8th International IEEE EMBS Conference on Neural Engineering
Shanghai, China, May 25 - 28 , 2017

Scaling Models for Microfabricated *In Vivo* Neural Recording Technologies

J. Scholvin[1], C.G. Fonstad[1], E.S. Boyden[1]
[1]Massachusetts Institute of Technology, Cambridge, MA, USA, email: esb@media.mit.edu

Abstract—**Microfabrication technology can enable extracellular neural recording electrodes with unprecedented wiring density, and the ability to benefit from continued CMOS technology scaling. A neural recording electrode consists of recording sites that sense electrical activity inside the brain, and wiring that routes these signals to neural amplifiers outside the brain. We here introduce a scalable circuit model for recording sites and signal routing, valid for different amplifier integration approaches. We define noise and cross-talk requirements, and analyze how future CMOS technology scaling will drive the ability to record from increasingly large number of sites in the mammalian brain. This analysis provides an important step in understanding how advances of MEMS and CMOS fabrication can be utilized in large-scale recording efforts of many thousands to possibly millions of neurons.**

I. Introduction

Extracellular recording of electrical activity in the brain provides an important tool to understand neural codes at single-spike resolution. Silicon-based recording technologies (reviewed e.g. in [1]) have enabled an increasing number of recording sites. A neuron can be recorded from multiple points in space, a concept first introduced by stereotrode and tetrode wire recordings [2], [3], and implemented in Si-based probes as polytrodes [4] and close-packed electrodes [5]. The components of an extracellular neural probe are shown in Fig. 1. The close packing of recording sites allows the spatial oversampling of neural activity, where at least locally there are more recording sites than nearby neurons. Oversampling can assist with the automation of data analysis, an important requirement for increasing the total number of recording sites.

Neural amplifiers can be integrated with probe shanks both monolithically (active probes) or heterogeneously (passive probes) [6]–[9]. In general, a neural probe consists of many individual shanks arranged into a 2-D or 3-D array [1]. Each shank can contain a dense grid of many recording sites along its length (Fig. 2). The amount of tissue displaced by inserting a probe should be small (e.g. <1% of brain volume). This places an upper bound on the cross-sectional area for each shank, and requires reduction of wiring dimensions when the number of recording sites on the shank's surface increases. Microfabricated CMOS metal wiring is well suited to provide the scaled and dense multi-layer wiring necessary to route recorded signals out of the brain and towards low-noise neural amplifiers (regardless of the choice of active or passive probe architecture). Structural support for the wiring is provided by a thinned-back Si substrate (e.g. 10 to 15 µm thick), reducing tissue displacement while allowing shanks to be inserted into the brain without buckling [10]. The required length for a shank and its wiring depends on the target brain region. Shank lengths can range from 1 mm (mouse cortex), to 1 cm (human cortex or deep mouse brain), and possibly up to 10 cm (deep human brain). Each shank can have multiple columns of recording sites (Fig. 2). To minimize shank width and tissue displacement, the wiring can be routed below the recording sites. Each recording site column thus represents an identical unit design, and we can simplify the analysis by focusing on a single column, regardless of how these unit columns are arranged into multi-column shanks and subsequently into 2-D and 3-D probe structures. We derive an equivalent circuit model and analyze how submicron wiring technology scaling influences neural probe performance metrics.

II. Technology Modeling

We first define an equivalent circuit for neural probe shanks (Fig. 3), by combining models for the recording sites and the scaled submicron wiring connecting to neural amplifiers. The model in Fig. 3 implements the wiring as a cascade of 6-port unit blocks. This arrangement simplifies our

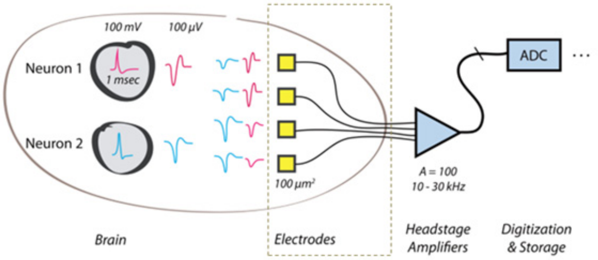

Fig. 1: Principle of extracellular recording. Microelectrodes are inserted into the brain and exposed recording sites measure electrical activity of nearby neurons. Insulated wires route the signals out of the brain, to amplify, digitize and process. Our analysis focuses on the scalability of the recording sites and their wiring, highlighted by the dotted box.

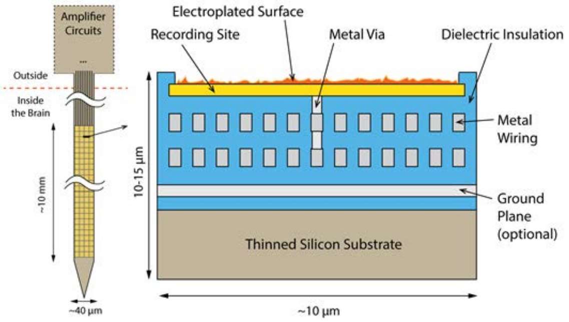

Fig. 2: Diagram and cross-section of a microfabricated neural recording shank, with typical geometries indicated as examples. The shank contains a dense array of recording sites in 4 columns, ideally utilizing the entire surface area (*left*) with minimal vertical cross-section (*right*). Layers of dense wiring connect the recording sites to neural amplifier circuits outside the brain.

978-1-5090-4603-4/17/$31.00 ©2017 IEEE

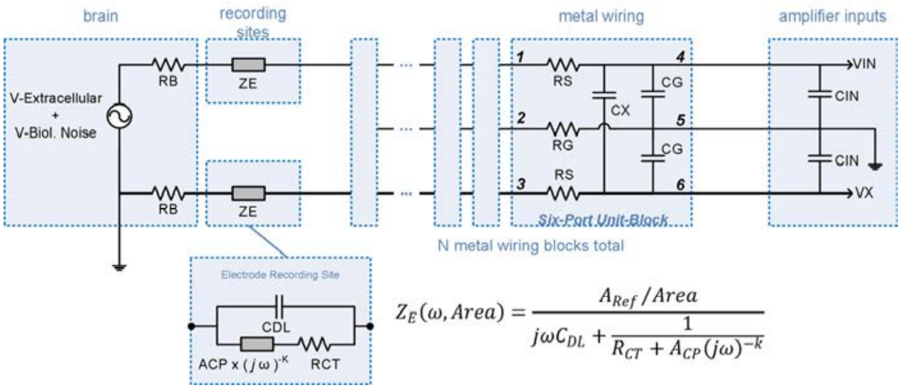

TABLE I
RECORDING SITE MODEL

Parameter	Value
Site material	Gold
Electroplated	PEDOT
A_{Ref}	9x9 μm^2
A_{CP}	16.8 MΩ
C_{DL}	11.9 pF
R_{CT}	52.8 kΩ
k	0.577
Calculated Noise	**Value**
Model	4.31 μV_{rms}
Data, plated	4.23 μV_{rms}
Data, unplated	5.92 μV_{rms}

Fig. 3: Equivalent circuit model of extracellular recording, based on [10], [17], [19]. A neuron's extracellular signal ($V_{Extracellular}$) is picked up by the recording site (Z_E) and routed to the amplifier (V_{in}) through dense metal wiring. The signal can couple to neighboring wires and result in cross-talk (V_x/V_{in}), modeled by including a second recording channel (V_x). We simulate the circuit in Matlab using the S-Parameter toolbox. A six-port S-parameter model was used to simulate both a lumped (one block) or distributed circuit (many blocks). Cross-talk is caused by the capacitive voltage divider of C_G and C_X, but is significantly shunted out by the comparatively low impedance of Z_E. Electroplating of recording sites is important to help reduce cross-talk (from C_X/C_G to C_X*ω/Z_E). Signal attenuation is set by the voltage divider defined approximately by the elements (Z_E+R_S) and (C_G+C_X+C_{IN}).

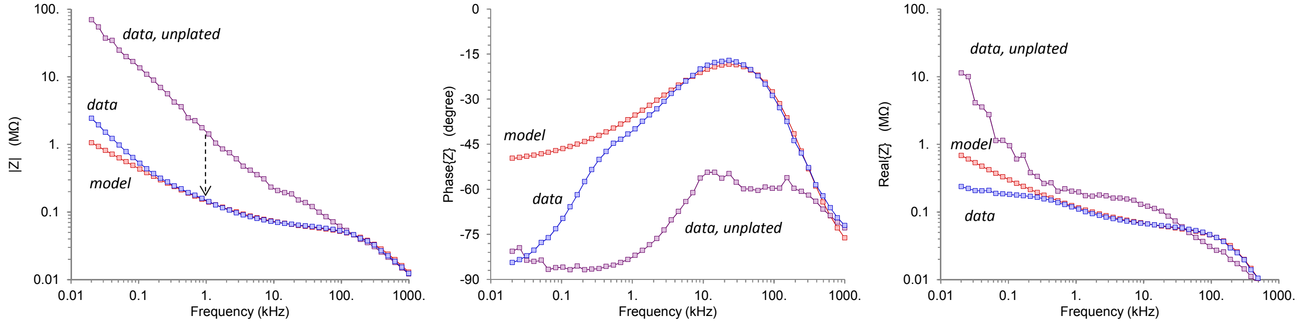

Fig. 4: Measurements of recording site impedance (in 0.9% saline) and model fit (Z_E defined in Fig. 3). The 3-parameter recording site model (see [17]) accurately describes the electroplated recording site across the relevant frequencies of 0.1-15 kHz. Electroplating can reduce impedance by an order of magnitude or more (*left*). Because unplated recording sites are mostly capacitive (near -90° phase, *center*), and a recording site's electrical noise depends on Real{Z} [18] (*right*), electroplating may only moderately reduce the recording site noise (see calculations in Table I).

TABLE II
WIRING MODEL FOR FIG. 3

Parameter	Value	Details
C_{IN}	10 pF	Amplifier input capacitance
C_G	2 pF/cm * (3 + $S_{\%}$)	Equivalent ground capacitance
C_X	2 pF/cm * (1 – $S_{\%}$)	Coupling capacitance
$S_{\%}$	0 to 1	Amount of wire shielding
R_G	10 Ω/cm	Ground plane resistance
R_B	0 Ω	Tissue resistance (negligible)
Z_E	See Fig. 3	Recording site impedance
d	10 nm to 10 μm	Wire width (feature size)
R_S	ρ_{Cu} $f_w(d)/(A/R*d^2)$	Line resistance (see Fig. 7)
$f_w(d)$	1 + 40 nm / d	Size effect, curve fit to ITRS
A/R	See Fig. 6	Wire aspect ratio (height/width)
ρ_{Cu}	2.2 μΩ-cm	Bulk resistivity of metal wire

circuit simulations, and we can use a multi-port network analysis approach that allows us to consider both lumped element models (using a single 6-port block) and distributed element models (by cascading many 6-port blocks) in the same simulation setup.

A. Recording Sites

Recording sites consist of a small exposed electrode (e.g. of 100 μm^2 area), typically electroplated to improve the electrochemical interfacing with the brain. Impedance measurements and their model fit for a 9x9 μm^2 gold recording site are shown in Fig. 4, before and after electroplating with PEDOT ([11]). Table I summarizes model parameters for the equivalent circuit (inset of Fig. 3). In addition, we measured impedances for a broad range of recording site areas (Fig. 5) to confirm that the recording site impedance scales with the inverse of area, allowing us to apply the model fit of Fig. 3 to a wide range of recording site areas.

B. Wire Routing

The dense wiring along the shank (Fig. 2) uses metal wires fabricated at minimum width and spacing. The number of metal layers is limited by practical considerations, and we use layer counts and wire aspect ratios typical for commercial microfabrication processes (Fig. 6). Wire resistance per unit length increases with scaling, not only because of smaller wire cross-sections, but also because of submicron size effects [12]. Fig. 7 shows measurements of the increased resistance against expected bulk resistivity. For deeply scaled technologies, the parasitic capacitance between adjacent wires (per unit length) is independent of scaling, since the cross-sectional geometry remains the same (though slight reductions can occur from different dielectric materials used in scaled technologies). This wire capacitance scaling behavior is different in single-layer fabrication technologies [10]. When the wiring enters the base of the probe (Fig. 2), the available routing space increases and we can treat the routing parasitics outside the probe shank as negligible. For that reason, our results apply to both passive and active architectures, because the majority of the wiring parasitics are contributed by the deeply scaled shank wiring, which is architecture independent.

C. *Neural Amplifiers*

Once the signals are routed out of the brain, they connect to neural amplifier circuits (e.g. [6], [9], [13], [14]), which must minimize power consumption, noise, and circuit area. These amplifiers can be based on highly scaled *in-vitro* neural amplifiers [15], [16], and typically are located outside the brain (but possibly still monolithically integrated with the probe shanks), where area and power constraints are more easily mitigated. We model the amplifier noise as 5 μV_{rms} between 0.1-15 kHz (across similar frequency bands, 2-10 μV_{rms} is common [15]), and model the neural amplifier input impedance with a 10 pF capacitor.

III. Figures of Merit

To understand how scaling impacts device performance, a design has to meet performance requirements for total probe noise (<15 μV_{rms}), signal cross-talk (<1%), and signal attenuation.

A. *Noise*

Noise is contributed by each of the neural probe's building blocks: the recording site, wiring, and neural amplifier. We will refer to the combined noise as the "probe noise". Biological noise is present (e.g. on the order of a few µV) and is caused e.g. by the random activity of many distant neurons (neural "background chatter" in the brain). Therefore, biological noise adds to the input signal, and is independent of the probe noise. When the probe noise is sufficiently reduced, the biological noise will eventually dominate, limiting the benefit of further probe noise reduction. When optimizing design parameters in our simulations, we chose a maximum permissible probe noise of 15 μV_{rms} (0.1-15 kHz).

B. *Cross-talk*

Capacitive coupling between adjacent wires results in signal cross-talk (defined as V_x/V_{in} in Fig. 3). Cross-talk should be low enough so that even a strong signal (e.g. 500 µV) couples into an adjacent wire below the probe noise floor. Therefore, we require cross-talk to be <1%. Cross-talk can be reduced by interleaving ground with signal wires, shielding neighboring signals at the expense of a ~50% lower signal wiring density. We can also isolate a signal wire for only a portion of its length (we will call this approach "fractional shielding"), enabling a more gradual tradeoff between wiring density and cross-talk.

C. *Attenuation*

Signal attenuation in the probe increases the input referred noise contribution of the neural amplifier (and to a lesser extent the input referred noise from backend wiring resistances). In our model, we use a neural amplifier noise of 5 μV_{rms}. We do not place any constraints on signal attenuation, because its influence is accounted for by referring the amplifier noise to the input, i.e. the recording site. Strong attenuation will thus increase the neural amplifier's contribution to the overall input referred probe noise.

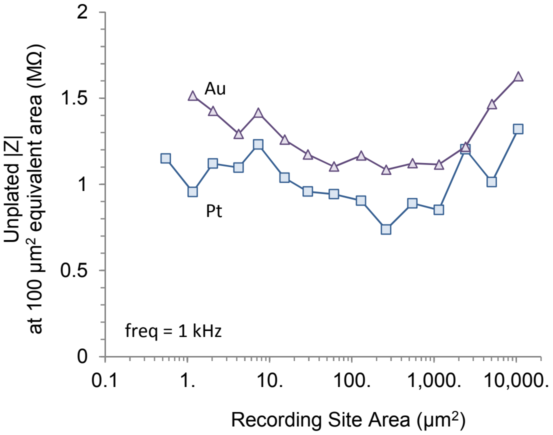

Fig. 5: Measurements of unplated Au and Pt recording sites in 0.9% saline for different site areas, expressed as the equivalent impedance of a 100 µm² site. We fabricated test devices with a constant *Site Count * Area* product. To improve measurement accuracy and eliminate parasitic capacitances, we fabricated the devices on fused silica wafers.

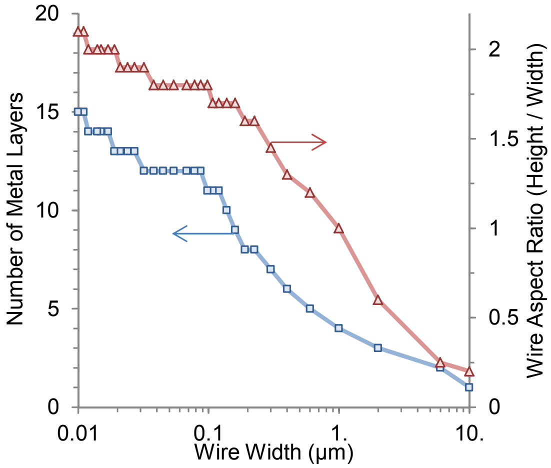

Fig. 6: Model assumptions for layer count and wire aspect ratios, partially based on the 2001, 2003 and 2009 ITRS roadmaps (for wire widths of 0.23 µm and below). Data for larger wire widths were interpolated to avoid unrealistically thick metal wires for the larger geometries.

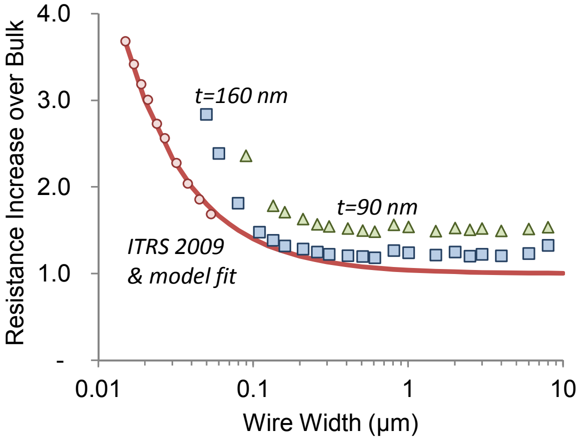

Fig. 7: Measurements showing the presence of size effects (e.g. sidewall scattering on narrow wires), resulting in a strong increase of resistivity below 100 nm. We fabricated Au wires with two wire thicknesses (90 and 160 nm) using electron-beam lithography, and observed similar effects to the ITRS roadmap model. For our wires, surface roughness and grain sizes are not optimized (compared to a commercial process), giving rise to a more accelerated resistance increase (see [12]).

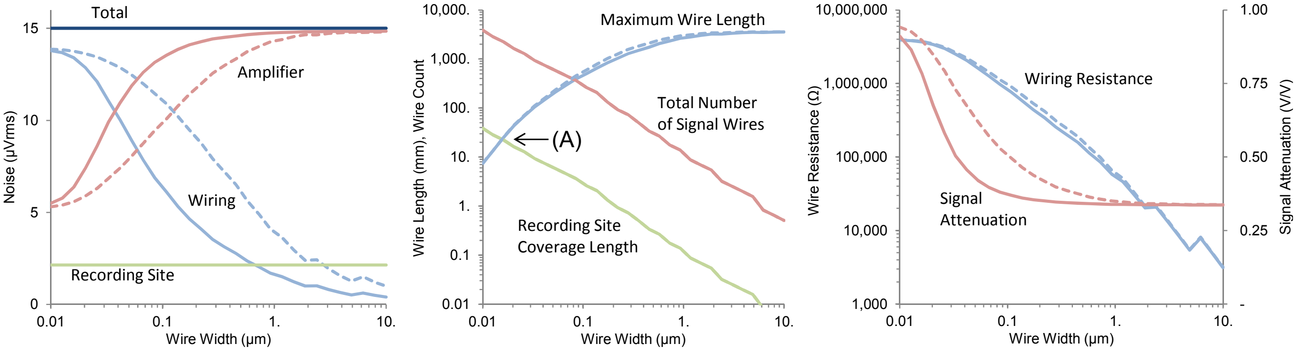

Fig. 8: Simulated impact of CMOS wire scaling for a constant recording site area of 10x10 µm^2, showing noise components (*left*), maximum number of wires and their lengths (*center*) and wire resistance and signal attenuation (*right*). Differences between lumped (solid) and distributed (dashed) circuit models have some impact on the relative noise contribution, but do not significantly change the maximum number of wires or wire lengths. A probe shank can record over a length of (10 µm x Total Number of Signal Wires), defined as the "coverage length". For small wire widths, wire resistance dominates performance and limits the wire's length, while permitting a large number of signal wires. This results in a cross-over point (marked "A"), below which wires cannot be made sufficiently long enough to route the entire coverage length out of the brain (here, approximately 20 mm). Thus, for the above simulations, wire widths below 14 nm would not be able to accommodate all of the recording sites, when using 10x10 µm^2 recording sites and a 15 µV$_{rms}$ noise requirement (the 1% cross-talk requirement is included in the simulation). Curve roughness for large wire widths arises from the steps in layer count and wire aspect ratios (see Fig. 6).

IV. Results

With the model topology of Fig. 3 and the parameters of Tables I and II, we simulate different scenarios of technology scaling: first, the ability to route wires over long distances (i.e. how to record from deep and large brain volumes) and second, the ability to increase recording site packing densities (i.e. how to record as densely as possible).

A. Technology Wiring Capability

Fig. 8 shows the impact of wire width scaling for constant 10x10 µm^2 area recording sites. As the wire width is reduced, more wires can be routed below each recording site column (Fig. 8, center), increasing the total number of possible recording sites per column (and therefore, allowing the probe shank to cover more depth). However, smaller wire widths increase the wire resistance (Fig. 8, right) and its contribution to probe noise (Fig. 8, left). The small wire's resistance places an upper limit on how long each wire can be made (Fig. 8, center) while maintaining the total probe noise below our constraint of 15 µV$_{rms}$.

For very relaxed wire geometries, the noise contribution from wire resistance is not a concern. The wires can be made very long, until eventually parasitic capacitance causes significant signal attenuation (Fig. 8, right). Consequently, the input referred noise of the neural amplifier becomes the bottleneck for increasing wire lengths (Fig. 8, left). For our choice of 5 µV$_{rms}$ neural amplifier noise and a total probe noise constraint of <15 µV$_{rms}$, the relaxed wire geometry will show this bottleneck in very long wires, but for lengths beyond the size of the brain (or a Si wafer). Of course, relaxed wire geometries restrict us to a small number of recording sites and very low recording site densities.

For large-scale recording, we want to use the smallest wire geometries and place as many recording sites per shank column as possible, to maximize the depth over which the shank can record in the brain (we call this the "coverage length"). As we reduce the wire geometry, however, the maximum possible wire length is decreasing (Fig. 8, center). This leads to a cross-over between an increased coverage length and a decreased maximum wire length (Fig. 8, point "A"). The exact cross-over depends on our choice of recording site area (here, for 10x10 µm^2). Geometries smaller than the cross-over point do not offer any benefits: we may be able to add more wires and more recording sites, but we would not be able to make such wires long enough to route the deepest sites out of the brain without exceeding our probe noise constraint.

B. Density Scaling

The previous section analyzed how much coverage length a probe can achieve for a fixed recording site size. In this section, we invert the analysis: we instead hold the coverage length constant and analyze how small (and dense) the recording sites can be when we scale the wire geometry. Reducing the recording site area allows for a tighter packing of recording sites along the probe's length, and thus a higher degree of spatial oversampling. The simulation results in Fig. 9 show the smallest possible recording site area that permits a 1 cm long probe shank fully covered by recording sites along its length. We chose the 1 cm length for its relevance in covering the depth of the entire mouse brain.

Scaling down the wire width results in a quadratic increase of resistance (Fig. 9, right) due to a roughly constant wire aspect ratio (Fig. 6). The recording site area, on the other hand, is linearly proportional to the wire width: as the recording site area is reduced, wire width must be reduced not only to accommodate a larger number of recording sites that fit into the same 1 cm of shank length, but the wires must also fit within a now decreased column width. These geometric considerations drive the general scaling behavior in Fig. 9.

For our choice of a 1 cm probe length, the probe noise only becomes significant for very narrow wires, e.g. at 50 nm and below (Fig. 9, left). The quadratic increase in wire resistance with scaling means that wire resistance eventually dominates the noise contribution as we reduce the wire width. Fig. 9 shows that we reach our ≤15 µV$_{rms}$ requirement at around 15 nm wire size. At that limit, each 1 cm long column is 5 µm wide and contains about 2,000 recording sites.

Based on the analysis of Fig. 9, we can also estimate the maximum number of recording sites that could penetrate a 1

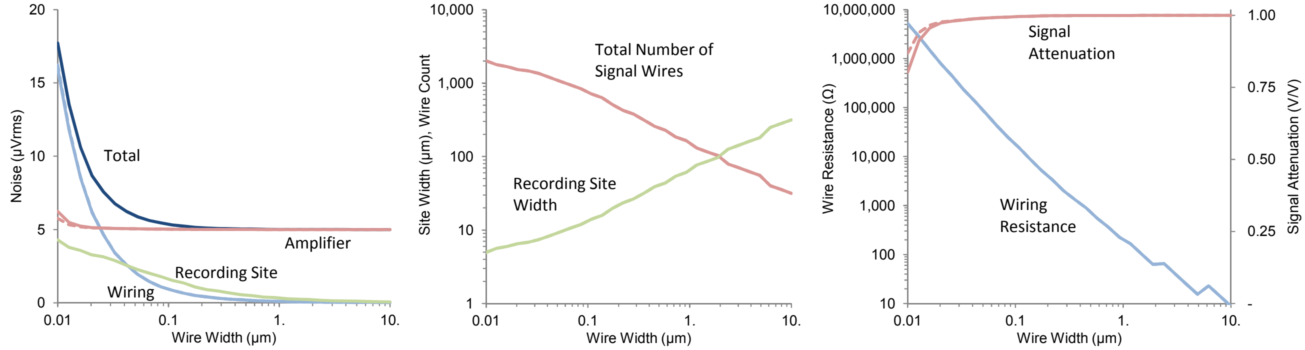

Fig. 9: Simulated impact of CMOS wire scaling, when maintaining constant recording site coverage length of 1 cm (along the length of the shank). The choice between lumped (solid) and distributed (dashed) circuit model shows no practical impact on the results. Smaller wire dimensions allow a larger number of wires (*center*), and consequently one can pack a larger number of smaller recording sites into the 1 cm long shank. We optimize the recording site size to maintain a maximum wire length of 1 cm (i.e. operating the wiring at the cross-over point "A" in Fig. 8). The 1 cm wire length is too short for the parasitic capacitances to create significant attenuation. The wiring and recording sites only contribute a relevant amount of noise when wire widths are scaled below 50 nm. Additional routing schemes can be devised that reduce the maximum wire resistance (e.g. widening metal lines as they reach the tip of a shank, as fewer wires are needed towards the tip, can achieve a 50% reduction in wire resistance). But the ability to implement such schemes may depend on a specific technology's design rules.

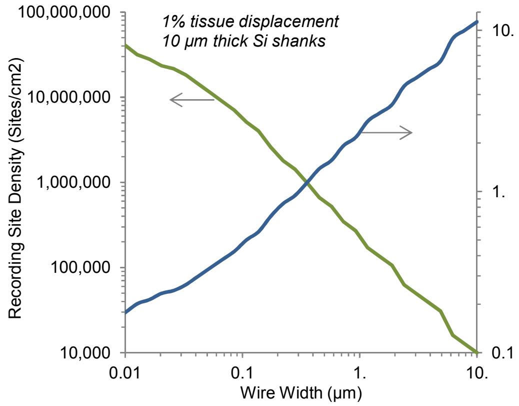

Fig. 10: Estimate of the number of recording sites that can be inserted through a 1 cm^2 surface, based on the probe shank designs of Fig. 9, and assuming a 10 µm thick silicon shank that displaces 1% of tissue. Even in the presence of a thinned 10 µm Si substrate (to provide the structural support for the wiring), the total cross-section per wire is minimal. In principle, semiconductor backend technology can enable neural recordings from millions of sites at 1% tissue displacement (approaching the number of neurons in the mouse brain at 100 million).

cm^2 surface area of the brain (e.g. the entire mouse brain). The result, shown in Fig. 10, shows a roughly linear scaling of the maximum number of sites as we reduce the wire width. But further increases can be possible, depending e.g. on the probe shank's substrate thickness.

V. Conclusion

We derived an equivalent circuit model for neural probe scaling, and carried out simulations that show how a deeply scaled CMOS technology may be able to provide an extremely dense recording infrastructure (Fig. 10) that can access on the order of as many recording sites in the mouse brain as neurons (e.g. 40 vs. 100 million, respectively). Deeply scaled future CMOS technologies can provide the extremely small and dense wiring infrastructure necessary for future large-scale neural recording architectures.

Acknowledgment

This work was supported by the Simons Center for the Social Brain, NIH, NSF, and NYSCF. Test structures were fabricated at MIT's Microsystems Technology Laboratories.

References

[1] H. W. Steenland and B. L. McNaughton, "Silicon Probe Techniques for Large-Scale Multiunit Recording," in *Analysis and Modeling of Coordinated Multi-neuronal Activity*, 2015, pp. 41–61.

[2] B. L. McNaughton et al., "The stereotrode: A new technique for simultaneous isolation of several single units in the central nervous system for multiple unit records," *J. Neurosci. Methods*, vol. 8, pp. 391–397, 1983.

[3] C. M. Gray et al., "Tetrodes markedly improve the reliability and yield of multiple single-unit isolation from multi-unit recordings in cat striate cortex," *J. Neurosci. Methods*, vol. 63, pp. 43–54, 1995.

[4] T. J. Blanche et al., "Polytrodes: high-density silicon electrode arrays for large-scale multiunit recording.," *J. Neurophysiol.*, vol. 93, no. November 2004, pp. 2987–3000, 2005.

[5] J. Scholvin et al., "Close-Packed Silicon Microelectrodes for Scalable Spatially Oversampled Neural Recording," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 1, pp. 120–130, 2016.

[6] P. Ruther and O. Paul, "New approaches for CMOS-based devices for large-scale neural recording," *Curr.Opin.Neurobiol.*,vol. 32, pp. 31–37,2015.

[7] J. Scholvin et al., "Heterogeneous Neural Amplifier Integration for Scalable Extracellular Microelectrodes," in *IEEE Eng. in Med. and Bio. Soc. 38th Ann. Conference (EMBC)*, pp. 2789-2793. August, 2016.

[8] K. Najafi and K. Wise, "Implantable multielectrode array with on-chip signal processing," *1986 IEEE Int. Solid-State Circuits Conf. Dig. Tech. Pap.*, vol. XXIX, no. December, 1986.

[9] C. M. Lopez et al., "An implantable 455-active-electrode 52-channel CMOS neural probe," *IEEE J. Solid-State Cir.*,vol.49,no.1,pp.248–261,2014.

[10] K. Najafi et al., "Scaling limitations of silicon multichannel recording probes," *IEEE Trans. Biomed. Eng.*, vol. 37, no. 1, pp. 1–11, 1990.

[11] X. Cui and D. C. Martin, "Electrochemical deposition and characterization of poly (3 , 4-ethylenedioxythiophene) on neural microelectrode arrays," *Sensors Actuators B Chem.*, vol. 89,pp.92–102,2003.

[12] T. Kuan and E. Al., "Fabrication and Performance Limits of Sub-0.1 µm Cu Interconnects," *Mat.Res.Soc.Symp.Proc*, 2000, vol.612,no.1,pp.1–8.

[13] T. Jochum et al., "Integrated circuit amplifiers for multi-electrode intracortical recording.," *J. Neural Eng.*, vol. 6, p. 12001, 2009.

[14] R. R. Harrison and C. Charles, "A low-power low-noise CMOS amplifier for neural recording applications," *IEEE J. Solid-State Circuits*, vol. 38, pp. 958–965, 2003.

[15] D. Tsai et al., "High - channel - count , high - density microelectrode array for closed - loop investigation of neuronal networks," pp. 1–4.

[16] A. Hierlemann et al., "Highly integrated CMOS microsystems to interface with neurons at subcellular resolution," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, vol. 2016–Feb, p. 13.2.1-13.2.4, 2016.

[17] T. Ragheb and L. A. Geddes, "The Polarization Impedance of Common Electrode Metals Operated at Low Current Density," *Ann. of Biomed. Eng.*, vol. 19, pp. 151–163, 1991.

[18] R. C. Gesteland et al., "Comments on Microelectrodes," *Proc.Ire*,1959.

[19] U. Rammelt and G. Reinhard, "On the Applicability of a Constant Phase Element (CPE) to the Estimation of Roughness of Solid Metal Electrodes," *Electrochem. Acta*, vol. 35, no. 6, pp. 1045–1049, 1990.