

BRAIN- SPINE- MUSCLE INTERFACES

BY: JACKSON POWELL
FOR: **FUTURE** JACKSON

root@/last_updated: **September 9, 2023**

Once you know the way broadly, you can see it in all things.
– The Book of Five Rings by Miyamoto Musashi

Contents

0.1	Overview	1
0.1.1	Purpose	1
0.2	To Do:	1
I	Electronics	2
1	Background Information	3
1.0.1	Overview	3
1.1	How Electricity Works	3
1.1.1	The misconception	3
1.2	Laws and Devices	4
1.2.1	Thevenin and Ideality	5
1.2.2	LED Circuits and PNP	7
1.3	Resistor Lattice Digression	8
2	Capacitors, Inductors, Filters	14
2.1	Overview	14
2.2	Capacitor Behavior	15
2.3	Filtration Conceptually	16
2.3.1	Integration and Differentiation	16
2.4	Frequency Dependence	18
2.4.1	Imaginary Numbers Digression.	19
2.5	Impedance	20
2.6	Filtration Quantitatively	21
2.6.1	Corner Frequency and Phase Shift	22
2.6.2	Summarizing Thoughts	22
2.7	Inductors	23
2.8	Peak Detection	25
2.9	Input and Output Resistance	25
3	Opamps	27
3.1	Introduction and Gold	27
3.1.1	Golden Rules	28
3.2	Simple Opamp Circuits	28
3.3	Opamp Imperfections	33
4	Transistors	35
4.0.1	Transistor Overview	35

4.1	Transistor Circuits	35
4.1.1	A Follower	35
4.1.2	Amplifiers	38
4.1.3	Current Sources	39
5	Writing Hardware	41
5.1	Introduction	41
5.1.1	Creating Modules	41
5.1.2	Bus Notation	42
5.2	Constants and Variables	43
II	Math and Models	44
5.3	Perspective	45
6	Math Essentials	46
6.1	Linear Algebra	46
6.1.1	Cross Products, Dot Products, and Gradients	46
6.2	Markov-chains	46
7	Neuron Modeling	48
7.1	Hodgkin-Huxley	48
7.1.1	The Main Form	48
7.1.2	Gating and Conductance	48
7.2	Fitzhugh-Nagumo Reduction	49
7.2.1	Why would we simplify this system?	49
7.2.2	How to Reduce	50
7.3	Diffusion	52
7.3.1	Forward Euler's	52
7.3.2	Backward Euler's	53
8	Modeling Circuits	55
9	Cell Motility	57
9.1	Introduction	57
9.2	Cell Motility	57
9.3	Fluid Mechanics	57
9.3.1	Conservation of Mass	60
9.3.2	Stress Tensor	61
9.3.3	Conservation of Momentum	62
9.3.4	Navier-Stokes Equations	63
9.3.5	Propulsion	65
III	Physiology	67
9.4	Perspective	68
10	Biochemistry	69
10.1	Membranes, Ions, and Potentials	69

10.1.1	Action Potentials	71
10.2	Ion Channels	72
10.2.1	Voltage Gated Channels	72
11	Muscles	74
11.1	Skeletal Muscle	74
11.1.1	A Cellular Level	74
11.1.2	Disorders Digression	75
11.2	Smooth Muscle	76
11.2.1	A Cellular Level	76
11.3	Cardiac Muscle	77
11.3.1	A Cellular Level	77
12	Sensory Processing	79
12.1	Various Senses	79
12.1.1	Touch	79
12.1.2	Temperature	79
12.1.3	Pain	79
13	The Central Nervous System	80
13.1	Spinal Cord General Structure	80
13.2	Autonomic Nervous System	80
13.2.1	Sympathetic Nervous System	81
13.2.2	Parasympathetic Nervous System	82
13.3	Somatic Nervous System	82
13.4	Acetylcholine, Nicotinic Receptors, and Muscarinic Receptors	82
13.5	Cerebrospinal Fluid	83
13.5.1	CSF Contacting Cells	83
13.6	Myelin in Health and Regeneration	85
13.6.1	Initiaion of Demyelination	85
13.7	Central Pattern Generation	86
IV	Spinal Cord Injury	87
14	The Injury Itself	88
14.1	Cell Specific Responses	88
14.1.1	Immune Response	88
14.1.2	Neural Response	89
15	In The Clinic & Therapeutic Approaches	91
15.1	Clinical Presentation	91
15.2	Auxiliary Management	92
15.3	Treatments	93
15.3.1	Electrical Stimulation	93
15.3.2	Biomaterials	94
15.3.3	Drug Treatment	94
15.3.4	Surgery	94

15.3.5	Rehabilitation	95
15.4	Complications After SCI	95
16	Approaches to Researching SCI	97
16.1	SCI Injury Models	97
16.2	The Basso Scale	98
16.3	Traumatic and Nontraumatic SCI	99
17	Broad Coverage of Axonal Regeneration	101
17.1	Intrinsic Factors	101
17.2	Extrinsic Factors	103
18	Metabolism of Neurons	105
18.1	Programming Mitochondrial Maintenance	105
18.1.1	Mitochondrial Generation and Degredation	105
18.1.2	Mitochondrial Trafficking	106
18.1.3	Response to Stress	106
18.1.4	Neuronal Injury and Regeneration	107
19	Circuit Reorganization and Plasticity	108
19.1	Reorganization After Injury	108
19.1.1	Interneurons and KCC2	110
19.2	Non-coding RNA in Rewiring and Plasticity	111
20	Stem Cell and Tissue Engineering	113
20.1	Cell Transplantation Therapy	113
20.1.1	Long Distance Growth of Stem Cells	114
20.1.2	3D Printed Spinal Scaffolds	114
V	Brain-Spine-Muscle Interfaces	116
20.2	Perspective	117
21	Interfacing with Nerves	118
21.1	Historical Dealings	118
21.2	Electrostimulation	119
21.2.1	Background	119
21.2.2	Biomaterials Used	119
22	Peripheral Nerve Stimulation	121
22.0.1	Adaptive, Conductive, and Electrotherapeutic Scaffolds	121
22.0.2	Vagus Nerve Modulates Circuits via Acetylcholine	122
22.0.3	Hand Movement Recovery via Intrafascicular Stimulation	123
23	Spinal Cord Stimulation	124
23.0.1	Application to Parkinson's Disease	124
23.0.2	Electronic Dura Mater	125
23.0.3	Electronics with Shape Actuation	126
23.0.4	SCI Specific Cells	127

23.0.5 Clinical Applications	127
24 Brain Stimulation	129
24.0.1 Living Electrodes	129
25 Reading Thoughts	131
25.0.1 Different Reading Devices	131
25.0.2 Reading LFPs Without Surgery	132
25.0.3 Neurograin	133
25.0.4 WIMAGINE ECoG	133
26 Synthesis	134
26.1 Lorach et al. 2023	134
26.1.1 Overview	134
26.1.2 Shortcomings	134
VI Project Proposal	136
27 Ramblings	137
27.1 Idealized World	137
27.2 Device Idea	138
27.2.1 Motivation	138
27.2.2 Framework	138
VII Addendum	141

0.1 Overview

0.1.1 Purpose

The purpose of this book is to aggregate all of the content and tools I will need in forging the future I desire: building better brain-spine-muscle interfaces. Interestingly, I already find myself referencing the information here quite a bit! It has already paid back itself in full.

0.2 To Do:

1. More neuron regeneration models (SCN, etc.).
2. Wrap-up electronics intro (Thev., Nort., etc.).
3. Diodes section :(
4. Markov Models :(
5. Sensory processing :(
6. Better breakdown of cellular localization in the spinal cord.
7. Clinical outcomes section.
8. Giovanni age-dependent regeneration article.
9. Sections for: Minassian, Rossignol, and the other titans.

Part I

Electronics

Like many of the vagabonds who live in the fields, stray horses seemed to him to be good-natured things. When you're through with them, they ask for nothing; they just go off quietly somewhere by themselves.

– Musashi by Eiji Yoshikawa

Chapter 1

Background Information

1.0.1 Overview

The bulk of this information comes from the Lab Electronics course at the University of Pennsylvania, taught to me by Professor Ashmanskas^{1,2}. As a comment on notation: V represents a variable value, while V represents a unit (or at least, that's what it should be; sometimes I forget proper notation). Graphics like this were made using CircuitTikz, whose manual can be found [here](#).

1.1 How Electricity Works

1.1.1 The misconception

One often represents electricity with the canonical metaphor of water flowing in a tube, equating the flow of water, which powers a wheel of some sort, as being the equivalent of electrons pushing through a wire. Incidentally, this simplifying schema illustrates a key misconception in the nature of electricity. This is easiest exemplified in the mode with which electricity reaches one's house from a power-plant, which before arriving will be subject to breaks in the circuit (transformers). In the traditional viewing of electricity, that is taught in early education, this is disconcerting as if electrons can not physically go from a power plant to the lights in your home, then how can their kinetic energy be transferred to them and turn them on?

Regarding the flow of *energy*: when a battery sits without wires attached, around it is an electric field. This field does not dissipate because no electrons flow from it. When wires are attached, charge accumulates on the surface of the wires. This causes a small electric field within the wires, but the drift velocity of electrons within the wire is quite slow—nowhere near the speed of light that you might expect electricity to flow. However, the flow of electrons within the wire is sufficient to drive an electric field which exists all around the wires. From this, we can determine the direction that energy will flow by taking the cross product of the electric and magnetic fields. In fact, if an lightbulb is attached, this means that energy flows from the battery to the bulb in all directions, not through the wires itself. This energy flow induces the vibration of electrons within the bulb's filament, thereby causing light. This means that a net flow of electrons is not required to power a bulb—but rather only their vibrations. Thus is illustrated next:

¹https://www.hep.upenn.edu/Courses/Phys364_spring23/

²<http://www.hep.upenn.edu/~ashmansk/>

Let us consider now what will happen with an AC circuit (120V AC outlets around your home), where the electromotive force flips with each cycle. In this case, both the electric and magnetic fields switch directions, meaning that their cross product will remain the same and, again, energy flows in all directions to power the lightbulb. Notably, the electrons do not move much (if at all) in this setup—but this is not a surprise, as it is not the electrons that carry the energy anyway. Now, it is still essential to recognize that it is the movement of electrons within the filament of a lightbulb that creates light. This is, indeed, from kinetic energy transferred from electrons bouncing against the metals lattice, dissipating energy in the form of light. The necessary distinction is that it is not electrons that flow all the way from the battery, but rather it is vibrations of those that were, and always will be, within the bulb itself. When you consider it like this, it is straightforward—as the electric field derived from the battery is what provides the electrons with enough kinetic energy to power the bulb.

Interestingly, comparisons to the “water flow” model fail dramatically in the traditional sense, but the Venturi Effect used to describe fluid flow actually succeeds. In adding a bulb you add a resistor, which is comparable to adding a part of a pipe with a smaller diameter. As water will flow faster in this section of the tube, so too will electrons. In order to maintain the same current as is through the rest of the tube, the drift velocity, V_d , must be higher. V_d is proportional to the electromotive force, E , meaning the force is highest within the bulb. Things like V_d are simplified into Ohm’s Law ($V = IR$) and not often discussed.

Of course you may say: “well, then why do we use wires at all?” The answer is that wires are helpful in channeling the fields, thereby making them more efficient. But, we do not *need* wires, per se. Think of wireless charging, for example. Knowing all of this, in this work I will almost invariably describe the flow of current as electrons moving through a wire. This is because it is much easier to think of electricity in this way, hence the ubiquitous misconception.

Nuances in the Fields.

It is worth explicitly highlighting that the electric field that causes the actual flow is from charges along the wires, rather than the battery. This is notable because if this were not the case, the proximity of the bulb to the battery would dictate its brightness. One may wonder how this type of charge distribution can be established so rapidly, and the reason is that the distance an electron needs to travel in order to create such a distribution is subatomic in size—meaning that with movement at the speed of light, the time it takes to establish a surface charge is effectively zero.

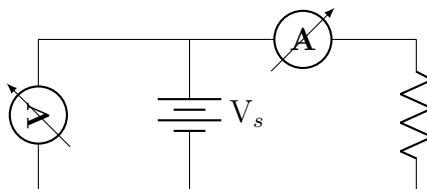
The whole idea is quite unintuitive, so it is appropriate to keep it smushed in the back of your mind, and to only draw it out when encountering things that are otherwise strange—like the aforementioned wireless charging, which should now be much more comprehensible.

1.2 Laws and Devices

The currents flowing in and out of a node will always be equal. The sum of the voltages over an entire circuit will always be equal. Kirchoff’s Voltage Law (KVL) can be used to show that wires connected in parallel will have the same voltage across them, and the current flowing in and out of a node will always be the same, defined in Kirchoff’s Current Law (KCL). In this way, we can predict the current flowing through a circuit to be $V = IR$. The formula for power is given as: $VI = P$, which means

that when one solves for voltage, and knowing that current is in units of charge/time and power in work/time, voltage is work per unit charge.

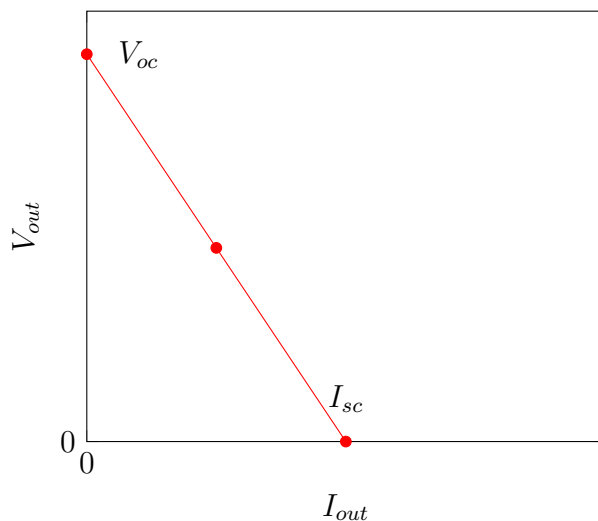
Voltage is measured using a voltmeter, a device in parallel with the load you are interested in measuring. An ammeter is used to measure current, which will be in series with the current you are trying to measure. This means that the voltmeter should have an extremely high resistance, so as to not draw any current, and an ammeter to have a low resistance, so as to not have any voltage drop. These are important considerations, as if the resistance of the load you are measuring is large (say, $1\text{M}\Omega$), it is possible that the voltmeter will have some non-negligible current flow through it. The same goes for if your circuit has very low resistance and you use an ammeter. To illustrate, one would measure the voltage and current coming from a battery as seen below:



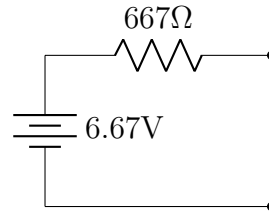
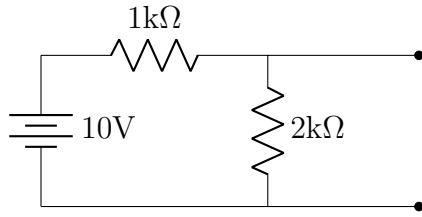
1.2.1 Thevenin and Ideality

A Thevenin Circuit simplifies the circuit to have a single resistance, R_{th} , and a single voltage, V_{th} . R_{th} can be calculated via replacing all of the voltage sources with a wire, and disconnecting all of the current sources. This “short circuits” your circuit and leaves you with only resistors, which can be used to calculate R_{th} using the familiar resistor rules. One can also short circuit the terminals, and determine the current flow, giving us $R_{th} = V_{th}/I_{sc}$.

R_{th} can be measured in a circuit by varying the R_{load} added to a circuit. In this case, you will see the voltage supplied (and corresponding current) change. The slope of this change ($\Delta V/\Delta I$) will equal R_{th} . If you vary the load through the two terminals and measure the voltage across it, you will get a graph that looks something like this:

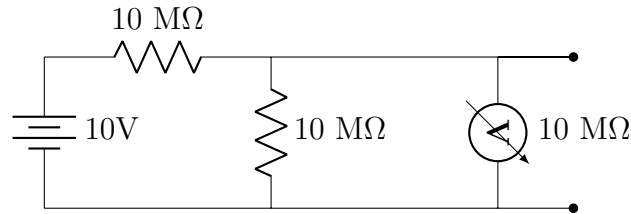


Again, the slope is what gives you R_{th} . I_{sc} is an important value which allows you to calculate V_{th} . The current that flows when you short circuit the load (I_{sc}), multiplied by R_{th} , gives you V_{th} .



The ideality of sources.

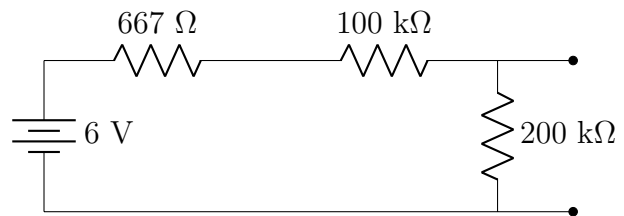
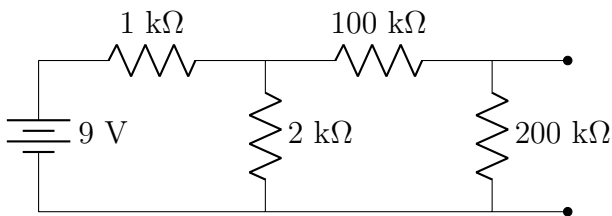
This coaxes us lightly into the topic of source ideality. Imagine all voltage or current sources as having a resistor in parallel with it, but inside of the component itself. An ideal battery, or voltage source, would be able to drive the same voltage, irrespective of the resistor/current. This would be like having a battery whose internal resistance is 0, causing the entirety of the voltage drop to occur on the circuit fragments outside of the battery. In the real world, batteries are not ideal. The canonical illustration of a batteries ideality is in trying to use a 9V battery to start your car. Naturally, the voltage dwindles as the current supplied increases. You can calculate the internal resistance of a battery by adding increasingly large loads to it, thereby giving you the batteries IV curve. It is worth considering this, as if your R_{load} is only $\approx 10 \times R_{th}$, then you may see drooping in the voltage supplied. Another way to state this is to make sure that the *input resistance* of your voltage source is much smaller than the *output resistance* of the upcoming circuit fragment you are attempting to drive. An example of when this fails is as follows:



Because the circuit has a non-negligible resistance relative to the voltmeter, you should expect to read something inconsistent with using an ideal voltmeter. As we are adding a 10 MΩ voltmeter in parallel to our 10 MΩ resistor, we expect that the “ R_{load} ” in this case will now be 5 MΩ, so the voltage divider at V_{out} will now be $1/3 \times 20V$, or 6.67V.

This contrasts to a current source, whose desired internal resistance is ∞ , as you will want no current to flow through it, and to flow entirely through the circuit fragments outside of the component. Once again, as the real world is not ideal, your goal in this case will be to have downstream circuit components whose input resistance is much smaller than the components output resistance.

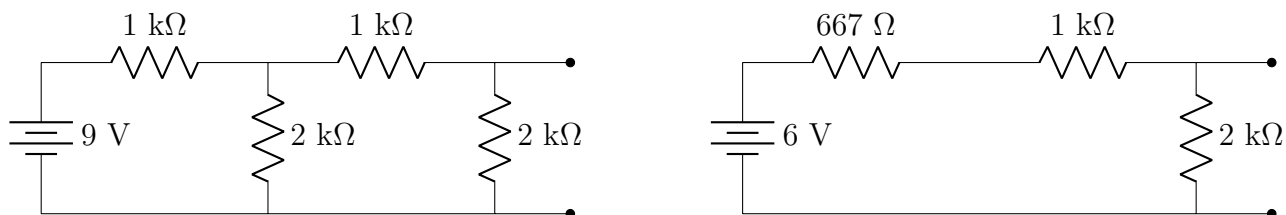
Let us consider another example of input resistance vs. output resistance:



If we draw a black box around the first voltage divider, we can convert it to the circuit schematic on the right because $1\text{ k}\Omega || 2\text{ k}\Omega = 667\text{ k}\Omega$ just as before. We can find the V_{th} by first finding I_{sc} , when

we short the black box's load. So, $I_{sc} = 9\text{V}/1\text{k}\Omega = 0.9\text{mA}$. We multiply I_{sc} (9 mA) by R_{th} (6.67 k Ω) to get a V_{th} of 6 V. If you were to compare the output resistance of the black box, 6.67 k Ω , to the input resistance of the upcoming voltage divider, 300 k Ω , you would find that it is much smaller. This naturally means that the entirety of the 6 V drop will occur over this part of the circuit.

Another way to think about this is as two successive voltage dividers, and qualitatively noting that the second's total resistance is much higher allows us to simplify things greatly. As a reminder, a voltage divider can be solved as $R_A I = R_A \times V_{total}/R_{total} = V_{total} \times R_A/(R_A + R_B)$. If you redraw as the Thevenin equivalent, the first voltage divider can effectively be ignored, because the voltage drop across this component will be minimal. Not by coincidence, since the first voltage divider is a 2/3 divider, and the second is a 2/3 divider, the voltage measured between our two terminals is $2/3 \times 2/3 \times 9\text{V}$, or alternatively, $2/3 \times 6\text{V}$ (V_{th}). This equivalency does not work when the input and output resistances of each fragment are comparable, here is an example:

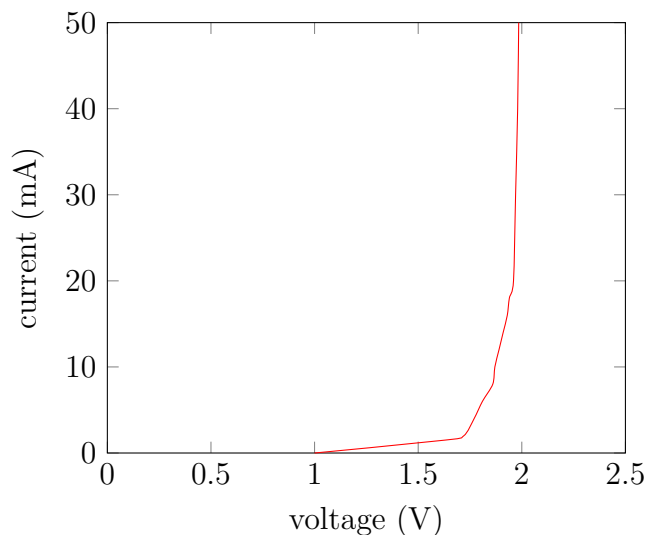


In this case, R_{th} will be the same, and R_{load} will be 3 k Ω . $V_{th} \times R_A/(R_A + R_B) = 3,000/3,667 \approx 5.45\text{V}$ for the output of the first divider (i.e., between R_{th} and the 1 k Ω resistor). And then naturally, if you were to measure the voltage between the 1 k Ω and 2 k Ω resistors, it would be $2/3 \times 5.45\text{V} \approx 3.63\text{V}$ for the voltage at the second divider.

This kind of Thevenin analysis only works when you have a linear IV curve. When might you have a non-linear IV curve?

1.2.2 LED Circuits and PNP

The IV curve across a light emitting diode (LED) should look something like this:

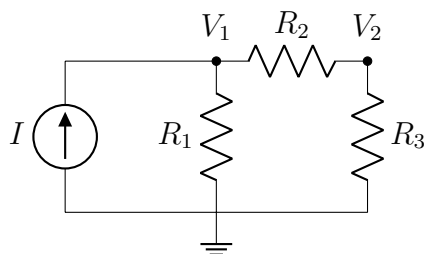


The IV curve for a diode, like an LED, is exponential in that the current slowly increases after the voltage across a diode hits some “threshold,” after which the current rises exponentially with voltage. Why is this the case? A diode is a P-N junction bridged by some depletion zone. The P side of the diode contains positively charged elements that act as “holes” (a silly way to say there is an absent electron position). The N side contains elements whose outer layers are loosely filled with electrons (i.e., low ionization energy). Effectively, the P side is devoid of electrons, while the N side has many free to give. What does this mean with regard to current and voltage? It means that the “depletion zone” between the two requires electrons to be able to bridge the gap. This really can’t happen unless they have a certain amount of energy, so increasing the voltage helps reach the “threshold” energy requires to pass the depletion zone (think of $P = IV$). Thus, as the electrons somewhat saturate the diode, you can theoretically pass an infinite current through it, as it will be effectively a short circuit.

1.3 Resistor Lattice Digression

One issue with modern BSIs is the usage of ECoGs, or EEGs, or other large measuring devices³. Too, they are almost universally hard electronics that require intense surgeries to implant. Therefore, if we could replace with soft electronics, we can cover considerable ground. For example, if one could drill a small hole into a patient’s skull and spread over the cortex a fabric that contained electrodes, one could achieve a similar amount of readings with a minimally invasive surgery. It is probable that there will be electrodes small enough to accomplish this, but let’s say there aren’t. Another way that this could be solved is using a lattice of resistors, with probes at either corner. These corners can exit the brain and be the points at which a computer interfaces with them.

Don’t be annoying, just go with the process.



Please, don’t be annoying—the solution is trivial but we will be talking about methods you can generalize. This example is from the SPICE method⁴. Considering KCL at the nodes gives us these two equations:

$$\begin{aligned} I &= \frac{V_1}{R_1} + \frac{V_1 - V_2}{R_2} \\ \frac{V_1 - V_2}{R_2} &= \frac{V_2}{R_3} \end{aligned} \tag{1.1}$$

As it goes, we can reformat this so as to easily turn it into a matrix in the following way:

³This is expanded on in the later parts.

⁴Thank you Prof. Ashmanskas, <http://www.ecircuitcenter.com/SpiceTopics/Overview/Overview.htm>

$$\begin{aligned}
I &= \frac{1}{R_1}V_1 + \frac{1}{R_2}(V_1 - V_2) \\
I &= \frac{1}{R_1}V_1 + \frac{1}{R_2}V_1 - \frac{1}{R_2}V_2 \\
I &= \left(\frac{1}{R_1} + \frac{1}{R_2}\right)V_1 - \frac{1}{R_2}V_2
\end{aligned} \tag{1.2}$$

$$\begin{aligned}
\frac{V_1 - V_2}{R_2} &= \frac{V_2}{R_3} \\
0 &= -\frac{1}{R_2}V_1 + \frac{1}{R_2}V_2 + \frac{1}{R_3}V_2 \\
0 &= -\frac{1}{R_2}V_1 + \left(\frac{1}{R_2} + \frac{1}{R_3}\right)V_2
\end{aligned} \tag{1.3}$$

These sorts of equations will usually be simplified using conductance as below:

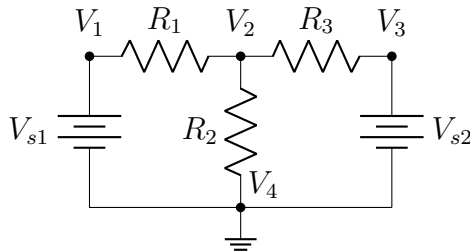
$$\begin{aligned}
(G_1 + G_2)V_1 - G_2V_2 &= I \\
-G_2V_1 + (G_2 + G_3)V_2 &= 0
\end{aligned} \tag{1.4}$$

$$\begin{bmatrix} G_1 + G_2 & -G_2 \\ -G_2 & G_2 + G_3 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} I \\ 0 \end{bmatrix} \tag{1.5}$$

So if $R_{1,2,3} = 100\Omega$, and $I = 1\text{mA}$, then:

$$x = A^{-1}B = \begin{bmatrix} 67\text{mV} \\ 33\text{mV} \end{bmatrix} \tag{1.6}$$

This method is called *nodal analysis*. You'll find more trouble trying to use a voltage source rather than a current source, and in fixing this, a method of *modified nodal analysis* was invented. Let us take the below example⁵:



In this system, we have 4 nodes, but node 4 largely functions as a reference for the other 3 nodes. A small note to make this a tad clearer is that in thinking about the current flowing into a node, such as into V_1 , you will view this as current going from V_2 to V_1 , but because everything in electronics is backwards, this is calculated as $(V_1 - V_2)/R_1$, and the opposite for current flowing into V_2 . Annoyance aside, using KCL, and then our definitions, we can gather these equations:

⁵<https://cheever.domains.swarthmore.edu/Ref/mna/MNA2.html>

$$\begin{aligned}
I_1 + \frac{V_1 - V_2}{R_1} &= 0 \\
\frac{V_2 - V_1}{R_1} + \frac{V_2}{R_2} + \frac{V_2 - V_3}{R_3} &= 0 \\
I_2 + \frac{V_3 - V_2}{R_3} &= 0 \\
V_1 &= V_{s1} \\
V_3 &= V_{s2}
\end{aligned} \tag{1.7}$$

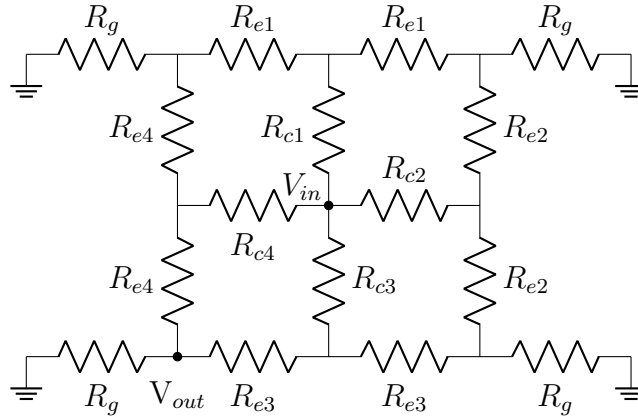
Which we convert to:

$$\begin{aligned}
I_1 + G_1 V_1 - G_1 V_2 &= 0 \\
G_1 V_1 + (-G_1 + G_2 + G_3) V_2 - G_3 V_3 &= 0 \\
I_3 - G_3 V_2 + G_3 V_3 &= 0 \\
V_1 &= V_{s1} \\
V_3 &= V_{s2}
\end{aligned} \tag{1.8}$$

And then:

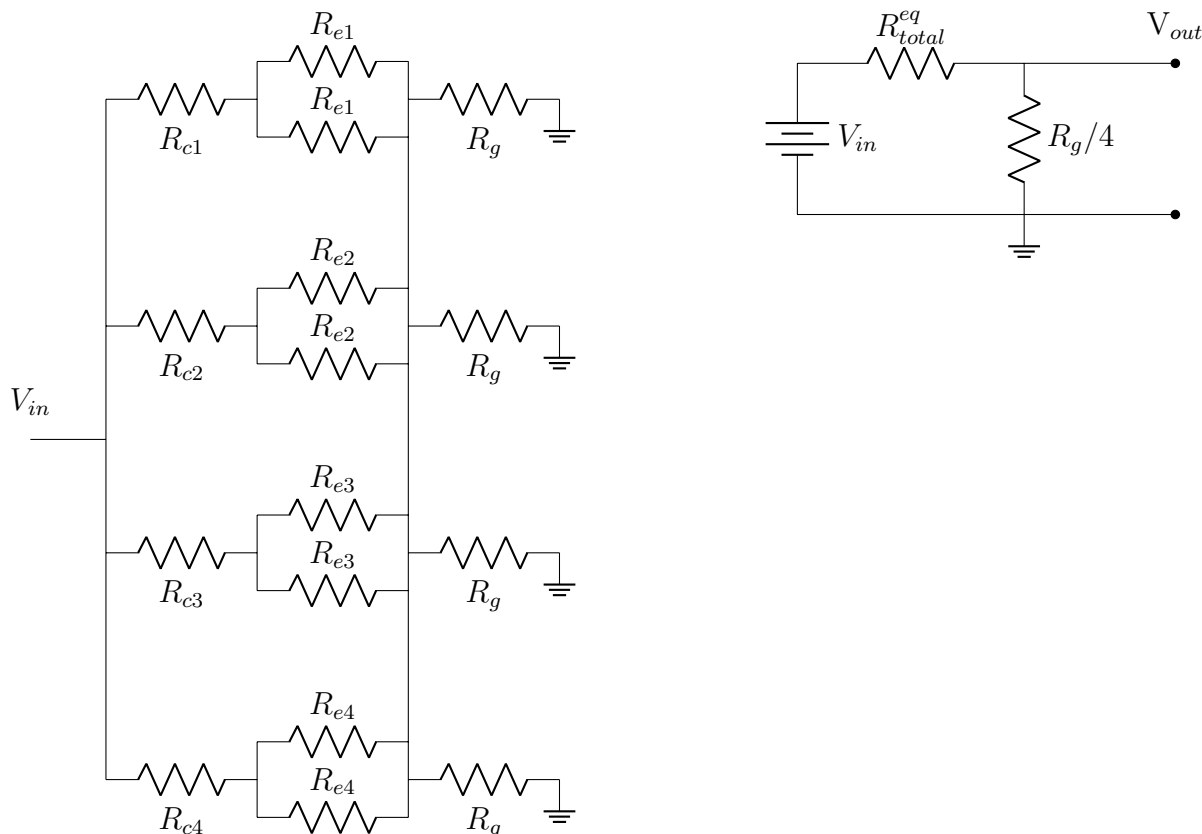
$$\begin{bmatrix} G_1 & -G_1 & 0 & 1 & 0 \\ -G_1 & G_1 + G_2 + G_3 & -G_3 & 0 & 0 \\ 0 & -G_3 & G_3 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ I_1 \\ I_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ V_{s1} \\ V_{s2} \end{bmatrix} \tag{1.9}$$

Now, I realize that one doesn't need to build a system of equations for most problems of small size, and they can be generally solved at a glance. Too, in many cases building a system is more hassle than it's worth. But, this will be important to keep in mind for the upcoming chapter **Modeling Circuits**. Let's look at some examples that are easier dealt with in the traditional way:

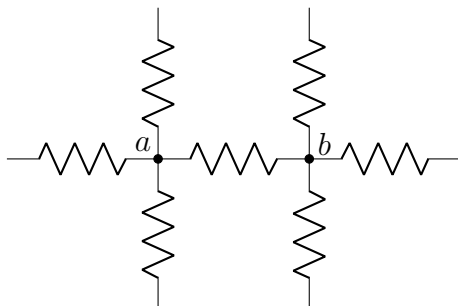


Firstly, apologies for the mess in labeling. The most difficult thing about this problem, in actuality, is how you look at it. Redrawing the circuit helps a great deal. I'll guide you through how one

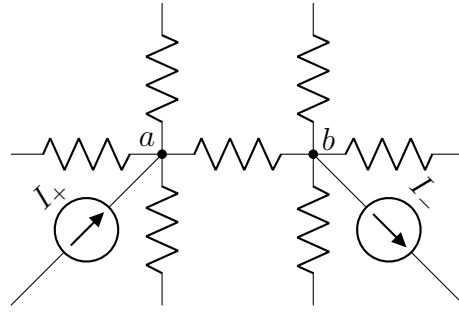
might do this. In truth, the four grounded resistors are in parallel and can be simplified to one wire ($R_g || R_g || R_g || R_g$, or $R_g/4$). You can probably realize that the edge resistors, for example the pair of R_{e1} , are also in parallel and are in series with R_{c1} . Therefore, we have $(R_{e1} || R_{e1}) + R_{c1}$. I'll call this R_1^{eq} . Then you'll realize simply that all of the equivalent R are also in parallel. This gives us $R_1^{eq} || R_2^{eq} || R_3^{eq} || R_4^{eq}$, which I will call R_{total}^{eq} . Now, our circuit simplifies to:



You may have had trouble at first glance because of all the “looped” sections of the circuit. It appears that current can flow any which way. Keep KVL in mind, and recall that current will only flow from high to low voltage. Therefore, you would never have current flowing from, say, R_{e2} to R_{e3} . A popular problem of the same flavor is the *infinite resistor grid*, which looks something like this:

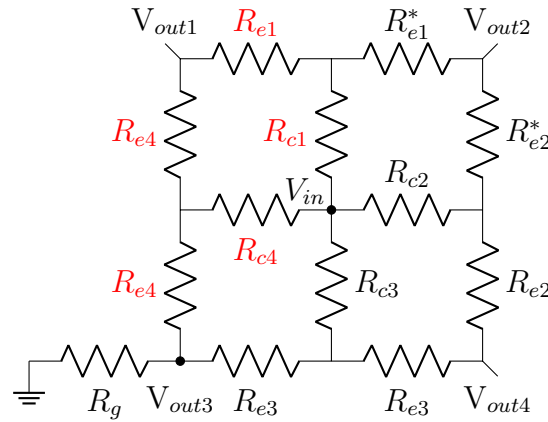


Imagine that the lattice of resistors continues out in every direction infinitely, in a grid-like fashion. The problem asks you to determine the equivalent resistance between points a and b . Considering the infinite nature, the more immediate method to solve would be to take some limit of parallel and series resistors. I actually think this is an important solution to keep in mind, as it can be generalized to other systems better. However, in this case the easiest solution is to use *superposition*. It goes as follows:

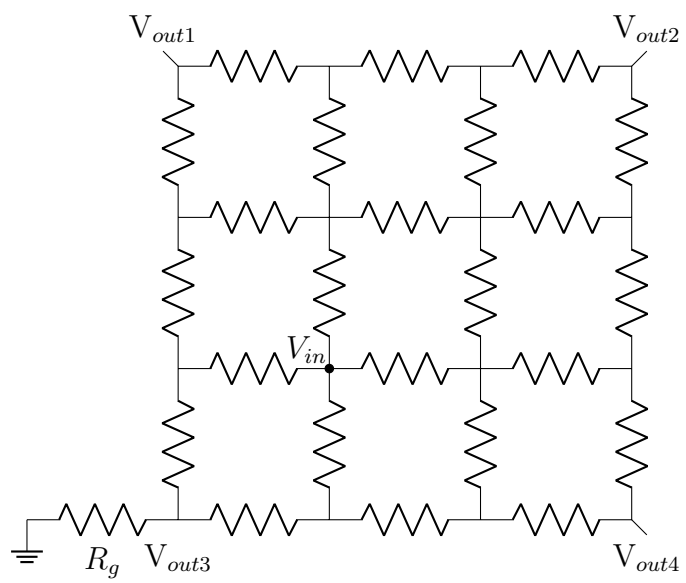


If you were to wire a positive current source to node a , you can immediately know that the current will be split equally in 4 ways, as the lattice is infinite. Then, if you apply a negative current source to node b , you can immediately know that it will draw $1/4I$ from each resistor as well. The total current flowing between nodes a and b will then be $1/2I$. Given that $V = IR$, and by superposition, we can know that $V_{ab} = (I/2)R$, so $R_{ab} = V_{ab}/I = (I/2)R/I = R/2$. The in-line math is kind of ugly, but I think you get the gist.

Another option of a similar flavor is:



In this case, we will think about what each V_{out} will read. One thing to note immediately is that the net current at V_{out2} will be 0. Therefore, R_{e1}^* and R_{e2}^* don't really matter, and can be neglected in our calculations. This once again is easiest to solve when re-visualized. The gist being that you have two quadrants (the R_{c1}, R_{c4} , which I colored to have red labels, and the R_{c2}, R_{c3} quadrants) where V_{in} is split into. The voltage drop across these quadrants being $3R||R$. We then add another R , for the one that connects to R_g , and voila, we are done. Let's call this R_q , so to find the equivalent resistance from V_{in} to V_{out3} it is $R_{q1}||R_{q2}$. I think you get the idea. The reason for going through these simple models is to prepare ourselves for some of the upcoming, more complicated ones. Let us level up now:



Chapter 2

Capacitors, Inductors, Filters

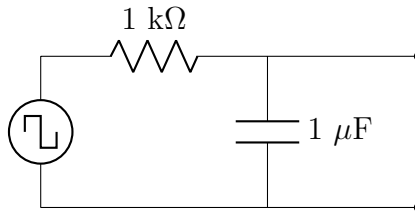
2.1 Overview

Getting a good intuition of capacitor behavior is an essential part of understanding electronics. A capacitor is a storer of charge, which allows it maintain a voltage for a period after the voltage source originally supplying the circuit is cut off. The basic structure of a capacitor is an anode and a cathode separated by a dielectric plate. Charge will be stored across it, and is released when the voltage supply dwindles. Once a capacitor is fully charged, there will be no current. However, if the voltage oscillates, the capacitor will follow V_{in} . Effectively, capacitors filter out DC voltage.

Capacitor discharge will be exponential decay in the form:

$$\begin{aligned}V_C &= V_S \times e^{-t/RC} \\ \tau &= RC \\ V_C &= V_S \times e^{-t/\tau}\end{aligned}\tag{2.1}$$

This is quite similar to neurons, isn't it! Let's consider some examples to further our intuition. Let's say you have the following circuit:

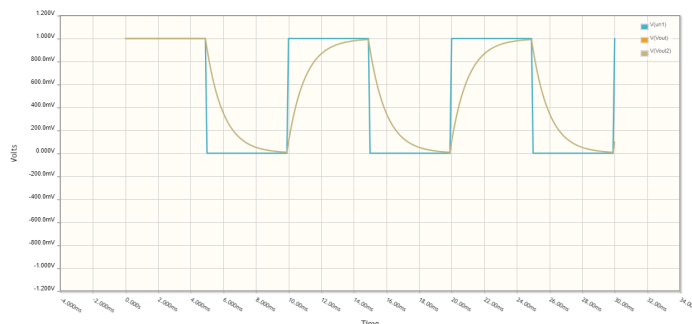
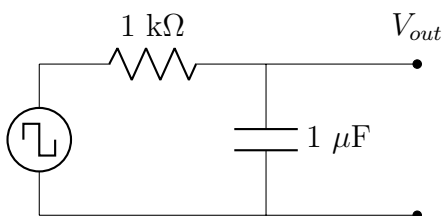


RC will be $1\text{k}\Omega \times 1\mu\text{F} = 1 \times 10^{-3}$, or $\tau = 1\text{ms}$. What does this actually mean, though? It means that if you charge this capacitor up to 10V , then $V_C = V_S \times e^{-t/RC} \rightarrow 10\text{V} \times e^{-t/1\text{ms}}$. So, 1 ms after voltage is removed, and or, one time constant, τ , after voltage is removed, the voltage at the capacitor will be $10 \times e^{-1} \approx 10 \times 0.367 = 3.67\text{V}$. If you were using a square wave which oscillated between 0 and 10V with a very high frequency (100 kHz , for example), then the period would be $1/100\text{kHz} = 0.01\text{ ms}$. Therefore, you would not expect the capacitor to ever fully discharge, and it would maintain a constant, high voltage. Perhaps you can intuitively predict that it would get stuck near a voltage of 5V , in the middle of our square wave input.

This circuit happens to be what is called a *low-pass filter*, which we will discuss in-depth later. But, you can see how it might get this moniker, as this very high frequency is not allowed to pass due to the capacitor's time constant. This was meant to serve as a basic intro to what a capacitor is. Now we can begin discussing the nuances.

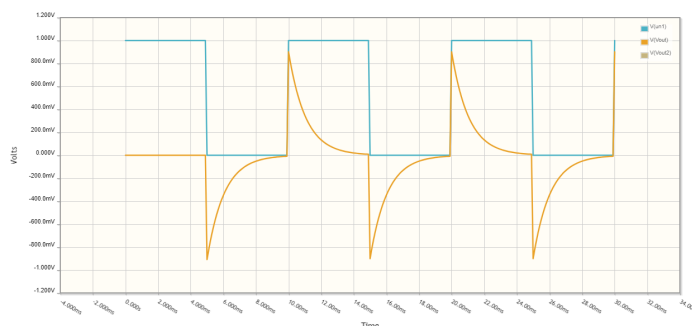
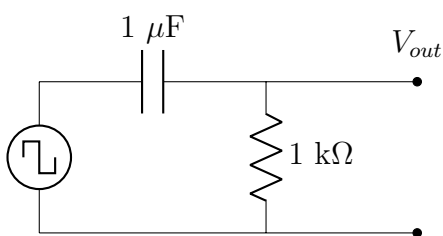
2.2 Capacitor Behavior

Let's look at the circuit below and its output (the same circuit as before)⁶. Only think about the shape of the output, no number. Only work on building up the conceptual understanding. This time our V_{in} (the blue square wave) is 100Hz.



Think about how the capacitor will behave. At time $t = 0$, the capacitor is fully charged. No current will flow at this time, meaning it will act as an open circuit, and therefore have a “resistance” of ∞ . You can think of the circuit as being like a voltage divider whose $R_2 = \infty$, and or $V_{out} = V_{in}$. At $t = (1/2)\omega$, the capacitor is able to discharge, and dissipates charge according to its decay curve. At $t = \omega$, the capacitor is subject to voltage again and begins to charge up. At first it acts like a wire, since it draws current proportional to the input voltage. As it becomes fully charged, it will again act like an open circuit.

Let's now consider swapping positions:



At $t = 0$, the capacitor is fully charged, so no current flows at all, and $V_{out} = 0$. This is expected, since capacitors are meant to filter out DC voltage. At $t = (1/2)\omega$, the capacitor shoots down to $-V_{pp}$. This is because V_{in} becomes negative relative to the voltage that the capacitor is held at. I.e., one plate of the capacitor is held at this V_{pp} value, so when V_{in} goes to 0, the voltage is negative between the plates of the capacitor, which draws negative current to the plate on the right side, making the voltage drop

⁶I really dislike needing to include .png images in works like this, but I believe I have no choice in this case.

across the $1k\Omega$ resistor equal to $-IR$. It then undergoes its standard decay curve, following this rule. Again, the decay will end at 0 because capacitors filter out DC voltage.

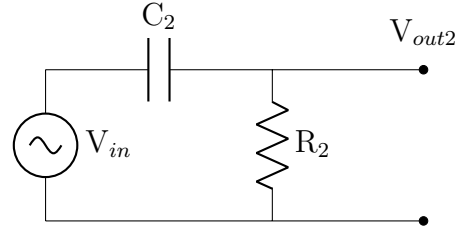
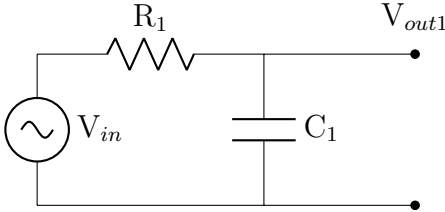
2.3 Filtration Conceptually

Firstly, what is filtration? Filtration refers to our filtering out of background signal, or any undesired signal, in order to isolate our signal of interest. For example, perhaps you are interested in a sine wave whose period is 1 ms, but this is superimposed on a bunch of other sine waves of variable periods. You can extract the 1ms sine wave with some clever circuits, and build your downstream circuit fragments based on it. It's like an analog version of the Fourier Transform!

Let's consider a specific example given by Professor Ashmanskas: imagine you are building a sensor that tests the water level in a pool. You don't want the pool to overflow when it rains, or get too low when the Summer comes and water evaporates—so you devise an automatic system to add or take out water as needed. One could simply add a sensor to the side of the pool, but it will be subject to the constant waves formed by people swimming, and its readings will be horribly variable and uninterpretable, as they do not reflect the so-called *steady-state* value of the water level. Each time someone cannon-balls in and splashes it, the sensor will think that the pool is greatly overflowing. Thus, one can devise a circuit that filters out all of these little fluctuations (this would be called a *low-pass filter*, for passing variations that occur on a long time-scale—low frequency).

2.3.1 Integration and Differentiation

Capacitors have this incredible ability to perform complex math, including taking derivatives or integrals of your wave form. Let's think about how this may occur using the circuit mentioned in the previous part. Firstly, recall that the current flowing through the resistor must equal the current flowing through the capacitor, and that $Q = CV$:



We can then solve for the voltage drop across R_1 as:

$$\begin{aligned} Q &= C_1 V_{out1} \\ \frac{d}{dt} Q &= \frac{d}{dt} C_1 V_{out1} \\ I &= C_1 \frac{dV_{out1}}{dt} \end{aligned} \tag{2.2}$$

$$\begin{aligned} IR_1 &= V_{in} - V_{out1} \\ \frac{dV_{out1}}{dt} &= \frac{1}{R_1 C_1} (V_{in} - V_{out1}) \end{aligned} \tag{2.3}$$

Therefore, if V_{out} is very small compared to V_{in} , you get:

$$\begin{aligned}\frac{dV_{out1}}{dt} &= \frac{1}{R_1 C_1} V_{in} \\ \int \frac{dV_{out1}}{dt} &= \int \frac{1}{R_1 C_1} V_{in} \\ V_{out1} &= \frac{1}{R_1 C_1} \int V_{in}\end{aligned}\tag{2.4}$$

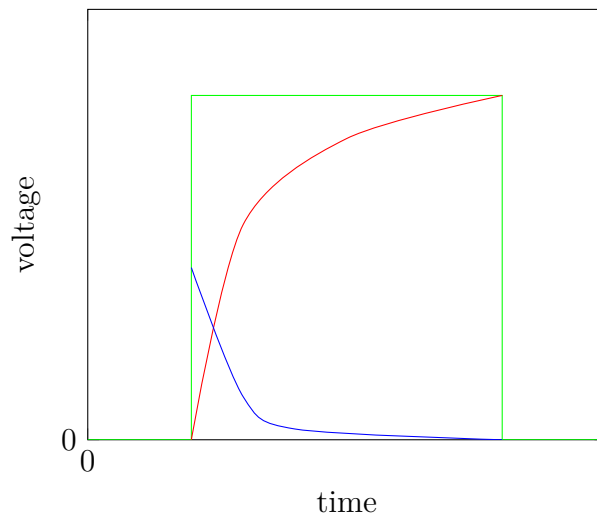
And or, that V_{out} integrates V_{in} . How will this change if we swap the positions of the resistor and capacitor in circuit 2? Once again, consider when V_{out} is much smaller than the input.

$$\begin{aligned}I &= C_2 \frac{d}{dt} (V_{in} - V_{out2}) \\ \frac{V_{out2}}{R_2} &= C_2 \frac{d}{dt} (V_{in} - V_{out2}) \\ \frac{V_{out2}}{R_2} &= C_2 \frac{dV_{in}}{dt} \\ V_{out2} &= R_2 C_2 \frac{dV_{in}}{dt}\end{aligned}\tag{2.5}$$

Thus, in this case V_{out} approximates the derivative of V_{in} .

Qualitative Thinking.

The best lead in to thinking about filtration, to me, is thinking simply about how the graphs look like when something integrates or differentiates a square wave, combined with our understanding of capacitor behavior from above. Let us not do any math, and think only qualitatively.

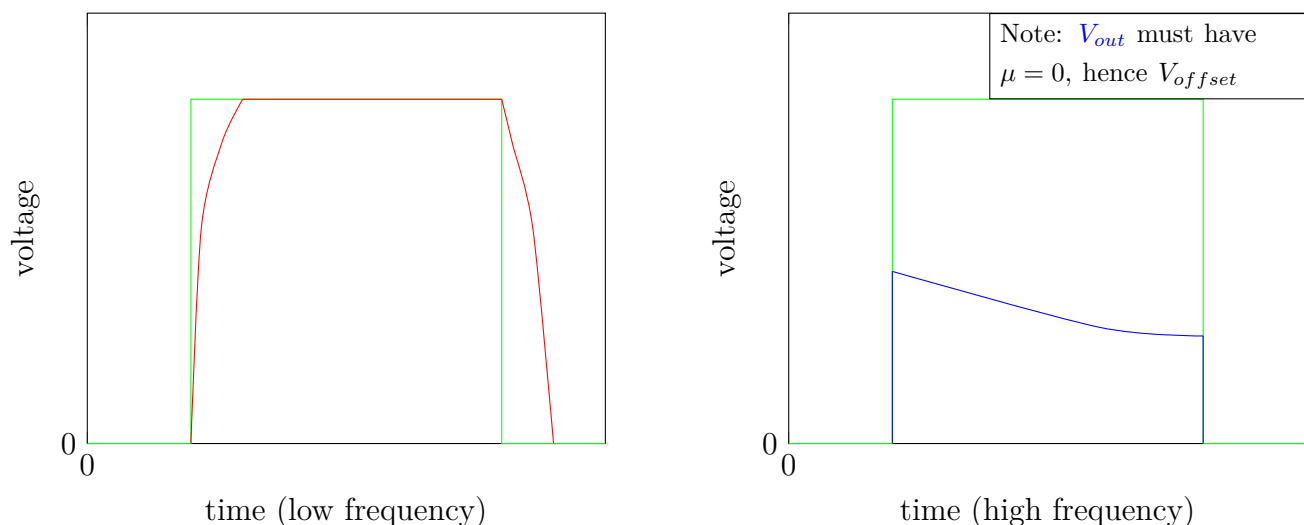


If the green waveform is our V_{in} from the previous example, then our red curve is similar to how V_{out1} may look, and the blue curve is similar to how V_{out2} . That is, the blue curve *kind of* differentiates

the green curve, because the relatively instant rise signifies a very positive derivative, marked by this blue spike. Similarly, the red curve *kind of* integrates the green curve, because the area under the green curve slowly accumulates, hence the overall positive slope of the red curve.

Something interesting you might notice, though, about the blue curve is that it seems centered around 0, while the red curve seems centered around the midway point of the green V_{in} . You can realize that the capacitor being before V_{out} means it'll filter out all DC voltage, thereby centering V_{out} about 0. Thus, you can expect there to be a symmetric, negative curve on the downward edge of the green curve, as mentioned and discussed in section 2.2.

So if we say that the frequency is extremely slow, and the red curve is the voltage being measured at V_{out2} , how might the red curve look?



Similarly, if the frequency is very high, how might the blue curve look? Take a second to ponder the two graphs above and gather a bit of intuition on it. This should allow you to qualitatively state that an integrating circuit will be a *low-pass filter* (because V_{out} matches V_{in}), and a differentiating circuit will be a *high-pass filter* (except that there's an offset when the long-term average of $V_{in} \neq 0$). As in the left graph, as the frequency gets smaller and smaller, the red plot will more and more closely match the green plot. This is the basis of capacitor filtration, an essential tool in electronics!

2.4 Frequency Dependence

This section will primarily be quantitative. The important bits, though, are the qualitative understandings, and the final results of this quantitative section. What you find in the interim is likely not worth understanding fully. Let's begin by supposing we apply a cos wave to a resistor. The relationship between current and voltage is as follows:

$$\begin{aligned}
 f &= \frac{\omega}{2\pi} \\
 V(t) &= V_{pp} \cos(\omega t) \\
 V(t) &= I(t)R \\
 I(t) &= (V_{pp}/R) \cos(\omega t)
 \end{aligned} \tag{2.6}$$

Therefore, if you were to solve something like V_{out}/V_{in} your pesky cos terms will cancel. This simply means that V_{in} and V_{out} will be in phase, only changed by some scaling factor. Unfortunately, this is not so for capacitor equations. We can see this below:

$$\begin{aligned} Q &= CV \\ I(t) &= C \frac{d(V_{pp} \cos(\omega t))}{dt} \\ I(t) &= -\omega CV_{pp} \sin(\omega t) \end{aligned} \tag{2.7}$$

This means that the current and the voltage will be 90° out of phase. Another way would be to write this as $A \cos(\omega t + \phi)$. Let us think about what will happen with our standard low-pass filter. Using KVL, we can state that:

$$\begin{aligned} V_{pp} \cos(\omega t) &= RC \frac{dv_C}{dt} + v_c \\ v_c &= A \cos(\omega t + \phi) \\ V_{pp} \cos(\omega t) &= RC \frac{d}{dt} A \cos(\omega t + \phi) + A \cos(\omega t + \phi) \end{aligned} \tag{2.8}$$

2.4.1 Imaginary Numbers Digression.

Maybe looked at the previous equation and thought it is solvable (probably?) but that it would be not worth your while to do so, and that there is likely a better way to go about it. Recall Euler's Relation⁷:

$$e^{j\omega t} = \cos(\omega t) + j \sin(\omega t) \tag{2.9}$$

One of the requirements of a linear circuit is superposition, and that is *kind of* the argument that allows us to use imaginary numbers. You can think that if you are using math that includes both a real and imaginary component, as long as you keep track of the real, your output will still be correct—but, don't think too hard about this. In the upcoming sections, I will use \mathbf{v}_c to denote the complex number. Let us examine:

$$\begin{aligned} V_{pp} e^{j\omega t} &= RC \frac{d\mathbf{v}_c}{dt} + \mathbf{v}_c \\ \mathbf{v}_c &= A e^{j\omega t} \\ V_{pp} e^{j\omega t} &= RC \frac{d}{dt} A e^{j\omega t} + A e^{j\omega t} \end{aligned} \tag{2.10}$$

We can differentiate, and then obtain an expression for A , and solve for the voltage at the capacitor as:

⁷Electrical Engineers use j instead of i for imaginary numbers. The claim is that it is easier to keep track of in the math. Other's claim it's because i is sometimes used to represent currents, particularly small ones.

$$\begin{aligned}
V_{pp}e^{j\omega t} &= j\omega RC A e^{j\omega t} + A e^{j\omega t} \\
V_{pp} &= j\omega RC A + A \\
V_{pp} &= A(1 + j\omega RC) \\
\frac{V_{pp}}{(1 + j\omega RC)} &= A \\
\mathbf{v}_c &= \frac{V_{pp}}{1 + j\omega RC} e^{j\omega t}
\end{aligned} \tag{2.11}$$

We now want to find the real component, which begins by rewriting the expression in its polar form:

$$\begin{aligned}
\mathbf{v}_c &= \left(\frac{1}{\sqrt{1 + \omega^2 R^2 C^2}} e^{j\phi} \right) V_{pp} e^{j\omega t} \\
\phi &= \tan^{-1}(-\omega RC) \\
\mathbf{v}_c &= \frac{1}{\sqrt{1 + \omega^2 R^2 C^2}} V_{pp} e^{j(\omega t + \phi)}
\end{aligned} \tag{2.12}$$

From here we can simply take the real part and be on our way:

$$v_c = \frac{V_{pp}}{\sqrt{1 + \omega^2 R^2 C^2}} \cos(\omega t + \phi) \tag{2.13}$$

This is equivalent to saying:

$$A_{out} = A_{in} \sin(\omega t + \phi) \tag{2.14}$$

Where A is the amplitude at either time. This is one way to go about this problem. Another way, which will be discussed in the next section, is using impedance.

2.5 Impedance

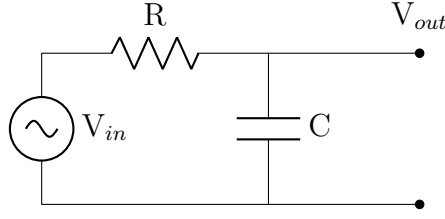
We mentioned earlier in equation (2.11) that the relationship between an input voltage and what is measured at a capacitor for a low-pass filter is:

$$V_{in} \frac{1}{1 + j\omega RC} = v_c \tag{2.15}$$

If we were to divide this fraction by $j\omega C$, and simplify using the representation Z_C , we would get:

$$\begin{aligned}
V_{in} \frac{1/j\omega C}{1/j\omega C + R} &= v_c \\
V_{in} \frac{Z_C}{Z_C + R} &= v_c
\end{aligned} \tag{2.16}$$

This looks just like a voltage divider equation! The concept of *impedance* is used to summarize other circuit fragments, like capacitors or inductors, using some resistance equivalent Z . The impedance of a resistor, Z_R , is simply R . Z_C is $1/j\omega C$. Let's think of how this pertains to our low-pass filter.



If we measure at V_{out} , we can think of it like a voltage divider, giving us:

$$V_{out} = \frac{Z_C}{Z_C + R} V_{in} \quad (2.17)$$

You may wonder if $Z_C = 1/j\omega C$ has any real basis, or if it is simply an extraction from the above math. In reality, you can find it quite simply through:

$$\begin{aligned} I_c &= C \frac{dv_c}{dt} \\ I_c e^{j\omega t} &= CA \frac{d}{dt} e^{j\omega t} \\ I_c e^{j\omega t} &= j\omega CA e^{j\omega t} \\ I_c &= j\omega CA \\ I_c \frac{1}{j\omega C} &= A \end{aligned} \quad (2.18)$$

This is the equivalent of Ohm's law, where $1/j\omega C$ is the resistance. Understanding this, we can find our high-pass filter's equation to be:

$$V_{out} = \frac{R}{Z_C + R} V_{in} \quad (2.19)$$

2.6 Filtration Quantitatively

We are finally prepared to talk about filtration quantitatively! As was mentioned in the impedance discussion, resistor-capacitor (RC) circuits can be formulated as voltage dividers. We know that the impedance of a capacitor is $1/j\omega C$, which has some frequency dependence from ω (or, $2\pi f$). Therefore, we can get the sense that V_{out} may change depending on this frequency. With a bit of re-writing, we find:

$$\begin{aligned} \frac{V_{out}}{V_{in}} &= \frac{Z_C}{Z_C + R} \\ \frac{V_{out}}{V_{in}} &= \frac{1}{1 + j\omega RC} \\ \frac{V_{out}}{V_{in}} &= \frac{1}{\sqrt{1 + (2\pi f RC)^2}} \end{aligned} \quad (2.20)$$

We can see that as the frequency goes up, this converges to $1/\infty$, or $1/1$ as frequency goes down. Whereas, for a high pass filter:

$$\begin{aligned}
\frac{V_{out}}{V_{in}} &= \frac{R}{Z_C + R} \\
\frac{V_{out}}{V_{in}} &= \frac{j\omega RC}{1 + j\omega RC} \\
\frac{V_{out}}{V_{in}} &= \frac{2\pi f RC}{\sqrt{1 + (2\pi f RC)^2}}
\end{aligned} \tag{2.21}$$

Which converges to ∞/∞ as frequency goes up, and $0/1$ as frequency goes down.

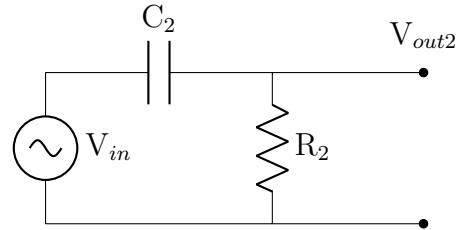
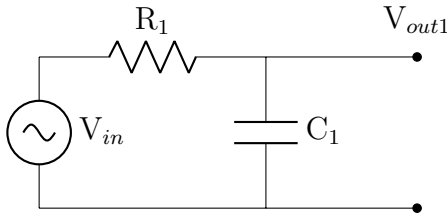
2.6.1 Corner Frequency and Phase Shift

The way one quantifies this is with the “corner frequency,” which you will more often hear as f_{3dB} . You’ll notice that for both the high-pass and low-pass filter equations described above, the ratio of V_{out} to V_{in} is $1/\sqrt{2}$ when the frequency inputted is $f = 1/2\pi RC$. $1/\sqrt{2}$ corresponds to about 0.7, meaning that at this f_{3dB} , the output is about 70% the amplitude of the input. Thus, it is a good marker of how well your filter will work. If you are trying to pick a low pass filter that filters out anything at 1000 kHz and higher, you’ll want to pick an RC circuit combination that is well below it.

Notably, filtration causes a phase shift. This shouldn’t be surprising when you recall that a low-pass and high-pass filter also integrate and differentiate respectively. Let us take, for example, a sin input. If it is passed through a low-pass filter, would expect it to be integrated to $-\cos$. If passed to a high-pass, it would be differentiated to \cos . Therefore, as $-\cos$ is -90° ($\sin(x - \frac{\pi}{2}) = -\cos(x)$) relative to \sin , and \cos is $+90^\circ$ ($\sin(x + \frac{\pi}{2}) = \cos(x)$), we would expect a low-pass filter to generate an output that lags V_{in} , while a high-pass will generate one which precedes V_{in} .

The filtration at f_{3dB} is 45° . I don’t think this has a straightfoward math explanation, but conceptually it is when the real and imaginary parts are equal. Maybe don’t think about this too hard.

2.6.2 Summarizing Thoughts



Let us look at these two circuits again, and see what we can tell just from a glance.

For the left circuit, since the current through the resistor and capacitor must be the same, we can say $Q = CV_{out}$, and thus $I = CV'_{out}$, that $V_{in} = IR = RCV'_{out}$, so V_{out} integrates V_{in} . Knowing that it integrates, we can intuit that it must be a low-pass filter, based on the graphs presented earlier. We can also know it is a low-pass filter if we recall that $Z_C = 1/j\omega C$, and that a voltage divider must look like $Z_C/R + Z_C$, which would give us $1/j\omega C + 1$. And, since it integrates, we know that $\int \sin = -\cos$, so there will be a -90° phase shift at very high frequencies.

For the right circuit, since $V_{in} - V_{out} = V_C$, we know $CV'_C = I$, so $V_{out} = RCV'_{in}$. Therefore, it differentiates, which means it must be a high-pass filter. Too, it must be of the form $R/Z_C + R$. And if it differentiates, then $\sin' = \cos$, so there must be a $+90^\circ$ phase shift at very low frequencies.

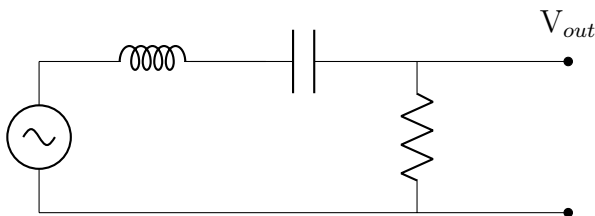
In other words, you can tell a lot just at a glance! This is without any math!

2.7 Inductors

An inductor stores energy in a magnetic field. An inductor is usually a tightly coiled bit of wire, and as you know a wire always has a magnetic field, so by coiling them and running current through it, you are summing these magnetic fields.

To conceptualize an inductor's behavior, as we will consider a water-wheel. An inductor is like a water-wheel in a pipe. As water flows through the pipe and reaches this large, heavy wheel, the speed of flow will be decreased as this wheel is pushed. Eventually, the wheel gets 'up to speed' so-to-speak, and its momentum carries it. The water's flow will not be impeded at this point. If you were to stop the flow of water, the momentum of the wheel would allow it to continue spinning, thereby continuing the water's movement for a time. If you were to have an LED in parallel with an inductor, you would expect that the LED would initially shine very brightly, as the inductor functions similarly to a large resistor its early stages. You'll see the LED dim as the inductor becomes fully charged, and now acts like a wire. If you remove the battery from the circuit, the LED will remain on for a time, as the inductor loses its charge. The rate at which the inductor dissipates its charge will depend on the total resistance of the circuit.

In fact, the magnetic field that forms as a reduce of current generates an electromotive force in the opposite direction, essentially opposing the current's flow. This field is held for some time constant, similar to a capacitor. An inductors time constant is $\tau = 2L/R$. The field is generated at at a rate of $V = L\frac{d}{dt}I$, compared to a capacitor's curve of $I = C\frac{d}{dt}V$. Therefore, you can imagine that an inductor may work opposite of a capacitor. Indeed, the impedance of an inductor is $j\omega L$, compared to a capacitors $1/j\omega C$. The f_{3dB} in this case is $R/2\pi L$. Inductors act like resistors, in that you can add them up in series like resistors, and treat them as you would with resistors in parallel. However, when we consider only impedance, everything functions like a resistor. That makes analyzing our below circuit much nicer.



This can be solved as $Z_2/Z_1 + Z_2$, where $Z_2 = R$, and $Z_1 = j\omega L + 1/j\omega C$. Let me try a bit of math here:

$$\begin{aligned}
Z_1 &= j\omega L + \frac{1}{j\omega C} \\
Z_1 &= \frac{-\omega^2 LC}{j\omega C} + \frac{1}{j\omega C} \\
Z_1 &= \frac{1 - \omega^2 LC}{j\omega C}
\end{aligned} \tag{2.22}$$

The impedance, Z_1 , goes to zero when $\omega^2 LC = 1$. In other words, at some frequency, f_0 , the impedance will be minimized. This is called the resonant frequency, and is found at $f_0 = 1/2\pi\sqrt{LC}$. You can think about what would happen when the inductor and capacitor are in parallel in the circuit above:

$$\begin{aligned}
Z_1 &= j\omega L \parallel \frac{1}{j\omega C} \\
Z_1 &= \frac{j\omega L / j\omega C}{j\omega L + 1/j\omega C} \\
Z_1 &= \frac{j\omega L}{1 - \omega^2 LC}
\end{aligned} \tag{2.23}$$

Now, impedance will go to infinity at the described f_0 —and you can filter out a specific frequency. It is useful to know how this filtering will affect the power dissipated by your circuit. At $\pm\Delta f = R/2\pi L$, the power halves, so one would call the bandwidth of this circuit to be $2\Delta f$. A halved power corresponds to an amplitude decrease of $1/\sqrt{2}$ (since $P = V^2/R$). Thus, at the two sides of the bandwidth, you'd expect to see an amplitude that is about 70% as large as at f_0 .

Either or?

So why are capacitors much more widely used than inductors? To start, inductors are usually large and heavy. As it must be a coil of wires, usually wrapped around some magnetically-permeable metal, it doesn't easily fit into circuit boards. Too, inductors will continuously dissipate energy, while capacitors that are charged do not.

2.8 Peak Detection

An interesting application of a low-pass filter is peak detection. There may be times where you are only interested in extremes, and not the little fluctuations in between. For example, perhaps you're interested in neural activity. You probably don't care too much about individual peaks, but instead you want to get a measure of the overall activity. This is one such way to do so⁸:

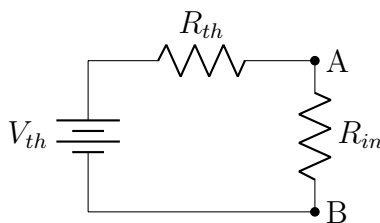


You can conceptually imagine that to create some form of peak detection, you want the RC decay value to be slow relative to the minute fluctuations, but fast relative to the longer term fluctuations you are interested in capturing.

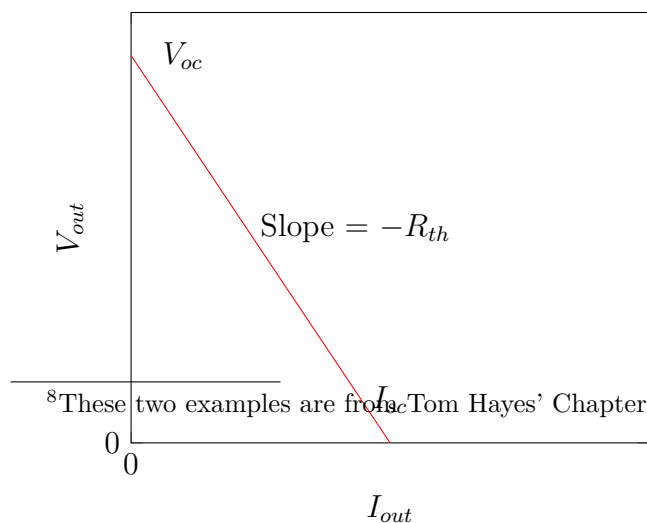
2.9 Input and Output Resistance

Now that we have a better understanding of impedance, let's discuss the idea of input and output resistance again. This concept is actually quite interesting and important, because when you are supplying a voltage to the body (say, in stimulating the spinal cord) you'll see impedance drifts, and phase shifts over time. Therefore, having a strong conceptual understanding of voltage sources & sinks, and input & output resistances is useful in quantifying this.

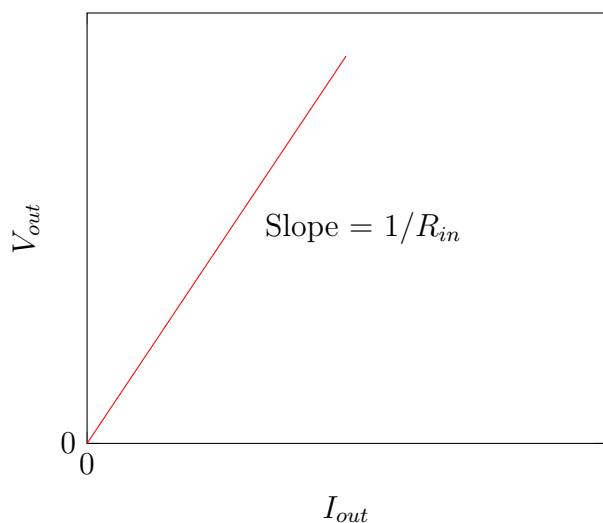
So for:



We get:



⁸These two examples are from Tom Hayes' Chapter 3N.



The left graph was presented in an earlier section. The important takeaway is that for a linear system, $R_{th} = -\Delta V_{AB}/\Delta I_A$. For a non-linear system, we can generalize this to include the impedance:

$$Z_{out} = -\frac{dV_{AB}}{dI_A} \quad (2.24)$$

Where V_{AB} is across the voltage source, and I_A is from the voltage source, and when we are varying the load resistance. Then, the input resistance of the load can be defined as:

$$\frac{1}{Z_{in}} = \frac{dI_A}{dV_{AB}} \quad (2.25)$$

For the I_A into the load and the V_{AB} across the load when we are varying the voltage source.

The next section should be diodes/LEDs, but I find those to be just so boring. So I'd rather skip to opamps.

Chapter 3

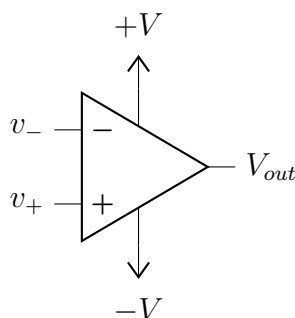
Opamps

Operational amplifiers (opamps) are one of the most ubiquitous tools in electronics, just behind classics like the resistor and capacitor. While they themselves are amazing, making CircuitTikz diagrams for them is a nightmare! So wish me luck in the coming parts!

As there are countless opamp applications, it can become muddled why you'd want a circuit to do such things. Thus, I'll do my best to think of interesting biological applications for the less obvious ones.

3.1 Introduction and Gold

Opamps are active components, in that they can take a circuit's power and increase it. They have the ability to amplify signal, act as a voltage clamp, do addition and subtraction, and even integration and differentiation. Let's discuss:



Inside of an opamp is a large network of transistors. It isn't worth trying to understand all the internals, but it is worth understanding all of the externals. An opamp has two inputs (v_- and v_+), and two power rails ($+V$ and $-V$). In general, an opamp's goal will be to minimize the difference between the two inputs, and it has some range of power to work with in doing so, which is limited by $+V$ and $-V$. It tries to minimize this input through its V_{out} . Therefore, one would often connect the v_- to V_{out} as a form of negative feedback. This will likely make more sense in the later sections. If there is no connection between V_{out} and either of the two inputs, your opamp will oscillate between the two power maximums set by $+V$ and $-V$. As in, if $v_- > v_+$, then $V_{out} = +V$. This is an opamp functioning as a *comparator*, in that it makes the somewhat binary comparison of which input is larger, and outputs either high or low voltage.

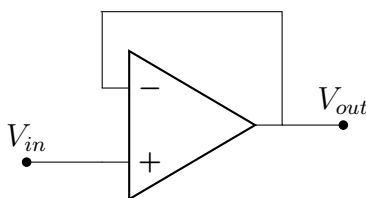
3.1.1 Golden Rules

In adding some form of negative feedback, and keeping your opamp from being saturating (i.e., clamping V_{out} to either of the two power rails)⁹, an opamp will adhere to two essential rules:

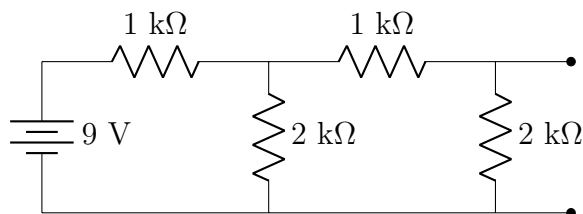
1. V_{out} does whatever needed to ensure $v_+ = v_-$.
2. v_+ and v_- draw no current.

3.2 Simple Opamp Circuits

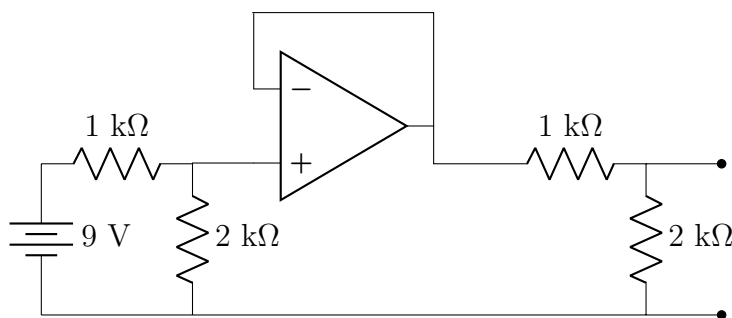
Implicit in rule two is that an opamp has immensely high resistance. That means that the input resistance seen by earlier circuit components is effectively an open circuit ($R = \infty$) for an ideal opamp. The simplest opamp circuit is seen as follows¹⁰:



This is called a *follower*, or *buffer*. In such diagrams, usually $-V$ and $+V$ are omitted. Based on our golden rules, v_- should match v_+ , and our circuit shows $v_+ = V_{in}$, so therefore V_{out} will be increased by the opamp until it reaches V_{in} , because v_- is tied to V_{out} . The point being, the circuit follows as $V_{in} = V_{out}$. You might say “so it’s just a wire then?” In many ways, you are correct! But here is a nice application that shows why a follower would be useful:



Recall this circuit from before. The important principle we learned was that because the R_{th} of the first divider was similar to the R_{eq} of the second, you could not treat this as $V_{in} \times \frac{2}{3} \times \frac{2}{3}$. However, if we do the following:

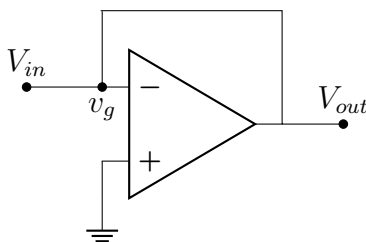


⁹These two specifications are sometimes called the 0th rule.

¹⁰Get it?

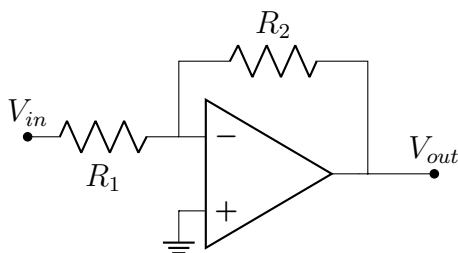
Now, the first voltage divider will “see” an input resistance of the opamp as being infinite. The opamp’s output will be the voltage of the first voltage divider, and apply it to the second voltage divider. As such, we do get $V_{in} \times \frac{2}{3} \times \frac{2}{3}$.

v_- is commonly called the *inverting input*, because one can invert with it. An inverting follower would be seen as follows:



Intuitively, since v_+ is grounded (or, $V = 0V$), to make $v_+ = v_-$, v_- must equal $-V_{in}$, and or, $V_{out} = -V_{in}$. Because of this, this node is sometimes called a *virtual ground*, or v_g .

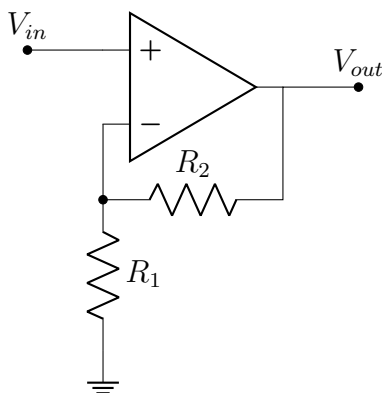
Given some of these ideas, we can imagine some special opamp uses. Like:



This is called an *inverting amplifier*. Sometimes you’ll see circuits like this overcomplicated in explanation. In reality, you should just recognize that the current across R_1 must equal the current across R_2 , since no current flows into v_- . Automatically, this means since $V_{in} = R_1 I$, and $V_{out} = -R_2 I$ (due to direction of flow), you’ll find that:

$$V_{out} = -\frac{R_2}{R_1} V_{in} \quad (3.1)$$

A slightly more complex application is:



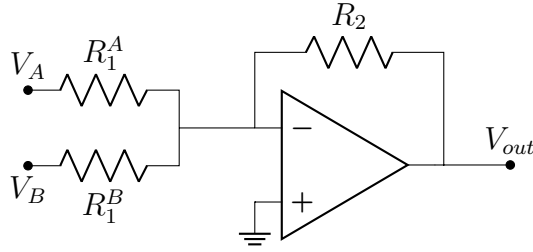
This is called a *non-inverting amplifier*. In this case, you can see that R_1 and R_2 form something of a voltage divider. That is, $V_- = R_1/(R_2 + R_1) \times V_{out}$. Knowing that an opamp always wants to ensure $v_- = v_+$, we can say:

$$\begin{aligned} V_{in} &= \frac{R_1}{R_2 + R_1} V_{out} \\ V_{in} \left(\frac{R_2}{R_1} + 1 \right) &= V_{out} \end{aligned} \quad (3.2)$$

The output resistance of an ideal opamp should be 0. That is, ideally an opamp would be able to drive any voltage irrespective of the current. The input resistance of the inverting amplifier will be whatever value we have for R_1 , while the input resistance for the non-inverting amplifier is ∞ , as it is directly connected to the opamp's v_+ .

Another thing you may wonder is if we can drive an infinite voltage at V_{out} . Recall that we are limited by whatever voltage is provided to the opamp's power rails, shown at the start of this section.

An expansion of the inverting amplifier is the *summing amplifier*:

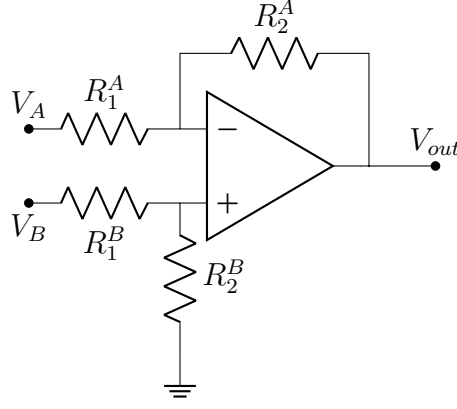


The interpretation is largely the same. Given that all of the currents through individual resistors will sum together through R_2 :

$$\begin{aligned} V_{out} &= -R_2 \left(\frac{1}{R_1^A} V_A + \frac{1}{R_1^B} V_B \right) \\ V_{out} &= -\frac{R_2}{R_1^A} V_A - \frac{R_2}{R_1^B} V_B \end{aligned} \quad (3.3)$$

Perhaps one application would be if you were interested in measuring the overall activity within the brain, and you are not so interested in discriminating individual electrodes in an EEG. Because the voltage changes measured by an EEG are so small, you'll need to amplify it. Thus, you throw them all together into a summing amplifier.

Next is a *differential amplifier* (not to be confused with derivative).

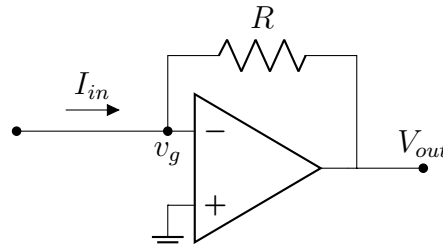


Knowing that $v_- = v_+$, and that $v_- = V_A - I_A R_1^A$ and $v_- = V_{out} + I_A R_2^A$:

$$\begin{aligned}
 \frac{V_A - v_-}{R_1^A} &= \frac{v_- - V_{out}}{R_2^A} \\
 R_2^A(V_A - v_-) &= R_1^A(v_- - V_{out}) \\
 R_2^A V_A + R_1^A V_{out} &= R_1^A v_- + R_2^A v_- \\
 \frac{R_2^A}{R_1^A + R_2^A} V_A + \frac{R_1^A}{R_1^A + R_2^A} V_{out} &= v_- \\
 v_+ = v_- &= V_B \left(\frac{R_2^B}{R_1^B + R_2^B} \right) \\
 \frac{R_2^A}{R_1^A + R_2^A} V_A + \frac{R_1^A}{R_1^A + R_2^A} V_{out} &= V_B \left(\frac{R_2^B}{R_1^B + R_2^B} \right) \\
 R_2^A V_A + R_1^A V_{out} &= R_2^B V_B \\
 R_1^A V_{out} &= R_2^B V_B - R_2^A V_A \\
 V_{out} &= \frac{R_2^B}{R_1^A} V_B - \frac{R_2^A}{R_1^A} V_A
 \end{aligned} \tag{3.4}$$

So that if $R_2^A = R_2^B$, you get $V_{out} = \frac{R_2}{R_1}(V_B - V_A)$. An important application of this is most physiological signals. Physiological signals are accompanied by an incredible amount of background noise, so it is useful to subtract an electrode placed on your location of interest with some other reference electrode. For example, if someone sustained damage to their left bicep, and doctors were using EMG recordings to measure recovery, an electrode should be placed on the left bicep and the right bicep. While the left bicep flexes, the right bicep can remain relaxed, and one can subtract the two signals to get the signal specific to the left bicep's flex. Because we are reading muscle depolarization through the skin, the measurable ΔV will be extremely small, hence the need to amplify it.

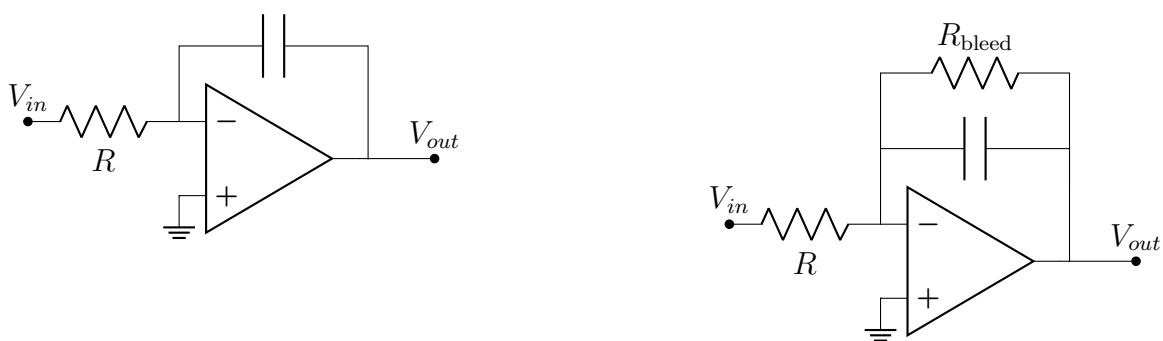
Next we will show the *current-to-voltage amplifier*. V_{out} will simply be $I_{in} \times R$:



Using an opamp is particularly beneficial, in this case, because if your current source is weak, you will not further inhibit it by adding a resistor in its path. Whatever current source you use will essentially drive a short circuit, due to the opamp's virtual ground (v_g)—and current sources love short circuits.

You may ask why you would ever need to convert current to voltage? Or, when would you be able to read current as opposed to voltage? An example from Prof. Ashmanskas is in PET scanners—photons emitted from fluorodeoxyglucose (18F) hit the walls of the PET scanner, and excite atoms enough to release electrons, generating an extremely small current. Really all piezo-electric compounds can be measured in this way. You can imagine one mode of stimulation would be implanting some piezo-electric compound and inducing electron release through low-intensity ultrasound. In testing, you'll need to measure such electron release using a current-to-voltage amplifier.

The next circuits are both integrators.

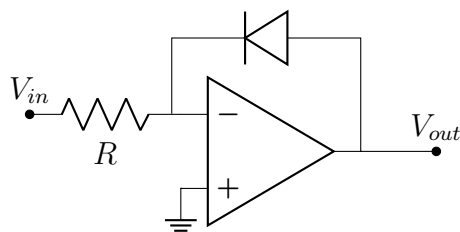


Both function as integrators (and low-pass filters), much like the low-pass filter discussion from long ago. That is, the current through the resistor is the same as through the capacitor, allowing us to use $I = CVdt$ to solve. Usefully, though, there is not a voltage limit as there is with the low-pass filter of the earlier parts. If you recall our opamp Golden Rules from section (3.2) you'll recall that one of the requirements is that there will be feedback. You may then realize that if V_{in} 's average over time is non-zero, there cannot be proper feedback in the left circuit, as the capacitor will block DC. Therefore, the second circuit uses a bleed resistor to allow for feedback when the capacitor is fully charged. This can be a bit dubious, as it is not expressly clear how this will affect your integration, but you can imagine that if your $R_{bleed}C$ value is large, it will drain charge from the capacitor over some long time scale. Intuitively, as the frequency applied to V_{in} increases, the amplification converges to R_{bleed}/R .

Another option is to use a switch to replace R_{bleed} . Therefore, whenever you want to begin integration you can flip the switch to zero the charge on the capacitor. Of course, the capacitor will saturate eventually, but you can “reset” it as desired.

As for applications, one interesting thing you can imagine is using an analog circuit to determine total muscle activity over time. Rather than digitally recording thousands of data points using an EMG and averaging them, you can use an integrator and record one voltage value at the end of your flexion—thus, giving some idea of total muscle output on a relative scale.

The last of this section will be the *logarithmic amplifier*:



I'll not explain the math so much, but you can think about the LED example from the very first section. A diode's current curve is exponential in nature, so the voltage will approach asymptotically toward the diode's diode-drop¹¹.

Biological applications of this are more obscure, but perhaps you can imagine a scenario where you'd really only care about the lower bounds of some stimulus. For example, if measuring the sub-threshold oscillations of a neuron, you may not want a linear curve—because once the neuron reaches its threshold, the voltage rises dramatically. However, a logarithmic graph of the voltages between hyperpolarization and threshold might be quite useful, and you can configure it so that your circuit approaches its asymptote as the neuron approaches its threshold voltage.

In Summary.

Some of the circuits here are pretty neat. We learned some ways in which analog circuits can perform relatively complex functions, like multiplication, subtraction, addition, integrals and even logs.

3.3 Opamp Imperfections

It is important to understand the slight opamp imperfections that cause them to deviate from being a perfect voltage source, etc. The first two clear deviations from ideality are that the opamp's *input resistance* is not truly infinite, but is something on the order of $10^{10}\Omega$. Secondly, an opamp does not have infinite gain, A , that it can use to drive V_{out} . I.e., $V_{out} = A(v_+ - v_-)$ has some limit in that $A \neq \infty$.

Another difference is a slight offset that will always exist between v_+ and v_- , called the *offset voltage*. In modern opamps, this number might be immeasurably small, but in the canonical LM741 used in most electronics classrooms, it is close to 1mV. This is simply due to difficulty in creating two distinct nodes of exactly the same potential.

As mentioned, the opamp's input resistance cannot be infinite, meaning it will draw some small but non-zero current. This current is called the *bias current*, or I_{bias} , or I_{b+} and I_{b-} for individual inputs. I_{b+} and I_{b-} are slightly different, so I_{bias} typically refers to $1/2(I_{b+} + I_{b-})$. There is also a current due to the offset voltage, which is typically an order of magnitude or two below I_{bias} . I_{bias} for the LM741 is $\approx 100\text{nA}$. The LM741's internal make-up is largely bipolar junction transistors (BJT), while higher-end opamps will use field effect transistors (FET), which allow for many orders of magnitude higher input resistance.

The LM741 is only capable of driving a small current of about 25mA. This is actually a very considerable point to keep in mind. Of course, this means an opamp cannot directly power any meaningfully

¹¹I still haven't done a Part on diodes yet, so the term "diode-drop" is not yet explored.

high power devices without the help of some external transistors.

Notably, all transistors have some finite capacitance, so it implies that the inner workings of an opamp may demonstrate some of the same phase shift or filtration we saw in the capacitors section. This is indeed true, and is worth considering at exceptionally high frequencies. Fascinatingly, at such exceptionally high frequencies, where the phase shift extends beyond 180° , the negative feedback switches to positive feedback. This effectively switches your opamp to comparator mode as V_{out} oscillates from its V_+ to V_- . Intentionally, opamp designers include a low-pass filter within it to avoid this, reducing the gain, A , before you can reach 180° . This is called *frequency compensation*, and again is an intentional feature. This is potentially quite interesting, as the small, immensely fast oscillations at physiological levels could theoretically exceed the opamp's 180° . So, it is curious to consider if such situations would cause positive feedback.

Opamps also have a *slew rate* which defines some $\max V_{out}dt$. That is, if the input is a square wave, the V_{out} cannot instantaneously follow it, and there will be some slight slope. For the LM741 it is $\approx 0.5\text{V}/\mu\text{s}$. The result is related to the internal capacitance and its charge time, making the effect non-linear and giving it some amplitude dependence. Driving a high amplitude and high frequency can therefore introduce considerable distortion.

NOTE: It would maybe be useful to make comments about the *output resistance*. Though, it is not immensely interesting.

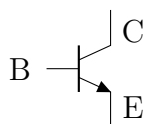
Chapter 4

Transistors

4.0.1 Transistor Overview

The prime purpose of a transistor is to use a small current to control a large one. A straightforward example being using the turning of a key to start a car. Naturally, you wouldn't want a humongous current, capable of powering your car, to be connected to the keyhole—so you use this small current to control the flow of a larger one.

Bipolar junction transistors (BJTs) are 3-terminal devices either composed as NPN or PNP. The three terminals are the emitter (E), base (B), and collector (C). The large current you are interested in controlling flows from the collector to the emitter, and is operable by flow at the base. Though, the current at E will indeed be $I_C + I_B$, but you'd expect I_B to be many orders of magnitude below I_C . The relationship between I_C and I_B is called β , where $I_C = \beta I_B$.

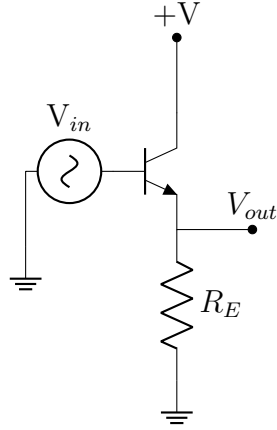


For an NPN transistor, it can either be off, where $I_B = 0$ (in this case, the voltage at the base will be below a diode drop), it can be active, where $I_B \neq 0$ and $I_C \leq \text{max}$, or it can be saturated, where $I_B \neq 0$ and $I_C = \text{max}$. You can probably imagine: active mode is useful if you'd like to make a transistor-based amplifier, while the saturation mode might be more useful in making a switch.

4.1 Transistor Circuits

4.1.1 A Follower

A follower, as with the opamp follower, is useful in building a high-input impedance, low output impedance element in your circuit. Again, this is useful as it will draw minimal current and preserve the voltage at that point in the circuit.



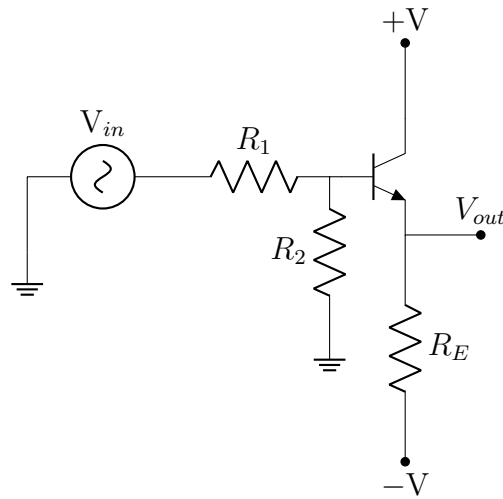
Alrighty, so in active mode, if this functions properly as a follower, we would expect that as V_{in} changes by ΔV , so too would V_{out} . Too, it will function very much like a diode—meaning V_{out} will feature a diode drop and cannot drop below 0. In later circuits, we'll show how to fix this by tying E to some negative voltage, $-V$. Thus, we get that: $\Delta I_B = \Delta I_E / (\beta + 1) = \Delta V / R_E (\beta + 1)$. Since $R_{in} = dV_{in} / dI_{in}$, we get:

$$R_{in} = (\beta + 1)R_E \quad (4.1)$$

Therefore, from the perspective of V_{in} , the load, which would be attached at V_{out} now appears to be much higher. More explicitly, it appears as:

$$R_{in} = (\beta + 1)(R_E || R_{load}) \quad (4.2)$$

Again, this is preferable for the source. As we explored in the very early voltage divider section, if you were to have the following circuit:



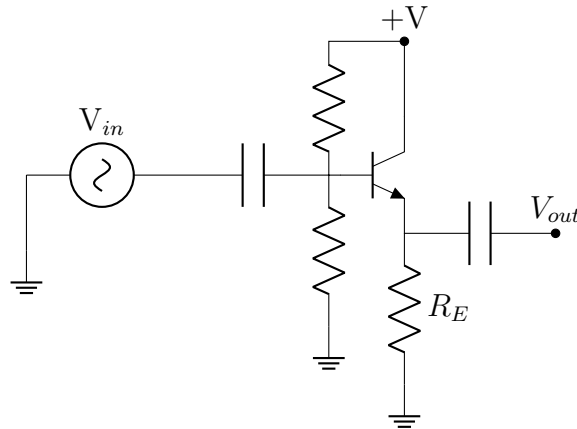
Now we have $R_2 || (\beta + 1)R_E$, which we would expect to be $\approx R_2$. Therefore, the voltage divider acts as it would if there were no load there at all. The output impedance will be that which exists at V_{out} . Since the current at ΔV_{out} varies with ΔV_{in} , and ΔI_{out} varies with $(\beta + 1)\Delta I_B$. So, $\Delta V_{out} = \Delta I_B (R_1 || R_2)$ or $\Delta V_{out} = \Delta I_B R_{th}$. Since $R_{out} = dV_{out} / dI_{out}$, we get:

$$R_{out} = \frac{R_{th}}{\beta + 1} \quad (4.3)$$

Which we expect to be quite small. Therefore, the output resistance is quite small. Since technically R_{load} is again in parallel with R_E , it should actually be:

$$R_{out} = \frac{R_{th}}{\beta + 1} || R_E \quad (4.4)$$

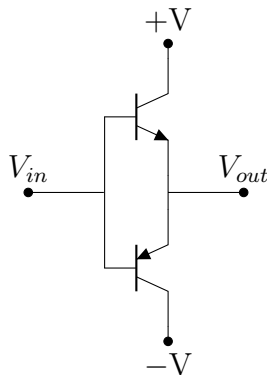
Which again we expect to be $\approx R_{th}/\beta + 1$, and or, a very small value. You could build a more complete BJT follower in this way:



I'll not walk through the entire idea, using explicit values, but you should be able to get the gist. The capacitors are used to remove the DC. The $+V$ adds a DC-offset, so that the oscillations of V_{in} can be all positive, removing the restriction of keeping V_E above 0. The two resistors are used as a voltage divider, stemming from $+V$, allowing us to pick our DC offset. One of the largest takeaways from this is simply that it is much more complex, and annoying, than opamp-based followers. However, it should give insight into what might be within an opamp.

Push-Pull Follower.

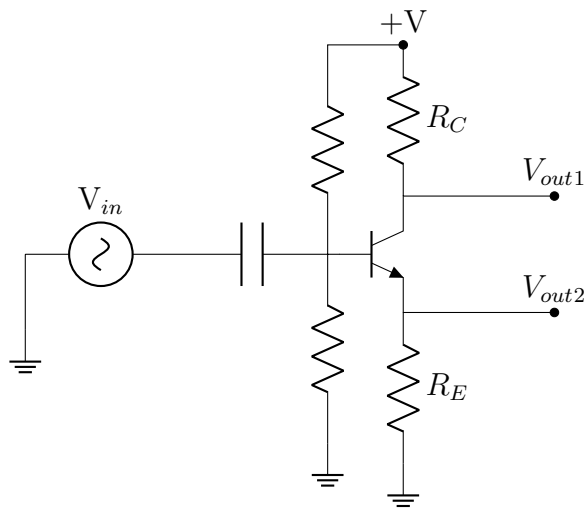
A neater way to avoid some of the drawbacks to the preceeding designs is to use what is called a *push-pull follower*. It takes advantage of using two BJTs. Where the first NPN handles the positive half, and the second, now a PNP, handles the negative half. It is seen as below:



The problem, of course, is that you will still have two diode drops in either direction. This is called *crossover distortion*.

4.1.2 Amplifiers

The first iteration is the common emitter amplifier, seen as follows:



Let's again analyze solely by glancing at the circuit. We know that the voltage at V_E will be one diode drop (say, 0.6V) below V_B in the active mode. We know that the current through R_E must therefore be $(V_B - 0.6)/R_E$. We know that $I_C \approx I_E$, therefore $I_C = (V_B - 0.6)/R_E$. If $V_B = V_{in}$, this must mean that the voltage at C:

$$V_{out1} = +V - (V_{in} - 0.6) \frac{R_C}{R_E} \quad (4.5)$$

Since the voltage drop across R_C will be $I_C R_C$. This, $dV_{out1}/dV_{in} = -R_E/R_C$. Using a capacitor to remove the DC offset will further make it function like the amplifier you desire.

4.1.3 Current Sources

Let's take this circuit:



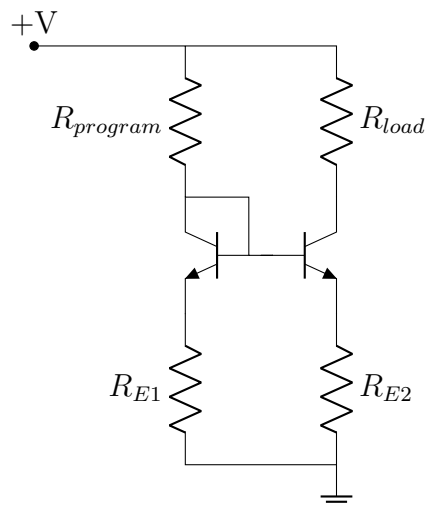
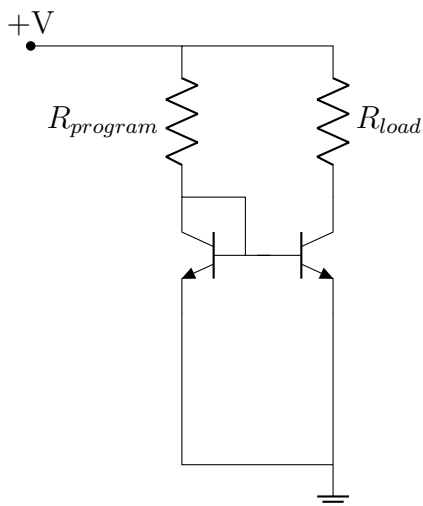
Recall that the current through the collector is set by the emitter. The voltage at the emitter is set by the base. V_E will be one diode drop below the base (for the leftward circuit, that is—for the rightward, it will be one diode drop above). So if the resistance at the emitter, R_E , is 1 k Ω , then we would want V_E to be 2 V, and V_B to be 2.6 V. Note that the transistors input resistance is $\beta(r_e + R_E) \approx 100k\Omega$.

Keeping a small R_E is important here, as the larger that R_E gets, the closer you will be to saturating your R_{load} . Let's try to consider an example. Covered much, much later in this book, we discuss a Courtine lab paper that used electronic dura mater. The load value appeared to vary all the way from 5k Ω to 100k Ω . If you wanted to supply a constant 2mA over this entire impedance range, you would need the following:

$$\begin{aligned}
 V_B &= 2.6\text{V} \\
 R_E &= 1\text{k}\Omega \\
 I_C &= 2\text{mA} \\
 V_C &= +V - R_{load}I_C \\
 V_C &= +V - (100k\Omega \times 1\text{mA})
 \end{aligned}
 \tag{4.6}$$

It looks like you'll need a 100V supply to keep your transistor from saturating. That is... not good!

Current Mirror.



On the left circuit, your $R_{program}$ sets the current via the fact that the voltage drop between $+V$ and ground happens only across $R_{program}$, so the current flowing through the emitter is $+V/R_{program}$. The same current will be pulled through R_{load} . Adding in the R_E on the rightward circuit gives a path to bleed and allows your programmed current to stay flat for longer. However, even in this case, R_{load} cannot exceed $R_{program}$.

So why would you want a current mirror? Here's an example from this video¹². Amplifiers can be biased using current sources, which allows the biasing current to become very stable. This is important, because as voltage or temperature fluctuates, the current can be maintained. Adding a current source to every amplifier would require too much annoying circuit building, so one way to avoid this is using a current mirror.

¹²https://www.youtube.com/watch?v=VnJHXQCPIvs&ab_channel=ALLABOUTELECTRONICS

Chapter 5

Writing Hardware

5.1 Introduction

Verilog¹³ is a language used to describe electronics, and allows you to avoid the physical action of wiring. This is the reason for the designation **writing hardware**. In this way, you can pick your poison: debugging code, or debugging breadboards. Importantly, though, Verilog is capable of computation and writing data files that go beyond circuit descriptions. Thus, it is not a “markdown” language and is Turing Complete¹⁴.

Tools like Field Programmable Gate Arrays (FPGAs) allow for this, as their internal composition is something of an array of transistors, which can be rewired through code in order to meet the demands of the programmer.

5.1.1 Creating Modules

In Verilog, a circuit is called a `module`. Each module is defined between a `module` and `endmodule`, which can be named as shown in the example below. Different ports connect the module to things outside of the module.

```
1 module example1(o, i1, i2);
2 // example1 is the name of our module, and o, i1, and i2 are our ports
3 // it is convention to list outputs first
4
5 output o; // this defines o as an output
6 input i1, i2; // this defines i1 and i2 as inputs
7
8 endmodule
```

Gates are also initialized like modules. The way to do this is with the built in primitives for AND and OR gates (`and` and `or` respectively). For example:

```
1 module example2(o, i1, i2);
2
```

¹³A large part of the background information and general syntax comes from the YouTube channel: CompArchIllinois.

¹⁴Or at least, I think it is. I can never remember the exact definition of Turing Complete :)

```

3 output o;
4 input i1, i2;
5 wire wire1, wire2; // this initializes two wires called wire1 and wire2
6
7 or or1(wire1, i1, i2);
8 // this makes an OR gate named or1 with inputs i1 and i2, and output called
  wire1
9 and and1(o, wire1, wire2);
10 // this makes a NOT gate named not1 with input i2, and output called wire2
11 // one of the outputs of the OR gate feeds into the AND gate (via wire1) in
   this example
12
13 endmodule

```

The order in which things are initialized do not matter. It is very important to not reuse wire or other variable names, as Verilog will read these as being connected irrespective of where they are intended to be. As mentioned, the code above uses modules built into Verilog, but you could make your own module in the following way:

```

1 module andgate(output o1, input i1, input i2);
2     assign o1 = i1 & i2;
3     // for OR you would use |, and for XOR you would use ^
4 endmodule;

```

5.1.2 Bus Notation

Bus notation is used to simplify the pins used (in Verilog, this is called a vector). For example, a multiplexer or an adder will have many inputs, which would be inconvenient to initialize individually. Instead, we can use something like this:

```

1 module adder(c, a, b); // a, b, c are 3 bus inputs we will use
2     output [3:0] c; // initializes 4 wires within our c bus
3     input [3:0] a, b; // initializes 4 wires within our a and b buses
4 endmodule

```

Firstly, note that Verilog is 0 indexed, so `[3 : 0]` includes 4 wires. In a circuit schematic, busses are drawn as thicker wires with a slash through them and a number denoting the amount of wires in the bus. If we wanted to call individual wires from our busses into the `andgate` module we declared earlier, we could do it as:

```

1 andgate(c[0], a[0], b[0]);

```

And we can connect busses together, or wires together, using the assign command like before. For example:


```

1 wire wire3;
2 assign wire3 = c[2];
3 assign c[2:0] = a[2:0]; // wire3 will now be connected to bus a[2] through
   bus c[2]

```

5.2 Constants and Variables

Verilog allows us to define constants using 3 parameters, defined as their size, method of encoding, and value. For example, 8'hd7 corresponds to a size of 8 bits, hexadecimal encoding, and the value d7 (equivalently, 11010111). You can use this in `boolean` comparisons, as below:

```

1 wire wire4;
2 'define CONST1 3'b011; // CONST1 is the name of the constant
3 wire4 = (a[2:0] == CONST1);
4 // This is a tad complicated. The gist is: all of the bit values in a[2:0]
   will be compared to CONST1 in a NXOR style statement. That will then be
   compared to all of the other bits in an AND style statement. So wire4
   will be on only if all wires in correspond to CONST1.
5 // The C++ equivalent would be something like:
6 //     (a[2] == 1'b0) && (a[1] == 1'b1) && (a[0] == 1'b1)
7 // Also, this could be completely wrong. I can't check any of this without
   an actual FPGA in front of me! So who knows!

```

This does bring up a worthwhile point, which is that everything you write in Verilog has a direct circuit component underlying it (I suppose the same is true for any program, but it's more... explicit with Verilog).

Part II

Math and Models

All truly strong people are kind.
– Vagabond by Takehiko Inoue

5.3 Perspective

Truly, I find it endlessly fascinating and miraculous that you can quantitatively describe physical or biological phenomena. Of course, it stands to reason, but its shine is not lost on me. I vividly recall being in my Human Physiology class and having our professor pose the question: *what will happen if two action potentials run into one another?* I knew that they'd terminate, so the result was not what drew my intrigue. What was miraculous is that, to answer this question, our professor pulled up MatLab and simulated it before our eyes. Indeed, the action potentials terminated.

Prior to this, I hadn't thought much about quantitative descriptions like this. Qualitatively, and or conceptually, it made sense to me that action potentials cannot propagate past one another. But naturally, not all answers can be intuited. After scrolling around through the amazing Keener and Sneyd *Mathematical Physiology* textbook, I realized the incredible breadth of phenoma that we have modeled. From membrane electrophysiology, enzyme kinetics, to disease spreading, to fluid dynamics of the circulatory system, to the movement of bacterium, to every other little itty-bitty thing you can become fascinated by—someone has build some differential equations for it.

If anyone ever reads this, besides me, I hope you too will be enthralled by both the brilliance and creativity required to describe something as complex as the biochemistry of a neuron in just a few equations.

Chapter 6

Math Essentials

6.1 Linear Algebra

6.1.1 Cross Products, Dot Products, and Gradients

Recall that the cross product is a vector value whose direction is perpendicular to both of the two vectors crossed. The formula is:

$$\vec{A} \times \vec{B} = \|\vec{A}\| \|\vec{B}\| \sin \theta \vec{n} \quad (6.1)$$

Where $\|\vec{A}\|$ is the length of \vec{A} , θ is the angle between the two vectors, and \vec{n} describes the unit vector between \vec{A} and \vec{B} .

The dot product is a scalar value achieved by multiplying the positions of vectors together, such as below:

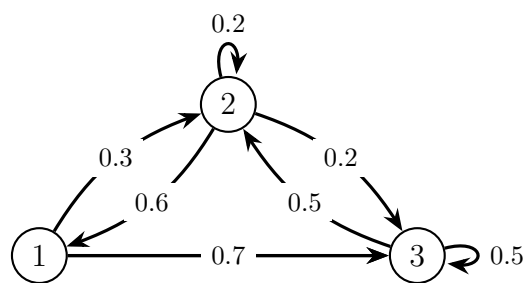
$$A = \begin{bmatrix} a \\ b \end{bmatrix} \cdot \begin{bmatrix} x & y \end{bmatrix} = ax + by \quad (6.2)$$

A gradient, ∇ , describes a derivative vector of a corresponding vector. For example, if vector v_1 exists in 3 dimensions, then:

$$\nabla \cdot \vec{v}_1 = \begin{bmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{bmatrix} \cdot \vec{v}_1 \quad (6.3)$$

6.2 Markov-chains

Markov chains are useful in predicting the next state desired.



Chapter 7

Neuron Modeling

It would certainly be worth reviewing the electrophysiology contained in Part **III** before trying this, unless you already have a strong understanding of biochemistry.

7.1 Hodgkin-Huxley

7.1.1 The Main Form

The pair won the nobel prize for this model, which formed the basis of our understanding of action potentials. Beyond neurons, it was used in modeling pacemakers of the heart, and muscle cell depolarizations before better models existed. The basis is simply KCL:

$$C\dot{V} = I - I_{Na} - I_K - I_{Leak} \quad (7.1)$$

Because the equations can be found in nearly any textbook¹⁵ or Wikipedia page, I will focus on some of the conceptual understanding I had issues with at first. The complete equation Hodgkin and Huxley arrived at is as follows:

$$C\dot{V} = I - \bar{g}_{Na}m^3h(V - E_{Na}) - \bar{g}_Kn^4(V - E_K) - \bar{g}_L(V - E_L) \quad (7.2)$$

There are a few main points to make here. Firstly, this model considers only 3 currents. Na and K are self explanatory, but $Leak$ represents the small amount of current that will always occur in cells due to the many routes of charged particles passing through the membrane. It is restorative, in that it pushes the membrane potential back to the resting voltage.

7.1.2 Gating and Conductance

The \bar{g} represent the maximal conductance of these ions. Of course, you may ask, “*shouldn’t conductance be variable, depending on how many channels are open?*” Yes, which is what m , h , and n are

¹⁵Izhikevich, *Dynamical Systems Neuroscience*

for! These three variables are effectively kinetic fits of the opening and closing dynamics of sodium and potassium channels. Again, I will not mention these equations explicitly as they can be found anywhere. Conceptually, there are three things to know:

Firstly, m is an activation curve for sodium, and the power to the 3rd represents that there are three activation gates. h is an inactivation gate for sodium. Potassium has 4 activation gates, n , and no inactivation gate. Gating can be any number of things, for example, h could be a conformational change that occurs in the channel after it has been open for $0.1ms$ that closes it again. Naturally, the gating for every channel will be different. Because the *Leak* current is an ensemble of many channel types, it will not have "gating" per se.

Secondly, m , h , and n all are between 0 and 1 and represent the **proportion of channels open**. For instance, if $n = 1$, then 100% of potassium channels will be open. This is why we multiply by the maximal conductance.

Thirdly, m , h , and n are dependent upon voltage, which affords them a time constant τ . This is the conceptually most difficult part. The experimental explanation may be beneficial in understanding. Hodgkin and Huxley realized that these three gating variables will converge to different values depending on the voltage. This makes sense, because we know potassium channels are voltage gated, we would expect the gating variable n to converge to around 1 as the voltage increases. But, the rate at which channels open and close is different. Therefore, their experiments were done to vary the voltage and determine how long it took the conductance of the channels to converge to some value. Does this make sense? In simplest terms: channels open and close at different rates, and that depends on the voltage.

What is the implication of this? Again, look up the exact equations if you are interested. Otherwise, trust the following: m has a time constant τ_m which is very small compared to τ_h and τ_n . Meaning, sodium channels will open the fastest in response to a voltage increase, causing depolarization of the cell. After some delay, sodium channel inactivation (h) and potassium channel activation (n) will kick in, causing repolarization and then hyperpolarization.

These are all derivatives.

One of the most difficult conceptual understandings I had was that \dot{V} , m , h , and n are all rates that depend on different time constants, which take voltage as their input. So, the derivative of voltage depends on the derivative of m , h , and n , which depend on voltage. The cyclic nature of this makes it strange, but still doable. Use the general form of derivative, $x_{i+2} = x_{i+1} + (x_{i+1} - x_i)/t$, follow the math, and you will survive.

7.2 Fitzhugh-Nagumo Reduction

7.2.1 Why would we simplify this system?

Reduction implies we are reducing the amount of variables. But why would we do this? The system is already incredibly generalized. We only consider two ion channels and are looking at a static neuron. How can we be accurate if we simplify this system any further?

Let's start by doing a simple thought experiment regarding the previous model:

$$C\dot{V} = I - \bar{g}_{Na}m^3h(V - E_{Na}) - \bar{g}_Kn^4(V - E_K) - \bar{g}_L(V - E_L)$$

As mentioned, m , h , and n have their own time constants $\tau_{m,h,n}$. That means you'll need to do at least 6 calculations in order to determine \dot{V} , which, because it is a derivative, has its own time constant τ_v . Thus, the whole equation is 4th dimensional with respect to time and requires at least 7 or so calculations per time step. If you'd like to simulate an action potential for around 10ms with a time step of 0.01ms, that means you'll perform around 7,000 calculations. Which is not so bad!

However, let's say you want to attempt a propagating action potential. Many people would model this on an infinitely long neuron/wire, but for the sake of this thought experiment let's say you're just interested in a 1 cm neuron/wire for 10 ms. To account for this spatial consideration, you'll need to add in another term besides I which receives current input from the previous segment of the neuron. So, this brings us up to at least 8,000 calculations.

You'd probably want to divide up the neuron into segments on the order of 1 μm . This multiplies our 8,000 calculations by an additional 100,000, giving us 800,000,000 to worry about. Still, this is not horrendous. But, this considers a 1D wire. Neurons are 3D dimensional. We are already considering a system that is 4th dimensional with respect to time, and now we desire to consider 3rd dimensional with respect to space. Imagine trying to calculate the flux through a $1000 \times 1000 \times 1000$ resolution box (i.e., perhaps μm^3 with good resolution). The surface area of this box is thus 6×10^6 . Now extend this surface area to include the length of the wire and the area of the soma and dendrites, giving you thousands of millions of points to calculate per time iteration. And, we are still only considering two ion channels. Neurons have dozens and dozens of channels all with different gating kinetics. It does not consider things like lateral inhibition, birufcation, dendritic input, etc. I'll not bother telling you how many calculations we need to perform beyond this point—but it would be large.

7.2.2 How to Reduce

What do we know about the time constants mentioned in the previous section? Roughly speaking, some are fast and some are slow. The upswing of an action potential is on a fast time constant, and the repolarization is on a slow time constant. We also know that the upswing portion is roughly a positive feedback loop, so as voltage increases, so should the derivative of voltage.

This helps us arrive at least at the following:

$$\dot{V} = V \times f(x) \tag{7.3}$$

Simply meaning that the derivative should scale with voltage in some way. We also know that there are at least two “equilibrium points” in a neuron. Meaning, when the neuron is at rest, the \dot{V} will be zero. And, when the neuron reaches the peak of the action potential, the same is true. This will allow us to immediately assume something interesting:

$$\dot{V} = V(V - V_{rest})(V - V_{max}) \tag{7.4}$$

We are already almost there. What we have just done is said that when either $V = V_{rest}$ or $V = V_{max}$, the \dot{V} will not change. These are all on the aforementioned “fast” time scale, and as this is representative of the activation of the action potential, it is effectively a simplification of the sodium channel dynamics. This is also extremely easy to measure experimentally.

On our second time scale, the slow time scale, we have the inactivation/repolarization function. How will this look like? Just as with the first equation, we will want this curve to increase in magnitude with voltage. Because n represented the potassium channel activation in the previous segment, we can use that as our repolarization function here.

$$\dot{n} = V - \gamma n \quad (7.5)$$

What does this say? It says that our repolarization curve \dot{n} will increase with respect to voltage. But, it will also decrease with respect to itself according to some scaling factor γ .

Now we have reduced our function down to two dimensions and can combine terms:

$$\begin{aligned} \dot{V} &= V(V - V_{rest})(V - V_{max}) - n \\ \dot{n} &= V - \gamma n \end{aligned} \quad (7.6)$$

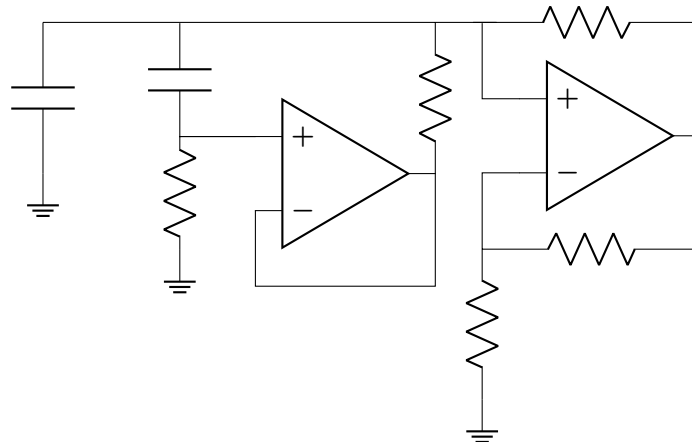
But, we still want voltage to be affected by an injected current, so we can simply add this term back in. And it is also in this equation that we will add our spatial dependence to reach the following:

$$\dot{V} = I_{app} + [V(V - V_{rest})(V - V_{max}) - n] + D \frac{\partial V^2}{\partial x^2} \quad (7.7)$$

D is our spatial dependence, which represents the diffusion of charge around the neuron membrane. And that's it, for now!

Circuits digression.

A lot of the original work done by Fitzhugh and Nagumo used circuit equivalents in order to model neurons. One such example is as follows¹⁶:



¹⁶This circuit is adapted from *Mathematical Physiology*, by James Keener & James Sneyd (1998).

The rightward opamp functions as a Schmitt Trigger, and the entire thing is effectively an opamp oscillator with a second opamp in the middle. The purpose is to simulate an excitable system, like a neuron that is continually firing. Excitable systems are those that fire and have some refractory period before firing again (for example forest fires, or even your toilet).

7.3 Diffusion

Note that in the coming sections, $[f]$ is used to describe the concentration of some molecule f . But, this principal can be applied to anything, including the spreading of voltage across some surface.

7.3.1 Forward Euler's

Diffusion is accomplished using some diffusion coefficient (D_c) multiplied by some measure of the proportion in one compartment versus another (often $\partial^2[f]/\partial x^2$). D_c can be tuned however desired. The important bit is the second derivative of $[f]$ with respect to space. This can be done using the general form of a second derivative, as written below:

$$f'' = \frac{f_{x+1} - 2f_x + f_{x-1}}{x^2} \quad (7.8)$$

One may wonder how one would solve for an edge case, as the general form of a second derivative requires three data points. There are some nuances, but in general the solution is simply the first derivative of the non-edge side. That is, since the second derivative is the difference in derivatives, that leaves simply the derivative of one side minus zero. This is like applying a closed end to your surface. You can ponder how to solve for an open end, if that ever arises.

This method has some slight issues in which the $[f]$ can occasionally go negative. The way in which this occurs is stated below (note that now $[f]$ is used instead of f to signify concentration at a value x and time t). We can first describe a simplified version of the problem:

$$\begin{aligned} [f]_i &= f_0 \exp(-mt) \\ \frac{d[f]}{dt} &= \frac{[f]_{i+1} - [f]_i}{\Delta t} = -m[f]_i \\ [f]_{i+1} &= (1 - m\Delta t)[f]_i = [f]_i - m\Delta t[f]_i \\ [f]_i &= (1 - m\Delta t)[f]_{i-1} = (1 - m\Delta t)^2[f]_{i-2} \\ [f]_i &= (1 - m\Delta t)^i[f]_0 \end{aligned} \quad (7.9)$$

You can see easily, from this, that if Δt is too big, you will abandon the characteristic decay you'd expect from $f_0 \exp(-mt)$, and instead get some diverging oscillatory function. How this applies to our interest in diffusion is described below:

$$\begin{aligned} \frac{[f]_x^{t+1} - [f]_x^t}{\Delta t} &= \frac{[f]_{x+1}^t - 2[f]_x^t + [f]_{x-1}^t}{x^2} \\ [f]_x^{t+1} - [f]_x^t &= \frac{\Delta t}{x^2} ([f]_{x+1}^t - 2[f]_x^t + [f]_{x-1}^t) \\ [f]_x^{t+1} &= \frac{\Delta t}{x^2} [f]_{x+1}^t + \left(1 - 2\frac{\Delta t}{x^2}\right) [f]_x^t + \frac{\Delta t}{x^2} [f]_{x-1}^t \end{aligned} \quad (7.10)$$

Therefore, if we want to ensure that the concentration is always positive, we are constrained by:

$$1 - 2\frac{\Delta t}{x^2} = 0$$

$$\Delta t < \frac{x^2}{2}$$
(7.11)

The relevance of this being that if one were interested in modeling on a very small Δx , then one would have to use a Δt that is not physiological, and thus waste a great deal of computing power in doing so. This can be avoided explicitly using some other methods, discussed next.

7.3.2 Backward Euler's

This form serves to solve the time-scale dilemma by swapping $[f]_{x+1}$, and can be used with any Δt . Let us consider the same example from above:

$$[f]_i = f_0 \exp(-mt)$$

$$\frac{d[f]}{dt} = \frac{[f]_{i+1} - [f]_i}{\Delta t} = -m[f]_{i+1}$$

$$[f]_{i+1} - [f]_i = -m\Delta t[f]_{i+1}$$

$$[f]_{i+1} = \frac{1}{1 + m\Delta t}[f]_i$$
(7.12)

Naturally, there is no longer a concern of the size of Δt . Though, one immediate concern is the difficulty of solving your equation for $[f]_{i+1}$. Getting back to the diffusion interest, we now have:

$$\frac{[f]_x^{t+1} - [f]_x^t}{\Delta t} = \frac{[f]_{x+1}^{t+1} - 2[f]_x^{t+1} + [f]_{x-1}^{t+1}}{x^2}$$

$$[f]_x^{t+1} - [f]_x^t = \frac{\Delta t}{x^2} ([f]_{x+1}^{t+1} - 2[f]_x^{t+1} + [f]_{x-1}^{t+1})$$
(7.13)

Which simply replaces the previous $[f]_x^t$ with $[f]_x^{t+1}$. This leaves us with three unknowns (those being $[f]^{t+1}$ at $x-1, x, x+1$). We must use linear algebra to solve by first rewriting the left and right side as vectors and a matrix in the following way:

$$\vec{[f]}^{t+1} = \begin{bmatrix} [f]_0^{t+1} \\ [f]_1^{t+1} \\ \vdots \\ [f]_x^{t+1} \end{bmatrix}; \vec{[f]}^t = \begin{bmatrix} [f]_0^t \\ [f]_1^t \\ \vdots \\ [f]_x^t \end{bmatrix}$$
(7.14)

and

$$A = \frac{\Delta t}{x^2} \begin{bmatrix} \dots & \dots & \dots & \dots & \dots \\ 1 & -2 & 1 & & \vdots \\ \vdots & 1 & -2 & 1 & \vdots \\ \vdots & & 1 & -2 & 1 \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$
(7.15)

The corners $1, 1$ and x, x were intentionally omitted, as what one desires to do with this is dependent on how they would prefer to treat their edges. As described before in the *Forward Euler's* method, one can use instead of the $(1, -2, 1)$ pattern, simply $(\emptyset, -1, 1)$ pattern, which signifies a closed edge. Therefore, together now we get:

$$\begin{aligned} [\vec{f}]^{t+1} - [\vec{f}]^t &= A[\vec{f}]^{t+1} \\ I[\vec{f}]^{t+1} - A[\vec{f}]^{t+1} &= [\vec{f}]^t \\ [\vec{f}]^{t+1} &= (I - A)^{-1}[\vec{f}]^t \end{aligned} \tag{7.16}$$

Where I is the identity matrix.

Remarks.

Forward Euler's is explicit, and will be preferred whenever the differential equations are non-stiff¹⁷. It is the more accurate of the two methods, and can be less computationally intensive if your decay rates are all slow. Backward Euler's is implicit, and can not be used to solve everything, but is doable in most cases. The extra computation required to solve the system of equations more than makes up for potential limitations in your Δt .

Ramblings.

I have this constant wonder if this method can be used to model the spread of voltage over a resistor lattice, by perfecting the D_c value. I am presuming it would work for a sufficiently large lattice, perhaps 100×100 . This would be a good exercise to do sometime in the future.

¹⁷On approximately the same time scale.

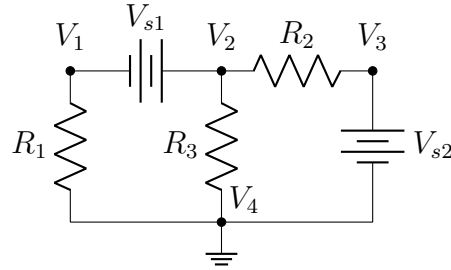
Chapter 8

Modeling Circuits

The idea of this algorithm comes from¹⁸ and is an expansion of what was discussed earlier in the **Electronics** part of this book. Not by coincidence, both of the examples in the earlier section had symmetric matrices. The general way to extract equations from a circuit is in the following pattern:

$$\begin{bmatrix} \mathbf{G} & \mathbf{V} \\ \mathbf{V}^T & \mathbf{Z} \end{bmatrix} \quad (8.1)$$

\mathbf{G} is essentially the way nodes are connected by resistors (represented as conductances). The diagonal represents the total number of resistors connected to a diagonal, and the other bits represents how nodes are connected by resistors. I will use the example provided in the footnote directly to illustrate:



There are 4 nodes, but as mentioned before, Node 4 is grounded, a reference, and thus not included in our calculations. This will mean our matrix \mathbf{G} must be $n \times n$, where n is $N - 1$. Node 1 is connected only to R_1 , resolving the $(0,0)$ ¹⁹ position of the matrix to be G_1 . Node 2 is connected to R_2 and R_3 , giving us $(1,1) = G_2 + G_3$. Node 3 to only R_3 , giving us $(2,2) = G_3$. Node 2 and Node 3 are connected via R_2 , meaning the 1,2 and 2,1 positions will have a G_2 , though notably, it will be $-G_2$:

$$\mathbf{G} = \begin{bmatrix} G_1 & 0 & 0 \\ 0 & G_2 + G_3 & -G_2 \\ 0 & -G_2 & G_3 \end{bmatrix} \quad (8.2)$$

There are two voltage sources ($V = 2$), and 3 nodes, meaning the matrix must be $n \times m$, where $n = N - 1$ and $m = V$. You fill the matrix as if the positive terminal of the j^{th} voltage source is connected to the i^{th} node, point $(i,j) = 1$, or $(i,j) = -1$ for the negative terminal. So for the above, since the negative terminal of V_{s1} is connected to Node 1, $(0,0) = -1$, and as the positive terminal connects to Node 2, $(1,0) = 1$. Lastly, as Node 3 connects to the positive terminal of V_{s1} , $(2,1) = 1$.

¹⁸<https://lpsa.swarthmore.edu/Systems/Electrical/mna/MNA3.html>

¹⁹You are a computer scientist in addition to a mathematician, so of course our matrices must be 0 indexed.

$$V = \begin{bmatrix} -1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (8.3)$$

$$V^T = \begin{bmatrix} -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (8.4)$$

And lastly, Z for zero is an $n \times n$ matrix of zeros where $n = V$:

$$V^T = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad (8.5)$$

Giving us an $Ax = B$ of:

$$\begin{bmatrix} G_1 & 0 & 0 & -1 & 0 \\ 0 & G_2 + G_3 & -G_2 & 1 & 0 \\ 0 & -G_2 & G_2 & 0 & 1 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ I_1 \\ I_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ V_{s1} \\ V_{s2} \end{bmatrix} \quad (8.6)$$

You may, rightfully, say “*Uh, who cares?*,” since we already know that each row corresponds to a different equation. Well, there are a few reasons. Firstly, doing this algorithmically allows us to avoid accidentally underdetermining our matrix using the ol’ eye-balling it technique. More importantly, though, the algorithmic approach allows us to solve this via code. Once I can write a bit more neatly, I’ll likely upload some snippets here.

Chapter 9

Cell Motility

9.1 Introduction

The life of small things is hard. Brownian motion reorients *E. coli* a full 90° approximately once per second. The objects which they are trying to chase after and eat, too, undergo Brownian motion. A good question to ask yourself, then, is what do they do to make it worth it?

9.2 Cell Motility

I have an idea. I'm wondering if you can use cell motility models, with a bit of adaptation, and use it to model growth cone motility. I.e., model an ensemble of neurons after injury. The motivation being that the architecture of the neurons after injury is quite important to recovery. This idea, sprouting, is explored in-depth throughout this work, such as in a discussed Courtine review. Similarly, the movement of axons on a single cell level is naturally a form of cell motility.

There have been a number of papers that have gathered the 3D structure of axons after injury—so it is not impossible to verify models^{20,21}.

9.3 Fluid Mechanics

There are two main open questions in the field of fluid mechanics. First, Newton's equations have been used for centuries and, in large part, model the world very well. However, we still haven't determined if these work completely, in all cases. Another open question is the origin of turbulence. Broadly speaking, we can define fluids as something that can be deformed, are continuous, and have no memory of its previous position. Complex fluids, like toothpaste, exhibit both fluid-like behaviors, but also retain their shape when left alone.

Knudsen number is used to define how continuous a fluid is:

$$K_N = \frac{\lambda}{L} \tag{9.1}$$

Knowing a system is continuous allows us to define dynamic, average quantities like $\rho(x; t)$, $p(x; t)$, $\mathbf{u}(x; t)$. Our pressure value, $p(x; t)$, is useful in defining the resistance of a fluid to force. Pressure is isotropic,

²⁰<https://www.sciencedirect.com/science/article/pii/S0896627321007753?via%3Dihub>

²¹<https://www.sciencedirect.com/science/article/pii/S2211124720308883?via%3Dihub>

meaning it is independent of direction. Defining the velocity, $\mathbf{u}(x;t)$, may be done by quantizing some small segment of the fluid and following its trajectory. This is the Lagrangian description. The trajectory that your quantized box follows is called the *pathline*. Alternatively, the Eulerian description treats the fluid as an ensemble, defining \mathbf{u} everywhere and producing a vector-field like array of streamlines. Conversion between the Lagrangian and Eulerian descriptions is:

$$\frac{Df}{Dt} = \frac{\partial f}{\partial t} + \mathbf{u} \cdot \nabla f \quad (9.2)$$

You can define movement of a fluid with three fundamental flows: translation, rotation (solid body), and extension. Translation and rotation both do not cause any deformation. The streamlines simply move in unison in some direction. Extension causes deformation, and have hyperbolic streamlines. A microfluidic chamber, in which 4 channels come together in a cross, is an example of extension flow. At the intersection, components will be stretched. Deformation like this will lead to friction.

The Reynolds Number.

A large part of this section will come from²² and²³.

The Reynolds number, R_e , describes the flow of a fluid and is used to characterize it as turbulent or laminar. Such turbulent or laminar flow is interesting and important in biology. For example, flow in the circulatory system is typically turbulent (unless we are discussing the capillaries). Areas of high turbulence are known to accumulate plaques more frequently. It is calculated as:

$$R_e = \frac{\rho L U}{\eta} \quad (9.3)$$

$$R_{e\omega} = \frac{\rho L^2 \omega}{\eta} \quad (9.4)$$

Where ρ describes the density of the fluid, L describes the characteristic length, U describes the flow speed, and μ describes the fluid's viscosity. It is common to see these variables interchanged with others, such as describing U as v , or η as μ , or others. Another common thing you may see is the replacement of ρ/η with a different constant called the *kinematic viscosity*, usually represented by ν .

The Reynolds number can be interpreted as a ratio of the inertial force to the viscous force. An R_e less than 10^3 is considered to be viscous dominated and laminar. An R_e greater than 10^4 is considered inertia dominated and turbulent. A more apt description may be that the viscous nature of the fluid holds significantly less weight for something at a high R_e —meaning that a blue whale's motion, which swims at an R_e of around 10^7 , is dictated nearly purely by inertia, as you would expect. In

²²<https://www.cambridge.org/us/universitypress/subjects/mathematics/fluid-dynamics-and-solid-mechanics/fluid-dynamics-cell-motility?format=HB&isbn=9781107174658>

²³<https://pubs.aip.org/aapt/ajp/article-abstract/45/1/3/1043148/Life-at-low-Reynolds-number?redirectedFrom=fulltext>

microbiology, R_e is $\ll 1$, making it highly viscosity dependent.

A nice equation you can find is the following:

$$\frac{\mu^2}{\rho} = F_{tow} \quad (9.5)$$

F_{tow} is the force required to “tow” an object in some liquid when $R_e = 1$. That is, any object with $R_e = 1$ can be towed by this force. The purpose of this is to demonstrate that small R_e is correlated to small forces. Another interpretation of R_e is that if you have an object, like a micro organism, with a very small R_e , you could push it with all your might, and nearly the instant you stop pushing, it will immediately halt. This is, again, because it is completely inertia-independent.

Swimming itself is a cyclic process, usually described as n cycles. Reciprocal swimming is when one’s motion is equal in either the forward or reverse direction, and where one ends up in the same state at the end of the cycle. Interestingly, at high R_e this cycle of swimming is acceptable. The example described in *Life at Low R_e* is a clam who opens its mouth to move and then closes it, shunting it forward and is carried by momentum. Of course, such momentum does not help at low R_e , so a swimmer must have multiple, non-reciprocal movements.

Lizard Digression.

A neat question you can ask is how exactly a large animal, like a lizard, are able to run on water? To do this, we’ll make a number of huge simplifications. First, let’s say the lizard’s foot deflects / indents the water in an approximately cube-like manner with a length R . The density of the fluid is what we’ll call ρ_f . The speed that they’re running we’ll call v , and the frequency of their steps to be ω . The force generated by stepping, also called the Shoke force, is due to the inertia of the fluid, and is summarized as:

$$F_{up} \approx R^3 \rho_f v \omega \quad (9.6)$$

The lizard is fighting the force of gravity pulling it down. If we take the lizard to be approximately cube like, with dimension L and density ρ_l , then:

$$\begin{aligned} F_{down} &\approx L^3 \rho_l g \\ R^3 \rho_f v \omega &\geq L^3 \rho_l g \end{aligned} \quad (9.7)$$

To simplify further, let’s assume $R \approx L$, and that $\omega \approx v/L$. Now we can simplify to:

$$\begin{aligned} \rho_f \frac{v^2}{L} &\geq \rho_l g \\ v &\geq \left(\frac{\rho_l}{\rho_f} g L \right)^{\frac{1}{2}} \end{aligned} \quad (9.8)$$

For the sake of not thinking too hard, you can intuit that body’s have physiological limits. Perhaps theoretically, an elephant can run on water if it moves its legs fast enough. However, you can probably

see that this isn't feasible; muscles have limits. Let's say the physiological constraint conforms to the following equation, where K is a constant describing muscle output:

$$\begin{aligned} Fv &\leq \rho_l L^3 K \\ R^3 \rho_f v^2 \omega &\leq \rho_l L^3 K \\ v &\leq \left(\frac{\rho_l}{\rho_f} L K \right)^{\frac{1}{3}} \end{aligned} \tag{9.9}$$

Therefore, the result you'll see is that to run on water, you are constrained between equation (9.8) and equation (9.9). At some point the result of (9.8) exceeds that of (9.9), making it so the burden of running on water exceeds physiological limits. This example wasn't explained in significant depth, and we made some grand simplifications, but the central theme is the ability to map the real, moving world onto equations and make some statements about what is possible.

9.3.1 Conservation of Mass

Let's discuss a fluid at point x and time t . Its velocity is $\mathbf{u}(x; t)$, and its density $\rho(x; t)$. If we are interested in some space, ω , we can determine the mass of fluid in this space by:

$$\int_{\omega} \rho(x; t) d\omega \tag{9.10}$$

If we are interested in the rate of change of fluid in our space ω , we can determine it by:

$$\int_{\omega} \frac{\partial}{\partial t} \rho(x; t) d\omega \tag{9.11}$$

Knowing that any change in mass of the fluid is due only to fluid flowing in and out of our space ω , we can rewrite this integral in terms of fluid passing the space's boundaries, which we'll call S , and let's define a vector normal to this space as \vec{n} . In other terms:

$$-\int_{S, \partial\omega} \rho \mathbf{u} \cdot \vec{n} dS = \int_{\omega} \nabla x \cdot [\rho \mathbf{u}] d\omega \tag{9.12}$$

The conversion between the two is a result of Gauss' theorem. This also means that, because ω is arbitrary:

$$\frac{\partial}{\partial t} \rho(x; t) = -\nabla x \cdot [\rho(x; t) \mathbf{u}(x; t)] \tag{9.13}$$

This is like saying the total change in density is equal to the change in density and velocity at all points x . Which we can rewrite in a way that is clearer, demonstrating mass is conserved:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0 \tag{9.14}$$

Where $\partial \rho / \partial t$ is the change in density and $\nabla \cdot (\rho \mathbf{u})$ is the advection term. For an incompressible fluid, where your density is fixed, you are left with:

$$\nabla \cdot \mathbf{u} = 0 \quad (9.15)$$

This, of course, is the first equation in the Navier-Stokes equations, discussed in section **9.3.4**.

9.3.2 Stress Tensor

Conservation of momentum can be solved using the forces applied to the system. This includes body forces ($\mathbf{F}(x; t)\delta V$ for some volume δV) and forces applied to a segment of the surface ($\tau(x; t)\delta S$ for some surface $\vec{n}\delta S$).

Body forces scale with $\mathcal{O}(\mathbf{F}L^3)$, and the canonical example is gravity. Forces applied to the surface include pressure, which scales with $\mathcal{O}(\mathbf{F}L^2)$. The force of friction (or viscosity), just as the friction of two solids, comes in two components: the normal force and the force which opposes movement laterally.

To think about this, consider some arbitrary object with a face. Regardless of the face, it has some normal vector \vec{n} , and let's say this surface has an area A . As mentioned before, body forces scale with $\mathcal{O}(\mathbf{F}L^3)$, while surface with $\mathcal{O}(\mathbf{F}L^2)$. This means that when $\lim_{V \rightarrow 0}$, body forces and surface forces must be equal. So let's think of τ as:

$$\tau(x; t) = \lim_{A \rightarrow 0} \frac{F}{A} \quad (9.16)$$

Let's first think of a cube whose faces are aligned with basis vectors $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$. All traction vectors can be summarized using these components such that if we have vector τ acting in the direction e_3 , we can describe it as:

$$\begin{aligned} \tau_{e_3} &= \sigma_{3,1}e_1 + \sigma_{3,2}e_2 + \sigma_{3,3}e_3 \\ \tau_i &= \sigma_{i,j}e_j \end{aligned} \quad (9.17)$$

This works nicely and succinctly when we are discussing a vector τ in the direction of a basis vector. Cauchy describes a more general case for some arbitrary tetrahedron. Forces applied to three of the faces on the tetrahedron can be described using $\{\sigma_{i,1}, \sigma_{i,2}, \sigma_{i,3}\}$, but the fourth face is misaligned with the bases. The 4th face of the tetrahedron has the normal vector \vec{n} , which can be defined as $\vec{n} = \vec{n}_1e_1 + \vec{n}_2e_2 + \vec{n}_3e_3$. A force acting on this face, τ is not aligned with any basis vector, but we can cheat by splitting τ into its components, $\tau = \tau_{n_1} + \tau_{n_2} + \tau_{n_3}$.

I'm going to yada-yada through the proof a little bit, but the gist is that if you take Newton's laws and describe them for each normal, you get something along the lines of:

$$\sum \mathbf{F}_{\mathbf{e}_1} = -\sigma_{1,1}n_1 - \sigma_{2,1}n_2 - \sigma_{3,1}n_3 + \tau_{n_1} + mg = ma \quad (9.18)$$

Since, as mentioned earlier, we are taking the limit as the volume goes to zero, we can say that:

$$\begin{aligned} -\sigma_{1,1}n_1 - \sigma_{2,1}n_2 - \sigma_{3,1}n_3 + \tau_{n_1} + 0 &= 0 \\ \sigma_{1,1}n_1 + \sigma_{2,1}n_2 + \sigma_{3,1}n_3 &= \tau_{n_1} \end{aligned} \quad (9.19)$$

And this holds true for all of the components of τ . This makes it more obvious that σ will be linear and homogenous. We can now write this as:

$$\begin{bmatrix} \tau_{n1} \\ \tau_{n2} \\ \tau_{n3} \end{bmatrix} = \begin{bmatrix} \sigma_{1,1} & \sigma_{2,1} & \sigma_{3,1} \\ \sigma_{1,2} & \sigma_{2,2} & \sigma_{3,2} \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_{3,3} \end{bmatrix} \begin{bmatrix} n1 \\ n2 \\ n3 \end{bmatrix} \quad (9.20)$$

Simplified, this is:

$$\tau_n = \sigma^T n \quad (9.21)$$

Where the transpose of σ multiplied by the normal gives us the force. So in short, to find the force acting on a plane you take some stress tensor and multiply it by the normal of that plane.

One component you may be interested in calculating is the force normal to the plane. Let's call this σ_n . This can be accomplished via a linear projection. From linear algebra, this is:

$$\begin{aligned} \sigma_n &= \text{proj}_n \tau_n = \frac{\tau_n \cdot n}{||n||} \\ \sigma_n &= \tau_n \cdot n \end{aligned} \quad (9.22)$$

Because the length of a unit vector, n , must equal 1. A shear stress, on the other hand, is parallel to the surface of the plane. For now, I'll call this π_n . Because the traction vector, τ_n , is the vector addition of σ_n and π_n . In other words, for either the vector or the scalar:

$$\begin{aligned} \vec{\pi}_n &= \tau_n - \sigma_n n \\ \pi_n &= ||\tau_n - \sigma_n n|| \end{aligned} \quad (9.23)$$

$\tau(x, t)dS$ describes the force exerted on a surface, called a traction. One component will be the normal force. The normal force will always point in the direction opposite of τ , which can be written as $\tau(x, t, -\vec{n}) = -\tau(x, t, \vec{n})$. To characterize this, you need to know the magnitude, direction, and orientation of the surface.

Defining $\tau(x, t)dS$ can be done with $\tau(\vec{n})dS = -\tau(e_1)dS - \tau(e_2)dS... = -[\tau(e_x)dS] \cdot \vec{n}$, where e_x is a surface. Somehow this leads to, via some linear projection, $dS_i = dS\vec{n} \cdot e_i$ and then $\tau = \sigma \cdot \vec{n}$. σ is your Cauchy stress tensor, in units of Pascals, which usually comes in the form of a 3×3 matrix.

This matrix, σ , is inherently symmetric.

9.3.3 Conservation of Momentum

To achieve the hydrodynamical equation of motion, we equate the rate of change of momentum within ω and out of the boundaries with the sum of forces acting on ω and the boundaries. The rate of change of momentum, in this case, is:

$$\int_{\omega} \frac{\partial}{\partial t} [\rho(x; t) \mathbf{u}(x; t)] d\omega \quad (9.24)$$

We can describe the outflow of momentum from the space ω by:

$$\int_{\S} \rho(x; t) \mathbf{u}(x; t) \mathbf{u}(x; t) \cdot dS = - \int_{\omega} \nabla x \cdot [\rho \mathbf{u} \mathbf{u}] d\omega \quad (9.25)$$

The sum of forces acting on the fluid within ω are loosely defined with X , where X is the force per unit volume from external forces:

$$\int_{\omega} X(x; t) d\omega \quad (9.26)$$

And the force force experienced, using σ as the stress tensor, is:

$$\int_S \sigma(x; t) \cdot dS = \int_{\omega} \nabla x \cdot \sigma d\omega \quad (9.27)$$

These are all combined and simplified into:

$$\frac{\partial}{\partial t} [\rho \mathbf{u}] + \nabla x \cdot [\rho \mathbf{u} \mathbf{u}] = X + \nabla x \cdot \sigma \quad (9.28)$$

We are allowed to make this simplification because ω is a nonexistent, arbitrary space.

$$\begin{aligned} \frac{\partial}{\partial t} \int_{\omega} \rho \mathbf{u} d\omega &= \int_{\omega} \mathbf{F} d\omega + \int_{\partial S} \sigma \cdot \vec{n} dS - \int_{\partial D} \rho \mathbf{u} (\mathbf{u} \cdot \vec{n}) dS \\ \rho \left[\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} \right] &= \rho g + \nabla \cdot \sigma \end{aligned} \quad (9.29)$$

At present, this cannot be solved because we do not know how σ behaves in the system. With an assumption, we can find the constitutive equation for σ :

$$\sigma = -pI + \mu(\nabla \mathbf{u} + {}^T \nabla \mathbf{u}) \quad (9.30)$$

Where $-pI$ is pressure multiplied by an identity matrix, giving us an isotropic term. The second term is viscosity, which was empirically determined by Newton. It was shown that the drag force experienced by an object scales with \mathbf{u} and A , but inversely with h . Thus, we get:

$$\frac{F}{A} = \mu \frac{\mathbf{u}}{h} \quad (9.31)$$

μ is a scaling factor, which is called the dynamic velocity and has units of Pascal·seconds.

9.3.4 Navier-Stokes Equations

In knowing the velocity vectors at all points in a fluid allows you to describe the movement of the fluid over time. Three key assumptions in this are that the fluid be Newtonian (meaning applying a shear force does not affect it in any way), incompressible, and isothermal. This brings us to:

$$\begin{aligned}\nabla \cdot \mathbf{u} &= 0 \\ \rho \frac{\partial \mathbf{u}}{\partial t} &= -\nabla p + \mu \nabla^2 \mathbf{u} + \mathbf{F}\end{aligned}\tag{9.32}$$

$\nabla \cdot \mathbf{u} = 0$ simply states that the mass is conserved. When describing a 3 dimensional velocity, we'll consider (x, y, z) . Thus:

$$\nabla \cdot \mathbf{u} = \frac{\partial \mathbf{u}}{\partial x} + \frac{\partial \mathbf{u}}{\partial y} + \frac{\partial \mathbf{u}}{\partial z}\tag{9.33}$$

Taking the gradient of \mathbf{u} is used to solve the divergence of the field. A nonzero divergence would describe a source of fluid flowing into the system (or a sink).

The second equation is a rewriting of Newton's second law, $\sum F = ma$. Some would describe it as being a conservation of momentum equation. Let's see how:

$$\begin{aligned}F &= ma/V \\ F &= \rho a \\ F &= \rho \frac{d\mathbf{u}}{dt}\end{aligned}\tag{9.34}$$

Because we are considering individual points, rather than the ensemble, we divide by the volume V , giving us the density. Acceleration can be rewritten as the change in velocity over time, $d\mathbf{u}/dt$. Next we will consider the forces relevant to this system.

$$-\nabla p + \mu \nabla^2 \mathbf{u} + \sum F = \rho \frac{d\mathbf{u}}{dt}\tag{9.35}$$

The ∇p describes one of two internal forces. It accounts for forces due to a change in pressure (recall that $p = F/A$). Again, in a 3D space, $\nabla p = \partial p/\partial x + \dots$, etc. The second term, $\mu \nabla^2 \mathbf{u}$, describes force of friction, and or viscosity.

Lastly, we can summarize all external forces ($\sum F$) within this big \mathbf{F} . We may see gravity as an external force, for example, in which case we may replace \mathbf{F} with ρg . Therefore, the Navier-Stokes equations are essentially the reapplication of fundamental concepts in physics for fluids.

\mathbf{u} we can define, perhaps overly complexly, as:

$$\mathbf{u}(r_s t) = \dot{r}_s + \mathbf{U}(t) + \Omega(t) \times r_s\tag{9.36}$$

Where r_s is defined as any point on the swimmer, and $\dot{r}_s = \partial r_s/\partial t$. $\mathbf{U}(t)$ describes the instantaneous linear velocity, and $\Omega(t)$ the angular.

As it goes, this equation can be made dimensionless as:

$$-\nabla p + \mu \nabla^2 \mathbf{u} + R_e \sum F = R_{ew} \frac{d\mathbf{u}}{dt} \quad (9.37)$$

Given that both R_e and R_{ew} are expected to be tremendously small for micro organisms, we can reduce the equations to the simplified Stokes equations. Importantly, this is simply the same as **removing the inertial terms**:

$$\begin{aligned} \nabla \cdot \mathbf{u} &= 0 \\ \nabla p &= \mu \nabla^2 \mathbf{u} \end{aligned} \quad (9.38)$$

Another slight confusion is that this entire equation seems to be zero. Maybe I am simply bad at math, but if $\nabla \cdot \mathbf{u} = 0$, then how does $\mu \nabla^2 \mathbf{u} \neq 0$?

9.3.5 Propulsion

Presently, I do not understand the theory behind this, but somehow everything that occurs at low R_e is linear. As it goes, force and torque are related by the following equations:

$$\begin{aligned} F &= AU + B\Omega \\ \tau &= CU + D\Omega \end{aligned} \quad (9.39)$$

We can use this to construct *propulsion* matrices like:

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \quad (9.40)$$

The proof is not shown here but interestingly $B = C$. To think of this matrix, let's consider dropping a cork-screw into a vat of syrup. You probably expect that the cork-screw will turn to some degree, but that there will still be some *slippage*. That is, it sinks at a rate faster than it spins. If something were to sink without spinning at all, you can imagine that the matrix would be completely diagonal, because, given our equation 9.39, that would mean the linear force, F , and the rotational force, τ , are completely decoupled. The propulsion efficiency is defined as (or proportional to) these off diagonal elements squared (i.e., B^2)—evidently, though, this is considered unimportant in the grand scheme of interactors.

Diffusion.

Diffusion coefficients describe how quickly an object moves around in a fluid. The formula, where a is the object's radius, is:

$$D = \frac{k_B T}{6\pi\eta a} \quad (9.41)$$

You may see the denominator rewritten as f , representing the friction coefficient of the object.

Fascinatingly, at very low R_e your movement inherently takes your surrounding environment with you. This actually has quite large consequences, and is called the *added mass phenomenon*. In short, if you move and carry a load of liquid along with you, it is as if you are propelling a much greater mass. It's estimated that:

$$\begin{aligned}\tau_{stir} &= \frac{l}{v_{stir}} \\ \tau_{diff} &= \frac{l^2}{D}\end{aligned}\tag{9.42}$$

Where τ is the time it takes you to move an object some distance l by either stirring or diffusion, and v_{stir} is the velocity of stirring performed. The ratio of such, i.e. lv/D (or LU/D), is called the Péclet number (P_e). Maybe you can recall from earlier that R_e can be written as lv/ν (equivalently LU/ν), so you can see how the Reynolds number and the Péclet number describe systems in a similar way. Stirring is only helpful if $lv/D > 1$, which gives us an indication of how much you can accomplish either by stirring or diffusion. When $P_e \ll 1$, diffusion dominates. For micro swimmers, like bacteria, stirring is virtually useless, and they're better off simply waiting for items to diffuse toward them. Another fascinating realization is that bacteria do not move to reach more nutrients—as we just stated that this would be useless. Instead, they travel to areas of a higher density, where the probability of diffusion is greater²⁴.

²⁴Okay—indeed there is something immensely fascinating here. Can the same be true for growth cones? I suppose... yes!

Part III

Physiology

Preoccupied with a single leaf, you won't see the tree. Preoccupied with a single tree, you'll miss the entire forest.

– Vagabond by Takehiko Inoue

9.4 Perspective

It is very important to avoid taking a quantized approach to studying physiology. You'll notice neuroscience majors may know some superb nuances regarding the role of HCN channels in overexcitable neural disorders, like epilepsy. However, they may be totally unaware of their canonical role as pace-makers in the heart. So too may they be unaware of the rich history of modeling I_h currents by computational biologists, preferring to look only at a channel as it functions in neurons and is testable in a cell culture. Hence, one must not become preoccupied with a single leaf.

Another bit of wisdom comes from Yoshikawa's *Musashi*. There is a chapter in which vagabonds are tilling a marsh's untenable land. Each time a storm came, all work was reset as flooding returned the land to mud. One realized that they had been tilling the field in the shape of a square, unconsciously aiming for a uniform, symmetrical crop that is typical of any farm. In doing so, they intended to bend nature from its gruff state into something it was not. Nature, not allowing this, cyclically reclaimed the land. But, if they tilled the land in an oblong fashion, asymmetrical and with the natural bends of the marsh, along the path that water would flow in a storm—perhaps the fields would survive the storm. Rather than trying to oppose the universe, one must first know its way, and be guided by it.

Surgical intervention is most often an opposition to the way of nature. Indeed, the most clear evidence of this is in today's BSIs which attempt to escape endogenous circuits all together. Perhaps one should focus more on enhancing nature's way first, before opposing it. And to enhance nature's way, one must understand it fully.

Chapter 10

Biochemistry

To be able to interface with the electrophysiology of neurons, one must fully understand it.

10.1 Membranes, Ions, and Potentials

This will be a general overview of some of the electrophysiology of neurons. This should probably precede the math and models section, but here we are anyway.

Of course, the ion composition of cells varies a bit, but a good enough approximation is as follows:

$$\begin{aligned}[\text{Na}^+]_i &= 15\text{mM}, [\text{Na}^+]_e = 150\text{mM} \\ [\text{K}^+]_i &= 120\text{mM}, [\text{K}^+]_e = 5\text{mM} \\ [\text{Ca}^{2+}]_i &= 100\text{nM}, [\text{Ca}^{2+}]_e = 2.5\text{mM} \\ [\text{Cl}^-]_i &= 15\text{mM}, [\text{Cl}^-]_e = 120\text{mM}\end{aligned}$$

Where $[\text{X}^+]_i$ and $[\text{X}^+]_e$ refer to the intracellular and extracellular concentrations respectively. Building up a strong intuition for electrophysiology is quite important to approaching the field. Some of the absolutely key facts to recall include:

1. **The intracellular and extracellular spaces are electro-neutral.** If this weren't the case, cell membranes would be under constant force. This means the sum of positive and negative ions must be equal between the intracellular and extracellular spaces.
2. **$[\text{Na}^+]$ and $[\text{Cl}^-]$ are high outside, $[\text{K}^+]$ is high inside.** This will tell you which directions the ions flow during an action potential—when the gates open, Na^+ wants to enter, K^+ wants to exit²⁵. In other words, $[\text{Na}^+]_i < [\text{Na}^+]_e$, $[\text{K}^+]_i > [\text{K}^+]_e$.
3. **The concentration of ions is approximately static.** The amount of ions that actually flow in an action potential and or through ion channels is quite small relative to the total amount.
4. **Intracellular Ca^{2+} is extremely regulated, and its concentration is small (around 100nM).** Ca^{2+} is the ion responsible for most cellular processes, so its tight regulation is essential. In other words, $[\text{Ca}^{2+}]_i \ll [\text{Ca}^{2+}]_e$.

²⁵A slight nuance is that this does not preclude ions from flowing against their gradient. It can certainly still happen, but the net movement will be out if channels are open in either direction.

Expanding on the calcium regulation, much of the free $[\text{Ca}^{2+}]_i$ is stored within the endoplasmic reticulum (ER), whose concentration will be many orders of magnitude higher than the cytosol. This is important for cells like myocytes who require a large influx of calcium to contract. Rather than attaining all of this Ca^{2+} from the extracellular space, the cells can use Ca^{2+} -induced- Ca^{2+} -release in order to drive the opening of the ER.

Ramblings.

As a complete aside, it is hypothesized that our cells control ions in this way because we originate from the ocean. As the ocean is salty, cells learned how to pump out as much Na^+ as possible, and based their electrophysiology off of this potential. When evolution allowed us to exit the ocean, we maintained this machinery and keep sodium in the extracellular space.

Direction of Flow.

Knowing the general concentrations of ions is a very important way to intuit ion behavior under different conditions. However, the Nernst Potential offers a way to quantify it as:

$$E_X = \frac{RT}{zF} \ln \frac{[X]_e}{[X]_i} \quad (10.1)$$

Where z is the charge of the ion. This allows you to solve the electromotive force, E , for your ion of interest. Knowing the E values for your ions allows you to determine the overall membrane potential using the Goldman–Hodgkin–Katz flux equation:

$$v = \frac{g_{n_1}E_{n_1} + g_{n_2}E_{n_2} + \dots + g_{n_k}E_{n_k}}{g_{n_1} + g_{n_2} + \dots + g_{n_k}} \quad (10.2)$$

The GHK formula is essentially like taking a weighted average of each ion's (n_i) membrane potential, giving us the total membrane voltage. The conductance, g , is the opposite of resistance (covered in more depth at section 7.1.2). In this case, you can think of conductance as being a measure of the membrane's permeability to an ion. That is, a neuron with a lot of Na^+ channels will have a very high total sodium conductance. Conductance is measured in terms of its maximum, meaning that although sodium channels will open and close, the value g_{Na} refers to when all of them are open. When calculating membrane voltage, generally we only use sodium and potassium (though, the GHK formula technically requests using all charged molecules). The reason being because the conductance of all of the other ions is essentially negligible in comparison.

In most cases, the GHK equation is simplified to only include Na^+ and K^+ . The reason is because these two ions have the highest conductances, by far, making them hold the most weight. For example, even though cells may contain HCO_3^- , there aren't tons and tons of ion channels for it like there are for sodium and potassium—thus it does not contribute to v to nearly the same degree. If this concept is confusing, simply consider $V = IR$. If there is no current, there is no voltage. Therefore, we can use the equation below as a good enough approximation of the resting membrane potential:

$$v = \frac{g_{\text{Na}}E_{\text{Na}} + g_{\text{K}}E_{\text{K}}}{g_{\text{Na}} + g_{\text{K}}} \quad (10.3)$$

Now, let's assume a membrane is at rest around -55mV and that the sodium concentrations are what we listed above (150mM and 15mM). We can use the Nernst equation as follows:

$$E_{Na} = \frac{(8.314)(310.15)}{(96,485)(1)} \ln \frac{150}{15} \rightarrow 0.062V \quad (10.4)$$

Because 0.062V (62mV) > the resting -55mV, we can assume Na⁺ will (on average) flow into the cell if sodium channels open. However, if the cell were to massively depolarize all the way to 100mV, we would actually expect Na⁺ to leave the cell, despite its gradient seemingly telling you otherwise. This is, of course, because ions will be subject to both their chemical and electrical gradients.

10.1.1 Action Potentials

Overview.

When considering an action potential, likely you'll want to focus on the flux of Na⁺, K⁺, and Ca²⁺. The standard description is that some kind of stimuli causes an influx of positive ions into the neuron's dendrites or soma, which leads to opening of voltage gated sodium channels in the initial segment of the axon. Sodium flows inward, which causes this segment of the axon to greatly depolarize (causing it to shoot up from $\approx -55\text{mV}$ to $\approx +20\text{mV}$). When the membrane reaches these more positive voltages, voltage gated calcium channels will begin to open and sodium channels will begin to close. More slowly (i.e., delayed in respect to sodium channel opening) in the depolarization process, potassium channels will open which allow for the efflux of potassium ions, causing the membrane to repolarize. Similarly, the depolarization will cause, the next set of sodium channels further along the axon to open, causing the depolarization to propagate down the axon. After the action potential is complete, active transport can somewhat quickly restore the ionic composition to exactly what it was before.

A Common Misconception.

The concentration of ions really does not change during an action potential. It is only a negligibly small amount of ions that are required to change the membrane voltage of the neuron. So, for example, [Na⁺]_e really only goes from 120mM to 119.99...mM during an action potential. Therefore, the membrane potentials that you calculate here are roughly true at any point in an action potential. What changes is the membrane voltage of the neuron, causing channels to open and close, which is what leads to the characteristic peaking of an action potential.

This idea is quite important. Many assume that ions flowing in and out substantially impacts their concentrations, and that is what leads to the in and outflowing of ions. It is often said that ion channels always pass ions "with their concentration gradient." While "transporters" pass ions or molecules against their gradient. But this is not true! Firstly, you can imagine that if channels are unidirectional, there will always be some stochastic passage against a concentration gradient. But more explicitly, many channels solely exist to passively pass ions against this gradient. For example: there are many "inward rectifying potassium channels," meaning they pass potassium ions into the cell (a famous example being the Kir family of channels). Importantly, the concentration of K⁺ outside of the cell will *never* exceed the concentration inside the cell. Therefore, any and all inward pointing K⁺ channels will inherently be against the chemical concentration gradient.

You may be slightly puzzled about why such a small concentration of ions is able to generate an action potential. This is because the voltage generated is hyper-localized to the membrane. The lipid composition of the membrane is also very strictly controlled. The hydrophobic heads of the lipids on the outside monolayer of the membrane (thus, those facing the extracellular space) are most often neutral in charge—still polar, though. Those on the inner monolayer (the cytoplasmic side) have a higher proportion of negatively charged heads. This helps create a partial membrane voltage and is incredibly important for ion behavior, as well as how transmembrane proteins orient themselves in the membrane.

Let's motivate this a little bit: the capacitance of a cell (C_m) is around $1\mu\text{F}/\text{cm}^2$. If the radius of the cell is $10\mu\text{m}$. Given that the voltage swings by around 100mV during an action potential, we can determine that:

$$\begin{aligned}C_{total} &= C_m \times 4\pi r^2 \\q &= C_{total}\Delta V \\q &= 1.3 \times 10^{-12}\text{C}\end{aligned}\tag{10.5}$$

Dividing this number by Faraday's constant, F , gives you the moles of charge, m . You can divide the moles by the volume to approximate ΔC (i.e., the change in total charge):

$$\Delta C = \frac{m}{4\pi r^3/3} \rightarrow 3.1 \times 10^{-7}\text{mM}\tag{10.6}$$

Don't forget to convert the volume into units of liters.

$3.1 \times 10^{-7}\text{mM}$ is a very small number, many orders of magnitude below the concentrations of either sodium or potassium. Therefore, in theory, to generate an action potential you'd require a nearly 0% change in ion concentration. However, what this actually describes is the *net* movement. That is, if the concentration of sodium inside the cell increases by x when sodium channels open, the concentration of K^+ must decrease by no more than $x - 3.1 \times 10^{-7}\text{mM}$ to depolarize the cell by 100mV . Thus, by this logic, x can be nearly any value—and perhaps the change in sodium and potassium is actually quite large? To verify that the general idea of this exercise is true, I simulated an action potential using the Hodgkin-Huxley model to find some approximations of ion currents, and found that the sodium concentration changes by $\approx 0.002\text{mM}$. This is only a small fraction of a percentage of the total concentration, so the principle holds.

10.2 Ion Channels

This section will be filled in better throughout the semester, through more Biochem courses.

10.2.1 Voltage Gated Channels

Voltage Sensing.

You may ask yourself "how does an ion channel know when the voltage is positive or negative?" Structure informs us of almost every puzzle in biology, and this case is no different! Let us consider one potassium channel called KvAP. KvAP is hypothesized to be voltage gated due to the presence arginine side chains (as positively charged amino acid)²⁶. Because of the positive charge on arginine,

²⁶<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1253646/>

you can imagine that if the outer membrane voltage becomes especially negative, it will pull the arginines outward, almost like opening the lid of the channel.

Chapter 11

Muscles

A large part of this information comes from Dee Silverthorn's *Human Physiology* textbook—arguably the best textbook of all time.

11.1 Skeletal Muscle

11.1.1 A Cellular Level

Structure.

Skeletal muscles are composed of muscle fibers, which are composed of myofibrils. Muscles use a silly nomenclature, in which “sarco” is added to words. For example, the whole structure sits within the sarcoplasm (cytoplasm), fibers are intertwined with the the sarcoplasmic reticulum (equivalent to endoplasmic reticulum) and surrounded by the sarcolemma (equivalent to the cell membrane). An individual myofibril is made up of overlapping actin and myosin segments, held together, to some degree, by titin. The myosin heads are the canonical structure you imagine, which bud off of the end of a chain like leaves on a branch and bind to actin. Undeniably, the most uninteresting part of muscle physiology is as follows: The “centerline” of the myosin networks is called the M line, while the “centerline” of the actin network is called a Z disk. Titin branches off from the Z disks to bind to the ends of myosin chains, providing both some elasticity and support for myosin. Actin chains are centered around a line of nebulin, which too provides structural support and organization. The I band is considered to be the unbound part of the actin structure, centered around the Z disks. The H zone is the unbound part of the myosin structure, centered around the M line. The A band is the entire length of the myosin fibers, thereby encapsulating both the bound part of the actin-myosin complex and the H zone, meaning it too is centered around the M line. Therefore, when contraction occurs, the size of the H zone and I band decreases, while the size of the A band stays the same.

Contraction-relaxation.

Myosin heads desire to bind to actin, but are blocked by tropomyosin. Ca^{2+} can bind to troponin, bound to tropomyosin, to cause conformational changes resulting in the exposure of the actin to the myosin heads. In doing so, binding can occur, followed by myosin's power stroke. The energy for the stroke comes from hydrolyzing ATP, which had already occurred by the time myosin attached to the actin. The powerstroke allows the release of the ADP and P_i . ATP then can bind to the empty active site of myosin, which causes the release of the head and prepares it for another cycle. Notably, when ATP does not bind to myosin, the muscle will be stuck in the rigor state.

In skeletal muscles, the source of Ca^{2+} is a combination of the extracellular Ca^{2+} flowing inward, and further release from the sarcoplasmic reticulum. The story goes as follows: a motor neuron releases acetylcholine onto the motor end plate (an area which has a high density of sodium channels). This causes a depolarization, which propagates down the muscle fiber. Structures called T-tubules sink lower into the tissue, allowing for more direct access to the inner processes (visually, these look similar to gyri in the brain or the crypts of the intestinal wall). The T-tubules are lined with dihydropyridine (DHP) channels, an L-type VGCC (specifically $\text{Ca}_v1.1$). DHP and ryanodine receptors (RyR) can be mechanically coupled, which influx of Ca^{2+} in through DHP mechanically opens RyR channels of the sarcoplasmic reticulum (a large store of Ca^{2+}). The free calcium is lowered through things like active pumping out of the sarcoplasm.

A steady supply of ATP is needed to maintain pumping, and it is said that at any given time, there is 8 or so twitches worth of ATP within the muscle fiber. Therefore, frequent production and alternate stores are required for continuous movement. One such storage is phosphocreatine, whose phosphate group can be quickly transferred to ADP through creatine kinase. Muscles therefore contain high levels of this enzyme, and **testing for it in the bloodstream can be a good proxy for muscle damage**.

Notably, it is very difficult to fully deplete a muscle of its ATP. Other forms of fatigue begin before this can possibly occur, which include CNS or PNS feedback. An example of this may be that acetylcholine is not synthesized fast enough to continually stimulate muscle fibers. Continual stimulation of muscle fibers, beyond what is allowable under normal conditions, **will fully deplete ATP levels and therefore cause damage to muscles**. Another consideration is the continuous use of ATP may result in P_i buildup in the sarcoplasm, making release of $\text{ADP} + \text{P}_i$ from myosin less likely to occur. Too, this opens the possibility of calcium phosphate forming, which can be quite damaging if it crystallizes further. There are also ion concentration changes to consider, and continued stimulation can result in tetanus.

11.1.2 Disorders Digression

There are many ways one can lose control of their muscles. Nerve damage, for example, will halt the release of acetylcholine onto the motor plate. In a similar manner, botulism is a result of the botulinum toxin blocking release of acetylcholine, resembling the effect of nerve damage. In McArdle's disease, muscles simply cannot convert glycogen to glucose-6-phosphate, causing the energy supply to be limited.

Muscular Dystrophy.

Muscular dystrophy is an umbrella term²⁷. There are more than 30 types, and in modern times the lines between muscular dystrophy and other disorders is more blurred. Duchenne muscular dystrophy (DMD) is the canonical type, which presents in children. Early signs include weakness in the shoulders or hips—i.e., the larger muscles. In DMD, the protein dystrophin is absent, which would normally attach actin to the cell membrane. This can result in membrane permeability, calcium influx, and thus activation of digestive enzymes that breakdown muscle fibers.

Treatment of muscular dystrophy is centered on symptom management. For DMD, steroid usage

²⁷https://www.youtube.com/watch?v=S6gPsYVmIEI&ab_channel=MayoClinic

seems to help. It is thought that muscle breakdown leads to immune responses, which can be aided by steroids. But, there is no cure, and eventually these steroids become less effective.

Multiple Sclerosis.

Multiple sclerosis (MS) is a CNS disorder, thought to be autoimmune in origin, and resulting in degradation of the myelin. It can be identified by plaques, and or, scar-looking lesions in the brain.

MS is rapidly developing. Over the course of days or weeks, one might slowly find numbness in a limb. Some light, tingling feeling is a typical first symptom. A clinician would follow-up with an MRI of the brain and spinal cord, looking for lesions within the myelin. The scarred regions are characterized in 4 ways: (1) an *active lesion*, which is acute and marked by inflammation and continual de- and remyelination, (2) an *inactive lesion*, which is chronic and marked by demyelination and limited inflammation, (3) a *mixed lesion*, where the center is primarily demyelinated, but the borders include microglia aiding in remyelination, and (4) a *remyelinated lesion* where a significant degree of healing has occurred.

The furthest downstream causes are classified as (1) a reduction in myelin during the active lesion, (2) total oligodendrocyte loss during the mixed and inactive lesions, and (3) poor differentiation of progenitor cells into remyelinating oligodendrocytes.

11.2 Smooth Muscle

Smooth muscle is much more variable than skeletal, differing by location, contraction pattern, required inputs, and structure.

11.2.1 A Cellular Level

Structure.

Smooth muscle is not considered to have sarcomeres, despite it having the same basic structural components of skeletal muscles. Smooth muscle contains much more actin than does skeletal muscle, and notably does not contain troponin like skeletal muscle does. Smooth muscle networks are connected through intermediate filaments, which usually attach to dense bodies within the cytoplasm. Actin also attaches to dense bodies, maintaining the actin-myosin network within the cell as well. Smooth muscles do not have T-tubules like skeletal muscles. A comparable structure may be caveolae, which do indent into the membrane and seem spatially associated with the sarcoplasmic reticulum. The autonomic nervous system stimulates fibers through neurotransmitter release from varicosities, or bulbous stores of the chemical. The neurons may innervate the muscle fibers, allowing multiple muscle cells to be stimulated at once, or they may be release to a few fibers, which are connected through gap junctions and stimulate the nearby ones, causing a propagating wave to stimulate others. The first case describes a multi-subunit muscle, and the second a single subunit.

Contraction-Relaxation.

Initiation of contraction begins the same as skeletal muscle, in that calcium enters and the concentration is further driven up by sarcoplasmic calcium release. Though, in this case Ca^{2+} may enter either through gap junctions or membrane ion channels. As such, there are many more modes of entry than

in skeletal muscle. For example, voltage-gated Ca^{2+} channels may open, but there are also ligand gated channels or stretch-activated channels, adding extra layers of possible regulation. Intracellular differences arise beginning from release from the SR. Firstly, is is no longer a mechanically gated RyR channel which allows its release. The release mechanism is now Ca^{2+} activated RyR release (commonly called Ca^{2+} -induced- Ca^{2+} -release (CICR)), and the IP_3 path. GPCRs activate phospholipase C, driving IP_3 production which binds to SR channels and causes them to open. The IP_3 path is usually considered the greatest way to drive up intracellular Ca^{2+} (or at least, that is what computational biologists seem to think). When Ca^{2+} is available, it binds to calmodulin (CaM), which then binds to the myosin light chain kinase (MLCK). MLCK phosphorylates myosin to increase myosin ATPase activity. Importantly, once contraction occurs, it stays stiff until released by a different mechanism. Because after contraction, no work is being done in the stiff state, smooth muscle is able to stay contracted for long periods. This explains why sphincters in the body are able to stay closed all the time, while one's bicep fatigues after carrying groceries for just a little while. Relaxation begins when Ca^{2+} is either pumped out of the cell through a Ca^{2+} ATPase pump, or sodium transporter. This causes CaM to unbind, myosin light chain phosphatase (MLCP) to dephosphorylate myosin, and the myosin heads to release from actin. Interestingly, diacylglycerol (DAG), another product of the IP_3 path, inhibits MLCP and thereby enhances muscle contraction.

The calcium stored in the SR is maintained in a number of ways. One example being the protein STIM1 responding to lower Ca^{2+} levels within it, moving toward the cell membrane, and activating store operated Ca^{2+} channels, such as Orai1.

11.3 Cardiac Muscle

11.3.1 A Cellular Level

Structure.

Cardiac cells contain sarcomeres, but the fibers themselves are smaller and often mono-nucleated. Cell edges are called *intercalated disks*, held together by desmosomes and permeated by gap junctions. The role of gap junctions is obvious but essential, and key to the rhythmic firing of the heart. Despite not requiring nervous system input, cardiac muscles have large T-tubules, and a significant amount of mitochondria.

Contraction-Relaxation.

Initiation of cardiac muscle contraction requires extracellular calcium, and generation can be graded. The entrance of Ca^{2+} powers the opening of non-mechanical RyR channels, as a method of CICR. Because of this, the sarcoplasmic reticulum is smaller in the myocardium, because some of the Ca^{2+} required for contraction comes from extracellular entry in about a 10:1 ratio of sarcoplasmic reticulum:extracellular origin. Ca^{2+} is removed using the Na^+ - Ca^{2+} exchanger (NCX).

The electrophysiological profile is marked by Na^+ entry but maintained by Ca^{2+} in a plateau phase. The fall occurs via K^+ entry. The refractory period is long due to this plateau phase, which prevents Na^+ channels from resetting.

Pacemakers.

Hyperpolarization cyclic nucleotide (HCN) gated channels induce an inward hyperpolarization (I_h , or I_f) current. Within the hyperpolarization range, HCN channels open and cause steady depolarization rhythmically. HCN channels are typically permeable to multiple cations, and in this case Na^+ is the biggest player. When the membrane potential is high enough, VGCCs will open and cause depolarization required to propagate the signal between myocytes. At peak, slow opening K^+ channels open and repolarize the membrane. Notably, voltage gated Na channels do not play a role in myocyte depolarization.

Expectedly, as β -blockers will inhibit a rise in cAMP, HCN channels will be less active. This decreases heart rate, among other things.

Chapter 12

Sensory Processing

Somatic senses are picked up by receptors on *primary sensory neurons*, whose cell bodies reside in the dorsal root ganglia and project to CNS interneurons. Such interneurons are considered *secondary sensory neurons* and their location depends on function. Fast traveling senses, like proprioception, are sent to the medulla before reaching interneurons (cross the body's midline in the brain), while slower traveling senses, like temperature, synapse onto interneurons upon entering the spinal cord (immediately crossing the midline). Neurons of the thalamus are considered *tertiary sensory neurons* and project to the somatosensory cortex.

12.1 Various Senses

12.1.1 Touch

12.1.2 Temperature

12.1.3 Pain

Nociceptive neurons respond to a variety of stimuli that can cause damage to cells. Afferent pain signals are carried either by myelinated $A\delta$ and unmyelinated C fibers. $A\delta$ carried pain is faster and perceived as sharp, local pain. C fiber carried pain is less sharp and less localized.

Transient receptor potential (TRP) channels are the most canonical class of ion channels, and are responsible for a great deal of pain sensation. Vanilloid ($TRPV_1$) channels respond to heat, while $TRPM8$ respond to cold. An immediate response to cell damage is lysing, causing marked increase in K^+ in the surrounding areas. Other released chemicals include histamine and prostaglandins. Upon activation, nociceptive neurons can trigger two responses. The first being reflexes that are integrated at the spinal cord, and the second being somatosensory cortex integration²⁸.

²⁸This ended up being much less interesting to write than I thought it would be.

Chapter 13

The Central Nervous System

We will primarily discuss the spinal cord.

13.1 Spinal Cord General Structure

A cross section of the spinal cord would reveal meninges just like the brain and skull; an outer layer of dura, then arachnoid, and finally pia mater. Too, there is an outer layer of white matter followed by a grey matter interior, which is centered around a “central canal.” The spinal cord has 33 vertebrae, beginning with cervical (C₁ to C₇), then thoracic (T₁ to T₁₂), then lumbar (L₁ to L₅), then sacral (S₁ to S₅), and finally one coccyx (Co₁ to Co₄). The sacral, and especially the coccygeal, vertebrae are fused, so you may see them depicted as a single unit. Adding slightly to the confusion is that there are only 31 pairs of spinal nerves. These are C₁ - C₈, T₁ - T₁₂, L₁ - L₅, S₁ - S₅, and Co₁. **Note: The difference in the nerve location and vertebrae labels is actually quite important, as it is not always that a fracture at some vertebrae leads to nerve damage at the same site**—this is expanded on in later chapters. The spinal nerves exit the spinal cord on either side, i.e., either dorsal or ventral roots. Dorsal entry neurons carry sensory information to the CNS, while ventral exit zones carry information from the CNS to the muscles. Though, importantly, their axons can not necessarily be found on either side, this will depend on where crossing over occurs. The section of grey matter which connects to the dorsal root is called the dorsal horn, and the same is true for the ventral horn. The lateral horn is in between the dorsal and ventral horns.

Note to self: It might be fun to add a more detailed breakdown at some point, such as the general arrangement of all the nuclei in the spinal cord.

Another note to self: It would be good to integrate more clinical outcomes into this. For example, what happens in the case of SCI at various vertebrae, or in vagotomy, etc.

13.2 Autonomic Nervous System

Pathways of the autonomic nervous system require two neurons, one that originates in the CNS and terminates at a ganglion, and a post-ganglionic neuron which terminates at the tissue of interest. While the circuit is often considered to have only two neurons, in fact many pre-ganglionic neurons synapse onto many post-ganglionic neurons, which means one neuron can affect many target tissues. Adding to this is the mode of release. The connection between neuron and target tissue is called the *neuroeffector junction*. At this junction, transmitters are secrete indirectly from bulbous varicosities

into the interstitial space, meaning neurotransmitters can diffuse over a larger area. Fascinatingly, these neurotransmitters are actually often synthesized within the axon / varicosities of neurons.

13.2.1 Sympathetic Nervous System

Pre-ganglionic neurons of the sympathetic nervous system originate in the hypothalamus or reticular formation and extend down to the mid-sections of the spinal cord, from neuron pairs T₁ to L₂. Sympathetic ganglia are directly beside the spinal cord in a long chain (often called the sympathetic chain). Generally speaking, neurons closer to the T₁ section correspond to organs higher in the body, such as the heart or lungs, while neurons closer to the L₂ section corresponds to those lower in the body, like one's reproductive organs²⁹.

Pre-ganglionic neurons release acetylcholine onto nicotinic receptors, and post-ganglionic neurons release norepinephrine onto adrenergic receptors. As all adrenergic receptors are GPCRs, their cellular response will be slower than ion channel transmission and occur at the protein-level primarily. Too, they will occur through secondary messenger, for example through the IP₃ path or cAMP. Norepinephrine is a tyrosine derivative and, like other catecholamines, can be broken down either by monoamine oxidase (MOA) in the neuron's mitochondria, requiring re-uptake, or in the liver by catechol-o-methyltransferase (COMT).

Dysregulation of the sympathetic nervous system most often manifests as cardiovascular pathologies, such as low blood pressure. Though, incontinence or impotence may also occur. On a cellular level, a process called *denervation hypersensitivity* may also occur, where in response to decreased sympathetic neurotransmitters, the body increases expression adrenergic receptors. This is easily measured as increased response from small doses of norepinephrine.

Adrenal Medulla Digression.

It's worth noting that the adrenal medulla secretes epinephrine in a hormone-like manner. However, the adrenal medulla is often considered a collection of pre-ganglionic neurons, meaning epinephrine acts globally like a hormone, but is also a neurotransmitter.

Let us look at the different adrenergic receptors³⁰:

Adrenergic Receptors			
Receptor	Location	Sensitivity	Effect
α_1	Most tissues	NE > E	Increase Ca ²⁺
α_2	Gut and pancreas	NE > E	Decrease cAMP
β_1	Heart and kidney	NE = E	Increase cAMP
β_2	Select smooth muscle	NE < E	Increase cAMP
β_3	Adipose	NE > E	Increase cAMP

²⁹Memorizing which nerves respond to which segments is likely not a good use of your time at this stage..., although it might have some clinical use to know which organs may be affected after SCI.

³⁰This is ripped directly from Silverthorn.

13.2.2 Parasympathetic Nervous System

Parasympathetic neurons originate from the uppermost and lowermost neurons, either leaving directly from the cranial nerves or from the S_2 to S_4 nerves. Ganglia in the parasympathetic nervous system are close to the target tissue, meaning post-ganglionic neurons are very short compared to the pre-ganglionic ones. One of the most paths being from the vagus nerve (a cranial nerve) that carries the majority of parasympathetic signals to organs including the heart, lungs, liver, stomach, intestines, and pancreas. Like the sympathetic nervous system, pre-ganglionic neurons release acetylcholine onto nicotinic receptors. Notably, parasympathetic post-ganglionic neurons also secrete acetylcholine, but onto muscarinic receptors instead.

Cholinergic Receptors		
Receptor	Location	Effect
N_N	Postganglionic neurons	Opens cation channels
N_M	Skeletal muscle	Opens cation channels
M_1, M_3, M_5	Target tissues	Increase Ca^{2+}
M_2, M_4	Target tissues	Decrease cAMP, open K^+ channels

Exceptions.

There are some exceptions to these rules, such as some sympathetic postganglionic neurons that terminate on sweat glands secrete norepinephrine rather than acetylcholine. These are called “sympathetic cholinergic neurons.” There are also neurons which secrete none of these, are called “nonadrenergic, noncholinergic neurons.” There are also neurons which secrete multiple types of neurotransmitters.

13.3 Somatic Nervous System

The somatic motor division is exclusively excitatory, and neurons project from the CNS all the way to muscles. Inhibition can occur at the neuronal level (i.e., motor neurons are inhibited) but not at the muscle level. Cell bodies of motor neurons are either within the ventral horn of the spinal cord or within the brain. Close to the muscle itself, the axons branch and connect to the muscle at the neuromuscular junction (NMJ), which includes both the target muscle and Schwann cell projections. The “synapse” equivalent is called the *motor end plate*, which is an area on the muscle with a high density of nicotinic receptors. The ECM contains acetylcholinesterase, responsible for breaking down acetylcholine. Notably, the nicotinic receptors on muscle cells are considered N_M , a slightly different version than is found on neurons (N_N). This difference has proven to be crucial in examples such as α -bungarotoxin, which binds only to N_M ³¹.

13.4 Acetylcholine, Nicotinic Receptors, and Muscarinic Receptors

Acetylcholine (which I’ll call ACh for this section) is unquestionably one of the most important molecules in the body. It is synthesized from choline (a hydroxyl azanium) and acetyl-CoA. Typical us-

³¹Dee Silverthorn quite possibly wrote the greatest textbook known to humanity, didn’t she?

age goes as follows: ACh is released from vesicles and binds to cholinergic receptors. Upon unbinding, acetylcholinesterase (AChE) breaks it down into acetate and choline. Choline is brought back into the pre-synaptic neuron through sodium cotransporters so that it may be reused. It is then re-combined with acetyl-CoA and repackaged and vesicles.

Funnily enough, nicotinic receptors (nAChRs) have been described as the most well understood membrane receptor³². Nicotinic receptors are fast opening non-specific cationic channels. They are expressed throughout much of the major structures in the brain, peripheral nervous system, and skeletal muscle. As such, drugs targeting these receptors are of considerable interest in many neurological disorders. The binding site for ACh is composed of aromatic amino acids, namely W and Y. This paper describes continual exposure to drugs treating nAChRs as causing their eventual desensitization, while exposure to nicotine increases their expression greatly.

Both of these are of key interest in treating Alzheimer's disease, as loss of cholinergic synapses is one of its features (including significant reduction of both muscarinic and nicotinic receptors). Interestingly, Alzheimer's shows varying loss of nicotinic receptor subtypes across the brain. There is no shortage of drugs currently in development for reversing this.

Myasthenia Gravis Digression.

Myasthenia Gravis, an autoimmune disorder, is an example of an acetylcholine dysregulation caused muscle disorder. As a digression within a digression, autoimmune disorders most often target endocrine organs, and the belief is that they are to protect against mutations causing hypersecretion³³. Upon first glance, you may think that Myasthenia Gravis evades this generality, but in fact it does not, as it is well associated with patients that dually have a thymoma (tumor of the thymus). Immune cells begin attacking cells with ACh receptors, tagging them with antibodies. Therefore, Myasthenia Gravis presents itself usually in the weaker muscles, such as those that control the eyes, as they have less ACh receptors and therefore a diminished response to signals. Thus, patients may have drooped eyelids (ptosis), double vision (diplopia), or trouble following moving objects with their eyes. The disease worsens with increased activity, as this causes more antibodies to be released onto ACh receptors—but improves on rest. Fascinatingly, men and women “get” the disease at different times (women typically under 40, while men over 60).

Antibodies may also target other genes (such as MUSK or LRP4) which are involved in ACh receptor localization, or other forms of regulation.

13.5 Cerebrospinal Fluid

13.5.1 CSF Contacting Cells

This topic may appear niche, but it is indeed of considerable interest! To start, epidural electrical stimulation (EES), which is covered in considerable detail later, does not stimulate specific neurons but rather depolarizes the entire CSF. Therefore, the cells in direct contact with the CSF may be the

³²<https://www.nature.com/articles/nrd2927>

³³[https://www.cell.com/immunity/pdf/S1074-7613\(20\)30180-1.pdf](https://www.cell.com/immunity/pdf/S1074-7613(20)30180-1.pdf)

key modulators in restoration and regeneration.

To begin, let us discuss this paper³⁴. CSF contacting neurons (CSF-cNs) are described as being chemosensitive, mechanosensitive, and may even act as stem cells. As mechanosensors, they inhibit motor neurons (i.e., in normal walking they are used to prevent over-compression of the spinal cord). Canonically, they are ciliated and function in tandem with the Reissner fiber to sense spinal curvature. The Reissner fiber is a protein fiber that extends along the spinal cord, and its breakdown is a known cause of scoliosis in model organisms³⁵. CSF-cNs cells modulate excitatory interneurons and reticulospinal neurons. Because of their location, CSF-cNs cells are very difficult to study. They are seated in the ependymal layer of the central canal. The Review cited above discusses their investigation primarily in zebrafish. An important advancement in their study was in discovering the marker polycystic kidney disease 2-like 1 (PDK2L1), which labels all CSF-cNs.

Given their hair-like structures that extend into the apical side of the central canal, their discoverers compared them to hair cells of the ear (though, these hairs are disordered, unlike inner or outer cells). Interestingly, though, these cells are GABAergic, and thus inhibit rather than excite. In spinal curvature tests, to the surprise of the experimenters, only those that were on the concaved side of the spinal cord responded, meaning they response to compression. Restated, those cells on the left side of the body respond to bending of the spinal cord toward the left side.

Remarkably: PDK2L1 is an ion channel with a quite high conductance (hundreds of pS) and the high membrane resistance of CSF-cNs neurons (on the order of GΩ)... the opening of a single PDK2L1 is able to depolarize a cell! It was demonstrated that mechanical force opens PDK2L1 channels, and that in the absence of such channels, CSF-cNs do not respond. Another responder to mechanical force seems to be acid-sensing ion channels (ASICs), which is indirectly affected when an ASICs modulator is altered by force. An unanswered question was what PDK2L1 is actually responding to, as the response did not correlate with the calculated CSF flow. Instead, evidence suggests it is coupled to the Reissner fiber. Structural analysis revealed vesicles on both the apical and basolateral sides, suggesting they can secrete both neurotransmitters, namely GABA, onto neurons or peptides into the CSF. Notably, after SCI it is said that CSF-cNs secrete monoamines. While neurotransmitter containing vesicles have been shown to release after a single depolarization, the root of apically released vesicles is not known.

CSF-cNs-connected neurons ascend ipsilaterally a couple hundred microns. Those that are particularly rostral may connect directly to the midbrain. Driving CSF-cNs also modulate activation of the central pattern generator, suggesting neural connection there as well. Their connection to V2a spinal interneurons is not especially conclusive yet, but is quite a promising possibility. Similar connections have been shown with reticulospinal neurons and propriospinal neurons.

Fascinatingly, despite CSF-cNs being inhibitory, optogenetic stimulation of them can lead to locomotion in zebrafish—indicating they have a multiplexed role. This data seems quite contradictory and unclear, as contrastingly, somatostatin expressing cells (specifically Sst1.1) are said to synapse onto excitatory interneurons and inhibit them. Sst knockout zebrafish exhibited longer durations of spontaneous locomotion. Their ability to promote locomotion seems much less supported than their ability to inhibit it.

³⁴<https://www.nature.com/articles/s41583-023-00723-8>

³⁵Very interestingly, no studies have been done to show the existence or lack of a Reissner fiber in humans.

As a lasting note, CSF-cNs show characteristics of immature neurons, and indeed seem to be able to differentiate into astrocytes or oligodendrocytes. Similarly, they seem to be able to recover after injury, and this was shown in a SCI model with EES.

13.6 Myelin in Health and Regeneration

Here we'll discuss the role of microglia in regulating myelin³⁶. Microglia are the macrophages of the CNS, and are derived from erythromyeloid progenitor cells. During embryogenesis, they enter the CNS. Notably, they have a long lifespan in non-pathogenic environments. Microglia respond to CNS damage via receptors for damage-associated molecular patterns (DAMPs) or pathogen-associated molecular patterns (PAMPs) and proliferate. Interestingly, microglia phagocytose oligodendrocyte progenitor cells before they mature, preventing myelination during development which helps “organize” myelin wraps depending on neural activity. After injury, microglia are activated to clear debris and force an environment in which oligodendrocyte progenitors can migrate and re-wrap neurons. Perhaps more fascinatingly, upon aging, microglia lose the ability to prune myelin sheaths, resulting in hypermyelination that in turn becomes degenerative.

Discussing such matters in the context of degenerative diseases is some of the most helpful. For example, multiple sclerosis is touched on a bit in section 11.1.2, and Alzheimer's is discussed in depth in the Addendum (Part VII, which is supposed to be a secret). Of course, in MS the topic of myelin is core, but demyelination is a common feature of many degenerative diseases, including AD. Fascinatingly, changes in the myelin / white matter structures can be seen decades before AD symptoms arise. Such regions are called white matter hyperintensities (WMHs) and can be seen by MRI—and splotches in the parietal lobe are especially predictive of AD. Though, it is presently unclear if axonal changes precede such myelin changes.

13.6.1 Initiation of Demyelination

The factors leading to demyelination are largely unknown. In the case of MS, it is thought to be an immune response which triggers myelin degradation. However, it is also hypothesized that transcriptional profile alterations lead to oligodendrocyte instability, perhaps predisposing them to degradation. Similarly, hypermyelination precedes demyelination in many cases, and is thought to be triggered by an autophagy pathway³⁷. In a mouse AD model, preceding a period of degeneration was months of oligodendrocyte growth, in which myelin sheaths were thicker than wildtype neurons. Follow-up work showed both de- and remyelination occur throughout the pathology's progression, but in the end demyelination prevails. Inhibiting demyelination via promoting oligodendrogenesis through inhibition of oligodendrocyte muscarinic receptor 1s seem to improve the cognitive impairments associated with AD. If interested, muscarinic receptors are discussed throughout section 13.2. These authors suggest monitoring lipid profiles as an early diagnostic tool of AD.

³⁶<https://www.nature.com/articles/s41577-023-00907-4>

³⁷This is actually interesting. I wonder if this is linked to the broad endocrinology hypothesis regarding quickly growing cells

13.7 Central Pattern Generation

Defining the central pattern generator is still somewhat contested. Computational models have been used to explore its existence³⁸, and for now we should take a few unifying assumptions: **(1)** the central pattern generator exists and is used to generate rhythmic moving, such as walking, **(2)** in many cases it is initiated by the CNS, but aside from that is largely devoid from CNS input, and **(3)** the CNS can work in tandem to compute integrated moving that requires coordination and balance.

³⁸<https://www.nature.com/articles/s41598-021-91714-1>

Part IV

Spinal Cord Injury

Maybe it sounds like I'm just spouting moral platitudes. But from a vagabond like me, it's not that. I can't begin to tell you how lonely I feel when I come across a beautiful view, then suddenly realize there's no one to enjoy it with me.

– Musashi by Eiji Yoshikawa

Chapter 14

The Injury Itself

Spinal cord injury is composed of the primary injury, prototypically, but not restrictively, due to some kind of high impact action. This is usually unpredictable and highly variable. Secondary injury, resulting from inflammation, oxidative stress, and other biological responses is much more predictable and potentially lends itself better to therapeutic intervention.

The lesion's composition is categorized in three ways: **(1)** the non-neural core, **(2)** the astrocytic scar around the core, and **(3)** the spare reactive neural tissue. In the mix of immune cell influx and scar formation, no neural cells can survive at the center of the lesion. On a neuronal level, the rostral end retracts in a process of Wallerian degeneration. The caudal end dies away. Growth from the cell body is limited both by the damaged cell's biochemistry and by the physical barriers which now present themselves in front of the axons. The physical barriers that immediately succeed injury are often called damaged axon-glia complexes (AGCs). Discussed further later, immune cell influx causes astrocytes to form a scar, meant to save the spare surrounding neural tissue, which is composed of both glia and neurons.

The traditional aim of treatment is to bridge the corticospinal tract with distant neurons through a therapeutic combination of inhibiting anti-regenerative and promoting regenerative factors. As I have commented many times, getting neurons to regenerate alone is insufficient in many cases, as reformation of the correct synapses will not necessarily follow. Forcing axon regeneration alone is, incidentally, not too hard—one can pump neurons full of metabolites or simply implant stem cells. The issue being that they do not know where to grow to. One possible route to solving this is remodeling neural circuits using interneurons to bridge these connections. There are also attempts to use biomaterials to simulate a pro-regenerative environment, hopefully enhancing plasticity of the circuits.

14.1 Cell Specific Responses

The discussion, for the moment, will mostly use information gathered from³⁹.

14.1.1 Immune Response

As SCI breaks the spine-blood barrier, influx of immune cells can cause further damage. Evidently, the nature of the immune response being helpful or harmful is still largely contested.

³⁹<https://www.nature.com/articles/s41392-023-01477-6>

Neutrophils.

Neutrophils compose part of the immediate response to injury, which are recruited by cytokines and chemokines secreted by cells damaged in the primary injury. They essentially initiate the secondary injury, and reach their peak around 1 DPI. Like most cells, the role of neutrophils cannot be characterized as solely pro- or anti-regenerative. While a high influx of neutrophils is associated with poor patient outcomes, so too are neutrophils associated with guiding macrophages to damaged tissue, suggestive of better recovery.

Microglia.

In mouse SCI models, it seems that there are two peaks of microglial activity. The time course is remarkably long and disparate, reported 7 DPI and 60 DPI. Microglia can either promote inflammation, thereby worsening the secondary injury (called the M1 phenotype) or decrease inflammation, and promote repair (called M2). It is likely that this response depends on the subtype of microglia, which varies depending on the environment. Regardless, it is true that the earlier one treats SCI, the more likely one is to avoid negative microglial effects. Fascinatingly, in a neonatal setting microglia are able to heal SCI almost entirely through their role secreting fibrinogen, which is able to connect damaged axons back together.

Macrophages.

Macrophages are considered to be the dominant immune cell located around the injury site. Microglia, conversely, are scattered around the borders of the injury. Depending on the type of glial scar that is formed, different types of macrophages have been found. Macrophages mediate the corraling⁴⁰ of cells around the injury site. The phagocytotic abilities of macrophages are of key importance, as loose fragments of cells must be removed, and microglia are incapable of keeping up such a high demand for removal. As a large part of this includes the destroyed myelin of oligodendrocytes, macrophages uptake great amounts of lipids. This can result in the formation of lipid droplets, which causes macrophages to become “foamy.” This foamy phenotype impairs further repair.

Lymphocytes.

The adaptive immunity is fairly universally regarded as harmful to regeneration (with some exceptions, of course). T cells further break down the spine-blood barrier and increase immune cell invasion. Evidently, T cell entry is also a major source of neuropathic pain in SCI patients. In general, it seems established that the overall immune response after injury impairs further regeneration, and a good example of this can be found here⁴¹. **NOTE: It would be good to read this entire article.**

14.1.2 Neural Response

Apoptosis of neurons initiates at around 4 h AI, but peaks at only 8 h AI⁴².

⁴⁰Corraling is a term used to describe the formation of a barrier around the injury, preventing further injury. It is composed astrocytes, and other cells, and is important in repair.

⁴¹<https://www.science.org/doi/full/10.1126/science.abd5926>

⁴²It is worth noting that these times are likely quite inaccurate, or very injury-type specific. But, they do give a good indication of the approximate timeframe—such as that the majority of this apoptosis occurs within the first day or so after injury.

Interneurons.

Fascinatingly, it was shown that the ability of neonatal mice to fully recover from SCI was due, in part, to interneurons maintaining excitatory conditions. In adult mice, these interneurons switch to inhibitory after SCI, which dampens signals to motor neurons. A paper investigating this can be found here⁴³. One must wonder, can you electronically mimic the excitatory interneurons in fully grown mice? Similar approaches have been done therapeutically, such as with potassium-chloride cotransporter-2 (KCC2) agonist CLP290⁴⁴ which seems to dampen the overexcited, inhibitory interneurons. Perhaps one could simply use something like DBS (or in this case, DSS?) on these interneurons.

Astrocytes.

Astrocytes, being the dominant supportive cell, plays an essential role in SCI. After injury, astrocytes form a physical barrier that is supposedly intended to limit the secondary injury. The old perspective was that astrocytes, like the fibrotic scar, impair recovery as a physical barrier. It is now more accepted that the astrocytic scar (or border) is a necessary mechanism of limiting inflammation. This occurs after astrocytes become activated, and are helpful in the initial stages but later form a glial scar, impairing regeneration. Astrocytes may either be activated by inflammation, causing them to be neurotoxic (called A1 cells) or by ischemia, causing them to be neuroprotective (A2 cells). The first transformation occurring through the NF- κ B path, and the second through STAT3. Microglia are said to be the greatest contributor to activating astrocytes through release of signaling molecules. Another important factor is type 1 collagen upregulation, which results in astrocytic adhesion through cadherin, causing activation and eventual scar formation.

The astrocyte scar is surprisingly thin, only a few layers of cells. Though, its importance is not to be underestimated. When ablated, mice with SCI were worse off by almost every metric. A cornerstone paper on the topic seems to be here⁴⁵. While scar tissue is primarily astrocytic in origin, it is worth mentioning that pericyte derived scar tissue (sometimes called the fibrotic scar) too play a role. Their positive roles include boosting tissue integrity, but so too do they seem to block axon regeneration as a physical barrier.

Oligodendrocytes.

Oligodendrocytes reportedly begin apoptosis around 1 DPI and it peaks around 8 DPI. Oligodendrocyte precursor cells (OPCs) may differentiate into oligodendrocytes or Schwann cells after SCI. OPCs have been shown to remyelinate neurons after SCI, but fascinatingly, it has been shown that locomotor recovery after SCI does not necessarily require remyelination by oligodendrocytes⁴⁶. Though, plenty of other research suggests it is required—so it is likely context dependent.

⁴³<https://www.nature.com/articles/s41593-022-01067-9>

⁴⁴<https://www.sciencedirect.com/science/article/pii/S009286741830730X>

⁴⁵<https://www.nature.com/articles/nature17623>

⁴⁶<https://www.nature.com/articles/s41467-018-05473-1>

Chapter 15

In The Clinic & Therapeutic Approaches

15.1 Clinical Presentation

Death rates from SCI are still as high as 20% in some countries. There is a strong age dependence to this, as the probability of walking again after SCI in those older than 50 is much lower than those under.

Diagnosing Injury.

Immediate diagnosis is done through scanning, such as X-ray or CT scans, combined with general neurological exams, including voluntary or involuntary motor control tests. An alternative technique in neural evaluation, if the patient is not responsive, is electrophysiological recordings (either through EEG or EMG). X-rays are often used to immediately see large fractures, and follow-up CT scans for investigating the more possible hairline fractures. CT angiography may be used to investigate vascular destruction. More soft-tissue damage follow-up is done through MRIs. Interestingly, a trade-off exists in an MRI vs. immediate surgery. An MRI may be used to detect non-obvious issues, like disc herniations away from the primary injury site—which, if not fixed, will cause further degeneration. However, doing an MRI delays one's ability to decompress the spine.

Classifying Injury.

You will often hear injuries as being complete or incomplete. Another helpful distinction is discomplete, where the injury is considered clinically complete, but one can still observe connections through electrophysiology. The main method of classifying injury and tracking progress of patients is called the American Spinal Injury Association (ASIA) impairment scale. ASIA scores are broken into the following categories:

1. Grade A: Complete impairment, where there is no motor or sensory information being transmitted below the injury site⁴⁷.
2. Grade B: Incomplete impairment, where there is no motor information being transmitted, but some sensory information is preserved.
3. Grade C: There is some motor activity preserved, but more than half of the key muscles are too weak to move against gravity (Grade 3 muscles).

⁴⁷It is not clear to me if this is measured by EMG or movement.

4. Grade D: A fair amount of motor activity is preserved, where at least half of the key muscles are above muscle Grade 3.
5. Grade E: There is no impairment at all!

The most common and complete method of classification is the International Standards for Neurological Classification of Spinal Cord Injury (ISNCSCI). The ISNCSCI uses the ASIA impairment scale, with ASIA motor and sensory scores. It is recommended that assessment be performed immediately upon hospital admission⁴⁸, with follow-ups in the future to assess improvement. Another important metric is quality of life (QOL) assessment.

Importantly, the injury site is often designated by the vertebrae that was fractured, but symptoms are due to the nerve pair that is damaged, which may be at a different location than the primary site of bone damage⁴⁹. This discrepancy seems to be exacerbated the more caudal you go. Injury in the cervical portions can lead to severe bradycardia and hypotension, due to dysregulation of brain-heart communication, particularly regarding baroreceptor feedback. Too, damage to the vagus nerve can occur here, leading to dysregulation of most organs. Injury in the thoracic part may have widespread affects on the symathetic nervous system due to damage both of the spinal cord nerves and the nearby ganglia. Interestingly, an unconsidered byproduct of lower thoracic SCI is damage to motor signals to the legs. The main focus of such being that one loses their ability to walk, but accompanying this is reduced venous return, as veins rely on muscle movement to get blood back to the heart. Dampening of CNS-cardiovascular system communication seems to be one of the primary indicators of poor prognosis.

In evaluating the severity of the injury, a *spinal shock*, marked by temporary paralysis, may muddy the waters. While one may temporarily lose their reflexes, it can sometimes be regained soon later. However, the ability to define this state, and its duration, remains problematic. *Neurogenic shock* manifests similarly, but the cause is hypotension after SCI. This may be caused by hypovolaemia from blood loss, or pooling of blood due to reduced venous return. This occurs most often in SCI above T₆, as it is these sympathetic nervous which maintain vascular tone.

A few named pathologies exist, such as Central Cord Syndrome. This is the most common incomplete SCI. Often, this occurs in elderly patients who fall and already had some form of spondylosis⁵⁰. It is marked by more damage to upper extremities and possible incontinence. Brown-Séquard Syndrome occurs from penetrating SCI, such as a stab wound. It is usually characterized by sensory loss.

15.2 Auxiliary Management

Haemodynamics.

A steady blood supply to the damaged spinal cord is an essential component of treatment. Tool, system hypotension is a common symptom both from nerve damage and the surgery itself.

⁴⁸This is interesting, as in other sections I have detailed that various forms of shock can cause misunderstanding in the initial phases of hospital stay.

⁴⁹<https://www.nature.com/articles/nrdp201718>

⁵⁰Weathering of the vertebrae.

Methylprednisolone sodium succinate.

Evidently, this is controversial! Methylprednisolone sodium succinate (MPSS) has the traditional role of reducing inflammation. The controversy is due to mixed results. Some large clinical trials showed no benefit, while others suggest there was a considerable improvement in ASIA scores, particularly if administered within 8h of injury. The adverse side effects seemed to be minimal, so many surgeons routinely administered it. Follow-up investigations suggest that negative effects are more consistent than previously thought, so the updated guidelines are not to administer MPSS.

15.3 Treatments

This section will cover some of the basic, in a broad overview manner. Complex research, such as the discussion of stem cells and the research surrounding them, will be covered in chapter 20.

15.3.1 Electrical Stimulation

This will, of course, be covered in excruciating detail in later sections (and or, Parts, like Part V). Fascinatingly, electrical stimulation has been used in conjunction with physical therapy in the past with good results⁵¹. The reasons may be that this promotes stem cell differentiation⁵², or disrupts inhibitory interneuron signaling. The optimal electrical application for differentiation has been explored extensively⁵³. One may wonder if the benefits of BSI are in the interface itself, or simply the stimulation. Combining stem cell implantation and electronics is, likely, the future.

A slightly different rose by the same name is functional electrical stimulation (FES). Many trials have shown improvement in patients treated with either external stimulation or internal stimulation.

Light.

Light stimulation feels like a footnote in the electrical modulation story, to me. Though, if one wanted to control different neurons or enzymes on an alternate time course, optogenetic activation may be an option. The obvious issue being that one does not have genetic access to patients, and therefore would need to design (likely very complicated) targeted therapeutics.

Sound.

Another footnote is ultrasound simulation. In this case, it will be low intensity focused ultrasound. Some approaches have seen altered gene expression, but perhaps a more promising one is modulating mechanosensitive channels as was shown here⁵⁴. Notably, this paper found that many mechanically activated channels are affected in ultrasound, including Piezo, and many of the Trp family proteins.

Magnetics.

I would be extremely curious to know if the magnetic field itself has any unique properties beyond its manipulation of the electric field. Still, too, the story is the same. Some seem to enhance channel

⁵¹<https://www.nejm.org/doi/pdf/10.1056/NEJMoa1803588?articleTools=true>

⁵²<https://www.mdpi.com/2073-4409/11/5/846>

⁵³<https://www.frontiersin.org/articles/10.3389/fbioe.2021.591838/full>

⁵⁴<https://www.nature.com/articles/s41467-022-28040-1>

activation, while others expression. Interestingly, transcranial magnetic stimulation (TMS) has been used as a treatment with some success. Incomplete spinal cord injury has seen improvements from TMS.

15.3.2 Biomaterials

The overall goal in the use of biomaterials is to block a worsened immune response, scar formation, and promote neuron activity. Adding promise to stem cell implantation is the use of biomaterials that enhance proper network reformation⁵⁵. Theoretically, a perfect biomaterial could be a substrate preferable for neuron growth, contain molecules that inhibit the immune response, neurotrophic factors that enhance stem cell differentiation and recruitment, and ion channel agonists. Notably, to date there have been no major publications where a “cocktail” like this has been successful. These sorts of things are usually made from hydrogels, collagens, or select inorganic fibers.

An open question is how one could leverage biomaterials to help clear damaged parts of neurons/cells that would normally be cleared by phagocytosis. Perhaps, one could add materials that are easily oxygenated to dampen the blow of ROS.

To date, implantation of biomaterials have been relatively lackluster in treating patients. While some regeneration scaffolds have proven to improve some neurological function, no patient has regained motor function.

15.3.3 Drug Treatment

Drug treatment primarily follows the same paths, being reduction of inflammation and neuroprotection. Methylprednisolone (MP) is the only drug approved to treat SCI and works through reducing inflammation. Notably, some side effects have been observed and therefore MP has fallen out of favor for treatment.

15.3.4 Surgery

I think you’ll find that there is a disappointing lack of options—signaling the primitive nature of neurosurgery!

Decompression.

Anyway, surgical intervention aims to restabilize the spinal cord as quickly as possible, particularly through decompression. This begins with re-aligning the spine, usually with some kind of hardware to hold bone in place. Early surgery seems indicative of shorter ICU stays and better neurological recovery. The first day or so post SCI is the critical time. Interestingly, even after decompression, the pressure within the spine remains high due to fluid build-up within the dura matter. This makes blood reperfusion more difficult, leading to more problems.

⁵⁵<https://pubs.rsc.org/en/content/articlepdf/2022/bm/d1bm01744f>

Dura Matter Manipulation.

While durotomy is often a complication of surgery due to progressive CSF leakage after operation, in this case it can be helpful to lessen spinal pressure, which there evidently is a long tradition of⁵⁶. Duroplasty is a more modern and sophisticated alternative, and can allow opening of the dura matter without as much risk⁵⁷.

Myelotomy.

Incision directly into the spinal cord itself, myelotomy, has also been done with some success. The belief is that it helps drain some of the harmful dying tissue. There seems to be time dependence in this, where if performed too late after injury it will simply reinvigorate inflammation.

15.3.5 Rehabilitation

As you would intuit, exercise is the most common technique, as it preserves muscle mass and promotes circuit reorganization. Another rehabilitative technique is pumping in a significant amount of oxygen, as ischemia occurs after injury. BSIs have also become more popular. Fascinatingly, decoding of handwriting has been used to generate text⁵⁸. Though, these seem to require deep access to the brain.

15.4 Complications After SCI

The multiplexed complications of SCI are, in reality, unending, as any combination of nerve damage can result in any combination of bodily dysfunction. We can discuss some of the standouts below, but note that this covers just a fraction of a fraction.

Syringomyelia.

Syringomyelia, a cyst forming within the spinal cord, is relatively rare. The fluid filled sac can be large and span far beyond the injury site itself. The symptoms of which may not show for months-years later, and progressively worsen. Cystic cavities are not uncommon, but syringomyelia differs in its size and reach. Symptomatic patients usually undergo a second decompressive surgery, or attempts to connect this cystic space to a drainage site.

Neuropathic Arthropathy.

After SCI, a patient may become numbed, leading to unnoticed injuries. One such result of this is arthropathy (a joint disease). Over time, a patient's joints may slowly degrade, leading to deformity—which may not present for over a decade after injury.

Spasticity.

Spasticity is a common side effect of chronic SCI. Spasticity can lead to further complications, such as microfractures.

⁵⁶<https://www.sciencedirect.com/science/article/pii/S0020138388901325>

⁵⁷<https://www.liebertpub.com/doi/full/10.1089/neu.2014.3668>

⁵⁸<https://www.nature.com/articles/s41586-021-03506-2>

Cardiovascular.

As mentioned in previous sections, often the clearest symptom of SCI is in significant changes to the cardiovascular system do to sympathetic nervous system impairment. This manifests often as systemic hypotension. In most cases, this seems to resolve itself within a matter of months.

Autonomic Dysreflexia.

This condition is an immediate response to *secondary* injury, and most often presents when the SCI is above T₆, and some other injury occurs in some peripheral organ, which causes a sympathetic spinal reflex. For example, injury to the gut may cause vasoconstriction to be overstimulated in the absence of spinal cord feedback. Too, it can result in overcompensation by the parasympathetic nervous system, leading to a swift drop in blood pressure. Notably, this can occur either in the acute or chronic stages of SCI. It is most often resolved via resolving this secondary injury.

Respiratory.

Damage to nerves innervating respiratory muscles, such as the phrenic nerve which controls the diaphragm, can lead to long term reduced lung capacity. Damage to the respiratory system also has big consequences on the effectiveness of rehabilitation. It is said that respiratory complications are one of the leading causes of death after chronic SCI.

Genitourinary and Gastrointestinal.

Dysregulation of the urinary or digestive systems is often the largest psychological consequence of SCI (barring, of course, large locomotor deficits). Quality of life is greatly impacted by this complication.

Neurogenic Heterotopic Ossification.

Ectopic bone formation can occur on the larger joints⁵⁹. It is not known the exact cause, but physical therapy tends to be the best mitigator.

⁵⁹Funnily enough, this review sites that 10-53% of people with chronic SCI form ectopic bone. Only a 43% margin.

Chapter 16

Approaches to Researching SCI

The first ever modeling and examination of traumatic spinal cord injury dates back to over 100 years ago. Dr. Alfred Allen, at the University of Pennsylvania, devised dog models of spinal cord injury and performed their accompany histological investigations⁶⁰. I believe firms in finding the fundamental source, and appreciating where leaves meet branches, branches meet the trunk, and the trunk meets its roots (section 21.1).

NOTE: It would be good to include some of the other approaches to researching neural regeneration in general—such sciatic nerve crush, optic nerve crush, *Drosophila* axon injury, etc.

16.1 SCI Injury Models

The bulk of this information comes from⁶¹, and focuses on injuries applied to mice—but similar principles can be done in other mammals. Interestingly, most SCI models focus on thoracic level injury, while in humans cervical is more common.

Contusion.

Contusion is the quick application of a force. Models include dropping a weight on the spinal cord or using an air-gun. One such is called the *NYU (MASCIS) impactor*, where a piece of the bone is removed (laminectomy) and a weight is dropped directly onto the exposed nerves. Severity can be adjusted by the height from which the item was dropped. The most recent iteration is the *MASCIS III*, where an electromagnetic weight is dropped at the push of a button and includes digital measuring to ensure replication. One concern is that the weight dropping may bounce, causing multiple impacts.

Another option is called the *infinite horizon (IH) impactor*. Rather than dropping a weight, a machine applies a set force. A laminectomy is still required⁶², but one can avoid the bounce effect. The *Ohio State University (OSU) impactor* is essentially the same, but uses an electromagnetic force applier. A more recent development is in the air gun. This requires drilling through the vertebrae, exposing the dura, where air pressure will be applied.

⁶⁰<https://jamanetwork.com/journals/jama/article-abstract/448138>

⁶¹<https://www.nature.com/articles/sc201491>

⁶²I suppose repeatability is considered paramount, but I find that these extensive controls may impede investigation. At a certain point, one must wonder if this is an SCI or an axotomy model.

Compression.

Compression occurs over a longer time frame. One interesting application of such models is in first applying contusion, followed by a compression model. This replicates the effects of increased pressure within the spinal cord after SCI, which occurs particularly in the absence of decompression surgery.

A common method is using an aneurysm clip and applying it to an exposed section of the spine, clamping down with some force for a few minutes. The force applied is in the range of 20-50g⁶³. An interesting drawback is the velocity of the clip's closing is variable. In the right usage, it can be used to cut off blood supply to simulate ischemia and reperfusion after SCI. Another option is to insert a balloon on a catheter down the spinal cord and inflating it. Another method is called SC-STRAPPER⁶⁴. A string is inserted and wrapped around the spinal cord, and its benefits are that it is relatively noninvasive, and compression is applied to the entire circumference of the spinal cord.

Distraction.

Distraction is done by stretching the spine. This is more commonly done in larger animals, and is difficult to develop a reproducible model. The Harrington distractor requires laminectomy and the addition of hooks to the upper and lower sides of the removed vertebrae. Motor steps are then applied to pull the spinal cord apart. However, as this occurs slowly and is variable in application, it has drawbacks. The UBC multimechanism device uses a wedge applied between the vertebrae to abruptly apply force. Interestingly, their device can also be used to induce contusion or slipped disc injuries—so learning it provides you with considerable versatility, particularly if dislocated vertebrae is your area of study. The singular wedge applies unidirectional injury, and so the UTA distractor aims to improve this by using clips attached externally (i.e., without laminectomy) and pull apart by a motor.

Transection.

Transection requires surgical severing of the spinal cord. Transection models may target the ventral, dorsal, or subsections of the spinal cord. Such models are most useful for studying regeneration, as clean transections are rare clinical pathologies. Interestingly, some animals have impressive, natural recovery of locomotion due to spinal cord circuitry post SCI (i.e., more robust central pattern generation or spinal reflex systems). Therefore, it is often that one must re-transect after locomotor recovery.

16.2 The Basso Scale

The Basso scale⁶⁵ has a few variations, such as the Basso, Beattie, Bresnahan (BBB) scale. It is a mode of characterizing injury after SCI in mouse models, specifically in the hindlimbs. You will often see Basso scale comparisons between treatment groups, or as a means to track recovery over time after SCI. The scale is as follows (directly from the paper):

- Rating 0: No ankle movement.
- Rating 1: Slight ankle movement.

⁶³Perhaps I am naive, but this seems remarkably small to me. When you consider that human SCI may result from a car accident, a few grams seems inconsequential.

⁶⁴Take a look at Fig. 1: <https://www.sciencedirect.com/science/article/pii/S0165027008000368>

⁶⁵<https://pubmed.ncbi.nlm.nih.gov/16689667/>

- Rating 2: Extensive ankle movement.
- Rating 3: Plantar placing of the paw, without stepping, OR consistent or frequent dorsal stepping (without plantar stepping).
- Rating 4: Occasional plantar stepping.
- Rating 5: Frequent or consistent plantar stepping, without coordination, OR some coordination with paws rotated at contact and lift off.
- Rating 6: Frequent or consistent plantar stepping, some coordination, and parallel paws at contact, OR mostly coordinated, paws rotated at contact and lift off.
- Rating 7: Frequent or consistent plantar stepping, mostly coordinated, parallel paws at contact, rotated at lift off, OR mostly coordinated with, parallel at contact and lift off, severe trunk instability.
- Rating 8: Frequent or consistent plantar stepping, mostly coordinated, parallel paws at contact and lift off, mild instability, OR mostly coordinated with paws rotated at contact and lift off, normal trunk stability, and tail up or down.
- Rating 9: Frequent or consistent plantar stepping, mostly coordinated, parallel paws, normal trunk stability, and tail always up.

Slight is defined as less than half joint excursion. *Extensive* is defined as more than half joint excursion. *Plantar stepping* is defined as placing both the thumb and the last toe touching the ground. *Consistent* is defined as without missing more than 5 steps. *Frequent* is defined as more than half the time. *Occasional* is defined as less than half the time moving forward. *Coordination* is defined as proper alternations between feet, and hindlimb and forelimb movements occurring at proper times, and the mouse must move at a consistent, non-slow pace. Some coordination meaning less than half the movements are coordinated, while most is more than half. *Severe trunk instability* is defined as an extreme waddle and collapsing during the test.

It is almost certain that I misworded some of those (as in accidentally wrote most instead of some a few times). But, I think you get the overall gist! Essentially, above 5 seems to have some degree of uncoordinated walking. Around 7-9 seem to have regained functional walking.

16.3 Traumatic and Nontraumatic SCI

Here we'll focus on neuroimaging and the information from⁶⁶. A large focus of this review is the parallels between traumatic SCI and neurodegenerative diseases like degenerative cervical myelopathy (DCM). Traumatic spinal cord injury (tSCI) refers to impact or immediate stress applied to the spinal cord, while nontraumatic occurs over long periods through continuous compression or other. DCM is the most common example of nontraumatic SCI, and can occur through things like disc bulging. Naturally, months or years may go by without symptoms while degeneration is underway.

Animal models indicate considerable overlap in traumatic and nontraumatic SCI, but verification of these processes in humans is harder because metal implants immediately into tSCI patients makes

⁶⁶<https://www.nature.com/articles/s41582-019-0270-5>

MRIs impossible. Conventional MRI, too, is often not powerful enough to analyze microstructural changes, like demyelination, that are canonical to degeneration. A better method is diffusion tensor imaging (DTI), which captures the diffusion of water through the tissue. It relies on the assumption that water diffuses along the white matter tracts. Core to the benefit of DTI and improved imaging techniques is identifying markers which can inform our understanding of improved recovery after SCI treatment. For example, clinical trials can be better motivated by quantifiable, noninvasive metrics like DTI.

Quantifying DTI.

DTI is devised 3 eigenvectors, let's call them: $\vec{V}_a, \vec{V}_{r1}, \vec{V}_{r2}$. You would expect \vec{V}_a to be the longest of the three as it runs axially, and or with the axons. Its eigenvalue, let's call it λ_a , corresponds to the diffusivity in the axial direction (parallel to the myelin tracts). The average of λ_{r1} and λ_{r2} describes the radial diffusivity, and or that which is perpendicular to the myelin tracts. The average of all 3 eigenvalues is the mean diffusivity. A scenario when $\lambda_a \neq 0$ and $\lambda_{r1} = \lambda_{r2} = 0$ would be described as an *anisotropic* diffusivity, because there is complete directional dependence of diffusion. $\lambda_a = \lambda_{r1} = \lambda_{r2}$ would be *isotropic*, because there is no directional dependence. The idea is that the diffusion of water gives some key insights into the microstructure of the injury, as things like ECM proteins and membrane composition will have some affect on such water diffusion. The axons themselves are naturally the largest contributor to this directional dependence, and you can imagine that in their absence water is much more likely to diffuse in all sorts of directions. The degree of anisotropy is a strong way to measure degeneration, then. There are some limitations, of course, as diffusion is only a proxy for degeneration. Ulterior factors, like inflammation, may too limit diffusion and obscure the contribution by degeneration or demyelination. Similarly, DTI is not as effective at imaging lower spinal cord segments, and is primarily used in cervical injury models.

DTI in Degeneration and SCI.

Of considerable study is the parallels between SCI's secondary injury and degenerative diseases. After tSCI, retrograde and anterograde SCI initiate. Interestingly, in the first 2 years following SCI, tissue volume decreased by $\approx 14\%$, and $\approx 30\%$ in 5 years. Such a decrease was greater in the grey than in the white matter, which is similarly observed in DCM. Interestingly, no difference in diffusivity within the grey matter is observed after SCI, but it is believed this may be due to volume effects in the region.

Without going into the great detail on the vector changes associated with DCM and tSCI, the two show similar decreases in anisotropy. Notably, DTI is able to resolve changes in asymptomatic DCM patients, making it a possibly useful tool in predicting worsening or improving outcomes of SCI that are not detected functionally. Remotely increased radial diffusivity was observed in both tSCI and DCM, suggesting that the local injury site / stenosis can have distant effects in both. However, DCM showed increased axial diffusivity above and below the site of stenosis, differing from tSCI. The authors suggest that this may be because compression at these sites increased cell density and decreased the extracellular space available for diffusion. Notably, though, the increase in radial diffusivity significantly outweighed axial increase, so it is still believed that demyelination is the dominant process in spinal stenosis.

Really, the takeaway from this is that cell death patterns in tSCI and degenerative diseases like DCM are comparable, but better investigations and imaging techniques will be needed to fully resolve this.

Chapter 17

Broad Coverage of Axonal Regeneration

Let's discuss⁶⁷ and⁶⁸. During development, axons are guided by their growth cones to achieve functional connections at distant targets. A quintessential barrier to recovery after spinal cord injury is the central nervous system's inability to regain such functional connections. In general, neuron regeneration (or neuroregeneration) refers to axonal regeneration of an injured neuron. However, *sprouting* is another form of regeneration in which uninjured neurons may grow to replace the damaged synapses. Similarly, the use of stem cells can too be considered a form of regeneration.

While the CNS is relatively un-plastic after injury, some endogenous plasticity after spinal cord injury can and does occur, but in large part results in spasticity and neuropathic pain after SCI. An in-depth discussion on circuit reorganization is covered later (chapter 19)—so for the moment, let's consider endogenous plasticity to be negligible or harmful after spinal cord injury.

17.1 Intrinsic Factors

Growth Cone Manipulation.

The physical interactions and the cell matrix, which typically relate to growth cone extension, are very important in growth. Deletion of non-muscle myosin II has shown enhanced growth. Deletion of profilin 1, which binds to actin and is thought to enhance its polymerization and use in filopodia, halts regeneration. Stabilizing the growth cone, or trending microtubule dynamics toward polymerization, is key to axonal extension. Doublecortin-like kinases (DCLKs) promote regeneration and are thought to modulate microtubule dynamics. Epothilone B and D are microtubule stabilizers, which have been shown to promote recovery after SCI. Epothilone B is also an anti-cancer drug as it inhibits mitosis, and similarly seemed to reduce glial scarring likely by inhibiting glial mitosis. Other similar, anti-cancer microtubule stabilizing drugs, like taxol, have shown similar outcomes. Importantly, other studies have called these findings into question, and some have shown quite negative outcomes in their usage. Similarly, it is unclear if their benefit is neuronal in origin or if their role in scar formation is sufficient to drive recovery. However, given that they are already used in cancer treatment entices researchers to find a use for them in treating SCI. To this point, though, their conclusive function is unclear.

⁶⁷<https://www.nature.com/articles/s41582-019-0280-3>

⁶⁸<https://www.nature.com/articles/s41580-022-00562-y>

Anterograde transport, including integrins and their RAB carriers, are shown to support regeneration by supplying the growth cone with necessities. However, selective transport is an essential focus, as RABs are likely to bring both pro and anti-regenerative factors to this site.

RhoA GTPase, a protein localized to the growth cone and functions in microtubule activity through Rho kinase (ROCK), has also been studied in an attempt to maintain the growth cone. Inhibition of RhoA or ROCK has promoted regeneration—specifically in acute SCI. If it is helpful in treating chronic SCI remains unclear.

Promising Proteins.

Target of rapamycin (mTOR) is all too familiar now, as it has been identified as an active translator in the axon and growth cone. mTOR phosphorylates 4EBP1, thereby inhibiting it, preventing it from repressing translation. Phosphatase and tensin homologue (PTEN) is a negative regulator of mTOR. PTEN is likely the most famous regeneration blocker—and for good reason! Inhibition of PTEN even a year after SCI has been shown to promote enhanced recovery. Though, the regeneration is still typically insufficient for meaningful and complete functional recovery. Currently, developing therapeutics against PTEN are marred by worries that it will produce malignancies within the tissues, as PTEN is also an essential tumor suppressor. Interestingly, PTEN inhibition alone strikes somewhat variable results. Co-inhibition of PTEN with suppressor of cytokine signaling 3 (SOCS3) (mentioned in the next section) provided more consistent, enhanced growth—especially in the critical corticospinal tract.

Inhibition of certain calcium channel components, like the $\alpha_2\delta_2$ subunit of voltage gated calcium channels, promote axon regeneration. One such method of doing so is through pregabalin, which limits spasms. Further evidence by administration of gabapentinoids enhancing regeneration shows promise in using anti-convulsants to treat SCI. This work fits into the larger hypothesis of neuronal activity limiting axon regeneration. Further work demonstrates that active vesicle machinery, like the protein Munc13, limit axon regeneration. Other work has shown that partial preservation of synapses stalls regeneration when full loss would have furthered it—suggesting that axons tend not to regenerate unless their connections have been completely abolished.

Indirectly related, our lab has identified the pore forming subunit of L-type voltage gated calcium channels, α_1 , to be necessary for regeneration. Therefore, the role of VGCCs and neuronal activity in regeneration is likely multiplexed.

Other proteins, like KCC2 (**19.1.1**), are discussed in-depth in other sections.

Transcriptional Possibilities.

Many transcription factors can be used to boost regeneration. One such is the cAMP-responsive element-binding protein (CREB) is necessary for axon growth in development. CREB activation by TTK21 promotes recovery after SCI. Another is signal transducer and activator of transcription (STAT3). Determining the biochemical barriers to CREB or STAT3 activation has been done through investigating histone deacetylases (HDACs) and DNA methyltransferases. It was shown that in peripheral nerves, where regeneration occurs naturally, histones were acetylated at specific, pro-regenerative sites. TTK21's activation of CREB is done through CREB-binding protein (CBP) acetyltransferase.

Fascinatingly, constitutive STAT3 activation is insufficient to promote regeneration. The reason be-

ing SOCS3 is a negative regulator of the JAK-STAT transcriptional pathway. While STAT3 may be activated by therapeutics or intrinsic growth factors, SOCS3 knockdown is required to attain its benefits. Another key family of transcription factors within this umbrella are Krüppel-like factors (KLFs). Many members of the KLF family have been found to block neurite outgrowth, with KLF4 being the strongest. Conversely, overexpression of other KLFs, like KLF6, enhance regeneration. Co-Knockdown of KLF4 and SOCS3 further promoted growth. KLF4 is thought to bind with phosphorylated STAT3 and inhibit it, while KLF6 seems to further activate it.

Minorely related is the discovery of the dual leucine zipper-bearing kinase (DLK). DLK is necessary for retrograde transport of STAT3, suggesting it is used in injury signaling. Interestingly, evidence for the DLK pathway's injury response in glia has emerged. In this case, DLK knockout worsens the injury site as astrocytes do not properly respond. The same can be said for SOCS3-STAT3 and PTEN-mTOR, demonstrating a correlation between the intrinsic factors dictating neuronal and glial injury responses.

17.2 Extrinsic Factors

Growth Factors.

A canonical problem is the lack of growth factors that normally guide axonal growth in development. Such canonical examples include nerve growth factor (NGF), brain-derived neurotrophic factor (BDNF), neurotrophin 3 (NT3), and glia-derived neurotrophic factor (GDNF). Naturally, growth factors have different, neuron subtype-specific effects. NGF modulates nociceptive neurons, NT3 and GDNF modulates corticospinal neurons, BDNF modulates corticospinal and reticulospinal neurons, etc. Notably, matured neurons may not be receptive to such growth factors, and one mode of treatment includes supplementing both the growth factors and their accompanying receptors. Other options exist, such as in sensitizing the neurons to growth factors therapeutically.

A key finding in this department is that the insulin-like growth factor 1 (IGF1) promotes regeneration. This may be partly due to its ability to promote mTOR activity, but the many potential outcomes of IGF1 overexpression are unknown. IGF1 injection has been used to treat autism in the past with relatively safe and promising outcomes, however the fear of tumor production remains a barrier to its use in treating SCI.

Growth Cone Navigation.

The cell-specific responses surrounding neurons is covered in-depth in section 14.1. Many of the modifiers exist in the ECM, such as secretion of collagen protein (Cthrc1), a pro-regenerative protein⁶⁹. Conversely, WNT signaling molecules accumulate around the injury site and repels axon regeneration and sprouting.

Other non-neuronal cells, like astrocytes and fibroblasts, secrete chondroitin sulfate proteoglycans (CSPGs). CSPGs are composed of a protein core with chondroitin sulfate sugar exteriors, whose inherent variability makes their study difficult. After injury, CSPGs too accumulate around the injury site and limit growth. They can do so through the protein tyrosine phosphatase σ (PTP σ) receptor or Nogo, which may trigger RhoA.

⁶⁹[https://www.cell.com/developmental-cell/pdf/S1534-5807\(20\)30984-9.pdf](https://www.cell.com/developmental-cell/pdf/S1534-5807(20)30984-9.pdf)

Administration of chondroitinase ABC (ChABC) has been shown to promote plasticity and growth by degrading CSPGs. ChABC administration in non-human primates promoted hand dexterity recovery without negative side effects. ChABC has poor stability and is difficult to produce in large quantities, making its clinical applications somewhat limited. Too, there is some evidence of their benefit, such as through perineuronal nets which may provide a supportive structure to neurons in the grey matter.

Myelin Inhibition.

Interestingly, many myelin-associated proteins are inhibitory, called myelin-associated inhibitory (MAI) proteins. These proteins bind to the Nogo receptor complex, and blocking this interaction, such as by NogoA antibodies, promotes growth, especially in the acute phase. Examples include myelin-associated glycoprotein (MAG) and oligodendrocyte myelin glycoprotein (OMGP), which both utilize a NOGO to RhoA-ROCK pathway to inhibit regeneration.

Chapter 18

Metabolism of Neurons

18.1 Programming Mitochondrial Maintenance

This is one of my favorite Reviews ever, so I am pleased to write about it⁷⁰! A key component of this work is discussing the unique structural features of neurons that lead to trouble trafficking mitochondria and meeting injury needs after injury or in degeneration. An incredible amount of energy is consumed by the brain (20% of total body consumption), and approximately 1,000,000 molecules of ATP are used per action potential in an effort to restore ionic concentrations alone. Around 20,000 molecules of ATP are used per synaptic vesicle recycles, and its predicted that the concentration of ATP in the presynaptic terminal is around 2mM.

18.1.1 Mitochondrial Generation and Degredation

Notably, mitochondria are not spontaneously generated, but rather undergo fission to create more—which primarily occurs in the soma where such machinery is located. Thus, irrespective of how much mitochondria are present, the key is that it must be properly trafficked. Amazingly, it can take days for mitochondria to reach the tip of an axon. Late endosomes carrying RNA granules and ribosomes may dock on mobile mitochondria and serve as sites for mitochondrial protein synthesis. Peroxisome proliferator-activated receptor- γ coactivator (PGC)-1 α is a key transcription coactivator of nuclear-encoded mitochondrial genes. ALS and Parkinson's disease models both show decreased distal translation of mitochondrial proteins and build-up of the corresponding mRNA, suggesting the energy demands are not met in these diseases.

Mitochondrial fission is mediated by dynamin-related protein 1 (DRP1), a cytosolic GTPase. DRP1 acts with mitochondrial adaptors, including fission protein 1 (FIS1), mitochondrial fission factor (MFF), and mitochondrial dynamics proteins (MiD), to induce fission. While larger mitochondria are better at energy production, smaller mitochondria are more easily trafficked to distant sites. Notably, fission can lead either to biogenesis or to mitophagy, depending on where the fission occurs and which adaptors mediate it. FIS1 specifically tends more toward mitophagy, while MFF more toward biogenesis.

Mitochondrial fusion is mediated by factors like optic atrophy 1 (OPA1) and mitofusion 1 and 2 (MFN1 and MFN2). Both are GTPases and aid in inner membrane and outer membrane fusion respectively. Fusion is useful in limiting mitochondrial stress, as combining the contents of two dilutes

⁷⁰<https://www.sciencedirect.com/science/article/pii/S0896627322002513>

harmful agents and aids in the increased production of energy. Mutations in OPA1 are commonly linked to Parkinson's disease.

Fixing damaged components of the mitochondria may occur through proteases lining the inner membrane. Alternatively, mitochondria frequently bud off vesicles, called mitochondrial-derived vesicles (MDVs), that contain damaged proteins or reactive oxygen species. Such budding is achieved through mitochondrial Rho GTPase (MIRO) and DRP1, and are then targeted to lysosomes. PTEN-induced kinase 1 (PINK1) is also able to enhance degradation through mitophagy, along with Parkin (PARK2). Such pathways are dysfunctional in Parkinson's disease as well. PINK1, on the outer membrane, recruits Parkin, which recruits degradative machinery. Interestingly, axonal mitophagy is incredibly rare. Typically, mitophagy occurs only after a damaged mitochondrion is trafficked to the soma, and in the case of Parkin or PINK1 mutations, damaged mitochondria build up in the soma rather than the axon.

18.1.2 Mitochondrial Trafficking

Trafficking is necessary for both the removal of damaged mitochondria and the addition of them to energy depleted sites. The speed of trafficking is quite variable, and is capable of both stalling and changing directions. Such movement occurs along the polarized microtubules within axons, especially by motor proteins like the kinesin-1 family (KIF5) for anterograde and dynein for retrograde transport. On the outer membrane, the MIRO-trafficking kinesin-binding proteins (TRAK) complex connects to KIF5 motors. It seems that there is some functional redundancy in the MIRO-TRAK complex with other outer membrane proteins, but in large part it seems to be a strong regulator of mitochondrial placement.

In the case of injury, movement of the mitochondria to the damage site is a key component of repair. One way to enhance promote movement is by deleting syntaphilin (SNPH), a mitochondrial anchoring protein. SNPH is capable of competing with the MIRO-TRAK complex's binding with KIF5. It is thought that p21-activated kinase 5 (PAK5) inhibits SNPH by phosphorylating it, enhancing mitochondrial mobility. Overexpression of MIRO1 was found to increase mitochondrial trafficking to the distal axon segments, leading to larger growth cones. MIRO1 is mediated by neuronal activity, as Ca^{2+} influx binds with MIRO and disrupts complex formation. Though, this blocking seems transient and seems to require the presence of SNPH. This mechanism seems particularly important for halting mitochondrial movement around the Nodes of Ranvier, where a higher energy demand exists to restore membrane potential.

Mitochondrial trafficking is adaptable. For example, high concentrations of glucose are capable of arresting its movement through glycosylation of TRAK. Guidance of mitochondria through the complex branches of an axon require such intracellular cues, some of which are discussed in the next section.

18.1.3 Response to Stress

Stress induced AMP-activated protein kinase (AMPK) phosphorylates PGC-1 α to upregulate mitochondrial RNA. ATP depletion has been shown to activate the AMPK-PAK3 stress pathway, which uses myosin 6 to recruit mitochondria and SNPH to anchor it to surrounding F-actin. Therefore, exceptionally active synapses may experience mitochondrial buildup proportional to their firing. As it goes, stress induced anterograde trafficking in a highly branched axon may result in mitochondrial build-up at the branch sites, which is achieved too by SNPH. Parallely, AMPK is capable of

blocking retrograde transport of mitochondria, resulting in their accumulation at distal, stressed sites. Counteractive to trafficking, reactive oxygen species may result in increasing intracellular Ca^{2+} , which arrests mitochondrial movement. As mitochondria are one of the main sources of such species, this can become a self-fulfilling prophecy of positive feedback.

Neuronal Degeneration.

Most neurodegenerative diseases are accompanied by mitochondrial dysfunction. Build up of protein aggregates, for example, hinder trafficking. Too, mtDNA mutations can regularly cause neurodevelopmental impairments.

PINK1/Parkin, while traditionally known for its role in Parkinson's, defines many mitochondria-related pathologies. Its function is thought to be in degrading MIRO through ubiquitination, causing local mitochondrial arrest and eventual mitophagy. Conversely, a mechanism of remobilization where mitochondrial vesicles containing SNPH are released with the idea of enhancing retrograde transport after stress. This mechanism is activated in pre-symptomatic ALS models and Alzheimer's disease. It seems that early disease causes mitophagy primarily in the soma, while later and/or higher levels of stress may induce it within the axon. The site of mitochondrial degradation is therefore an indicator of disease progression.

18.1.4 Neuronal Injury and Regeneration

While degeneration is progressive and slow acting, injury is acute. Repair is inherently a highly energy intense process. The injury itself can cause mitochondrial structural damage, furthering the energy crisis. Preserving mitochondrial function is a mode of limiting Wallerian degeneration after injury. Axotomy quickly depolarizes mitochondria, and inhibiting SNPH serves as an important way to remobilize and restore their function. In certain injuries, the AKT path becomes activated which promotes PAK5 (mentioned above as inhibiting SNPH), serving as an energy-boosting mode of neuronal repair. Another such method is via ARMCX1, a protein that mobilizes mitochondria through MIRO1. Histone deacetylase 6 (HDAC6) is known to deacetylate the MIRO1 encoding gene, leading to lessened mobility. Conversely, hypoxia inducible factor 1 α (HIF-1 α) activates hypoxia upregulated mitochondrial movement regulator (HUMMR), which promotes MIRO1/2.

After injury resulting in limited blood supply, such as in a stroke, metabolism occurs through glycolysis, resulting in increased lactate production. However, after the blood supply is regained, repurfusion injury can overload the system with oxygen, resulting in production of reactive oxygen species, thereby further harming mitochondria.

Delivery of mitochondria to damaged neurons through vesicles or tunneling nanotubes (TNTs) is thought to be one mode of enhancing recovery. It has been shown that neural stem cells are capable of transferring mitochondria to damaged cells, and microglia use TNT to do so. Mitochondrial transplantation has been done, however the efficacy of such is heavily debated. Though, it is well supported that complex networks of metabolite shuttles exist in the nervous system. The astrocyte-neuron lactate shuttle (ANLS) is a prime example, thought to allow astrocytes to fuel neural oxidative phosphorylation. Glial regulation of neuronal metabolism too occurs, such as through release of sirtuin 2 (SIRT2) which promotes metabolism in neurons.

Chapter 19

Circuit Reorganization and Plasticity

We will cover circuit reorganization in considerable depth, not necessarily in the context of spinal cord injury, in this section.

19.1 Reorganization After Injury

Circuit reorganization naturally and with neuromodulation is covered in-depth here⁷¹. Part of this paper is framed on the fact that spinal cord circuits are much more complex than previously thought. That is, some computation is done at the level of the spinal cord. Too, the paper attempts to motivate investigating the specific neurons involved in recovery, and that previous investigations have not used the resolution necessary to investigate such matters. One constraint on the field is that, while the brain's cellular profile has been extensively investigated, detailed profiling of spinal cord cells is less explored—particularly in the injury condition.

Circuit reorganization can occur in many ways, both above and below the injury site. Examples seen in our own lab include damage to axons causing retraction and thus sprouting of the axon in seemingly random directions far above the injury site. Too, below the injury site, axons that once filled this space will fade away, and uninjured axons may grow into this now empty space. Interestingly, growth into this empty space has been shown to augment hand function after recovery⁷².

Of course, when damaged neurons regrow in their previous locations it can cause functional recovery. However, fascinatingly, even when all corticospinal tract neurons are damaged, motor cortex signals can be sent through alternative, surviving paths. Here are two brainstem examples: (1) After SCI, cortical neurons can synapse onto the reticular formation and recover walking. (2) Activation of the mesencephalic locomotor region (MLR) after SCI can recover walking. This region is evolutionarily ancient and unable to generate walking on its own, but is able to recruit reticulospinal neurons and generate walking.

In a similar vein, otherwise non-essential neurons, like propriospinal neurons in the cervical region, can be reprogrammed to improve hand dexterity. While one may think that information must be conveyed directly from the motor cortex to the muscles on interest, in this case propriospinal neurons relay information past the injury site. KCC2 downregulation leads to inhibition of such circuits. This

⁷¹<https://www.nature.com/articles/s41593-022-01196-1>

⁷²This is all very interesting when it comes to BSMIs—as one usually focuses on or around the injury site, or the dorsal root exit zones that generate walking. This suggests many sites along the spinal cord may be worth stimulating.

is expanded on in a later section. Much remains unknown about such neurons, what allows them to transmit beyond the injury site, and what recruits them to do so.

Rehabilitation proves to consistently aid circuit reformation, but lack of care causes poor reorganization that can worsen pathologies. Notably, when supraspinal⁷³ connections are severed, motor control relies on sensory signals only. One example being *V2a* neurons discussed in a previous section. Another similar set of neurons includes *dI3* neurons⁷⁴, another set of interneurons. While these aid in recovery, many changes are harmful to it. As a general trend, it appears that generally less activity, due to injury conditions like hypoxia, immune cells, etc. leads to greatly increased sensitivity of receptors and ion channels alike. Below the injury site often sees large increases in excitatory synapses. This combination leads to abnormal reflex responses, spasms, neuropathic pain, etc.

Effects of Neuromodulation.

Again, neuromodulation is covered in considerable depth in Part V. Both activity and artificial stimulation of the motor cortex promote regrowth of the corticospinal tract. This can be done non-invasively. Timing this with spinal cord stimulation enhances connections. There are a couple of paper's exploring this in depth in NHP⁷⁵, and in humans⁷⁶. **It would be good to look through these papers in the future.** Fascinatingly, indirect neuromodulation, such as through the vagus nerve, too has been shown to enhance response to rehabilitation. Too, deep brain stimulation has been done on the MLR⁷⁷, and when positive yielded results comparable to stimulating the motor cortex⁷⁸. Though, the MLR's region is less clearly defined, and results were more variable.

Stimulating the spinal cord is primarily done at the dorsal root entry zones or through EES. It is thought that low-frequency EES works through proprioceptive afferents, which help circuit reorganization. Damage to the sympathetic nervous system is an essential problem to solve, as it can lead to life threatening drops in blood pressure. One such way to do so is in stimulating the low thoracic spinal cord's pre-ganglionic sympathetic neurons, causing release of norepinephrine and thereby constrict vessels. Similarly, EES can restore the sympathetic response via large afferents recruiting sympathetic neurons.

Effects of Biological Manipulation.

Biological strategies seem most successful when attempting to enhance the projectome of surviving supraspinal neurons. One such way to do this is through removing damaged tissue and grafting in more malleable, healthy tissue. As part of this, physicians may attempt to ablate the fibrotic scar. Interestingly, partial ablation, and not complete ablation, has shown symptom improvement after SCI.

Pharmacological promotion of axon regeneration has inherent limitations. Such investigations into axon regeneration is covered in better depth in chapter 17. For example, the canonical player PTEN (discussed in other sections) has shown robust regeneration, capable of penetrating the astrocytic border, but only in certain neuronal populations. Too, even in this case neurons abruptly halt when

⁷³Supra meaning "above" the spine, i.e., brain-to-spine neurons.

⁷⁴<https://elifesciences.org/articles/21715>

⁷⁵[https://www.cell.com/neuron/pdf/S0896-6273\(13\)00762-9.pdf](https://www.cell.com/neuron/pdf/S0896-6273(13)00762-9.pdf)

⁷⁶<https://www.pnas.org/doi/abs/10.1073/pnas.1505383112>

⁷⁷Another example of DBS promoting rather than inhibiting—kind of interesting.

⁷⁸It would be nice to build a patient profile of those who may benefit the most from motor cortex vs. MLR vs. vagus nerve etc. stimulation.

reaching the fibrotic scar. Somehow, one must induce global, or brilliantly targeted, growth to achieve true functional recovery.

Another application commented in earlier sections is the conversion of inhibitory neurons into excitatory, or the deletion of inhibitory neurons. One way to do so is via chondroitinase (ChABC) which dissolves glycosaminoglycan side chains of chondroitin sulfate proteoglycans (CSPGs). CSPGs are known for their role in stem cell differentiation, and bookend the periods of neurogenesis. Another such attempt is in inhibiting nogo receptor reticulon 4 (RTN4).

Reversion to a state of biological development too is another core route of investigation⁷⁹. As proprioceptive neurons have been found to be possible relay neurons, these researchers targeted such neurons and were able to achieve growth across the injury site. This required the use of growth factors applied to the neurons, fibroblasts, and attracted molecules downstream of the injury site. Release was done in a spatially and temporally controlled manner. Fascinatingly, **no functional recovery was found**. Therefore, while neurons were able to bridge this gap, they were unable to form useful synapses. Authors postulate this may be due to inactivity or lack of myelin, among other things. In a similar vein, the grafting of stem cells is often successful in allowing for axonal growth across the injury site, but retains the inability to restore function.

19.1.1 Interneurons and KCC2

This is from the paper⁸⁰, but it seems KCC2 is an interesting topic of research in a couple different pathologies, such as in minimizing neuropathic pain, and downregulation increases spasticity after SCI. Some of the background for this paper is in better understanding how spinal cord stimulation leads to regain of function, with one hypothesis being that it helps reawaken “dormant” circuits.

To investigate, they used staggered, lateral hemisections performed at the T₇ and T₁₀ sites. This injury meant all neurons would be severed, except those crossing the midline between T₇ and T₁₀. They verified near complete paralysis after injury, and used this to conclude that while some neurons were spared, they must be “dormant” and/or unable to recover function. They performed a mini-screen of drugs known to modulate neural activity, and landed on CLP290 as the only one showing a phenotype. CLP290 is known to activate KCC2, among other things. They determined that CLP290’s function required some preserved axons—as in their complete lesion tests, no functional recovery was found. They further verified this by determining no enhanced regrowth was found due to CLP290 in their hemisection model. To determine if CLP290’s function was through KCC2, they tried a KCC2 overexpression line and verified similar functional recoveries.

They then used the Cre-Lox system to verify that KCC2’s expression specifically in inhibitory interneurons (driven by Vgat-Cre) leads to recovery. They presumed these neurons interact with / enhance the ability of propriospinal neurons to transmit information, and to resolve the spatial component, they tried local injection of AAV-KCC2 in the lumbar segments and did not find recovery. They believe it is specifically interneurons within the relay site between the two lesions that are required for repair. Their explanation being that when administering systemically, the blood-brain-barrier is compromised only at the lesion site.

⁷⁹<https://www.nature.com/articles/s41586-018-0467-6>

⁸⁰<https://www.sciencedirect.com/science/article/pii/S009286741830730X>

In development, intracellular Cl^- is high. During this period, Cl^- channel opening is excitatory, rather than inhibitory in fully-developed cells. KCC2 upregulation is important in removing this Cl^- , converting Cl^- channel opening into being inhibitory. They believe activating KCC2 helps restore a more physiological state. They also did work to inhibit inhibitory interneurons, and excite excitatory interneurons, and determined that only the former was beneficial in achieving recovery after SCI.

19.2 Non-coding RNA in Rewiring and Plasticity

Background.

Non-coding RNAs (ncRNAs) have been studied quite a bit in recent years⁸¹. ncRNAs are classified as small (less than 100 nt) or long (above 500 nt), and among the small ncRNAs are microRNAs (miRNAs) of $\approx 20\text{nt}$. They are typically produced through canonical means and play a role in post-transcriptional regulation within the cytoplasm.

As for miRNAs, gene silencing occurs by binding to a complementary untranslated region of an mRNA. This region is sometimes called a *seed region*. From here, a miRNA-induced silencing complex (miRISC) forms. Then, some of the standard RNA degradation mechanisms can occur like decapping. Some of the relevant long ncRNAs include long intergenic RNAs (lincRNAs), which are linear, circular RNAs (circRNAs), and antisense (AS) RNAs. CircRNAs are produced but noncanonical means, and generally seem to function in tandem with miRNA in some manner, and or, they regulate miRNAs. AS RNAs are born of the opposite DNA strand, most often modulate this specific locus, and may be thousands of nt long. They are particularly relevant in early development. ncRNAs work in a symphonious way, usually involving multiple types coordinating together. This Review makes it a point to state that recent data has been overwhelming, due to advancements in technology. Therefore, it remains difficult to stay up to date or understand the complexities as new functions continuously emerge.

Regulation of Local Growth.

Local translation was first determined to be mediated through miRNAs in the case of miR-9's ability to decrease the expression of microtubule protein MAP1B via the brain-derived neurotrophic factor (BDNF) path within growing axons. Similarly, BDNF increased presence of miR-132, leading to increased axonal branching. In *Xenopus*, miR-182 was shown to be necessary for axon targeting via multiple modes of tubulin and actin regulations and their usage in the growth cone's guidance.

In the dendrites, miR-9 has also been implicated. miR-9 is under a negative feedback loop, in which the repressor element 1 (RE-1) silencing transcription (REST) factor is activated by miR-9, which in turn decreases its expression. miR-9 downregulated the actin regulator Diap1, inhibiting further growth. In short, miR-9 inhibits growth of both dendrites and axons.

In mice, loss of miR-101 led to increased network excitability and impaired memory through increased expression of Na-K-Cl cotransporter 1 (NKCC1) and increased synaptogenesis, thought to be mediated through *Kif1a* and *Ank2*. Fascinatingly, miR-101 acute loss did not lead to such a result, but rather only when lost in development. The lincRNA ALAE was the first example of an lncRNA that

⁸¹[https://www.cell.com/neuron/fulltext/S0896-6273\(23\)00341-0](https://www.cell.com/neuron/fulltext/S0896-6273(23)00341-0)

locally regulates mRNA translation during growth by blocking the function of RNA-binding protein, and thereby allowing GAP43 synthesis, causing axons to grow⁸².

Circuit Remodeling.

In general, it appears that miRNA expression is bad for remodeling, and long-term potentiation (LTP) specifically. miR-134 expression was found to inhibit long-term potentiation through inhibition of CREB1. Under physiological conditions, SIRT1 inhibits miR-134 expression. miR-26a and miR-384 also inhibit LTP by different means. Many such other examples exist⁸³.

Interestingly, miR-153 was upregulated in the dentate gyrus during LTP. It seems to be required for exocytosis, suggesting it has a role in excitatory coupling. miR-218 and miR-137 were also named as promoting LTP. I think that's about all that's worth saying, for now.

⁸²I think part of the issue with this kind of paper is it is almost purposeless trying actually learn this. It is simply letters followed by numbers doing things redundantly.

⁸³I hate to kind of “yada-yada” through this, but again, it's just miR-*xxx* does different, harmful things.

Chapter 20

Stem Cell and Tissue Engineering

In this section we'll discuss a few labs in much depth, such as the Tuszynski lab at UCSD and the Cullen lab at UPenn—some of the pioneers in three dimensional tissue engineering. My understanding is that most researchers would agree stem cells alone are likely to be insufficient to attain functional recovery after SCI. Even the most staunch supporters would agree an integrated approach, combining neuromodulation and rehabilitation, is certain to be better. The ependyma is a source of stem cells, which contributes to glial scar formation after SCI. Connexin (gap junction) signaling between ependymal cells is a factor involved in their differentiation, notable because it opens up the possibility of manipulation through epidural electrical stimulation, discussed in chapter **23**.

Injections of both oligodendrocyte progenitor cells and spinal-cord derived neural stem cells have been grafted into patients without complications, demonstrating promise.

20.1 Cell Transplantation Therapy

In this section, we'll primarily cover⁸⁴. The majority of implantation investigation focuses on neural stem and progenitor cells (NSPCs, or NPCs), Schwann cells, oligodendrocyte precursor cells (OPCs), olfactory ensheathing cells (OECs), and mesenchymal stem cells (MSCs). NSPCs are multipotent progenitor cells and are able to differentiate into neurons and glia alike.

Naturally, one of the primary benefits of grafting stem cells is that they may replace damaged neurons and form new functional connections. Alternatively, they can form relay stations where endogenous axons synapse onto the newly grafted cells. The physical addition of cells itself plays the role of giving neurons something to synapse and grow onto. If, for example, a neuron was capable of growing, the unideal milieu, such as open cavities filled with interstitial fluid, do not provide anything to grow on. This is sometimes called a *bridge*, especially when it allows neurons to grow rostrally to caudally. Through production of ECM proteins like laminin and collagen, a new, more favorable environment can form. However, such bridges still have great limitations, causing heterogeneous regrowth due both to the neurons' inherent abilities and because the graft may be dissimilar from certain axon's needs.

It is also thought that many such implanted cells allow for neuroprotection after SCI. This will be commented on a bit later, but one issue with this is neuroprotection is strongest when the implantation is done in the acute phase. Implantation after the first day sparsely shows protection. Yet,

⁸⁴<https://www.nature.com/articles/nn.4541>

implantation is normally done at 1-2 weeks after injury, because when done in the acute phase, the graft tissue is more prone to dying. This creates a slight Catch 22, where to maximize protection one must implant early, thereby sacrificing the graft and limiting the other potential benefits. The proposed mechanisms include protecting oligodendrocytes via healthy signaling, production of myelin, and the sprouting of new axons. All of these factors may contribute to the observed enlargement of the white matter near the injury site after a graft.

Other mechanisms of study include revascularization. Revascularization begins around 3d after injury and reaches its maximum around 14d after injury⁸⁵. OECs, in particular, appear to increase vascularization and enhance its directionalization along the graft and or spinal cord. Immunomodulation is similarly of interest, as correlative studies demonstrate grafting decreases harmful inflammation.

A general concern is where implanted cells may migrate. Therefore, you'll likely see the mention of halting migration in further reading. One strategy to doing so is encasing the stem cells within a matrix, typically of fibrin.

20.1.1 Long Distance Growth of Stem Cells

Stated in the 2019 Giovanni Review: "We believe that the most promising preclinical strategy to promote long-distance regeneration of axons in the spinal cord has come from recent studies in the **Tuszynski laboratory**, which used human spinal cord-derived NSC grafts combined with a cocktail of growth factors." Indeed—let's begin with such a study⁸⁶. The idea of this paper is to use stem cells cultured in a fibrin matrix, containing a cocktail of growth factors, and then graft this entity into rats after SCI. Growth was mTOR dependent, but not Nogo dependent⁸⁷.

Rats underwent complete T₃ transection. Rat embryonic stem cells were labeled with GFP and differentiated primarily into neurons after grafting. Cells did not migrate much beyond the graft / lesion site either. Grafted neurons were capable of growing long distances, with axons extending multiple vertebrae away both rostrally and caudally. They determined approximately 29,000 axons grown after grafting, and a similar density among all subjects tested. In fact, grafted neurons even became wrapped by host oligodendrocytes! Their justification for the growth not being Nogo dependent is that grafted cells still expressed Nogo, so therefore it is unlikely that Nogo inhibition promoted the growth. They tried inhibiting mTOR and found a significant decrease in axonal growth.

They then repeated these studies using human-derived stem cells. I didn't see this question answered directly in the text (maybe I just missed it) but I am wondering how much the neurons actually matter, especially for functional recovery. I.e., the matrix contained growth factors—could the growth factors alone be promoting growth, independent of the stem cells?

20.1.2 3D Printed Spinal Scaffolds

The ability to create 3D structures of cells is becoming ever more popular. One such example is again from the Tuszynski lab⁸⁸. The goal of this paper is to generate a scaffold capable of promoting growth, more so than creating a CNS mimic. They essentially designed a cross section of the spinal cord with

⁸⁵Take this with a large grain of salt.

⁸⁶[https://www.cell.com/fulltext/S0092-8674\(12\)01018-5](https://www.cell.com/fulltext/S0092-8674(12)01018-5)

⁸⁷Dr. Strittmatter will be upset!

⁸⁸<https://www.nature.com/articles/s41591-018-0296-z>

pores / channels where axons can grow into, synapse onto neural progenitor cells, and those cells can grow in the opposite direction and synapse onto neurons below the injury site. It is kind of like a well structured relay station for signals. They call their method microscale continuous projection printing (μ CPP), and remarkably it is able to print such scaffolds in a matter of seconds. They showed an interesting potential application where a 3D model of a patient's SCI lesion was taken and then a corresponding scaffold was printed to fill the gap.

The device itself is quite impressive. They are able to print on a $1\mu\text{m}$ level. In their proof of concept runs, they printed scaffolds that resemble different modes of SCI. Notably, this is meant as an almost complete replacement for the spinal cord—i.e., it requires a very invasive surgery to implant and the similar removal of leftover nervous tissue. The ones they used were 2mm in length and made of polyethylene glycol-gelatin methacrylate (PEG-GelMA). Fortunately, this paper had some of the most in-depth controls, including comparisons to stem cell only injections and scaffolds made of agarose. Another key control was that the implantation did not worsen inflammation. However, at the inflammation site there was still a reactive astrocyte layer. While it was less prominent in the PEG-GelMA scaffold than in agarose, it was still present.

Very interestingly, implantation of stem cells often requires waiting many weeks. This is because dissociated cells like this, which do not have support from blood vessels and are exposed to reactive oxygen and immune hurdles, do not survive well. The group wondered if by providing a supportive scaffold, if implantation can be done earlier after injury—and indeed it could. Acutely implanted scaffold retained their stem cells a month later. It seems like in general, the results were a bit inconsistent. In long term studies they found that some neural progenitor cells became myelinated and the tubes became vascularized. In fact, they even claim that astrocytes' feet lined the vessels, signaling a restored blood brain barrier. Interestingly, the walls of the scaffold thinned by $\approx 50\%$ at the 60 months post implantation mark. It is unclear if this has any meaningful impact, though, as once neural connections are made, perhaps the scaffold itself is not needed.

Onto the part that matters most. Animals with the scaffolds exhibited significant functional recover, reaching a Basso score of $\approx 6 - 7$, compared to the controls reaching $\approx 1 - 2$ at 5 months after injury⁸⁹. They also used an electrophysiological method where they stimulated the motor cortex and read motor neuron outputs, and again found the scaffold to exhibit significantly better outcomes⁹⁰.

⁸⁹This is a sidenote, but geez—it is kind of inhumane to keep a mouse alive with complete immobility for 5 months, solely to ensure it doesn't recover.

⁹⁰As another sidenote, this whole thing is really... odd, in a sense. Of course, the FDA and medical boards cannot allow researchers to do whatever they want. But, these mice had a complete transection and were able to get significant recovery. Surely it can do some good in humans? This is some serious motivation for an enhanced Right to Try bill. If one is completely paralyzed, the option should be open.

Part V

Brain-Spine Interfaces

When we stand at the edge of the ocean, we can not understand its vastness. We know only that it is big. Ittō Ittosai is big. And what about Kojirō?

– Reworded from Vagabond by Takehiko Inoue

20.2 Perspective

There is an interesting part of Yoshikawa's Musashi which details how a vagabond had a poor reputation, despite being an incredible swordsman and a generally giving and selfless person. Their reputation came from those who had long-standing, baseless grudges against them, or those that disagreed with their rise to fame, either out of jealousy or other. A monk in the story comments that a vagabond under 30 should almost never have a pure reputation, as they couldn't possibly be established enough to be beyond critique or ridicule. It is implied that for one to accomplish anything great, it is inherent that one must take risks, be bold, and likely ruffle feathers along the way. Irrespective of one's identity, it is certain that there will be those that dislike them or their approach.

I mention this specifically in this section because it is likely that in advancing technology, you will encounter people who want to belittle your ideas. It is only natural that those who have worked in a field for a long time may be inclined to retain their current way of life, thereby pushing against unseen advancement. The canonical example being those surgeons that resisted the endoscope. I add this to say that it is not unlikely that you will do all the right things, but still find those who disapprove.

Pay them no mind, or better yet, learn from them. Perhaps if you learn from all of those that are unwilling to learn from you: your growth will be unmatched, and real progress can be made. Best of luck.

Papers to Read:

- Neuroimaging guidelines: <https://www.nature.com/articles/s41393-019-0309-x>
- munc13 <https://www.sciencedirect.com/science/article/pii/S0896627321007753>
- courtine in NHP <https://www.nature.com/articles/nature20118>
- Myelin: <https://www.nature.com/articles/s41577-023-00907-4>
- Another scaffold: <https://www.science.org/doi/10.1126/science.abh3602>
- CD8 t cell paper <https://www.science.org/doi/full/10.1126/science.abd5926>
- Growth cone review 1 <https://www.nature.com/articles/nrm2679>
- Growth cone review 2 <https://www.nature.com/articles/nrn3176>
- a third, focusing on mechanics <https://www.frontiersin.org/articles/10.3389/fncel.2015.00244/full>
- literally focuses on glia-neuron growth cone lol <https://www.frontiersin.org/articles/10.3389/fnins.2020.00203/full>

Chapter 21

Interfacing with Nerves

21.1 Historical Dealings

A proverb in a writing called *Testament* by Daitō Kokushi reads:

I beg you, try to find the fundamental source.

...

Like our great predecessors,

Do not merely pinch off the leaves

Or concern yourselves only with the branches.

The field of neuron regeneration feels novel, as investigating its complexities appears to require new age technology. However, greatness spans time. In the 1930s, one of the most brilliant neuroscientists the world has ever known, Ramón y Cajal, suspected restoration of neural circuits after injury required the emulation of developmental conditions. Using electronics to study neural circuits, too, has a long and storied history. One can think of Hodgkin and Huxley, whose ubiquitous achievements evades a need for citation, using an operational amplifier to clamp the voltage of giant squid axons, they revealed the fundamentals of action potentials in the 50s. Or, one can think of Fitzhugh, whose work inspired Nagumo to devise analog circuits to model excitable systems in the 60s. In the case of solving spinal cord circuits, so too have electronics long been used, with examples dating back to the 40s⁹¹. Similarly, stimulating the spinal cord originates from the theory of central pattern generation, first characterized in the 1980s⁹². A large part of what we must do as modern scientists is to not forget the questions posed by our predecessors. While they were limited by technology in their ability to investigate, the validity of their curiosities can never be lost.

Note: These next two paragraphs deserve their own sections. These labs are instrumental. Find the fundamental source!

Restoration of the signals which pervade severance in spinal cord injury through therapeutics has been equally explored. For example, drugs such as clonidine, α_2 receptor agonist, can promote walking in cats with spinal cord injury⁹³. This effect can be blocked completely by yohimbine, an α_2 receptor antagonist⁹⁴.

⁹¹<https://journals.physiology.org/doi/epdf/10.1152/jn.1943.6.2.111>

⁹²https://nyaspubs.onlinelibrary.wiley.com/doi/full/10.1111/j.1749-6632.1998.tb09062.x?casa_token=squhz5qGCxsAAAAA%3A1_eFc1105AQAUdyU1ZoGqmSoCeMnjfpNpvBrnf5NezmG-AwmwujHWj04NNH51HWlrYRurC3H1Tbh8phW

⁹³<https://journals.physiology.org/doi/full/10.1152/jn.1998.79.6.2941>

⁹⁴<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2278596>

The ability to generate walking-like movements from electrical stimulation of the lower spine, the T10-L1 region, has been long known^{95,96}. In these instances, “tonic” stimulation was able to generate step-like movements. Although the term “tonic” is dubious because other papers by the same lab identified frequency dependence in their stimulation⁹⁷. While one can generate step-like movements at once frequency, continued extension can be generated with another.

21.2 Electrostimulation

21.2.1 Background

Stimulating nerves is not straightforward. The amount of current necessary to stimulate a nerve is highly dependent on the location. For electrodes especially close to myelinated fibers, currents as small as $0.1 \mu A$ have been successful in depolarizing neurons⁹⁸. It seems that as a general trend, the faster conduction velocity of a neuron, the lower amount of current is required to simulate it. Too, throwing off one’s measurements is often that you may indirectly stimulate a neuron of interest by depolarizing nearby cells. For direct stimulation, the distance from electrode to the closest Node of Ranvier seems to be essential metric. Figure 3 of the paper footnoted earlier in this paragraph shows a fascinating plot. The current supplied by an electrode resembles something of a double-sided Lenhert-Jones plot. Therefore, for an anode⁹⁹, one would actually expect a core of cells to be depolarized, while those of the periphery would actually be hyperpolarized. The inverse would be true for a cathode. This makes me wonder if oscillating, $\pm A$ has been applied, and the outcomes. If one were to use two electrodes (i.e., an anode and a cathode nearby) then one has the option of longitudinal stimulation or transverse stimulation (i.e., either parallel or perpendicular to the axon).

Note: Perhaps it would be good to go through the different electrode options? I.e., where one places ground, whether current or voltage is applied, etc. Though, I have had trouble finding information on this.

Ramblings.

In a strange sense, the human body can be considered the infinite resistor lattice mentioned in the electronics section.

21.2.2 Biomaterials Used

A primary focus in finding suitable biomaterials are those that are conducting, flexible, and non-reactive¹⁰⁰. Polypyrrole (PPy) is one such option. PPy alone is quite rigid, but is a good accompaniment with other materials. For example, you may coat some glass or fabric in PPy in order to make it more biocompatible and conducting. Too, PPy seems a good candidate for boosting cell adhesion to your biomaterials. Others have used PPy in a tubular fashion, applied some current, and used it to

⁹⁵<https://www.sciencedirect.com/science/article/pii/B9780444521378000188>

⁹⁶<https://journals.sagepub.com/doi/pdf/10.1177/1073858417699790>

⁹⁷<https://link.springer.com/article/10.1007/s00422-004-0511-5>

⁹⁸<https://www.sciencedirect.com/science/article/pii/S0006899375903649>

⁹⁹Fascinatingly, a quick Google search for ‘anode’ provides two definitions, one stating it is positively charged and one stating it is negative. Regardless, you only need to know that it supplies electrons.

¹⁰⁰<https://onlinelibrary.wiley.com/doi/epdf/10.1002/term.383>

guide regenerating axons. Another similar product is polyaniline (often called PANI). Polyaniline can be oxidized and reduced, providing an additional usage. PPy and PANI are two of the most studied conducting polymers, but many more exist.

Carbon nanotubes (CNTs) are another way one can manipulate the direction of growth. Interestingly CNTs appear to be cytotoxic when suspended in media, but strongly promote growth when integrated into a scaffold, keeping them immobilized¹⁰¹. CNTs frequently have diameters similar to axons themselves.

The use of piezo-electric materials too provides interesting avenues of study. Piezo-electric materials being those that convert mechanical energy to electrical energy. An example is poly(vinylidene fluoride) (PVDF). Generally, piezoelectric materials are considered inferior to standard conducting polymers because external control is limited and activation is often too localized.

Ramblings.

The axon-like size of CNTs seems very interesting to me, as they provide a neat route of usage, i.e., as artificial myelination. Perhaps one can develop a CNT with a membrane permeable enough for extracellular signaling to get through, but conductive enough to function like myelin. I am curious if you could produce a multi-layered scaffold, where certain tracts fit axons, certain tracts fit glia, etc, and it can act like repaired oligodendrocytes after SCI.

Another interesting thought is the deep implantation of Piezoelectric materials and then activating them using sonication. It would be neat to develop some kind of material that can be mechanically stimulated at specific sonic frequencies, preventing random stimulation from body movement.

¹⁰¹I am a bit confused by this. Who would ever leave the tubes just floating around in the media? What would be the point?

Chapter 22

Peripheral Nerve Stimulation

Peripheral nerve stimulation¹⁰² has shown promise in treating nerve damage, but naturally is greatly limited by technology. One such limitation is the difficulty in targeting peripheral nerves that may be lodged deep within the body. Too, once reached, one must avoid off-target stimulation, movement of the electrodes over time, and swelling in surrounding regions. Removal of leads after the completion of neurorehabilitation can cause further damage as well, as tissue may become attached to it, resulting in a second injury.

22.0.1 Adaptive, Conductive, and Electrotherapeutic Scaffolds

Adaptive, conductive, and electrotherapeutic scaffolds (ACESs) is one implementation of electrical neurorehabilitation after injury¹⁰³. The primary structure is electrodes embedded in alginate (an anionic polymer, which creates a maleable gel-like structure when combined with water, and is often used for creating molds) and polyacrylamide (a cationic polymer most known for its use in SDS page gels, and can be used to suspend solvents at various stiffnesses)—which can be dissolved, allowing for removal of leads with minimal damage to tissue. Because the electrodes are located within the gel, glial encapsulation is not to occur (or, if it does, it will be at the borders of the gel and not the electrodes themselves)—therefore, stimulation should be un-dampened. Gold particles were used to increase conductivity of the gel.

In rats, in order to test accessible nerves, sciatic nerve transection was performed, and electrodes were used for 6 weeks. Stimulation was delivered every other day for 30 minutes in **anesthetized** rats. It was delivered to each electrode at 20 hz (every 200 ms), 2 mA¹⁰⁴ amplitude, and a 100 μ s pulse width. They used a stainless steel electrode by Plastics One¹⁰⁵. Stimulation was powered using the Intan system¹⁰⁶. As two electrodes were used, their pulses were offset by 100 ms. The two stimulating electrodes were proximal and distal to the injury site, and a third, ground electrode was located at the tibialis anterior muscle. EMG recordings and response to sensory stimulation were used to validate functional success of regeneration, as well as axon counts. To test percutaneous nerve stimulation, they performed a similar procedure on the vagus nerve of swine.

¹⁰²Annoyingly, peripheral nerve stimulation is often abbreviated as PNS, so do not get confused with peripheral nervous system vs. nerve stimulation.

¹⁰³<https://www.sciencedirect.com/science/article/pii/S2666634023001654?via%3Dihub>

¹⁰⁴It is not clear to me why they choose to deliver current rather than voltage.

¹⁰⁵Interestingly, I can't seem to find a product that resembles the one found in their supplemental images. Still it is probably comparable: <https://www.protechp1.com/search?q=MS303&type=products>

¹⁰⁶https://intantech.com/RHS_system.html

22.0.2 Vagus Nerve Modulates Circuits via Acetylcholine

Vagus nerve stimulation seems to be much more complex than you'd expect on surface. One such example is¹⁰⁷. Vagus nerve stimulation (VNS) has shown to enhance motor recovery after nerve injury when paired with rehabilitation. Canonically, VNS is used to treat epilepsy and depression—but new treatment options are expanding, such as recent papers looking into treatment after a stroke. Notably, VNS leads to widespread brain activity, but seemingly had a narrow, targeted effect. This paper identifies M1 neurons as being partly responsible for enhanced motor control following VNS. **Importantly, this was tested only in healthy mice, and not any form of nerve injury.** Their work found that VNS specifically *after* successful completion of a task helped reinforce learning of this task—therefore, it may not work as a therapeutic in the case of total injury, but could help recover partial injuries back to normal levels.

Awake, active mice were stimulated via cuff which wrapped the Vagus nerve. Machinery was mounted on the skull, from which wires would extend. They tested a few different methods, and found stimulation following success to be the only positive difference from the sham. Stimulation was done with 100 μ s pulses at a frequency of 30hz (or, every 33.3ms) a total of 15 times (entire duration summing to 500ms) at an amplitude of \approx 0.5mA. Looking at their figure 1, there are some questions. Firstly, what they decide as early and late in the training days seems arbitrary (i.e., first 4 days is early, second 10 is late). I am curious if you thresholded this differently, if their significance levels would change. Secondly, the VNS stimulation results were significant, but the % increase in task completion from WT to stimulated seems only to be about 10-15%. Similarly, the difference in early vs. late trials seems to be about 10% as well—making me wonder how physiologically relevant this is¹⁰⁸.

To determine what is downstream of VNS, they used a tetrode implanted into the brain to read the response of neurons in the basal formation, as they wondered if it was here that encoded for reinforcement. They found that about 43% of neurons showed some response (either suppression or activation), and about 28% showed an increased activity due to VNS. Now, I'm no expert, but those don't seem like great odds to me. That seems like a coin toss. I didn't see much mention in this paper of trying different VNS parameters. I would postulate that they'd get a much better response in tweaking this. An experiment they could run right away is using the tetrode, and varying peak duration, amplitude, and frequency of stimulation to maximize this increased activity parameter. They could then repeat their behavioral studies and see if it increases success rate. Interestingly, they did something similar in the motor cortex, investigating temporal relationship of activation or suppression of neurons over a range of amplitudes. They found that neuronal activation and suppression occurred at disparate times succeeding VNS. Fascinatingly, to test if this observed interaction was ACh dependent, they injected ACh antagonists and allowed them to act globally—effectively invalidating the entire experiment¹⁰⁹.

As a concluding remark, it seems from their data that VNS may contribute to motor learning reinforcement, but certainly is not the essential driving factor. Similarly, I did not see a “negative control” anywhere in their paper (i.e., a vagotomy, for example) demonstrating that motor learning was severely hindered when this pathway was stopped¹¹⁰. Many of the decisions made in this paper are quite confusing to me.

¹⁰⁷[https://www.cell.com/neuron/fulltext/S0896-6273\(22\)00555-4#%20](https://www.cell.com/neuron/fulltext/S0896-6273(22)00555-4#%20)

¹⁰⁸I.e., would a patient consider brain surgery to have a 10% increase in some motor activity? Absolutely not.

¹⁰⁹I must be missing something, because this experiment makes no sense.

¹¹⁰Perhaps it is buried in the supplements, or done in a previous paper, though.

22.0.3 Hand Movement Recovery via Intrafascicular Stimulation

This group utilized intrafascicular, rather than intramuscular, stimulation as a means to recover fine motor skills in the hand of primates¹¹¹. Of course, Courtine is also on this paper. A drawback to intramuscular surgery is it requires electrodes placed on multiple muscles and is typically quite invasive. Very interestingly, intramuscular stimulation has been found to recruit large efferent nerves before small, contrary to normal function and leading to faster fatigue. In other words, faster fatiguing muscles are activated earlier than expected in this case. Stimulating an entire nerve fascicle requires high amplitude and is prone to overstimulation or recruitment of unintended nerve fibers. Intrafascicular stimulation is an opportunity to resolve this.

Such a method requires first identifying the ideal implantation site and building an electrode design accordingly. They call their device a transverse intrafascicular multichannel electrode (TIME). They developed a model of the monkey's fascicles in order to model recruitment¹¹². They settled on a 16-electrode design, withheld in a polyimide shaft.

The bulk of the paper is stimulating individual or combination fibers to initiate a variety of different grips. The end is when they take it to the next level, so-to-speak, in giving the monkeys some voluntary control. They pharmacologically block the monkey's spinal cord circuitry. They used intracortical recordings to modulate hand stimulating, using what they describe as a simple program. Their design was too simplified to activate complex grips, but did prove to work in simple grasping tasks.

¹¹¹<https://www.science.org/doi/10.1126/scitranslmed.abg6463>

¹¹²I don't know how they did this, but it would be good to learn.

Chapter 23

Spinal Cord Stimulation

A large part of spinal cord stimulation, particularly epidural electrical stimulation (EES), is in fact not neural stimulation directly, but rather CSF stimulation¹¹³. EES requires the implantation of electrodes onto the dorsal surface of the spinal cord. As was identified through this computational model, depolarizing signals spreading through the CSF are capable of triggering large, afferent fibers. Such fibers are most easily accessible at their dorsal root entry zones. Their activation leads to motor neuron activation, or the conversion of an inactive circuit to an excitatory one. The bulk of this investigation is done at the lumbar level, as its focus is on restoring walking after SCI. However, other studies have replicated the idea in hand dexterity when stimulating the cervical dorsal root entry zones.

An unsung hero which operates by similar means is transcutaneous stimulation—done through the skin, making it non-invasive and easily performed. Of course, it lacks the precision of EES, but has been used to promote motor recovery after SCI.

23.0.1 Application to Parkinson’s Disease

In large part, tremors induced by PD have already been well solved through deep brain stimulation and high intensity focused ultrasound. However, this paper focuses on some of the non-tremor symptoms of PD, and the idea that one can use spinal cord stimulation to fix degenerative diseases is generally interesting, so let us learn from¹¹⁴. The premise of this experiment is somewhat surprising, as stimulation is often thought to interrupt signals (as is the case in deep brain stimulation for PD, and SCS for neuropathy). In this case, though, they believe SCS **enhances locomotor signals**. That is, they are not treating PD tremors, but rather are treating PD induced akinesia. Experiments were performed in mice using dopamine-transporter knockout (DAT-KO) as their PD model. They also injected tyrosine hydrolase inhibitor (called AMPT)¹¹⁵ to further deplete dopamine. Dorsal root stimulation was performed using electrodes into the upper thoracic levels of the spinal cord. The greatest locomotor improvement in locomotion was found at 300 Hz stimulation.

Notably, some hospitals are currently attempting clinical trials for spinal cord stimulation’s use in treating Parkinson’s. Some efforts include blocking pain and treating akinesia.

¹¹³<https://www.jneurosci.org/content/33/49/19326.short>

¹¹⁴<https://www.science.org/doi/10.1126/science.1164901>

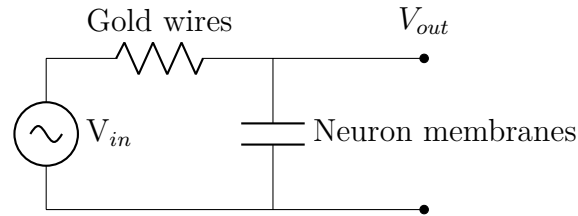
¹¹⁵I didn’t bother Googling what this is, but intuition would suggest it aids in the conversion of tyrosine to catecholamines.

23.0.2 Electronic Dura Mater

A couple years ago now, Courtine and Lacour groups developed a framework for electronic dura mater¹¹⁶. One of the primary problems addressed here is that dura mater is hard and protective while neural tissue is soft—so if one were to remove the dura and replace it with an electronic, it would need to meet both of these criteria. Their “invention” is called e-dura. It is meant to go beneath the dura mater, and is composed of silicone with gold wiring¹¹⁷. The silicone has a similar kPa to the dura itself. Most implant are done above the dura, so the soft nature allows for implantation beneath it.

One of their comparisons is in implantation of soft vs. hard electronics in locomotion (this is without spinal cord injury). Notably, the soft electronics performed the same as the sham—while the hard electronics were much worse. They similarly quantified damage to the spinal cord, and expectedly found that soft electronics did less damage, and similarly that there was more inflammation in hard electronics. They also made a model of the linak cord using a hydrogel to show similar effects—perhaps useful to keep in mind.

I’ve noticed this same thing in a few papers. But, they make the claim that impedance stays constant over the course of a few weeks. However, the error bars show an impedance of $\approx 50\text{k}\Omega \pm 30\text{k}\Omega$. Firstly, this is quite high impedance—but that may be a protective feature. Secondly, and more crucially, the spread is humongous. My wonder is if the Reviewers did not notice or understand this, and saw that the average stayed around the same. Impedance does seem to be quite variable / strange in every paper. So it also may be that controlling this is quite hard. There is also quite a large phase angle, reaching $\approx -30^\circ$. Notably, *in vitro* there is no phase shift nor is the impedance near $50\text{k}\Omega$. The $\approx -40^\circ$ shift is suggestive of this being a low-pass filter at their tested 1kHz . Very fascinatingly, the circuit may look something like this¹¹⁸:



In vitro, you would see this phase shift because there are no capacitors to ground. Let us take this a step further, even. If:

$$\begin{aligned} A_{out} &= A_{in} \sin(\omega t + \phi) \\ \frac{V_{out}}{V_{in}} &= \frac{1}{\sqrt{1 + (2\pi f RC)^2}} \\ \sin(\omega t + \phi) &= \frac{1}{\sqrt{1 + (2\pi f RC)^2}} \\ \sin(-40^\circ) &= \frac{1}{\sqrt{1 + (2\pi \times 1\text{kHz} \times 6\text{k}\Omega \times C)^2}} \end{aligned} \tag{23.1}$$

¹¹⁶<https://www.science.org/doi/abs/10.1126/science.1260318>

¹¹⁷I didn’t realize this until just now, but the increased impedance over time is due to micro-cracks in the gold. Interesting to keep in mind.

¹¹⁸This is why you should read the electronics section.

I estimated R to be $6\text{k}\Omega$ by eyeballing their graphs. I couldn't find any raw data in the paper. Solving this gives us a C value of $3.16 \times 10^{-8}\text{F}$, or 31.6nF . This does seem a bit low, as I believe the standard estimate for a cell membrane is around $100\mu\text{F}$. However, there are likely many factors at play.

I digress. The functional tests they did were using a 3×3 array of electrodes, and the motor cortex of mice were stimulated using channelrhodopsin-2. The e-dura demonstrated the ability to record signals sent from the cortex. They then tried delivering stimulation after spinal cord injury, induced via severe contusion. They delivered serotonergic replacement therapy through microfluidic chambers in the e-dura and delivered continuous stimulation at 40Hz , 0.2ms , and 50 to $150\mu\text{A}$. Naturally, stimulation improved locomotion.

23.0.3 Electronics with Shape Actuation

The goal of this paper is to address the gap in stimulation devices¹¹⁹. That is, one can choose either percutaneous electrodes (i.e., puncturing through the skin and covering a small area) or a large paddle requiring invasive surgery. This paper developed a flexible paddle, which can be rolled up to a diameter of less than 2mm , and injected similar to an electrode but expands upon entrance. Too, the flexible shape is fitted with microchannels. Applying air pressure to such channels causes expansion. It is composed of materials like gold, parylene-C, silicone, polyethylene, and polyimide.

The device is made in two stages, but is monolithic. The first silicone layer is laid, and before it is fully cured, secondary materials are added. Then, the final layer of silicone is laid, and the entire device cures together. To validate it electrically, they used “electrical impedance spectroscopy techniques¹²⁰” Interestingly, the impedance was calculated to be around $1000\ \Omega$! And it was quite variable; their box plots show an IQR of about 600 to $1200\ \Omega$. This is, indeed, not ideal. One concern is how much power will be dissipated by current running through the device, thereby generating heat (and possibly causing device expansion). They also tracked increase in impedance due to mechanical stress, and it seems to increase from about 500 to $1000\ \Omega$ —representative of an approximately new device vs. approximately 6 years old within the body¹²¹.

The device has a diode-like IV curve¹²². They look exactly like a diode would in the positive half (i.e., current exponentially increases at about 0.8V). But the negative voltage range looks quite strange. To apply voltage, they used the Intan device, like in the ACES paper.

The claim is that they are able to implant the entire device using a syringe (presumably under the lamina) into cadavers and then expand the device over the dura mater. Somehow, the how the device is able to spread over the spinal cord while, despite all of the connective tissue and fat filling the lamina-dura space¹²³. To find the desired site of implantation required fluoroscopy, meaning the device itself had to be visible under fluoroscopy. Another confusing aspect of this paper is the biphasic nature of SCS to treat neuropathy. As one always undergoes a trial period, the question is will this device complement the trial phase properly?

¹¹⁹<https://www.science.org/doi/10.1126/sciadv.abg7833>

¹²⁰It is not clear to me what they are referring to, beyond simply driving different currents. There is a paper discussing this [here](#), though.

¹²¹How they could approximate a 6 year old device is questionable to me.

¹²²For lack of a better word, the IV curves look insane.

¹²³All of this is very odd to me. I am having trouble understanding how this is possible mechanically.

The big takeaways, to me, are that expansive devices seem possible, and fitting it with microfluidic channels is a fascinating way to apply therapeutics in addition to electrical stimulation. The next question is, of course, can we do the same thing with ECoG—thereby not requiring a highly invasive procedure to record brain waves?

23.0.4 SCI Specific Cells

Using a simplified device similar to what will be discussed in the upcoming Lorach *et al.* 2023 section, the Courtine group used electrical therapy to re-establish walking in a number of SCI patients¹²⁴. Their paper goes into depth on the cell-type specific response to this electrical modulation. They called this model epidural electrical stimulation (EES) + rehab, or EES^{REHAB}. They first approached by measuring metabolic consumption in the spinal cord before and after EES^{REHAB} using PET scanners, and found that metabolism actually **decreases** in the face of EES^{REHAB}, despite regaining the ability to walk. Therefore, they hypothesized that EES^{REHAB} “cleans up” the circuits a bit, causes activity to become targeted for walking.

To investigate which specific neurons undergo changes, they used a mouse model and performed RNAseq. They found excitatory interneurons as being enriched upon EES^{REHAB}. These neurons were identified by the markers *Vsx2* and *Hoxa10*, and thought to be *V2a* neurons. These neurons projected exclusively to the ventral spinal cord and formed synapses with varying neuron types (glutamatergic, GABAergic, and cholinergic in no preferred proportion). The synapses were found in dense appositions (i.e., many synapses side-by-side) and SCI caused a reduction in such appositions. Ablating/inactivating these neurons alone in non-SCI mice did not halt their ability to walk. However, inactivating them did halt walking under EES^{REHAB}. Therefore, it is hypothesized that these neurons are specific to the injury condition and required for recovery after SCI.

The main takeaway from this paper, in my opinion, is simply the recapitulation that excitatory interneurons play a key role in recovery from SCI. This has been established now through many routes, and is certainly of considerable interest in the future. There are a few open questions, with the biggest being: how exactly does electric modulation drive circuit reformation through these interneurons?

23.0.5 Clinical Applications

Reduction of Pain.

This is from this video¹²⁵. In reducing neuropathic pain, the mechanism by which it works is postulated to be Gate-Control Theory¹²⁶. That is, painful signals can be overridden by other pathways. Namely, fast, large $A\beta$ nerves can override the smaller, pain carrying $A\delta$ and C fibers. It is thought to occur at the level of the dorsal horn through inhibitory interneurons, but this is not known with certainty.

Failed back surgery is the most common need for spinal stimulation of this kind. Stimulation is a good treatment option if the patient’s pain is localized, and no identifiable structural defects are found. That is, if one’s spinal cord is still improperly aligned, stimulation should not be considered. It is essential to attain extensive pre-operative imaging, as structural issues can also lead to hardware

¹²⁴<https://www.nature.com/articles/s41586-022-05385-7>

¹²⁵https://www.youtube.com/watch?v=MULbgftczf8&ab_channel=CongressofNeurologicalSurgeons

¹²⁶<https://pubmed.ncbi.nlm.nih.gov/5320816/>

malfunction. For example, a slipped disc may cause unforeseen warping of hardware. Interestingly, psychiatric symptoms is a common indicator of poor patient outcome. Repeated operations usually indicate less effects as well.

Implanting an electrodes is done in two stages:

1. Trial: A lead is inserted in order to find the site of pain. That is, exactly where to place the electrode permanently is found first in a trial. In the trial phase, patients are usually kept awake.
2. Permanent: After such place is found, a permanent surgery can be scheduled for long-term treatment.

A variety of stimulators exist, ranging from linear leads to paddles. Paddles will require laminectomy by a neurosurgeon, but less invasive leads do not necessarily need this.

The stimulation itself has changed greatly. That is, the frequency, amplitude, etc. has been widely investigated. For example, bursts of stimulation, as opposed to continuous stimulation, has been recently tried. Notably, burst stimulation is usually imperceptible to patients, making it preferable in many cases. It is hypothesized that burst stimulation may inhibit both the physical and emotional aspects of pain.

Regaining Motor Control.

This is from this video¹²⁷. EES is performed on SCI patients following a long period of physical therapy. The physical therapy has two purposes. First, it is to determine whether or not someone can recover at all without surgery. Two, it is to “prime” the cells for rehabilitative EES.

The operation requires complete anesthesia, placement of the electrode on the outside of the dura (i.e., epidural stimulation). It requires a laminectomy. The device they used is the same one as is used in treatment of chronic neuropathies. Importantly, the electrode is not placed nearby the injury site. In this patient, the injury was in the upper thoracic area, but the electrode was placed in the T_{12} area. They also used an EMG with electrodes in the muscles in order to trial and error the placement of the electrode. The entire surgery took ≈ 6 h. The battery pack is also added within the patient a few days later.

Absolutely remarkably, the patient was able to regain some motor control the first time the device was turned on.

¹²⁷https://www.youtube.com/watch?v=OubpMu-2EcE&ab_channel=MayoClinic

Chapter 24

Brain Stimulation

Noninvasive methods like transcranial direct current stimulation and transcranial magnetic stimulation have been done with some success, especially when combined with rehabilitation, particularly in recovery of hand dexterity after injury. It is thought that recovery occurs through growth of descending corticospinal tract neurons.

24.0.1 Living Electrodes

This is from the work by Dr. K. Cullen, and featuring Dr. I. Chen¹²⁸. Part of the problem they are trying to solve is the rigidity of standard, mechanical electrodes. To solve it, they developed implantable cortical neurons embedded within a hydrogel tract. This paper mentioned another drawback of standard, inorganic designs causing overheating of surrounding tissues—this is the first time I’ve seen this mentioned, and I have frequently wondered about it.

The platform they use is called microtissue engineered neural networks (μ TENN). The gist is they make a hydrogel tube, culture neurons in kind of a cluster (or, aggregate), place this aggregate at one end of the tube, fill the tube with ECM-like compounds to promote axonal growth across it, and then implant the tubes where the synapses will be formed at the end opposite of the cell bodies. The columns are $\approx 400\mu\text{m}$ wide, with only the inner $\approx 180\mu\text{m}$ being made up of cells. They are made between 2 and 9mm in length. Action potentials can propagate bi-directionally in some models, or uni-directionally, depending on how it is made. Their experimental setup in this paper, for example, is to use a uni-directional setup with optically activated neurons, and record the induced activity through a second, bi-directional column (that is not optically activated). Channelrhodopsin-2 (ChR2) was used as their optic input, while RCaMP was used as their output reader.

In vitro, interestingly, they show that the max $\Delta F/F$ increases with increased optic power. However, they do not show it levelling off at any point—which would have good to know (i.e., where the point of max stimulation is). They also didn’t show the baseline intensity over time (for more than 20 seconds, that is), which might have been interesting to see. The fluorescence plots over time seem quite variable. That is, RCaMP $\Delta F/F$ is not uniformly affected at the point of each stimulation. *In vivo*, the non-uniformity is even worse. I would assume there is some accumulation of RCaMP fluorescence over time, as repeat stimulations slowly build up more calcium. Parallely, one might expect the baseline intensity to go down as RCaMP becomes bleached.

¹²⁸<https://www.science.org/doi/full/10.1126/sciadv.aay5347>

There are many limitations to consider. For example, RCaMP's k_{on} or k_{off} , reset time for ChR2, refractory period of neurons, the time it takes to grow neurons (which seems to be closer to a week or more), necrosis within the aggregate, etc. Too, you lack any real understanding of the neuron if using living electrodes as sensors. That is, you don't get much localization or intensity information (if any). Confusingly, from watching their supplemental videos, the potential seem to be wave-like, flowing horizontally across the columns rather than one instantaneous pulse. Perhaps this is due to the k_{on} of the calcium indicator used, but it seems to indicate some strange temporal differences in activation—i.e., the depolarizing of one neuron within the aggregate may cause the depolarization of another nearby that is independent of its intent as a sensor. This would be comparable to a spreading depression one might expect in a migraine, but instead of being within a brain, it is within this cultured cluster.

Ramblings.

What is fascinating about reading papers like this is I go into them every time thinking “Wow, this method sounds remarkable! Perhaps this will be the secret to next-gen BCIs.” Yet, I always finish the paper with more questions / critiques than I began with.

Chapter 25

Reading Thoughts

At present, I am unsure if reading one's thoughts is a requirement at all for overcoming SCI. I am unconvinced that electrical stimulation controlled by one's brain is better than electrical neurorehabilitation—or if it is better by any measurable margin. In investigating this, I believe the limiting factor in bridging from the stimulation applied by Tator and Minassian is a reliable way to read one's objective through their brain waves. The Lorach *et al.* 2023 solved this using the WIMAGINE system¹²⁹. But, let's begin by discussing the options and a bit of history.

25.0.1 Different Reading Devices

As you would imagine, there is a tradeoff between the resolution and the damage you will confer upon one's brain. This is described well by Schwartz in 2006, so notably technology may have advanced since then¹³⁰. The four classes of devices are:

1. Electroencephalography (EEG)
2. Electrocorticography (ECoG)
3. Local field potentials (LFPs)
4. Single Unit AP

EEGs, naturally, are non-invasive and have a reading range on the order of a few centimeters. ECoG is on the tenths of centimeters level, while LPS is millimeters level. Single unit AP is as the name describes, and reads individual action potentials. EEGs sit far above the cortex, outside of the skull, and ECoG come in contact with it. Both LPS and single unit require the electrodes to be buried within the cortex itself, thus causing damage to the brain tissue. Therefore, EEG is typically preferable, as it does not require surgery. It is highlighted by Schwartz that many negatives can occur long term from invasive procedures, such as degeneration, volume displacement, or glial encapsulation. Therefore, it is for this reason that I presume the ability to read one's objective was the limiting factor in progressing the field.

For the moment, I will avoid describing individual technologies between the four groups, as it is likely that the tech progresses greatly each year, and it would be pointless to attempt to describe the reading ranges of any without being totally up-to-date. Instead, we will discuss the WIMAGINE in depth here.

¹²⁹<https://www.frontiersin.org/articles/10.3389/fnins.2019.00847/full>

¹³⁰<https://www.sciencedirect.com/science/article/pii/S0896627306007264>

Soft vs. Hard Electronics.

The importance of the electronic's physical form has been considered by the Courtine group as well¹³¹. In their case, they looked specifically at the electrode paddly inserted into the spinal cord. Notably, they did not use this device in their recent, Lorach *et al.* 2023 paper. One problem I personally find is that in using hard electronics, one immediately resolves to doing massive cranioplasty in order to add their device. If one goes the soft-only electronics route, perhaps you can insert the electronics beneath the skull itself.

25.0.2 Reading LFPs Without Surgery

A very recent article devised injectable electrodes for reading brain activity¹³². It is injected through blood vessels, and requires vessels of $\approx 100\mu\text{m}$. For reference, the smallest arteries are around $\approx 100\mu\text{m}$, and arterioles below this threshold. Therefore, this is usable in a wide range of vasculatures. Such a device is considered a micro-endovascular (MEV) probe. Other attempts at this have been done in the past, but a primary limiting feature is that metal electrodes are inflexible and thus cannot be navigated through the array of vessels in the brain. One of this paper's primary advancements, then, was making it ultra-flexible. I'm actually having trouble understanding how many electrodes the device has—in some graphs they show 16, in others only 5. It seems that they too use an Intan device to read and process data.

The device is injected through a microcatheter, which itself has a much wider diameter. Therefore, the location of implantation is limited in that it must be accessible by a wider vessel nearby. In their example, a large vessel branched off into two directions. To select between the two branches, they had to change the mechanical properties of the device itself¹³³. My understanding of their method of inducing this was by stressing part of the mesh making up the device.

Their testing was done in anesthetized mice in both wildtype and disease models (such as induced seizures). In such circumstances, they were able to record local LFPs. In some instances, though, they were able to get single AP recordings on individual electrodes. Interestingly, this paper included both the filtered and unfiltered readings, providing some nice insight into the benefit of bandpass filtering!

Some confusions: I'm not sure exactly if this is meant to be a permanent addition or not, or if it is removable atraumatically. Similarly, I'm not understanding how they are actually reading from the device. Their setup in the supplements seems to require anesthetizing the mice and doing surgery to attach the electrodes to the Intan device—which I assume cannot be true, lest the whole purpose is defeated. So my wonder is: are the mice walking around with wires hanging from their skulls, or are the wires inaccessible without surgery? I should probably read more closely to understand.

NOTE: A K. Cullen paper cited in the above article¹³⁴. Read later.

¹³¹<https://www.science.org/doi/10.1126/science.1260318>

¹³²<https://www.science.org/doi/10.1126/science.adh3916>

¹³³This is likely a considerable limitation. It is unlikely that this sort of device will ever become generalized or patient specific. However, its use in research might be considerable.

¹³⁴<https://www.frontiersin.org/articles/10.3389/fnins.2019.00269/full>

25.0.3 Neurograin

Neurograin (a type of ECoG) is an interesting development that includes both the ability to read and stimulate the brain from small, rice-grain sized chips¹³⁵. Neurograins are individual pieces, providing the immensely powerful ability to scale your device. They verified their method using 48 electrodes, but suggested the ability to scale to up to 770—providing quite impressive resolution.

The core achievement of this paper is in developing microchips of less than 1mm in size, that are able to read and transmit information. However, I am slightly confused by the intended use. In their Figure 1, they show the neurograin array beneath the dura, and a wireless power/download port wirelessly connect from above the skull. It isn't clear to me how they intend to use this—as they'll need to remove the dura in order to implant the device. A possibility application would be to make a network that can be “woven” into the spinal cord, allowing you get to readings with minimal (if any) laminectomy.

One large question I have is how they are processing this data. In the case of a PET scanner, which has a similar 1GHz+ sampling, instead of overloading the computer with billions of data points a second, it has a 2-step sampling approach, where data is discarded unless it is “non-zero.” That is, data is transmitted only when light hits a receiver. In this case, I'm not sure if that is possible, since brain waves will not be binary.

The bulk of the paper is very electronics heavy, so it will be good to reference in future builds.

25.0.4 WIMAGINE ECoG

As the name suggests, the device requires the skull to be torn away and must sit directly upon the dura, above the motor cortex. It features 64 electrodes, and a human patient would need to have two implanted in order to read both sides of one's cortex. The device lacks a bit of the spatial resolution, as it does not directly touch the brain—but it is surely an improvement on current clinical devices.

The device was tested longest in sheep, which was a 10 month trial. Surprisingly, at the end of the 10 month trial, through GFAP staining, they still found a great deal of glial migration/build-up around the site. They did not quantify this, but in my personal opinion, this does not bode well for our patient—as they will have to spend a lifetime with the device. Secondly, calcification of the skull over the device had begun. This is not inherently surprising, and perhaps is a good sign that the bone was not irreversibly damaged. However, as the device is wirelessly charged, and wirelessly transmits brain information, one must wonder how years of calcium buildup may impact the ability to send or receive information. Too, if the technology needs a dust-up, the surgeons will have to re-destroy these calcified layers. There is quite a lot of fuss made about the fact that the device is wireless, and how it was a decision made to best serve patient comfort. But, in my opinion, it may spell out more long-term harm than good. It is not clear to me yet whether the 8×8 resolution of the electrodes is sufficient to delineate all necessary information. Though, this is the trade-off one must make in choosing between EEG, ECoG, and LPS. The WIMAGINE system digitizes information with 12 bits of resolution. Data is processed through pwelch spectral analysis. From what I can gather, this is simply another implementation of the FFT.

¹³⁵<https://www.nature.com/articles/s41928-021-00631-8>

Chapter 26

Synthesis

The purpose of synthesizing the aforementioned parts together to create a brain-machine interface is so that you may skip over the injury altogether, and transmit from brain to body.

26.1 Lorach et al. 2023

26.1.1 Overview

This is the most recent iteration of an ongoing project by the Courtine group. The group used surgical implantation into the skull, over the motor cortex, to record patient movements and wirelessly transmit this to the spinal cord. The technology used to capture these electrocorticographic (ECoG) signals is the WIMAGINE. The design features two antennas, and the second of which transmits motor signals that are to be decoded and sent to a pulse generator. The pulse generator is ACTIVA RC, the same that is used in deep brain stimulation (DBS) or pacemakers in the heart. There is no mention of having the patient replace this in future years, but my assumption is that a new battery will require a new surgery in years to come.

26.1.2 Shortcomings

The three shortcomings explicitly mentioned in the **Introduction** are that **(1)** one using this BSI must have motion sensors on in order to compensate (a sort of PID controller, perhaps), **(2)** that the patient's movement was not perceived as natural, and **(3)** that there was still considerable hurdles in traversing variable terrain. Many more shortcomings exist than this, and we will go through them below.

Motion Sensors.

I suppose it is obvious why one would use a motion sensor rather than an EMG—as it would be too difficult to discern whether the motion itself, or strength of contraction, is too high using an EMG alone. Still, though, I wonder if this is a feasible alternative. Perhaps one can simply predict motion based on EMG patterns.

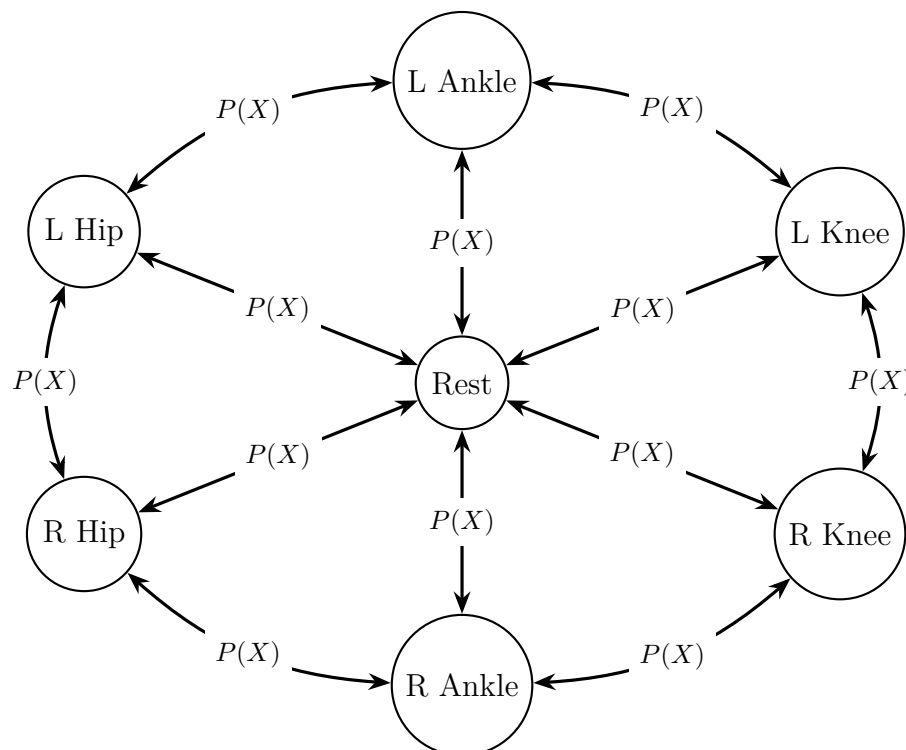
The largest drawback, in my opinion, is their desire to make it quickly programmed/calibrated. As with DBS, one would actually prefer longer calibration time. That is, one would be happy to wait a few extra hours or days if it makes their ability to walk improve.

Similarly, they commented that the patient tested was able to walk without the BSI turned on (using crutches) after some time. This signals three things to me, **(1)** this patient's injury was not, per se, insurmountably severe, **(2)** the plastic abilities of a damaged spinal cord may be better than we had thought—especially since the patient's injury predated this treatment by 10 years, **(3)** that there was likely some important tuning being done by the brain-spine endogenous connection (i.e., their BSI was not doing all of the work), and **(4)** in some cases, no BSI is likely needed at all, and electrical stimulation may be sufficient. This in itself is a quite interesting, alternative patient state. For example, one can imagine a scenario where a patient's brain-spine communication is deeply dampened, but not entirely ceased by damage. Therefore, one would only need amplification of signals as opposed to complete generation of new ones. There may be some of that here. Another comment to make is that their signals are incredibly un-sophisticated, as presented in this paper. That is, they use 16 random electrodes, which do not have connections to synapses themselves. Perhaps this design is to hijack a more reflex-based approach to walking—hence why one can not traverse variable terrain that goes beyond reflexive movement.

A comment regarding the surgery itself: The patient's are having two large, 50mm arrays placed into their skull. Naturally, most would agree to this in order to restore their ability to walk. However, it surprises me that they chose to make the device wireless, given that the surgery is already quite invasive.

Seven States.

Data is processed and used to edit the probability distribution of a Markov chain, which progresses the electrode paddle through seven defined states defined below.



Notably, they did not explicitly mention that one of the progressions could be from state to self state, but one can only assume this is true. So too, only seven states seems to limit the mobility of a patient.

Part VI

Project Proposal

Man takes up the sword in order to shield the wound in his heart sustained in a far-off time beyond remembrance. Man wields the sword so that he may die smiling in some far-off time beyond perception.
– Berserk by Kentaro Miura

Chapter 27

Ramblings

This section will largely be messy—likely for the coming years. It is for me to jot down ideas I have through reading.

27.1 Idealized World

Regarding ECoGs.

Better ECoGs are really the path to incredible options. Of course, we have the mechanics already. Theoretically, one could simply build a mech-suit to accomplish any task—the trouble is getting it to respond. Somehow, it seems certain groups are able to resolve speech patterns and handwriting from brain data—so certainly it is within the realm of possibility. In this same way, ECoGs must become less invasive. The route to accomplishing this is likely through extremely intelligently made soft electronics. The resolution must be greatly improved by the inclusion of many, many more electrodes. In this way, powerful machine learning will be needed to process this data¹³⁶.

There are a few things to consider. First, glial encapsulation will probably destroy you if you try to use too small of electrodes. Second, hard electronics seem to have many drawbacks. If you are using ultra-micro electronics, they are still limited by their ability to read within the gyri, given that the electronics will be hard. Similarly, you simply cannot afford to do massive cranioplasties on a regular basis. It'll need to be improved. Accompanying this issue is the ability to test on humans. One must first test on mice, which parallelly means that the device must be scalable.

Some Experiments.

1. Inhibit L-type VGCCs—likely need to use some kind of hydrogel with antagonists built in.
2. Optimize stiffness of paddles—or hydrogel, if we use the ACES system.
3. Optimize ECoG—that is, take up less space, maintain resolution, etc. to minimize invasiveness of surgery.

¹³⁶I should really start teaching myself ML.

Inhibit L-type VGCCs.

It is well known by now that electrical stimulation combined with rehabilitation improves patient's ability to walk after SCI.

ACES Method.

I like the ACES method as a starting point. It seems easy enough to implement, and can be built upon to arrive at some SCI method. Of course, I like that it is already integrated into a hydrogel setup. Some mild concerns are the 2 mA applied. It seems like quite a bit.

Biomaterials.

The biomaterials possibilities are quite cool too.

CSF-cNs.

A great, easy experiment is CSF-cNs ablation with EES, compare vs. WT, and see if recovery is similar.

Waveforms for Neuropathic Pain.

Perhaps one can use a feedback mechanism, where the frequency of stimulation can be altered in order to dampen neural signals. You'd need to both read and write, essentially. Surely you can record brain waves responsible for feeling pain, and have a circuit edit itself to minimize them? — > actually, dampening neural signals is a great example of a time to use the peak detection method, because you're only interested in magnitude.

27.2 Device Idea

27.2.1 Motivation

Much is made of the frequency and amplitude of neuropathic pain stimulators. Dr. Ben-Haim even called the search for such as being the “waveform revolution.” Perhaps one can avoid this investigation altogether by simply connecting the entire pain pathway to the feedback loop of an opamp. That is, an opamps v_+ is grounded, the V_{out} is tied to a electrode in the spine, and the v_- is connected to some form of ECoG in the brain. Theoretically, the opamp's V_{out} will do whatever it can to cause the signal read at the brain to go to 0. Of course, you cannot assume the opamp's feedback will be predictive, nor can you assume it will fix sharp pains. But, it may be able to fix chronic pain. Let's think of an overall build.

27.2.2 Framework

Processing Brain Waves.

To process the brain waves we will likely:

1. Filter out background from individual electrodes.
2. Subtract and amplify from some reference.

3. Do some summing amplification from each of our electrodes.
4. Rectify our signal.
5. Convert via peak detection.

Feedback.

To implement, we will then:

1. Feed processed signal to a PID controller of some sort.

The only thing I am not sure about is if we will need a separate PID controller for each electrode in the spinal cord. I am not sure if we can just take all of the brain signals and combine them together, but I would assume that is fine. Another question is how much current we will need. If too low, we can add a push-pull to boost, but I doubt we will need a high current.

THANK YOU.

A FINAL NOTE:

The world is always full of the sound of waves.

The little fishes, abandoning themselves to the waves, dance and sing and play, but who knows the **heart** of the sea, a hundred feet down? Who knows its depth?

– Musashi by Eiji Yoshikawa

Part VII

Addendum

Overview

Hello there! You've reached not only the end of the book, but actually the parts beyond its conclusion! At least for the moment, there are topics I'd like to discuss, but they're only tangentially related (if at all) to the idea of brain-machine interfaces. Therefore, consider this the bonus parts that are not worth reading, unless you are immensely inclined.

One such topic is that of neurodegeneration. Naturally, there are not many direct parallels to be drawn from diseases like Alzheimer's and spinal cord injury. Yet, I do believe strongly in Musashi's epithet, that once you know the way broadly, you can see it in all things. For this reason, I think there is unknowable value in learning about degeneration. As it turns out, there are considerable parallels between the pathologies of traumatic SCI and neurodegenerative diseases, like degenerative cervical myelopathy (DCM) as covered in other sections like **16.3**. Thus, it is certainly worth understanding in-depth. Oh, and I because I am taking a class on it and I need somewhere to take notes ;)

Neurodegeneration

Alzheimer's

Alzheimer's disease (AD) is a degenerative disorder present in many demented, elderly patients. It must be diagnosed after death because of considerable symptom overlap. Diagnosis after death is on the basis of extracellular amyloid plaques and intracellular tau tangles.

Holtzman Lecture

This is from a video featuring Holtzman called Sleep & the Pathophysiology of Alzheimer's Disease¹³⁷. To start, dementia is considered a decline in cognitive abilities sufficient to impair daily life. Approximately 50% of people over the age of 80 have some form of dementia.

The role of glia has been understated for a long time in the progression of AD. It seems that Tau and β -amyloid buildup activates a glial / autoimmune response. Like most neurodegenerative diseases, symptom onset marks the end stages of it. Pathology beginnings occur far before symptom onset.

APP is enriched at synapses in the brain. Interestingly, endocytosis occurring at the membrane leads to membrane bound APP being taken back up by vesicles into the synapse. It is here that, while in endosomes, APP is cleaved, leaving behind its β unit. These are then exocytosed. This can occur both pre- and postsynaptically.

The $A\beta$ peptide is somehow higher in the dark and lower in the light. The $A\beta$ presence is also closely correlated with the amount of lactate, which sometimes serves as a proxy for synaptic activity. $A\beta$ can be measured in the CSF. Chronic sleep deprivation leads to significantly increased $A\beta$ plaque buildup, and the converse was true in sleep induced mice. Essentially, this idea was repeated in humans and was conserved. These researchers also determined that it was due, in fact, to production and release as opposed to clearance (done using labeled $A\beta$). This is quite interesting because ApoE is thought to help clear $A\beta$. Also fascinatingly, $A\beta$ seems to further worsen sleep.

While $A\beta$ accumulation precedes symptoms by many years, tau buildup seems to correlate strongly with symptom onset. Acute neuronal activity increase and chronic increase both lead to buildup of tau. As wakefulness increases neuronal activity, so too does it increase tau.

¹³⁷<https://www.youtube.com/watch?v=zCmngDk9VDU>

Overview by Bonini

Regarding drugs and tools, Alois Alzheimer's used dyes to stain tissues and reveal their histology. Aricept blocks acetylcholine degradation, which can slow symptoms for a short time. Measuring amyloid buildup, or other markers, in the CSF is a tool to determine disease progression and or drug efficacy. Congo red is used to stain amyloid buildup (any β -pleated sheet)—i.e., it is non-specific to β -amyloid of AD. It also seems capable of staining tau tangles. Another method is using (thioflavin T) ThT, which also interacts with β -pleated sheets. This has been used to make a radioactive version called Pittsburgh compound B (PiB) and quantify the amount of plaques. This helps identify dementia years before it begins. This is primarily used to track disease progression.

Trisomy 21 patients almost ubiquitously get AD, and specifically early onset AD—which suggested there must be some association with Chr21. β -amyloid was purified in 1984 by Glenner and Wong, which allowed it to be sequenced early. They then generated an antibody against it, allowing them to stain certain regions of the brain. After the β -amyloid protein was identified, other groups sequenced the surrounding nucleotides and found it to be part of a larger transmembrane protein—which was eventually called the amyloid precursor protein (APP), mapped to Chr21. Thus, it was intuited that this little peptide was cleaved from the larger transmembrane protein.

Goate et al, 1991

Overview / Background

A β was first sequenced in 1984. In 1987, four groups cloned the APP gene.

Previously, Chr21 was found to contain an AD locus. The authors identified a site on Chr21 that seems to segregate, sometimes, with early onset Alzheimer's disease and in the gene encoding amyloid precursor protein (APP), specifically at the β -amyloid site. However, as this mutation does not seem to be ubiquitous, they concluded that there is considerable heterogeneity even among early AD patients. This is an important note, because it was originally believed that the APP gene was not disease causing, as some linkage studies did not find APP to be causative. Though, they've identified two families with this mutation and believe it to be causative in some cases.

Table 1

Table 1 contains the LOD scores for markers along Chr21 against the proposed AD locus. APP clearly exhibited the most likely among the markers.

Figure 1

Figure 1 is a pedigree chart. The majority of it is uninformative, as many of the family members have of one two alleles for the APP gene and surrounding markers. Depending on the allele, family members either do or do not have AD. However, two family members exhibit recombinant alleles at different sites. The first individual shares the same allele at the telomeric end of the chromosome to the affected parent. This suggests the AD causing gene is not, in this family, telomeric to the APP gene. Second, another individual shared the same allele as the affected parent only centromeric to APP, and at the time of the study was unaffected. It's possible this individual would develop AD

later in life. Therefore, based on their current pedigree, APP was the clearest possible disease causing gene, which motivated their Fig. 2 work.

Figure 2

Figure 2A used PCR to identify the site of mutation, and determined a valine to isoleucine polymorphism specifically in the β APP encoding portion. They sequenced the β unit first because it appeared causative in other diseases. Figure 2B used a *BclII* restriction site induced by the polymorphism in order to cut DNA and detect, by the number of base pairs, the presence of the mutation on a gel. That is, a non-affected person will have an uncut sequence of DNA of 319 bp in length, while those affected individuals will have a band at 199 bp and 120 bp. This allowed them to quickly screen for affected individuals in other families, of which they found. While sequencing may take closer to weeks, such enzyme gels can be done in a day. To verify these families were not related, they checked other polymorphic sites and found divergence, suggesting they are not related.

As some final comments, they did not sequence the entire APP gene. Thus, it's possible that the true disease-causing mutation is not within the β encoding portion. They suggest that APP's lack of causative support may be due to previous studies misdiagnosing AD (and or phenocopies) and or AD's heterogeneity. They also made a somewhat silly, in retrospect, suggestion that the mutation makes the APP transmembrane domain anchored more tightly to the membrane. They did correctly postulate that the deposition of β amyloid is associated with a worsened disease, and based on the phenotype of trisomy 21, is somehow dose dependent.

Follow-Up

After this work, many of the labs around the world began sequencing from their brain banks, and a host of mutations was found. There seems to be a correlation between the location of mutations and the site that a protease cleaves APP. α -secretase precedes γ -secretase cleavage, which leads to two fragments—this is not pathogenic. When β -secretase precedes γ -secretase, that is when problems arise! γ -secretase is not hyper-specific to a given site, and thus some variable cleavages too lead to worsened disease.

Sherrington et al, 1995

Overview / Background

Genetic studies identified an AD3 gene as being associated with an aggressive form of AD. AD3, along with ApoE and β APP are the three loci so far identified as being genetically related to AD. Within the defined region is a transcript they call S182, encoding for a newly identified, transmembrane protein. They identified mutations in this new protein that are conserved between families, indicating these defects are causative.

Figure 1

Using a group of familial pedigrees and previously published data, they analyzed a set of genetic markers on Chr14. By analyzing the recombination events, they were able to conclusively narrow the causative region to between two markers. To clone the gene, they started by generating contigs. **Table 1** expands on this, denoting the shared alleles within boxes (i.e., the identical partial halotypes). Two clusters of genes were found to be shared between families of common ethnic origins. This indicated to

them the heritability of the gene, and allowed them to make more direct comparisons as the families will be under similar “genetic backgrounds” so to speak. They analyzed the cDNA clones of this region and determined 151 of them to reflect spliced mRNA. Using RT-PCR from AD affected brains and non-affected brains allowed them to narrow their field of interest down to 7 transcripts. Among these was a novel gene, represented by fragment S182.

Figure 2

Using a Northern Blot, they identified 2 major transcripts and expression in most regions of the brain. Notably, mouse tissue only identified one of the two, specifically the 3kb region—which the authors suspect may be the normal mRNA, and the other the other transcript may represent a more rare, alternatively spliced version.

Figure 3 & 4

Figure 3 shows the sequence homology between mice and humans, and some other features geneticists probably care about. Algorithmic predictions show 7 transmembrane domains, and an example structure is shown in Figure 4. Two other non-polar regions were found that appear unstructured and unlikely to form transmembrane helices. They identified potential glycosylation sites, which they suspect may help order these regions into structures along the membrane. Not shown in a figure directly, but they did comment on sequence homology with *C. elegans* proteins. No great candidates stand out among those identified. SPE-4 seemed to be the most similar, which plays a role in spermatogenesis specifically through a golgi derived organelle formation—thus the authors throw out the possibility of S182 altering β APP or Tau golgi trafficking. Of course, we know in hindsight that this is untrue. A similarly funny hypothesis is that this S182 is the mammalian equivalent of Ca- α 1D¹³⁸!

Figure 5

Within the ORF of this gene was mutations in multiple AD affected individuals. Among families with such mutations, no unaffected family members were ever found to have these mutations.

Strul & Greenwald, 1999

Background

Soon after PSEN1 was identified, another gene was found on Chr1, which happened to be quite similar to PSEN1! We now know this to be PSEN2. Then, a *C. elegans* paper came out in which a worm homolog was found. It was identified through their investigation of Notch signaling. In hindsight, we know that γ -secretase and PSEN are indeed one and the same (or rather, that it is the proteolytic subunit of the complex).

Notch signaling is a key developmental path in many organisms. Notch is especially important in cell-to-cell communication during development, aiding in cross talk between cells and driving cell differentiation. A ligand released by a cell will bind to Notch’s extracellular domain, inducing a conformational change and allowing Notch’s intracellular domain to be cleaved. Cleavage is done by γ -secretase, which cleaves the Notch intracellular domain (NICD). The NICD is trafficked to the nucleus, which it interacts with g inhibitors, turns them off, and allowed genes to be transcribed.

¹³⁸This is funny to no one but myself, as my first published paper was on Ca- α 1D function in *Drosophila*, a cool 28 years after this paper came out.

Overview

Prenesilin (PSEN) is a transmembrane protein known to contribute to AD progression, and thought to play a role in amyloid processing. In *Drosophila*, PSEN modulates Notch activity, which thereby can influence transcription. This paper identifies PSEN as aiding in Notch cleavage, allowing it to exit the membrane and enter the nucleus where it can exert its control.

By this point, it was known that β -secretase is required for APP's conversion to β -amyloid and such cleavage occurs in the extracellular space. Similarly, it was known that γ -secretase cleaves APP in the transmembrane domain. It was shown that Notch cleavage occurs in a similar way. PSEN mutations cause AD by increasing the amount of A β . PSEN knockdown inhibits the activity of γ -secretase. It was not known at this point if PSEN was an activator or a component of γ -secretase. Notably, they specifically comment on PSEN in relationship to an A β 42 (42 aa in length) variant, which is mentioned in other literature.

Figure 1

Figure 1 analyzes the phenotypes of PSEN (PS in flies) knockdown. They generated two truncated alleles, which both showed identical phenotypes. PS⁻ embryos exhibited clustered neuroblasts, which should typically be spread out. Midline cells, thought to be controlled by Notch, did not differentiate as they do in WT. Homozygous PS loss is pupal lethal. Clones of PS⁻ cells causes abnormal wing development, thought to be through Notch's role in dorsal-ventral cell communication. Finally, embryos of PS⁻ flies show normal accumulation / membrane localization of Notch, which indicates that it is not upstream of Notch expression—suggesting a role in Notch signaling.

Figure 2

I'm a tad confused by the exact method they used. My rough understanding is they performed a Gal4 knock-in into the intracellular side of the Notch protein. In the absence of a ligand, and in PS⁻, Notch is not trafficked to the nucleus. This is marked by a UAS, which drives β -Gal expression that they can stain for. My understanding is this: They have 3 lines, which they call N⁺-GV3, N^{ECN}-GV3, and N^{intra}-GV3. N⁺-GV3 is like their WT control, which functions normally. In the absence of either PSEN (PS⁻) or Notch's ligand (Delta, Dl⁻), no nuclear transport occurs, marked by no β -Gal staining. N^{ECN}-GV3 includes a large deletion of the extracellular domain, which essentially decouples Notch from its ligand and makes it constitutively active. Thus, in this case they found Dl⁻ embryos to still have considerable trafficking to the nucleus, while PS⁻ did not. Finally, N^{intra}-GV3 is missing both the extracellular and transmembrane domains—therefore suggesting it will never be trafficked to the membrane at all, causing it to remain cytosolic. In this case, neither Dl⁻ nor PS⁻ prevented Notch trafficking to the nucleus, so all embryos could be stained by the β -Gal.

These experiments led them to the conclusion that PSEN was specifically involved in cleavage of Notch, as opposed to acting as a ligand or aiding in its downstream signaling cascade. They also concluded that PSEN is cleaving Notch somewhere in its transmembrane domain. So naturally, the big takeaway is the suggestion that PSEN may play a similar role in APP processing, producing A β in a similar way.

Ye et al. 1999

Figure 1

To investigate the role of PSEN in Notch signaling, they developed 5 *Df* lines. To verify the phenotypes of these lines are PSEN dependent, they used a variety of DNA rescues, one of which omitted the PSEN DNA and did not phenotypically rescue such fly lines.

Figure 2

Based on their mutants, they identified a pupal lethal phenotype they considered to be similar to a Su(H) mutant, an effector of the Notch signaling path. Using a few different larva wing disc markers (driven by expression of various genes related to wing development), they determined β -Gal expression was missing in their *Df* lines, but not elsewhere.

Similarly, sensory organ precursor (SOP) cells were found to cluster together with limited lateral inhibition, determined by enlarged territories and increased *achaete* and *scabrous* expression (early SOP markers).

Figure 3

They determined, using ELAV staining, that paternal PSN partially rescues neurodevelopmental troubles.

Figure 4

Notch protein localization and expression was similar across their PSEN mutants and wildtype larva. This was consistent between both the intracellular and extracellular components of Notch, and for the Notch ligand, Delta. In retrospect, the finding that intracellular PSEN is unchanged may be considered surprising, since PSEN is now known to cleave Notch. The authors attribute this to the small intracellular quantity that is actually trafficked to the nucleus, and which remains undetectable using immunostaining.

Figure 5

To investigate how PSEN may process notch, the authors used Western blots. Figure 5A suggests the largest, ≈ 120 K piece migrates more slowly in the PSEN mutants, suggesting the piece is slightly larger. Figure 5B shows that among the 120K fragments, the largest one dominates in the PSEN knockout, compared to the WT. Figure 5D & E serves to prove the phenotype is specific to PSEN knockdown, as in the Su(H) mutant background, this phenotype is not observed (i.e., Notch is processed as WT is), and similarly in mutant Notch background the large 120K band is absent. Figure 5F showed that the same fragments are unaltered in other proteolytic knockdown (specific to Notch's processing in the golgi), demonstrating a specificity for PSEN and that the cleavage observed occurs after its membrane targeting. Figure 5G & H verified the intracellular domain is unaffected by the PSEN knockdown. Finally, Figure 5I shows that PSEN knockdown did not affect Delta levels.

Figure 6

Figure 6A shows that inactive Notch confers proneural clustering in both WT and PSEN mutants. Figure 6B demonstrates, using only the intracellular Notch domain, that no SOP differentiation

occurs in both WT and PSEN mutants. Figure 6C shows suppression of most SOP differentiation in a membrane tethered background. It is supposed by the authors that PSEN is not required for membrane trafficking nor nuclear trafficking of Notch. Rather, they supposed PSEN is needed in some activation mechanism that follows ligand binding.

Games et al. 1995

Despite some attempts, no successful AD model had been established. This group used a modified, transgenic APP gene in mice which reproduced many of the AD phenotypes. Such a model drove increased expression in many mouse tissues, but especially within the brain. They consider the promoter they used, PDGF- β , and the familial AD mutation they used (V717F), to be the two prime contributors to success.

Figure 1

Figure 1 serves to validate their model. Figure 1B shows 3 different alternatively spliced forms of APP mRNA. Figure 1C & D show via western blots that the amount of protein is greatly increased in the transgenic mice vs. wildtype mice. Figure 1E demonstrates that A β was indeed being produced too.

Figure 2 & 3

A β buildup occurs only at 6-9 months of age and worsened from there. It was primarily local to the hippocampus, corpus callosum, and cerebral cortex. A β buildup took a variety of forms, including irregularly shaped bundles and other more characteristic, dense plaques. Figure 3D demonstrated that reactive astrocytes often surrounded such plaques via GFAP staining.

Immensely important: tau was not found. It was commented that tau is notoriously absent from rodent tissues, which is surprising because I'd certainly expect mice to have similarly structured tau.

Figure 4

Using more immunostaining, A β plaques were associated with synaptophysin, suggesting their specific localizing to sprouting axons. Staining for MAP-2 in addition to synaptophysin showed plaques disrupting and de-densifying nearby regions. Plaques also seemed to compress the surrounding neurophil (the unmyelinated, cell dense regions of neural tissue).

Tau

To summarize to this point. At present, there is the belief that the ratio of A β 42 to 40 indicates the progression of the disease. This is part of the amyloid cascade hypothesis. All of this work has been focused on the plaques, but no insight whatsoever into the tangles.

Tangled up in Blue

In the 60s, tangles were purified to the point of visual analysis, indicating they were helical in nature and likely two filaments coiled around one another. This gave them the name paired helical filaments (PHF). In 1992, this PHF was tested against tau antibodies and stained with Coomassie Brilliant Blue, showing different forms of tau at variable molecular weights. Normal, wildtype tau was largely

unaffected by dephosphorylation by a generic phosphatase. But dephosphorylating PHF substantially lowers the molecular weight of tau. In short, this suggested PHF is a heavily phosphorylated version of tau. This principle is generally conserved in diseases, where a protein becomes abnormally phosphorylated. Notably, tau mutations are never associated with familial AD—but are associated with other degenerative diseases.

Tau binds to microtubules and is thought to help stabilize them. When phosphorylated, it unbinds from the microtubules, thought to cause axonal regeneration. Notably, tangle burden progresses with the disease more so than does plaque burden. In general, this created a rift among scientists on what the root of the disease was— $A\beta$ or tau.

Lewis et al. 2000

The paper overexpressed mutant tau protein that can be found in frontotemporal dementia and parkinsonism (FTDP). The mice exhibited a wide variety of degeneration linked to the tau protein buildup. This included symptoms even in the spinal cord and peripheral nerves.

Figure 1-4

Their Figure 1 was a composite of behavioral assays, including escape responses, motor defects, and the inability to right (flipping over from its back onto its feet again). Eventually they were unable to walk.

Tangles were found all throughout the body. Notably, the fibrils were irregular in shape.

Jonsson et al. 2012

This paper identifies an APP mutation that protects against AD. Fascinatingly, they claim it also protects against age-related dementia in non-AD patients. This seems very fishy to me.