Introduction

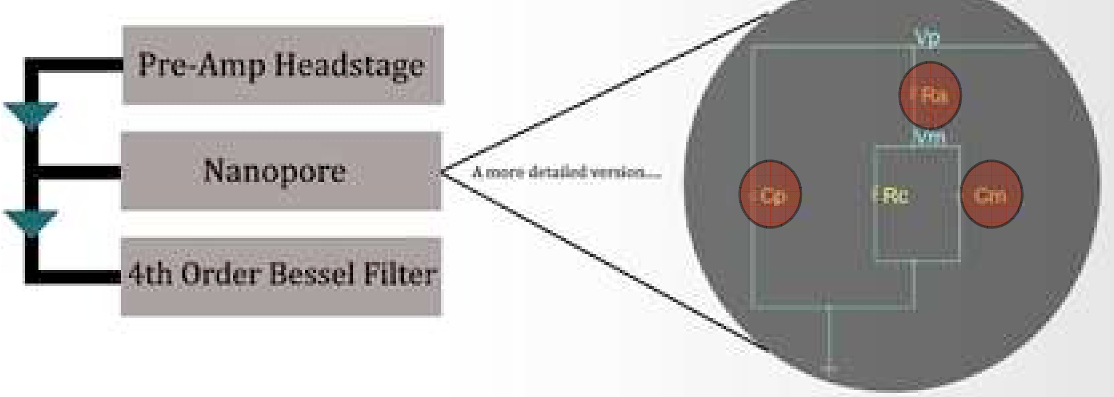
A nanopore can be aptly described as a nano-size pore within two cell membranes. There are a variety of types with an assortment of purposes, but the one worked on this summer is an alpha-hemolysin nanopore, which carries the potential to accurately sequence DNA cheaper and faster then any other existing techniques. The ability to do this rests upon a firm knowledge of the nanopore system itself.

Greater knowledge begets greater manipulation that could hopefully lead to the ultimate goal of DNA sequencing. Unfortunately, the ability to measure some of the values of the biological circuitry created by the alpha hemolysin nanopore is obscured and inhibited by the limitations imposed by its size. In other words, a model of the nanopore system is unable to be attained because of the limitations of measurement techniques. A method to overcome this barrier is System Identification.

System Identification is a methodical process involving mathematical algorithms that build models from data. What System Identification theoretically could yield are values for the inaccurately measured values for the biological circuit. There are two varieties of System Identification - Black Box and Grey Box Modeling. Black Box Modeling has no prior models available, but Grey Box Modeling involves building a model where some parts are unknown, but enough is known to construct a base model. The Nanopore System yields a Grey Box Model.

There are three simple steps involved with the System Identification process - know the system, pick a model, and validate and optimize. The first step involves collecting data as well as knowing as much about the system as possible. The system, shown below in Figure 1, portrays the three units - Preamplifier Headstage, Nanopore, and the 4th Order Bessel Filter - as well as a blown up picture of the Nanopore biological circuit. The full system is not completely shown, only the pre-amplifier headstage and the Bessel Filter are parts of the Axopatch 200B Patch Clamp shown, as the other portions omitted are negligible when performing the System Identification. The values circled in red are the unknown parameters that were estimated by System Identification. Ra is the access resistance from the electrode to the nanopore, Cp is the stray parasitic capacitance to the ground, and Cm is the capacitance across the lipid membrane.

Figure 1: Nanopore System



The second step consists of picking a model to represent the system. There are a wide variety of models one can pick from, but the one chosen was a state space model representation. There are a variety of reasons that contribute to this decision. A state space model can describe the relationship between input, noise, and output signals as a system of first-order differential equations in the form of a state vector x(t), has a great ability to model noise, can represent continuous time, single-input/single-output systems very well, and can easily be represented in computer simulation.

All three of the of the circuit elements produced several differential equations that defined their operations that were used to convert them to state space form. The initial final state space form of the Nanopore system had 6 states, 2 from the Nanopore and Preamplifier Headstage and 4 from the Bessel Filter shown in Figure 2.

4 from the 4th Order Bessel Filter are shown below in Figure 2. “Aux” means that the value is not dependent on any of the circuitry values, but on the value of the frequency of the Preamplifier Headstage, which can be subject to change.

Figure 2: Final State Matrix

$$\begin{bmatrix} -\frac{1}{\tau} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{C_m \cdot R_m} & \left(-\frac{1}{C_m \cdot R_m} - \frac{1}{C_m \cdot R_e}\right) & 0 & 0 & 0 & 0 \\ \left(\frac{1}{R_m} - \frac{C_2}{\tau}\right) & -\frac{1}{R_a} & Aux & Aux & Aux & Aux \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Results

The manner in which to validate the values retrieved through System Identification using the System ID Toolbox within Matlab is to match the actual output with the output generated by feeding real input into the Model. Figure 3 shows this.

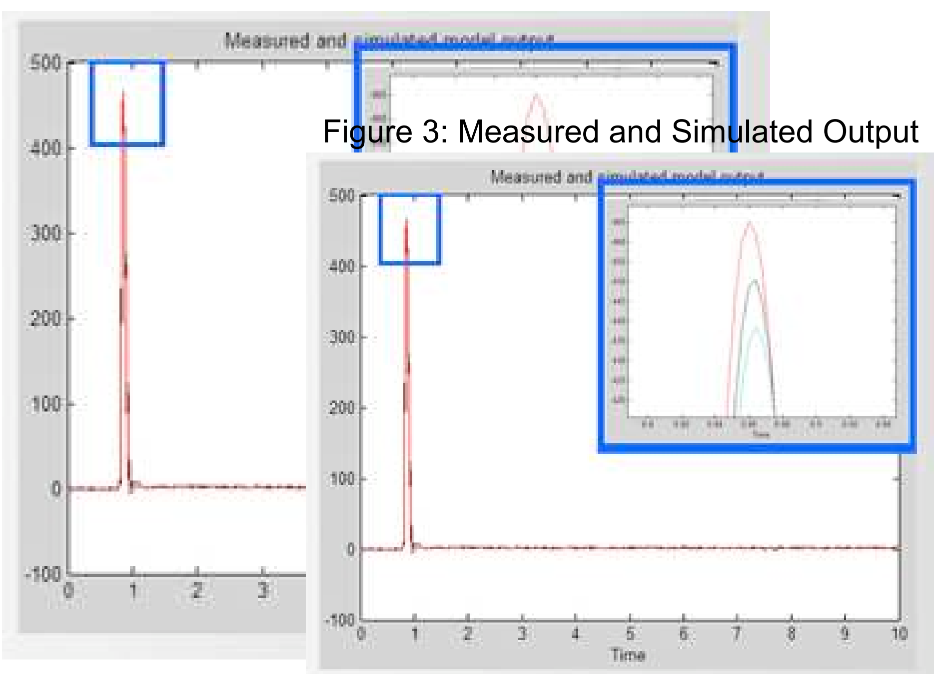


Figure 3: Measured and Simulated Output

As can be observed, the outputs are extremely close to each other. The Black line represents the actual output, the Red (90.07% fit) represents the model formulated using the methods described above, and the Blue (94.57% fit) represents the red model but optimized further using the Perimeter Estimation Method (PEM) within Matlab. The values yielded for the unknown parameters were 0.0116 GigaOhms for Ra, 2.0022 PicoFarads for Cp, and 4.7847 PicoFarads for Cm.

Discussion and Future Work

At first, the results were promising but illogical. Further investigation led to the observation that the time delay was not accounted for within the model. Once this was done, the percentage fits of the Red rose to around 94% and the Blue to 97%. All these figures were drawn from having the Preamplifier Headstage be set at 1 kilohertz. The model was also tested at a 5 kilohertz frequency and the results proved puzzling. The percentages are only assurances of the values, but if the values are illogical something must be accounted for. Cp and Cm's values were literally switched when going from 1 kilohertz to 5 kilohertz.

It was then found that through System Identification, only the sum of Cp and Cm, which accounts for this error. This was found by condensing the transfer function of the Preamplifier Headstage and Nanopore and removing mathematical terms that were rendered negligible, and the result yielded an equation that only contained the sum of the capacitances.

Taking this into account, a 5 State Matrix was made instead of a 6 and two models were made with the unknown parameters numbering 2 and 3 respectively. It was found that the sum of the capacitances were consistently around 6.6 PicoFarads and the Rc was consistently around 3.3 GigaOhms. The third unknown parameter was the time constant of the Preamplifier Headstage, and once again showed consistency at logical figures. These models were tested with both positive and negative square waves and sawtooth waves, and yielded output fits of around 95%, which provides the assurance needed to accept the given values.

Overall, the work proved previous estimates on Rc, the time constant, as well as the sum of the capacitances. Due to the limitations found during debugging on the ability to find the values of the separate capacitances, the need has arisen to found at the very least a value for one of the capacitances and System Identification could then be used to find the other.

Acknowledgements

National Science Foundation
SURF-IT REU Program
Reference: System Identification Theory for the User by Lennart Ljung