A COMPREHENSIVE EXAM SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
ELECTRICAL ENGINEERING

JOHN DICECCO

UNIVERSITY OF RHODE ISLAND
KINGSTON, RI
MARCH 26, 2007

PhD Comprehensive Exam Problems for John DiCecco

March 17–26, 2007

Take-home exam. Open book/notes. You may use library and Internet resources. Please cite all papers/resources that you use. You may not discuss these problems with anyone else other than the professors who made up the questions. Please contact Ying Sun first if you need clarification on any of the problems.

Electrophysiology

Prof. Robert Hill

1. What do you make of "Hodgkin and Huxley model – still standing?" (Nature 4 January 2007, p. v).

2. Write down and look at the GOLDMAN EQUATION. (Goldman, 1943; Hodgkin and Katz, 1949) Is it a problem that there are sharp constraints in the derivation of the GOLDMAN EQUATION? Summarize 5 constraints which may lead one to question applicability of the GOLDMAN EQUATION to living membranes.

3. Discuss and diagram experimental rigs for determining the ionic basis of action potentials.

    a. With an axial wire in a squid axon.
    b. By applying a sucrose gap to a squid axon or another tissue.
    c. By sharp microelectrode recording.
    d. By use of patch clamp measurements in the whole-cell current clamp mode.

4. Invent and discuss your own comprehensive question in the area of electrophysiology.

Signals and Systems

Prof. Leland Jackson

See questions on the next two pages.

Comprehensive Exam for John DiCecco March, 2007

(Open book, MATLAB, computer, notes, etc.) L. Jackson

The discrete Daubechies wavelets are based on a 2-channel FIR filter bank with power-complementary HPF and LPF. Letting $H_{pp}(z)$ be the LP analysis filter, $H_{pp}(z^{-1})$ is the LP synthesis filter, $H_{pp}(-z^{-1})$ is the HP analysis filter, and $H_{pp}(-z)$ is the HP synthesis filter. Let

$$S_{pp}(z) = H_{pp}(z)\,H_{pp}(z^{-1}) \qquad \text{(LP)}$$

and

$$\hat{S}_{pp}(z) = H_{pp}(-z)\,H_{pp}(-z^{-1}) \qquad \text{(HP)}$$

The PR property is then

$$S_{pp}(z) + \hat{S}_{pp}(z) = 1$$

For example, (From Jackson, red book, pp. 314-18,

$$H_{11}(z) = \tfrac{1}{2}(1+z^{-1}), \quad H_{11}(-z) = \tfrac{1}{2}(1-z^{-1})$$

and $$S_{11}(z) = \tfrac{1}{4}(z+2+z^{-1}), \quad \hat{S}_{11}(z) = \tfrac{1}{4}(-z+2-z^{-1})$$

Thus, $S_{11}'(w) = \cos^2\left(\frac{w}{2}\right)$ and $\hat{S}_{11}'(w) = \sin^2\left(\frac{w}{2}\right)$.

It can be shown from Hermite polynomials that

$$S_{pp}(z) = \left[S_{11}(z)\right]^p \sum_{k=0}^{p-1} \binom{p+k-1}{k} \left[\hat{S}_{11}(z)\right]^k \quad \text{from (9.4.15)}$$

1) Find $S_{22}(z)$ and plot the zeros in the z-plane and the squared magnitude $S'_{22}(\omega)$ using MATLAB.

2) Show that $S_{22}(z)$ and $\hat{S}_{22}(z)$ are power-comp.

3) Write a MATLAB program to generate (a causal version of) $S_{pp}(z)$ and use it to generate $S_{33}(z)$. Plot these zeros and $S'_{33}(\omega)$.

4) Factor $S_{22}(z)$ to get $H_{22}(z)$, and factor $S_{33}(z)$ to get $H_{33}(z)$. Can $H_{33}(z)$ be made quasi-linear-phase? Can $H_{44}(z)$?

5) Estimate experimentally (from plots of $S'_{pp}(\omega)$) the order p required to approximate the second-order IIR Butterworth case.

Biomedical Engineering and Nonlinear Dynamics
Prof. Ying Sun

1. Define and compare (strict) stationarity and wide-sense stationarity.

2. Identity at least three different methods reported in the literature for testing the stationarity (either strict or wide-sense) of data.

3. Name three most useful methods, in your opinion, for analyzing the nonlinear dynamics of a one-dimensional data segment. Describe how these methods are implemented. Comment on the strength and weakness of these methods.

4. Identify three significant papers that are concerned with the applications of nonlinear dynamics to biomedical or biological data. Briefly summarize each paper. Comment on why you think each study is significant.

Electrophysiology
Prof. Robert Hill

1. What do you make of "Hodgkin and Huxley model - still standing?" (Nature 4 January 2007, p. v).

The article outlines a difference of interpretation of the waveform associated with action potentials (AP) generated by cortical pyramidal cells in the cat. Of critical note is the location of origin of action potentials in the cortical pyramidal cells. Briefly, the action potential is formed in the axon initial segment (AIS) and not the cell body. How this affects the generation and appearance of APs in the cell body and dendrite, as well as their specific form, is the subject of debate in the article. There are two main points of contention: 1) there is a rapid rate of rise (kink), in the action potential onset 2) there is high threshold variability [1,2]. The existence of neither the kink nor the variability is under debate, but rather it is the explanations proposed for their existence. Naundorf, et al, propose an explanation due to a yet undiscovered intercooperativity of sodium channels in the cell membrane. McCormick, et al, propose that the observations in the waveforms are simply a result of current back propagation from the AIS to the soma. The initiating paper in this contention, written by Naundorf, et al, contains insufficient explanation in the Methods section to ascertain the location of the recording site [3]. That is, the section mentions both sharp microelectrode and patch clamp recording methodologies, however it fails to mention whether the recordings were done at the AIS. Going on the presumption that recordings were made at the soma and not the AIS, as the Methods section suggests, it is likely that Naundorf failed to take into account the effects of back propagation. The axon and the soma certainly do not share the same membrane potential so no

conclusions about the threshold variability in the AIS while recording from the cell body should be made. Further, by taking into account the reduction in membrane potential as the current back propagates from the AIS to the soma, the rapid onset of the action potential seems to be substantiated and predicted by the Hodgkin-Huxley (HH) model [4]. Mathematically, the reduction in the membrane potential as the AP (current) back propagates from the AIS to the soma causes a floor effect (the membrane potential being the dependent variable) which allows the derivative $\left(\frac{dV}{dt}\right)$ to climb quickly, resulting in the rapid onset waveform seen in the papers by both Naundorf and McCormick.

$$I = C_M \frac{dV}{dt} + \bar{g}_K n^4 (V - E_K) + \bar{g}_{Na} m^3 (V - E_{Na}) + \bar{g}_{leakage} (V - E_{leakage}) \quad (1)$$

$$\frac{dn}{dt} = \alpha_n (1 - n) - \beta_n n \quad (2)$$

$$\frac{dm}{dt} = \alpha_m (1 - m) - \beta_m m \quad (3)$$

$$\frac{dh}{dt} = \alpha_h (1 - h) - \beta_h h \quad (4)$$

$$\alpha_n = \frac{0.01(V + 10)}{e^{\frac{V+10}{10} - 1}} \quad (5)$$

$$\beta_n = 0.125 e^{\frac{V}{80}} \quad (6)$$

$$\alpha_m = \frac{0.1(V + 25)}{e^{\frac{V+25}{10} - 1}} \quad (7)$$

$$\beta_m = 4 e^{\frac{V}{18}} \quad (8)$$

$$\alpha_h = 0.07 e^{\frac{V}{20}} \quad (9)$$

$$\beta_n = \frac{1}{e^{\frac{V+30}{10} + 1}} \quad (10)$$

I believe McCormick, et al, have established that the observations can easily be attributed to known physiological models and predicted by the HH model and therefore, the HH model IS still standing.

[1] McCormick, D. A. Shu, Y, and Yu, Y. Hodgkin and Huxley model – still standing? *Nature* (2007) 445, E1-E2

[2] Naundorf, et al. reply. *Nature* (2007) 445, E2-E3.

[3] Naundorf, B., Wolf, F., Volgushev, M. Unique features of action potential initiation in cortical neurons. *Nature* (2006) 440, 1060-1063.

[4] Hodgkin, A. L., and Huxley, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* (1952) 117, 500-544.

2. Write down and look at the GOLDMAN EQUATION. (Goldman, 1943; Hodgkin and Katz, 1949) Is it a problem that there are sharp constraints in the derivation of the GOLDMAN EQUATION? Summarize 5 constraints which may lead one to question applicability of the GOLDMAN EQUATION to living membranes.

If we limit our discussion to sodium, potassium, and chloride ions, the Goldman equation is

$$E_{(Na^+,K^+,Cl^-)} = \frac{RT}{F}\ln\left(\frac{P_{Na^+}[Na^+]_{out} + P_{K^+}[K^+]_{out} + P_{Cl^-}[Cl^-]_{in}}{P_{Na^+}[Na^+]_{in} + P_{K^+}[K^+]_{in} + P_{Cl^-}[Cl^-]_{out}}\right)$$

where E is the membrane potential, R is the ideal gas constant, T is the temperature (Kelvins), F is Faraday's constant, P is permeability (measured in centimeters per second and defined as $u\beta RT/aF$ where u is the mobility of the ions in the membrane, *β* is the partition coefficient between the membrane and the solution, a is the thickness of the membrane, and R, T, and F are previously defined), and the brackets define concentration either in or out of the cell, as indicated [1].

Often, the Goldman equation is referred to as the Goldman-Hodgkin-Katz equation, or GHK equation. The Goldman equation first appears as a variation of the above form in 2 and was later used by Hodgkin and Katz in 3. The fact that the equation supercedes Hodgkin and Katz' work is not a trivial one, since much of the substantiation for their work is based on the relationships established in the Goldman equation.

In Goldman's original work on the subject he wrote *"Obviously radical simplifications will have to be adopted thereby vitiating the results somewhat. Nevertheless rough agreement may be hoped for in some cases and such an*

analysis may also help in clarifying the influence of some of the controlling factors." It is clear that Goldman was well aware that there were constraints that needed to be placed on the equation and placed a level of caution on its use.

The Goldman equation is an equation used to calculate the potential across the cell membrane that accounts for the ionic concentrations inside and outside the cell. Assumptions are made in using the Goldman equation that influence the final result, most notably the way in which ions diffuse across the membrane. The problem in using such assumptions (constraints) is that the final result yields an approximation based on the assumptions. If the assumptions are grossly inaccurate, the equation will yield inaccurate results. Of course this is a problem, but the assumptions are simple and reasonable, and attempt to *model* cellular responses. Like any model, it is not meant to provide an absolute quantification, but rather a reasonable approximation to what is observed naturally. The equation assumes that ions in the membrane diffuse in a similar manner to those in free solution [2,3]. This, of course, is not true given the mechanisms of ion transport and the gating of ion channels and pores in general. This gives rise to another assumption that the electric field (as described by the Nernst-Planck equation) is equally distributed across the membrane [4]. This would be a reasonable assumption if the only contribution to the electric field was provided by the ions, and only the ions. As the number of ions gets very large, the probability that they are equally diffuse approaches one (they share the same probability distribution function). However, it is well known that there are a number of structures embedded in the cell which are electrogenic and those

structures are not uniformly distributed [1]. Another difficulty when attempting to apply the equation is that the equation assumes the cellular membrane is homogenous, i.e. that it is equally probable throughout the membrane that the rate of diffusion is the same, regardless of the location. It is clear to see why this assumption is necessary. In order for the equation to hold, the concentration of ions at the surface of the inside of the cell must be proportional to the concentration of ions at the surface outside the cell. This is a direct consequence of the assumption that the equation is applied under steady state conditions, such that the change in potential over the distance of the membrane is constant (linear) [2]. And here is the crux of the issue. The equation attempts to apply equilibrium (steady state) calculations to a nonequilibrium process [3]. These five constraints (ions in cells behave as ions in free solution, the electric field is equally distributed, the cell membrane is homogenous, the concentration at the inside boundary of the cell is proportional to the concentration at the outside boundary, and that the change in membrane potential across the distance of the cell is a constant) must be observed when applying the equation. More importantly, the equation should be viewed as a model of the electrophysiological properties of the cell, not an absolute description.

[1] Aidley, D. J. The physiology of excitable cells, 3rd edition. Cambridge University Press, Cambridge, UK, 1989. pp 23-25.

[2] Goldman, D. E. Potential, impedance, and rectification in membranes. *J. gen. physiol.* (1943) 27, 37-60.

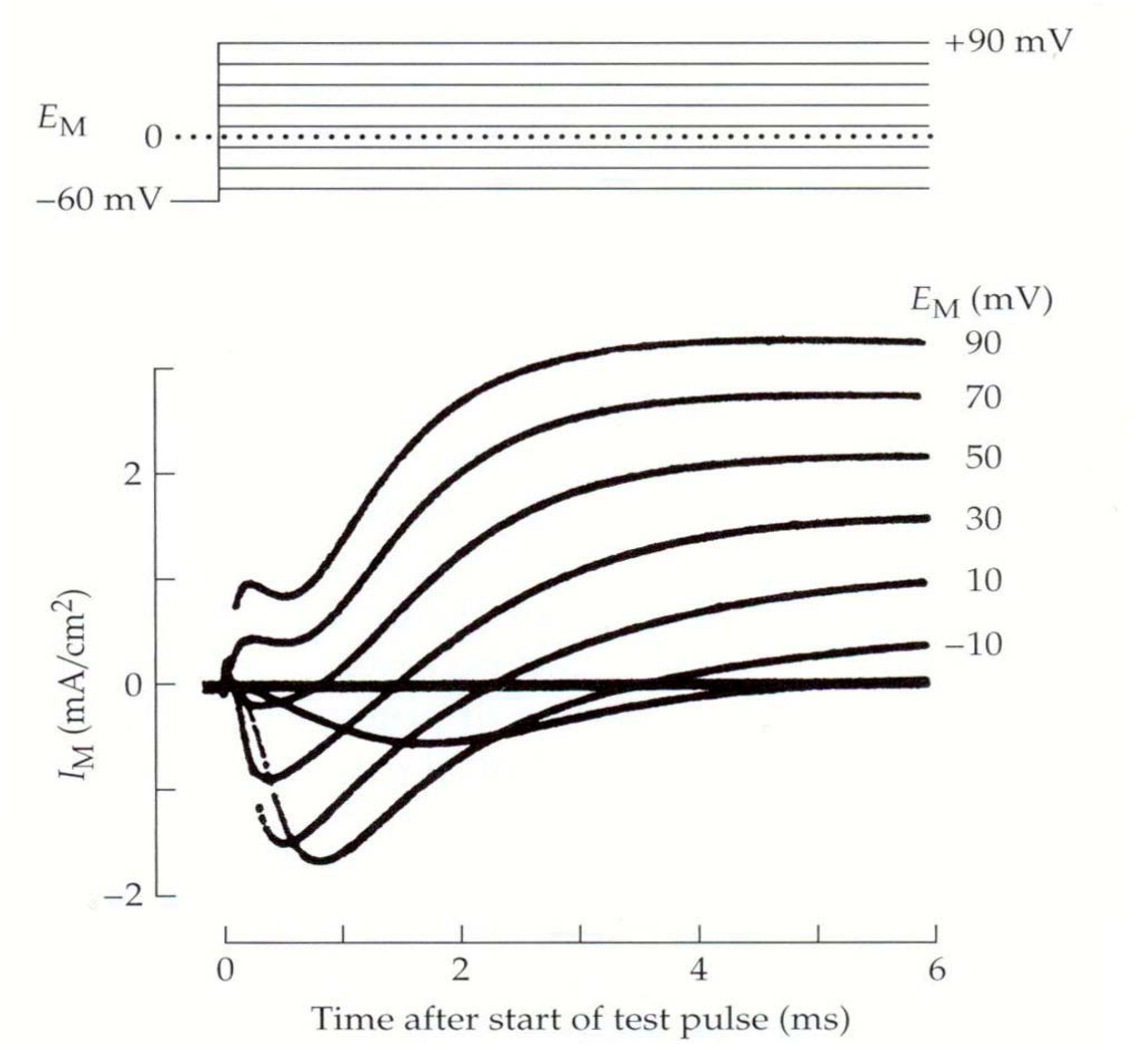
[3] Hodgkin, A. L. and Katz, B. The effect of sodium ions on the electrical activity of the giant axon of the squid. *J. physiol.* (1949) 108, 37-77.

[4] Hille, B. Ion Channels of Excitable Membranes, 3rd edition. Sinauer Associates, Inc, Sunderland, MA, 2001. pp 445-449.

3. Discuss and diagram experimental rigs for determining the ionic basis of action potentials.

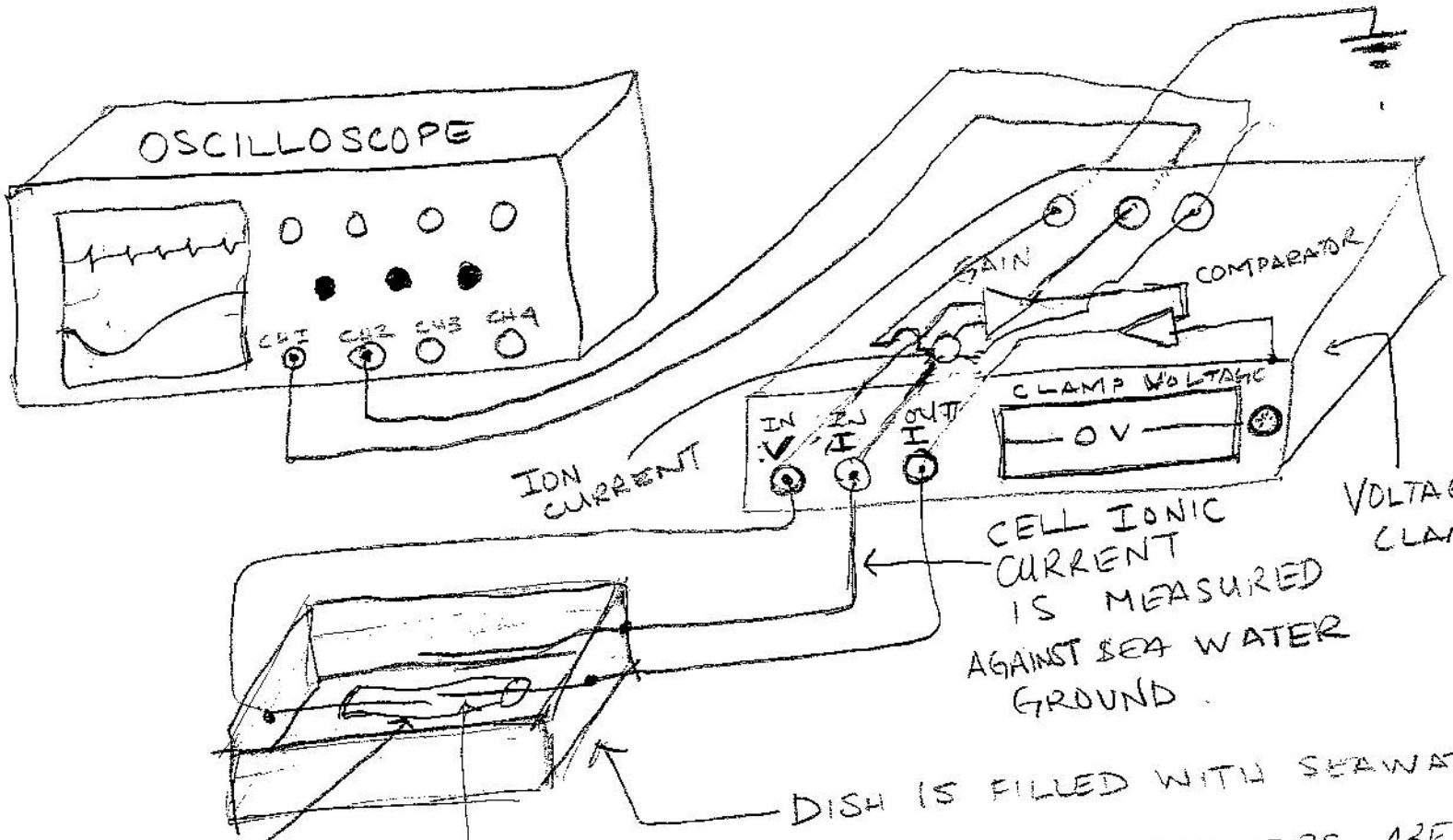
a. With an axial wire in a squid axon.
b. By applying a sucrose gap to a squid axon or another tissue.
c. By sharp microelectrode recording.
d. By use of patch clamp measurements in the whole-cell current clamp mode.

Before a discussion of specific experimental rigs for determining ionic contributions to action potential generation, it is prudent to look at the techniques common all experimental rigs. First and foremost, a primer on the voltage clamp technique is in order. The voltage clamp, developed by George Marmount and Kenneth Cole (independently) in 1949, has been transformed many times over since its inception, but the fundamental principal remains: undesired capacitive currents only flow when the membrane potential is changing and therefore the membrane potential must be held to some fixed value via a network of feedback amplifiers in order to isolate the ionic currents responsible for the action potential. An internal (inside the cell) electrode measures the membrane potential versus the extracellular voltage, usually fixed to ground. The electrode is connected to an amplifier, such as the Axon Instruments Gene Clamp 500, and a reference voltage (clamped voltage) is programmed to the amplifier. A feedback circuit (comparator op amp) compares the measured potential to the reference voltage and amplifies the difference, or error, and passes current into the cell through another electrode in order to minimize the error.



In the above diagram, the membrane is clamped to 0 volts. The membrane is then stepped through a series of voltages through hyperpolarizing (negative) and depolarizing (positive) currents, resulting in changes to the membrane potential. Since the cell membrane is clamped, there are no capacitive currents flowing across the cell. The ionic responses seen above correspond to the cell adjusting its ionic concentration. Downward deflections are representative of current moving into the cell (or membrane) and upward defections represent current moving out of the cell [1].

AXIAL WIRE
AXON



(1) (IN) MEMBRANE POTENTIAL OF AXON IS MEASURED BY CLAMP

(2) POTENTIAL IS COMPARED TO CLAMP VOLTAGE

(3) (OUT) CURRENT PASSED BACK TO MEMBRANE TO MINIMIZE DIFFERENCE FROM VOLTAGE CLAMP VALUE

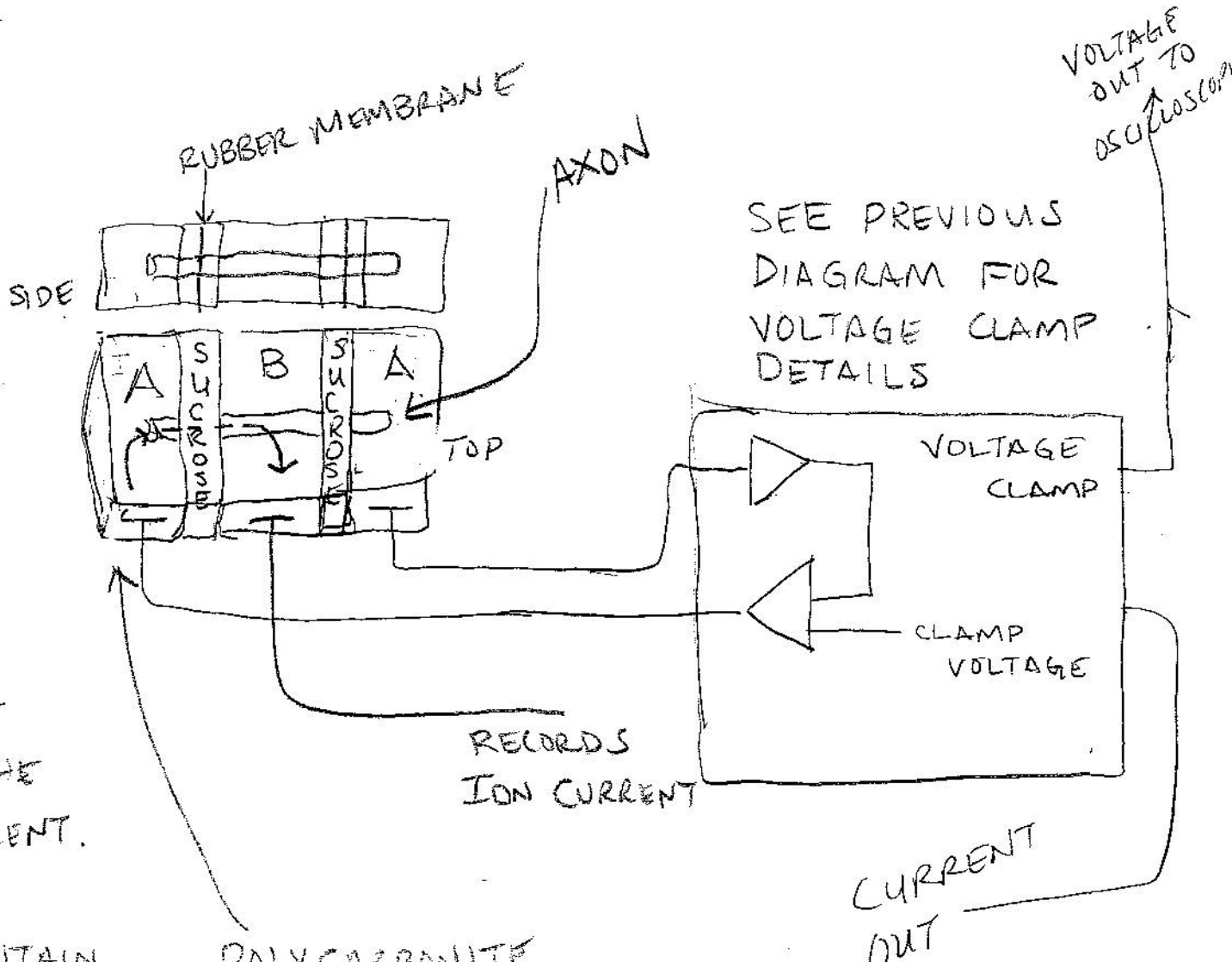
SUCROSE GAP

A ISOTONIC WITH INTRACELLULAR SOLUTION

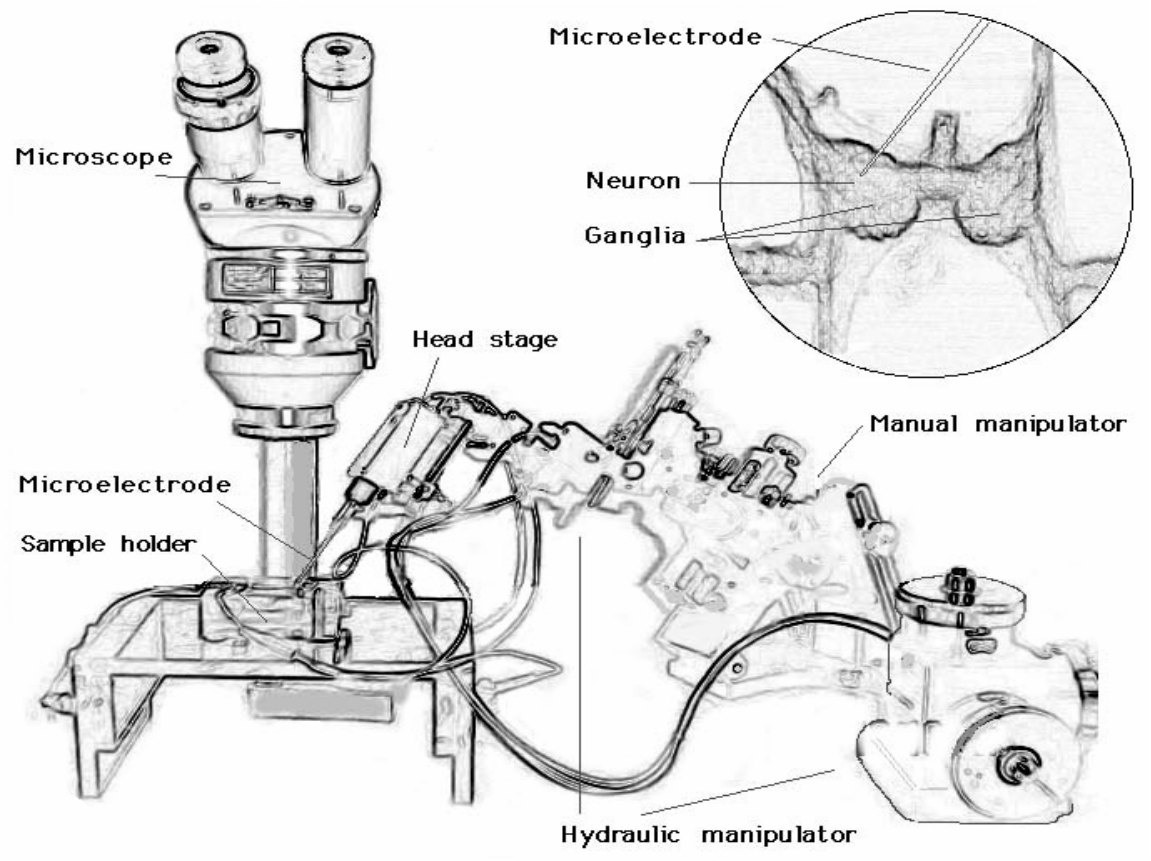
B ISOTONIC WITH EXTRACELLULAR SOLUTION

SUCROSE ACTS AS A BUFFER (INSULATOR) TO FLOW OF IONS. A RUBBER MEMBRANE ALLOWS ONLY THE TISSUE TO CONDUCT CURRENT.

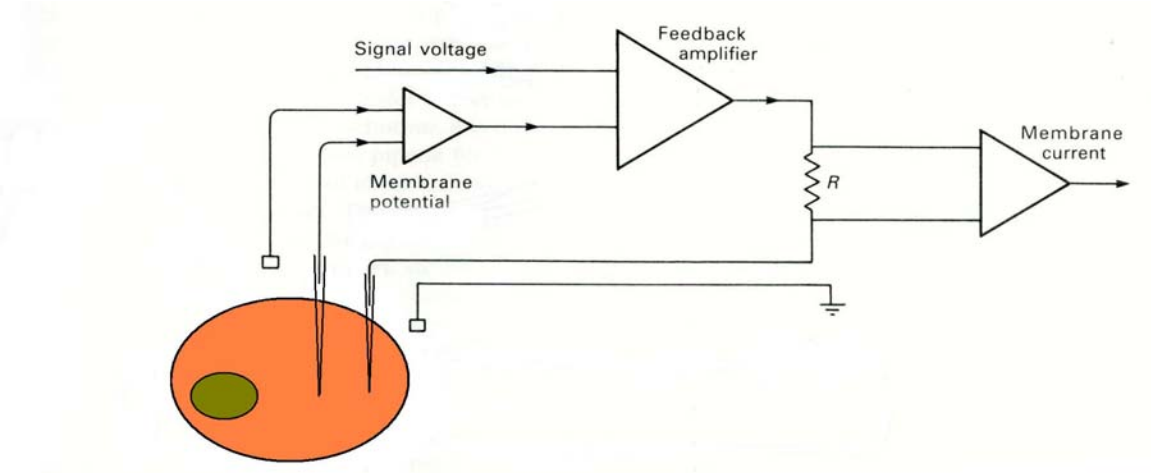
PERFUSION SYSTEMS MAINTAIN STEADY CIRCULATION OF FRESH SOLUTIONS.



Two Microelectrode Rig



A Faraday cage surrounds the entire microelectrode recording to shield the experiment from the 60Hz noise generated by the ambient light and recording equipment.



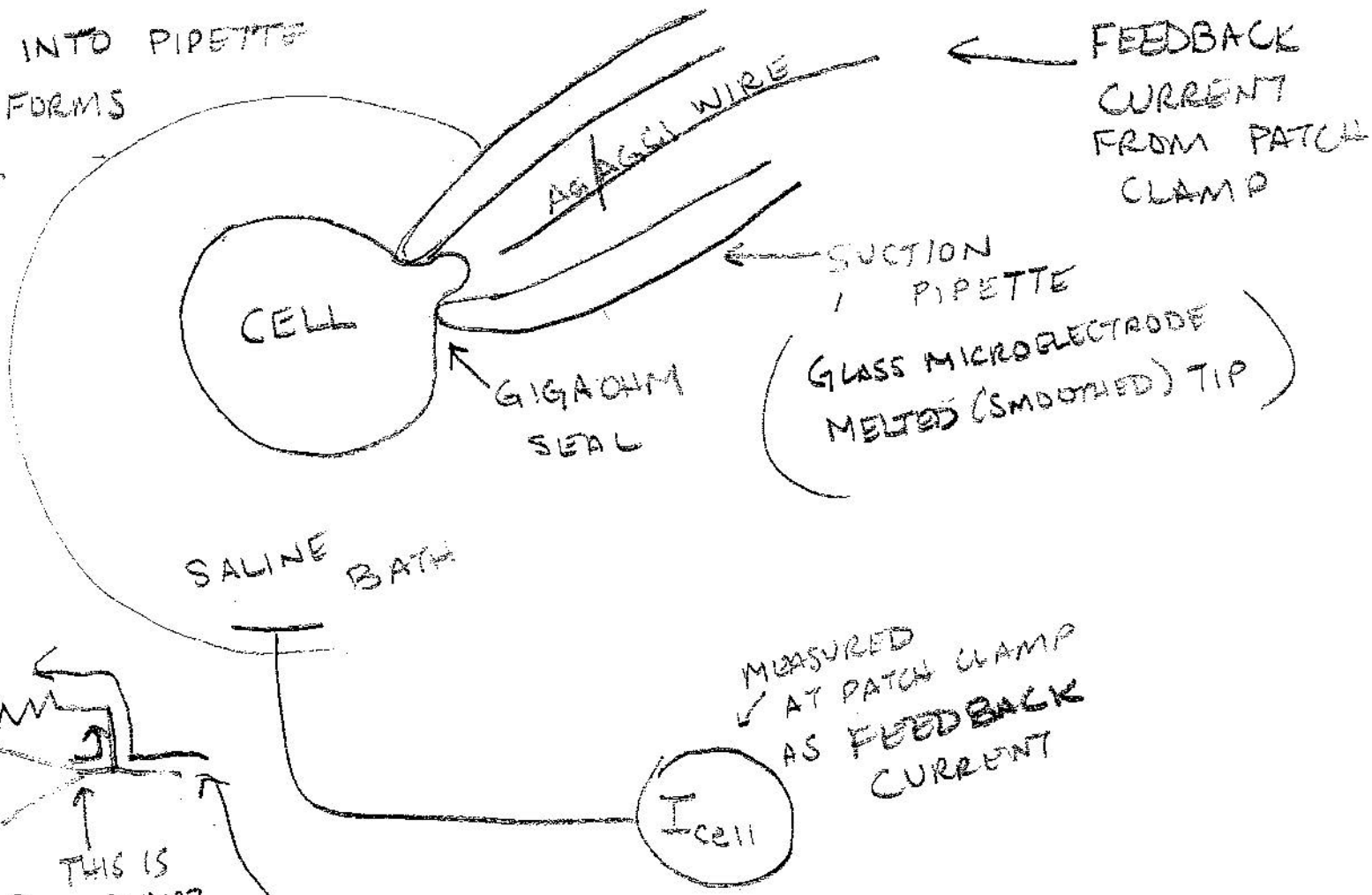
See Axon Rig for Voltage Clamp details.

PATCH CLAMP

(SEE RIG FROM 2 MICROELECTRODE SET UP)

① CELL IS SUCKED INTO PIPETTE (PARTIALLY) & FORMS GIGAOHM SEAL

② FEEDBACK IS MEASURED AS THE CURRENT ACROSS THE CELL DUE TO IONS



A note regarding Silver/SilverChoride wire electrodes. If electrons are flowing (i.e., there is current) from copper (Cu) to a solution, there is an inherently difficult exchange of electrons [2]. We use chloride coated silver (Ag) wire to smooth the transition. Under current, AgCl is stripped of its Cl-, and the Cl- ions enter solution, typically 3M KCl. If the current reverses direction, Ag atoms in the AgCl wire give up their electrons and combine with Cl- in the solution to make insoluble AgCl. This is, therefore, a reversible electrode, i.e., current can flow in both directions. This is a critical point in the ability to interface standard electrical devices in a wet environment.

[1] Hille, B. Ion Channels of Excitable Membranes, 3rd edition. Sinauer Associates, Inc, Sunderland, MA, 2001. pp 445-449.

[2] The Axon CNS Guide. Molecular Devices, 2006.

4. Invent and discuss your own comprehensive question in the area of electrophysiology.

On December 16, 1997, Japanese children had settled in to watch the highest rated cartoon in that nation's history. This particular episode of *Pokemon* featured a bomb that exploded first in white, then in red, and repeated at a rate of 12 times per second. Thousands of children across the country became sick, at least 600 were taken to area hospitals, and one child fell unconscious for over half an hour [1]. The condition is known as photosensitive epilepsy and it is believed, though not known, to be caused by the firing of neurons in lockstep, induced by the strobe of intense light at specific frequencies. The question is "What are the dynamics of cellular interaction that allow a cell, or group of cells, to assume a specific pattern and synchronize?" It is well known that epilepsy is a disease of dynamics and not of structure [2]. That is, there are usually no visible physical defects in brain architecture that cause the electrical activity to transition from normal to ictal. There are instances, however, when a region of the brain is identified as the epileptic focus. When that portion of the brain is removed, it often results in considerable relief of epileptic episodes for a patient. What is fascinating about epilepsy is that even in the same patient, the epileptic event is ectopic, originating in areas of the brain not associated with the epileptic focus. This indicates a low level synchronization of electrical activity that spans regions of the brain. It is this synchronization that interferes with the normal working of the brain and gives rise to the epileptic seizure [2,3]. Such electrical coupling must be able to be observed and described. Many researchers have adopted the analysis of time series recordings to predict when epileptic events will occur.

Researchers use measures of low dimensional chaos and nonlinear dynamics, such as entropy, Lyapunov exponents, attractor reconstruction, etc [3,4]. This does little to solve the issue of how cells synchronize, only to detect when they do or predict when they will. There are at least three states in the epileptic brain, preictal, interictal, and postictal, so intuitively, brain activity transitions from normal to epileptic and back to normal. The dynamics of synchronization that lead to the seizure and the dynamics that restore normal function are caused at the cellular level. Through cooperative neural networks, in an epileptic brain, such synchronizing dynamics cause disastrous consequences. Paroxysmal disorders, such as this, are difficult to fully understand because of the complex nature of the dynamics, whereas the same set of dynamical activity in the normal brain produces no effect whatsoever [2].

The comprehensive question is then “What are the cellular characteristics, that when coupled with other cells in neural networks, produce cooperative synchronization.” In order to answer this question, simultaneous recordings would need to be made from a neural network known to attain a synchronous state. That, in and of itself, is insufficient to contribute significantly to the understanding of epileptic events, since the synchronized behavior propagates through the brain, causing destructive electrical wave fronts [4].

However, it would help to elucidate the mechanisms of communication during synchronized activity which, in turn, could reveal the reasons why some neural networks are susceptible to pathological events while others are not.

[1] Strogatz, S. Sync. Hyperion, New York, 2003. pp 276-277.

[2] Lopes da Silva, et al. Epilepsies as Dynamical Diseases of Brain Systems: Basic Models of the Transition Between Normal and Epileptic Activity. *Epilepsia*. (2003) 44(Suppl.12) 72-83.

[3] Rieke, et al. Discerning Nonstationarity From Nonlinearity in Seizure-Free and Preseizure EEG Recordings From Epilepsy Patients. *IEEE Trans. Biomedical Engineering*. (2003) 50, 634-639.

[4] Sprott, Julian. Chaos and Time-Series Analysis. Oxford University Press, New York, 2003.

PhD Comprehensive Exam Question 2

Professor Leland B. Jackson

For the unity gain first order FIR lowpass filter $H_1(z) = (1+z^{-1})/2$, and $S(z) = H(z)H(z^{-1})$, where S is the z transform of the autocorrelation function of h(n), i.e. r(n) = h(n)*h(-n). This gives $S_1(z) = \frac{1}{4}(z+2+z^{-1})$. Similarly, for the unity gain first order FIR highpass filter $H_2(z) = (1-z^{-1})/2$, giving $S_2(z) = \frac{1}{4}(-z+2+-z^{-1})$.

Evaluating $S_2(z)$ around the unit circle, we have $S_2(z) = \frac{1}{4}\left(-z+2+-z^{-1}\right)$, for $S_2(e^{j\omega})$

$$S_2(e^{j\omega}) = \frac{1}{4}\left(-e^{j\omega}+2-e^{-j\omega}\right)$$
$$= \frac{1}{2}(1-\cos\omega) = \sin^2\omega$$

For perfect reconstruction, $H_1(z)H_1(z^{-1}) + H_1(-z)H_1(-z^{-1}) = 1$. Of course, this is equivalent to $S_1(z) + S_2(z) = 1$, since $H_2(z) = H_1(-z)$, by construction. For the class of polynomials $(1-x)^p$, where $(1-x)^p$ is a lowpass filter, it follows that a corresponding highpass filter can be derived by 1 - $(1-x)^p$. These orthogonal polynomial sequences are termed Hermite interpolation polynomials and can be recursively derived as

$S'_{pp}(z) = [S_{11}(z)]^p \sum_{k=0}^{p-1} \binom{p+k-1}{k} [\hat{S}_{11}(z)]^k$, where $S_{11} = \frac{1}{4}\left(z+2+z^{-1}\right)$

and $\hat{S}_{11} = S_2 = \frac{1}{4}\left(-z+2-z^{-1}\right)$.

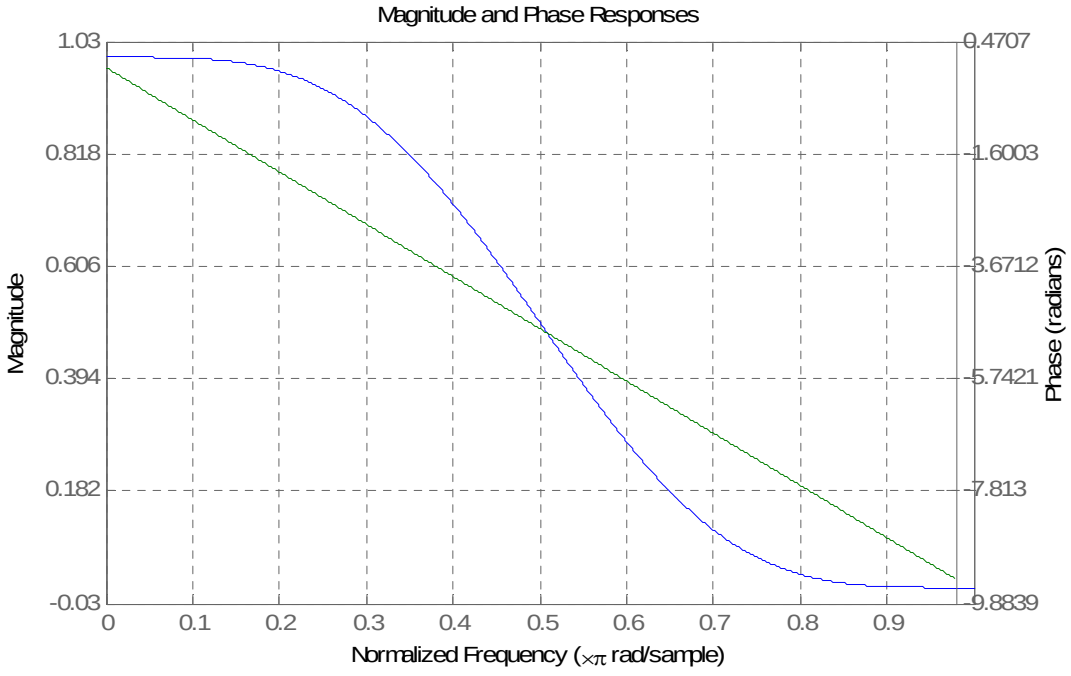
As an example, the derivation for p = 2 follows.

$$S_{22}^{'}(z) = [S_{11}(z)]^2 \sum_{k=0}^{1} \binom{2+k-1}{k} [\hat{S}_{11}(z)]^k$$

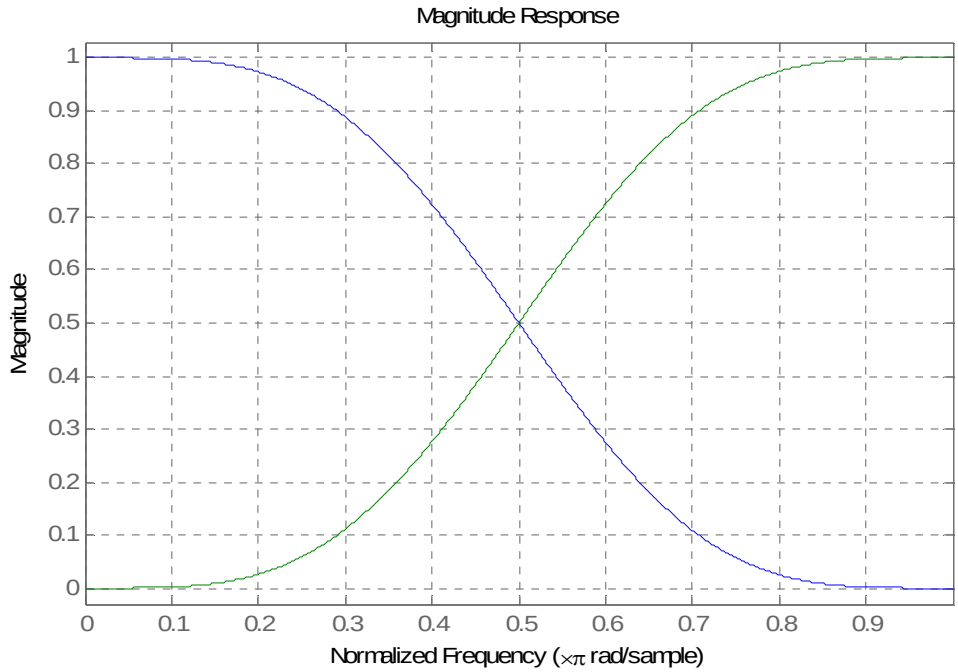
$$S_{22}^{'}(z) = \left(\frac{1}{4}[z+2+z^{-1}]\right)^2 \sum_{k=0}^{1} \binom{2+k-1}{k} [\hat{S}_{11}(z)]^k$$

$$S_{22}^{'}(z) = [0.0625z^2 \quad 0.25z \quad 0.375 \quad 0.25z^{-1} \quad 0.0625z^{-2}] \sum_{k=0}^{1} \binom{2+k-1}{k} [\hat{S}_{11}(z)]^k$$

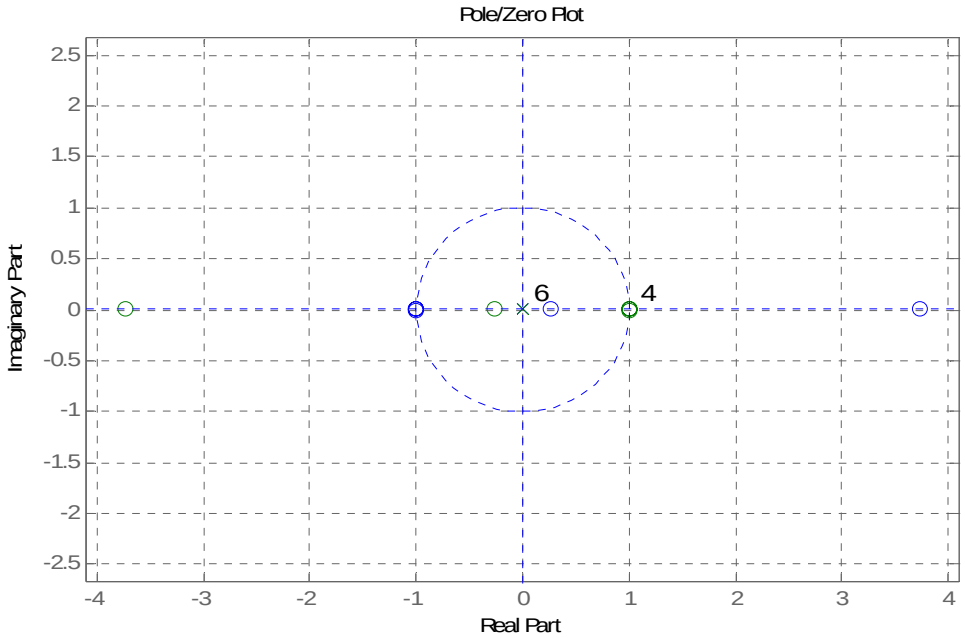
The summation is just $1+\frac{1}{4}[-z+2-z^{-1}] = [-0.5z+2-0.5z]$, since any number choose 0 is equal to 1 and any polynomial raised to the 0^{th} power is 1. The final result is then $S_{22}^{'}(z) = \left[-0.0313z^3 \quad 0z^2 \quad 0.2813z \quad 0.5 \quad 0.2813z^{-1} \quad 0z^{-2} \quad -0.0313z^{-3}\right]$. Of course, the result is symmetric and so it is linear phase. A simple shift, z^{-3}, makes it causal.



It is then a trivial exercise to construct the high pass filter as the samples of the impulse response about the midpoint are odd negative, i.e. $(-1)^k$, for $k = 1, 2, \ldots floor(length/2)$.



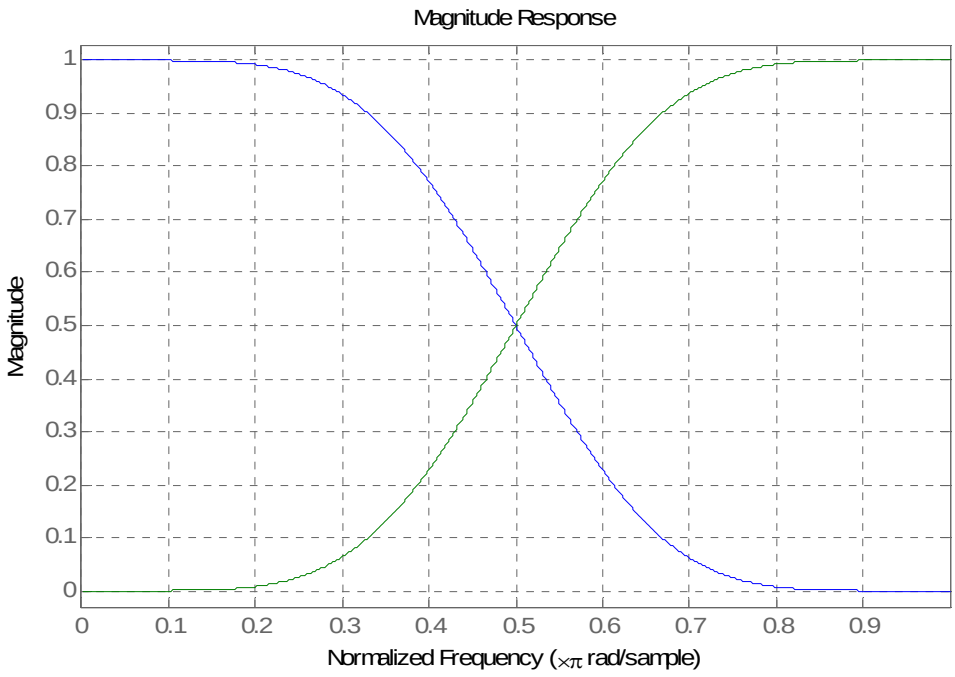
It is clear that the filters are power complimentary, as designed. The pole-zero plot for the designed quadrature mirror filter (QMF)



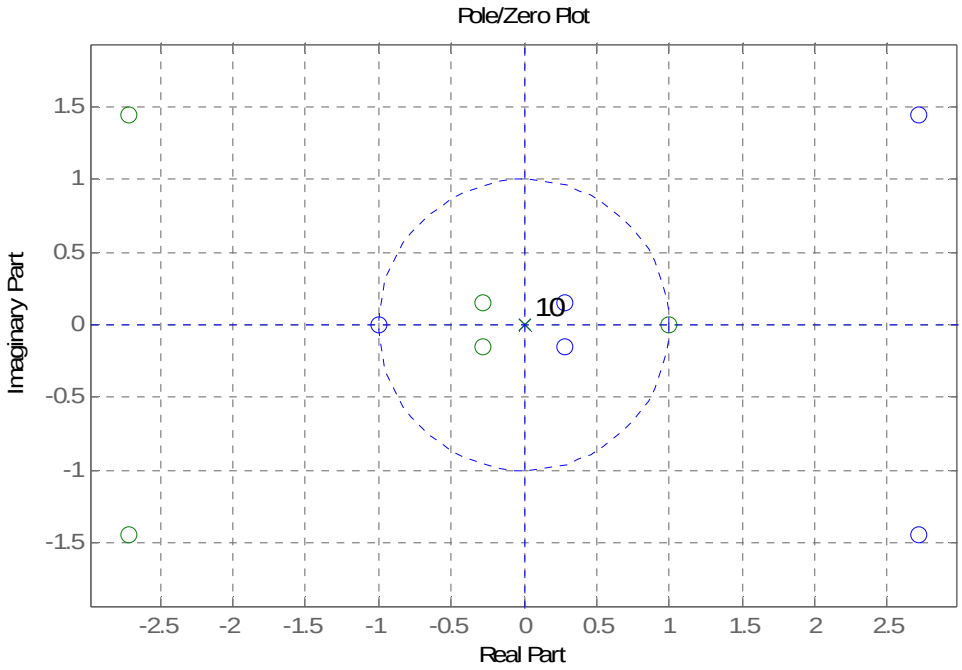
and filter information given by,

```
% Filter #1
% ---------------------------------

Discrete-Time FIR Filter (real)
---------------------------------
Filter Structure     : Direct-Form II Transposed
Numerator Length     : 7
Denominator Length   : 1
Stable               : Yes
Linear Phase         : Yes (Type 1)

% ---------------------------------
% Filter #2
% ---------------------------------

Discrete-Time FIR Filter (real)
---------------------------------
Filter Structure     : Direct-Form II Transposed
Numerator Length     : 7
Denominator Length   : 1
Stable               : Yes
Linear Phase         : Yes (Type 1)
```

For p = 3, a simple recursive program with an adapted convolution command yields



with pole-zero plot



and filter information given by

```
% Filter #1
% ------------------------------

Discrete-Time FIR Filter (real)
-------------------------------
Filter Structure     : Direct-Form II Transposed
Numerator Length     : 11
Denominator Length   : 1
Stable               : Yes
Linear Phase         : Yes (Type 1)

% ------------------------------
% Filter #2
% ------------------------------

Discrete-Time FIR Filter (real)
-------------------------------
Filter Structure     : Direct-Form II Transposed
Numerator Length     : 11
Denominator Length   : 1
Stable               : Yes
Linear Phase         : Yes (Type 1)
```

The following is a portion of the code used to generate the filter coefficients.

```
p = 3;
binom = 1; %because any N choose 0 = 1 and any polynomial^0 = 1; 1*1 =1
for k = 1:p-1
    binom = [0 binom 0]+nchoosek(p+k-1,k)*nconv(k,s11_hat_b);
end
sss = conv(nconv(p,s11_b),binom);
half = floor(length(sss)/2);
for k = 1:half
    vec(k) = (-1)^k;
end
sss_hat = sss.*[vec 1 vec];
```

Where the function nconv is

```
function polynomial = nconv(n,x);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%nconv takes in an integer n and
%convolves the vector x with itself
%n times
%
%J. DiCecco  3/22/2007
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

conv_hold = 1;
for j = 1:n
    conv_hold = conv(conv_hold,x);
end
polynomial = conv_hold;
```

Recall for the case when p = 2

$$S_{22}^{'}(z) = \left[-0.0313z^{3} \quad 0z^{2} \quad 0.2813z \quad 0.5 \quad 0.2813z^{-1} \quad 0z^{-2} \quad -0.0313z^{-3} \right]$$

To factor this polynomial, first normalize with respect to the leading coefficient.

$S_{22}^{'}(z) = \left[1z^{3} \quad 0z^{2} \quad -9z \quad -16 \quad -9z^{-1} \quad 0z^{-2} \quad 1z^{-3} \right]$. $S_{22}^{'}(z) = H_{22}(z)H_{22}(z^{-1})$, so it must factor into two 3rd order polynomials.

Several points arise from this. First and foremost, since the polynomial is in the frequency domain, we must invoke the techniques associated with spectral factorization. There are a number of techniques to do this, and for the stated question, the method of roots will suffice. Also note, that the even order polynomials, except for 0, are zero. This is due to the following derivation.

$S_{pp}^{'}(z) = [S_{11}(z)]^p \sum_{k=0}^{p-1} \binom{p+k-1}{k} [\hat{S}_{11}(z)]^k$, where $S_{11} = \frac{1}{4}\left(z + 2 + z^{-1}\right)$

and $\hat{S}_{11} = S_2 = \frac{1}{4}\left(-z + 2 - z^{-1}\right)$. This results in

$S_{pp}^{'}(z) = \left[\frac{1}{4}\left(z + 2 + z^{-1}\right)\right]^p \sum_{k=0}^{p-1} \binom{p+k-1}{k} \left[\frac{1}{4}\left(-z + 2 - z^{-1}\right)\right]^k$. Now, if we

normalize $S_{22}^{'}(z) = \left[1z^3 \quad 0z^2 \quad -9z \quad -16 \quad -9z^{-1} \quad 0z^{-2} \quad 1z^{-3}\right]$ and rewrite as

$S_0(z) = \left(\frac{1}{4\sqrt{2}}\right)^2 \left[1 \quad 0z^{-1} \quad -9z^{-2} \quad -16z^{-3} \quad -9z^{-4} \quad 0z^{-5} \quad 1z^{-6}\right]$, the product filter can

be expressed as $S_0(z) = z^{2p-1} S_{pp}^{'}(z)$. Substituting $S_{pp}^{'}(z)$ into this equation yields

$S_0(z) = (1 + z^{-1})^{2p} \frac{1}{2^{2p}} \sum_{k=0}^{p-1} \binom{p+k-1}{k} (-1)^k z^{-(p+k-1)} \left(\frac{1 - z^{-1}}{2}\right)^{2k}$ and it is now clear that the

term in summation cancels the odd orders.

As mentioned, there are many methods we could invoke to solve the spectral factorization. The most intuitive is the root method. Once we generate the sequence $S_{pp}^{'}(z)$, we root the polynomial, and determine which roots are inside the unit circle and on the unit circle (within some error neighborhood). These are the roots responsible for minimum phase. (FIR filters based on Daubechies wavelets are of course orthogonal, but they are not linear phase.) However, since we are looking for a factorization, we need only half the zeros, the other half will be generated in the square. So for the case of

$S_0(z) = \left(\frac{1}{4\sqrt{2}}\right)^2 \left[1 \quad 0z^{-1} \quad -9z^{-2} \quad -16z^{-3} \quad -9z^{-4} \quad 0z^{-5} \quad 1z^{-6}\right]$, the roots are

[3.7321 -1.0001 -1.0000 + 0.0001i -1.0000 - 0.0001i -0.9999 0.2679] according to MATLAB. We know of course that the error in placement of the zeros at -1 (the

+0.0001i) can be ignored and as expected there are 4 roots at -1, one root inside the unit circle and one root outside the unit circle (reflected). The roots used to generate the two polynomials $H(z)$ and $H(z^{-1})$ are then [0.2679 -1 -1]. We use MATLAB's poly command to expand the roots to form [1.0000 1.7321 0.4642 -0.2679]. Finally, the polynomial is normalized by multiplying by the square root of the maximum value of S_0, 0.5 in this case, divided by the square of the sum of the absolute value of the expanded roots. That is

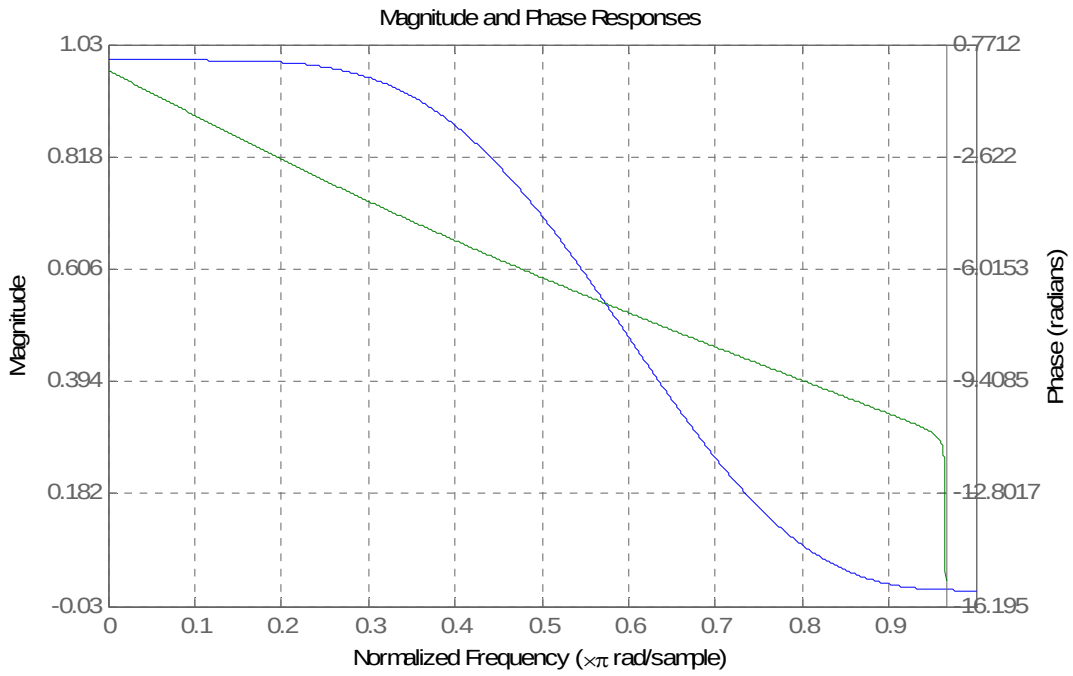
$$[1.0000\ \ 1.7321\ 0.4642\ \ -0.2679]\left(\sqrt{0.5\Big/\sum|[1.0000\ \ 1.7321\ 0.4642\ \ -0.2679]|}\right),$$ which gives [0.3415 0.5915 0.1585 -0.0915]. This is $H(z)$. $H(z^{-1})$ is obtained simply by flipping the polynomial from right to left, as this is equivalent to time reversal, making the causal sequence the required anticausal counterpart.

Using this technique, the spectral factorization for p = 3 is

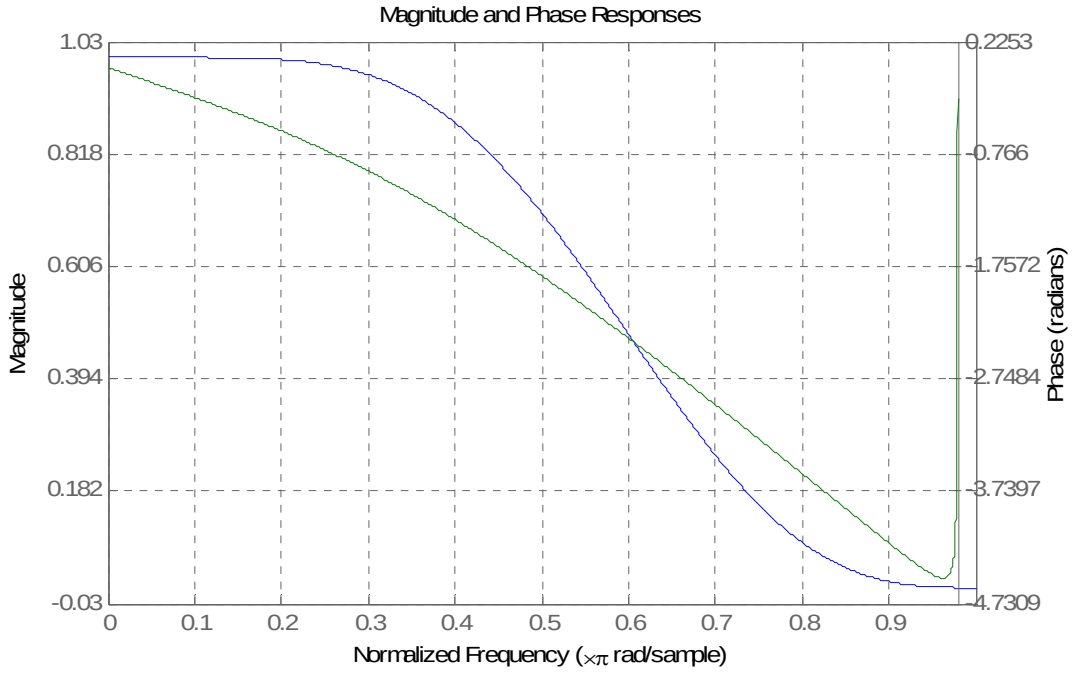
$H_{33}(z) = [0.2359 \quad 0.5708 \quad 0.3242 \quad -0.0957 \quad -0.0601 \quad 0.0248]$ and

$H_{33}(z^{-1}) = [0.0248 \quad -0.0601 \quad -0.0957 \quad 0.3242 \quad 0.5708 \quad 0.2359]$. Clearly, neither $H_{33}(z^{-1})$ nor $H_{33}(z)$ are linear phase; there is no symmetry. They are a combination of a minimum phase (all zeros on or in the unit circle) and a mixed phase (some zeros are on, some are out), however, and the phase plot does a nice job of quantifying this.

Mixed phase…

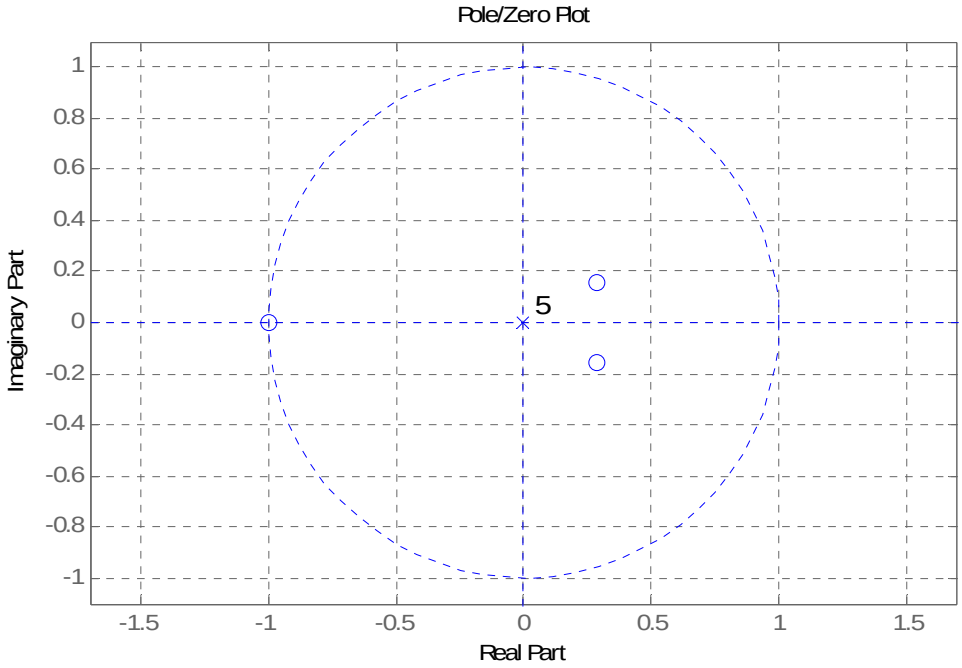


Minimum phase …

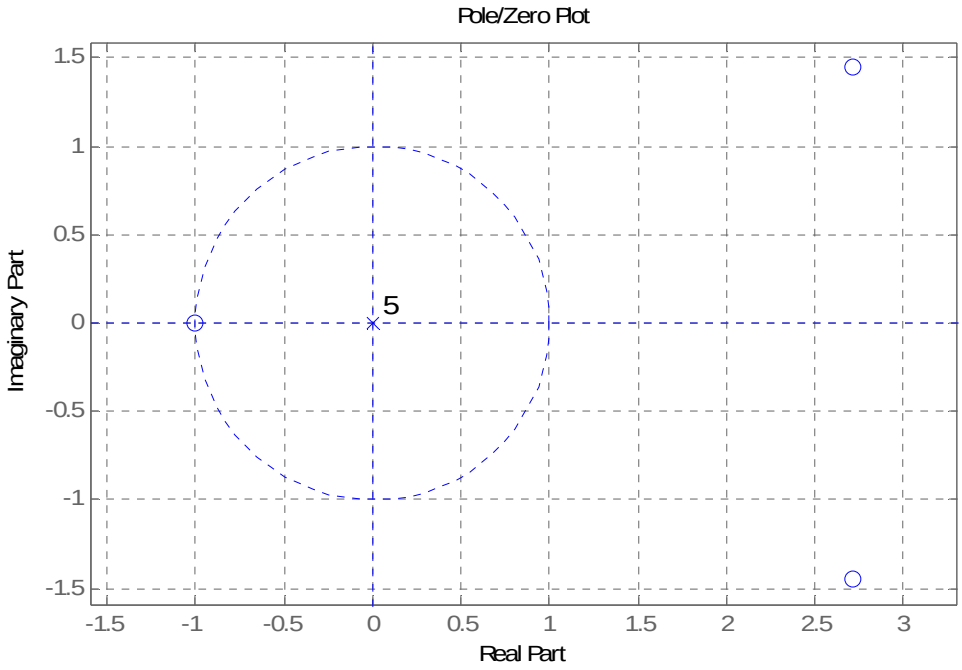


The pole-zero plots for both $H_{33}(z^{-1})$ and $H_{33}(z)$ show the reciprocal nature of the factorization.

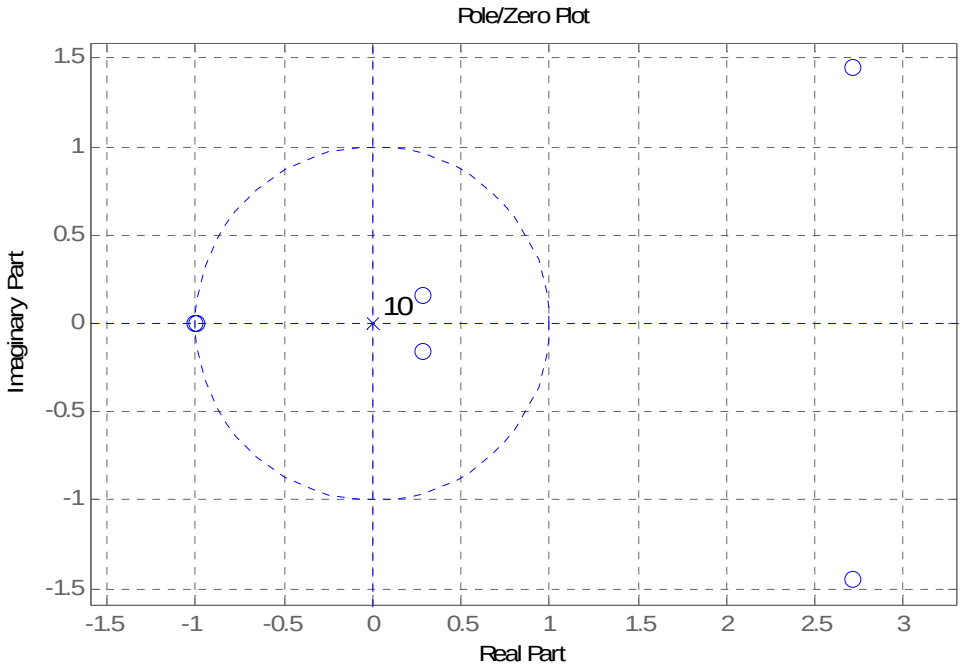
$H_{33}(z)\ldots$



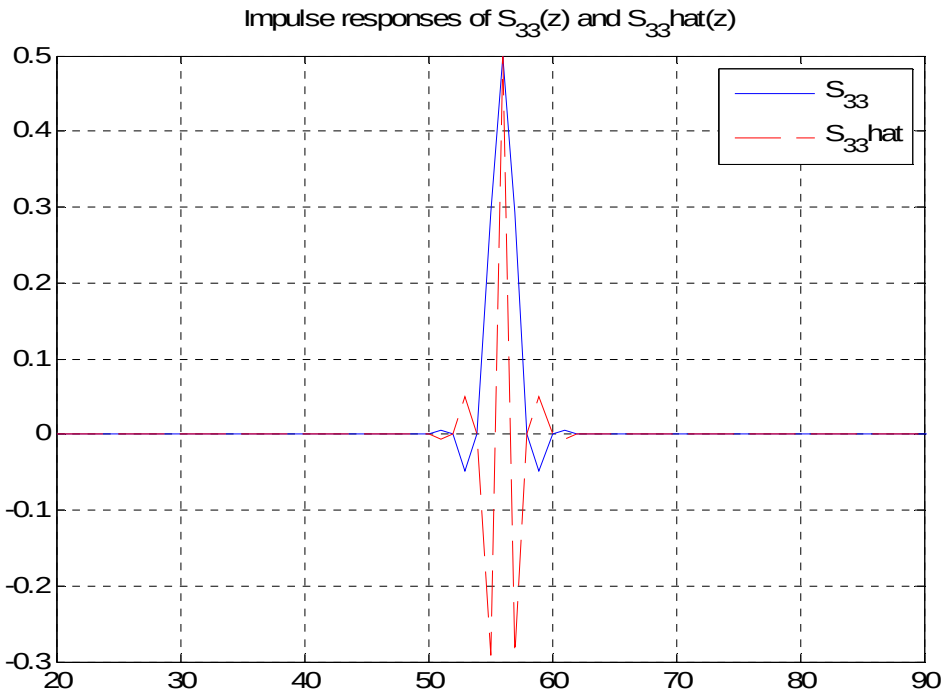
and $H_{33}(z^{-1})\ldots$



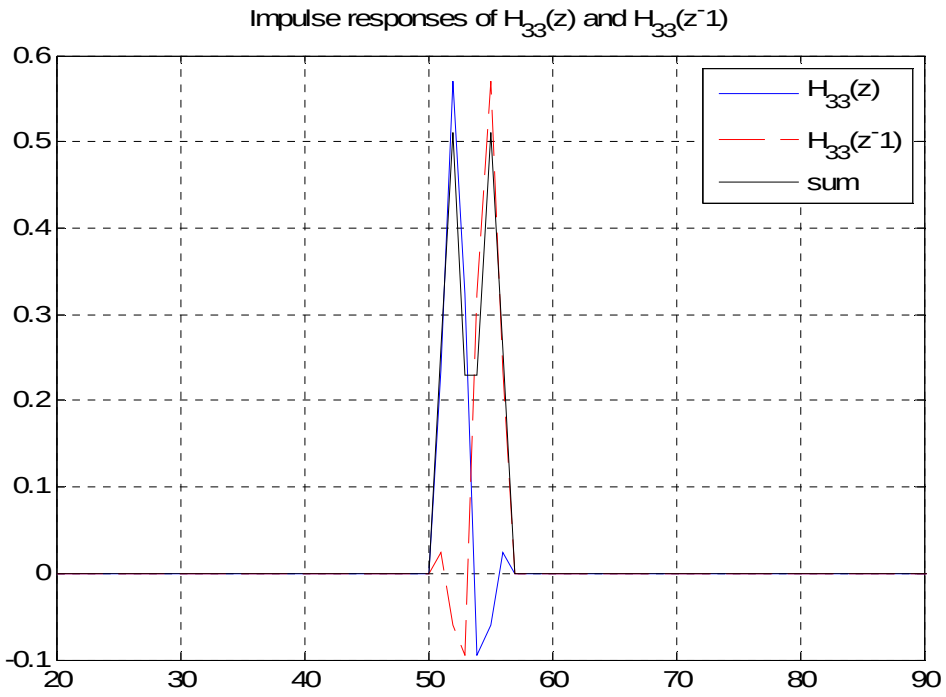
Compare these to the pole-zero plot of S_{33}, and it is clear that the factorization is correct.



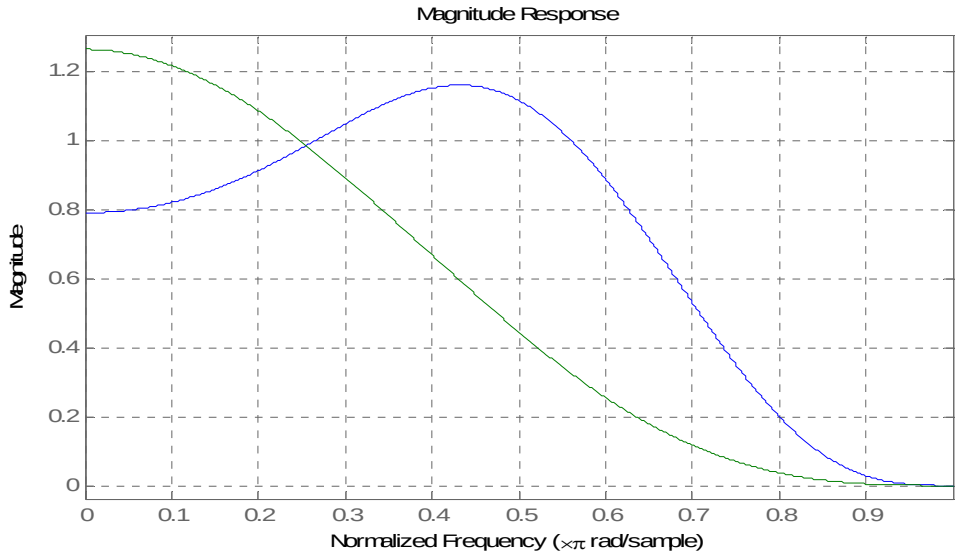
Certainly, there are many ways that $S_{pp}^{'}(z)$ can be factored to give $H_{pp}(z^{-1})$ and $H_{pp}(z)$. Some of those factorizations will give mixed order lengths and linear phase solutions. If the filters are to remain orthogonal, as is in the Daubechies case, one has to be more selective about the zero placements in order to approximate linear phase. For visualization, consider the impulse responses of $S_{33}(z)$ and $\hat{S}_{33}(z)$.



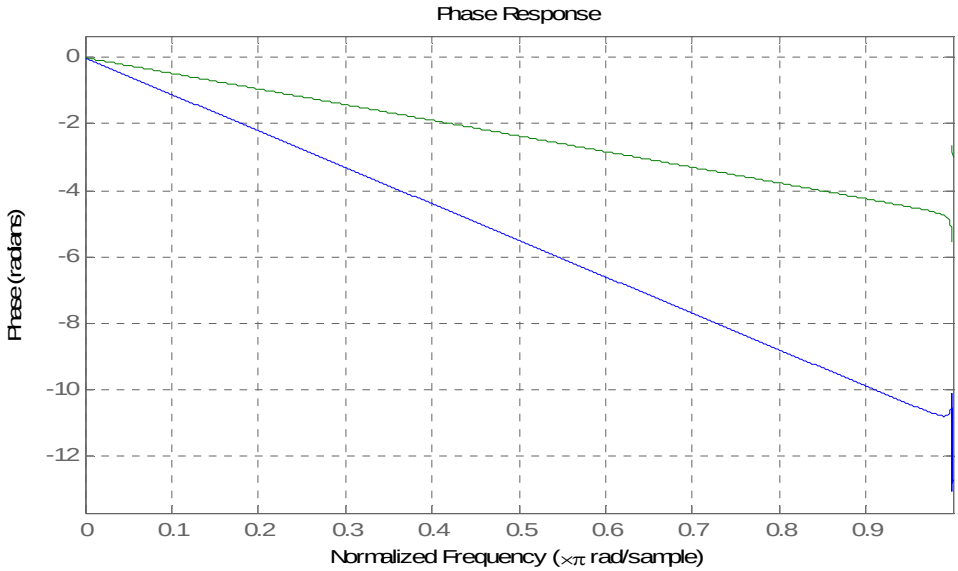
By inspection, they are symmetric and so have linear phase. Now consider the impulse responses of the factorization of $H_{33}(z^{-1})$ and $H_{33}(z)$.



Combined, of course, it is easy to see the symmetry. But separately it is clear that they do not have symmetry independently and will influence the delay (phase) of the energy in the signal. Now, if we wish to force symmetry by isolating the zeros on the circle as *H(z)* and the reflective zeros as $H(z^{-1})$, the magnitudes are different

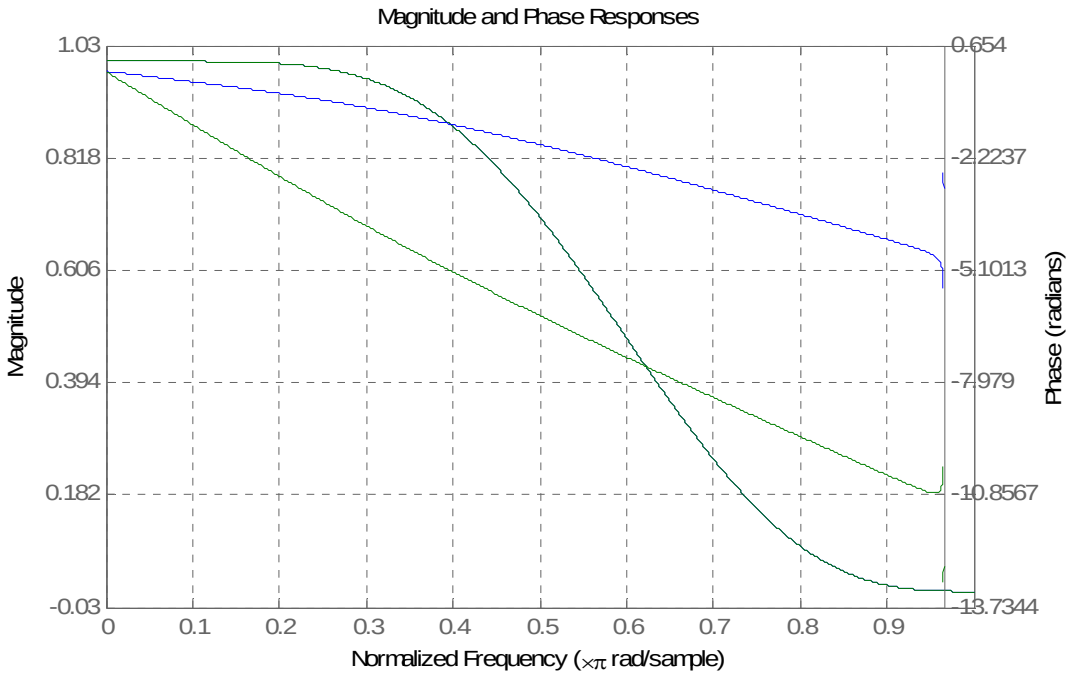


but the resulting phase is quasilinear. (It may look linear, but it's not!)



We have done so, however, at the expense of destroying the causal, anticausal relationship as required by the formula $S_1 = H_1(z)H_1(z^{-1})$. An important consequence of

the symmetry issue is that it destroys orthogonality. But by preserving the orthogonality, we lose the symmetry, and we are back to nonlinear phase.



Consider now the case for p = 4.

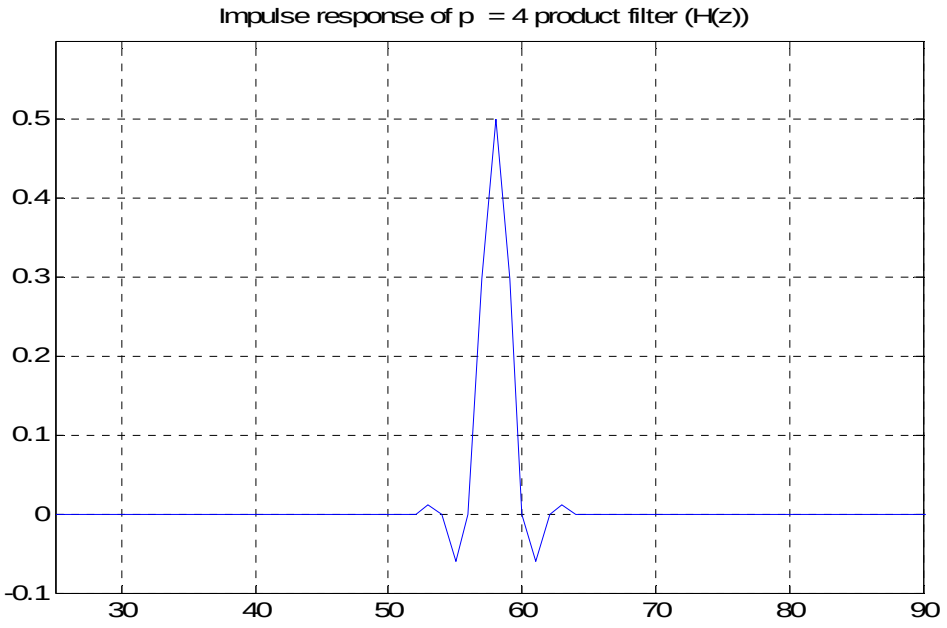
$S_{44} = [-0.0012 \;\; 0 \;\; 0.0120 \;\; 0 \;\; -0.0598 \;\; 0 \;\; 0.2991 \;\; 0.5000 \;\; 0.2991 \;\; 0 \;\; -0.0598 \;\; 0 \;\; 0.0120 \;\; 0 \;\; -0.0012]$

As expected, it has symmetry and the odd order z terms are zero. The orthogonal decomposition (Daubechies) factors S_{44} into

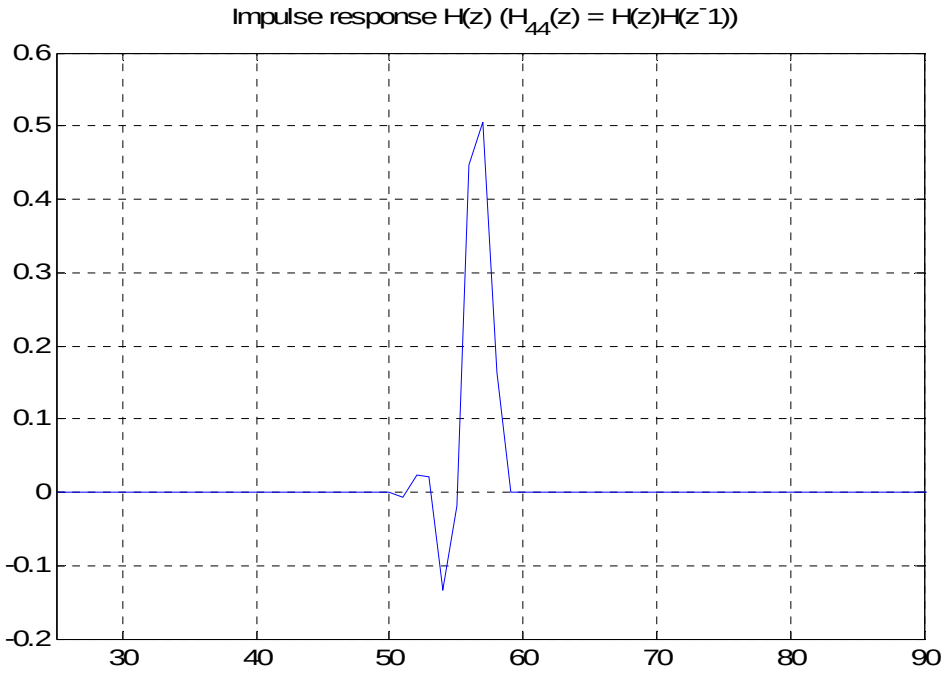
$H_{44}(z) = [0.1629 \;\; 0.5055 \;\; 0.4461 \;\; -0.0198 \;\; -0.1323 \;\; 0.0218 \;\; 0.0233 \;\; -0.0075]$ and

$H_{44}(z^{-1}) = [\; -0.0075 \;\; 0.0233 \;\; 0.0218 \;\; -0.1323 \;\; -0.0198 \;\; 0.4461 \;\; 0.5055 \;\; 0.1629]$, which is the flipped (time reversed) version of $H_{44}(z)$. (As a side note, it can easily be demonstrated using MATLAB's deconv command that the deconvolution of either $H_{44}(z)$ or $H_{44}(z^{-1})$ from $S_{44}(z)$ will yield the flipped version of either $H_{44}(z)$ or

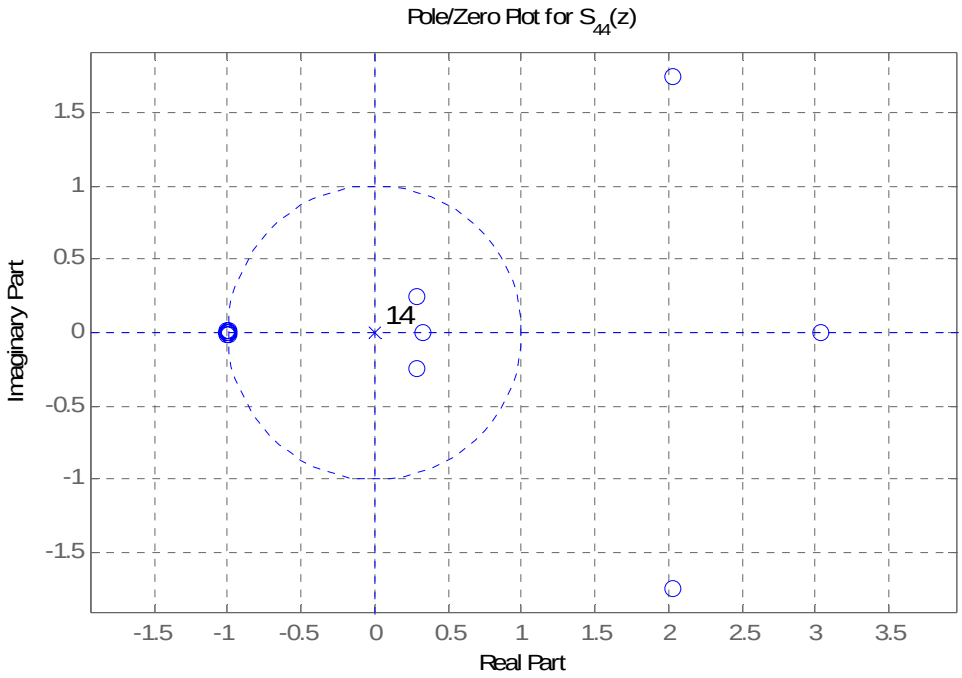
$H_{44}(z^{-1})$, respectively.) Again, we see a symmetric impulse response, indicating linear phase.



When we perform the orthogonal spectral factorization, again we see the impulse response become left or right skewed, corresponding to advance or delay, respectively.



In order to approximate linear phase, the impulse response from $H_{44}(z)$ must be made to be as close to symmetric as possible. To accomplish this, the zeros from $S_{44}(z) = H_{44}(z)H_{44}(z^{-1})$ must be manipulated. Consider the pole-zero plot for $S_{44}(z)$



where the orthogonal spectral factorization gives the roots for $H_{44}(z)$ as $[-1 \quad -1 \quad -1 \quad -1 \quad 0.2841+0.2432i \quad 0.2841-0.2432i \quad 0.3289]$ and for $H_{44}(z^{-1})$ as $[3.0407 \quad 2.0311+1.7390i \quad 2.0311-1.7390i \quad -1 \quad -1 \quad -1 \quad -1]$. We can redistribute the zeros from $H_{44}(z)$ to form $h_{44\_1} = [-1 \quad -1 \quad 0.3289]$ and $h_{44\_2} = [-1 \quad -1 \quad 0.2841+0.2432i \quad 0.2841-0.2432i]$. Similarly, we redistribute the zeros from $H_{44}(z^{-1})$ to form $h_{44\_3} = [-1 \quad -1 \quad 3.0407]$ and $h_{44\_4} = [-1 \quad -1 \quad 2.0311+1.7390i \quad 2.0311-1.7390i]$. Clearly, h_{44_1} and h_{44_3} are inverses of one another, as are h_{44_2} and h_{44_4}. By convolving the polynomials associated

with h_{44_1} and h_{44_4}, the new filter is

$$H_{44\_1}(z)H_{44\_4}(z) = [1.0000 \quad -0.3904 \quad -3.0789 \quad 9.2419 \quad 24.9488 \quad 15.4616 \quad -0.9031 \quad -2.3465]$$

After normalizing for unity gain

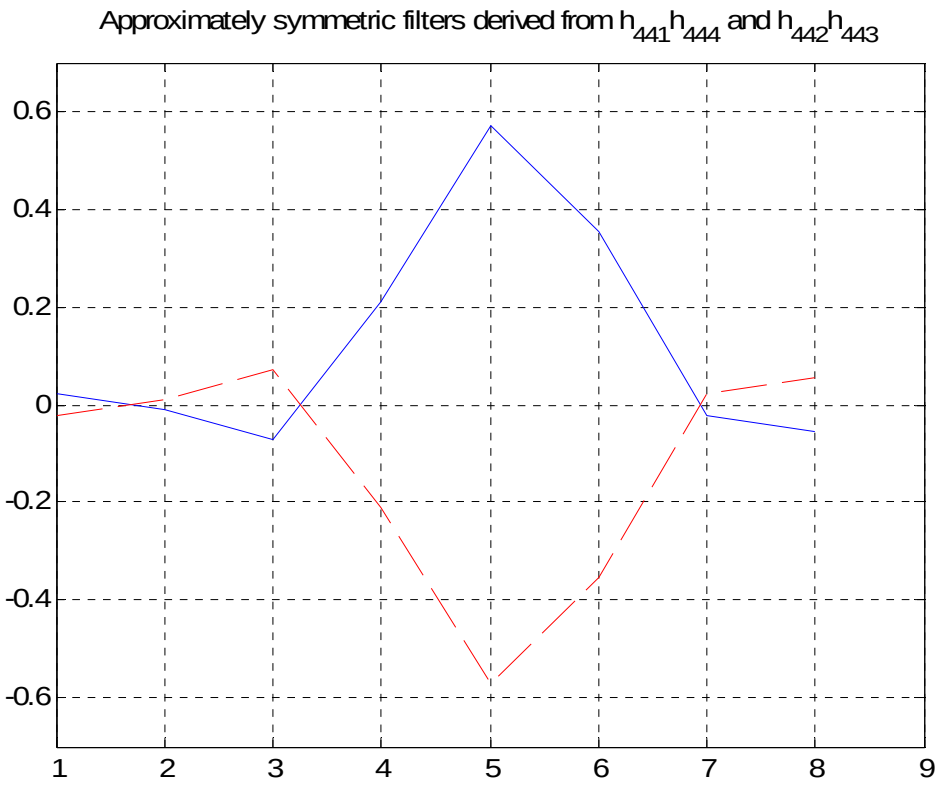
$$H_{44\_1}(z)H_{44\_4}(z) = [0.0229 \quad -0.0089 \quad -0.0705 \quad 0.2116 \quad 0.5713 \quad 0.3541 \quad -0.0207 \quad -0.0537]$$

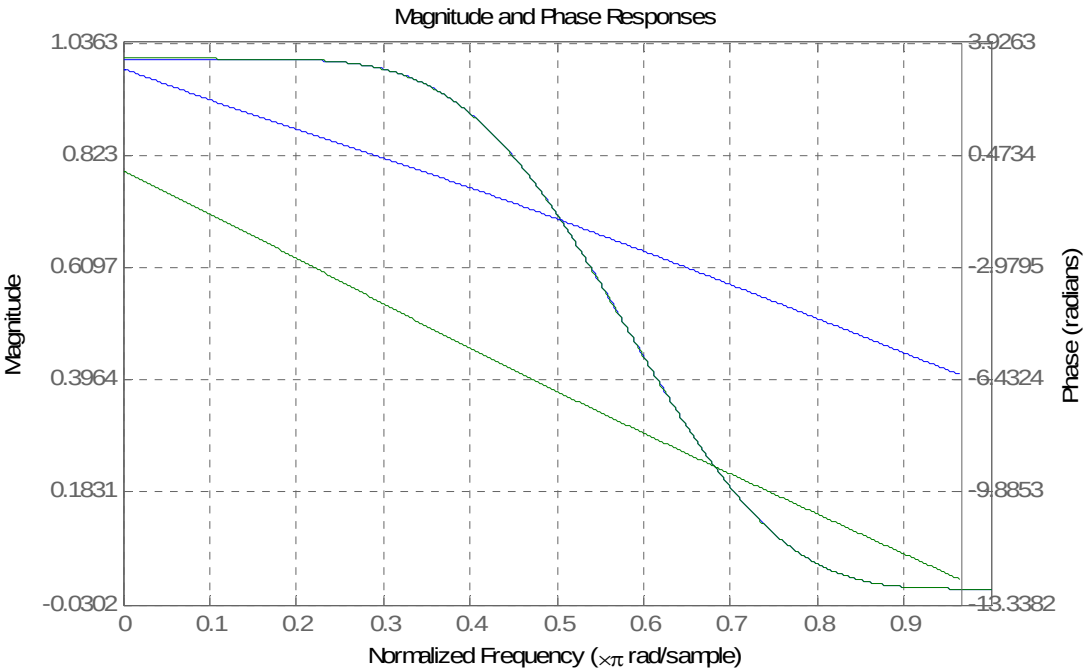
We do the same for h_{44_2} and h_{44_3} which yields

$$H_{44\_2}(z)H_{44\_3}(z) = [0.0539 \quad 0.0211 \quad -0.3538 \quad -0.5714 \quad -0.2118 \quad 0.0706 \quad 0.0090 \quad -0.0229]$$

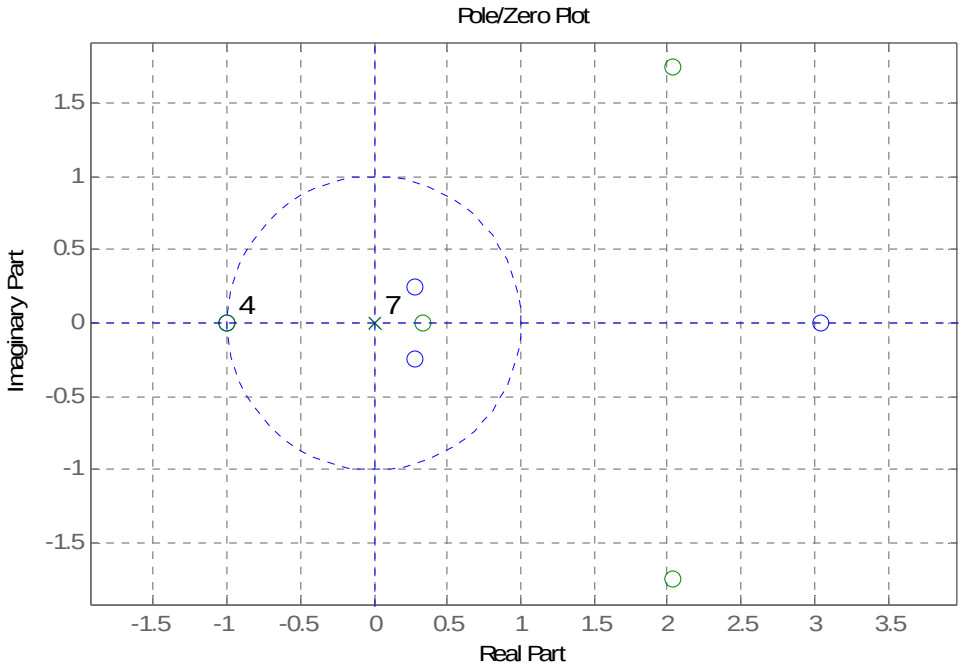
Clearly, these two filters are time reversed and negatives of one another.

The resulting polynomials are approximately symmetric and so give approximately linear phase.





All the original zeros are represented and accounted for, and in their original positions



Note the distribution of the pairings (green and blue).

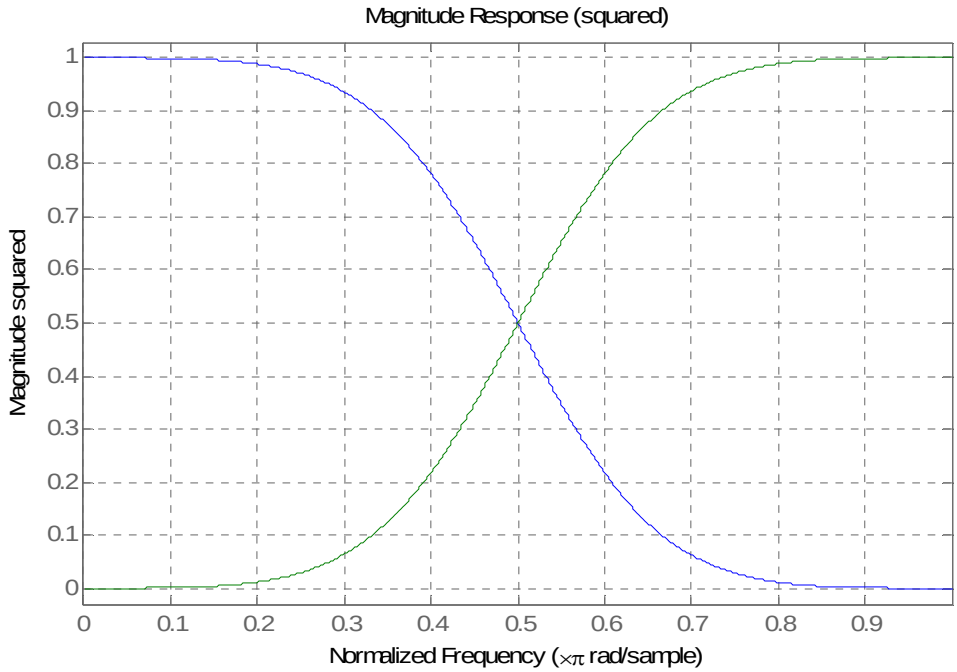
To illustrate the difference in the number of filter coefficients needed for an FIR versus an IIR filter, consider the case of the 2nd order halfband Butterworth filter. Starting with the lowpass filter we have

[b,a] = butter(2,0.5)

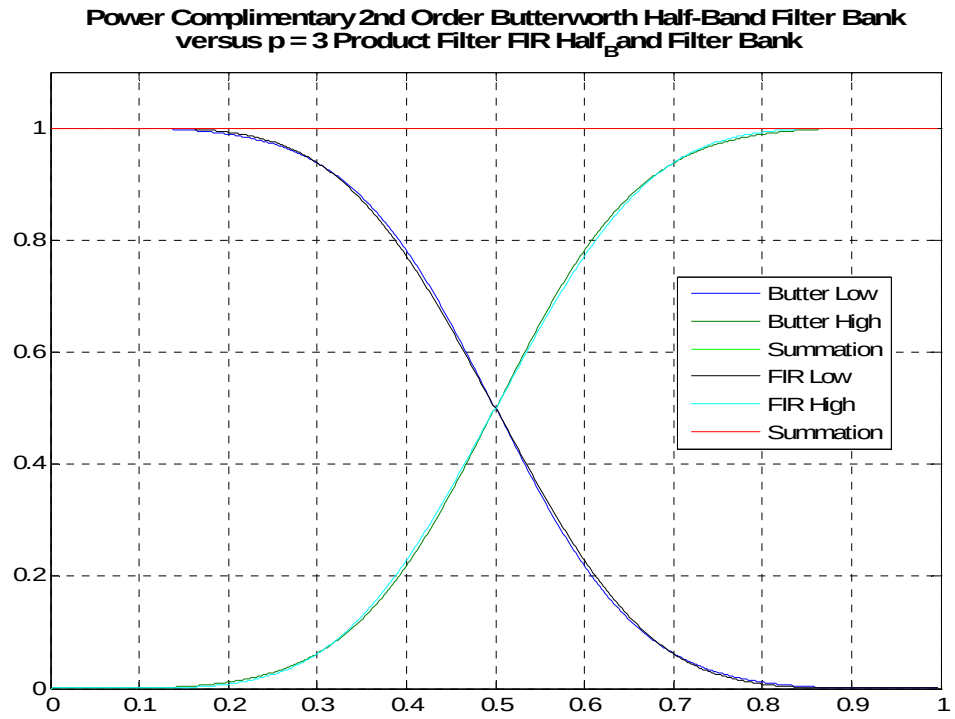
b = 0.2929 0.5858 0.2929

a = 1.0000 -0.0000 0.171

where the coefficients have been normalized to give unity gain. The highpass filter is trivially obtained by negating the middle term in the b coefficients while the a coefficients are left unchanged.



The plot is generated by squaring the magnitude response since the relationship required for reconstruction is $H_1(z)H_1(z^{-1}) + H_1(-z)H_1(-z^{-1}) = 1$. For $|H_1(z)| = |H_1(z^{-1})|$, this is equivalent to $|H_1(z)|^2$. If we superimpose the result from the p = 3 product filter, we have



These correspond to roughly the same filter. The difference is the number of coefficients necessary to achieve the same result. Recall for the p = 3 filter,

$S_{33}(z) = [0.0059 \quad 0 \quad -0.0488 \quad 0 \quad 0.2930 \quad 0.5000 \quad 0.2930 \quad 0 \quad -0.0488 \quad 0 \quad 0.0059]$,

while for

$\hat{S}_{33}(z) = [-0.0059 \quad 0 \quad 0.0488 \quad 0 \quad -0.2930 \quad 0.5000 \quad -0.2930 \quad 0 \quad 0.0488 \quad 0 \quad -0.0059]$

.

We can, of course squeeze out the zeros by adding a delay, but we still need to alternate the signs with the exception of z^0, which for the causal system, is now z^{-3}. Importantly, there is no feedback needed since it is FIR. We look to the Butterworth IIR case and we have the coefficients previously listed b = 0.2929 0.5858 0.2929

a = 1.0000 -0.0000 0.171 for the lowpass. For the highpass, the only change is the negated z^{-1} term for the b coefficients. Of course now we need a feedback loop for the poles. For the FIR, there are 7 multiplications with 4 extra delays (no feedback),

resulting in a tenth order system. For the IIR configuration, we have a 2nd system order with feedback.

REFERENCES

Jackson, L. B. Digital Filters and Signal Processing, 3rd edition. Kluwer Academic Publishers, Norwell, MA, 1996.

Herley, C., and Vetterli, M. Wavelets and Recursive Filter Banks. *IEEE Trans. Sig. Proc.* (1993) **41,** 2536-2556

APPENDIX

Matlab code

```
clear all
close all

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%s11(z) = 1/4[z + 2 + z^-1] and s11_hat(z) = 1/4[-z + 2 + -z^-1]
%s11_prime(w) = e^jw/4 + e^-jw/4 + 1/2 = 1/2(1 + cos(w)) = cos^2(w/2)
%similar for s11 hat
%
%spp(z) = [s11(z)]^p(sum((p+k-1 choose k)*[s11_hat(z)]^p)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

s11_b = [1 2 1]/4;
s11_hat_b = [-1 2 -1]/4;

s22 = conv(s11_b,s11_b);
binom = [0 1 0] + nchoosek(2,1)*s11_hat_b

s22 = conv(s22,binom);
s22_hat = s22.*[-1 0 -1 1 -1 0 -1];

[h22,w] = freqz(s22);
[h22_hat,w] = freqz(s22_hat);

plot(w/pi,abs(h22),'k'),hold on,plot(w/pi,abs(h22_hat))
plot(w/pi,abs(h22)+abs(h22_hat),'r')
grid

p = 3;
binom = 1; %because any N choose 0 = 1 and any polynomial^0 = 1; 1*1 =1
for k = 1:p-1
    binom = [0 binom 0]+nchoosek(p+k-1,k)*nconv(k,s11_hat_b)
end
sss = conv(nconv(p,s11_b),binom);
half = floor(length(sss)/2);
for k = 1:half
    vec(k) = (-1)^k;
end
sss_hat = sss.*[vec 1 vec];


[hpp,w] = freqz(sss);
[hpp_hat,w] = freqz(sss_hat);
```

```
function polynomial = nconv(n,x);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%nconv takes in an integer n and
%convolves the vector x with itself
%n times
%
%J. DiCecco  3/22/2007
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

conv_hold = 1;
for j = 1:n
    conv_hold = conv(conv_hold,x);
end
polynomial = conv_hold;



function b = daubech(s)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%daubech takes in a PSD Spp and
%performs spectral factorization based
%on the root locations
%
%J. DiCecco  3/22/2007
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

r = roots(s)
eps_d = 0.01; %Matlab always has that little circle of zeros not on -1
in = r(abs(r)<(1-eps_d));
on = r((abs(r)>=(1-eps_d)) & (abs(r)<=(1+eps_d)));

[a,k] = sort(angle(on))
on = on(k(1:2:end));%just take half


r = [in; on];
%r = [on];
b = poly(r);

if isreal(s)  %if the signal is real -> complex conjugate roots
    b = real(b); %just need the real parts since this is the poly
end

b = b*sqrt(max(s)/sum(abs(b).^2));
```

Biomedical Engineering and Nonlinear Dynamics
Prof. Ying Sun

1. Define and compare (strict) stationarity and wide-sense stationarity.

If for the time series X(t), the joint distribution of

$[X(t_1), X(t_2), ...X(t_n)]$ and $[X(t_1+\tau), X(t_2+\tau), ...X(t_n+\tau)]$ is the same for all t_n and τ, the process is said to be stationary in the strict sense [1,2]. In words, given a time series, the joint distribution depends only on the difference in time (or sample) τ and not the sample time (or sample number) t_n. That is to say, that the probabilistic structure is completely invariant under a shift of the time (or sample) origin [1]. This is a remarkably difficult condition to prove in working with real, finite time series recordings since we cannot say for certain that based on one realization (the recording itself), we have captured the statistics of the time series. Stationarity is a function of the process, not the data [3]. However, we can sometimes accurately determine a certain number of moments, mean and variance, for example, of a time series. Under such circumstances we can define a weaker, less strict, definition of stationarity. Wide-sense stationarity, or weak-sense stationarity, characterizes a time series such that the joint distribution of $[X(t_1), X(t_2), ...X(t_n)]$ and $[X(t_1+\tau), X(t_2+\tau), ...X(t_n+\tau)]$ is the same up to order m, where m is the order of moments [1]. For instance, a time series could be stationary up to order 1, if its mean μ were some constant less than infinity and did not change throughout the time series. Further we could define stationarity up to order 2, if both the mean μ and variance σ^2 were a constant

less than infinity, and the covariance were strictly a function of the time difference between samples. Mathematically, we can say for a time series {X(t)}, if the following conditions are met, it is stationary in the wide-sense.

$$\mu_X = E[X] = \sum_k x_k p_X(x_k) = \overline{X} = \text{constant}$$
$$\sigma_X^2 = E\{[X - \overline{X}]^2\} < \infty = \text{constant}$$
$$\text{Cov}[X(t_1), X(t_2)] = E[X(t_1)X(t_2)] - \mu^2$$
$$= \text{function of } (t_1 - t_2) \text{ only}$$

For the purposes of nonlinear time series analysis, stationarity is a necessary, though not sufficient, condition for establishing whether a dynamical system is chaotic [3]. Further, one needs to determine the nature of the time series, since nonlinear data can often mimic nonstationary data [4].

[1] Priestley, M. B., Spectral Analysis and Time Series. Academic Press Limited, London, UK. 1989.

[2] Kantz, H. and Schreiber, T. Nonlinear Time Series Analysis, 2nd edition. Cambridge University Press, Cambridge, UK. 2005.

[3] Sprott, J. Chaos and Time-Series Analysis. Oxford University Press, New York. 2003.

[4] Reike, et al. Discerning Nonstationarity From Nonlinearity in Seizure-Free and Preseizure EEG Recordings From Epilepsy Patients. *IEEE Trans. Biomedical Engineering*. (2003) 50, 634-639.

2. Identity at least three different methods reported in the literature for testing the stationarity (either strict or wide-sense) of data.

There are a number of methods for testing for stationarity in data which, although trivial, should be mentioned before listing more the more rigorous tests found in literature. The reason for this is simple. Should you have a time series that fails a simple test, there is no need to spend precious time trying to prove otherwise. For instance, a time series can easily be divided in half to yield two new time series. Given a time series x(n), for n = 1,2…N, the two new time series are $x_1(n)$, for n = 1, 2, …(N/2)-1 and $x_2(n)$, for n = N/2, N/2 +1, N/2 +2, …N. If the means and variances of $x_1(n)$ and $x_2(n)$ are different, within statistical significance, the original time series is nonstationary and it is not necessary to invoke more elaborate tests. Other simple tests, such as the autocorrelation function, can also yield quick results, but even if these tests fail, one cannot conclude that the time series is stationary, and must move on to more rigorous tests.

KPSS (1992)

The Kwiatkowski, Phillips, Schmidt, Shin (KPSS, 1992) test is one of several stationary tests that employ the method of testing an autoregressive (AR) model for unit root. Given a time series x(n), assume it can be decomposed as $x[n] = r[n] + \beta n + \varepsilon[n]$, where r[n] is a random walk (r[n]= r[n-1] + u[n], for u[n]~N(0, σ_u^2); βn is a deterministic trend; ε[n] is a stationary error.

The KPSS test can be thought of as other statistical tests in that it yields a number which will determine a particular confidence level for the null hypothesis. For KPSS the statistic is $KPSS = N^{-2}\sum_{n=1}^{N} S^2[n]/\hat{\sigma}^2(p)$, where, N is the length of the time series, S is the sum process of the error sequence ($error = x_t - \overline{x}$), and $\hat{\sigma}^2(p)$, is the consistent estimator of σ^2. For large N, $KPSS = N^{-2}\sum_{n=1}^{N} S^2[n]/\hat{\sigma}^2(p) \rightarrow \int_0^1 V_1(r)^2 dr$ under the null hypothesis of level stationarity (stationarity without a trend), and will approach $\int_0^1 V_2(r)^2 dr$ under the null hypothesis of stationarity with a trend. (V_1 and V_2 are standard Brownian bridge distribution). Below are the statistics established by Kwiatkowski, et al.

Distribution	Upper tail percentiles			
	0.1	0.05	0.025	0.01
$\int_0^1 V_1(r)^2 dr$	0.347	0.463	0.574	0.739
$\int_0^1 V_2(r)^2 dr$	0.119	0.146	0.176	0.216

The difficulties in applying this test are the same as those in all unit root test statistics. Namely, the test is often dealing with non-uniform and non-standard distributions, and the distributions are severely affected by deterministic additions (constants, trend lines, magic numbers, etc.).

Witt, et al, "Testing stationarity in time series" (1998)

This technique is probably the most intuitive (statistically) test for stationarity checking. Again, the time series is divided into smaller parts, and new sets of data are produced. This is done using a windowing technique, though any segmentation scheme will work. Essentially, two tests are preformed. The hypothesis is that A) the one dimensional (1D) probability density is independent of time, and B) the power spectral density is independent of time [].

The first test involves a simple calculation of means and variances that will be used to test the *time independence of probability distributions* of the windowed segments. For weak sense stationary, distributions can be fully categorized by the first two moments, μ and σ^2 (in fact under the normally distributed Gaussian assumption, this is always the case). The distributions of each window are binned and compared using a modified χ^2 test statistic. One can compare one window against another sequentially or randomly, or one can compare multiple windows. In the case of comparing two windows, the chi-square statistic is

$t_{A,2} = \sum_{k=1}^{r} \frac{(R_k^i - R_k^j)^2}{\sigma^2(R_k^i) + \sigma^2(R_k^j)}$, where R^i is the number of elements in the i^{th} and R^j is the number of elements in the j^{th} bin. But the bins are separated by some distance D. To account for this, Witt, et al, have modified the statistic such that

$t_{A,2} = n_w^2 \sum_{k=1}^{r} \frac{(R_k^i - R_k^j)^2}{(R_k^i)^3 \sigma^2[D_k^i(m)_{m=1}^{R_k^i-1}] + (R_k^j)^3 \sigma^2[D_k^j(m)_{m=1}^{R_k^j-1}]}$, where σ^2 [D] represents the variance in the index distance []. Once this test shows time independence for the first central moments, the test continues to check the *time independence* of the second order noncentral moments. The autocorrelation function of each

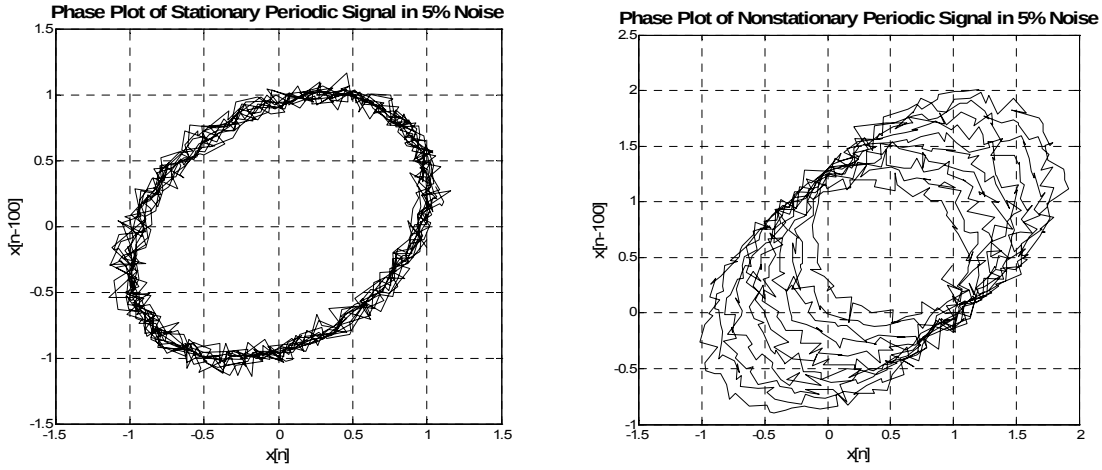
windowed segment is approximated by $\rho^j(k) = \frac{\sum_{\tau=1}^{n_w - r} (\mathrm{x}^{\mathrm{j}}_{\tau} - \mu^{\mathrm{j}})(\mathrm{x}^{\mathrm{j}}_{\tau+k} - \mu^{\mathrm{j}})}{(n_w - k)\sigma^j}, k = 0,1,...n_\rho$,

where n_w is the size of the window []. Taking the Fourier transform gives the power spectral density, which is then binned according to distribution. The chi-square test is performed as in the first test, and statistical significance is attached. If the distributions are the same, the signal is stationary (in the weak sense), otherwise, it is nonstationary.

Rieke, et al, "Measuring Nonstationarity by Analyzing the Loss of Recurrence in Dynamical Systems" (2002)

The most important difference in this method is that there is no need to partition the data. Instead, the technique utilizes the concept of "loss of recurrence." Consider a time series x[n], for n = 1, 2 …N. If we create a phase plot by plotting x[n] versus a delayed version of itself, i.e. x[n - 10], the resulting structure is termed an attractor. A neighborhood on the attractor is established by $U_\varepsilon(\vec{x}_r) = \{\vec{x}_n : \left\|\vec{x}_n - \vec{x}_r\right\| \le \varepsilon\}$, where $\vec{x}_r$ is a reference vector (point in state space), $\vec{x}_n$ is the vector (point in state space) we are judging the distance from the reference vector, and ε is the smallest area (neighborhood) we wish to define. Alternatively to the neighborhood being a size, we can refer to the neighborhood as a specific number of vectors (points), such that once the number is reached, the neighborhood is closed. In the time evolution of the attractor, the trajectory of evolution follows a path around the attractor. The time lag between successive visits to the same neighborhood on the attractor is given by

$E(l_r) = \frac{N}{2} - \frac{(r-1)(N-r)}{(N-1)}$, assuming that all vectors have the same probability of revisiting the neighborhood in some observed lag time $l_r = \frac{1}{k}\sum_{\vec{x}_n \in U} |n-r|$, where k is the number of neighbors in the neighborhood []. This is a reasonable assumption if the time series is stationary. This revisiting is referred to as recurrence, and the loss of recurrence is an indication of nonstationarity, since a nonstationary time series will yield an attractor that does not revisit the same state space as often as a stationary time series will.



Above is a graphic illustration of the difference between an attractor of a stationary time series and an attractor of a nonstationary time series. It is clear that if you were to define a neighborhood on the attractor of the stationary signal, many revisits (recurrences) would occur, whereas in the case of the nonstationary attractor, there would be a reduction in the number of revisits in a similar sized neighborhood. This is what is meant by a loss of recurrence.

The issue then is to establish a distribution based on the deviation from the expected lag E(l_r) to the observed l_r. This is possible if we treat the values l_r as independent. For the case of a stationary signal, again, this is a reasonable assumption, since the lag times will be uniformly distributed. The drawback to this assumption is that the reference vector is, of course, also a neighbor and so

the independence criteria cannot be asserted. However, large enough sample spaces (large k values) will account for this.

The strength of this test lies in the ease of application and the ability to leave the time series intact.

There are several other noteworthy tests, namely Kennel's test for stationarity, Huang's empirical mode decomposition (EMD), and Bispectrum analysis. Kennel's test is similar to that proposed by Rieke in that it relies on statistical assumptions about nearest neighbors in state space [4]. It is not as intuitive as Rieke's method and still requires the data to be segmented. Empirical mode decomposition, while not actually testing for nonstationarity, isolates the trend line in the final intrinsic mode function (IMF) in the decomposition. Clearly, if one wanted to determine whether a trend existed, the EMD will reveal it in the IMF decomposition [5]. Bispectrum analysis will be discussed under question 4.

[1] Kwiatkowski, D. Testing the null hypothesis of stationarity against the alternative of a unit root. *J. Econometrics*. (1992) **54** 159-178

[2] Witt, et al. Testing stationarity in time series. *Phys. Rev. E.* (1998) **58** 1800-1810

[3] Rieke, et al. Measuring Nonstationarity by Analyzing the Loss of Recurrence in Dynamical Systems. *Phys. Rev. Lett.* (2002) **88** 244102-1 - 244102-4

[4] Kennel, M. B. Statistical test for dynamical nonstationarity in observed time-series data. *Phys. Rev. E* (1997) **56**, 316-321

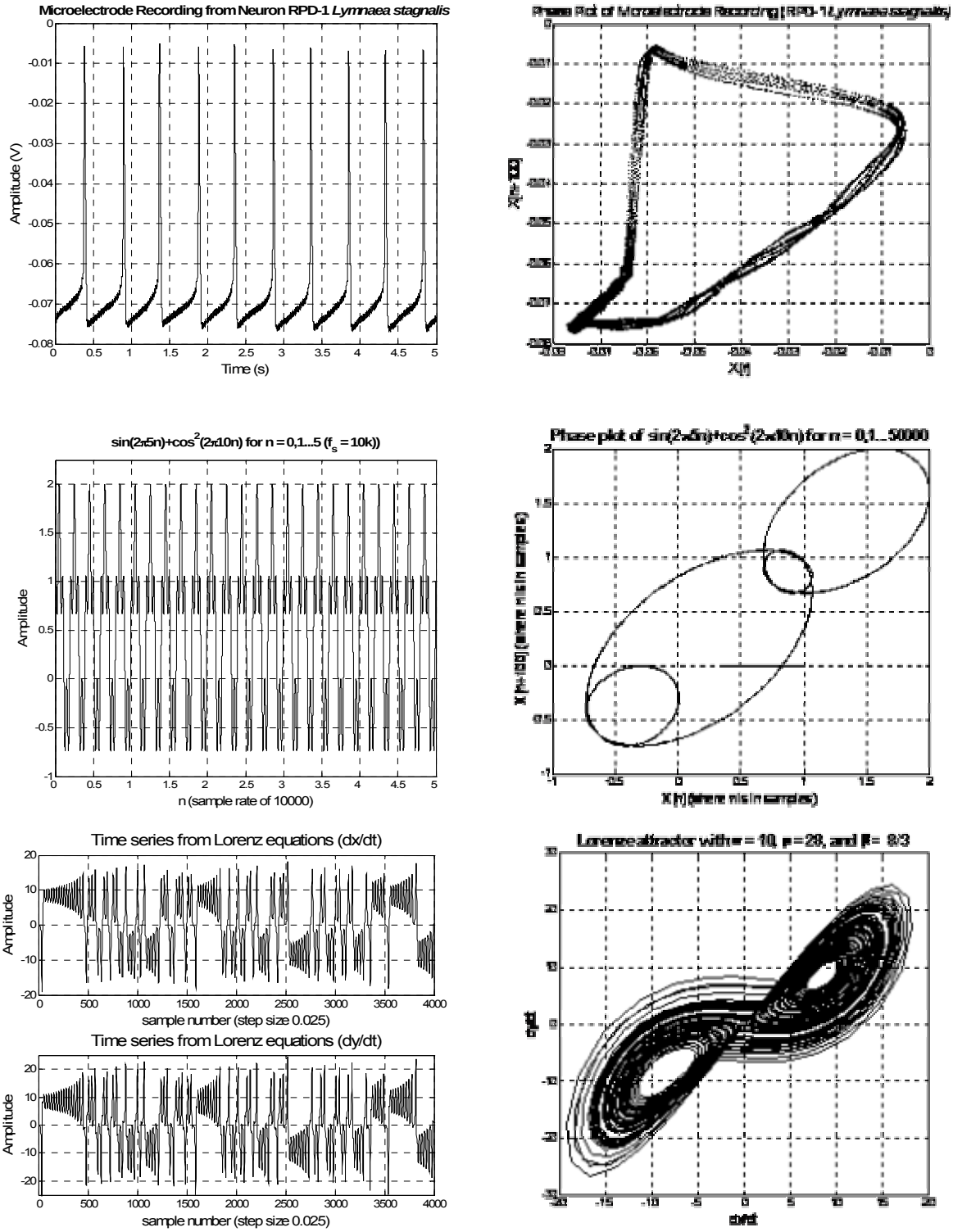
[5] Huang, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond*. A (1998) **454**, 903-995

3. Name three most useful methods, in your opinion, for analyzing the nonlinear dynamics of a one-dimensional data segment. Describe how these methods are implemented. Comment on the strength and weakness of these methods.

Before listing nonlinear analysis tools, it is worth mentioning that linear tools can often provide great insights into the time series under review. Methods such as Fourier analysis, autocorrelation, linear filters and predictions, are all valuable tools in determining the nature of the time series. Previously, we discussed methods for establishing stationarity, which is a linear process. These tools, however, can only take us so far and eventually, the nonlinearities, if they exist, will render the application of linear tools useless. When that happens, we turn to the following.

Phase plots and time-delay embedding

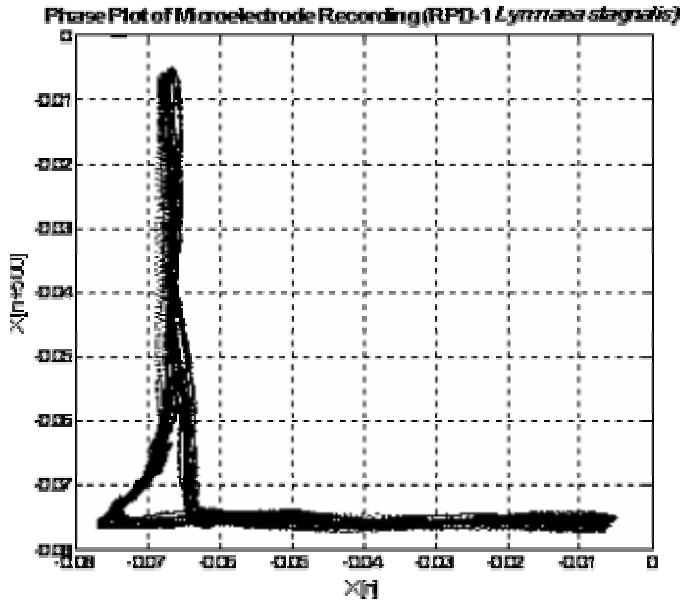
In the discussion of detecting nonstationarity, we showed an example of a phase plot to illustrate the concept of loss of recurrence. We will now look at an example dealing directly with interpreting nonlinear systems versus linear systems. Consider the following time series. The first set is a recording of action potentials from the pond snail *Lymnaea stagnalis* and its corresponding phase plot. (Simply, the phase plot is a plot of a signal versus a delay version of itself.) The second set is linear combination of a sine and cosine function. And the third is a chaotic time series, in this case the Lorenz attractor with the known chaotic parameters $\sigma = 10$, $\beta = 8/3$, $\rho = 28$ [1,2].



Top left. A time series recording from live tissue. All biological systems are nonlinear, and the phase plot (top right) does a nice job exposing features such as the bends, multiple trajectories, and the transitions from one feature in the time series to the other. Contrast that with a linear system (middle left) and its corresponding phase plot (middle right). Clearly the system is linear (periodic) as the phase plot has regular circular and elliptic shaping with no deviation from the attractor. Lastly, the Lorenz equations for $\sigma = 10$, $\beta = 8/3$, $\rho = 28$. Chaos!

A distinction is made here that phase plots and time delay embedding are not mutually exclusive. For a time series for which one has a set of coupled equations, phase plots can easily be made by plotting one equation versus another, as in the Lorenz case. However, biological recordings come with no such set of equations and so the coupling is revealed by embedding a time delay. (Takens embedding theorem) [2,3,4]

One immediate limitation of this form of analysis comes to mind. The dimension of the attractor in experimental data is unknown and must be guessed. This leads to a host of difficulties when trying to choose a delay time. For instance, if we increase the delay from the action potential recording from 100 to 500, the result collapses, and no information can be gleaned.



This is the same recording as the previous page, except that the delay has been increased from 100 to 500 samples. Note how the phase plot simply follows the rise and fall of the action potential without revealing any fine structure.

Empirical Mode Decomposition (EMD)

In EMD, the signal is decomposed into intrinsic mode functions (IMF), each of which is linear but need not be stationary. An IMF is defined only if (1) the number of extrema and the number of zero-crossings are equal or at most differ by one, and (2) the mean of the envelope of the maxima and the envelope of the minima is zero at all points. These criteria eliminate riding waves and help to smooth uneven amplitudes [5]. IMFs have properties conducive to signal processing, namely that they are linear and they are often stationary. Even if they are not stationary, it is known that IMFs have well behaved Hilbert Transforms [5].

Intrinsic mode functions are the product of empirical mode decomposition. The decomposition itself assumes the signal being decomposed has a maximum and a minimum and that it has a time scale defined by the delay between successive maxima or minima. The EMD is achieved through the following steps:

1) Upper and lower envelopes are formed from the maxima and minima, respectively.

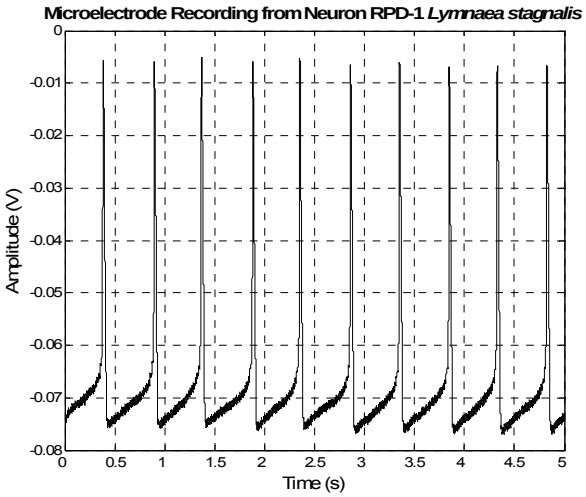
2) A cubic spline fits the envelopes.

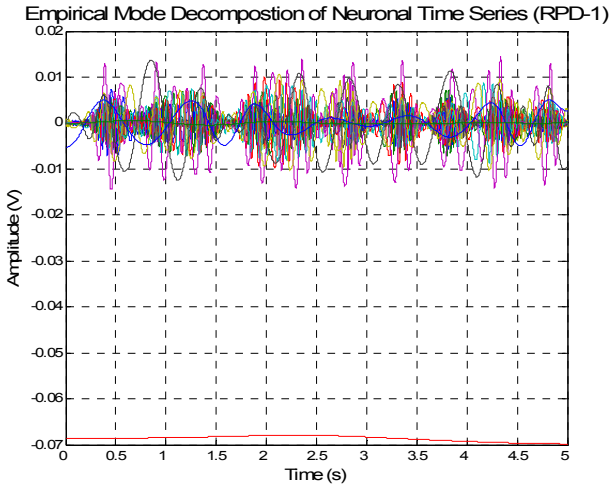
3) Calculate the mean of the two envelopes and subtract it from the data.

4) If the result satisfies the criteria for an IMF, it is the first IMF, it is subtracted from the data and process is repeated until a stopping criterion is met.

5) If the result does not satisfy the criteria for an IMF, the mean is mean is subtracted and the resulting signal is taken as the data. This process is known as sifting and continues until the IMF criteria are met.

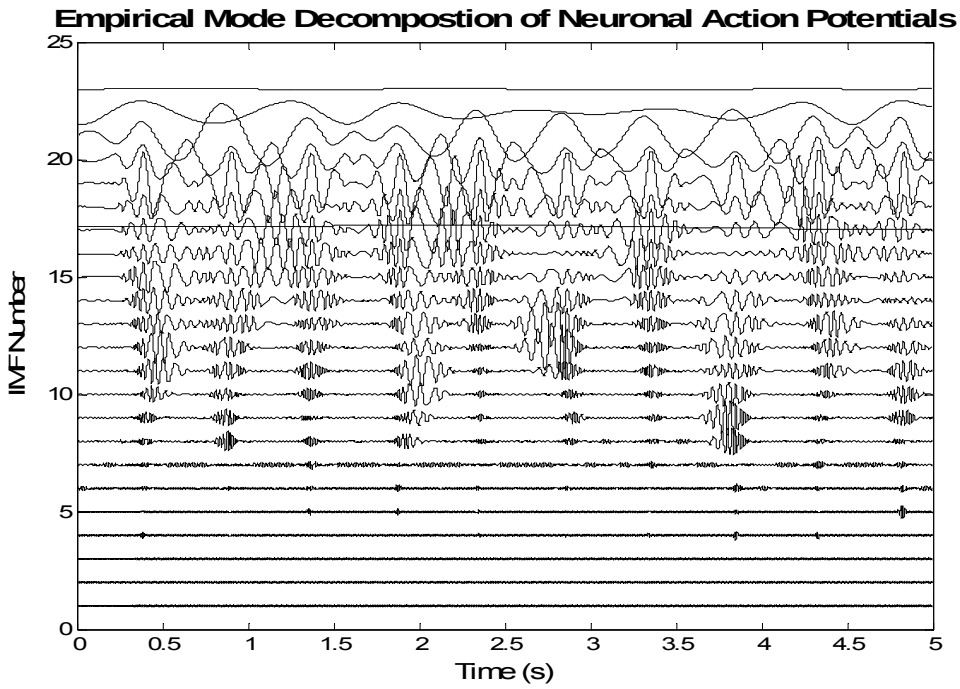
For visualization, suppose we have the following nonlinear, nonstationary signal. This is the familiar sharp microelectrode recording from neuron RPD-1 in *L. stagnalis.*



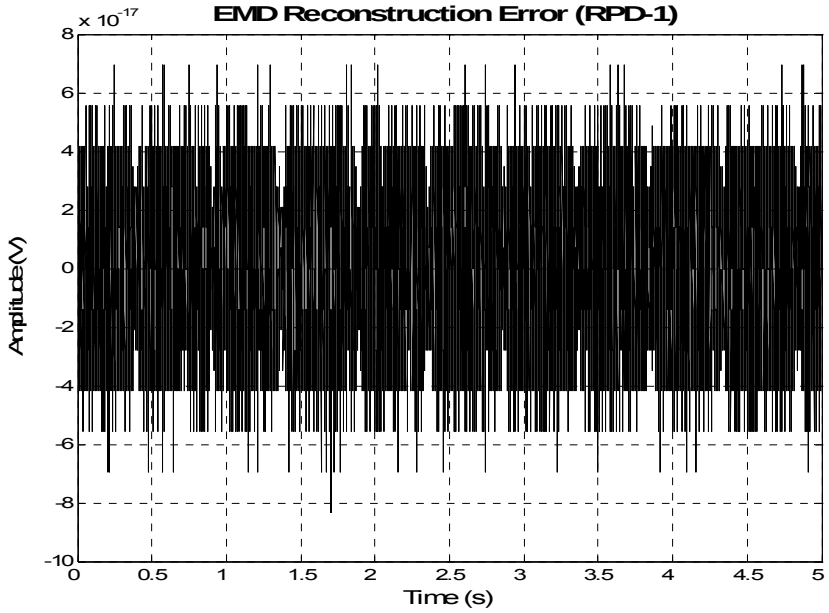


Though not glaringly clear, one can see a trend in the data with a downward slope at the base of the recording. It becomes quite clear in the IMF decomposition of the signal (right). (It is the last IMF, in red).

For a better visualization, we separate the IMFs and it is revealed that the first several IMFs contain the higher frequency components, whereas the final few IMFs contain the low frequency components.



The benefit to this type of decomposition is that it will take a nonlinear, nonstationary signal and decompose it into linear, stationary signals. This is not always the case, however, and some times the best the algorithm can achieve is orthogonal, approximately stationary signals. In the above case, the error in reconstruction is at the limit of MATLAB®'s machine error.



Lyapunov exponent

There is perhaps no single number more important to the study of nonlinear systems. Of the many criteria necessary to determine if a system is chaotic, the Lyapunov exponent characterizes the most important chaotic property – sensitive dependence on initial conditions. For a deterministic system, i.e. one that is not driven by non-random variables, the Lyapunov exponent quantifies the sensitivity of the system to initial conditions [2,3]. The exponent is a measure of the exponential rate of divergence (separation) from two vectors in state space as they follow their trajectories around the attractor. A positive exponent implies that the rate of separation is positive exponential and the system has a strong chaotic signature, provided there are at least 3 dimensions in state space [4]. A negative exponent implies that the separation is negative exponential and so will converge. This condition implies a dissipative system and the larger the absolute value of the negative exponent, the more stable the system. There is more than one Lyapunov exponent for a given system. In fact, there is one Lyapunov exponent for each dimension of the phase space. Collectively this is the Lyapunov spectrum . Usually, however, we are interested in the largest Lyapunov exponent, referred to as the maximum Lyapunov exponent (MLE) [2]. The calculation of the Lyapunov exponent from experimental data is not an easy task. First, one has to determine the embedding dimension to accurately capture the dynamics of the system that produced the time series. Of course, we cannot know this a priori and so we must test directly for the exponential divergence of nearby trajectories. The technique proposed by Rosenstein, et al, advances point by point along a trajectory in m dimensional time-delay embedded phase

space. Each point X_n is determined to have a point X_l that is closest to it within the state space. The logarithmic rate of separation of these two points over k number of time steps is averaged [2].

$$L_k = \left(\frac{1}{2(N-k-m+1)} \right) \sum_{n=m}^{N-k} \log \sum_{j=0}^{m-1} \left(X_{l-j+k} - X_{n-j+k} \right)^2$$

The slope of the linear portion of the curve generated by $\frac{dL_l}{dk}$ at intermediate values of k is the largest (average) Lyapunov exponent. Small values of k should be ignored since the points are not aligned along the maximum expansion vector. Likewise, large k should be ignored since eventually k will be as large as the entire attractor where no further separation can take place. Clearly the larger the value of k, the better the approximation to the true Lyapunov exponent will be. There are inherent difficulties with this approach, however. Obviously, it is computationally extensive for large data sets. It is also prone to spurious results since the attractor may not be well formed for a given embedding dimension m.

[1] Kantz, H. and Schreiber, T. Nonlinear Time Series Analysis, 2nd edition. Cambridge University Press, Cambridge, UK. 2005.

[2] Sprott, Julian. Chaos and Time-Series Analysis. Oxford University Press, New York, 2003.

[3] Hilborn, R. C. Chaos and Nonlinear Dynamics, 2nd edition. Oxford University Press, Oxford, UK. 2000.

[4] Strogatz, S. Nonlinear Dynamics and Chaos. Westview Press, Cambridge, MA, 1994

[5] Huang, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond*. A (1998) **454**, 903-995

4. Identify three significant papers that are concerned with the applications of nonlinear dynamics to biomedical or biological data. Briefly summarize each paper. Comment on why you think each study is significant.

Elson, et al, "Synchronous Behavior of Two Coupled Biological Neurons" (1998)

The research leading to this paper was conducted by very well respected and well known scientists in both electrophysiology and nonlinear dynamics.
The paper attempts to answer the questions relating to neuronal dynamics and synchronization. The researchers connected two biological neurons via a dynamic clamp that introduced an artificial synapse between them. The neurons were pyloric dilator neurons from the stomatogastric ganglion of the California spiny lobster, *Panulirus interruptus*. The neurons are known to be electrically coupled and exhibit weak synchronization, or state-dependent synchronization [1]. The group showed that by varying parameters such as coupling strength and membrane conductance, the neurons could be forced to synchronous behavior of not just the slow oscillations of reciprocal inhibition, but the fast spikes as well.
What makes this paper significant is the application of nonlinear systems analysis to biological systems. Biological systems are of course inherently nonlinear but before this paper, investigations of neural networks were studied using probabilistic models. Synchronous behavior is responsible for a host of pathological conditions including epilepsy and heart attacks. The ability to accurately describe the coupling effect as it relates to synchronization has led to predictive models and a greater understanding of neural systems.

Rieke, et al, "Discerning Nonstationarity From Nonlinearity in Seizure-Free and Preseizure EEG Recordings From Epilepsy Patients" (2003)

Earlier in this examination, a technique for analyzing the loss of recurrence was discussed in application to detecting nonstationarity [2]. This technique was applied to electroencephalogram (EEG) recordings. It is known that nonlinearity and nonstationarity are independent properties of the EEG and that nonstationarity, in the presence of nonlinear signals, provides valuable biological information. A formal treatment of the nonlinear detection algorithm, iterative amplitude adjusted Fourier transform surrogate data (IAAFT), is beyond the scope of this discussion, but the basic idea is provided for reference. It is well known that many different signals can have the same power spectrum density distribution and share the same autocorrelation function. It is possible then to maintain these properties of a time series and search for a random (Gaussian) signal that shares the same properties. This is the basis of surrogate data and is used as a comparison to test determinism and linearity. The reason for this is quite simple: nonlinear systems very rarely generate Gaussian distributions. The null hypothesis is that the original time series can be reconstructed with random surrogates. Many improvements exist that allow rescaling of the Gaussian linear process [3].

The paper is significant in that it provides a combined measure for nonlinearity and nonstationarity, and demonstrates that important information is contained in those properties. Too often we rush to linearize and detrend. This paper provides clear evidence that to do so would be to "throw out the baby with the bathwater."

Salisbury, J. and Sun, Y., "Assessment of Chaotic Parameters in Nonstationary Electrocardiograms by Use of Empirical Mode Decomposition" (2004)

The knee jerk reaction to choosing this paper would be a belief that the author was trying to curry favor with the reviewer. While on the surface that may be plausible, I submit that is not the case. The author began graduate studies in the fall of 2004 and was given this paper to read. It was the author's introduction to the Empirical Mode Decomposition (EMD) and began, what is to date, a "several" year march toward a PhD based on nonlinear dynamics.

A summary treatment of the EMD technique was provided earlier in this work. In the Salisbury-Sun paper, the application of this technique to the American Heart Association (AHA) database is used to determine the chaotic parameters present in electrocardiograms (ECG). It is well established that biologically generated time series are inherently nonlinear. Certain features of the ECG, namely the RR interval, are known to be chaotic. However, it is often difficult, if not impossible, to apply meaningful analytic metrics to nonstationary data. The EMD decomposes nonstationary data into linear (or at least orthogonal) and stationary IMFs and so is well suited to apply chaos analysis, as that analysis requires the data be stationary.

The paper is significant in that it touches on a host of techniques for handling nonlinear dynamical systems, namely Lyapunov exponents, nonlinear detection via the bispectrum analysis, phase-space reconstruction, and correlation dimension.

[1] Elson, R. C., et al. Synchronous Behavior of Two Coupled Biological Neurons. *Phys. Rev. Lett.* (1998) **81** 5692-5695.

[2] Rieke, et al. Measuring Nonstationarity by Analyzing the Loss of Recurrence in Dynamical Systems. *Phys. Rev. Lett.* (2002) **88** 244102-1 - 244102-4

[3] Rieke, et al. Discerning Nonstationarity From Nonlinearity in Seizure-Free and Preseizure EEG Recordings From Epilepsy Patients. *IEEE Trans. Biomedical Engineering*. (2003) **50**, 634-639.

[4] Salisbury, J. and Sun, Y. Assessment of Chaotic Parameters in Nonstationary Electrocardiograms by Use of Empirical Mode Decomposition. *Annals of Biomedical Engineering*. (2004) **32**, 1348–1354.