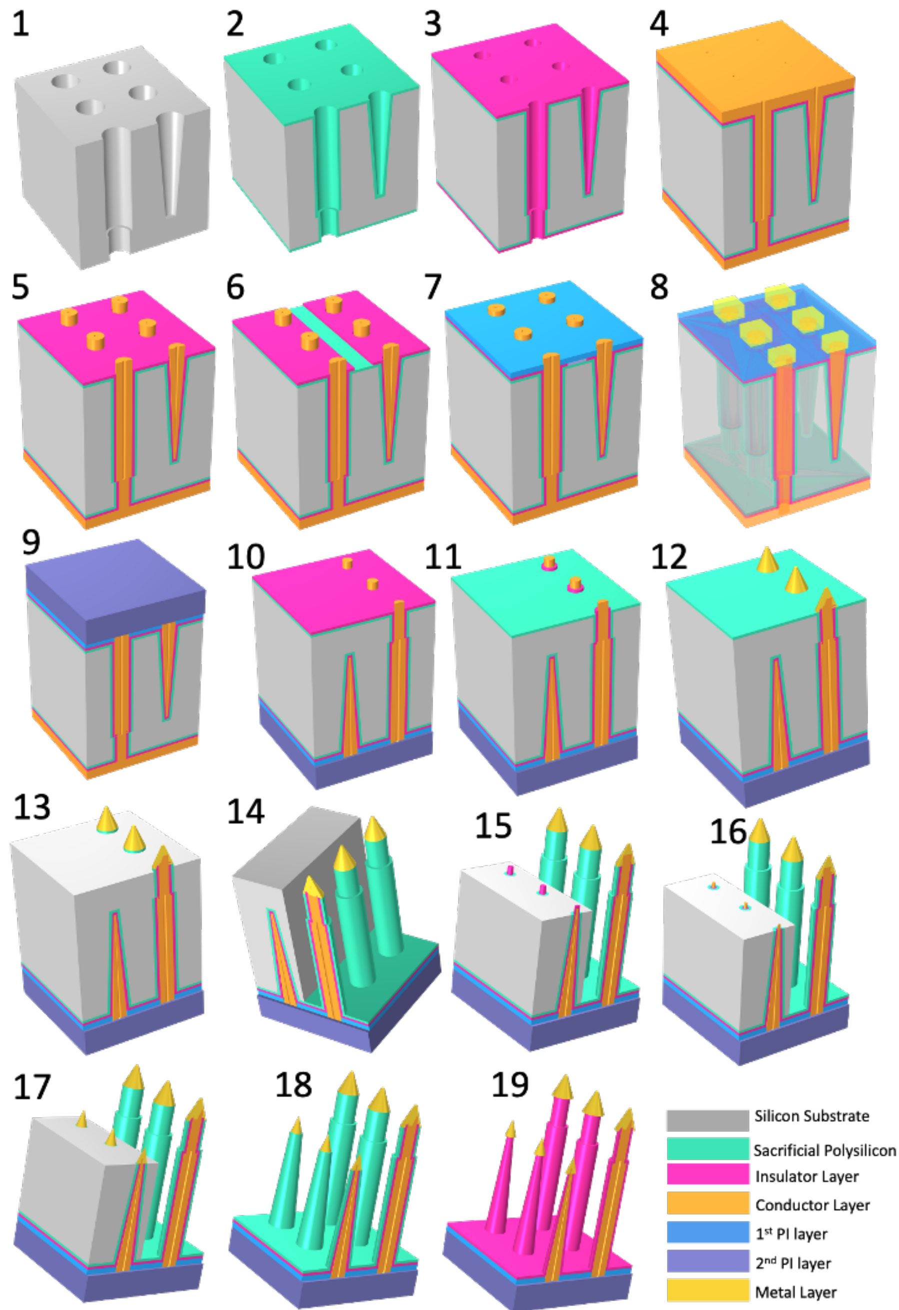

Figure 3.27: Process flow diagram illustrating the fabrication steps for both DSE and SSE types of non-planar neural probes. Steps 1-9 detail the process with the left side representing DSE and the right-side representing SSE. From steps 10-19, the wafer is flipped, showcasing the reverse perspective with the right side now illustrating DSE and the left side portraying SSE.

The diagram comprises 19 steps, covering steps 1 through 9, where the left-side array represents the DSE method, and the right-side array corresponds to the SSE type. However, from step 10 to 19, the wafer is flipped, displaying the backside at the top. Consequently, the right-side array depicts the DSE type, while the left-side array shows the SSE type.

Figure 3.28 shows SEM images of the high-count high-density non-planar arrays, alongside the integrated cables, following the dry etching of the substrate. The images highlight the diversity in sizes and densities among the samples.

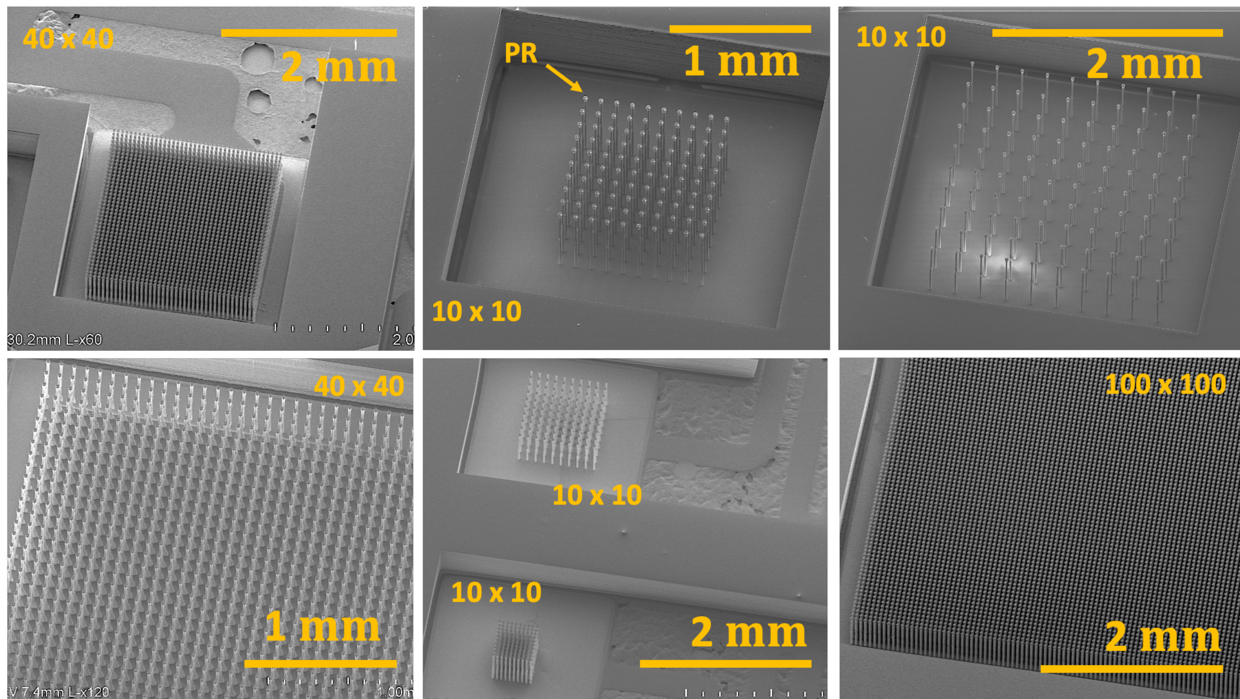

Figure 3.28: SEM images of high-count high-density non-planar arrays, integrated cables, after dry etching the substrate, showing variations in sizes and densities among the samples.

These samples were all fabricated on the same 4" wafer, which was mounted with crystal-bond on a carrier wafer. After etching the substrate, the next step is to strip the PR from the tips of the needles. To do that, we need to protect the polyimide cables with a layer of photoresist (SPR 220(3)) locally to cover only the cables. Then, the next step after stripping the PR from the tips is

to release the devices from the carrier wafer and clean the cables by soaking them in Acetone and warm remover PG.

All samples are intentionally designed with bridges connected to designated frames around them. These frames, along with the polyimide thin bridges, serve to maintain structural integrity while the sample is released from the carrier wafer. Once released, the frames and bridges can be easily broken, facilitating the release of the sample.

The bridges contain a small piece of metal that can be easily broken by a laser, further aiding in the release process. This design feature helps protect the samples and prevents them from being released prematurely in the solution, thereby safeguarding their delicate needles. Figure 3.29 illustrates the wafer containing samples before and after release from the carrier wafer, as well as the polyimide bridges and frames surrounding the devices. As depicted, the flexible cables are fully released.

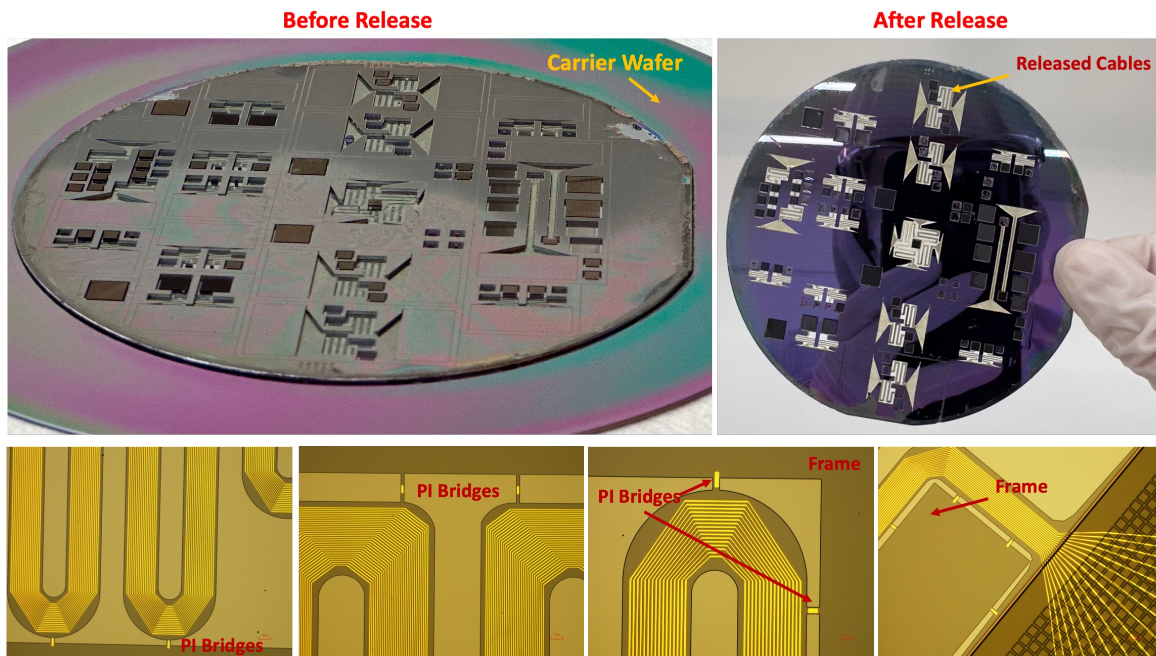

Figure 3.29: Illustration of the wafer containing samples before and after release from the carrier wafer, showing the polyimide bridges and frames surrounding the devices.

Figure 3.30 presents images of several released samples, juxtaposed with a finger or a penny to provide a size comparison. Figure 3.30 (a) shows a 100 × 100 array on an index finger. The inner image is an SEM of a part of the same array. Figure 3.30 (b) shows a 40 × 40 array connected to a straight-style flexible cable juxtaposed with a penny. Also, an inner image shows an SEM image of the same array. Figure 3.30 (c) shows the backside of an array connected to a meandered-style flexible cable juxtaposed with a penny.

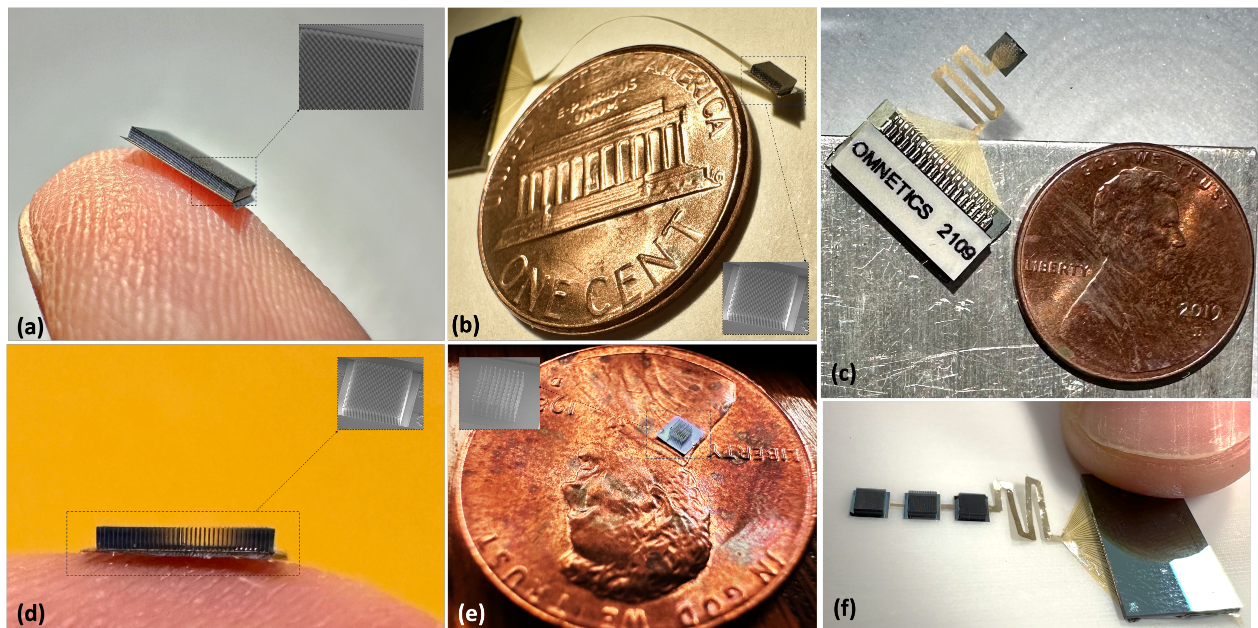

Figure 3.30: Comparison of released samples with a finger and a penny for size reference. The Omnetics connector is attached to one of the backends, illustrating the use of the flexible cable's backend as a PCB for easy and fast connector soldering post-release.

Notably, an Omnetics connector is attached to the backend of this sample. This configuration demonstrates that the backend of the flexible cables serves as a PCB, allowing for the connectors to be readily soldered onto it immediately after the samples are released, making the packaging process much easier and faster compared to wire bonding or other conventional packaging methods. Figure 3.30 (d) and (e) show 40 × 40 and 10 × 10 arrays. Figure 3.30 (f)

shows three 40 × 40 arrays connected to the same cable, a configuration to give more flexibility to the neuroscientists to do measurements from different parts of the brain simultaneously.

SEM images in Figure 3.31 provide a detailed view of various needle components, including the tips and the junction between the top and bottom holes in the DSE version. Notably, there is no misalignment, and the junction is almost ramped, not step-shaped. This design improvement (aligning dissimilar sizes of holes) addresses the breaking point due to misalignment observed in SEA arrays. Furthermore, the arrays feature high-count and high-density, with sharp gold-plated tips and customizable features.

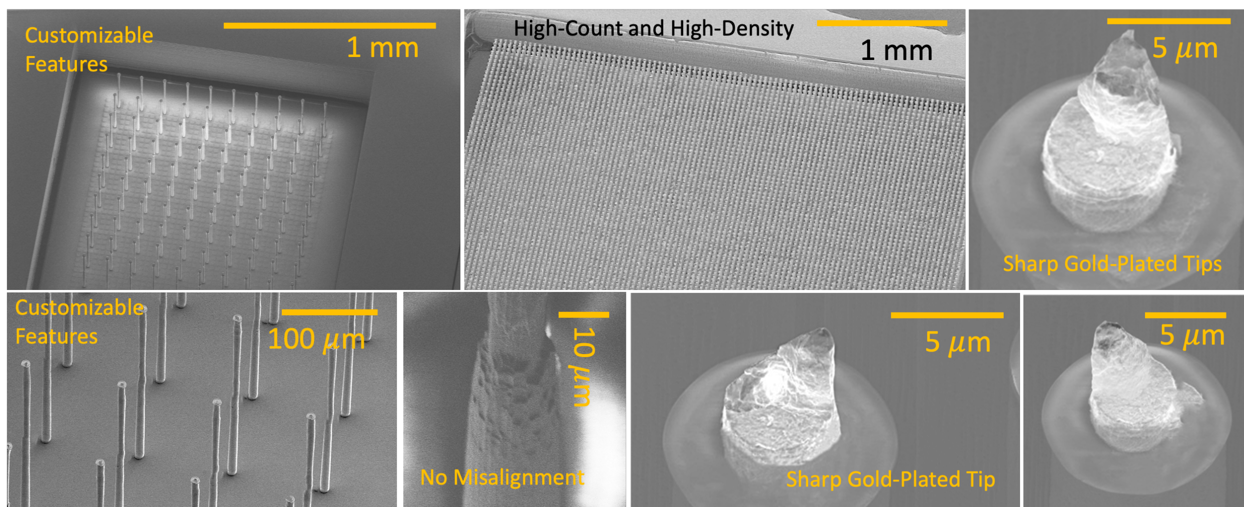

Figure 3.31: SEM images of needle components in the DSE version, showing ramped junctions and no misalignment between the top and bottom holes. This design improvement addresses breaking points seen in SEA arrays. The arrays feature high-count, high-density, sharp gold-plated tips, and other customizable features.

It is noteworthy that, similar to the SEA fabrication method, both the DSE and SSE approaches share the ability to generate longer shanks. This is achieved through a process that includes etching, aligning, and bonding multiple wafers, allowing the creation of deeper holes and, consequently, longer shanks.

Moreover, in the SEA and SSE methods, varying shank lengths within a single array is achievable by creating holes with different diameters at the top of the substrate, a phenomenon influenced by the DRIE lag effect. Conversely, in the DSE method, where holes are etched through the wafer, all shanks in an array exhibit a uniform final length.

3.2.4 Tip Metallization Technique for Varying Shank Lengths Neural Probe Arrays

The tip metallization suggested for the DSE and SSE method is not effective when there are varying shank lengths within a single array. The SEA array method offers tip metallization on the shoulders of the shanks, not at the tips (Figure 3.4- 1). Therefore, in this section we propose an alternative method called electroplated photoresist (EP) method to de-insolate and metallize the tips of shanks, either with the same length or varying shank lengths within a single array.

This technique involves employing electroplated photoresist as a thin, conformal mask layer to expose the tips. We utilized the EP metallization technique for de-insulation and metallization of the tip of Michigan's SEA arrays, but it can be used for any out-of-plane neural probe array with any lengths of shanks.

Out-of-plane neural probes consist of an array of needles, with each needle supporting a conductor layer covered by an outer insulator layer that needs to be removed at the needle tip to expose the conductor layer. Once the conductor layer is exposed at the tip, it must be coated with a metal layer to create an ohmic contact and form recording/stimulating electrodes. First, the needle arrays are coated conformally with EP (MicroChem InterVia 3D-P) with a thickness of 10–15µm (Figure 3.32 (b-2)).

Second, the EP layer is directly exposed in a standard mask aligner from the topside for an optimum amount of time. The vertical sidewalls of the needles are not exposed as much as the non-vertical tips, therefore only the photoresist around the tips receives sufficient exposure to be

developed and removed to expose the insulating layer at the tip (Figure 3.32 (b-3)). Third, the insulator layer at the tip is etched using RIE to expose the underlying conductor layer (Figure 3.32 (b-4)). Once the conductor at the tip has been exposed, it is coated with a metal layer, which also covers the patterned EP layer (Figure 3.32 (b-5)). Finally, the metal-coated EP is lifted off, leaving behind the needle array with their metallized tips (Figure 3.32 (b-6)).

We used InterVia 3D-P, an electroplated photoresist that can be deposited through a voltage-controlled process. The proposed electroplating process on SEA arrays was modeled with COMSOL. SEA arrays with needles with lengths of 1200μm and diameters of 20μm and ultra-sharp tips were used in our simulation.

Each needle consists of an outer layer of silicon oxide/silicon nitride/silicon oxide (ONO) with a total thickness of 400nm as an insulator layer and a core layer of n-polysilicon with a thickness of 2μm as a conductor layer. Figure 3.32 (a) shows the FEA simulation of InterVia 3D-P electroplating on a SEA array. The rainbow-colored legend in Figure 3.32 (a) indicates the thickness range of the electroplated resist. This deposition is voltage-controlled, and the applied voltage is 150 V.

According to Figure 3.32 (a), the maximum thickness of the deposited 3D-P occurs in the convex regions because of the high current density at sharp edges; the maximum thickness after 5 minutes is 10μm. Deposition thickness is maximum at the tip of the probes, which is undesirable. Since we will be exposing the tips, the less photoresist there is at the probe tip, the better for de-insulating them. While this may seem problematic, in practice we observed that the tips of the arrays were not completely covered by EP after electroplating.

This is due to the fact that SEA needles are only covered by an insulating layer at the tip (see the close-up of the tip in Figure 3.32). As a result, there is no conductor layer underneath the

insulating layer at the very tips. This significantly reduces the thickness of the EP layer, which is desirable.

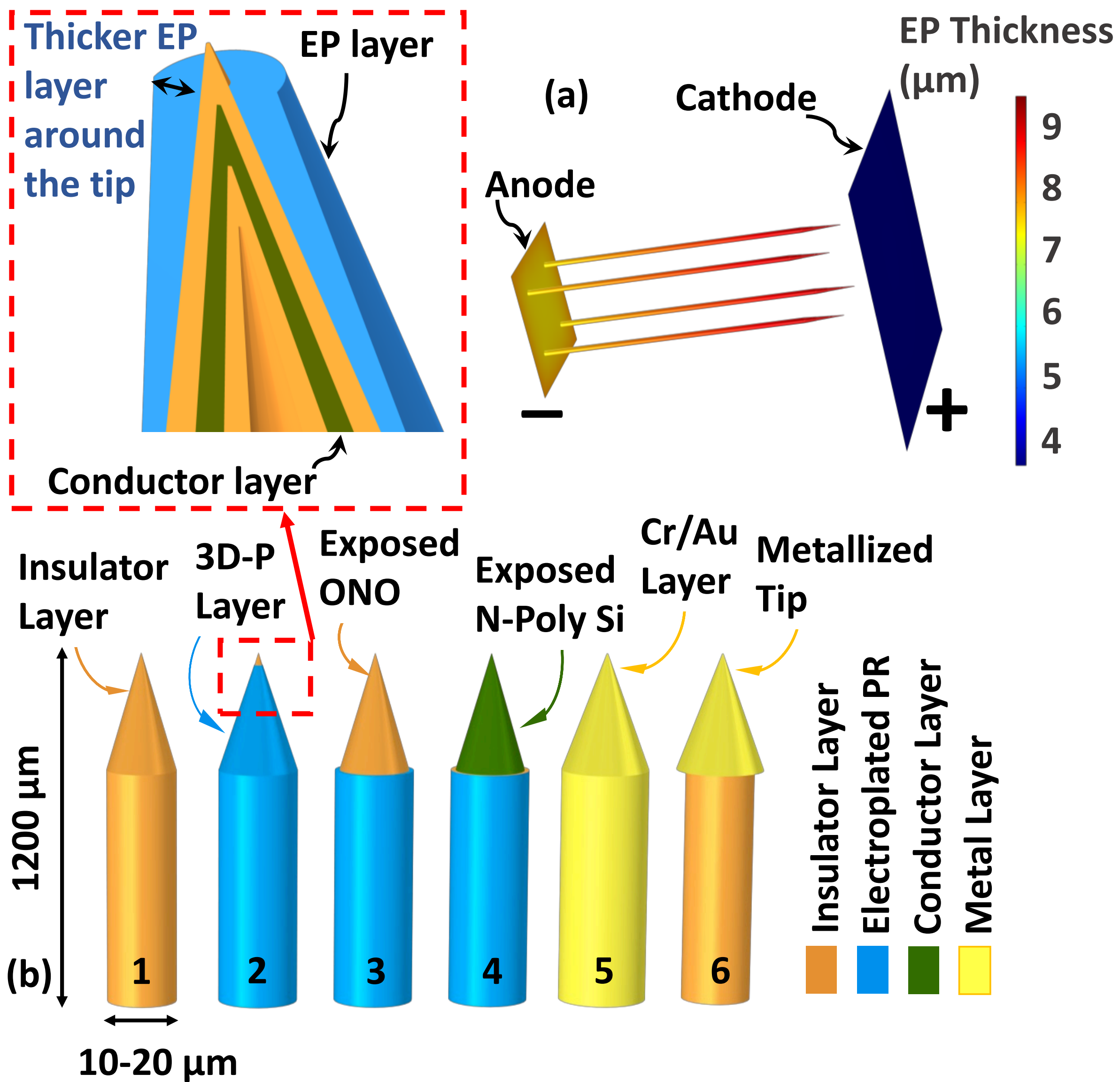

Figure 3.32: EP tip metallizing process, (a) The COMSOL simulation of 3D-P electroplating on a 2 × 2 SEA array. The rainbow-colored legend shows the thickness of EP on the surface of the array, (b-1) a 3D neural probe (b-2) electroplating 3D-P, (b-3) exposing the EP, (b-4) DRIE the insulator layer, (b-5) depositing metal layer, (b-6) lift-off process to create the metallized tips.

The EP process has been demonstrated on the SEA arrays. The arrays consist of needles of 1200μm length, 20μm diameter, and ultra-sharp tips. First, we diluted the InterVia 3D-P with DI water to a 1:1 ratio. The temperature was 20-30°C.

During electroplating, the substrate with arrays mounted on it served as the anode and was treated with O2 plasma prior to electroplating. The sample was soaked in the EP solution for at least five minutes before plating. Following that, the voltage was ramped up to 150 V. The electroplating process took about five minutes to complete. Afterward, we rinsed the sample with DI water, dried it, and baked it for about 8 minutes at 90 °C.

Figure 3.33 shows SEM images of 2×2 and 3×3 SEA arrays both before (Figure 3.33 (a)) and after the electroplating process (Figure 3.33 (b-d)). The needles on these arrays are 1200μm long and 20μm in diameter, separated by 250 - 500μm. As shown in Figure 3.33 (b), all parts of the needles in the array are clearly covered by a reasonably uniform layer of EP shown in green.

There is some bumpiness around the needles shank mostly due to the topography on the shank itself. The resist thickness however is not completely uniform. In addition, the EP is clearly thicker near the tip of needles than in other sections, which is due to the higher current density in the convex regions.

At the tips, however, we do not see a thick EP layer, which makes it easier to expose and etch the tips. To remove the EP from the needle tips, the coated sample was exposed for 10 seconds without a mask, followed by 60 seconds of development with InterVia 3D-P Developer.

Figure 3.33 (c) shows needles after exposing and developing the EP. The pink tips indicate the exposed insulator layer. Exposure and development of the EP is followed by de-insulating the tips and metalizing the arrays by Cr/Au (300/3000Å), and lifting off the EP layer, (Figure 3.33 (d)).

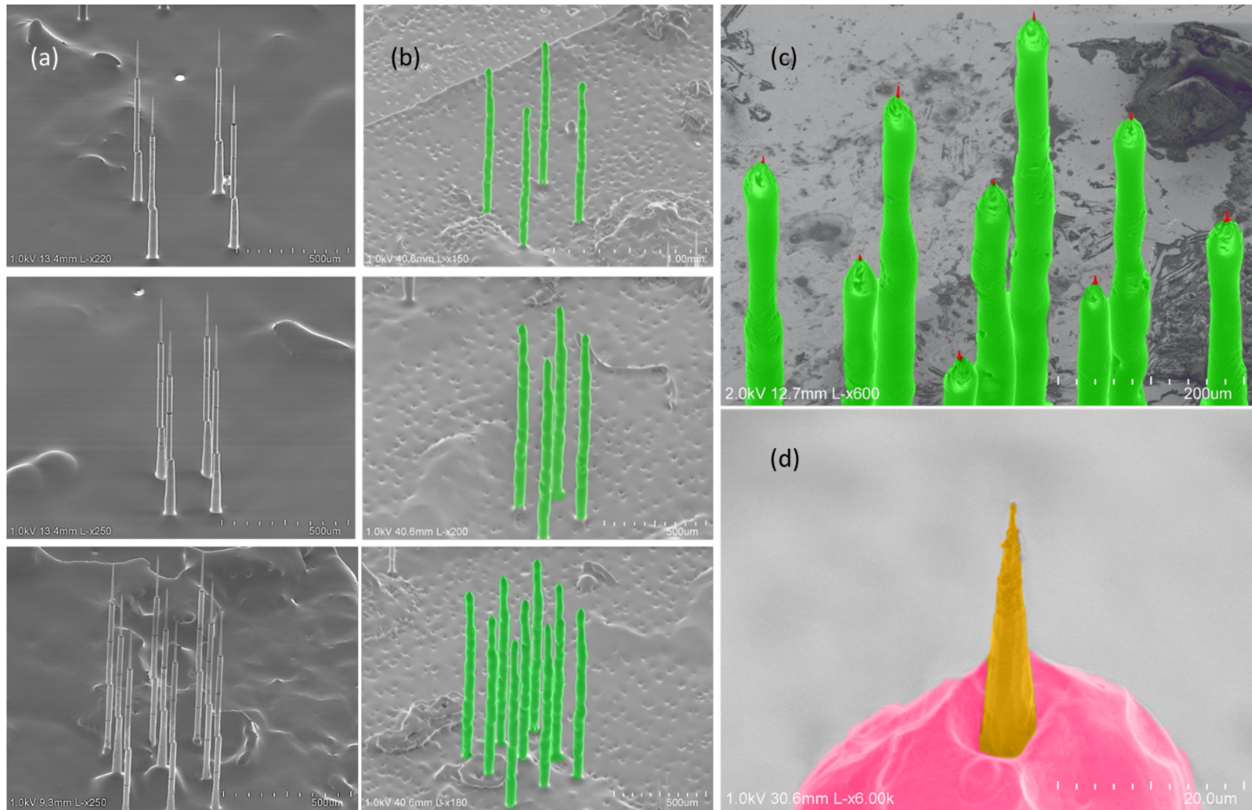

Figure 3.33: Implementing the EP metallization technique on SEAs, (a) SEM of 2×2 and 3×3 SEA arrays before electroplating photoresist, (b) After electroplating 3D-P, (c) Needles after exposure and developing the EP, (d) Close-up of a de-insulated and metallized tip.

The proposed EP metallization technique is a repeatable, uniform, and flexible process for de-insulating and metallizing the tips of out-of-plane probes, which consist of arrays of sharp needles with arbitrary sizes in length, width, and shape.

The proposed new fabrication techniques for creating high-count and high-density out-of-plane neural probe arrays, the DSE and SSE methods, offer significant improvements over the existing SEA method in several key areas.

First, the DSE and SSE methods offer improved packaging compared to the SEA method. In DSE and SSE approaches, arrays are monolithically integrated with flexible cables and backends, eliminating the need for wire bonding of contact pads on the array base to a separate

PCB for channel access, as required by the SEA method. This not only simplifies the packaging process but also increases efficiency.

Second, recording site placement is crucial in neural probes, and both the DSE and SSE methods allow for recording sites at the shank tips, which is a more desirable configuration compared to the SEA method, where sites were at the shank shoulders. Additionally, the EP method is proposed that is suitable for metallizing the tips of arrays with arbitrary shank lengths, diameter, or shape.

Third, in the SEA and SSE methods, varying shank lengths within a single array is achievable by creating holes with different diameters at the top of the substrate, a phenomenon influenced by the DRIE lag effect. However, in the DSE method, where holes on the substrate are etched through the wafer, the final length of all shanks in an array is the same and cannot be varied.

Fourth, to achieve longer shanks, each of the three methods provides viable solutions. This includes utilizing thicker wafers or employing the process of etching and bonding multiple wafers, akin to the approach employed in the SEA method. In the case of a thicker wafer, the DSE method exhibits improved wall coverage during the deposition of multiple layers into ultra-deep, high aspect ratio holes. This advantage arises from the holes being open from both sides of the substrate, distinguishing the DSE from SEA and SSE methods.

Finally, all three methods—SEA, DSE, and SSE—are customizable and capable of creating large count, high-density arrays with arbitrary density and distribution of shanks, which is valuable for diverse applications in neuroscience studies.

Table 3.1 provides a comprehensive comparison of the capabilities of the proposed new fabrication techniques for creating high-count and high-density out-of-plane neural probe arrays with the SEA method.

Table 3.1: Comparative Analysis of Fabrication Methods: SEA, DSE, and SSE.

Properties	SEA	DSE	SSE
Length of Shanks in an Array	Variable	Consistent with substrate thickness	Variable
Hole Wall Coverage	Adequate	Improved	Adequate
Packaging	Wire Bonding	Integrated Flexible Cable and Backend	Integrated Flexible Cable and Backend
Recording Site Metallization	Conventionally on Shank Shoulders; Alternatively, SSE or EP methods for tip metallization	At the Tip using DSE Method	At the Tip using SSE or EP methods
Capability for Longer Shanks	Use Thicker Wafer or Bond Multiple Wafers	Use Thicker Wafer or Bond Multiple Wafers	Use Thicker Wafer or Bond Multiple Wafers
Misalignment Issue	Arises from Uniform Hole Size During Alignment	No Misalignment	No Misalignment
Large Count and High-Density Arrays	✓	✓	✓
Arbitrary Densities and Distributions in an Array	✓	✓	✓

3.3 Mechanical and Electrical Characterizations

After completing the design and microfabrication phases, the subsequent step is to conduct mechanical and electrical characterization tests on the probes. The innovative technique we have demonstrated involves integrated flexible cables and Polyimide-based backends that function as a PCB. Consequently, once the microfabrication of the devices is completed, they are immediately ready to be connected to Omnetics connectors using soldering. We proceeded by soldering connectors onto the integrated PCB using solder paste and a hot air gun. This method is significantly easier and faster compared to wire bonding or flip-chip bonding, especially when the number of sites increases beyond one hundred.

Upon completion of the packaging process, we proceeded to assess the electrical functionality of the devices. Initially, we characterized the impedance of the recording sites using an Agilent vector network analyzer. The impedance measurements obtained from many samples showed promising results, providing insights into the phase and magnitude of impedance for the

measured sites. One of these samples had 32 available channels connected to the Omnetics connector. The impedance measurements for this sample are shown in Figure 3.34.

The figure displays the average magnitude and phase with red lines, while the error bars for all the measurements are shown in pink. The measurement results at 1 kHz are represented by green dots. Notably, at 1 kHz, the average impedance of these 32 sites was found to be around 1MΩ with a phase of -75 degrees.

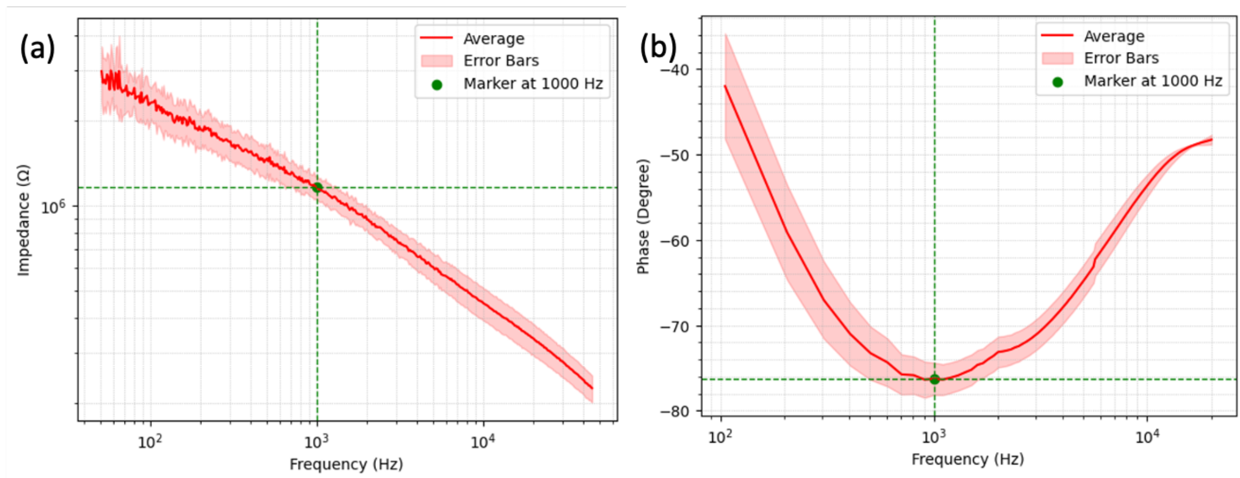

Figure 3.34: The impedance for 32 recording sites, (a) Magnitude, (b) Phase.

A lumped circuit model, comprising resistors and capacitors, was derived using the average impedance obtained from the experimental setup. The simulated data shown in Figure 3.35 closely matches the actual impedance measurements shown in Figure 3.34.

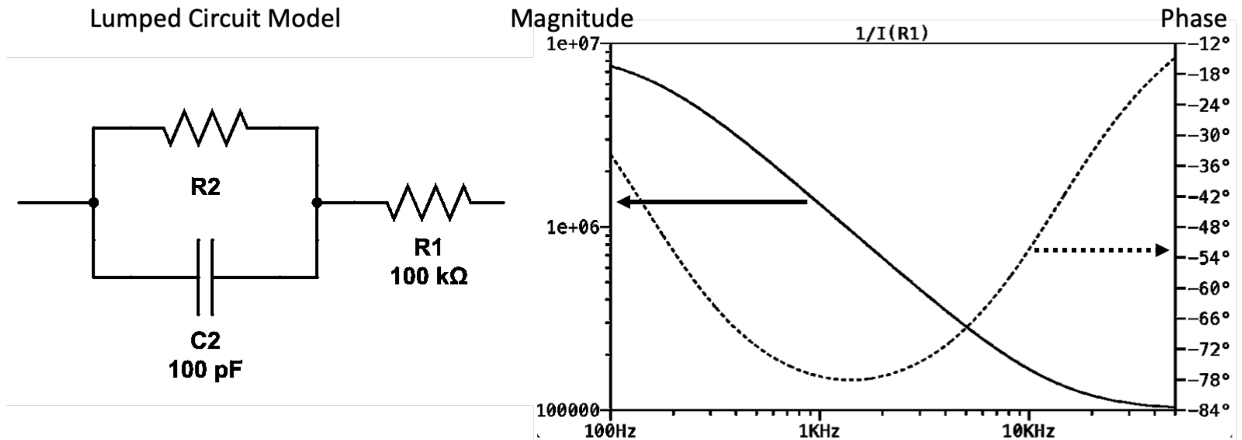

Figure 3.35: Lumped circuit model derived from the impedance measurements, the close alignment between the simulated and measured impedance data indicates the accuracy of the model in predicting the electrical behavior of the system. These results closely match the actual impedance measurements shown in Figure 3.34.

In addition, in-vitro tests were conducted to evaluate the recording and stimulation capabilities of these probes. In the initial test, a spike train mimicking a neural action potential was generated at a frequency of 40 Hz.

This signal, produced by passing a spike train through an RC lowpass filter (with a peak-to-peak value of approximately 100μV), was injected into one of the recording sites of a sample with 16 available channels, and the resulting signal sensed by neighboring recording sites in the same PBS solution was recorded. Figure 3.36 shows the stimulated signal passed through the site marked by ‘1’ and the recorded signals from the other 15 adjacent recording sites.

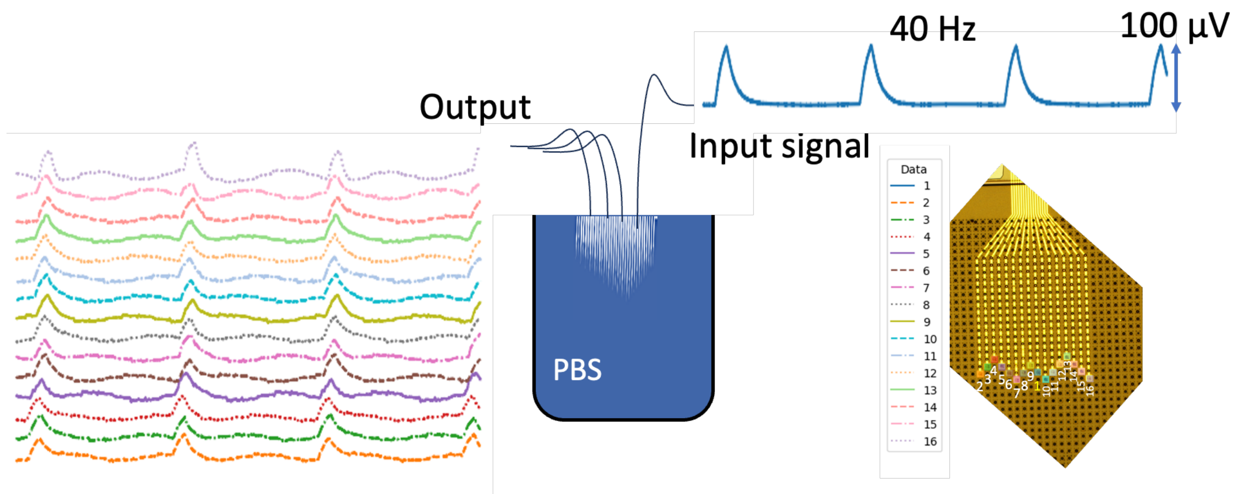

Figure 3.36: Assessment of stimulation capability in an in-vitro test. Stimulated signals marked by '1' and recorded signals from 15 neighboring sites demonstrate the probes' responsiveness.

For a deeper analysis of the results, the modeled impedance of the recording sites in the PBS solution was used to simulate the equivalent circuit of the test. In AC conditions, the electrical double layer at the electrode/electrolyte interface functions as a capacitor, denoted as Cp, and polarization resistance is represented as Rp, which is a very large parallel resistance and can usually be neglected.

For a specific electrolyte, Rs is determined by the resistance between the reference and working electrodes. In practical terms, this ohmic resistance also includes the ohmic resistance of the connection cables and the working electrode.

A lumped circuit model, comprising resistors and capacitors, was also derived using the transient responses obtained from this experimental setup. The parameter values in the model are extracted from the measured traces, and these values match those obtained through impedance measurements, as illustrated in Figure 3.36 and 3.37.

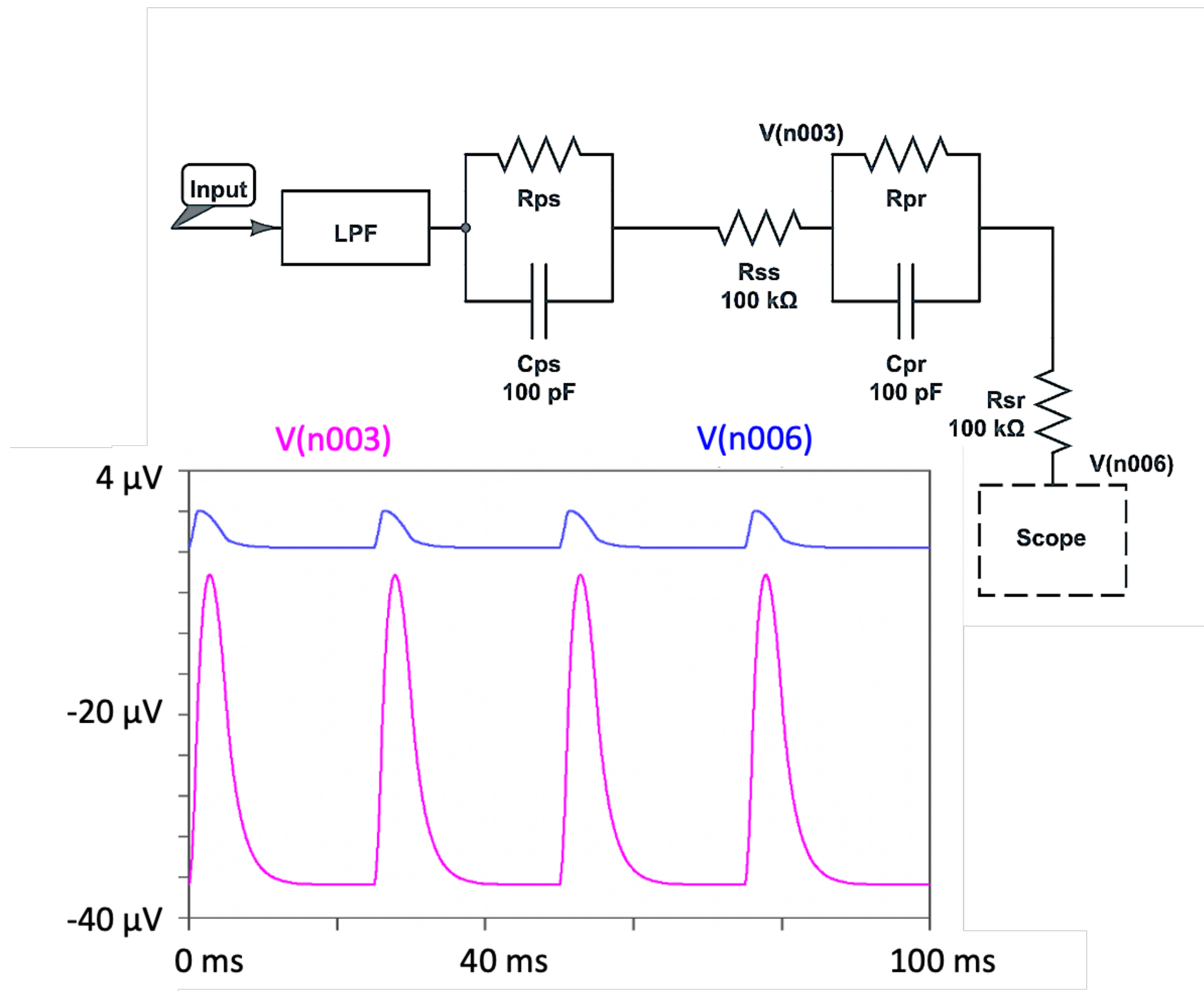

Figure 3.37: A lumped circuit model, comprising resistors and capacitors, was also derived using the transient responses obtained from this experimental setup. The simulated data closely matched the actual measurements.

The shanks of these high-density electrode arrays are remarkably robust, despite their small size (10-15 um diameter and 500 um length). To assess their mechanical strength and their suitability for brain tissue implantation, we used tofu as a material that mimics brain tissue. Figure 3.38 shows the implantation of these probes into tofu. Additionally, the figure demonstrates the flexibility of the meandered polyimide cables. For both 10 × 10 and 40 × 40 arrays, no electrodes broke during any of the implantation tests. Figure 3.39 displays a 40 × 40 array after implantation in tofu, with all shanks remaining intact.

Figure 3.38: Implantation of high-density electrode arrays into tofu, showing the shanks robustness and the flexibility of the meandered cables.

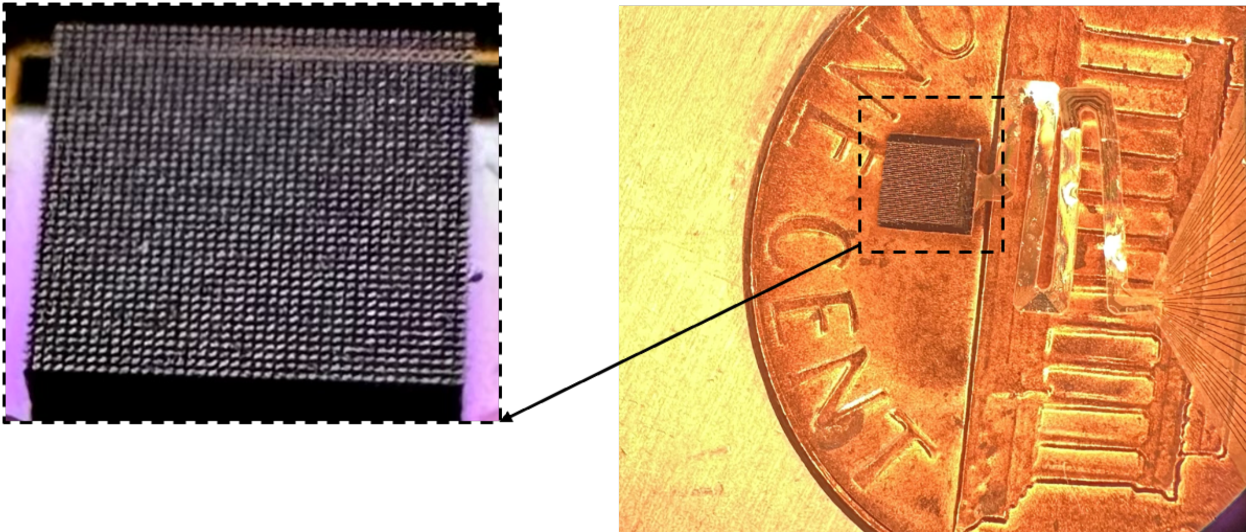

Figure 3.39: A 40 × 40 array after implantation in tofu, demonstrating the robustness of the electrodes with all shanks remaining intact.

3.4 Conclusion

Two new approaches for microfabricating highly non-planar, large count and high-density arrays are presented, allowing for customization as well as enhanced spatial resolution and future detailed mapping of brain functions. Their design versatility enables customization of array size, density, and distribution to suit specific experimental needs. The microfabrication process creates ultra-small needles with sharp tips, enabling precise recording and stimulating sites at the tips of the shanks. These arrays have needles with a length of 500μm and a diameter ranging from 10 to 15μm. The needles are distributed with arbitrary pitches, spanning up to 100μm. Finally, technologies such as electroplated photoresist (EP) and backside patterning using either the DSE or SSE methods ensure the formation of precisely defined sites at the tips. Additionally, we conducted electrical and mechanical characterization tests on the fabricated probes to verify their functionality.

Furthermore, we introduced a packaging method that incorporates a Polyimide-based PCB, facilitating the easy and cost-effective connection of connectors or active electronics to the rear end of the probes, eliminating the necessity for wire bonding or conventional flip-chip techniques. At the time of the writing of this thesis, in vivo testing had not been performed. However, it was demonstrated that these arrays could easily record electrical spikes, mimicking neural spikes in amplitude and frequency, delivered through PBS solution from another site, thus illustrating their eventual utility for high-density in vivo recording.

Chapter 4 Computational Approaches for Large-Count High-Density Neural Interfaces

4.1 Introduction

Extracellular neural recordings offer a unique window into understanding how neurons respond during sensory-motor or cognitive tasks at the individual cell level. Through the analysis of this recorded data, researchers gain valuable insights into the functioning of the brain. However, the data obtained consists of a series of spikes embedded within various sources of noise, necessitating researchers to extract and organize these spikes through a process known as spike sorting.

Fundamentally, neurons communicate via neurotransmission at the synaptic level, with spikes serving as a sign or evidence of this communication. These spikes are captured using neural probes implanted in the brain, resulting in raw data comprising a mixture of signals from different neurons.

Spike sorting entails the isolation and categorization of these individual neural spikes, assigning them to specific neurons. This task of spike sorting can be analogized to the familiar 'cocktail-party problem,' wherein the goal is to discern and pinpoint individual conversations amid a bustling gathering [115].

The significance of spike sorting lies in its ability to differentiate signals recorded by implanted probes and precisely associate them with the neurons that produced them, thereby facilitating a clearer comprehension of neural activity and avoiding conflating specific signals that are generated from individual neurons.

This is pivotal for elucidating how individual neurons contribute to information processing, learning, and memory. Furthermore, precise spike sorting plays a crucial role in mapping intricate neural circuits within the brain, aiding in the identification of functional connections between neurons and the comprehension of the complex network dynamics influencing cognitive functions and behaviors.

In the realm of neurological disorders such as epilepsy, Parkinson's disease, or psychiatric disorders, spike sorting becomes indispensable for investigating abnormal neural activity. Understanding these changes to how individual neurons propagate signals is vital for the development of targeted treatments.

4.1.1 Methods: Manual and Automated

Various methods for spike sorting have been developed to extract and analyze neural activity from extracellular recordings, ranging from manual techniques performed by researchers to automated algorithms designed to streamline the process. Manual sorting entails visually inspecting the recorded data and manually assigning spikes to individual neurons based on their distinct features, such as waveform shape and amplitude.

While some laboratories opt for manual sorting techniques, utilizing tools like XClust [116], SimpleClust [117], and Plexon Offline Sorter [118], standardizing these approaches proves challenging due to the inherent biases and subjective nature of individual operators [83].

In contrast, automated spike sorting algorithms leverage computational techniques to automatically detect and classify spikes within recorded data. These algorithms commonly employ clustering methods to group spikes exhibiting similar waveform features and often employ dimensionality reduction techniques such as principal component analysis (PCA), effectively discerning neuronal activity from background noise.

Various spike sorting algorithms have been developed to date. An overview of the common steps involved in spike sorting, shared among most algorithms, is depicted in Figure 4.1. The initial step involves recording changes in the extracellular electrical potential caused by nearby neural activity (see Figure 4.1 (a-b)). To enhance the data, a high-pass or band-pass filter is often applied to remove low-frequency components (see Figure 4.1 (c)). Spike detection, particularly challenging in high-density multielectrode recordings, is commonly performed using methods like voltage thresholding (see Figure 4.1 (d)).

Following detection, spike waveforms are extracted and temporally aligned based on a common feature, such as the position of the voltage peak (see Figure 4.1 (e)). Dimensionality reduction techniques, like PCA, are then applied to reduce data dimensionality (see Figure 4.1 (f)). Clustering is a crucial step where spikes are grouped based on their features, and each cluster ideally represents spikes from a single neuron (see Figure 4.1 (g-h)).

After clustering, the output of automated spike sorters typically includes a set of clusters, each representing the activity of a distinct neuron. These clusters are accompanied by various metrics, such as waveform shape, amplitude, interspike interval (ISI) — which is the time interval between two consecutive action potentials (spikes) generated by a neuron — and firing rate, providing valuable insights into the characteristics of individual neurons.

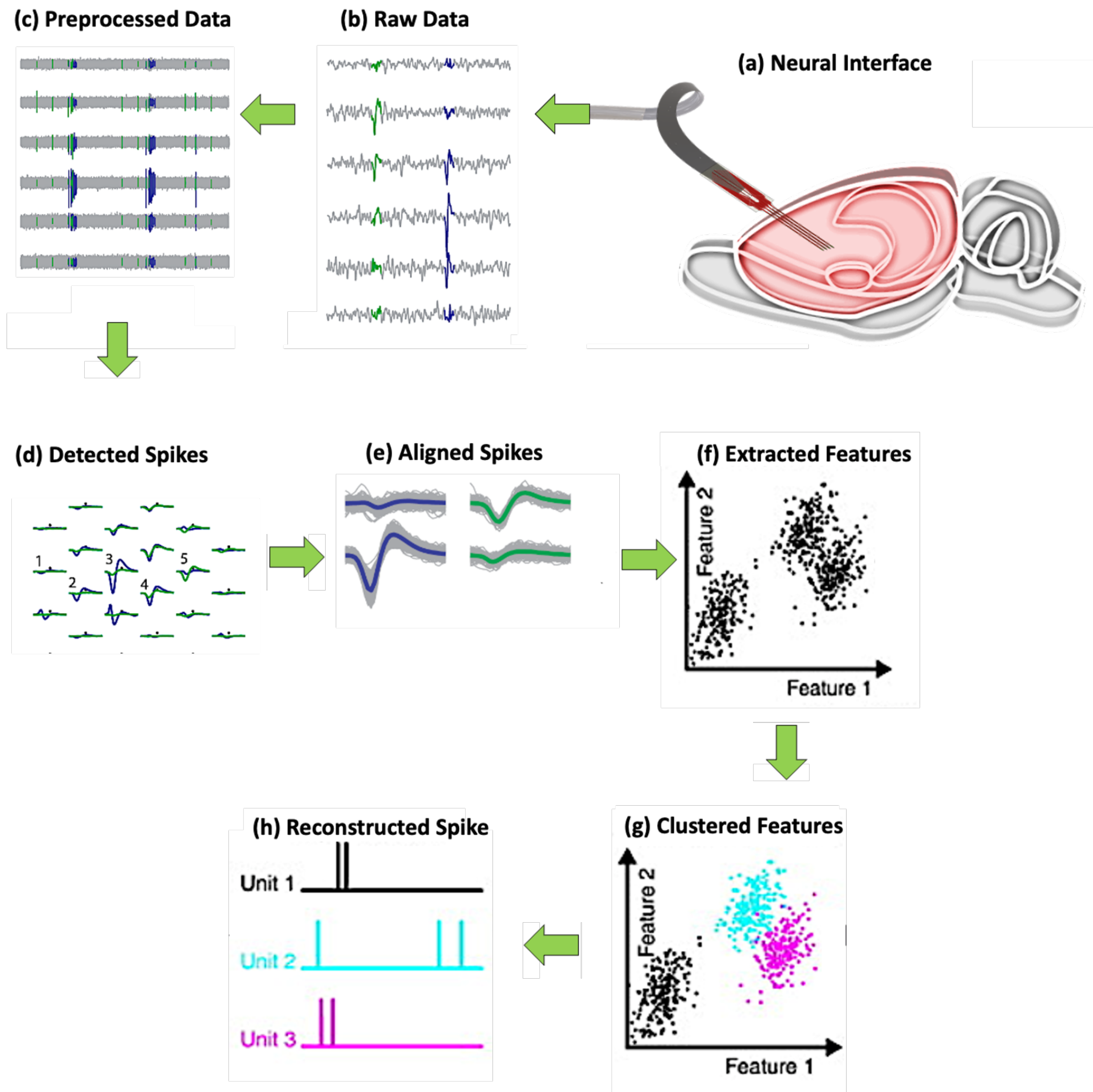

Figure 4.1: Comprehensive overview of steps involved in spike sorting, (a-b) Recording changes in extracellular electrical potential resulting from nearby neurons' activity. (c) Preprocessing data (d) Spike detection (e) Detected sikes are aligned based on a common feature, such as the position of the voltage peak. (f) Feature extraction techniques, including principal component analysis, (g-h) The critical steps of clustering to group spikes based on their features, with each cluster ideally representing spikes from a single neuron.

In the output of automated spike sorting algorithms, clusters are typically categorized into "good" and "bad" clusters based on their quality and reliability in representing neuronal activity. Good clusters exhibit distinct waveform characteristics and consistent features such as ISI

throughout the recording, indicating reliable neuronal signals. Conversely, bad clusters may arise due to various factors such as noise, artifacts, or overlapping waveforms, leading to inaccurate neuronal identification.

Identifying false positives (spikes that are incorrectly labeled as spikes from specific neurons), and false negatives (missed spikes from specific neurons) is crucial in evaluating the performance of spike sorting algorithms. False positives can introduce erroneous interpretations of neural activity, while false negatives result in the loss of valuable data, potentially masking important neuronal events. Therefore, minimizing false positives and false negatives is paramount for obtaining accurate and meaningful results from spike sorting. The depiction of the confusion matrix, outlining True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), is presented in Figure 4.2.

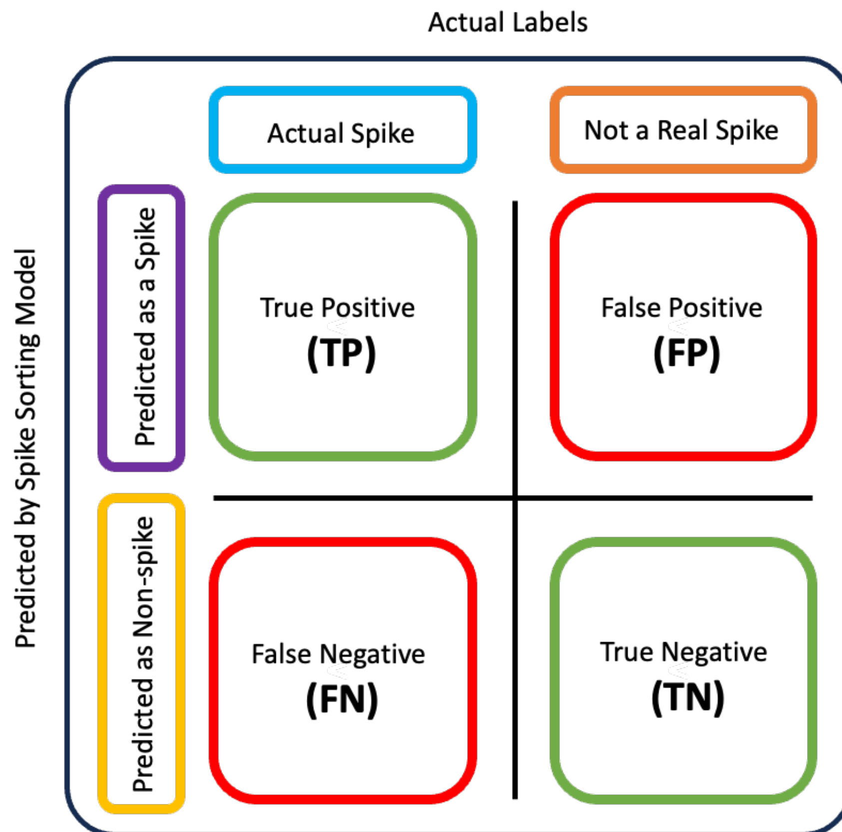

Figure 4.2: Confusion matrix for spike sorting models.

4.1.2 Advancements in Automated Spike Sorting for High-Density Neural Interfaces

The field of spike sorting has seen significant progress in both algorithmic sophistication and software refinement over the years, aimed at improving accuracy and automation [119]. This evolution has been especially notable with the introduction of high-density recording devices in recent times [37, 67, 120-128]. These devices allow for the simultaneous capture of neuronal activity from hundreds to thousands of neurons, rendering manual sorting impractical and increasing the need for accurate and scalable automated spike sorting algorithms [79, 82, 129-134].

However, large-count, high-density neural interfaces with closely-packed electrodes present challenges such as spatiotemporally overlapping spikes from nearby neurons and the generation of large datasets containing signals from numerous neurons simultaneously, complicating accurate sorting. Despite these challenges, automated methods offer improved efficiency and objectivity in spike sorting. Commonly-used automated spike sorting algorithms in neuroscience research for analyzing extracellular electrophysiological data obtained from multi-electrode arrays include KiloSort [135], SpykingCircus [132], and Klusta [82].

KiloSort [135], developed by the laboratory of Joshua Siegle and the Allen Institute for Brain Science, is widely utilized for its template matching-based approach and efficient clustering of spikes based on waveform similarity. Template matching in spike sorting involves comparing the shape of each spike in an extracellular recording to a set of predefined templates, typically created by averaging multiple instances of spikes from the same neuron.

The spike is then assigned to the template with the highest similarity score, often determined using a metric like the Euclidean distance or correlation coefficient. This process is

iterated, with templates updated based on assigned spikes, until all spikes have been assigned. The templates are then classified into different clusters based on their similarity to each other. While template matching can be effective for well-separated and distinct spikes, it may struggle with overlapping or poorly separated spikes, and other techniques like clustering-based methods or machine learning algorithms may be more suitable in such cases. Its speed and scalability make it suitable for sorting large datasets from high-density electrode arrays. KiloSort 2 [129], an advanced version, introduces automation for tasks like unit merging, streamlining the process and reducing manual intervention, although it may occasionally over-merge spikes from multiple cells.

SpykingCircus [132], developed by the laboratory of Pierre Yger, combines template matching and clustering techniques to handle recordings with high channel counts and dense spike trains, offering features for automatic spike detection, clustering, and noise estimation. Klusta [82], developed by the laboratory of Leslie Smith, combines automatic clustering with manual curation steps, performing automatic clustering based on waveform shape and amplitude while providing tools for manual inspection and refinement of results, known for its flexibility and adaptability to various recording conditions.

4.1.3 Curation after Spike Sorting

Following automated spike sorting, a manual curation stage has typically been necessary, during which a human operator corrects any imperfections in the automated results using a Graphical User Interface (GUI). Manual curation plays a pivotal role in the spike sorting process, ensuring the accuracy and reliability of the extracted units.

This step allows human operators to apply their expertise and judgment to the data, which can be particularly valuable in cases where automated methods may struggle, such as with overlapping or poorly separated spikes. Human operators can also take into account contextual

information, such as the presence of artifacts or noise in the recording, which may not be captured by automated algorithms. Additionally, manual curation allows for the incorporation of subjective biases, such as whether an operator tends to "clump" or "split" clusters, which can influence the final sorting results. Though time-consuming, it is an indispensable step in guaranteeing the high quality of spike sorting results, enabling researchers to confidently address their research inquiries. It would be desirable to eliminate this manual curation step if spike sorting algorithms could provide much improved reliability.

Phy [82] is an open-source software package specifically designed for the analysis of electrophysiological data. Its graphical user interface (GUI) serves as a specialized tool tailored for the manual curation of extracted units derived from electrophysiology recordings. This process involves examining the waveforms and other features of the extracted units, such as waveform shape, peak amplitude, number of spikes, and autocorrelogram, to assess their classification accuracy and determine if they belong to the same neuron. Phy can open datasets spike-sorted with the following programs: KiloSort, SpykingCircus, and Klusta.

The Phy GUI, depicted in Figure 4.3, consists of several components, each serving a specific purpose with defined criteria and standards for cluster management:

- Cluster Scatter View: This component presents all clusters in a scatter plot, positioning them based on amplitude (x-axis), depth (y-axis), and marker size indicating firing rate.
- Cluster View: Providing a comprehensive list of all clusters contained within it.
- Similarity View: Similar to the cluster view but with an additional column for similarity to clusters selected in the cluster view, sorted by decreasing similarity.

- Waveform View: Displays waveforms of selected spikes on relevant channels, considering amplitude and proximity to the peak waveform amplitude channel.
- Amplitude View: Shows the amplitude of selected spikes belonging to the chosen clusters, accompanied by vertical histograms on the right.
- Correlogram View: Presents auto-correlograms and cross-correlograms between all pairs of selected clusters, with horizontal lines indicating baseline firing rate and vertical lines representing the refractory period (defaulting to 2 ms).
- ISI View: Displays inter-spike intervals computed using all spikes for the selected clusters.
- Firing Rate View: Illustrates the instantaneous firing rate computed using all spikes for the selected clusters.
- Probe View: Offers an illustration of recording sites on high-density multielectrode arrays, typically containing hundreds to thousands of recording sites.

This structured breakdown enhances the efficiency of navigating and utilizing the GUI for electrophysiological data analysis. Additionally, Phy can display features calculated by various other software tools, such as the isolation distance computed through the electrophysiology pipeline.

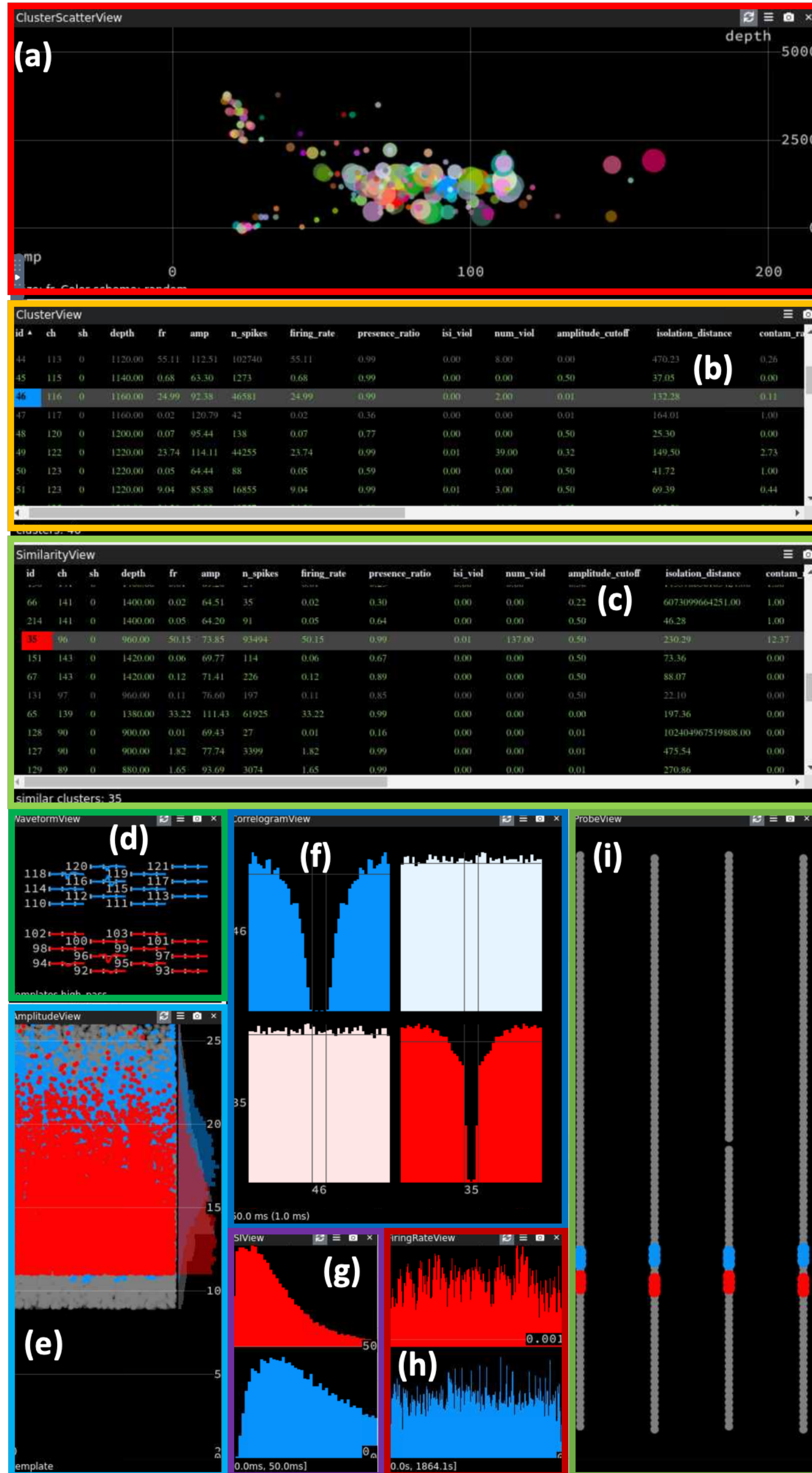

Figure 4.3: Phy GUI, (a) cluster scatter view, (b)cluster view, (c) similarity view, (d) waveform view, (e)amplitude view, (f) correlogram view, (g) ISI view, (h) firing rate view, and (i) probe view.

4.1.3.1 Manual Curation

Manual curation involves several critical considerations to ensure the accuracy of spike sorting results. These rules of thumb represent common practices employed by neuroscientists during manual curation. They serve as guiding principles to ensure the accuracy and reliability of spike sorting results, aiding researchers in effectively identifying and classifying neuronal spikes in electrophysiological data.

First, it is essential to establish a clear refractory period between spikes, indicating distinct neuronal origins and avoiding confusion with refractory artifacts. Additionally, identifying spikes with a sufficiently large amplitude is crucial as it often signifies robust neuronal activity, thereby improving the signal-to-noise ratio.

Furthermore, selecting spikes with dissimilar waveforms compared to neighboring spikes helps distinguish distinct neuronal sources, reducing the risk of contamination or overlap between clusters. Consistency in waveforms across multiple recordings is also important, as it boosts confidence in clustering assignments and minimizes the inclusion of noise or artifacts.

Finally, spikes must surpass a predefined threshold to ensure the exclusion of low-amplitude noise or irrelevant activity from the analysis. Adhering to these criteria during manual curation enables researchers to effectively identify and classify neuronal spikes, leading to more accurate and reliable analyses of electrophysiological data.

4.1.3.2 Automated Curation

The expansion of the number of channels through the utilization of high-density and large-count neural probes significantly increases the complexity of manual curation in electrophysiological data analysis. Manual curation is not only time-consuming and labor-intensive but is also susceptible to human biases, requiring a high level of expertise [83]. Operator

bias can manifest itself in various forms, including preconceived notions about expected outcomes, subjective judgments regarding waveform similarity, and inconsistencies in decision-making across different operators.

Addressing these challenges necessitates the development of automated or semi-automated curation methods that can efficiently handle the burgeoning volume of data while mitigating the impact of operator bias. These methods often leverage machine learning algorithms and advanced signal processing techniques to expedite the curation process and enhance objectivity in spike sorting, such as XClust [116], SimpleClust [117], and Plexon Offline Sorter [118].

(a) Automated Curation: Consensus Method

One potential automated method for evaluating spike sorting results is the so-called "consensus method," as proposed by Buccino et al. [136]. This method relies on combining the results of multiple spike sorting runs or algorithms to improve the reliability and accuracy of cluster assignments. This approach aims to mitigate variability and uncertainty by aggregating information from diverse sources. First, spike sorting algorithms are applied to the same dataset multiple times, generating individual sets of cluster assignments. Then, a consensus clustering is created by integrating these assignments, often through voting or averaging. Spikes consistently assigned to the same cluster across multiple runs are deemed more reliable and contribute to the final cluster assignments.

By leveraging the collective information from multiple runs or algorithms, the consensus method enhances the robustness of cluster assignments and facilitates a more accurate analysis of neuronal activity. However, running spike sorting with the consensus method may require more memory compared to running only one algorithm, primarily due to the need to store results from multiple runs and perform the consensus-building process.

(b) Automated Curation: Evaluation Metrics

Given that recorded neural datasets lack "ground truth" information, any sorting algorithm is prone to errors. Quality metrics serve to identify the types and severity of these mistakes, providing insights into the sorting process. Automated curation methods in spike sorting leverage evaluation metrics to objectively assess and refine the results. These metrics play a pivotal role in quantifying algorithm performance, providing a basis for comparison, and guiding the automated curation process. Various key evaluation metrics are commonly employed in the context of spike sorting to assess clustering quality and improve the accuracy of sorted data.

In general, evaluation metrics attempt to quantify the "goodness" of a unit by considering factors such as the signal-to-noise ratio in a signal or the geometric and temporal separation of spikes from one another [67, 83]. Some of the common quality metrics generated by KiloSort 2 are listed below:

- Firing Rate: Measures the average rate at which a neuron fires spikes, indicating its activity level. A high firing rate indicates robust neuronal activity and is typically associated with well-isolated units. Low firing rates may indicate noisy or poorly isolated units.
- Presence Ratio: Calculates the proportion of spikes in the dataset that are assigned to a particular cluster, providing an indication of how well the cluster represents the neuronal activity. A high presence ratio suggests that a cluster contains a significant proportion of spikes in the dataset, indicating good representation of neuronal activity. Conversely, a low presence ratio may indicate underrepresentation or contamination by noise.

- ISI Violations (Inter-Spike Interval): Identifies spikes that violate the refractory period between subsequent spikes, indicating potential contamination by noise or artifacts. Excessive ISI violations suggest that spikes are not well-separated in time, indicating potential contamination by noise or artifacts.
- Signal-to-Noise Ratio (SNR): Quantifies the ratio of the amplitude of the spike to the background noise level, indicating the quality of spike detection. A high SNR indicates that spike waveforms are well-separated from background noise, suggesting high-quality spike detection.
- Isolation Distance: Measures the distance from each spike to its nearest neighboring cluster, providing an indication of how well-isolated a cluster is from other clusters. A high isolation distance indicates good separation between clusters, suggesting well-isolated units. Low isolation distances may indicate clusters that are not well-separated from others.
- L-ratio: Assesses cluster separation by comparing the number of false positive spikes to the number of true positive spikes in a cluster. A low L-ratio indicates better separation between clusters, suggesting higher-quality sorting results.
- d-prime: Evaluates the discriminability of spike waveforms between clusters, providing a measure of the ability to distinguish between neuronal units. Higher d-prime values indicate better discriminability between clusters, suggesting more reliable spike sorting.

As outlined here, each evaluation metric offers valuable insights that neuroscientists utilize to assess the outcomes of automated spike sorting algorithms. However, these metrics do not encompass dynamic changes or temporal patterns within the recorded signal, which could be

significant, particularly when addressing issues like drifting. Drifting refers to the phenomenon where neurons shift in relation to the recording electrodes. Despite several algorithms being proposed to address drifts, there is still a lack of systematic comparison regarding the performance of these algorithms.

Although automated curation methods expedite the refinement of spike sorting results, their outputs may not always be dependable, especially in the presence of common measurement issues such as drifting and noise. Despite several algorithms being proposed to address drifts, there is still a lack of systematic comparison regarding the performance of these algorithms [137]. Having access to dynamic information about clusters over time could facilitate the detection of dynamic occurrences at specific points within the time domain that impact multiple clusters equally.

Among the metrics mentioned previously, our primary focus is on the L-ratio and its optimization. The L-ratio serves as a pivotal measure, evaluating the compactness and separation of clusters by indicating the proximity of noise or non-unit waveforms to a cluster's center. A low L-ratio value suggests minimal presence of noise waveforms near the unit's cluster, signifying better separation. This metric is crucial as automated algorithms can utilize it to pinpoint clusters needing refinement for improved isolation. Typically, neuroscientists establish a threshold value (usually 0.1) for the L-ratio to discern between good units and noisy ones.

In this chapter, we delve into the L-ratio and present a modified version tailored for dynamic time context, enriching the calculation with more temporal information. This advanced approach yields dynamic and detailed insights into clusters, furnishing neuroscientists with a powerful tool to evaluate, refine, and curate sorted spike clusters with improved accuracy and reliability.

4.2 Dynamic L-ratio

The concept of L-ratio was initially introduced by Schmitzer-Torbert and Redish in 2004 [138], primarily applied to measure how well each cluster (i.e., each putative cell) was separated from other clusters and noise events recorded on the same probe. For each cluster, the squared Mahalanobis distance (D^2) from the center of the cluster is calculated to every spike in the data set using the covariance matrix based on spikes assigned to the cluster. The Mahalanobis distance has the effect of scaling the spikes in the cluster to unit variance in all dimensions.

Under the assumption that the spikes in the cluster distribute normally in each dimension, D^2 for spikes in a cluster will follow a chi-squared (χ^2) distribution with df degrees of freedom. The contamination level of a given cluster is denoted by L and is calculated as the sum of the probabilities that each spike, which is not a cluster member, should actually be a part of the cluster. The probability of cluster membership for each spike is determined as the inverse of the cumulative distribution function (CDF) for the χ^2 distribution with df degrees of freedom. Spikes that are close to the center of the cluster (C) will have probabilities approaching 1, whereas spikes far from the center of the cluster (C) will have probabilities approaching 0. For each cluster (C), L is calculated as follows [138]:

$$L(C) = \sum_{i \notin C} 1 - CDF_{\chi^2_{df}}(D^2_{i,C})$$

In this equation, $i \notin C$ is the set of spikes that are not members of the cluster (C). $CDF_{\chi^2_{df}}$ is the cumulative distribution function of the χ^2 distribution with df degrees of freedom. $D^2_{i,C}$ is the squared Mahalanobis distance of spike i from the center of cluster (C).

The Mahalanobis distance, a measure between a sample point and a distribution, is defined as follows: Given a vector y and a distribution characterized by its mean μ and covariance Σ, the squared Mahalanobis distance from y to this distribution is calculated as [138]:

$$D^2 = (y - \mu)^T \Sigma^{-1} (y - \mu)$$

The superscript "T" means "transpose". This distance indicates the distance of y from the mean (μ) in terms of the number of standard deviations.

In $L(C)$ calculation, spikes from other clusters or noise events that are close to the center of cluster (C) will have high probabilities and contribute significantly to this sum, whereas spikes far from the center of cluster (C) will contribute little. A low $L(C)$ value indicates effective separation, suggesting a well-isolated cluster. Conversely, a high L value suggests poor separation, implying potential inclusion of non-cluster spikes and exclusion of true cluster spikes. The L-ratio is obtained by dividing $L(C)$ by the total number of spikes in the cluster, n_C:

$$L_{ratio}(C) = \frac{L(C)}{n_C}$$

By employing a criterion based on L-ratio instead of L alone, clusters with a larger spike count (higher n_C) can accommodate a higher level of contamination while still ensuring effective separation.

A modified version of the L-ratio metric for cluster separation was introduced and evaluated by Omar J. Ahmed, et. al. [139] over the course of seizures. During the seizure, clusters move in a non-monotonic way. In this method, they computed Mahalanobis distances to a surrogate cluster for various subsets of the entire cluster at 8 equal divisions throughout the seizure and 3 divisions within the 10-minute period occurring two minutes after the seizure concluded. They calculated the sum of false negative assignment probabilities under a Gaussian model for this reference cluster, represented by the value L, for each seizure and post-seizure division. Then,

they dynamically normalized L by the number of spikes assigned to the cluster in each time division to estimate contamination rates relative to the number of spikes assigned during that division, termed the dynamic L-ratio.

We introduce a Python toolbox that employs the same dynamic concept to generate a customizable dynamic L-ratio with a definable time division. We then explore the capabilities of the dynamic L-ratio and demonstrate how it can provide additional information about the temporal occurrences in the recorded neural signal.

Figure 4.4 illustrates the overarching concept of dynamic L-ratio calculation. In part (a), we show cluster locations over time, dividing the total time span into small time divisions (Δt). Black line indicates the cluster center [140]. In part (b), we create a scatter plot of the spike features for each time division. Part (c) displays spike waveforms for six example units recorded from four channels, while part (d) shows the distribution of the Mahalanobis distance from the fitted cluster center to the observed spikes. The dashed lines indicate the theoretical distributions for Gaussian distributed spikes.

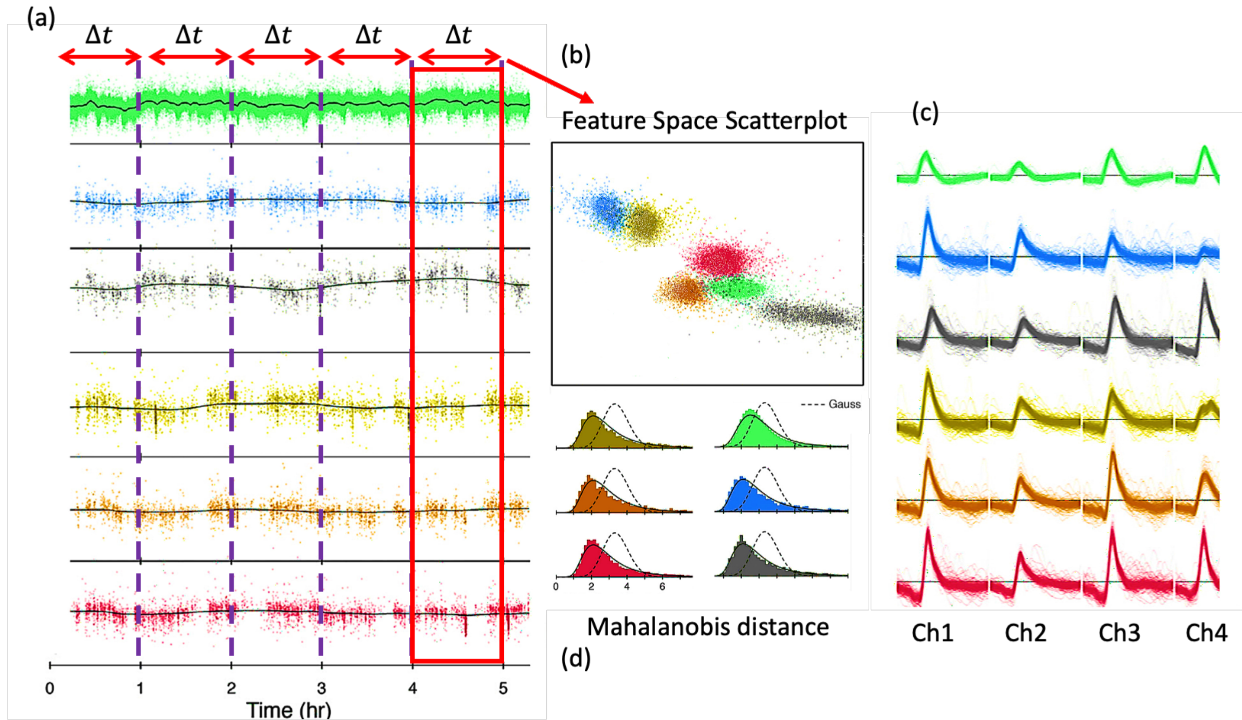

Figure 4.4: Illustration of the dynamic L-ratio calculation process. (a) Cluster locations over time, with the time span divided into small time divisions (Δt). (b) Scatter plot of spike features for each time division. (c) Spike waveforms for six example units recorded from four channels. (d) Distribution of the Mahalanobis distance from the fitted cluster center to the observed spikes, with dashed lines indicating theoretical distributions for Gaussian distributed spikes.

Using this toolbox, the inputs include the "window_duration", which is the time division length defined by the user (Δt), as well as some of the KiloSort 2 outputs. These outputs are as follows:

- pc_features: A matrix containing the principal components (PCs) of the spike waveforms. These PCs represent the shape of the spike waveforms in a lower-dimensional space, aiding in clustering similar spikes together. The pc_features matrix typically has dimensions (number of spikes) x (number of PCs), where each row corresponds to a spike and each column corresponds to a PC.
- pc_feature_ind: A vector containing the indices of the principal components (PCs) used to represent the shape of the spike waveforms. This vector typically has

dimensions (number of PCs), where each element corresponds to a PC and indicates whether it was used in the clustering process. A value of 1 indicates that the PC was used, while a value of 0 indicates that the PC was not used.

- spike_times: A vector containing the times at which each spike occurred. This vector typically has dimensions (number of spikes), where each element corresponds to a spike and indicates the time at which it occurred. The times are usually expressed in seconds or milliseconds, depending on the sampling rate of the recording.
- spike_clusters: A vector containing the cluster assignments for each spike. This vector typically has dimensions (number of spikes), where each element corresponds to a spike and indicates the cluster to which it belongs. The cluster assignments are usually represented as integers, with each integer corresponding to a different cluster.
- spike_templates: A vector containing the template assignments for each spike. This vector typically has dimensions (number of spikes), where each element corresponds to a spike and indicates the template to which it belongs. The template assignments are usually represented as integers, with each integer corresponding to a different template.

The outputs of the dynamic L-ratio function presented in this toolbox consist of a series of plots for each cluster. These plots visually illustrate the changes in the L-ratio for each time division over time, along with the number of spikes for each time division represented as bar plots on the same plots.

Figure 4.5 illustrates the transformation of input data by the toolbox into output data. The number of generated plots is equal to the number of clusters, and each output plot includes the number of spikes as blue bars and the dynamic L-ratio calculated for each time division as red dots.

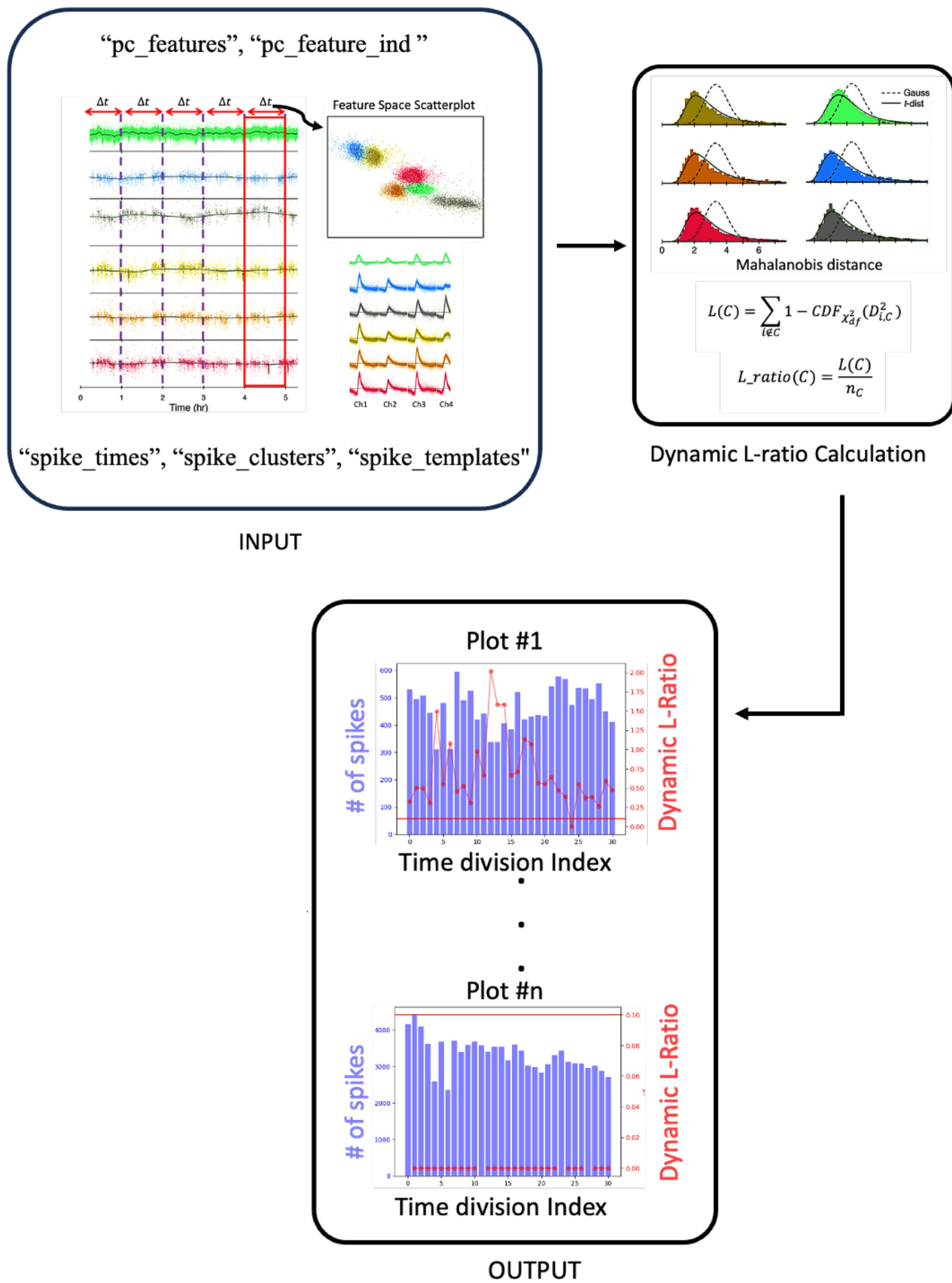

Figure 4.5: A graphical representation of how the input data is transformed or processed by the toolbox to produce the output data.

Adjusting the time window width results in varied output plots. A wider time window reduces computation time but offers fewer temporal details. The choice of time window size depends on the desired level of detail, considering memory and computational capacity constraints.

Figure 4.6 presents various plots for the same cluster, derived from a dataset recorded from a mouse hippocampus provided by Professor Ahmed's lab, University of Michigan, Ann Arbor. The initial cluster, generated by KiloSort, undergoes analysis using the dynamic L-ratio calculation across four distinct window durations (30, 60, 120, and 180 seconds). As illustrated, a reduction in window duration yields access to more nuanced information.

For instance, with a window duration of 30 seconds, the L-ratio exceeds 0.1—considered the threshold for determining whether a cluster is sufficiently isolated from neighboring clusters—only around 40% of the time. This percentage increases to approximately 65% for a 60-second window, and more than 90% for 120-second and 180-second windows. In contrast, if we calculate the L-ratio using the classic method, which considers the entire time span, the resulting L-ratio value of 0.22 surpasses the 0.1 threshold, which fails to provide detailed insights into the changes in cluster separation over the time span.

When analyzing neural data, it is common to encounter noise or drift over time. Noise can arise from various sources, such as electrical interference or physical movement of the recording electrode, while drift refers to slow changes in the recording conditions or the neural signals themselves. These factors can make it challenging to accurately identify and isolate individual neurons or clusters of neurons.

The dynamic L-ratio calculation can be particularly useful in such scenarios because it evaluates cluster isolation over time. By using shorter window durations, the analysis can capture

changes in cluster isolation caused by noise or drift. This allows researchers to identify periods of time when a cluster is well-isolated and when it may be contaminated by noise or affected by drift.

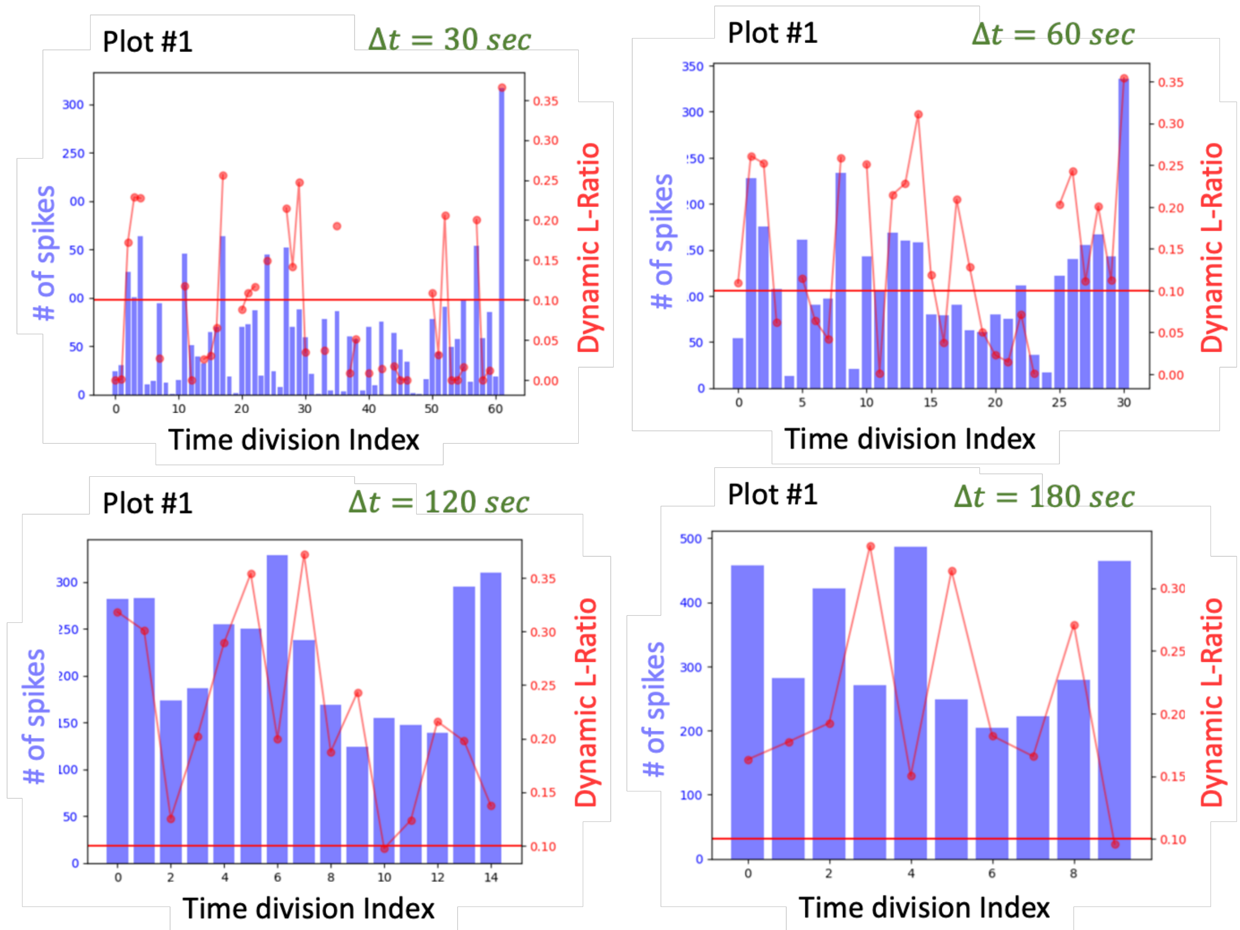

Figure 4.6: Dynamic L-ratio Analysis of a KiloSort Cluster Over Time Windows of Varying Durations (30, 60, 120, and 180 seconds) in a Mouse Hippocampus Dataset provided by Professor Ahmed's lab, University of Michigan, Ann Arbor.

4.3 Spike Sorting using Dynamic L-ratio

As previously described, the dynamic L-ratio can offer a wealth of internal temporal information. This information is invaluable for determining whether a cluster initially labeled as non-spike or noise is mislabeled (a false negative). Additionally, it can reveal instances where a

cluster initially sorted as good is mislabeled and not isolated during certain time segments. These insights are unattainable when using the classic L-ratio metric.

To evaluate the additional information provided by the dynamic L-ratio compared to the classic L-ratio, we have developed a learning algorithm. This algorithm utilizes the feature matrix provided by KiloSort 2, along with the dynamic spike rate and dynamic L-ratio, instead of the firing rate and classic L-ratio data.

This approach allows us to create a new feature matrix for our learning algorithm. By training it with a manually sorted and labeled dataset, which distinguishes between good and bad clusters, we can then apply the algorithm to sort the remaining datasets. This enables us to compare the results obtained using the classic L-ratio vs the dynamic L-ratio.

Figure 4.7 illustrates the supervised learning process for the spike sorting algorithm. Initially, the raw data undergoes sorting via Kilosort 2. Subsequently, the 'pc_features', 'pc_features_ind', 'spike_time', 'spike_clusters', and 'spike_template' outputs from Kilosort 2, along with the pre-defined window duration, are utilized as inputs for the dynamic L-ratio computation toolbox.

The resulting outputs consist of arrays of dynamic L-ratio data and dynamic firing rates, which are employed as the feature matrix for the learning algorithm. As this is a supervised learning algorithm, it is initially trained using a manually curated and labeled dataset. Once trained, it is capable of predicting labels for clusters within new datasets.

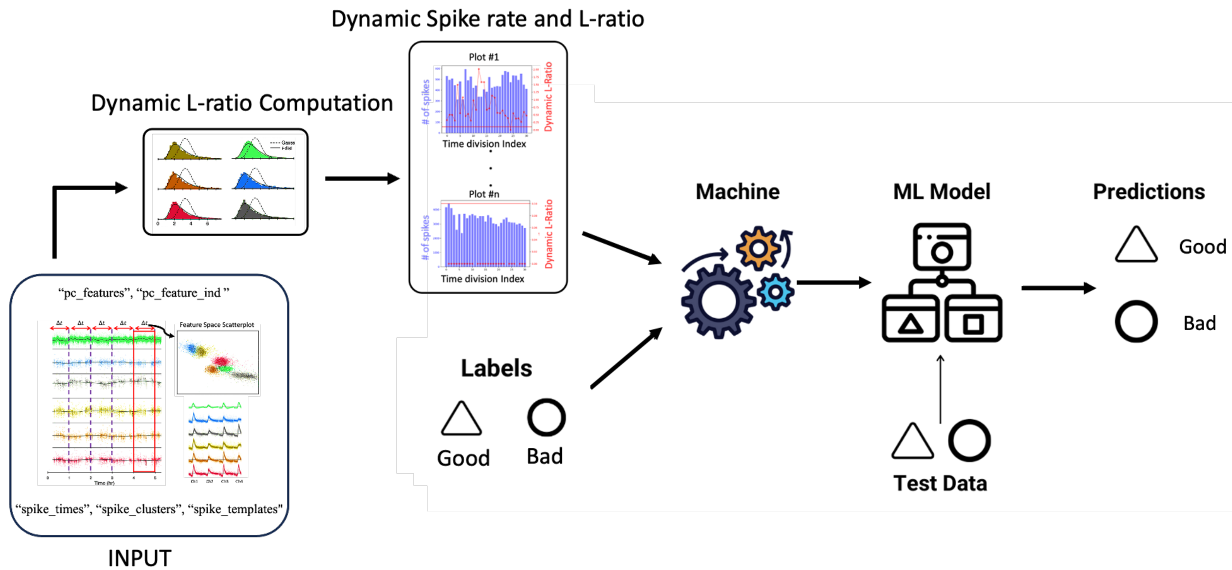

Figure 4.7: Illustration of the supervised learning process for the spike sorting algorithm, showcasing the sequential steps from raw data sorting to the generation of dynamic L-ratio data and dynamic firing rates, ultimately leading to the classification of clusters within new datasets.

The labels 'Good' and 'Bad' are assigned to clusters based on the presence or absence of actual spikes, respectively. 'Good' clusters contain significant or meaningful data points, while 'Bad' clusters are primarily composed of noise or irrelevant data.

The learning algorithm employed here is a Multi-Layer Perceptron (MLP) classifier, a neural network model tailored for classification tasks. It operates as a feedforward neural network, featuring multiple layers of nodes, each intricately connected to the subsequent layer.

The MLPClassifier is part of the sklearn.neural_network module in Python's scikit-learn library. The model is highly customizable, allowing users to specify the number of neurons in each hidden layer, the activation function used for the hidden layers, the optimization algorithm used to train the model, the maximum number of iterations for training, and the random seed used for initializing the weights of the model.

The employed MLP classifier is configured with a single hidden layer consisting of 900 neurons, utilizing the ReLU (Rectified Linear Unit) activation function for the hidden layer, the Adam (Adaptive Moment Estimation) optimization algorithm for training, and a maximum of 1000 iterations for training.

ReLU is a popular choice in deep learning models due to its computational efficiency and ability to mitigate the vanishing gradient problem. It is defined as

$$f(x) = \max(0, x)$$

It returns the input value if it is positive, and zero otherwise. This simple non-linear function has several advantages. Firstly, ReLU is computationally efficient to compute and differentiate, making it a popular choice in deep learning models. Secondly, ReLU helps mitigate the vanishing gradient problem, which can occur when using activation functions like the sigmoid or tanh functions. The vanishing gradient problem can make it difficult to train neural networks, but ReLU's linear behavior for positive inputs can help prevent the gradients from vanishing. Lastly, ReLU is a sparse activation function, meaning that it can output zero for a large portion of its input space. This can help reduce the computational cost of training the neural network. The rectified linear unit function is used for addressing the problem of optimization by mapping nonlinearity into the data [141]. The reason we chose ReLU was that it could be used to alleviate the problem of gradient disappearance.

Gradient disappearance, also known as the vanishing gradient problem, is a common issue in deep learning models. It occurs when the gradients of the loss function with respect to the weights of the network become very small as they are propagated backward through the layers of the network during training. As a result, the weights of the network are not updated effectively, and the model fails to learn. In a single layer neural network, the vanishing gradient problem can

occur if the input values to the network are very large or very small, causing the gradients of the activation function to become very small. This can make it difficult for the optimization algorithm to update the weights effectively, and the model fails to learn.

The Adam optimization algorithm is an adaptive learning rate optimization algorithm that is well-suited for training deep neural networks. The Adam optimizer is a popular optimization algorithm used in training deep learning models. The Adam optimizer maintains two moving averages: the first moment (mean) and the second moment (uncentered variance) of the gradients. These moving averages are used to adjust the learning rate for each parameter during training.

The Adam optimizer is defined by the following update rule: (number equations and cite reference)

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

$$\widehat{m}_t = \frac{m_t}{(1 - \beta_1^t)}$$

$$\hat{v}_t = \frac{v_t}{(1 - \beta_2^t)}$$

$$\theta_t = \theta_{t-1} - \alpha \frac{\widehat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}$$

- m_t is the first moment (mean) of the gradients at time step t,
- v_t is the second moment (uncentered variance) of the gradients at time step t,
- $\widehat{m}_t$ and $\hat{v}_t$ are bias-corrected estimates of the first and second moments, respectively,
- θ_t is the parameter being updated at time step t,
- g_t is the gradient of the loss function with respect to θ_t,
- α is the learning rate,

- β_1 and β_2 are the exponential decay rates for the first and second moments, respectively,
- ϵ is a small constant added to the denominator to prevent division by zero.

The goal of the optimization algorithm is to minimize the loss function by updating the network parameters. The loss function measures how well the model is performing, and the goal is to find the set of parameters that minimize this function. Here, the optimization process starts with initializing two moving averages: the first moment (mean) and the second moment (uncentered variance) of the gradients. These moving averages are used to adjust the learning rate for each parameter during training.

Adam adaptively adjusts the learning rate for each parameter during training, which can lead to faster convergence and better performance. Intuitively, it can be thought of as descending a high-dimensional landscape. If we could project it onto a 2D plot (Figure 4.8), the height of the landscape would represent the value of the loss function, and the horizontal axis would represent the values of the neural network parameters. Ultimately, the goal is to reach the bottom of the landscape by iteratively exploring the space around us.

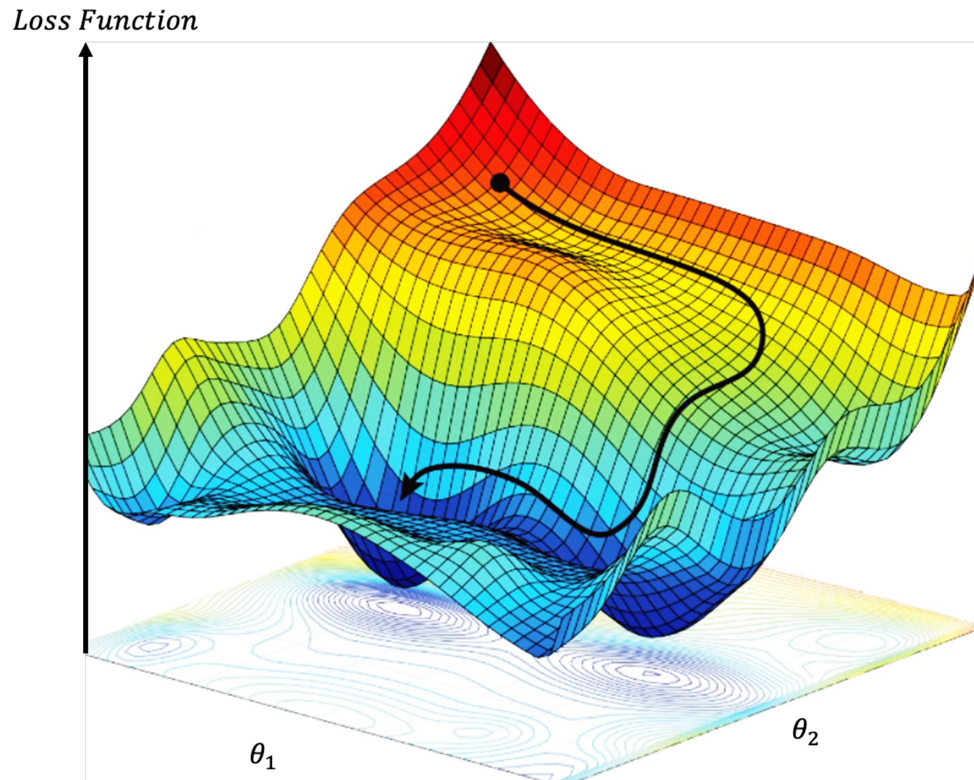

Figure 4.8: Visualization of the “loss” landscape travelled by the Adam stochastic optimization algorithm. Adam seeks to minimize loss by traversing the network and adjusting parameters. Reproduced from [142].

The MLP classifier trains iteratively, as at each time step, the partial derivatives of the loss function with respect to the model parameters are computed to update the parameters. The maximum number of iterations is set to 1000 to ensure that the model has sufficient time to converge to a satisfactory solution.

The learning algorithm applied to the clusters to classify them into 'Spike' and 'Non-Spike' classes is summarized in Figure 4.9. The first step involves balancing the unbalanced training dataset. Subsequently, the training dataset is utilized to train a neural network with a single hidden layer comprising 900 nodes and a ReLU activation function. The Adam optimization algorithm is employed with 1000 iterations to predict the labels of the new dataset.

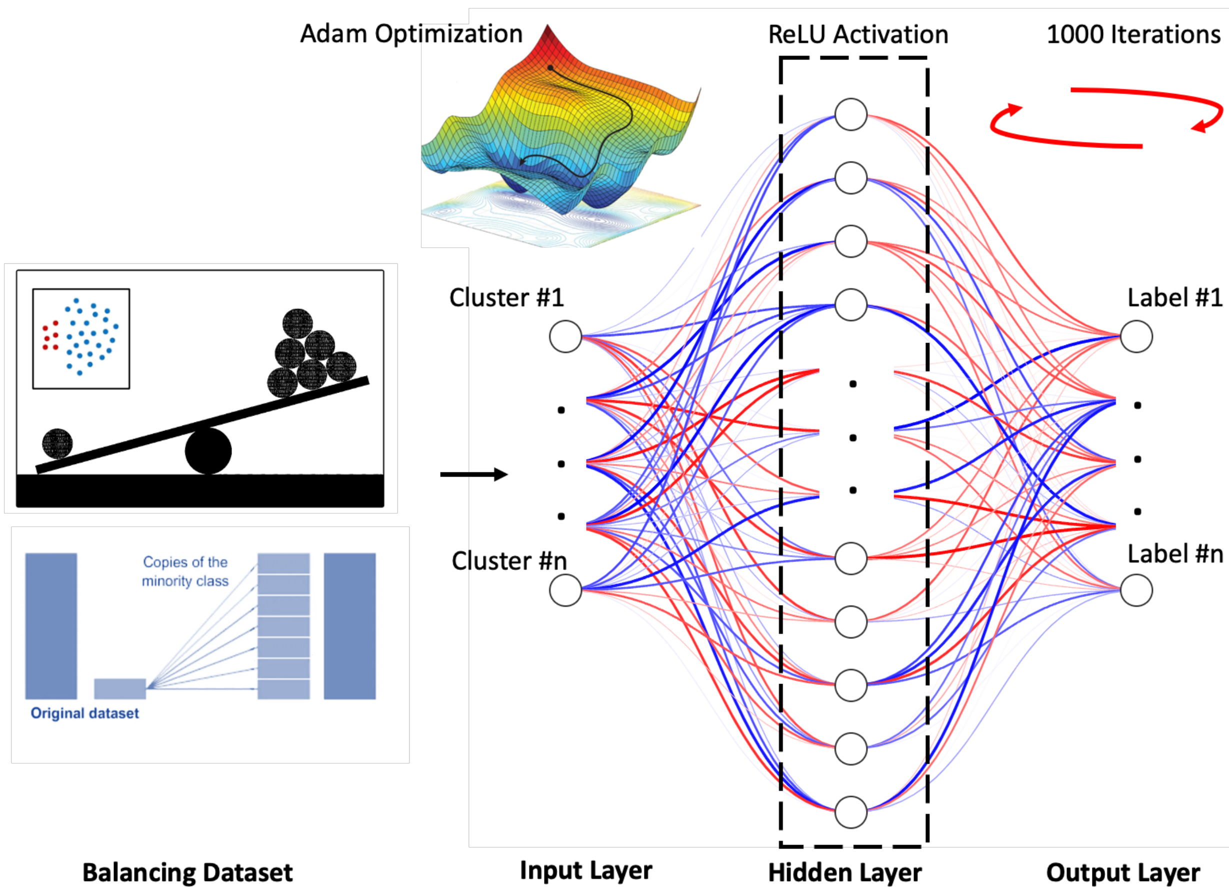

Figure 4.9: Illustration of the learning algorithm applied to classify clusters into 'Good' and 'Bad' classes.

By applying the trained algorithm to 20 datasets and comparing their predicted labels to their manually curated labels, we could calculate several metrics, including ‘accuracy’, ‘precision’, ‘recall’, and ‘F1-score’. The definitions of these metrics are as follows:

- Accuracy: The proportion of correctly classified instances out of the total instances. It is calculated as the number of correct predictions (True positive and true negative) divided by the total number of predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- Precision: The proportion of true positive predictions (correctly predicted 'Good' instances) out of all positive predictions (both true positive and false positive). It is calculated as the number of true positive predictions divided by the sum of true positive and false positive predictions.

$$Precision = \frac{TP}{TP + FP}$$

- Recall: The proportion of true positive predictions (correctly predicted 'Good' instances) out of all actual positive instances (both true positive and false negative). It is calculated as the number of true positive predictions divided by the sum of true positive and false negative instances.

$$Recall = \frac{TP}{TP + FN}$$

- F1 Score: The harmonic mean of precision and recall. It is a single metric that combines both precision and recall, providing a balanced measure of a classifier's performance. It is calculated as:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Accuracy is commonly used when the classes are balanced, meaning there are roughly equal numbers of instances in each class. It is a good overall measure of the classifier's performance but can be misleading when the classes are imbalanced. Precision is useful when the cost of false positives is high. For example, in a medical diagnosis scenario, a false positive could lead to unnecessary treatment or surgery. In such cases, precision is a more important metric than recall. Recall is useful when the cost of false negatives is high. For example, in a fraud detection scenario, a false negative could result in a fraudulent transaction going undetected. In such cases, recall is a more important metric than precision. F1-score is a balanced measure of precision and

recall and is useful when there is an uneven class distribution or when both false positives and false negatives are equally important. It is often used in binary classification tasks.

Here, as both false positives and false negatives are equally important, we calculate all the metrics and compare them to when we classify the clusters using the feature matrix provided by KiloSort 2, along with the classic L-ratio data instead of dynamic L-ratio and dynamic firing rate data, then compare the results as shown in Figure 4.10.

Based on the provided F1-scores for both the dynamic L-ratio and classic L-ratio, we can observe that the dynamic L-ratio consistently outperforms the classic L-ratio in terms of F1-score. The F1-scores for the dynamic L-ratio range from 0.79 to 0.96 (maximum of 1), while the F1-scores for the Classic L-ratio range from 0.39 to 0.69.

This suggests that training the same learning algorithm using dynamic data from dynamic L-ratio and dynamic firing rate results in much more accurate predictions compared to when it is only trained using classic L-ratio data. All the datasets are provided and manually curated by Professor Ahmed's Lab at the University of Michigan, Ann Arbor.

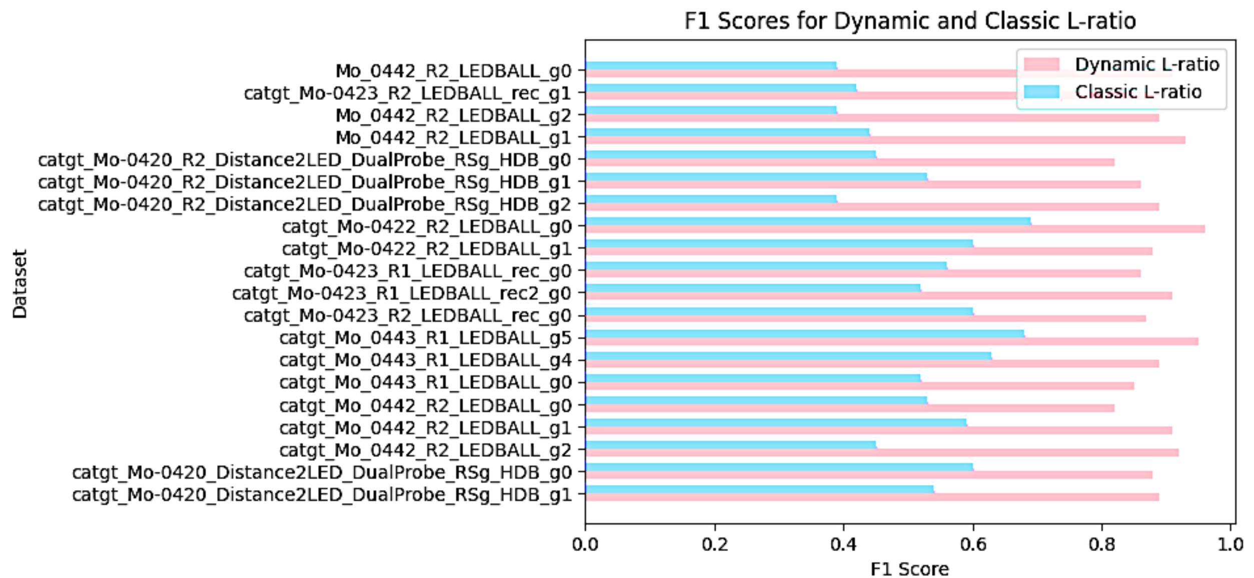

Figure 4.10: Performance comparison of the prediction results when training the same learning algorithm using dynamic data from dynamic L-ratio and dynamic firing rate vs. when it is only trained using classic L-ratio data.

4.4 Conclusion

Spike sorting is a crucial step in neuroscience research, allowing for the extraction and organization of neural spikes from noisy extracellular recordings. This method enables the differentiation of signals from various neurons, offering valuable insights into individual neuron activity and its role in information processing, learning, and memory.

When dealing with large datasets from high-count and high-density neural interfaces, manual spike sorting becomes impractical. This necessitates the use of automatic software and algorithms to cluster the recorded data into spike and non-spike classes. After automatic spike sorting, the results must be curated either manually or with the assistance of quality metrics such as the L-ratio, which indicates the isolation of the clusters.

In this chapter, we introduced a Python Toolbox designed to calculate the dynamic L-ratio for clusters generated by KiloSort 2. This new evaluation metric provides more temporal information compared to the classic L-ratio. We then demonstrated that leveraging this new information, which includes dynamic firing rate and dynamic L-ratio for each cluster, can be used to train a neural network learning algorithm in a supervised learning process. Subsequently, the trained algorithm can be employed to automatically curate the sorted clusters by KiloSort 2. It is worth noting that this approach could potentially be extended to other sorters, although we have not yet explored this possibility. The results of our study indicate that the dynamic L-ratio consistently outperforms the classic L-ratio in terms of F1-score. Specifically, the F1-scores for the dynamic L-ratio range from 0.79 to 0.96, while the F1-scores for the classic L-ratio range from 0.39 to 0.69. This suggests that training the same learning algorithm using dynamic data from both dynamic L-ratio and dynamic firing rate results in approximately 50-100% better in predictions compared to when it is only trained using classic L-ratio data, all while taking almost the same

computation time. Overall, the introduced dynamic L-ratio toolbox is a valuable tool for neuroscientists, aiding in the precise tracking of neurons recorded on large-scale recording devices over time. This helps to address potential drift in neuronal isolation, ensuring accurate analysis and monitoring of neuronal activity.

Chapter 5 Conclusion and Future Works

This research significantly contributes to the development of high-density and high-count neural interfaces by addressing several key challenges and developing innovative solutions. The primary focus is on technological and computational advancements to overcome these challenges. The main contributions of this research are:

1. Technological Innovations for Microfabricating High-Count, High-Density Probes:
    a. Development of innovative methods for microfabricating high-count, high-density probes with user-defined features such as density, size, shape, and distribution.
    b. Creation of a new class of large count (> 64 recording sites), high-density (6μm vertical separation and 10.5μm horizontal separation between recording sites), multiple shanks (at least 4-shanks), silicon-based 2D planar neural probes.
    c. Introduction of a new class of 3D non-planar (500μm height) silicon-based neural probes consisting of slender and tall needles each supporting one recording/stimulation site, capable of accommodating ultra-high count recording sites (100×100) with design versatility enabling user-defined array size, density (400 electrodes per mm^2), and distribution.
2. Incorporation of Mechanically and Electrically Engineered Features:
    a. Engineered T-bar and π-bar shaped cross sections of the silicon-based planar probes to achieve sufficient mechanical stiffness for each probe shank for precise implantation without buckling, while reducing the volume and cross-sectional area of each shank.

b. Incorporation of fine shanks, small stiffeners, and unique cross-sectional designs to improve mechanical performance without the need for thicker shanks, resulting in probes with remarkable straightness, compact cross-sectional areas, and flexibility.
c. Reducing the mechanical stress around recording site junctions in planar probes by using a stepped recoding site design to improve long-term recording reliability.
d. Optimization of probe geometry to be more mechanically compliant with brain tissue, that could potentially reduce foreign body response and ensuring probes can be implanted at the targeted region without buckling.
e. Creation of technologies for forming precisely defined recording/stimulating sites on extremely high-density and large count (100×100) 3D non-planar probes using electroplated photoresist and single-sided etch and double-sided etch techniques.

3. Packaging and Integration Innovations:
   a. Introduction of Polyimide and Parylene-C based integrated flexible cables for easier handling of the 3D and 2D probes, respectively.
   b. Development of a packaging method that incorporates a Polyimide-based PCB, facilitating the easy and cost-effective connection of connectors or active electronics to the rear end of the 3D non-planar probes, eliminating the need for wire bonding.
4. Electrical and Mechanical Characterizations:
   a. Electrical characterizations, along with in-vitro tests for stimulation and recording. In planar probes, at 1 kHz, the magnitude of impedance for the sites after gold plating varied between 300 kΩ and 400 kΩ, while the phase of the impedance varied between -80 and -63 degrees. For non-planar probes, the average impedance at 1 kHz was approximately 1 MΩ, with a phase of -75 degrees.

b. Chronic impedance measurement tests have been performed in vitro on planar probes (for almost a year) to assess the durability of the recording sites. To evaluate the long-term durability of the two-step coverage site metallization method, the impedance of recording sites in a sample was continuously measured over an 11-month period, during which the probes were immersed in a PBS solution. Among the ten recording sites analyzed, seven exhibited consistent performance over time. All the recorded sites are functional up to 9 months, and 70 percent of them are functional for almost one year.

c. Both planar and non-planar probes are mechanically tested by being implanted into tofu, which mimics brain tissue mechanically. This testing demonstrated that the probes can be implanted and removed without breaking the shanks, indicating their durability and suitability for use in vivo.

5. Computational Techniques for Processing Large Neural Datasets:

   a. Introduction of a Python Toolbox to calculate a dynamic version of the L-ratio, which provides more temporal information compared to the classic L-ratio.

   b. Demonstration that leveraging this new information can be used to train a neural network learning algorithm in a supervised learning process, which can then be employed to automatically curate the sorted clusters by KiloSort 2.

   c. The results of our study indicate that the dynamic L-ratio consistently outperforms the classic L-ratio in terms of F1-score. Specifically, the F1-scores for the dynamic L-ratio range from 0.79 to 0.96, while the F1-scores for the classic L-ratio range from 0.39 to 0.69. This suggests that training the same learning algorithm using dynamic data from both dynamic L-ratio and dynamic firing rate results in more accurate predictions compared to when it is only trained using classic L-ratio data.

Overall, this research significantly advances the development of high-density and high-count neural interfaces by providing innovative technological and computational solutions, improving mechanical and electrical performance, and enhancing data processing techniques for more accurate and efficient neural recording and stimulation.

Future work suggested to further demonstrate the capabilities presented in this research include:

a. Different length and size of non-planar probes: Designing and fabricating non-planar probes with longer shanks (either by using a thicker substrate or bonding multiple substrates on top of each other) to reach deeper brain regions or to accommodate specific experimental requirements. This would expand the application range of the probes and enable more comprehensive neural recordings.

b. Non-planar probes with variant shank sizes: Developing non-planar probes with shanks of different sizes to allow for selective targeting of specific brain regions or cell types. This would provide researchers with more flexibility in their experimental designs and enable more precise neural stimulation and recording.

c. In-vivo tests: Conducting in-vivo tests with both planar and non-planar probes to assess the performance and biocompatibility of the developed neural interfaces in real physiological conditions. This would provide valuable insights into the acute and long-term functionality and safety of the probes when implanted in living organisms.

d. Python toolbox to generate other metrics: Expanding the Python Toolbox to generate additional metrics beyond the dynamic L-ratio, such as spike waveform features, firing rate variability, or spike train correlations. This would enhance the capabilities of the toolbox and provide researchers with a more comprehensive set of tools for analyzing neural data.

These future works would further demonstrate the potential of the developed probes as well as advance the field of high-count high density neural interface technology and contribute to the development of more effective and versatile technological and computational tools for neuroscience research and clinical applications.

To sum up, I hope that the research outlined in this thesis will assist neuroscientists and technologists in unraveling the complexities of brain circuitry. The technological and computational solutions presented here have the potential to generate innovative clinical, therapeutic, and prosthetic solutions for a range of neurological conditions.

Bibliography

1. van Middendorp, J.J., G.M. Sanchez, and A.L. Burridge, *The Edwin Smith papyrus: a clinical reappraisal of the oldest known document on spinal injuries.* Eur Spine J, 2010. **19**(11): p. 1815-23.
2. Galvani, L., *De viribus electricitatis in motu musculari commentarius.* Bononiensis Scientiarum et Artium Instituto atque Academia Commentarii, 1791. **7**: p. 363-418.
3. Zhou, Y., et al. *Ultra-flexible neural probes with electrochemical modified electrodes for reliable, chronical recording*. in *2020 IEEE 33rd International Conference on Micro Electro Mechanical Systems (MEMS)*. 2020. IEEE.
4. Guan, S., et al., *Elastocapillary self-assembled neurotassels for stable neural activity recordings.* Science advances, 2019. **5**(3): p. eaav2842.
5. Wen, X., et al., *Flexible, multifunctional neural probe with liquid metal enabled, ultra-large tunable stiffness for deep-brain chemical sensing and agent delivery.* Biosensors and Bioelectronics, 2019. **131**: p. 37-45.
6. Berényi, A., et al., *Large-scale, high-density (up to 512 channels) recording of local circuits in behaving animals.* Journal of neurophysiology, 2014. **111**(5): p. 1132-1149.
7. Gourikutty, S.B.N. and R. Lim. *Flexible Probe for Electrical Neural Signal Recording*. in *2019 IEEE 69th Electronic Components and Technology Conference (ECTC)*. 2019. IEEE.
8. Krucoff, M.O., et al., *Enhancing nervous system recovery through neurobiologics, neural interface training, and neurorehabilitation.* Frontiers in neuroscience, 2016. **10**: p. 584.
9. Clark, G.M., *The multi-channel cochlear implant: Multi-disciplinary development of electrical stimulation of the cochlea and the resulting clinical benefit.* Hearing research, 2015. **322**: p. 4-13.
10. Rheinberger, M.B. and H.H. Jasper, *Electrical activity of the cerebral cortex in the unanesthetized cat.* American Journal of Physiology-Legacy Content, 1937. **119**(1): p. 186-196.
11. Robinson, D.A., *The electrical properties of metal microelectrodes.* Proceedings of the IEEE, 1968. **56**(6): p. 1065-1071.
12. Skrzypek, J. and E. Keller, *Manufacture of metal microelectrodes with the scanning electron microscope.* IEEE Transactions on Biomedical Engineering, 1975(5): p. 435-437.
13. Wise, K.D., J.B. Angell, and A. Starr, *An integrated-circuit approach to extracellular microelectrodes.* IEEE transactions on biomedical engineering, 1970(3): p. 238-247.
14. Buzsáki, G., *Large-scale recording of neuronal ensembles.* Nature neuroscience, 2004. **7**(5): p. 446-451.
15. Wise, K.D. and K. Najafi, *Microfabrication techniques for integrated sensors and microsystems.* Science, 1991. **254**(5036): p. 1335-1342.
16. Du, J., et al., *Multiplexed, high density electrophysiology with nanofabricated neural probes.* PloS one, 2011. **6**(10): p. e26204.
17. Banks, D., et al. *Microengineered interfaces with the nervous system*. in *IEE Colloquium on Medical Applications of Microengineering*. 1996. IET.

18. Urban, G.A., O. Prohaska, and F. Olcaytug, *Early BioMEMS multi-sensor neuroprobes*, in *BioMEMS*. 2006, Springer. p. 1-13.
19. Banks, D., *Neurotechnology.* Engineering Science & Education Journal, 1998. **7**(3): p. 135-144.
20. Pearce, T.M. and J.C. Williams, *Microtechnology: meet neurobiology.* Lab on a Chip, 2007. **7**(1): p. 30-40.
21. Campbell, P.K., et al., *A silicon-based, three-dimensional neural interface: manufacturing processes for an intracortical electrode array.* IEEE Transactions on Biomedical Engineering, 1991. **38**(8): p. 758-768.
22. Wise, K.D., et al., *Microelectrodes, microelectronics, and implantable neural microsystems.* Proceedings of the IEEE, 2008. **96**(7): p. 1184-1202.
23. Sandoughsaz Zardini, S.A., *Sea of Electrodes Array (SEA): Customizable 3D High-Density High-Count Neural Probe Array Technology*. 2019.
24. Zardini, A.S., et al., *Sea of electrodes array (SEA): extremely dense and high-count silicon-based electrode array technology for high-resolution high-bandwidth interfacing with 3D neural structures.* Biorxiv, 2021: p. 2021.01. 24.427975.
25. Stevenson, I.H. and K.P. Kording, *How advances in neural recording affect data analysis.* Nature neuroscience, 2011. **14**(2): p. 139-142.
26. Urai, A.E., et al., *Large-scale neural recordings call for new insights to link brain and behavior.* Nature neuroscience, 2022. **25**(1): p. 11-19.
27. Vázquez-Guardado, A., et al., *Recent advances in neurotechnologies with broad potential for neuroscience research.* Nature neuroscience, 2020. **23**(12): p. 1522-1536.
28. Schrödel, T., et al., *Brain-wide 3D imaging of neuronal activity in Caenorhabditis elegans with sculpted light.* Nature methods, 2013. **10**(10): p. 1013-1020.
29. Ahrens, M.B., et al., *Whole-brain functional imaging at cellular resolution using light-sheet microscopy.* Nature methods, 2013. **10**(5): p. 413-420.
30. Yamamoto, W. and R. Yuste, *Whole-body imaging of neural and muscle activity during behavior in Hydra vulgaris: effect of osmolarity on contraction bursts.* eneuro, 2020. **7**(4).
31. Mann, K., C.L. Gallen, and T.R. Clandinin, *Whole-brain calcium imaging reveals an intrinsic functional network in Drosophila.* Current Biology, 2017. **27**(15): p. 2389-2396. e4.
32. Aimon, S., et al., *Fast near-whole–brain imaging in adult Drosophila during responses to stimuli and behavior.* PLoS biology, 2019. **17**(2): p. e2006732.
33. Marblestone, A.H., et al., *Physical principles for scalable neural recording.* Frontiers in computational neuroscience, 2013. **7**: p. 137.
34. Kleinfeld, D., et al., *Can one concurrently record electrical spikes from every neuron in a mammalian brain?* Neuron, 2019. **103**(6): p. 1005-1015.
35. Ji, J. and K.D. Wise, *An implantable CMOS circuit interface for multiplexed microelectrode recording arrays.* IEEE Journal of solid-state circuits, 1992. **27**(3): p. 433-443.
36. Najafi, K., K. Wise, and T. Mochizuki, *A high-yield IC-compatible multichannel recording array.* IEEE Transactions on Electron Devices, 1985. **32**(7): p. 1206-1211.
37. Jun, J.J., et al., *Fully integrated silicon probes for high-density recording of neural activity.* Nature, 2017. **551**(7679): p. 232-236.

38. Raducanu, B.C., et al., *Time multiplexed active neural probe with 1356 parallel recording sites.* Sensors, 2017. **17**(10): p. 2388.
39. Zhou, P., et al., *Efficient and accurate extraction of in vivo calcium signals from microendoscopic video data.* elife, 2018. **7**: p. e28728.
40. Salatino, J.W., et al., *Glial responses to implanted electrodes in the brain.* Nature biomedical engineering, 2017. **1**(11): p. 862-877.
41. Jeong, J.-W., et al., *Soft materials in neuroengineering for hard problems in neuroscience.* Neuron, 2015. **86**(1): p. 175-186.
42. Khilwani, R., et al., *Ultra-miniature ultra-compliant neural probes with dissolvable delivery needles: design, fabrication and characterization.* Biomedical microdevices, 2016. **18**: p. 1-20.
43. Szostak, K.M., L. Grand, and T.G. Constandinou, *Neural interfaces for intracortical recording: Requirements, fabrication methods, and characteristics.* Frontiers in Neuroscience, 2017. **11**: p. 665.
44. Hong, G. and C.M. Lieber, *Novel electrode technologies for neural recordings.* Nature Reviews Neuroscience, 2019. **20**(6): p. 330-345.
45. Song, E., et al., *Materials for flexible bioelectronic systems as chronic neural interfaces.* Nature materials, 2020. **19**(6): p. 590-603.
46. Li, H., J. Wang, and Y. Fang, *Bioinspired flexible electronics for seamless neural interfacing and chronic recording.* Nanoscale Advances, 2020. **2**(8): p. 3095-3102.
47. Acarón Ledesma, H., et al., *An atlas of nano-enabled neural interfaces.* Nature nanotechnology, 2019. **14**(7): p. 645-657.
48. Schander, A., et al. *In-Vitro and In-Vivo Longevity Evaluation of Free-Floating Intracortical Silicon-Stiffened Neural Probes*. in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*. 2019. IEEE.
49. Kil, D., et al., *Dextran as a resorbable coating material for flexible neural probes.* Micromachines, 2019. **10**(1): p. 61.
50. Liu, J., et al., *Syringe-injectable electronics.* Nature nanotechnology, 2015. **10**(7): p. 629-636.
51. Xie, C., et al., *Three-dimensional macroporous nanoelectronic networks as minimally invasive brain probes.* Nature materials, 2015. **14**(12): p. 1286-1292.
52. Luan, L., et al., *Ultraflexible nanoelectronic probes form reliable, glial scar–free neural integration.* Science advances, 2017. **3**(2): p. e1601966.
53. Barz, F., et al., *CMOS-compatible, flexible, intracortical neural probes.* IEEE Transactions on Biomedical Engineering, 2019. **67**(5): p. 1366-1376.
54. Skousen, J.L., et al., *Reducing surface area while maintaining implant penetrating profile lowers the brain foreign body response to chronically implanted planar silicon microelectrode arrays.* Progress in brain research, 2011. **194**: p. 167-180.
55. Seymour, J.P. and D.R. Kipke, *Neural probe design for reduced tissue encapsulation in CNS.* Biomaterials, 2007. **28**(25): p. 3594-3607.
56. Loeb, G.E., et al., *Parylene as a chronically stable, reproducible microelectrode insulator.* IEEE Transactions on Biomedical Engineering, 1977(2): p. 121-128.
57. von Metzen, R.P., et al., *Diffusion-limited deposition of Parylene C.* Journal of microelectromechanical systems, 2010. **20**(1): p. 239-250.
58. Bhandari, R., et al., *A wafer-scale etching technique for high aspect ratio implantable MEMS structures.* Sensors and Actuators A: Physical, 2010. **162**(1): p. 130-136.

59. Ji, J., et al., *Microfabricated microneedle with porous tip for drug delivery.* Journal of Micromechanics and Microengineering, 2006. **16**(5): p. 958.
60. Bhandari, R., et al., *A novel masking method for high aspect ratio penetrating microelectrode arrays.* Journal of Micromechanics and Microengineering, 2009. **19**(3): p. 035004.
61. Yoo, J.-M., et al., *Excimer laser deinsulation of Parylene-C on iridium for use in an activated iridium oxide film-coated Utah electrode array.* Journal of neuroscience methods, 2013. **215**(1): p. 78-87.
62. Shandhi, M., et al. *A novel method of fabricating high channel density neural array for large neuronal mapping*. in *2015 Transducers-2015 18th International Conference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS)*. 2015. IEEE.
63. Shandhi, M.M.H., et al., *Reusable high aspect ratio 3-D nickel shadow mask.* Journal of Microelectromechanical Systems, 2017. **26**(2): p. 376-384.
64. Cho, J., et al. *Ultra conformal high aspect-ratio small-gap capacitive electrode formation technology for 3D micro shell resonators*. in *2017 IEEE 30th International Conference on Micro Electro Mechanical Systems (MEMS)*. 2017. IEEE.
65. Heschel, M. and S. Bouwstra, *Conformal coating by photoresist of sharp corners of anisotropically etched through-holes in silicon.* Sensors and Actuators A: Physical, 1998. **70**(1-2): p. 75-80.
66. Rostami, B. and K. Najafi. *Forming Tip Electrodes on 3D Neural Probe Arrays Using Electroplated Photoresist*. in *2022 IEEE Sensors*. 2022. IEEE.
67. Lopez, C.M., et al. *22.7 A 966-electrode neural probe with 384 configurable channels in 0.13 μm SOI CMOS*. in *2016 IEEE International Solid-State Circuits Conference (ISSCC)*. 2016. IEEE.
68. Foster, J.D., et al., *A freely-moving monkey treadmill model.* Journal of neural engineering, 2014. **11**(4): p. 046020.
69. Walker, R.M., et al. *A 96-channel full data rate direct neural interface in 0.13 μm CMOS*. in *2011 Symposium on VLSI Circuits-Digest of Technical Papers*. 2011. IEEE.
70. Norlin, P., et al., *A 32-site neural recording probe fabricated by DRIE of SOI substrates.* Journal of Micromechanics and Microengineering, 2002. **12**(4): p. 414.
71. Herwik, S., et al., *Fabrication technology for silicon-based microprobe arrays used in acute and sub-chronic neural recording.* Journal of Micromechanics and Microengineering, 2009. **19**(7): p. 074008.
72. Lee, Y.-T., et al., *A pseudo 3D glass microprobe array: glass microprobe with embedded silicon for alignment and electrical interconnection during assembly.* Journal of Micromechanics and Microengineering, 2009. **20**(2): p. 025014.
73. Gunasekera, B., et al., *Intracortical recording interfaces: current challenges to chronic recording function.* ACS chemical neuroscience, 2015. **6**(1): p. 68-83.
74. Barrese, J.C., et al., *Failure mode analysis of silicon-based intracortical microelectrode arrays in non-human primates.* Journal of neural engineering, 2013. **10**(6): p. 066014.
75. Scholvin, J., et al., *Close-packed silicon microelectrodes for scalable spatially oversampled neural recording.* IEEE Transactions on Biomedical Engineering, 2015. **63**(1): p. 120-130.
76. Steinmetz, N.A., et al., *Challenges and opportunities for large-scale electrophysiology with Neuropixels probes.* Current opinion in neurobiology, 2018. **50**: p. 92-100.

77. Garcia, S., A.P. Buccino, and P. Yger, *How do spike collisions affect spike sorting performance?* Eneuro, 2022. **9**(5).
78. Hazan, L., M. Zugaro, and G. Buzsáki, *Klusters, NeuroScope, NDManager: a free software suite for neurophysiological data processing and visualization.* Journal of neuroscience methods, 2006. **155**(2): p. 207-216.
79. Jun, J.J., et al., *Real-time spike sorting platform for high-density extracellular probes with ground-truth validation and drift correction.* BioRxiv, 2017: p. 101030.
80. Petersen, P.C., et al., *CellExplorer: A framework for visualizing and characterizing single neurons.* Neuron, 2021. **109**(22): p. 3594-3608. e2.
81. Rossant, C. and K.D. Harris, *Hardware-accelerated interactive data visualization for neuroscience in Python.* Frontiers in neuroinformatics, 2013. **7**: p. 36.
82. Rossant, C., et al., *Spike sorting for large, dense electrode arrays.* Nature neuroscience, 2016. **19**(4): p. 634-641.
83. Wood, F., et al., *On the variability of manual spike sorting.* IEEE Transactions on Biomedical Engineering, 2004. **51**(6): p. 912-918.
84. Barnett, A.H., J.F. Magland, and L.F. Greengard, *Validation of neural spike sorting algorithms without ground-truth information.* Journal of neuroscience methods, 2016. **264**: p. 65-77.
85. Hill, D.N., S.B. Mehta, and D. Kleinfeld, *Quality metrics to accompany spike sorting of extracellular signals.* Journal of Neuroscience, 2011. **31**(24): p. 8699-8705.
86. Magland, J., et al., *SpikeForest, reproducible web-facing ground-truth validation of automated neural spike sorters.* Elife, 2020. **9**: p. e55167.
87. Kozai, T.D., et al., *Mechanical failure modes of chronically implanted planar silicon-based neural probes for laminar recording.* Biomaterials, 2015. **37**: p. 25-39.
88. Rostami, B., et al., *Development of an enhanced porosity AgAgCl reference electrode with improved stability.* Engineering Research Express, 2019. **1**(1): p. 015039.
89. Rostami, B., et al., *Design and Fabrication of an Ultra-low Noise Ag-AgCl Electrode.* arXiv preprint arXiv:1903.10311, 2019.
90. Chen, X., et al., *Shape perception via a high-channel-count neuroprosthesis in monkey visual cortex.* Science, 2020. **370**(6521): p. 1191-1196.
91. Micera, S., J. Carpaneto, and S. Raspopovic, *Control of hand prostheses using peripheral information.* IEEE reviews in biomedical engineering, 2010. **3**: p. 48-68.
92. Roh, H., et al., *Fabrication of high-density out-of-plane microneedle arrays with various heights and diverse cross-sectional shapes.* Nano-Micro Letters, 2022. **14**: p. 1-19.
93. Rousche, P.J. and R.A. Normann, *A method for pneumatically inserting an array of penetrating electrodes into cortical tissue.* Annals of biomedical engineering, 1992. **20**: p. 413-422.
94. Obaid, A., et al., *Ultra-sensitive measurement of brain penetration with microscale probes for brain machine interface considerations.* BioRxiv, 2018: p. 454520.
95. Khanna, P., et al., *Sharpening of hollow silicon microneedles to reduce skin penetration force.* Journal of Micromechanics and Microengineering, 2010. **20**(4): p. 045011.
96. Boonma, A., R.J. Narayan, and Y.-S. Lee, *Analytical modeling and evaluation of microneedles apparatus with deformable soft tissues for biomedical applications.* Computer-Aided Design and Applications, 2013. **10**(1): p. 139-157.

97. Vandekerckhove, B., et al., *Technological challenges in the development of optogenetic closed-loop therapy approaches in epilepsy and related network disorders of the brain.* Micromachines, 2020. **12**(1): p. 38.
98. Narayanan, R.T., D. Udvary, and M. Oberlaender, *Cell type-specific structural organization of the six layers in rat barrel cortex.* Frontiers in neuroanatomy, 2017. **11**: p. 91.
99. Bhandari, R., et al., *A novel method of fabricating convoluted shaped electrode arrays for neural and retinal prostheses.* Sensors and Actuators A: Physical, 2008. **145**: p. 123-130.
100. Lee, K.-W., et al., *Pillar-shaped stimulus electrode array for high-efficiency stimulation of fully implantable epiretinal prosthesis.* Journal of Micromechanics and Microengineering, 2012. **22**(10): p. 105015.
101. Xiang, Z., J. Liu, and C. Lee, *A flexible three-dimensional electrode mesh: An enabling technology for wireless brain–computer interface prostheses.* Microsystems & Nanoengineering, 2016. **2**(1): p. 1-8.
102. Yu, H., et al., *A parylene self-locking cuff electrode for peripheral nerve stimulation and recording.* Journal of Microelectromechanical Systems, 2014. **23**(5): p. 1025-1035.
103. Lebedev, M.A. and M.A. Nicolelis, *Brain–machine interfaces: past, present and future.* TRENDS in Neurosciences, 2006. **29**(9): p. 536-546.
104. Moosavifar, M., Y.K. Cherivirala, and D.D. Wentzloff. *A 320μW Receiver with-58dB SIR Leveraging a Time-Varying N-Path Filter*. in *2022 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*. 2022. IEEE.
105. Moosavifar, M., J. Im, and D.D. Wentzloff, *An Interference-Resilient Bit-Level Duty-Cycled ULP Receiver Leveraging a Dual-Chirp Modulation.* IEEE Journal of Solid-State Circuits, 2023.
106. Moosavifar, M., et al. *A $110\mu\mathrm {W} $2.5 kb/s− 103dBm-Sensitivity Dual-Chirp Modulated ULP Receiver Achieving− 41dB SIR*. in *2022 IEEE International Solid-State Circuits Conference (ISSCC)*. 2022. IEEE.
107. Edell, D.J., et al., *Factors influencing the biocompatibility of insertable silicon microshafts in cerebral cortex.* IEEE Transactions on biomedical engineering, 1992. **39**(6): p. 635-643.
108. Bjornsson, C., et al., *Effects of insertion conditions on tissue strain and vascular damage during neuroprosthetic device insertion.* Journal of neural engineering, 2006. **3**(3): p. 196.
109. Sharp, A.A., et al., *In vivo penetration mechanics and mechanical properties of mouse brain tissue at micrometer scales.* IEEE transactions on biomedical engineering, 2008. **56**(1): p. 45-53.
110. Li, Y., et al., *Fabrication of sharp silicon hollow microneedles by deep-reactive ion etching towards minimally invasive diagnostics. Microsys. Nanoeng. 5, 41 (2017)*. 2019.
111. Wark, H., et al., *A new high-density (25 electrodes/mm2) penetrating microelectrode array for recording and stimulating sub-millimeter neuroanatomical structures.* Journal of neural engineering, 2013. **10**(4): p. 045003.
112. Hoogerwerf, A.C. and K.D. Wise, *A three-dimensional microelectrode array for chronic neural recording.* IEEE Transactions on Biomedical Engineering, 1994. **41**(12): p. 1136-1146.
113. Patel, P.R., et al., *Chronic in vivo stability assessment of carbon fiber microelectrode arrays.* Journal of neural engineering, 2016. **13**(6): p. 066002.

114. Tang, Y., et al., *Ultra deep reactive ion etching of high aspect-ratio and thick silicon using a ramped-parameter process.* Journal of Microelectromechanical Systems, 2018. **27**(4): p. 686-697.
115. Einevoll, G.T., et al., *Towards reliable spike-train recordings from thousands of neurons with multielectrodes.* Current opinion in neurobiology, 2012. **22**(1): p. 11-17.
116. Härdle, W., et al., *XClust: clustering in an interactive way.* XploRe: An Interactive Statistical Computing Environment, 1995: p. 141-168.
117. Voigts, J. *Simpleclust*. 2012; Available from: https://jvoigts.scripts.mit.edu/blog/simpleclust-manual-spike-sorting-in-matlab/.
118. *Plexon offline sorter*. 2020; Available from: https://plexon.com/products/offline-sorter/
119. Rey, H.G., C. Pedreira, and R.Q. Quiroga, *Past, present and future of spike sorting techniques.* Brain research bulletin, 2015. **119**: p. 106-117.
120. Eversmann, B., et al., *A 128× 128 CMOS biosensor array for extracellular recording of neural activity.* IEEE Journal of Solid-State Circuits, 2003. **38**(12): p. 2306-2317.
121. Berdondini, L., et al., *High-density electrode array for imaging in vitro electrophysiological activity.* Biosensors and bioelectronics, 2005. **21**(1): p. 167-174.
122. Frey, U., et al., *Switch-matrix-based high-density microelectrode array in CMOS technology.* IEEE Journal of Solid-State Circuits, 2010. **45**(2): p. 467-482.
123. Ballini, M., et al., *A 1024-channel CMOS microelectrode array with 26,400 electrodes for recording and stimulation of electrogenic cells in vitro.* IEEE journal of solid-state circuits, 2014. **49**(11): p. 2705-2719.
124. Muller, E., et al., *Python in neuroscience*. 2015, Frontiers Media SA. p. 11.
125. Müller, J., et al., *High-resolution CMOS MEA platform to study neurons at subcellular, cellular, and network levels.* Lab on a Chip, 2015. **15**(13): p. 2767-2780.
126. Yuan, X., et al. *A microelectrode array with 8,640 electrodes enabling simultaneous full-frame readout at 6.5 kfps and 112-channel switch-matrix readout at 20 kS/s*. in *2016 IEEE Symposium on VLSI Circuits (VLSI-Circuits)*. 2016. IEEE.
127. Bron, F., *Why not record from every channel with a CMOS scanning probe?*
128. Angotzi, G.N., et al., *SiNAPS: An implantable active pixel sensor CMOS-probe for simultaneous large-scale neural recordings.* Biosensors and Bioelectronics, 2019. **126**: p. 355-364.
129. Marius Pachitariu, N.A.S., Jennifer Colonell. *Kilosort2*. 2018.
130. Lee, J.H., et al., *YASS: yet another spike sorter.* Advances in neural information processing systems, 2017. **30**.
131. Chung, J.E., et al., *A fully automated approach to spike sorting.* Neuron, 2017. **95**(6): p. 1381-1394. e6.
132. Yger, P., et al., *A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings in vitro and in vivo.* Elife, 2018. **7**: p. e34518.
133. Hilgen, G., et al., *Unsupervised spike sorting for large-scale, high-density multielectrode arrays.* Cell reports, 2017. **18**(10): p. 2521-2532.
134. Arrays, D.M., *Automatic Spike Sorting for High.*
135. Pachitariu, M., et al., *Fast and accurate spike sorting of high-channel count probes with KiloSort.* Advances in neural information processing systems, 2016. **29**.
136. Buccino, A.P., et al., *SpikeInterface, a unified framework for spike sorting.* Elife, 2020. **9**: p. e61834.

137. Garcia, S., et al., *A modular implementation to handle and benchmark drift correction for high-density extracellular recordings.* Eneuro, 2024.
138. Schmitzer-Torbert, N. and A.D. Redish, *Neuronal activity in the rodent dorsal striatum in sequential navigation: separation of spatial and reward responses on the multiple T task.* Journal of neurophysiology, 2004. **91**(5): p. 2259-2272.
139. Ahmed, O.J., et al., *Two modes of inhibitory neuronal shutdown distinctly amplify seizures in humans.* MedRxiv, 2020: p. 2020.10. 09.20204206.
140. Shan, K.Q., E.V. Lubenov, and A.G. Siapas, *Model-based spike sorting with a mixture of drifting t-distributions.* Journal of neuroscience methods, 2017. **288**: p. 82-98.
141. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning.* nature, 2015. **521**(7553): p. 436-444.
142. Hutson, M., *AI researchers allege that machine learning is alchemy.* Science, 2018. **360**(6388): p. 861.