MARGINAL OSCILLATOR SENSITIVITY ENHANCEMENT USING FULL-STATE NONLINEAR FEEDBACK

BY

MARK D. GINSBERG

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Professor Peter W. Sauer, Chair
Professor Naira Hovakimyan
Professor Arne J. Pearlstein
Professor William D. O'Brien, Jr.
Associate Professor Jeffrey L. Schiano, The Pennsylvania State University

ABSTRACT

A marginal oscillator is a tank circuit with nonlinear output feedback applied to maximize the change in amplitude with respect to the circuit's internal resistance. Although used in many applications, the marginal oscillator is most commonly used in continuous wave magnetic resonance (CW-MR). Continuous-wave is useful under two circumstances. The first is when attempting to find previously undocumented magnetic resonances over a wide range of frequencies. An individual resonance may yield a peak that is only a few kilohertz wide where the search space may span many megahertz. Hence a search may require many hours to complete. Second, CW requires much less power than Fourier or pulsed techniques; this is very useful in field applications, and to avoid quenching superconducting search coils.

The currently accepted mathematical model describing a marginal oscillator leads to transcendental analytical expressions that can only be approximated. It also lacks a known path to optimize the nonlinear feedback policy.

This dissertation describes a redesign of the marginal oscillator using state-space modeling and feedback of all state variables (i.e. full-state feedback).

This achieves several goals, all of which were unachievable using previous analysis. First, the resulting mathematical model, although still nonlinear, can be described in closed form. Second, the circuit model can be revised to better resemble laboratory instrumentation and can be implemented in hardware or software. Third, for this and previous designs, it had been observed that conversion-gain is proportional to the settling time of the circuit. Under very loose constraints, this observation is now proved as a theorem. Alternative measurement methods using the marginal oscillator at smaller conversion gains are briefly discussed. Fourth, the state-space model is mapped to a dimensionless coordinate system inducing data collapse. Therefore, at each data sample the oscillation amplitude is well characterized, where current methods that estimate a signal envelope from the output voltage

are susceptible to phase noise. Fifth, the effect of parasitic resistance in the switched capacitor/varactor bank is analyzed. At frequencies near resonance, this is shown as equivalent to changing the resistance of the idealized lumped circuit model.

To my family: past, present, and future, for their love and support.

ACKNOWLEDGMENTS

It should be emphasized that I am older than the usual graduate student. Indeed it is unusual to obtain a doctorate after years of research both as principal investigator and as program manager. To be candid, I have caused more than my fair share of turbulence: for my various mentors, the Department of Electrical and Computer Engineering staff, and the Graduate College advisors. To be frank, the following list simply cannot be exhaustive.

For the folks who do require a sincere and formal "Thank you!" I am listing you in loose order of appearance with respect to this project.

Jeff Schiano was a former student employee at CERL. He prodded me to work on this project and pledged his support on the condition that I go through formal graduation so he could hood me. Thanks for letting me use the exhaustive database of papers you have collected on marginal oscillators. Thanks for providing a wealth of technical information and experience gained in years of laboratory experiments. Your vision for this project did not include a comprehensive redesign of a marginal oscillator using state-space methods. I'm gratified that you followed along patiently previous to the time it became clear that the new design would yield advantageous new properties.

Alexander Scheeline graciously agreed to be my "research mentor" when it became apparent that I was grappling with problems beyond the experience of my previous Corps of Engineers mentors. Thank you for the hours of kind counsel on many topics for which I only had self-training, informal training, or in many cases, none at all. Thank you for the fast editing of my drafts by chapter. Your quick responses meant more than I can say.

Sean Meyn served as my advisor until he took a new position at the University of Florida. He was unfamiliar with marginal oscillators, and was

brave enough to work with me anyway.

Andrew Singer served as my advisor until just shy of my preliminary exam. He took a significant risk working with me knowing that my chosen problem was somewhat outside his expertise.

Arne Pearlstein was the only person at UIUC able to read my chapter on Bioprotection of Facilities with genuine understanding and insight. On that basis he also agreed to be on my committee when others had ducked. Thanks for the vote of confidence when I needed it badly. His highly detailed review of the dissertation was exceptionally helpful and much appreciated.

Peter Sauer served as my academic advisor from the time of my preliminary examination onward. Thank you for being an infinite source of common sense when I needed it most.

Naira Hovakimyan served as my technical advisor even though my research did not represent any of her current projects. She was an absolute blessing at a moment when I felt totally abandoned. She welcomed me to her group meetings, giving me a temporary home where I could interact with some of her students.

William D. O'Brien served as committee member on very short notice. Afterwards, he was kind enough to welcome me to his group meetings as well.

To my friends and family who had to endure many moments when I simply could not be as attentive to their needs as I would have liked, thanks for your support and forgiveness.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ASHRAE	American Society of Heating, Refrigerating, and Air-Conditioning Engineers
CCD	Charge Coupled Device
CEE	Civil and Environmental Engineering
CERL	Construction Engineering Research Laboratory
COTS	Commercial Off The Shelf
CW	Continuous Wave
DARPA	Defense Advanced Research Projects Agency
DOD	Department of Defense
ERDC	Engineer Research and Development Center
EPA	Environmental Protection Agency
ESR	Equivalent Series Resistance
FET	Field Effect Transistor
HEPA	High Efficiency Particulate Air
HVAC	Heating, Ventilation, and Air Conditioning
ICR	Ion Cyclotron Resonance
IEEE	Institute of Electrical and Electronics Engineers
IR	Infrared
ISC3	Industrial Source Complex version 3
KCL	Kirchhoff's Current Law
MERV	Minimum Efficiency Reporting Value

NIST	National Institute of Standards and Technology
NMR	Nuclear Magnetic Resonance
NQR	Nuclear Quadrupole Resonance
ODE	Ordinary Differential Equation
ORISE	Oak Ridge Institute for Science and Education
PCB	Printed Circuit Board
PF	Protection Factor
RDX	Research Department eXplosive, or Trimethylenetrinitramine
RLC	Resistive Inductive Capacitive
SNR	Signal-to-Noise Ratio
SQUID	Superconducting Quantum Interference Device
TNT	Trinitrotoluene
UIUC	University of Illinois at Urbana-Champaign

LIST OF SYMBOLS

$\boldsymbol{A}$	Matrix
$\boldsymbol{B}$	Column vector
ω	Natural frequency
τ	Characteristic time constant
θ	Angle in cylindrical coordinates
$\widehat{R}$	Estimate of resistance R
A	Amplitude
a, m, γ	Scalar constants
C	Capacitance
E_T	Total energy at resonance
f	Frequency
$F(r)$	Feedback function in cylindrical coordinates
$G(v)$	Feedback function
$g(v)$	Feedback function
g_A	Transconductance slope A
g_B	Transconductance slope B
G_C	Conversion gain
$h(r)$	A linear function of r
$I(s)$	Current in the Laplace transform domain
$i(t)$	Current

L	Inductance
r	Radius in cylindrical coordinates
R	Resistance
s	Laplace parameter
S_R^A	Conversion gain of A with respect to R
t	Time
$V(s)$	Voltage in the Laplace transform domain
$v(t)$	Voltage
x	Cartesian coordinate and scaled voltage
y	Cartesian coordinate and scaled voltage

PREFACE

The articles resulting from the current research effort are listed in chronological order.

Bio-inspired Engineering Design: Exoskeletal Sensors

The material in Appendix B on page 131 originally appeared as "A case study in bio-inspired engineering design: defense applications of exoskeletal sensors," M. Ginsberg, J. Schiano, M. Kramer, and M. Alleyne, *Defense & Security Analysis,* vol. 29, no. 2, pp. 156–169, 2013 [1].

This article is motivated by my long-term interest in physiological foundations of psychology and material gleaned from courses taught by C. Trahiotis on audio perception, and J. Malpeli on visual perception. A central success story for the entire field of studying neurophysiology by direct comparison to engineered systems is echolocation in bats [2, 3]. Physiological processes commonly use nonlinear signal processing and feedback. At the suggestion of Prof. Andrew Singer, I polled Prof. Gene Robinson for a collaborator. Dr. Robinson recommended Marianne Alleyne, and this article resulted.

The article is mostly an exercise in phenomenology. As funding was unavailable for experimental follow-up, the results have not been enriched with dynamical descriptions and analysis.

However, this work represents a rich source of new research problems, and much follow-up work has already taken place [4], and will continue.

With this many co-authors, it is reasonable to ask about individual contributions. First, Jeff Schiano originally put forward the suggestion that the marginal oscillator might serve as an initial model for some natural forms of nonlinear feedback. The idea is not far-fetched when one realizes that the marginal oscillator requires nonlinear feedback to achieve a large conversion

gain. Second, while I served as a supervisor at the local and state levels of the Science Olympiad, I met Megan Kramer, who is a gifted undergraduate student and is on full scholarship at UIUC for membership on a team that won the national Science Olympiad. I offered her the opportunity to collaborate on the article as an act of mentorship. She served as a very capable editor, and made several helpful suggestions on clarifying the writing that were totally beyond her age and apparent experience. Finally, I had the honor of attending two class sessions of Marianne Alleyne's course on insect physiology where she covered the sensor material and listed many helpful original reference sources for further reading. She was a tremendous help in suggesting correct turns of phrase where my descriptions became needlessly wordy. She is also the artist who produced the biological illustrations that appear on the left-hand side of most of the figures.

I have a long-term goal to continue research in nonlinear feedback control in the hopes that the resulting information will eventually inform analysis of nonlinear feedback as observed in various physiological systems.

Bioprotection of Facilities

The material in Appendix A on page 105 originally appeared as "Bioprotection of facilities," M. D. Ginsberg and A. T. Bui, *Defense & Security Analysis*, vol. 31, no. 1, pp. 1, February 2015 [5]. It subsequently won the 2015 Engineer Research and Development Center, Research & Development Achievement Award for Technical Excellence.

Although originally not planned as a part of the dissertation, the result is clearly at the intersection between biology and control and decision theory, so it is a very natural addition. The article has been well received, resulting in invited talks. It is currently scheduled to inform changes to building design standards for the DOD (in process) [6]. As the analysis clearly exhibits the system identification methods directly from my control theory courses, it is included here because I am sincerely grateful.

The results of this work were somewhat surprising on several levels. Although the setup of the ordinary differential equations is considered utterly standard, all previous results describing an external contamination event in the open literature were run as simulations using numerical approxima-

tions (most commonly using a NIST software package CONTAM, using the Newton-Raphson method). W. R. Ott, formerly of the Stanford Statistics Department, lately consulting professor in the CEE Department, successfully made the connection to transform methods while analyzing cigar smoke traveling from room to room with no HVAC system modeled [7]. Actually solving the differential equation using transform methods and then examining the resulting algebra for patterns appears to be new. In the process it was shown that the *protection factor*, a term of art usually associated with gas masks, is the central quantity of interest in the algebraic solution.

Another surprising result is that the HVAC system is not a candidate for implementation of feedback control. This is established by several facts.

First, using commercial-off-the-shelf filters imposes a very small energy penalty as the filter is made more efficient. All commercial manufacturers are aware that air filters have to be designed to be compatible with the fan curve of the system's blower motor. As filters are made more efficient, they are simply made larger and folded into pleats to present a larger cross-sectional area to the air stream.

Second, while implementing the immune building program, the Army and DARPA were surprised that using military standard M-98 HEPA filters imposed markedly increased energy costs. However, the Navy designed M-98 filters not for energy efficiency, but for an ability to fit into pre-existing ship compartments. Hence they were poor candidates for fixed infrastructure where extra room is usually not a problem, and energy efficiency and life cycle cost are paramount.

Third, all current contaminant-specific sensors are far too slow to activate feedback control of a building's HVAC system. With delay times of about 10 minutes, the building's occupants will have breathed too much contaminated air before the feedback system is activated.

With a co-author listed, a quick comment on individual contributions is warranted. I was officially named as a mentor for my co-author, Alex Bui, under the Oak Ridge Institute for Science and Education (ORISE) internship program. Alex is a capable rising star in civil engineering, and was responsible for tracking down numerical guidelines for HVAC design of existing structures. He also produced the initial draft of the diagrams for the paper.

Marginal Oscillator Sensitivity Enhancement Using Full-State Nonlinear Feedback

Chapters 1 to 6 represent the analytical core of my dissertation. As originally conceived, it was expected that two themes would emerge. First, this is a *compositional problem* where one would like to optimize a nonlinear feedback function that performs a useful job, and where the same job is impossible with linear feedback. (In contradistinction to, for example, demonstrating an implementation of the Lorenz equations as a mere curiosity.) Second, optimization was initially thought to be an application of the calculus of variations.

As my subsequent work shows, this is an excellent example of a compositional problem. However, this problem is not yet a candidate for the fine tuning of the calculus of variations. Instead, what I show is that rigorous treatment of this problem requires nothing less than a paradigm shift. Instead of using output feedback, a solution using full-state feedback completely changes long held assumptions about CW spectrometry. In particular, for decades it was assumed that *time to detection* was proportional to sensitivity, and that this coupling fundamentally limited the speed of detection. The analysis presented clearly shows that full-state feedback provides a closed-form analytical description of all limit cycles. It also shows that the estimates of resistance from nearby limit cycles, which are unaffected by poor time constants, yield a sufficiently accurate measurement of the unknown resistance. Hence, there is no need to be concerned with the poor time constant yielded by a large conversion gain. The analysis here also yields a reasonably simple feedback law that is suitable for implementation in hardware. The simulation results show the improved performance over the existing two-slope limiter.

Publication Plan

This research has resulted in two peer reviewed publications [1, 5]. The remaining work has yielded one additional manuscript submitted to *IEEE Sensors Journal*.

In addition, the proposed method represents a general recipe for improving the sensitivity of existing sensors. This may help improve sensors applicable to diverse environmental problems in addition to those given in Section 1.2.

CHAPTER 1

INTRODUCTION

1.1 Overview

In the post 9–11 environment, there has been increasing emphasis on research directed toward rapid detection of standard explosives such as TNT and RDX. There is a need for detection devices that are fast, cheap, ubiquitous, and easy to use.

In recent years, new explosive materials have been developed at an alarming pace. As part of asymmetric warfare, terrorists and other non-governmental actors have become more resourceful at using widely available materials to synthesize new weapons, including explosives. These new materials will typically elude existing detection devices. Many governments are currently leveraging new computational techniques to design explosive compounds in simulation, and develop the most promising ones experimentally.

In this age of rapid development of new explosive compounds, it would be desirable to devise detection and quantification methods that work both for readily available explosive compounds, and would also be capable of rapid adaptation for newly developed compounds.

In my Army related research, I have been actively pursuing many different technologies for identification and quantification of materials. The ideal is to use technology that is rapidly and seamlessly adaptable as new threats arise. One such technology that holds much promise for future development is magnetic resonance. Magnetic resonance is a standard technique to detect and quantify the presence of specific chemical species. The resonance peaks associated with a particular chemical (i.e., amplitude as function of frequency) occur in readily identifiable patterns. These patterns are specific enough that the method can discriminate between chemical *isomers* (i.e., same chemical formula, but different molecular structure). Any new explosive compound

can be quickly assayed for its resonance peaks at one laboratory, and the resulting pattern could be sent to magnetic resonance detectors world-wide in a matter of hours.

In recent years, pulsed magnetic resonance techniques have been more commonly used in laboratory-based settings. However, compared with a pulsed magnetic resonance device, a continuous wave magnetic resonance device uses less power, is capable of detecting and quantifying a wider variety of solid materials, characterizes broad classes of new solid materials more rapidly, and is far cheaper to construct and run. Hence, it is well worth considering if the continuous wave technique should be updated to bring it into more common usage.

It is important to note that magnetic resonance measurements of solids are slow, lasting from seconds (detection) to many hours (complete characterization of a frequency signature). This is because the measured quantities are very small and susceptible to the presence of noise from many sources. These noise sources slow the measurement time by requiring that the same data be taken multiple times, and then averaged to achieve noise reduction. Therefore, each time one can eliminate or reduce a noise source, it can have a large effect on reducing measurement time.

I had served as program manager on several projects whose purpose was to continue development of continuous wave resonance. To be blunt, great as the promise was, scientific progress seemed slow and laborious. When it came my chance to select a topic for this dissertation, I felt it was a unique opportunity for me to dive into the technical details of the work and determine if I could facilitate faster progress.

The electronic device used in continuous wave magnetic resonance is called a *marginal oscillator*. The idea behind the device is simple to state. Create an oscillator tunable to any one frequency in a range of interest. The oscillator is constructed using an inductor/capacitor combination. One brings the material to be characterized into the oscillating magnetic field of the circuit's inductor. If the material has a magnetic resonance near the frequency of oscillation, the material will begin to draw energy out of the oscillating magnetic field. If this occurs, the marginal oscillator should be designed so that its amplitude changes markedly to indicate that energy is being absorbed by the sample.

The marginal oscillator was first mentioned in Pound and Knight [8]; however, it was not given a satisfactory theoretical description until the work

of Viswanathan et al. [9]. As it turns out, the work of Viswanathan et al. is an excellent description of why the nonlinearities associated with using vacuum tubes or field effect transistors (FETs) were required for the marginal oscillator to work. However, the mathematics used to describe the system were so complicated that they did little to inform researchers about how to change the circuit to achieve any design goal.

This dissertation updates the system analysis using more modern mathematical tools. By doing so, we gain several advantages over the previous work. (Only the most general are listed here.) First, the resulting mathematical expressions are far easier to interpret. They now facilitate the work of those interested in changing the circuit's design. Second, the new analysis is not limited to describing circuits made with FETs and vacuum tubes. It easily describes more general and flexible methods that can be implemented with digital electronics.

In addition we show that the new mathematical descriptions lead to immediate improvements in previously used data analysis techniques. First, previous work had attempted to estimate the magnitude of a time varying oscillating signal using the signal's envelope by interpolating the low frequency outline, or envelop, of the signal from high frequency raw data. Digital sampling of such signals creates additional noise (called *phase noise*) under most circumstances. The new analysis shows that obtaining a measurement of the total energy stored in the circuit eliminates this phase noise. This speeds up measurement by reducing the number of samples taken to achieve a given quality of noise reduction. Second, the analysis allows us to prove that the system's signal gain is proportional to the circuit's settling time. It shows that one can actively control the circuit's gain and settling time. It also shows that it is not necessary to always tune the circuit for high gain, thus imposing a slow settling time. Using this new flexibility, one can increase the speed of measurement. Third, a specific class of parasitic resistance in the circuit is analyzed. The net effect of these parasitic resistances is shown to be equivalent to a very simple change in the circuit model.

As significant as these developments are, they only represent the first fruits of what can be achieved with the updated analysis. The most significant advance is that the door is now open for much more adventurous ideas in how a marginal oscillator can be configured to achieve desired design goals.

1.2 Previous Engineering Applications

The marginal oscillator is sensitive to energy being drawn from the circuit by a sample exposed to the oscillating magnetic field of the circuit's inductor. Experienced researchers sometimes call materials that absorb energy *lossy*.

The following list gives a description of previous marginal oscillator applications. (Some of these descriptions are difficult to write about because the marginal oscillator is being used to test the condition of another sensor. Hence I will use "sensor" for the underlying application and "marginal oscillator" for the instrument being used to examine the sensor.)

NMR/NQR Transitions in Solids In these devices, energy is absorbed by the material under test because the material possesses a magnetic resonance peak somewhere near the frequency of oscillation. A spectrum of absorption as a function of frequency yields a pattern that can distinguish between isomers (same chemical formula, different molecular structure), and also distinguish between polymorphs (same molecular structure, but different crystal structure).

Roberts and Rollin credit Pound with developing the first marginal oscillator in the 1940s to observe nuclear magnetic resonance (NMR) transitions within solids [8, 10, 11]. The continuous wave resonance method has several advantages.

a) Measurements can be taken at much lower field strength, and therefore consume much less power. This enables applications requiring human exposure, e.g., detecting explosive material in a shoe that is still on a human, or using non-ionizing radiation to image the bones of pregnant women, etc. Further, at lower power, the technique is more suitable for field applications, e.g., measuring environmental impact of energetic materials.

b) For materials having a long spin-lattice relaxation time, T_1, the continuous wave technique is faster than pulsed techniques.

c) For search coils possessing a high Q-factor, such as high-temperature superconducting coils, using the lower power greatly reduces the risk of quenching the superconductor.

Defects in Irradiated Silicon In high-energy accelerators, it was known

that neutron scattering degraded high purity silicon components, especially sensors. Under specific experimental conditions, the silicon exhibited a change in conductivity indicating damage. Alexiev et al. [12] found that using a marginal oscillator was superior to previous techniques that could only characterize silicon to a predetermined skin-depth. As the entire suspect sensor could be used as a paramagnetic sample (when placed in the field of the marginal oscillator's inductor), thus the marginal oscillator could characterize the conductivity of the entire sensor and assess radiation damage.

Skin-depth of Superconductors Superconducting materials can be used to build very sensitive instruments including superconducting quantum interference devices (SQUIDs). However, the synthesis of superconducting materials is a difficult process that requires careful characterization of the material produced. For many applications, it is important to obtain a measurement of the conducting skin-depth of the synthesized material. Gauzzi et al. [13] showed how marginal oscillators used in a noncontact application could characterize skin depth both in bulk samples and thin films of superconducting material. In some cases his apparatus was able to characterize the magnetic penetration depth to within $\pm 1\,\mathrm{pm}$. Again, because the superconducting material is placed within the marginal oscillator's magnetic field, the conductivity of the sample under test can be measured in bulk.

Ion Cyclotron Resonance ICR is still used in modern mass spectrometers to separate particles by their mass-to-charge ratio. Warnick et al. [14] applied a marginal oscillator to an ICR to characterize the amount of material present at a characteristic mass-to-charge ratio. Although he was able to show that his marginal oscillator based instrument outperformed the best commercial unit available at the time, limitations of the measurement circuitry prevented an exact analysis of the quantitative performance of his new unit. More recently, this application has come to rely on Fourier transform ion cyclotron resonance, where the mass-to-charge ratio of ions is measured as a function of their cyclotron frequency in a fixed magnetic field.

Thin Films Characterization G. L. Miller et al. [15] were interested in

characterizing thin films undergoing ion implantation, laser alloying, and ion beam mixing. They were able to fabricate a mechanical marginal oscillator using photo-lithography techniques. Here the material under test is formed into a tiny cantilever beam. The beam will have a characteristic resonance as a function of frequency. As the beam is subjected to various test conditions, the frequency of resonance will shift if the beam's mechanical properties are changed. The resulting devices were demonstrated to successfully measure surface changes to the cantilever beam, sometimes even phenomena with a skin-depth of less than 1 µm. Miller indicates that his devices were not used for quantitative measurements so much as they were used as a rapid survey to indicate changes in mechanical properties.

Curing of Plastics In a patent application, Thomas [16] used a marginal oscillator that was acoustically coupled to curing plastic, yielding non-contact measurements of the plastic's cure-rate in real time. The pre-cured plastic had liquid properties that caused acoustical losses, where the cured plastic was more solid and less acoustically lossy.

Precision Measurement of Capacitance Using a modified Wein bridge that included a marginal oscillator, Robinson [17] was able to measure capacitances in the range 0.1 µF to 1 µF with an accuracy of 0.1 %. He points out that many cheap instruments in his lab will measure capacitance to a precision of 0.1 %, but only the most expensive and exotic of these will yield an accuracy of 0.1 %. A strange feature of his device is that he uses simple circuitry, yet obtains a high accuracy capacitance measurement for three different capacitors simultaneously.

Precision Measurement of Low Temperatures At liquid helium temperatures, Betts et al. [18] used a sample of paramagnetic salt (cerium magnesium nitrate) to measure temperatures in the range of 0.1 K to 4 K. This works because the resonant frequency varies as inverse temperature, $1/T$, over a wide range of temperature. The limits of this temperature range are imposed by details of his experimental setup, not the salt itself.

1.3 Biological Sensors

Organisms exploit the nonlinear behavior of materials to increase the sensitivity of biosensors to stimuli [19]. For example, campaniform sensillum, a strain sensor found in insects, relies upon the nonlinear response of a membrane structure to deflection forces to amplify the strain-sensing property of the microstructure [20]. As another example, the fire beetle uses a thermo-pneumatic transduction mechanism that consists of a cavity covered with an ultra-thin film [21]. The deflection of the film is a linear function of temperature except near a small region centered near room temperature. Within this region, where the film deflection is a nonlinear function of temperature, the sensor has the greatest sensitivity to infrared light [22, 23]. While these studies have resulted in biometric strain-sensing microstructures and microfabricated cavities for IR micro-imaging [20, 22], they have not provided a fundamental understanding of how the nonlinear material behavior improves detection sensitivity.

At the outset, I had hoped to show a nontrivial connection between the marginal oscillator and a known nonlinear biological sensor; however, this initiative had to be dropped due to lack of funding. Hence, this conjecture still remains unproven. Therefore the contents of Appendix B have been published [1], but are, sadly, still disconnected from marginal oscillators.

1.4 Bioprotection of Facilities

During the course of my marginal oscillator research, I had the opportunity to write a paper that now appears in Appendix A [5]. Again, this paper is not directly involved with marginal oscillators, but at the same confluence between electrical engineering analysis, biology, and national security concerns regarding protection of infrastructure. I had been sheepish about including it at first. However, it subsequently won an award from the Army. So it is still included.

1.5 Technical Overview

1.5.1 Research Objective

The objective of this research is to do the following: gain a fundamental understanding of why incorporating a sensor into a nonlinear feedback loop yields a closed-loop system whose sensitivity to the measured variable is far greater than that of the standalone sensor; for such nonlinear feedback, determine the fundamental relationships between sensitivity and sensor bandwidth of the closed-loop sensor system; and test the predicted relationships using computer simulations.

We make the following hypotheses. First, one can add considerable mathematical richness to the marginal oscillator by analyzing the system's properties in phase space. Second, the phase-space description should yield a new marginal oscillator design based on a memoryless nonlinear feedback path around the sensor. Third, direct measurement of the circuit's amplitude requires fewer samples than detection of the signal envelope of the output voltage. This maintains the sensitivity of a marginal oscillator while greatly decreasing the time required to detect a particular quantity (or analyte) of interest.

1.5.2 The Central Application NMR/NQR

Magnetic resonance frequencies are sensitive measures of the chemical structure of the analyte being examined. In the case of nuclear magnetic resonance (NMR), the resonant frequencies are determined by the interaction between any nuclei with spin 1/2 and the nearest neighboring atoms in their molecule. In this way one can discriminate between hydrogen bound to oxygen versus hydrogen bound to carbon, etc. The most common measurements determine the molecular bonds for hydrogen- and phosphorous-bearing compounds.

Nuclear quadrupole resonance (NQR), is also a sensitive measurement of chemical structure, and yields the best information regarding nuclei that have a spin of 1, most typically nitrogen. In this sense it is similar to NMR in its ability to discriminate between chemical isomers (same chemical composition, but different molecular structure). In addition NQR is an order of magnitude more sensitive to temperature, pressure, and any statistically

consistent strain applied to a particular molecular species. Therefore it is also potentially a method for non-destructive evaluation of materials made from a particular molecule. As almost all energetic materials (explosives, fuzes, etc.) are nitrogen-bearing compounds, NQR is a potentially important method for detection and identification of explosives. In addition, specific types of chem-bio based poisons and narcotics are also nitrogen-bearing and, therefore, susceptible to detection via NQR.

The marginal oscillator has been used for "continuous wave spectroscopy." Historically, it has been superseded by pulsed spectrometry techniques. However, it still has some unique advantages that should be developed further. First, a CW spectrometer is approximately 100 times cheaper to implement in hardware. Most pulsed spectrometers require a massive radio frequency power amplifier. This drives up both cost and power consumption. In contrast, a viable continuous wave instrument can be assembled with off-the-shelf parts. Further, a CW instrument exposes a sample analyte to very low-power radio waves. CW measurements not only consume less power, but also dump much less energy into the search coil. Therefore CW is a far better technique for superconducting search coils, being much less likely to quench the superconducting material used.

1.5.3 Previous Analysis of the Two-Slope Limiter Based Marginal Oscillator

A conceptual representation of a marginal oscillator appears in Fig. 1.1b on the following page. The diagram shows a feed forward path representing the dynamics of the sensor, and a feedback path containing a memoryless nonlinear function. The nonlinear feedback element sustains a steady-state oscillation and increases the sensitivity of the closed-loop system to the measured variable. The representation in Fig. 1.1a is applicable to a wide range of sensors, including continuous-wave (CW) quadrupole resonance (QR) spectrometers.

In order to describe the operation of the marginal oscillator in further detail, we restrict attention to Fig. 1.1b on the next page, which represents a marginal oscillator that detects QR transitions in solids. In this application, the parallel RLC network represents the sensor, and the inductor contains

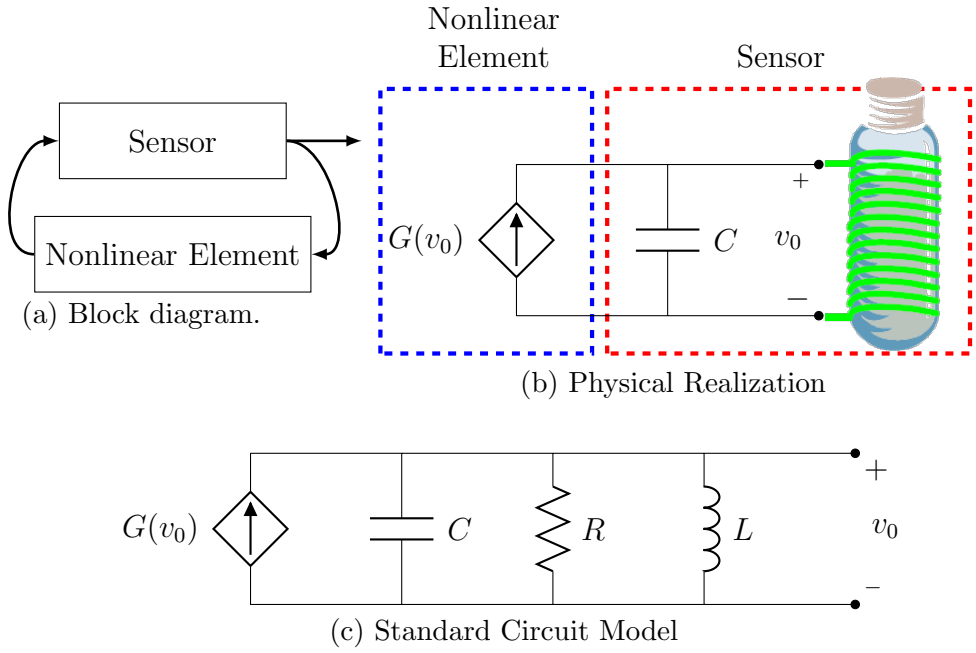


Figure 1.1: Conceptual representation of a marginal oscillator. (a) Block diagram. (b) Marginal oscillator for detecting QR transitions. (c) The standard circuit model.

a sample material possessing a QR transition. In response to a sinusoidal input current $i(t)$, the sensor generates an output voltage $v(t)$. The inductor generates an oscillating magnetic field that interacts with nuclei within the sample. When the frequency of the magnetic field approaches a QR transition frequency, the nuclei adsorb energy and as a result, the losses within the tuned RLC network increase. The increase in losses represents an effective decrease in the value of resistance R, and reduces the amplitude of the sinusoidal output voltage $v(t)$.

To sustain a steady-state oscillation near the natural frequency of the RLC network, the dependent current source $G(v)$ must appear as a negative resistance of value $-R$, so that the net energy dissipation in the marginal oscillator is zero. In theory, one could implement the dependent current source as a linear function $i = G(v) = g_A v$ of voltage, where the conductance g_A is chosen as $1/R$. In practice, it is not possible to realize a dependent current source whose conductance g_A exactly matches $1/R$. If g_A is either smaller or larger than $1/R$, then the oscillation amplitude exponentially grows towards infinity or decays to zero, respectively. We overcome this limitation and sustain a steady-state oscillation by implementing the dependent current

source $G(v)$ as a nonlinear function of v. The first marginal oscillator design by Pound and Knight uses the nonlinear characteristics of a vacuum tube [8]. As the input voltage to the vacuum tube increases, its gain, and hence the feedback current $i(t)$ to the RLC circuit, decreases.

A key figure of merit for a marginal oscillator is the *sensitivity* of the amplitude of oscillation with respect to the losses in the linear system. As we are expecting the measurement to cause small changes in the resistance R, the sensitivity is written as

$$S_R^A = \frac{\%\text{ change in A}}{\%\text{ change in R}} = \frac{\Delta A/A \times 100}{\Delta R/R \times 100} = \frac{R}{A}\frac{\Delta A}{\Delta R} \tag{1.1}$$

where the amplitude A and the resistance R represent the nominal value of oscillation and losses, respectively, while ΔA represents the change in amplitude due to a change ΔR in losses. In the limit as ΔR approaches zero,

$$S_R^A = \frac{R}{A}\frac{\partial A}{\partial R} \tag{1.2}$$

The sensitivity definition appearing in Eq. (1.2) is consistent with signal processing and control systems usage. In contrast, *instrument sensitivity* defines the smallest value of a variable the instrument can measure in the presence of noise. In order to avoid confusion between these two terms, this document uses the convention of Viswanathan et al. [9], and refers to the sensitivity defined in Eq. (1.2) as the conversion gain

$$G_C = S_R^A \tag{1.3}$$

To understand why a large conversion gain is desirable, consider the conversion gain of a Q-meter, where the dependent current source in Fig. 1.1b on page 10 is replaced by an independent sinusoidal current source

$$i(t) = I_0 \cos(\omega_n t) \tag{1.4}$$

with constant amplitude I_o. The frequency of the dependent source matches the natural frequency of the RLC circuit. As the RLC circuit appears as a pure resistance R at the natural frequency, the output voltage is

$$v(t) = RI_0 \cos(\omega_n t) = A\cos(\omega_n t) \tag{1.5}$$

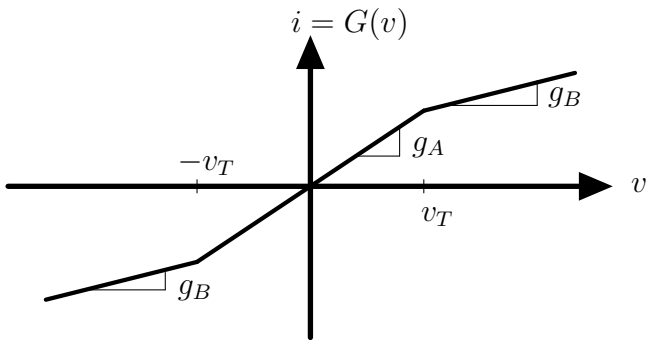


Figure 1.2: Input/output characteristics of the nonlinear dependent current source

It follows that the conversion gain of a Q-meter is unity:

$$G_C = \frac{R}{A}\frac{\partial A}{\partial R} = \frac{R}{RI_0}\frac{\partial (RI_0)}{\partial R} = 1 \tag{1.6}$$

In the magnetic resonance experiments performed by Pound and Knight [8], the normalized change in losses, $\Delta R/R$, is on the order of 10^{-6}. As the nominal amplitude A of oscillation is about 1 V and the conversion gain is unity, the change in A is about 1 µV. In order to improve the signal-to-noise ratio (SNR) when measuring a change in amplitude ΔA, a conversion gain greater than unity is desirable.

Pound's original motivation for replacing the Q-meter with a marginal oscillator was to eliminate the need for an external oscillator. In the Q-meter configuration, separate variable capacitors set the natural frequency of the RLC circuit and the frequency of the independent current source. Using this configuration, it is difficult to precisely match the oscillator frequency to the natural frequency of the RLC circuit while sweeping the frequency of both through a NMR transition. The marginal oscillator in Fig. 1.1b on page 10 eliminates this problem by using a single variable capacitor to adjust the oscillation frequency.

Serendipitously, Pound noted that the conversion gain of the marginal oscillator is significantly larger than that of the Q-meter. Pound did not explain why the marginal oscillator increases the conversion gain, nor did he determine the conversion gain as a function of the nonlinear transfer characteristic of the dependent current source.

It was not until the mid-1970s that Viswanathan et al. showed that the shape of the nonlinear feedback function determines the conversion gain of the marginal oscillator [9]. Viswanathan et al. represent $G(v)$ using the piecewise-linear curve in Fig. 1.2 on page 12. It is sufficiently close to what is observed in FETs and vacuum tubes, that it is considered useful.

To derive an expression for the conversion gain, Viswanathan et al. first represent the marginal oscillator in Fig. 1.1b on page 10 using the nonlinear differential equation

$$\frac{d^2v}{dt^2} + \frac{1}{C}\left(\frac{1}{R} - g(v)\right)\frac{dv}{dt} + \frac{1}{LC}v = 0 \tag{1.7}$$

where $-g(v) = -dG(v)/dv$ represents the conductance looking into the dependent current source. Using an approximation to the solution of a nonlinear second-order system developed by Krylov and Bogoliubov [24], Viswanathan et al. show that the conversion gain approximately satisfies

$$G_C = \frac{\pi}{2\left(1 - \frac{g_B}{g_A}\right) g_A R \sin(2\theta)} \tag{1.8}$$

where θ is the solution to

$$\sin 2\theta + 2\theta = \frac{\pi}{\left(1 - \frac{g_B}{g_A}\right)}\left[\frac{1}{g_A R} - \frac{g_B}{g_A}\right], \text{ where } \theta \in \left(0, \frac{\pi}{2}\right] \tag{1.9}$$

Furthermore, Viswanathan et al. show that the oscillation amplitude satisfies

$$A = v_T / \sin\theta \tag{1.10}$$

Equation (1.8) implies that one can achieve an arbitrarily large conversion gain by increasing the value of g_B towards g_A. Increasing the conversion gain does not come without a penalty. Viswanathan et al. approximate the transient response in oscillation amplitude A for small changes in losses R as a first-order linear differential equation, and show the time constant of the response is

$$\tau = \frac{2R}{L\omega_n^2} G_C \tag{1.11}$$

As measurement bandwidth is the reciprocal of the time constant, Eq. (1.10) reveals that the price for increasing the conversion gain is a proportionate re-

duction in measurement bandwidth. In other words, increasing the conversion gain comes at the expense of increasing the measurement time.

The work of Viswanathan et al. leaves several questions unanswered and still more unasked. This dissertation focuses specifically on the following questions. For the nonlinear function $G(v)$ appearing in Fig. 2.3 on page 24, and Eq. (1.10) on page 13 showing that the time constant associated with changes in oscillation amplitude is directly proportional to the conversion gain, does this linear relationship hold for all nonlinear feedback? Does this coupling matter? It will be shown that these two quantities are strongly coupled, but there is a straightforward work-around that will be developed in detail.

Recent work at Penn State addresses two other items [25]. First, using the describing function method [26], the results from Penn State show how to determine the conversion gain for a nonlinear system of arbitrary order. As for the method of Krylov and Bogoliubov, the describing function approach also requires approximations. Using the work of Mees and Bergen [27], the group at Penn State provides upper and lower bounds on the estimates of conversion gain, amplitude of oscillation, and measurement bandwidth.

1.5.4 Project Contributions

This dissertation presents many new results. The following is a partial list.

1. New analysis allows a much broader class of functions for implementing nonlinear feedback while still yielding conversion gains that are as large as those achievable with Pound's marginal oscillator. We will see that this objective is achieved with much less delicate laboratory calibration. The recursive algorithm given may, or may not, be more challenging to realize in practice.

2. Using full-state feedback allows all analysis to use closed-form algebra. Using this enhanced understanding, we can establish the exact functional dependence between conversion gain and the settling time of the oscillator. This understanding is in excellent agreement with previous methods based on transcendental approximations.

3. The marginal oscillator is normally tuned for high conversion gain.

This appears to increase the settling time. The observed correlation between settling time and conversion gain is now proved as a theorem. However, the following work-around is also developed. Each time the circuit settles to a limit cycle, even at a low conversion gain, we can accurately predict the exact internal resistance of the marginal oscillator to about five significant digits. Using previous approximate methods, such predictions were unachievable.

4. Using a second voltage source, we show that each data point yields an estimate of the energy of the oscillator and, hence, its amplitude. Previous methods have estimated the envelope of a modulated sinusoid, and required either high-speed sampling, or analysis methods that generate repeatable phase noise. As we do not require box-car-integration of a repeating signal, the number of data samples required to arrive at an estimate of the signal amplitude is greatly decreased.

5. For the variable capacitor used in actual lab experiments, we show (by several distinct methods) that any parasitic resistances can easily be accounted for by a suitable change in the circuit's resistive element. This is because the circuit is always used very close to resonance.

CHAPTER 2

MARGINAL OSCILLATOR SENSITIVITY ENHANCEMENT USING FULL-STATE NONLINEAR FEEDBACK

2.1 Summary

In many sensor applications of industrial and military interest, one requires an electronic sensor or transducer that translates small environmental changes into corresponding changes in the electrical properties of a resistor, inductor or capacitor. This chapter briefly summarizes previous applications of the marginal oscillator, then develops a new strategy and approach to sensing tiny changes in the resistance of a parallel *tank circuit*.

Since the late 1970s those using marginal oscillators have relied on output feedback to set the sensitivity (in the form of the conversion gain). The most common implementation takes advantage of the intrinsic nonlinearity of vacuum tubes, or FETs, to implement a nonlinear feedback loop. In all known circuits using this implementation, increasing the conversion gain also increases the amount of time required to detect the perturbation of interest. The analysis here shows that implementing the marginal oscillator using full-state feedback allows updates in the estimate of resistance that are fast compared to the relaxation time of the circuit. These new ideas also lead to a more realistic electrical model of the marginal oscillator circuit, a new schematic for full-state feedback, and a new strategy for control that leads to much simpler algebraic analysis of the expected performance. This chapter shows that there is no need to rely upon the averaging theory employed by Krylov and Bogoliubov [24]. That style of averaging theory results in transcendental approximations that are subsequently difficult to analyze. In contrast, full-state feedback yields closed-form expressions for settling time and conversion gain.

This chapter summarizes the mathematical strategies and concepts required to successfully implement full-state feedback. The detailed design arising

from this overall strategy will be developed in the next chapter.

2.2 Measuring Small Changes in Resistance Using an Oscillator

A marginal oscillator is used to measure small changes in the losses of a resonant circuit. Pioneered by Pound [8], the marginal oscillator uses a negative resistance converter to drive a lossy resonant circuit. In the case of a parallel resistor-capacitor-inductor (RLC) or *tank* circuit, where R models the losses in the circuit, powering the circuit to maintain steady-state oscillation requires that the power source appear as a negative resistance with the value $-R$. When the power source is applied, the remaining inductor-capacitor (LC) circuit behaves in an ideal manner maintaining steady-state and nearly sinusoidal oscillations indefinitely. Section 2.3.1 on page 23 will show that linear feedback only yields a fixed conversion gain (namely 1). Hence, the historical interest in nonlinear marginal oscillators.

Only under ideal circumstances can we match the lossy RLC circuit's exact value for R. Therefore, Pound also introduced the idea that below a fixed voltage v_T the marginal oscillator appears as a negative resistor of slightly greater magnitude than R in the circuit, thus inducing oscillations of steadily increasing amplitude. Above the threshold v_T, some nonlinear element of the circuit (in Pound's case, a vacuum tube) decreased the negative resistance below the value of R, thus keeping the oscillation amplitude from increasing without bound.

In Pound's work on NMR, the sample of interest is placed within the turns of the inductor coil. At a frequency where nuclei in the sample begin to absorb energy from the oscillating magnetic field in the coil, the resistance parameter of the parallel RLC circuit appears to decrease very slightly. This, in turn, shrinks the amplitude of oscillation A.

2.2.1 Previous Work

Following the example of Pound, marginal oscillators are usually implemented with nonlinear output feedback. The first fully successful analysis of how this worked is the two-slope limiter described in Viswanathan et al. [9].

However, taking a fresh look at the underlying assumptions shows that full-state feedback is a more natural way of handling the marginal oscillator and shows that multiple design goals can be supported with a two-state design. Although originally used for NMR and NQR applications, the new design may help solve the general problem of obtaining increased sensitivity from this electronic sensor. For the moment, this exposition will pursue the analysis from an NMR/NQR perspective.

A continuous wave oscillator is used to find NMR and NQR frequencies in materials where the time constants are not advantageous for pulsed spectrometers. An initial analysis of NQR for energetic materials used by the Army indicates a fairly even split between those best detected with pulsed versus CW NQR.

As analyzed in previous literature, the circuit model of interest is given in Fig. 2.1. A sample of material is placed in the inductor coil. When the sample goes into resonance, the resistance R changes slightly to $R - \delta R$. In other words, the sample begins to draw energy from the circuit. By using

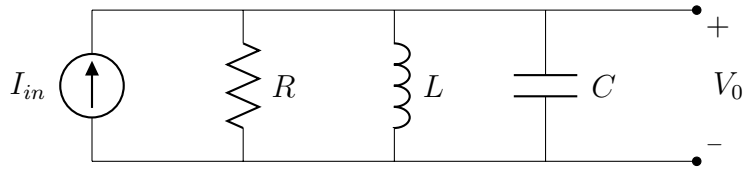


Figure 2.1: Open-loop CW-NQR circuit, with a sample of resonant material present $R \to R - \delta R$

Kirchhoff's current law (KCL) on the top rail of this diagram it is easy to write down the equations of interest in the Laplace domain.

$$I_{in}(s) = \frac{V_0}{R} + \frac{V_0}{sL} + \frac{V_0}{(1/sC)} \tag{2.1}$$

Rearranging terms readily gives the standard form

$$s^2 V(s) + \frac{sV(s)}{RC} + \frac{V_0(s)}{LC} = \frac{sI(s)}{C} \tag{2.2}$$

Before attempting to close the loop with nonlinear feedback, this expression has to be brought back into the time domain.

$$\ddot{v_0}(t) + \frac{\dot{v_0}(t)}{RC} + \frac{v_0(t)}{LC} = \frac{1}{C}\frac{d}{dt} i_{in}(t) \tag{2.3}$$

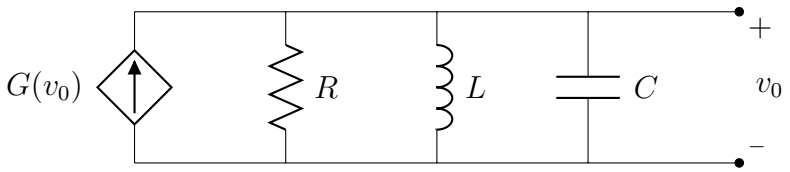


Figure 2.2: Closed-loop CW-NQR circuit, with a sample of resonant material present $R \to R - \delta R$

A cursory reading of terms in Eq. (2.3) on page 18 indicates a single input $i_{in}(t)$ and a single output $v_0(t)$. This is for historical reasons. Previous hardware implementations used vacuum tubes and FETs, each documented as yielding good conversion gain, without giving any detailed analysis.

One of the few accurate attempts at analysis is contained in a paper by Viswanathan et al. [9]. In their analysis, the first and third terms on the left-hand side of Eq. (2.3) on page 18 govern the oscillation and the circuit losses are governed by the second term. Therefore we replace $i_{in}(t)$ by $G(v_0)$, and further define $g(v_0) \equiv dG(v_0)/dv$, so the resulting equation has $g(v_0)$ appearing exclusively in the middle term.

$$\ddot{v}_0 + \frac{1}{C}\left[\frac{1}{R} - g(v_0)\right]\dot{v}_0 + \frac{1}{LC}v_0 = 0 \tag{2.4}$$

For more extensive coverage of analysis using this strategy of *output feedback*, see theses by Zhang [28], and Tyson [25]. By implementing output feedback, we claim that the resulting design problem of choosing $g(v_0)$ results in many forms of seemingly anomalous behavior of the closed-loop circuit. The most cogent way to explain the unfortunate properties of this nonlinear output feedback strategy is in the context of phase-space analysis as given in Section 2.3 on page 22.

2.2.2 Conversion Gain

In previous work, it has been held that there is a preferred figure of merit for any feedback strategy. It was held that it is desirable for the amplitude of oscillation to change as much as possible in response to a small change in the resistance R, i.e. $R \to R - \delta R$. In electrical engineering this would normally be called the *sensitivity*. However, this would be confusing to physicists where sensitivity is tied up with notions of finding signals in a noise floor. Therefore, the figure of merit is re-named *conversion gain*. For the moment,

the definition is given using A as amplitude and R as the resistor in our circuit, but the definition is independent of these choices.

The new analysis given in this and subsequent chapters will keep the conversion gain as an applicable figure of merit.

Definition 2.2.1 (Conversion Gain)**.**

$$\text{Conversion Gain} = S_R^A = \frac{\text{fractional change in A}}{\text{fractional change in R}} = \frac{\frac{\Delta A}{A}}{\frac{\Delta R}{R}} \tag{2.5}$$

In the limit as $\Delta R \to 0$ this expression becomes

$$S_R^A = \frac{R}{A}\frac{\partial A}{\partial R} \tag{2.6}$$

This definition can be seen as forcing the conversion gain to become dimensionless so that one cannot play games with units of measurement, such as setting either A or R to unnaturally large or small values. (It is important to note that, in a typical magnetic resonance experiment, $\Delta R/R \approx 10^{-6}$. Hence the historical emphasis on pushing the conversion gain to high values.)

The following theorem shows explicitly that conversion gain is independent of the units used for measurement, and will also be used many times in this dissertation when changing coordinate systems. As long as measurements in the two coordinate systems differ by a constant multiplicative factor, the conversion gain will be invariant.

Theorem 2.2.1 (Conversion Gain Invariance Under Transformation of Units)**.** *For real variables y and x, and constant m satisfying $y = mx$, the conversion gain of y with respect to x is 1.*

Proof. The following equation is given.

$$y = mx$$

Now apply the definition of conversion gain.

$$\begin{aligned} S_x^y &= \frac{x}{y}\frac{\partial y}{\partial x} \\ &= \frac{x}{mx}m \\ &= 1 \end{aligned}$$

□

Here is a more general theorem for later use.

Theorem 2.2.2 (Conversion Gain Relation Via Power Law)**.** *For real variables x and y, and real constants m and n satisfying $y = mx^n$, the conversion gain of y with respect to x is n.*

Proof. The following equation is given.

$$y = mx^n$$

Now apply the definition of conversion gain.

$$\begin{aligned} S_x^y &= \frac{x}{y}\frac{\partial y}{\partial x} \\ &= \frac{x}{mx^n} nmx^{n-1} \\ &= n \end{aligned}$$

□

Also observe that the conversion gain obeys a chain rule based on the properties of partial differentiation.

Theorem 2.2.3 (Chain Rule for Conversion Gain)**.** *Let $A = f(r)$ and $r = g(\gamma)$. If the applicable derivatives exist, then*

$$S_\gamma^A = S_r^A S_\gamma^r$$

Proof. Using the definition of conversion gain, simply substitute everywhere, commute terms, and cancel.

$$\begin{aligned} \frac{\gamma}{A}\frac{\partial A}{\partial \gamma} &= \left(\frac{r}{A}\frac{\partial A}{\partial r}\right)\left(\frac{\gamma}{r}\frac{\partial r}{\partial \gamma}\right) \\ \frac{\gamma}{A}\frac{\partial A}{\partial \gamma} &= \left(\frac{r}{A}\frac{\gamma}{r}\right)\left(\frac{\partial A}{\partial r}\frac{\partial r}{\partial \gamma}\right) \\ \frac{\partial A}{\partial \gamma} &= \frac{\partial A}{\partial r}\frac{\partial r}{\partial \gamma} \end{aligned}$$

The last line holds using the chain rule for partial derivatives. □

2.3 New Approach, Phase-Space Analysis

The overall strategy to update the analysis of this circuit is to proceed in two steps. First, realize that there is not just one output variable available for feedback but two. Taking advantage of all state variables available is termed *full-state feedback*. This observation is more straightforward if one realizes that equations describing Fig. 2.1 on page 18 must have two state variables, namely, the inductor current and the capacitor voltage. Once this analysis is performed, it will motivate a second step. The circuit Fig. 2.1 on page 18 has a very natural place to add an additional voltage source and, correspondingly, adds a dependent voltage source to Fig. 2.2 on page 19. This means the circuit will go from being single-input, single-output to being multi-input, multi-output.

Starting again with the KCL equation given earlier and referring to Fig. 2.1 on page 18:

$$I_{in}(s) = \frac{V_0}{R} + \frac{V_0}{sL} + \frac{V_0}{(1/sC)} \tag{2.7}$$

Only this time, we now recognize the second term on the right-hand side of Eq. (2.7) as the second state variable.

$$I_L = V_0/sL \tag{2.8}$$

So the Eq. (2.7) becomes

$$I_{in}(s) = V_0/R + I_L + sV_0C \tag{2.9}$$

With minor rearrangement of the terms the state-space ODE begins to emerge.

$$\begin{aligned} sV_0 =& \frac{-V_0}{CR} - \frac{I_L}{C} + \frac{I_{in}(s)}{C} \\ sI_L =& \frac{V_0}{L} \end{aligned} \tag{2.10}$$

These expressions are now readily put into state-space notation

$$\begin{pmatrix} sV_0 \\ sI_L \end{pmatrix} = \begin{pmatrix} -\frac{1}{CR} & -\frac{1}{C} \\ \frac{1}{L} & 0 \end{pmatrix} \begin{pmatrix} V_0 \\ I_L \end{pmatrix} + \begin{pmatrix} \frac{1}{C} \\ 0 \end{pmatrix} I_{in}(s) \tag{2.11}$$

Notice that this is now in standard state-space notation $\dot{x} = \boldsymbol{A}x + \boldsymbol{B}u$. By luck, the output V_0 is identical to the capacitor voltage. Therefore, we have

fulfilled the strategy of using the inductor current and capacitor voltage as our two states.

2.3.1 Linear Feedback

Consider the transfer function $X(s)/U(s) = (s\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{B}$:

$$\begin{pmatrix} V_0/I_{in} \\ I_L/I_{in} \end{pmatrix} = \begin{pmatrix} \frac{s/C}{s^2+s/RC+1/LC} \\ \frac{1/LC}{s^2+s/RC+1/LC} \end{pmatrix} \tag{2.12}$$

With the transfer functions now calculated, the system's conversion gain is easily calculated. Weirdly, the same result for the conversion gain emerges for both of the following equations.

$$\frac{R}{V_0}\frac{\partial V_0}{\partial R} = \frac{s/CR}{s^2 + s/CR + 1/LC} \tag{2.13}$$

$$\frac{R}{I_L}\frac{\partial I_L}{\partial R} = \frac{s/CR}{s^2 + s/CR + 1/LC} \tag{2.14}$$

At resonance, both expressions trivially simplify to 1. In previous analysis, this result has been used to show that the open-loop system is not a viable means of detecting tiny changes in R.

It is now straightforward to notice why the situation does not improve using linear feedback. By referring to Eq. (2.11) on page 22 it is straightforward to observe that the most that can be hoped for is to effectively drive the upper left-hand entry in the state matrix $\boldsymbol{A}$ to zero. Hence, any slight decrease in R rapidly kills the oscillation altogether, and any slight decrease will increase the amplitude of oscillation until it grows beyond the power supply rail. Hence linear feedback is not a viable strategy when implemented in an actual device.

2.4 Phase-Plane Analysis

Having now stated the problem in a two-dimensional phase space, we can compare and contrast work done previously to what is now possible.

Remark (The use of cartoons). The reasoning required to understand the value-added of using two state variables for feedback can be motivated by

a sequence of phase-plane portraits, pictures and explanations given in this section. It is important to note that if these diagrams were given as precise, scaled, mathematical pictures, the salient features would be imperceptible. Therefore, the features of these diagrams have to be greatly exaggerated. When this is done, the caption will list the figure as being a *cartoon*.

At this point it is easier to understand how feedback in the situation outlined in Eq. (2.4) on page 19 has been handled. The Schiano Group at The Pennsylvania State University [25,28] have been deeply interested in a method originally proposed by Viswanathan et al. [9]. For the feedback function $G(v_0)$, and hence $g(v_0)$, they have used a *two-slope limiter* as depicted in Fig. 2.3 and Fig. 2.4.

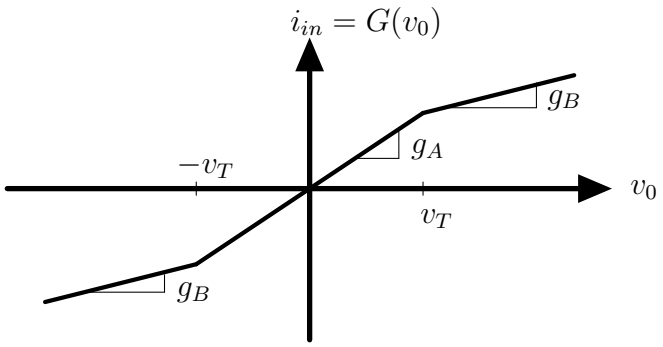


Figure 2.3: Cartoon "two-slope limiter" feedback $G(v_0)$ of Viswanathan et al.

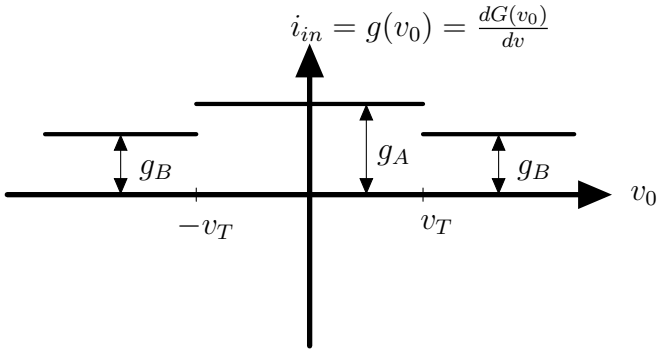


Figure 2.4: Cartoon "two-slope limiter" feedback $g(v_0)$ of Viswanathan et al.

Let us take a more detailed look at how this feedback scheme works. The main idea is that the slopes g_A and g_B are used to make a *negative resistor*

where $g_A > 1/R > g_B$. Amazingly, Viswanathan et al. predict that g_A and g_B should be chosen to *almost* match $1/R$ without being allowed to become equal to $1/R$. As g_A and g_B are allowed to approach $1/R$ the conversion gain can be made larger without theoretical limit. To get an intuitive feel for how this works, start with Eq. (2.11) on page 22. Using nonlinear feedback, it is required to start in the time domain as follows:

$$\begin{aligned} \dot{x} &= \boldsymbol{A}x + \boldsymbol{B}u \\ \begin{pmatrix} dv_0/dt \\ di_L/dt \end{pmatrix} &= \begin{pmatrix} -\frac{1}{CR} & -\frac{1}{C} \\ \frac{1}{L} & 0 \end{pmatrix} \begin{pmatrix} v_0 \\ i_L \end{pmatrix} + \begin{pmatrix} \frac{1}{C} \\ 0 \end{pmatrix} i_{in}(t) \end{aligned} \tag{2.15}$$

What becomes obvious with a small amount of exploratory work is that the off-diagonal terms of $\boldsymbol{A}$ on the right-hand side of this equation are responsible for oscillation and that the upper left-hand term of the state matrix $\boldsymbol{A}$ causes the equilibrium point at the origin to become the center of a stable spiral in the absence of feedback. Viswanathan et al. use a two-slope limiter so that when $I_{in}(t)$ is chosen to be $G(v_0)$ as in Fig. 2.3 on page 24, the upper left-hand term in the closed-loop state matrix can be driven to either a slightly positive or slightly negative value. When the value is slightly positive, the origin of the phase portrait becomes the center of an unstable spiral. When negative, the phase portrait is a stable spiral about the origin.

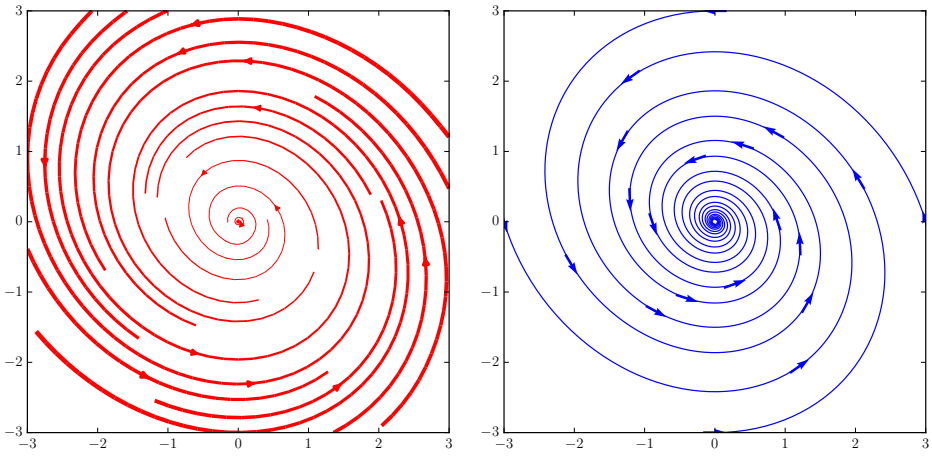
For those unfamiliar with state-space graphs, a stable spiral is depicted in Fig. 2.5 on the next page. Notice that all trajectories swirl toward the origin, hence the terminology *stable spiral*. An unstable spiral is depicted in Fig. 2.6 on page 27, where all trajectories move away from the origin.

The two-slope limiter of Viswanathan et al. results in a situation depicted in Fig. 2.7 on page 28 where any trajectory circulating near the origin will be an unstable spiral and be an expanding cycle. Far away from the origin, most of the trajectory exists in the part of the phase plane occupied by stable spirals and will tend to shrink. Hence, there exists an equilibrium where these two tendencies balance and will become a stable limit cycle.

The situation described by Fig. 2.7 on page 28 is graphically represented in the cartoon of Fig. 2.8 on page 29. For this cartoon the transition points are set to plus or minus one volt, $v_T = 1$. To make this cartoon, an abnormally large g_A and abnormally small g_B are used so that the human eye can interpret the area near the center $|v_0| < 1$ as an unstable spiral, and the outer area $|v_0| > 1$

as a stable spiral. Very near $v_0 = 1$ there is a limit cycle that has a rather complicated shape. Using averaging methods of Krylov and Bogoliubov [24] results in mathematically complicated transcendental approximations. In the real world where g_A and g_B get very close to $1/R$, the limit cycle is closer to being circular (or elliptical) but analysis of the distortions still require averaging theory that gives rise to transcendental approximations.

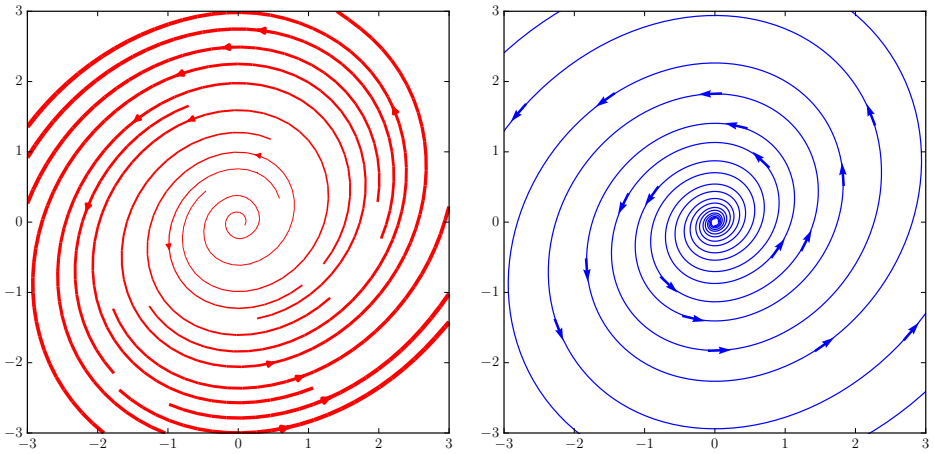
In the real world, NMR/NQR instruments built using this method are tuned in the following way. First, g_A is dialed in so that the spiral around the origin expands as slowly as possible. Then g_B is dialed in to hold this expansion in check. Invariably (independent of the choice of technology used for implementation: vacuum tubes, FETs, etc.), this procedure challenges the limit of resolution with which g_A and g_B can be tuned.



(a) Streamline plot (b) Integral curves

Figure 2.5: An example of a stable spiral $\dot{x} = \boldsymbol{A}x$, where $\boldsymbol{A} = \begin{pmatrix} -0.3 & -1 \\ 1 & 0 \end{pmatrix}$. Diagram a) is a streamline plot. Diagram b) shows more complete integral curves.

Remark (Integral curves versus streamline plots). When depicting the trajectories of solutions to an ODE in phase space, there are two useful graphical methods. The most common way, used in most mathematical texts, is the use of integral curves. One selects a set of integral curves that reasonably depict the important features of these trajectories. A partial list includes: critical points, limit cycles, specific lines or curves in the plane that a trajectory cannot cross, etc. There can be subtle questions regarding where solutions exist in a global or local sense. Selecting trajectories that depict all salient



(a) Streamline plot (b) Integral curves

Figure 2.6: An example of an unstable spiral $\dot{x} = \boldsymbol{A}x$, where $\boldsymbol{A} = \begin{pmatrix} 0.3 & -1 \\ 1 & 0 \end{pmatrix}$. Diagram a) is a streamline plot. Diagram b) shows more complete integral curves.

features can be a time consuming and exacting exercise. Solutions can be dense near a limit cycle or point of attraction. The visual density can vary a lot within the same graph.

A second method is to use a *streamline plot*. Here a computer algorithm determines how to depict trajectories in the phase space so that the visual density across the plot is more uniform. However, to achieve this effect, the computer is likely to begin trajectories where the visual density is too low, and to stop trajectories where the visual density is becoming too high. Where these trajectories start or stop has nothing to do with their existence globally, nor is it tied to the time-limits of integration.

An excellent example of this is Fig. 2.8 on page 29 where both a streamline plot and a plot of integral curves are given. The visual density of the streamline plot is more uniform and the eye can pick up the overall pattern pretty easily. The plot of integral curves is more forceful in showing the limit cycle. It also gives a better indication that all trajectories are moving toward the limit cycle. However, the area near the limit cycle is rather dense and difficult to see clearly. The structure of trajectories away from the limit cycle is also more difficult to see. It is tough to discern the areas where graph depicts an unstable versus a stable spiral.

For the purposes of this dissertation, the ODEs used are simple enough

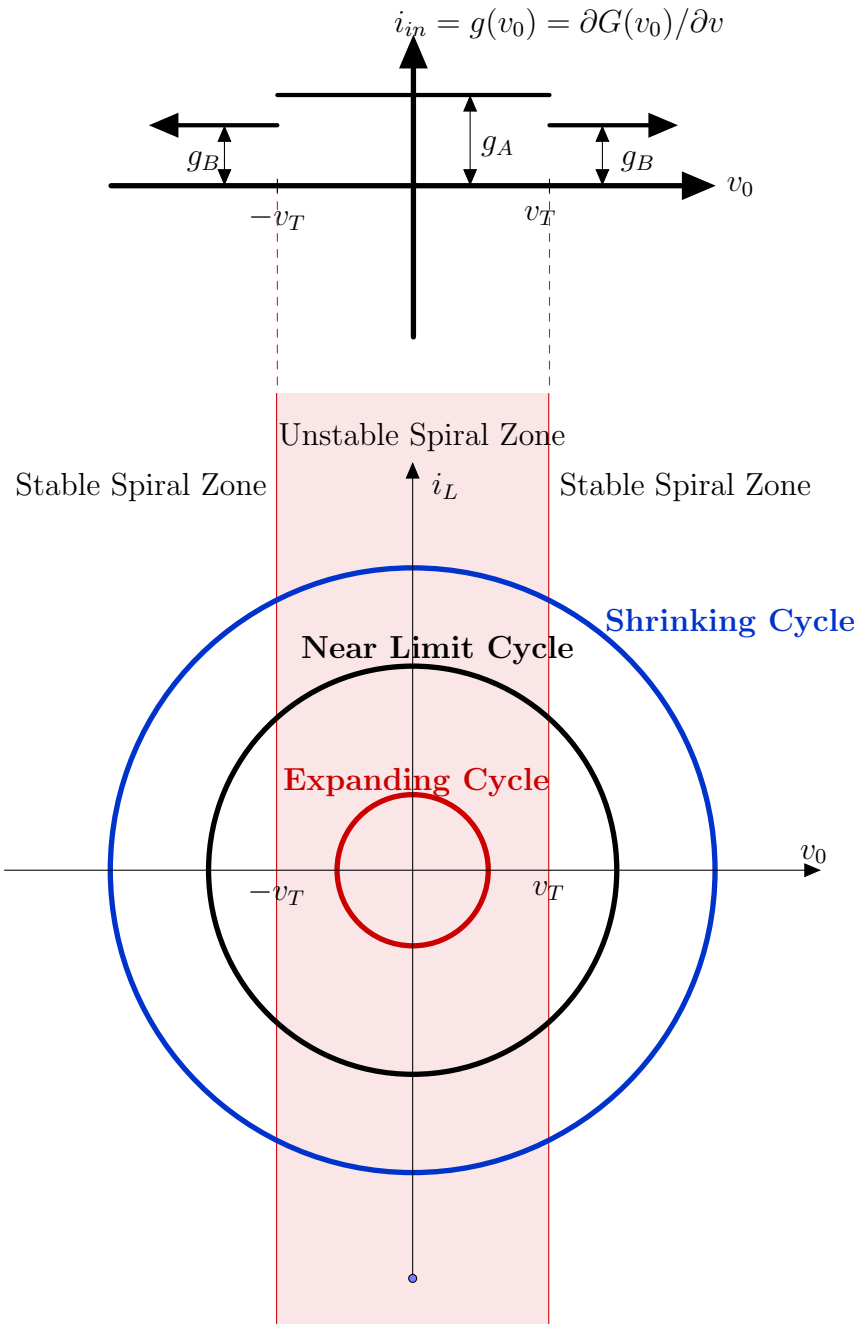
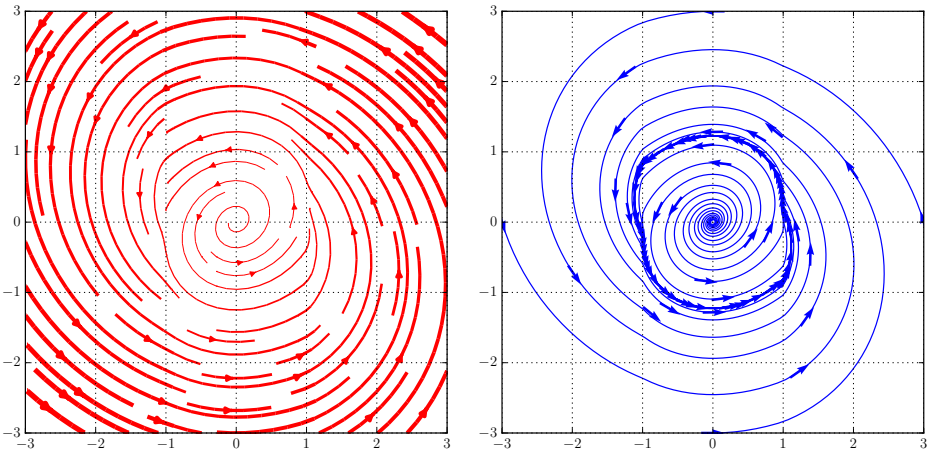


Figure 2.7: Cartoon of a gedanken experiment describing the effect of a two-slope limiter as viewed in the two-dimensional phase space



(a) Streamline plot (b) Integral curves

Figure 2.8: A cartoon of the phase space using the two-slope limiter of Viswanathan et al., with $v_T = 1$. In the actual phase space, the trajectories are nearly circular. Near the origin, the spiral moves outward. Away from the origin the spiral moves inward.

that complicated questions regarding existence and uniqueness never arise. So the choice between these two representations is a judgment call. To avoid confusion, every depiction of trajectories in the phase plane is explicitly labeled as either a streamline plot or set of integral curves.

2.5 Full-State Feedback

To motivate an approach requiring two states in feedback, notice that there are design goals that are simply impossible using schemes like the two-slope limiter of Viswanathan et al. Let us start with a simple example now and get more complicated later. In the two-slope limiter, as g_A gets exceptionally close to $1/R$, the electric circuit implementing the unstable spiral requires longer and longer times to spool up to the set-point where one can start an experiment. If there is an attempt to fix this by changing the characteristics of the two-slope limiter near the origin, the resulting circuit cannot discriminate between the start-up phase and the quiescent limit cycle where the NMR or NQR experiment can begin. The second way to see this is that any change in $g(v_0)$ will be exercised in every cycle of the oscillator. Hence the value becomes pinned to g_A near $v_0 = 0$ with no meaningfully different value available. This

is best seen by referring to the cartoon of Fig. 2.7 on page 28.

By contrast with two-variable feedback, it is possible to achieve the design goal of quick start-up. Here, refer to Fig. 2.9, where it is easy to separate different regions of the phase space as a function of radius. This allows a very natural way to implement a start-up region near the origin and still retain an exquisitely sensitive region near the limit cycle where measurements can be taken.

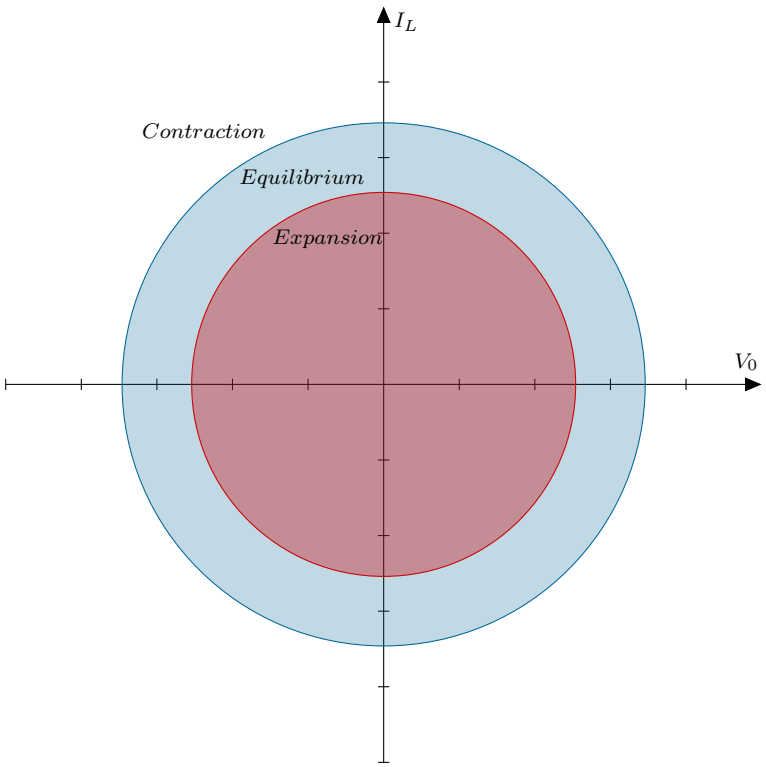


Figure 2.9: Cartoon summarizing a gedanken experiment using full-state feedback. Here there are 3 distinct zones: the start-up or expansion region to allow the instrument to get to its quiescent state quickly, the equilibrium zone where actual measurements are taken, and an outer zone for later use.

2.6 Solution Strategy

To get control of both state variables in the marginal oscillator circuit, a quick look at the state-space equation derived earlier and restated here motivates a

new design.

$$\begin{pmatrix} sV_0 \\ sI_L \end{pmatrix} = \begin{pmatrix} -\frac{1}{CR} & -\frac{1}{C} \\ \frac{1}{L} & 0 \end{pmatrix} \begin{pmatrix} V_0 \\ I_L \end{pmatrix} + \begin{pmatrix} \frac{1}{C} \\ 0 \end{pmatrix} I_{in}(t) \tag{2.16}$$

When applying feedback through $I_{in}(t)$ alone, it is easy to manipulate the terms describing dV_0/dt. To retain an axisymmetric approach, it is also required to add a circuit element that allows manipulation of the terms describing dI_L/dt.

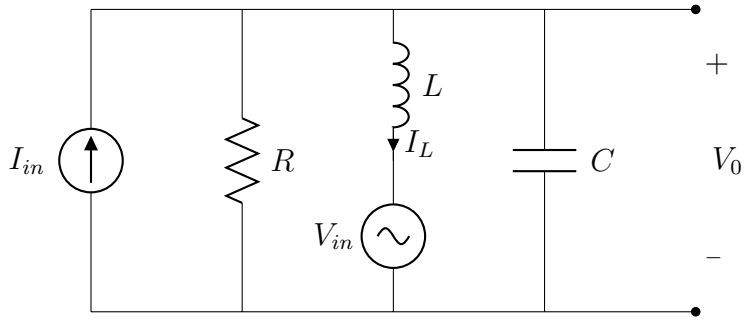


Figure 2.10: Two-state circuit open-loop

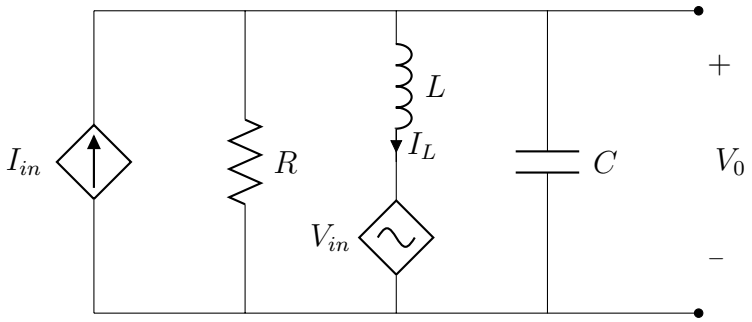


Figure 2.11: Two-state circuit closed-loop

Consider the circuit design given in Fig. 2.10. Again write KCL for the upper node.

$$I_{in}(s) = \frac{V_0}{R} + \frac{V_0 - V_{in}}{sL} + \frac{V_0}{(1/sC)} \tag{2.17}$$

This time, observe that the second term on the right-hand side is, again, I_L and analogously arrive at the equation

$$\begin{pmatrix} sV \\ sI_L \end{pmatrix} = \begin{pmatrix} -\frac{1}{CR} & -\frac{1}{C} \\ \frac{1}{L} & 0 \end{pmatrix} \begin{pmatrix} V_0 \\ I_L \end{pmatrix} + \begin{pmatrix} \frac{1}{C} \\ 0 \end{pmatrix} I_{in}(s) + \begin{pmatrix} 0 \\ -\frac{1}{L} \end{pmatrix} V_{in}(s) \tag{2.18}$$

And to pull everything into standard notation for a multi-input system.

$$\begin{pmatrix} sV \\ sI_L \end{pmatrix} = \begin{pmatrix} -\frac{1}{CR} & -\frac{1}{C} \\ \frac{1}{L} & 0 \end{pmatrix} \begin{pmatrix} V_0 \\ I_L \end{pmatrix} + \begin{pmatrix} \frac{1}{C} & 0 \\ 0 & -\frac{1}{L} \end{pmatrix} \begin{pmatrix} I_{in}(s) \\ V_{in}(s) \end{pmatrix} \tag{2.19}$$

By changing to a multi-input system, one gains a lot of flexibility in designing a control scheme that manipulates the state matrix terms of the closed-loop system.

By using full-state feedback with both state variables $v_0(t)$ and $i_L(t)$, the design is more powerful and flexible. No previous work recognized the possible advantage of this approach.

Because this is a compositional problem, this analysis will give rise to feedback laws that directly manipulate the system's amplitude. In the sequel, it is shown that the system amplitude can now be represented in cylindrical coordinates without loss of generality. An outline of how to compose new feedback laws in cylindrical coordinates will be given in Section 2.6.3 on page 35.

2.6.1 Actual Circuit Realization

A desirable side-effect of recasting the marginal oscillator in a full-state representation is that the circuit analysis now easily admits a more realistic representation of the circuit elements without increasing the complexity of the resulting analysis.

Alert readers will notice that, in the real world, the circuits of Fig. 2.10 and Fig. 2.11 on page 31 are not how the real NMR or NQR search coils work as the inductor L and the resistor R are actually the same element in a real device, as depicted in Fig. 2.12 on the following page, and would normally be analyzed as being in series. Using the full-state model this is now easily fixed, as is shown in Fig. 2.13 and Fig. 2.14 on page 34. The revised analysis presents no difficulty, and is given here.

Again, using KCL it is easy to arrive at the following equation.

$$\begin{pmatrix} sV_0 \\ sI_L \end{pmatrix} = \begin{pmatrix} 0 & -\frac{1}{C} \\ \frac{1}{L} & -\frac{R}{L} \end{pmatrix} \begin{pmatrix} V_0 \\ I_L \end{pmatrix} + \begin{pmatrix} \frac{1}{C} & 0 \\ 0 & -\frac{1}{L} \end{pmatrix} \begin{pmatrix} I_{in}(s) \\ V_{in}(s) \end{pmatrix} \tag{2.20}$$

Here we simply manipulate the feedback to effectively set the term in the

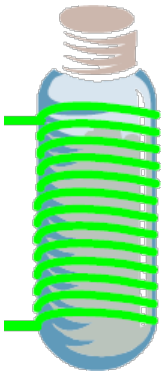


Figure 2.12: An inductor with a test sample placed within the coils

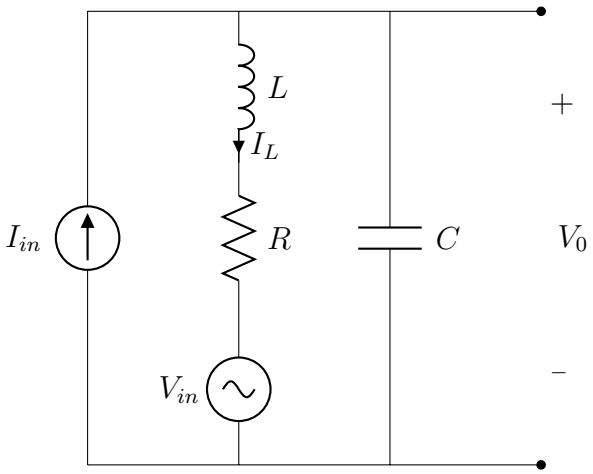


Figure 2.13: Open-loop NMR/NQR CW circuit as seen in a real world application

lower right corner of the state matrix $\boldsymbol{A}$ to zero and the strategy continues to work using analogous reasoning. The uninitiated reader will wonder why this circuit was not used earlier in the exposition.

Consider the closed-loop analysis when restricted to output feedback. The real-world circuit presents significant complications. One quick way to see this is that the input term using the voltage input V_i (i.e., the rightmost term) of Eq. (2.20) on page 32 is not available, and that controlling the input current I_{in} to cancel the lower right-hand term of the state matrix cannot be done directly. Instead the goal would be to generate a term in the upper left corner of the state matrix so that the trace of the state matrix is zero.

A second way to see why such an analysis is difficult without the state-space representation is as follows. If we use a series RL for the sample coil, this would result in the circuit depicted in Fig. 2.15 on the next page. The

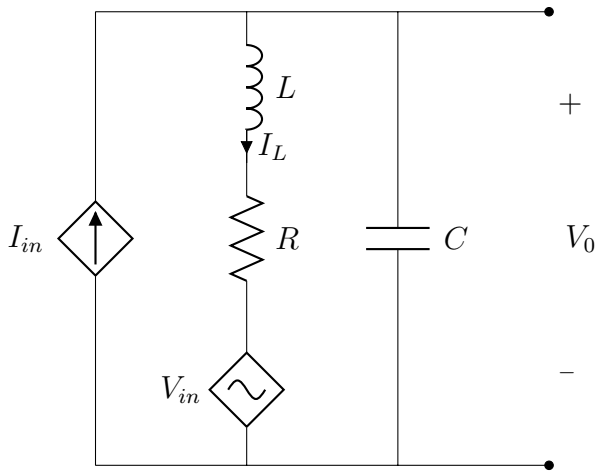


Figure 2.14: Closed-loop NMR/NQR CW circuit as seen in a real world application

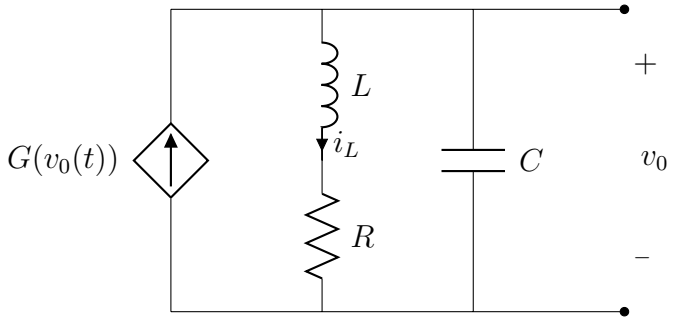


Figure 2.15: Higher fidelity closed-loop NMR/NQR CW circuit

differential equation suddenly gets more complicated.

$$\ddot{v}_0(t) + \frac{R}{L}\dot{v}_0(t) + \frac{1}{LC}v_0(t) = \frac{1}{C}\frac{d}{dt}G(v_0(t)) + \frac{R}{LC}G(v_0(t)) \tag{2.21}$$

Having the function $G(v_0(t))$ and its derivative appear on the right-hand side of Eq. (2.21) results in a more complicated analysis.

By contrast, the state-space representation given in Eq. (2.20) on page 32 shows a straightforward changeover.

2.6.2 Side Comment on Observability and Controllability

It is a very quick exercise to show that all state-space models in this dissertation are both observable and controllable. Therefore, there is no requirement to measure the current i_L directly. However, an extensive discussion regarding

a detailed derivation of an observer does not contribute to the central analysis. It is sufficient to note that obtaining information on the current i_L is not a difficult additional task.

2.6.3 Compositional Problem in Cylindrical Coordinates

Once we compose our system so that the limit cycle appears as a circle in phase space, it will be advantageous to write the system model in cylindrical coordinates. In the sequel, it will be shown that the model consists of two equations. One is for the quantity $\dot{r}$ governing the oscillation amplitude, and the second is $\dot{\theta} = \omega$ signifying that the system oscillates at frequency ω. Hence, we will have effectively reduced the state space from two dimensions to one dimension. Luckily, we will be able to formulate the feedback law in cylindrical coordinates and translate the result to rectangular coordinates. This section presents the mathematical machinery needed when this happens.

Let us put together some simple rules regarding how to dream up a desirable ordinary differential equation in a cylindrical coordinate system and then express it in a Cartesian coordinate system. Armed with this recipe, one can then freely discuss desirable qualities of a differential equation in cylindrical coordinates knowing full well that the translation to Cartesian coordinates, in this case the capacitor voltage and inductor current, can be achieved by straightforward algebraic rules.

2.6.4 Required Identities

All of these calculations are standard and well known, but will be needed in some detail in the subsequent analysis.

How is the following ODE correctly translated into cylindrical coordinates?

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \Phi(x,y) \\ \Psi(x,y) \end{pmatrix} \tag{2.22}$$

The standard coordinate system transformation relating Cartesian to cylindrical coordinates is

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \end{aligned} \tag{2.23}$$

Differentiating both sides with respect to time using the chain rule, one obtains

$$\begin{aligned} \dot{x} &= (\cos\theta)\dot{r} - r(\sin\theta)\dot{\theta} \\ \dot{y} &= (\sin\theta)\dot{r} + r(\cos\theta)\dot{\theta} \end{aligned} \tag{2.24}$$

So the derivation proceeds as follows:

$$\begin{aligned} \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} &= \begin{pmatrix} \Phi(x,y) \\ \Psi(x,y) \end{pmatrix} \\ \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} \begin{pmatrix} \dot{r} \\ \dot{\theta} \end{pmatrix} &= \begin{pmatrix} \Phi(r\cos\theta, r\sin\theta) \\ \Psi(r\cos\theta, r\sin\theta) \end{pmatrix} \\ \begin{pmatrix} \dot{r} \\ \dot{\theta} \end{pmatrix} &= \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}^{-1} \begin{pmatrix} \Phi(r,\theta) \\ \Psi(r,\theta) \end{pmatrix} \\ \begin{pmatrix} \dot{r} \\ \dot{\theta} \end{pmatrix} &= \frac{1}{r} \begin{pmatrix} r\cos\theta & r\sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \Phi(r,\theta) \\ \Psi(r,\theta) \end{pmatrix} \\ \begin{pmatrix} \dot{r} \\ \dot{\theta} \end{pmatrix} &= \begin{pmatrix} \cos\theta & \sin\theta \\ -\frac{1}{r}\sin\theta & \frac{1}{r}\cos\theta \end{pmatrix} \begin{pmatrix} \Phi(r,\theta) \\ \Psi(r,\theta) \end{pmatrix} \end{aligned} \tag{2.25}$$

Going back the other way from cylindrical to Cartesian coordinates is difficult because the equations

$$\begin{aligned} r &= \sqrt{x^2 + y^2} \\ \theta &= \arctan(y/x) \end{aligned} \tag{2.26}$$

quickly become very complicated using the same treatment. However the structure of the matrix

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\frac{1}{r}\sin\theta & \frac{1}{r}\cos\theta \end{pmatrix} \tag{2.27}$$

as seen in the last line of Eq. (2.25) gives a straightforward work-around given in the next section.

2.6.5 Compositional Rules

There is a sufficient recipe to get desired functions (in our case, a particular feedback law) to appear in the expression for $\dot{r}$ independently of $\dot{\theta}$ or vice versa. This allows us to always maintain separation of variables, and hence

leads to a much more tractable analysis.

1. Pick an arbitrary function of radius r, and call it $F(r)$. Consider the following ODE in Cartesian coordinates:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \frac{x}{r} F(r) \\ \frac{y}{r} F(r) \end{pmatrix} \tag{2.28}$$

   Here we are using r as a shorthand for $\sqrt{x^2 + y^2}$ to avoid becoming overly pedantic. The analysis proceeds precisely as in Eq. (2.25) on page 36 with the last line being

$$\begin{pmatrix} \dot{r} \\ \dot{\theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\frac{1}{r}\sin\theta & \frac{1}{r}\cos\theta \end{pmatrix} \begin{pmatrix} \cos(\theta) F(r) \\ \sin(\theta) F(r) \end{pmatrix} \tag{2.29}$$

   Carrying through the final multiplication we arrive at

$$\begin{pmatrix} \dot{r} \\ \dot{\theta} \end{pmatrix} = \begin{pmatrix} F(r) \\ 0 \end{pmatrix} \tag{2.30}$$

   Thus we have a recipe for making $F(r)$ appear exclusively in the expression for $\dot{r}$ in cylindrical coordinates.

2. Pick an arbitrary function of angle θ, and call it $F(\theta)$. Consider the following ODE in Cartesian coordinates:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} y F(\theta) \\ -x F(\theta) \end{pmatrix} \tag{2.31}$$

   Here we are using θ as a shorthand for $\arctan(y/x)$ to avoid becoming overly pedantic. The analysis proceeds precisely as in Eq. (2.25) on page 36 with the last line being

$$\begin{pmatrix} \dot{r} \\ \dot{\theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\frac{1}{r}\sin\theta & \frac{1}{r}\cos\theta \end{pmatrix} \begin{pmatrix} r\sin(\theta) F(\theta) \\ -r\cos(\theta) F(\theta) \end{pmatrix} \tag{2.32}$$

   Carrying through the final multiplication we arrive at

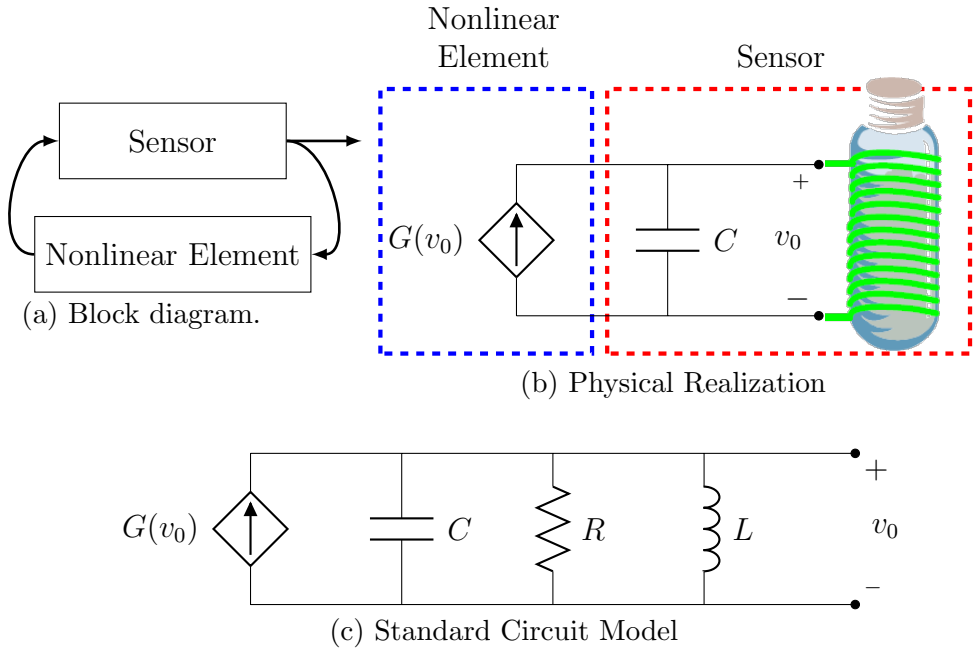
$$\begin{pmatrix} \dot{r} \\ \dot{\theta} \end{pmatrix} = \begin{pmatrix} 0 \\ F(\theta) \end{pmatrix} \tag{2.33}$$

Thus we have a recipe for making $F(\theta)$ appear exclusively in the expression for $\dot{\theta}$ in cylindrical coordinates.

This strategy is tuned to the fact that, in the chosen phase-plane, the oscillator's limit cycle is in the counter-clockwise direction. Similar but slightly modified rules apply when dealing with a clockwise limit cycle. Trivially, sums and differences of these compositions work as well.

2.7 Independence of *Circuit Settling Time* and *Conversion Gain*: A Heuristic Explanation

We can now appreciate why detection time and conversion gain are coupled under full-state feedback, but this coupling may not be nearly as strong as once thought. For a specialist in NMR with long experience using output feedback from the two-slope limiter, this independence is not immediately clear. For completeness the following longstanding result from Viswanathan et al. [9] is given.



(a) Block diagram.

(b) Physical Realization

(c) Standard Circuit Model

Figure 2.16: Conceptual representation of a marginal oscillator

To derive an expression for the conversion gain, Viswanathan et al. first represent the marginal oscillator in Fig. 2.16c using the nonlinear differential

equation

$$\frac{d^2v_0}{dt^2} + \frac{1}{C}\left(\frac{1}{R} - g(v_0)\right)\frac{dv}{dt} + \frac{1}{LC}v_0 = 0 \tag{2.34}$$

where $-g(v_0) = -dG(v_0)/dv$ represents the conductance looking into the dependent current source. Using an approximation to the solution of a nonlinear second-order system developed by Krylov and Bogoliubov [24], Viswanathan et al. show that the conversion gain satisfies

$$G_C = \frac{\pi}{2\left(1 - \frac{g_B}{g_A}\right)g_A R \sin(2\theta)} \tag{2.35}$$

where θ is the solution to

$$\sin 2\theta + 2\theta = \frac{\pi}{\left(1 - \frac{g_B}{g_A}\right)}\left[\frac{1}{g_A R} - \frac{g_B}{g_A}\right], \text{ where } 0 \leq \theta \leq \frac{\pi}{2} \tag{2.36}$$

Furthermore, Viswanathan et al. show that the oscillation amplitude satisfies

$$A = v_T / \sin\theta \tag{2.37}$$

Equation (2.35) implies that one can achieve an arbitrarily large conversion gain by increasing the value of g_B towards g_A. Increasing the conversion gain does not come without a penalty. Viswanathan et al. approximate the transient response in oscillation amplitude A for small changes in losses R as a first-order linear differential equation, and show that the time constant of the response is

$$\tau = \frac{2R}{L\omega_n^2}G_C \tag{2.38}$$

As measurement bandwidth is the reciprocal of the time constant, Eq. (2.37) reveals that the price for increasing the conversion gain is a proportionate reduction in measurement bandwidth. In other words, increasing the conversion gain comes at the expense of increasing the measurement time.

The demonstration by Viswanathan et al. of coupling is not pedagogically obvious, as the method requires use of the averaging theory and resulting transcendental approximations given by Krylov and Bogoliubov [24]. In contrast, by forcing near radial symmetry and using a cylindrical coordinate system, in the sequel it will be shown that there is a more direct way of stating when and why the coupling occurs between conversion gain and measurement

time.

2.7.1 Coupled Case

Let us now give an example of a differential equation in one variable, $dr/dt = F(r)$, choosing a homogeneous function $F(r)$ where coupling is straightforward to calculate.

Remark. The radius r is lower case, where the resistance R is upper case, and likely to cause confusion. For the moment we are using the variable r to denote the distance of the limit cycle from the origin. This focuses attention more directly on the definitions of: equilibrium, conversion gain, eigenvalue and time-constant.

Let $F(r) = e^{-r} - a$ and let a be a fixed constant in the range $0 < a < 1$. This function of r exponentially decays from positive values near the origin to an asymptote that is located at a negative value $-a$ as depicted in Fig. 2.17. There is an equilibrium point where the homogeneous expression $F(r)$ goes through zero at a value r_0, i.e., $f(r_0) = 0$. A simple calculation shows that $r_0 = -\ln(a)$.

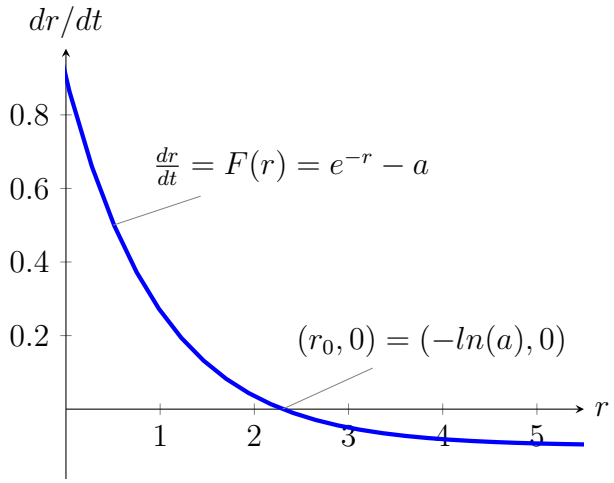


Figure 2.17: A more complicated differential equation

As the parameter a is allowed to get smaller, the equilibrium point r_0 will move more and more rapidly toward larger values. The conversion gain is calculated as

$$r_0 = -ln(a) \implies S_a^{r_0} = \frac{a}{r_0}\frac{\partial r_0}{\partial a} = \frac{a}{-\ln(a)}\left(\frac{-1}{a}\right) = \frac{1}{\ln(a)} \tag{2.39}$$

Further, one calculates the eigenvalue (or time constant) associated with the equilibrium point r_0 as

$$\left.\frac{dF(r)}{dr}\right|_{r=r_0} = \left.\frac{d}{dr}\left(e^{-r} - a\right)\right|_{r=r_0} = \left.-e^{-r}\right|_{r=r_0} = -e^{\ln(a)} = -a \tag{2.40}$$

Hence, the time constant for the solution trajectories is $1/a$, although this is an infinitesimal argument. For example, convergence from the left will be faster than from the right. Hence, it is important to realize that the infinitesimal definition of conversion gain is a linearization near an operating point, and not really the full story.

So this function $F(r)$ couples the *conversion gain* (position of $r_0(a)$) to the *time to detect* (eigenvalue at the equilibrium point $r_0(a)$). As $a \to 0$, the equilibrium point goes to infinity, $r_0 \to \infty$, and the time constant is driven to infinity as well, $1/a \to \infty$.

Hence, any feedback scheme, which in turn determines $F(r)$, necessarily couples the radius of the limit cycle to the time constant describing how rapidly the trajectory approaches the limit cycle radius, and would cause the same problem.

It only remains to be shown that one can exhibit a function that eliminates this undesirable coupling.

2.7.2 Decoupled Case

The best way to present the decoupling between the time-to-detection and the radius of the limit cycle is to show that the two quantities are quite similar to the eigenvalue and equilibrium point of a linear ordinary differential equation respectively.

Consider the diagram of a one-state, homogeneous ODE shown in Fig. 2.18 on the next page given by the equation of a line $dr/dt = m(r - r_0)$. As depicted, let the slope be negative, $m < 0$, and the equilibrium point be positive $r_0 > 0$. Therefore the point $r = r_0$ is a stable equilibrium point. Second, within the constraints given, the slope m and equilibrium point r_0

can be chosen independently of one another. The resulting trajectories of the solution with $r(t)$ graphed versus t will look similar to the those shown in Fig. 2.19 on the following page. Here, all solutions of the differential equation move toward an equilibrium point $r = r_0$ with a characteristic relaxation time constant of $1/m$.

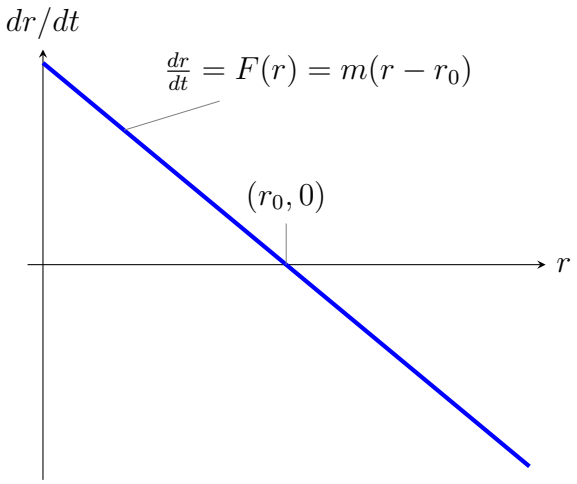


Figure 2.18: A simple differential equation

It is critical to realize that the conversion gain describes the value of the equilibrium point with respect to some parameter. In this case the decoupling is instantly obvious because the slope m (and hence, time constant $1/m$) does not appear in the expression describing the equilibrium point r_0. To get a parameter involved let us claim that $dr/dt = m(r - r_0(a))$. The first step in calculating the conversion gain is to find r_0 when $dr/dt = 0$. The expression is nearly a tautology.

$$0 = m(r - r_0(a)) \implies r = r_0(a) \tag{2.41}$$

It is already clear that the resulting expression can have no dependency on the slope m. Therefore the slope m does not appear in any of the subsequent equations for the conversion gain. This establishes that the time constant describing relaxation to the limit cycle is independent of the equilibrium point $r_0(a)$.

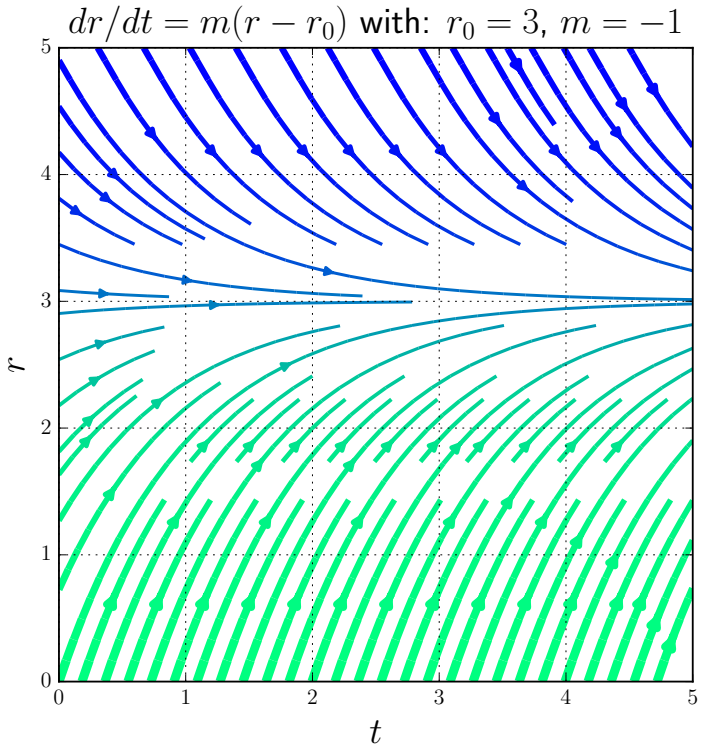


Figure 2.19: Streamline plot of the ODE given different initial conditions

2.7.3 How Decoupled Is Reasonable?

First, there needs to be a quick word of apology to those reading about marginal oscillators for the first time. In the last two subsections, it looks as though we are going to great lengths to state that a point on a line (in this case the ordinate dr/dt intercept with the abscissa r axis) is distinct from the slope of the line at that point. For those used to using the two-slope limiter, these two quantities have been strongly coupled in analysis spanning the last 40 years, and this presentation has to emphasize why this strong coupling is not strictly necessary. In principle, all that is needed is that the limit cycle radius changes in a way that is distinctly dissimilar from the way the slope at the limit cycle varies.

2.8 From a General Form to a Restricted Form of Feedback

We now formulate a prospective method of implementing nonlinear full-state feedback using a function (in this case, a control law). The overarching goal is to design a nonlinear control system that is easily implemented in hardware. We start with a raw guess regarding how to design the controller and systematically show where it gives us implementation difficulties. This allows us to isolate and then modify the controller to ease the details of implementation. This will lead to a more restrictive class of functions used for feedback. These restrictions could have simply been given up-front. But it is more instructive to use an example to show where the difficult parts arise and how they are readily fixed using specific restrictions.

Those wishing to skip over detailed analysis of this straw-man example in favor of starting with the restricted form are advised to resume reading at Section 2.9 on page 53.

2.8.1 Exemplar Implementation

The previous analysis now allows one to simply think in terms of controlling the radius of the one-dimensional state-space graphs such as Fig. 2.18 on page 42 and Fig. 2.17 on page 40. As an initial guess, we choose a cubic $F(r) = -\alpha(r - r_0)^3$ given in Fig. 2.20 on the following page. The guess is formulated based on the following strategy. The inflection point near $r = r_0$ gives a large conversion gain because any change in the loss term freely drives the zero crossing r_L either toward the origin (absorbing energy and potentially killing the oscillation) or toward infinity. However, as the limit cycle radius is driven away from the quiescent value of r_0, it is eventually caught by one of the two arms of the cubic that move toward plus and minus infinity more rapidly than a small perturbation of the loss term. Hence the oscillation will not be killed, and will not grow without bound.

Conjecture 2.8.1 (Linearity of error function in cylindrical coordinates)**.** *There will always be a residual error when we attempt to estimate the resistance R. The limit cycle occurs at the stationary point in the expression $\dot{r} = \varphi(r) - \gamma r$ where γ is an unknown constant and the function $\varphi(r)$ is a feedback*

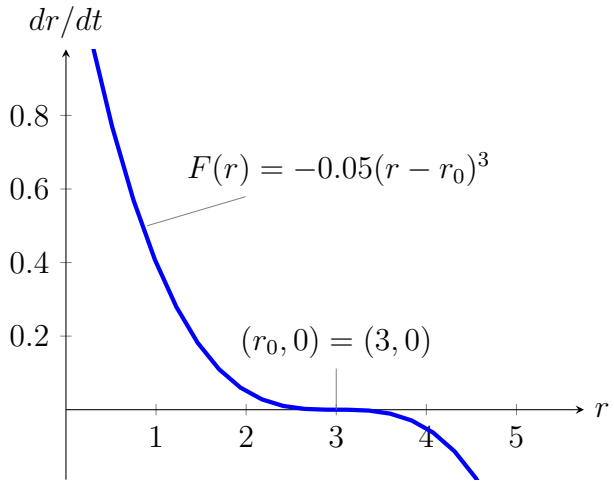


Figure 2.20: Exemplar function

law of our choice. The limit cycle will occur when $\dot{r} = 0$. *Equivalently,* $\gamma r = \varphi(r)$. *Proof is deferred until Chapter 3.*

Combining Theorem 2.8.1 on page 44 with Fig. 2.20 we obtain a more refined visualization depicted in Fig. 2.21 on the next page, where the small change in the resistance, and hence the loss term, is represented by the red line. As the red line must go through the origin, small values of slope (small losses or gains) will always intersect the cubic curve. The intersection point shows where the velocity outward (or inward) exactly matches the opposite velocity enforced by the cubic feedback scheme. Therefore the intersection is the radius of the limit cycle after a small change in the loss term. In practice, it is unrealistic to expect that the loss term can be canceled completely (i.e., the exact position of $r_0(0)$ can only be placed with finite precision). Therefore, in practice, it will be important to measure the radius of the limit cycle as an estimate of the residual loss term.

Also, curves with flatter inflection points at r_0 allow very large changes in the radius of oscillation due to tiny changes in the loss term, and hence will result in larger conversion gains. In practice, there will be a natural limit to how flat a curve can be implemented depending on choice of hardware or software used to implement the feedback law.

Using this line of reasoning, a sigmoidal function with an inverse, preferably with an inflection point at the r-intercept point, is a possibility for use in place of the cubic given. The slope of the red line represents tiny changes in the loss term. Recall that these changes are measured in parts per million.

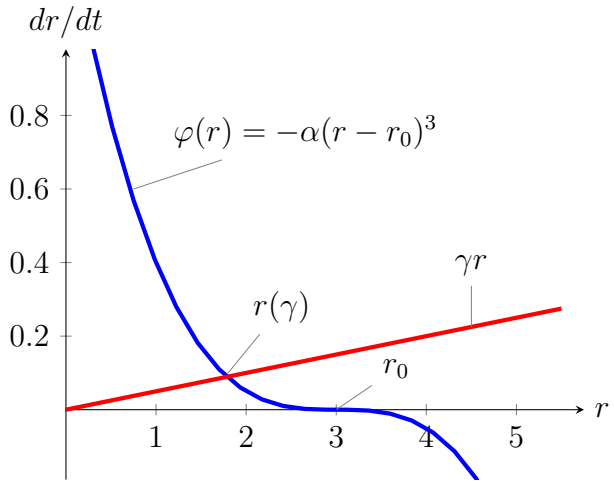


Figure 2.21: A cartoon showing that small changes (necessarily exaggerated) in loss or gain also change the radius of the limit cycle

It now only remains to estimate the quality of our exemplar function, and to use this result to algebraically express the feedback scheme, and examine if it can be implemented in the laboratory.

With this conjecture, the expected conversion gain can be calculated. The intersection point can be determined using the equilibrium expression

$$\gamma r = -\alpha(r - r_0)^3 \tag{2.42}$$

where γ is the slope of the red line of Fig. 2.21 and α is a positive constant. Normally one would solve for r, but the expression that results is too complicated to be useful. It is more useful to use geometric reasoning while staring at Fig. 2.21. For the moment let us approximate the answer by stating that the red line intercepts the blue curve at *approximately* an altitude above the r axis given by the lever arm r_0 times the slope of the red line γ, so we have

$$\gamma r_0 \approx -\alpha(r - r_0)^3 \tag{2.43}$$

This approximation will be discussed later in the exposition, so it is assigned the name *radial approximation*.

The notation here gets slightly confusing, so here we clarify. The quiescent limit cycle has a radius of r_0. When the quantity γ moves away from zero, the limit cycle radius changes to a new value $r(\gamma)$ which solves Eq. (2.43).

Therefore

$$r = r_0 - \left(\frac{\gamma r_0}{\alpha}\right)^{1/3} \tag{2.44}$$

From this we can readily calculate

$$S^r_\alpha = \frac{\alpha}{r}\frac{\partial r}{\partial \alpha} = \frac{\left(\frac{\gamma r_0}{\alpha}\right)^{1/3}}{3\left(r_0 - \left(\frac{\gamma r_0}{\alpha}\right)^{1/3}\right)} \tag{2.45}$$

$$S^r_{r_0} = \frac{r_0}{r}\frac{\partial r}{\partial r_0} = \frac{3r_0 - \left(\frac{\gamma r_0}{\alpha}\right)^{1/3}}{3\left(r_0 - \left(\frac{\gamma r_0}{\alpha}\right)^{1/3}\right)} \tag{2.46}$$

$$G_C = S^r_\gamma = \frac{\gamma}{r}\frac{\partial r}{\partial \gamma} = \frac{-\left(\frac{\gamma r_0}{\alpha}\right)^{1/3}}{3\left(r_0 - \left(\frac{\gamma r_0}{\alpha}\right)^{1/3}\right)} \tag{2.47}$$

To drive the conversion gain G_C, or any of these equations, to a maximum value, one would choose $\alpha \approx \gamma/r_0^2$, thus driving the denominator as close to zero as desired or practicable.

2.8.2 Exemplar Implementation

For all the ancillary machinery presented, the implementation step to translate the exemplar function to an algebraic description of the hardware implementation is almost anticlimactic. However, the places in the algebra where the exemplar function becomes difficult to implement in circuitry will directly motivate the *refined implementation* given in the sequel.

The following argument may be stated as if it were laboratory procedure; this may or may not be the case and is discussed in more detail later in Section 2.10 on page 55. The steps are presented to emphasize that finding successive approximations to the resistor value R has a sound mathematical foundation.

Start with the system description that is now restated here for the sake of clarity.

$$\begin{pmatrix} \frac{d}{dt}v_0 \\ \frac{d}{dt}i_L \end{pmatrix} = \begin{pmatrix} 0 & -\frac{1}{C} \\ \frac{1}{L} & -\frac{R}{L} \end{pmatrix} \begin{pmatrix} v_0 \\ i_L \end{pmatrix} + \begin{pmatrix} \frac{1}{C} & 0 \\ 0 & -\frac{1}{L} \end{pmatrix} \begin{pmatrix} i_{in} \\ v_i \end{pmatrix} \tag{2.48}$$

With no loss of generality, we take a guess at the value of the loss due to the resistance R, and call the guess $\widehat{R}$. The dependent voltage source is now set to mimic a *negative resistor* with the value $-\widehat{R}$, leaving a residual error

we shall name $R_{err} = \widehat{R} - R$.

Temporarily take the coordinate system as identifying the capacitor voltage v_0 as the x coordinate ($x(t)^2 = Cv_0(t)^2/2$) and the inductor current i_L as the y coordinate ($y(t)^2 = Li_L(t)^2/2$).

To force a limit cycle using the properties of the exemplar function, $dr/dt = F(r) = -\alpha(r-r_0)^3$. Using the compositional rules given earlier in Section 2.6.5 on page 36, we set things up as follows:

$$\begin{aligned}\begin{pmatrix}\cos(\theta)F(r)\\ \sin(\theta)F(r)\end{pmatrix} &= \begin{pmatrix}\frac{x}{\sqrt{x^2+y^2}}F(\sqrt{x^2+y^2})\\ \frac{y}{\sqrt{x^2+y^2}}F(\sqrt{x^2+y^2})\end{pmatrix}\\ &= \begin{pmatrix}\frac{\sqrt{\frac{C}{2}}v_0}{\sqrt{\frac{C}{2}v_0^2+\frac{L}{2}i_L^2}}F\left(\sqrt{\frac{C}{2}v_0^2+\frac{L}{2}i_L^2}\right)\\ \frac{\sqrt{\frac{L}{2}}i_L}{\sqrt{\frac{C}{2}v_0^2+\frac{L}{2}i_L^2}}F\left(\sqrt{\frac{C}{2}v_0^2+\frac{L}{2}i_L^2}\right)\end{pmatrix}\end{aligned} \tag{2.49}$$

For the moment, we set our estimate of the resistance R, called $\widehat{R}$, to the exact value of R to focus on the ideal case. So the closed-loop system uses the feedback law

$$\begin{aligned} i_{in} &= \frac{C\sqrt{\frac{C}{2}}v_0}{\sqrt{\frac{C}{2}v_0^2+\frac{L}{2}i_L^2}}F\left(\sqrt{\frac{C}{2}v_0^2+\frac{L}{2}i_L^2}\right)\\ v_{in} &= -Ri_L + \frac{-L\sqrt{\frac{L}{2}}i_L}{\sqrt{\frac{C}{2}v_0^2+\frac{L}{2}i_L^2}}F\left(\sqrt{\frac{C}{2}v_0^2+\frac{L}{2}i_L^2}\right)\end{aligned} \tag{2.50}$$

Substituting for the function $F(r)$ as given earlier

$$\begin{aligned} i_{in} &= -\frac{\alpha C\sqrt{\frac{C}{2}}v_0}{\sqrt{\frac{C}{2}v_0^2+\frac{L}{2}i_L^2}}\left(\sqrt{\frac{C}{2}v_0^2+\frac{L}{2}i_L^2}-r_0\right)^3\\ v_{in} &= -Ri_L + \underbrace{\frac{\alpha L\sqrt{\frac{L}{2}}i_L}{\sqrt{\frac{C}{2}v_0^2+\frac{L}{2}i_L^2}}}_{\text{Lead Radical}}\left(\underbrace{\sqrt{\frac{C}{2}v_0^2+\frac{L}{2}i_L^2}}_{\text{Embedded Radical}}-\underbrace{r_0}_{\text{Limit cycle radius}}\right)^3\end{aligned} \tag{2.51}$$

There are several points that should be emphasized. First, this is the feedback formulation only, and has to be substituted into the system equations. Second,

further simplification of these expressions is possible, but not meaningful in the current context, so they will be left in their present form. Third, the underbraces and labels in Eq. (2.51) on page 48 are required to unambiguously describe further restrictions placed on the class of functions that can be easily implemented in hardware. These restrictions are formulated in the next section.

A representative phase plane portrait is shown in Fig. 2.22. It has the properties that we have predicted. The zone of the limit cycle has a very gentle basin of attraction while trajectories that start farther away are rapidly drawn into the basin.

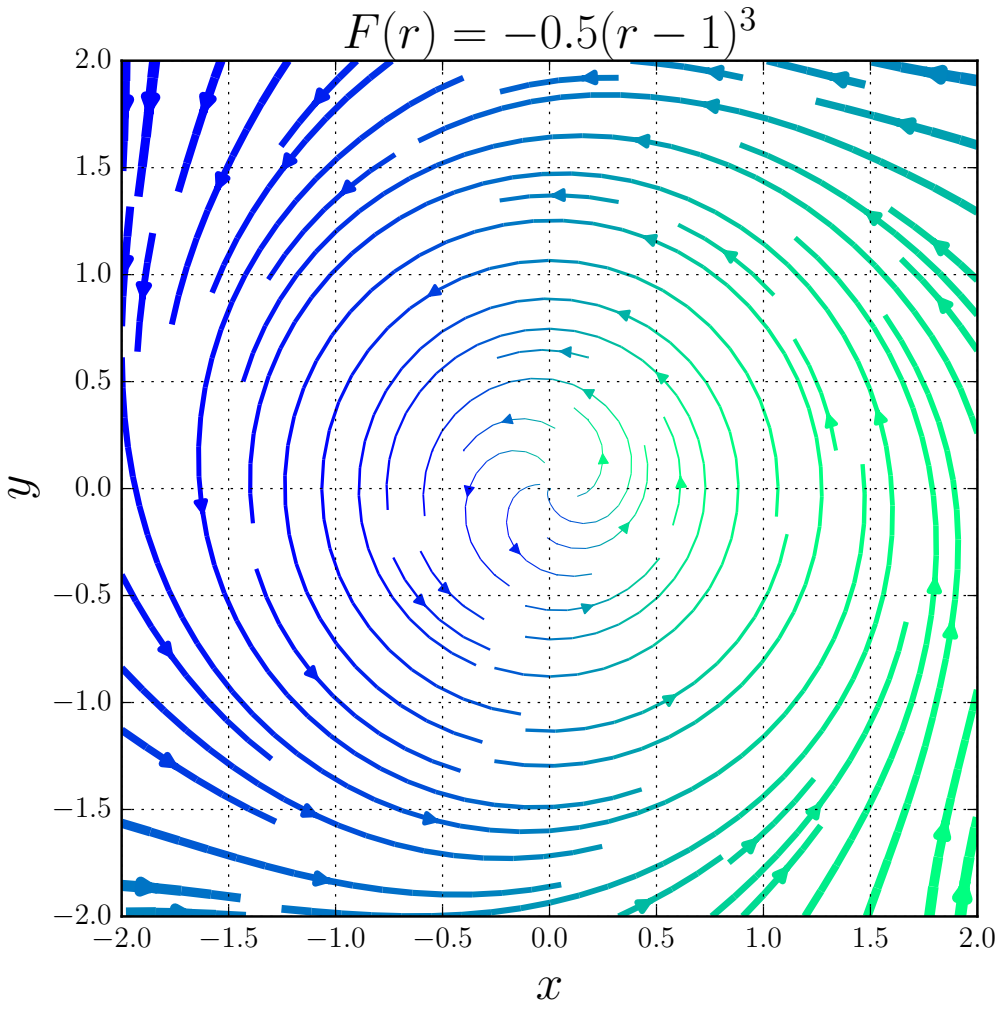


Figure 2.22: A cartoon streamline plot of the exemplar cubic function phase plane. Notice the limit cycle near a radius of 1. Here the constant α has been set to 0.5.

Both the algebra and the phase plane portrait are very encouraging. It appears one can iteratively make estimates of the resistance R and even when the precision of the implementation of the feedback loop is met, the residual error can be accurately measured by the distance between the expected and actual oscillation amplitude in the cylindrical coordinate system.

This example guess can be restricted in a couple of ways to make the feedback law much less difficult to implement. These restrictions still yield a broad class of functions $F(r)$ to support simplified implementation in hardware.

2.8.3 Restricting the Class of Functions to Usable Form

Following the mathematical analysis of the example function, listed above, gives invaluable guidance in restricting the feedback law to achieve functions that are easier to realize in hardware or software. Even with these restrictions applied, there is still an enormous class of functions available. We now motivate the restrictions and formulate a newly refined function within the new restrictions.

2.8.4 Eliminating Both the Lead Radical, and the Radial Approximation

Recall that in analog circuitry, addition and subtraction are standard building blocks. Multiplication of signals can also be achieved with a degree of care. However, division by a signal, or taking a root (of any order), are both comparatively difficult.

By fortunate coincidence we can neatly eliminate two different problems with the same restriction.

First, consider the last step of the analysis of the exemplar function given in Eq. (2.51) on page 48. The lead radical's placement in the denominator of the expression makes hardware implementation difficult. The lead radical comes from having to multiply by the term y/r or x/r when the coordinate system is changed from cylindrical to Cartesian. Second, consider that the radial approximation in the previous exposition flows from observing properties of the function depicted in Fig. 2.21 on page 46. It would have been an

advantage not to rely on the radial approximation of the equation

$$\gamma r = F(r)$$

by

$$\gamma r_0(0) \approx F(r)$$

However, both of these problems are neatly solved by only considering functions of the form

$$F(r) = rf(r) \tag{2.52}$$

This both eliminates the lead radical in Eq. (2.51) on page 48 and also means that the equation

$$\gamma r = rf(r) \tag{2.53}$$

immediately allows us to divide both sides by r cleanly obtaining $\gamma = f(r)$. Hence, this also eliminates using the radial approximation.

2.8.5 Eliminating the Embedded Radical

There is also a further simplification to be gained by restricting our attention to functions of the squared radius r^2. In this way the embedded radical term of Eq. (2.51) on page 48 will no longer require use of a square root.

To summarize, we will restrict our attention to functions of the form

$$F(r) = rf(r^2) \tag{2.54}$$

In the next section it will be shown that this has the desirable effect of relating the square of the radius, r^2, to the total energy stored in the passive circuit elements.

2.8.6 Eliminating the Limit Cycle Radius r_0

Here, we will fix the limit cycle radius to the value 1 in our coordinate system. Consider an ideal LC circuit at resonance. The total energy stored in the circuit at any one moment is a constant, E_T. Hence,

$$\frac{C}{2}v_0(t)^2 + \frac{L}{2}i_L(t)^2 = E_T \tag{2.55}$$

Expressed slightly differently,

$$\frac{C}{2E_T}v_0(t)^2 + \frac{L}{2E_T}i_L(t)^2 = 1 \tag{2.56}$$

This strongly suggests that the natural coordinate system is given by

$$x(t) = \sqrt{\frac{C}{2E_T}}v_0(t) \tag{2.57}$$

$$y(t) = \sqrt{\frac{L}{2E_T}}i_L(t) \tag{2.58}$$

And we shall adopt this convention for the balance of this exposition.

2.8.7 Use of Sigmoid Functions

A cursory examination of the overall shape of the cubic given in the example implementation can lead to the following guess. Suppose we take a known sigmoid function as our unknown $F(r)$, as depicted in Fig. 2.23. In practice, several sigmoid functions were attempted but do not lead to simple expressions for implementation in lab hardware. The polynomial expression of the next section gives a very simple expression lending itself to implementation more easily.

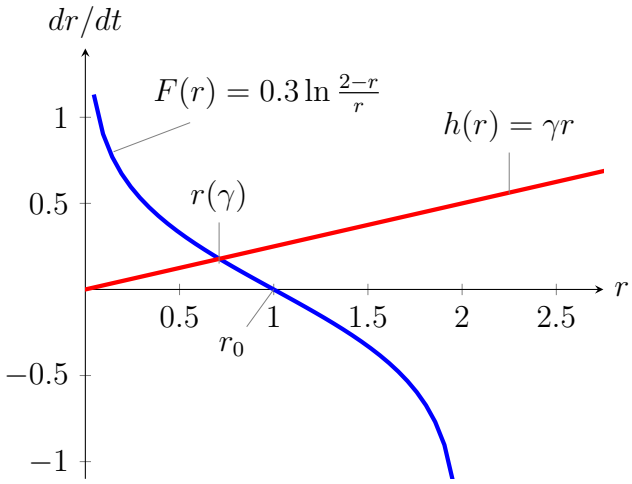


Figure 2.23: A sigmoid function

Therefore, the cubic used in the exemplar implementation turns out to be an inspired guess as the inflection point at the zero crossing continues to exist

independent of the choice of, say, the constant α.

2.9 A Refined Implementation

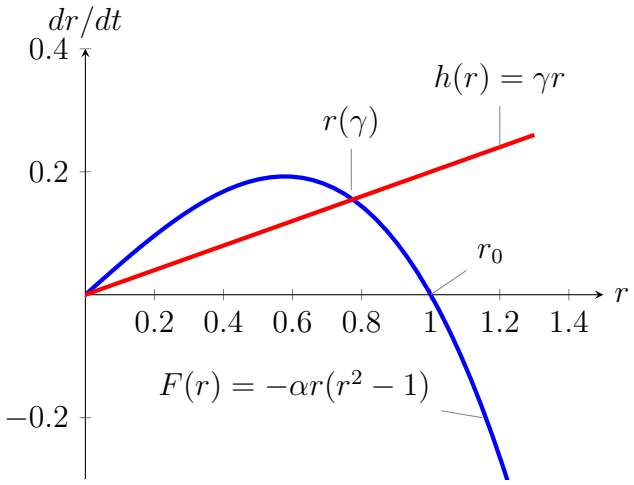


Figure 2.24: A cartoon showing that small changes (necessarily exaggerated) in loss or gain, γ, modify the radius of the limit cycle

Under the restrictions aforementioned we are limited to functions of the form

$$F(r) = rf(r^2) \tag{2.59}$$

Consider the function

$$F(r) = -\alpha r(r^2 - 1)^n \tag{2.60}$$

as depicted for $n = 1$ in Fig. 2.24.

Clearly, this function is chosen to yield many desired properties. For example, the point of inflection near the expected limit cycle radius $r = 1$ as the variable n is set to increasingly large odd integers yields a big conversion gain.

The conversion gain for $n = 1$ is:

$$S_\gamma^r = \frac{\gamma}{r}\frac{\partial r}{\partial \gamma} = \frac{-\left(\frac{\gamma}{\alpha}\right)}{2\left(1 - \left(\frac{\gamma}{\alpha}\right)\right)} = \frac{\gamma}{2(\gamma - \alpha)} \tag{2.61}$$

To drive the conversion gain S_γ^r to a maximum value, one would choose $\alpha \approx \gamma$ thus driving the denominator as close to zero as desired or practicable.

Details of the hardware implementation will set the lower limit on the user-selectable constant α. It is also important to note that this derivation is an approximation in the limit where γ goes to 0. At finite values for γ and α, the expression will not be accurate.

A graph showing the analog of Fig. 2.22 on page 49 is given in Fig. 2.25. The trajectories near the origin are being ejected to the limit cycle at lower velocity due to the extra prefactor r.

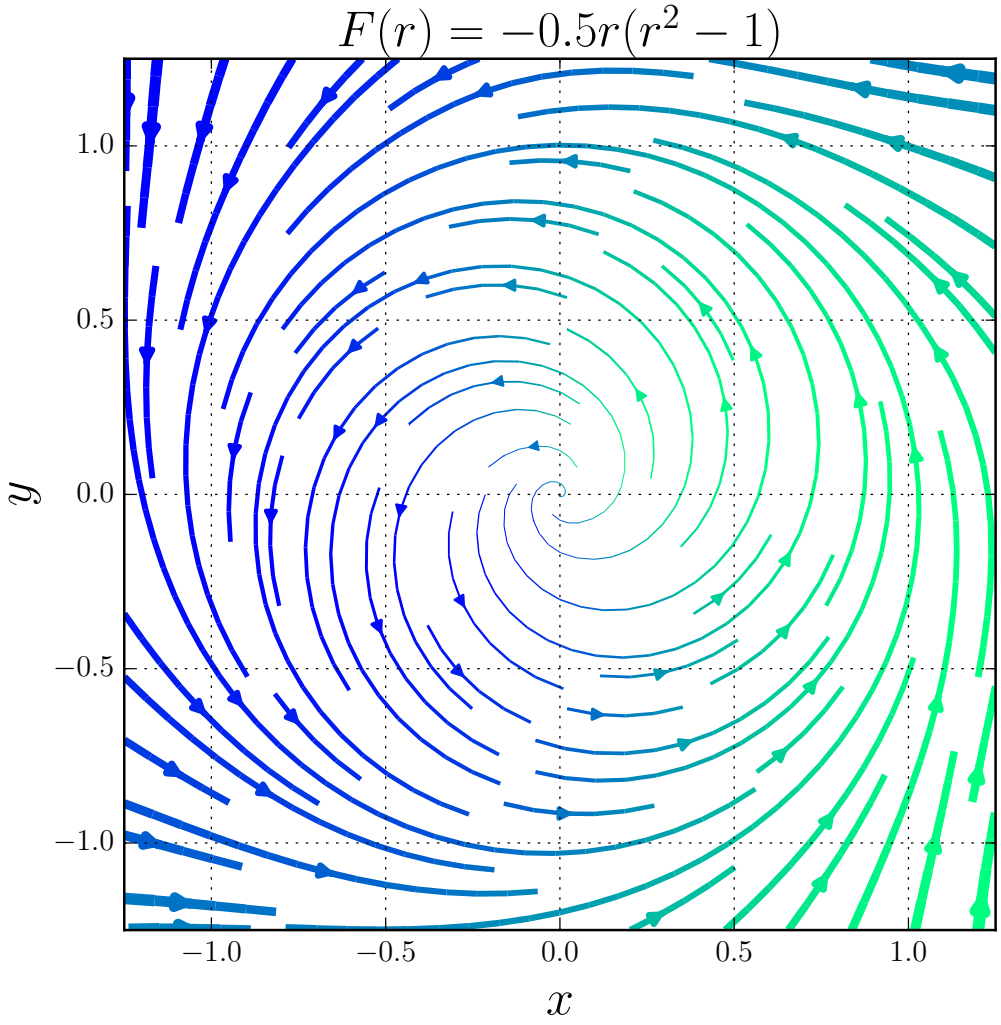


Figure 2.25: A cartoon streamline plot of the restricted and refined cubic function phase plane. Notice the limit cycle near a radius of 1. Here the constant α has been set to 0.5.

2.10 An Iteratively Tuned Implementation

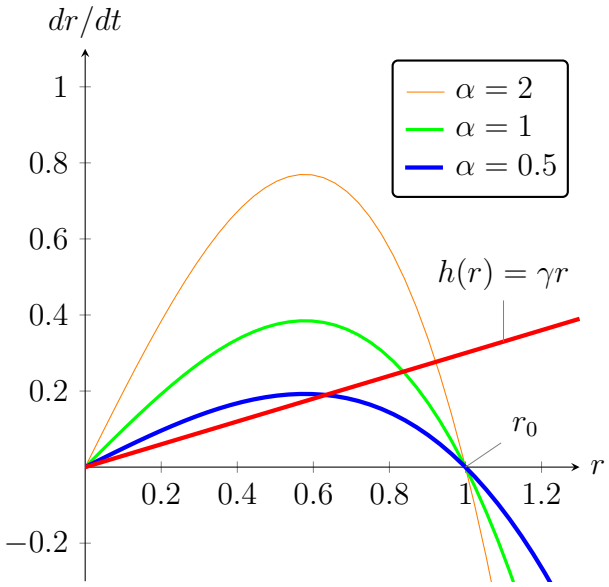


Figure 2.26: Phase plane with $F(r) = -\alpha r(r^2 - 1)$. Notice that as the constant α gets smaller, the crossover point between the two curves shifts to the left.

Again, consider the function $F(r) = -\alpha r(r^2 - 1)$, as depicted in Fig. 2.26. We are engaged in a guessing game where we have control over the parameter α and are attempting to guess the resistance R. Let us call our guess $\widehat{R}$.

Heuristically, the procedure is to feed our guess $\widehat{R}$ into the controller. The oscillator then settles on a limit cycle of radius r. From this radius, calculate the error in our estimate $\widehat{R}$, and update the estimate accordingly. The details of how to update $\widehat{R}$, and how to update the feedback loop constants, are discussed in the next chapter.

If our current estimate $\widehat{R}$ turns out to be poor, it may be best to make the parameter α larger. For example, if the oscillation is simply extinguished because our estimate is too small and the real resistor R causes trajectories in the phase space to be a stable spiral, boosting α drives the hump in the function $F(r)$ upward to assist in preserving the oscillation and obtaining a more meaningful estimate of the resistance R. The analogy in Fig. 2.26 is that the red line may be so steep as to not intersect the curve of our chosen function $F(r)$. Increasing the parameter α raises the height of the maximum of the function $F(r)$ until it does intersect the red line, and hence establishes

an oscillation at the amplitude of the crossover point between the line and the function.

Eventually the parameter α is big enough to establish an oscillation. The amplitude will give us a much better estimate of the resistance R. As our estimate improves, we will not need the parameter α to be set as high. Decreasing the parameter α increases the conversion gain and gives us a finer measurement of the resistance R. It may be an advantage to drop the parameter α to as low a level as is practical to obtain a maximally precise measurement of R.

CHAPTER 3

COORDINATE TRANSFORMATIONS AND SIMULATION RESULTS

3.1 Summary

Having now developed the mathematical techniques required to redesign a marginal oscillator using nonlinear full-state feedback, the current chapter simply summarizes the mathematics in detail. Any loose ends in the previous exposition will be relentlessly nailed down to closed-form algebraic expressions. I have erred on the side of taking small mathematical steps so that the exposition can be followed by others who would like to try out their own ideas.

3.2 The Marginal Oscillator Schematic

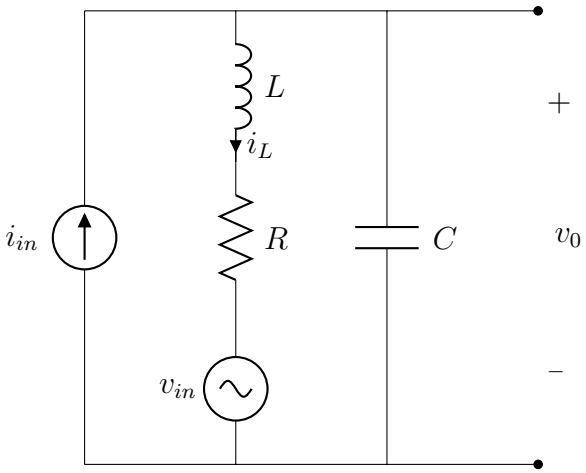


Figure 3.1: Open-loop marginal oscillator

Consider the schematic of an idealized marginal oscillator given in Fig. 3.1. The relationship between the input current I_{in} and output voltage V_0 is

determined using KCL on the top node. (As per previous discussion, all analysis can be accomplished in the Laplace transform space until we are ready to apply nonlinear feedback. At that point all equations will be taken back into time-domain.)

$$I_{in} = \frac{V_0 - V_{in}}{sL + R} + \frac{V_0}{(\frac{1}{sC})} \tag{3.1}$$

Identifying the first term on the right-hand side as the second state variable we can split the expression into

$$I_{in} = I_L + \frac{V_0}{(\frac{1}{sC})} \tag{3.2}$$

$$I_L = \frac{V_0 - V_{in}}{sL + R} \tag{3.3}$$

Analyzing each equation individually, we obtain

$$sCV_0 = -I_L + I_{in} \tag{3.4}$$

$$sLI_L = V_0 - RI_L - V_{in} \tag{3.5}$$

Hence the change to state-space matrix notation is straightforward:

$$\begin{pmatrix} sV_0 \\ sI_L \end{pmatrix} = \begin{pmatrix} 0 & -\frac{1}{C} \\ \frac{1}{L} & -\frac{R}{L} \end{pmatrix} \begin{pmatrix} V_0 \\ I_L \end{pmatrix} + \begin{pmatrix} \frac{1}{C} & 0 \\ 0 & -\frac{1}{L} \end{pmatrix} \begin{pmatrix} I_{in} \\ V_{in} \end{pmatrix} \tag{3.6}$$

Expressed in the state variables of inductor current $I_L(s) \Leftrightarrow i_L(t)$, and capacitor voltage $V_0(s) \Leftrightarrow v_0(t)$, we will call this the *lab coordinate system* and the equation is expressed in *lab units*.

In the next section this equation is taken into a dimensionless coordinate system where the limit cycle is a circle.

3.3 Transformation to Normalized Energy Coordinates

We take advantage of the concept that when equilibrium is established at a fixed limit cycle, one can take advantage of energy relation in an ideal LC

circuit at resonance. Recall that in an ideal LC circuit the energy equation is:

$$\frac{L}{2}i_L(t)^2 + \frac{C}{2}v_0(t)^2 = E_T \tag{3.7}$$

where E_T is a constant and represents the total energy in the circuit at a steady state resonance. The value of E_T is selected by the experimenter based on the operating range of various system components. By contrast, the variable $E(t)$ is defined to be the amount of energy trapped in the circuit at any one instant. The function $E(t)$ is not a constant for two reasons. First, the circuit may not have reached its limit cycle and may still be in a transient state. Second, even when the circuit relaxes to its limit cycle, the amount of energy trapped in the circuit reflects the quality of our resistance estimate $\widehat{R}$. Indeed the amplitude of the limit cycle gives us information about how big this error is.

So we identify a single coordinate with each term

$$\begin{aligned} x(t) = \sqrt{\frac{C}{2E_T}}v_0(t) \implies v_0(t) = \sqrt{\frac{2E_T}{C}}x(t) \\ y(t) = \sqrt{\frac{L}{2E_T}}i_L(t) \implies i_L(t) = \sqrt{\frac{2E_T}{L}}y(t) \end{aligned} \tag{3.8}$$

This immediately yields the relation $x(t)^2 + y(t)^2 = \frac{E(t)}{E_T}$. Therefore, we call this the *normalized energy coordinate system*, and the state quantities, $x(t) \Leftrightarrow X(s)$ and $y(t) \Leftrightarrow Y(s)$, are both dimensionless.

Under this coordinate transformation, let us determine what happens to the state equation in lab units Eq. (3.6) on page 58 when taken into normalized energy units.

First we calculate that

$$\begin{aligned} \frac{d}{dt}x(t) = \sqrt{\frac{C}{2E_T}}\frac{d}{dt}v_0(t) \\ \frac{d}{dt}y(t) = \sqrt{\frac{L}{2E_T}}\frac{d}{dt}i_L(t) \end{aligned} \tag{3.9}$$

In the Laplace domain we have:

$$\begin{aligned} sX = \sqrt{\frac{C}{2E_T}} sV_0 \implies sV_0 = \sqrt{\frac{2E_T}{C}} sX \\ sY = \sqrt{\frac{L}{2E_T}} sI_L \implies sI_L = \sqrt{\frac{2E_T}{L}} sY \end{aligned} \tag{3.10}$$

Now we substitute values of V_0 and I_L into Eq. (3.6) on page 58 to obtain

$$\begin{aligned} \begin{pmatrix} \sqrt{\frac{2E_T}{C}} sX \\ \sqrt{\frac{2E_T}{L}} sY \end{pmatrix} = \begin{pmatrix} 0 & -\frac{1}{C} \\ \frac{1}{L} & -\frac{R}{L} \end{pmatrix} \begin{pmatrix} \sqrt{\frac{2E_T}{C}} X \\ \sqrt{\frac{2E_T}{L}} Y \end{pmatrix} \\ + \begin{pmatrix} \frac{1}{C} & 0 \\ 0 & -\frac{1}{L} \end{pmatrix} \begin{pmatrix} \sqrt{\frac{2E_T}{L}} Y_{in} \\ \sqrt{\frac{2E_T}{C}} X_{in} \end{pmatrix} \end{aligned} \tag{3.11}$$

Gathering terms so that the new constants appear in the matrix and column vectors,

$$\begin{aligned} \begin{pmatrix} \sqrt{\frac{2E_T}{C}} sX \\ \sqrt{\frac{2E_T}{L}} sY \end{pmatrix} = \begin{pmatrix} 0 & -\frac{1}{C}\sqrt{\frac{2E_T}{L}} \\ \frac{1}{L}\sqrt{\frac{2E_T}{C}} & -\frac{R}{L}\sqrt{\frac{2E_T}{L}} \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} \\ + \begin{pmatrix} \frac{1}{C}\sqrt{\frac{2E_T}{L}} & 0 \\ 0 & -\frac{1}{L}\sqrt{\frac{2E_T}{C}} \end{pmatrix} \begin{pmatrix} Y_{in} \\ X_{in} \end{pmatrix} \end{aligned} \tag{3.12}$$

This allows dividing through by the lead constants on the left-hand side.

$$\begin{aligned} \begin{pmatrix} sX \\ sY \end{pmatrix} = \begin{pmatrix} 0 & -\frac{1}{C}\sqrt{\frac{C}{L}} \\ \frac{1}{L}\sqrt{\frac{L}{C}} & -\frac{R}{L} \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} \\ + \begin{pmatrix} \frac{1}{C}\sqrt{\frac{C}{L}} & 0 \\ 0 & -\frac{1}{L}\sqrt{\frac{L}{C}} \end{pmatrix} \begin{pmatrix} Y_{in} \\ X_{in} \end{pmatrix} \end{aligned} \tag{3.13}$$

Simplifying the radicals to obtain:

$$\begin{pmatrix} sX \\ sY \end{pmatrix} = \begin{pmatrix} 0 & -\frac{1}{\sqrt{LC}} \\ \frac{1}{\sqrt{LC}} & -\frac{R}{L} \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} \frac{1}{\sqrt{LC}} & 0 \\ 0 & -\frac{1}{\sqrt{LC}} \end{pmatrix} \begin{pmatrix} Y_{in} \\ X_{in} \end{pmatrix} \tag{3.14}$$

As the nonzero entries in the state matrix all have units of inverse time, the system is in homogeneous units. In other words X and Y must be in the same units. All of this can be checked using the quantities summarized in

Table 3.1: Dimensions and units of electrical and magnetic entities. $MLTQ$ is mass, length, time, charge. $LTI\Phi$ is length, time, current, voltage.

Name	System $MLTQ$	System $LTI\Phi$	Unit Name
Mass	M	$L^{-2}T^{3}I\Phi$	kg
Electric Charge	Q	TI	coulomb
Electric Inductive Capacity ϵ	$M^{-1}L^{-3}T^{2}Q^{2}$	$L^{-1}TI\Phi^{-1}$	farad/m
Magnetic Inductive Capacity μ	MLQ^{-2}	$L^{-1}TI^{-1}\Phi$	ohm sec /m
Electric Current Density J	$L^{-2}T^{-1}Q$	$L^{-2}I$	amp/m^2
Electric Current I	$T^{-1}Q$	I	amp
Electric Displacement D	$L^{-2}Q$	$L^{-2}TI$	amp sec/m^2
Electric Field Intensity E	$MLT^{-2}Q^{-1}$	$L^{-1}\Phi$	volt/m
Electric Potential	$ML^{2}T^{-2}Q^{-1}$	Φ	volt
Electric Capacitance	$M^{-1}L^{-2}T^{2}Q^{2}$	$TI\Phi^{-1}$	farad
Electric Resistance	$ML^{2}T^{-1}Q^{-2}$	$I^{-1}\Phi$	ohm
Magnetic Field Intensity H	$L^{-1}T^{-1}Q$	$L^{-1}I$	amp/m
Magnetic Induction B	$MT^{-1}Q^{-1}$	$L^{-2}T\Phi$	weber/m^2
Flux of Magnetic Induction	$ML^{2}T^{-1}Q^{-1}$	$T\Phi$	weber
Coefficient of Inductance $\mathcal{L}, M$	$ML^{2}Q^{-2}$	$TI^{-1}\Phi$	henry
Electric Energy	$ML^{2}T^{-2}$	$TI\Phi$	joule
Electric Power	$ML^{2}T^{-3}$	$I\Phi$	watt

Table 3.1. (In this table, units of measurement are summarized as mass M, length L, time T, charge Q, current I, and voltage Φ. In a dimensionless coordinate system, all entries in the state matrix must have units of inverse time. This provides an additional check on the algebra used thus far.) This expression can be simplified again using the resonant frequency $\omega = 1/\sqrt{LC}$:

$$\begin{pmatrix} sX \\ sY \end{pmatrix} = \begin{pmatrix} 0 & -\omega \\ \omega & -\frac{R}{L} \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} \omega & 0 \\ 0 & -\omega \end{pmatrix} \begin{pmatrix} Y_{in} \\ X_{in} \end{pmatrix} \tag{3.15}$$

3.4 Apply Feedback, Then Return to Lab Coordinates

We restate the last equation for clarity.

$$\begin{pmatrix} sX \\ sY \end{pmatrix} = \begin{pmatrix} 0 & -\omega \\ \omega & -\frac{R}{L} \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} \omega & 0 \\ 0 & -\omega \end{pmatrix} \begin{pmatrix} Y_{in} \\ X_{in} \end{pmatrix} \tag{3.16}$$

3.4.1 New Feedback Recipe

After a couple of iterations using various feedback functions, the following appears to yield the most compact algebraic expressions that are easily implemented in hardware.

$$F(r) = -\alpha r(r^2 - 1) \tag{3.17}$$

A quick calculation of the conversion gain starts with

$$\gamma r = -\alpha r(r^2 - 1) \tag{3.18}$$
$$\gamma = -\alpha(r^2 - 1) \tag{3.19}$$

Solving for r we obtain

$$r = \left(1 - \left(\frac{\gamma}{\alpha}\right)\right)^{1/2} \tag{3.20}$$
$$\frac{dr}{d\gamma} = -\frac{\left(\frac{\gamma}{\alpha}\right)}{2\gamma\left(1 - \left(\frac{\gamma}{\alpha}\right)\right)^{1/2}} \tag{3.21}$$
$$S_\gamma^r = \frac{\gamma}{r}\frac{dr}{d\gamma} = \frac{\gamma}{2\left(\gamma - \alpha\right)} \tag{3.22}$$

This can be made arbitrarily large by forcing the denominator to be tiny as $\alpha \longrightarrow \gamma$

3.4.2 Cylindrical Coordinates to Cartesian Coordinates

Start with the desired recipe for feedback.

$$F(r) = -\alpha r(r^2 - 1) \tag{3.23}$$

It is required to get out of the transform space to apply the nonlinear feedback. As per the previous result summarized in Section 2.6.5 on page 36, we use the transformation to Cartesian coordinates:

$$\begin{aligned} \frac{d}{dt}x(t) &= \frac{x(t)}{r}F(r) = -\frac{x(t)}{r}\alpha r(r^2 - 1) = -x(t)\alpha(r^2 - 1) \\ \frac{d}{dt}y(t) &= \frac{y(t)}{r}F(r) = -\frac{y(t)}{r}\alpha r(r^2 - 1) = -y(t)\alpha(r^2 - 1) \end{aligned} \tag{3.24}$$

where the squared radius is $r(t)^2 = x(t)^2 + y(t)^2$.

This feedback is applied in Normalized Energy coordinates. Additionally we include the feedback term intended to cancel the effect of the resistance R using an estimate of the resistance called $\widehat{R}$:

$$\begin{aligned} y(t)_{in} &= -x(t)\alpha(r(t)^2 - 1) \\ x(t)_{in} &= -\frac{\widehat{R}}{\omega L} y(t) + y(t)\alpha(r(t)^2 - 1) \end{aligned} \tag{3.25}$$

Therefore the state equation becomes

$$\begin{aligned} \begin{pmatrix} \frac{d}{dt}x(t) \\ \frac{d}{dt}y(t) \end{pmatrix} &= \begin{pmatrix} 0 & -\omega \\ \omega & -\frac{R}{L} \end{pmatrix} \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} \\ &\quad + \begin{pmatrix} \omega \\ 0 \end{pmatrix} \left[-x(t)\frac{\alpha}{\omega}(r(t)^2 - 1) \right] \\ &\quad + \begin{pmatrix} 0 \\ -\omega \end{pmatrix} \left[\frac{-\widehat{R}}{\omega L} y(t) + y(t)\frac{\alpha}{\omega}(r(t)^2 - 1) \right] \end{aligned} \tag{3.26}$$

For the purposes of looking at the vector field, here is the nonmatrix version of what is happening:

$$\begin{aligned} \frac{d}{dt}x(t) &= -\omega y(t) - x(t)\alpha(r(t)^2 - 1) \\ \frac{d}{dt}y(t) &= \omega x(t) + y(t)\frac{\widehat{R} - R}{L} - y(t)\alpha(r(t)^2 - 1) \end{aligned} \tag{3.27}$$

with $r^2 = x(t)^2 + y(t)^2$.

This representation is now sufficient to run simulations in the normalized energy coordinate system.

3.4.3 Normalized Energy Coordinates to Lab Coordinates

Substitute the coordinate transformations into Eq. (3.27) on page 63.

$$\begin{aligned}\sqrt{\frac{C}{2E_T}}\frac{d}{dt}v_0(t) =& -\omega\sqrt{\frac{L}{2E_T}}i_L(t)\\ &-\sqrt{\frac{C}{2E_T}}v_0(t)\alpha(r(t)^2-1)\\ \sqrt{\frac{L}{2E_T}}\frac{d}{dt}i_L(t) =& \omega\sqrt{\frac{C}{2E_T}}v_0(t)+\sqrt{\frac{L}{2E_T}}\frac{(\widehat{R}-R)}{L}i_L(t)\\ &-\sqrt{\frac{L}{2E_T}}i_L(t)\alpha(r(t)^2-1)\end{aligned} \tag{3.28}$$

with $r(t)^2 = x(t)^2 + y(t)^2 = \frac{C}{2E_T}v_0(t)^2 + \frac{L}{2E_T}i_L(t)^2 = \frac{E(t)}{E_T}$. Start the simplification by multiplying each term by $\sqrt{2E_T}$

$$\begin{aligned}\sqrt{C}\frac{d}{dt}v_0(t) =& -\omega\sqrt{L}i_L(t) - \sqrt{C}v_0(t)\alpha(r(t)^2-1)\\ \sqrt{L}\frac{d}{dt}i_L(t) =& \omega\sqrt{C}v_0(t) \quad +\sqrt{L}\frac{(\widehat{R}-R)}{L}i_L(t) - \sqrt{L}i_L(t)\alpha(r(t)^2-1)\end{aligned} \tag{3.29}$$

Divide both sides by the lead factor on the left-hand side.

$$\begin{aligned}\frac{d}{dt}v_0(t) =& -\frac{1}{C}i_L(t) - v_0(t)\alpha(r(t)^2-1)\\ \frac{d}{dt}i_L(t) =& \frac{1}{L}v_0(t) \quad +\frac{(\widehat{R}-R)}{L}i_L(t) - i_L(t)\alpha(r(t)^2-1)\end{aligned} \tag{3.30}$$

Now substitute $r(t)^2 = x(t)^2 + y(t)^2 = \frac{C}{2E_T}v_0(t)^2 + \frac{L}{2E_T}i_L(t)^2 = \frac{E(t)}{E_T}$ to obtain

$$\begin{aligned}\frac{d}{dt}v_0(t) =& -\frac{1}{C}i_L(t) - v_0(t)\alpha\left(\frac{E(t)}{E_T}-1\right)\\ \frac{d}{dt}i_L(t) =& \frac{1}{L}v_0(t) \quad +\frac{(\widehat{R}-R)}{L}i_L(t) - i_L(t)\alpha\left(\frac{E(t)}{E_T}-1\right)\end{aligned} \tag{3.31}$$

We are now right on the verge of having the original equation in lab coordinates for Eq. (3.6) on page 58. The only difference is that the feedback mechanism does not appear to enter the equations as is required by the circuit. In particular, the stack vectors containing ω do not exist in the original. However, this is easily fixed by noting that $\omega = 1/\sqrt{LC}$ and

substituting accordingly.

$$\begin{aligned}\frac{d}{dt}v_0(t) &= -\frac{1}{C}i_L(t) - v_0(t)\alpha\left(\frac{E(t)}{E_T} - 1\right)\\ \frac{d}{dt}i_L(t) &= \frac{1}{L}v_0(t) \quad + \frac{(\widehat{R} - R)}{L}i_L(t) - i_L(t)\alpha\left(\frac{E(t)}{E_T} - 1\right)\end{aligned} \tag{3.32}$$

Equivalently

$$\begin{aligned}\begin{pmatrix}\frac{d}{dt}v_0(t)\\ \frac{d}{dt}i_L(t)\end{pmatrix} &= \begin{pmatrix}0 & -\frac{1}{C}\\ \frac{1}{L} & -\frac{R}{L}\end{pmatrix}\begin{pmatrix}v_0(t)\\ i_L(t)\end{pmatrix}\\ &+ \begin{pmatrix}\frac{1}{C}\\ 0\end{pmatrix}\left[-Cv_0(t)\alpha\left(\frac{E(t)}{E_T} - 1\right)\right]\\ &+ \begin{pmatrix}0\\ -\frac{1}{L}\end{pmatrix}\left[-\widehat{R}i_L(t) + Li_L(t)\alpha\left(\frac{E(t)}{E_T} - 1\right)\right]\end{aligned} \tag{3.33}$$

The last matrix equation has been carefully factored to mimic the structure of the original lab coordinate system equation Eq. (3.6) on page 58. Therefore the feedback law is isolated within the square brackets of the expression of Eq. (3.33).

3.5 Does the Limit Cycle Occur as Predicted?

At long last, we are now in a position to provide a definitive proof of the theorem that was stated in the previous chapter without proof. To avoid confusion, it is stated again here.

Theorem 3.5.1 (Linearity of error function in cylindrical coordinates)**.** *There will always be a residual error when we attempt to estimate the resistance R. The limit cycle occurs at the stationary point in the expression $\dot{r} = \varphi(r) - \gamma r$ where γ is an unknown constant and the function $\varphi(r)$ is a feedback law of our choice. The limit cycle will occur when $\dot{r} = 0$. Equivalently, $\gamma r = \varphi(r)$.*

Proof. Consider the equations

$$\begin{aligned}\frac{d}{dt}x(t) &= -\omega y(t) - x(t)\alpha(r(t)^2 - 1)\\ \frac{d}{dt}y(t) &= \omega x(t) + y(t)\frac{(\widehat{R} - R)}{L} - y(t)\alpha(r(t)^2 - 1)\end{aligned} \tag{3.34}$$

with $r(t)^2 = x(t)^2 + y(t)^2$.

We now ask if the feedback schema will behave as predicted. In particular, the radius will stop changing when $dr/dt = 0$. This implies that $d(r(t)^2)/dt = 0$. Using the chain rule,

$$\frac{d(r(t)^2)}{dt} = \frac{d(x(t)^2 + y(t)^2)}{dt} = 2x(t)\frac{dx(t)}{dt} + 2y(t)\frac{dy(t)}{dt} = 0 \tag{3.35}$$

Without loss of generality one can neglect the factor of 2 so that

$$x(t)\frac{dx(t)}{dt} + y(t)\frac{dy(t)}{dt} = 0 \tag{3.36}$$

Substituting we obtain

$$\begin{aligned} x(t)[-\omega y(t) \quad & -x(t)\alpha(r(t)^2 - 1)] \\ +y(t)[\omega x(t) + y(t)\frac{(\widehat{R} - R)}{L} \quad & -y(t)\alpha(r(t)^2 - 1)] = 0 \end{aligned} \tag{3.37}$$

$$\frac{\widehat{R} - R}{2L} = \alpha(r(t)^2 - 1) \tag{3.38}$$

or

$$\frac{R - \widehat{R}}{2L} = -\alpha(r(t)^2 - 1) \tag{3.39}$$

Comparing this equation with Eq. (3.19) on page 62 we find that

$$\gamma = \frac{(R - \widehat{R})}{2L} = \frac{-R_{err}}{2L} \tag{3.40}$$

thus yielding an expression for the slope γ in lab coordinates. □

Please notice that for later use of the Conversion Gain Chain Rule, we will need the following: $S^{\gamma}_{R_{err}} = 1$.

3.6 Analysis of Restricted Functions

For technical considerations, the feedback we wish to apply is best restricted to functions of the form:

$$\frac{dr(t)}{dt} = rF(r^2) \tag{3.41}$$

In particular, let us choose:

$$\frac{dr}{dt} = -\alpha r(r^2-1)^n \quad \text{for odd } n > 0. \tag{3.42}$$

Here, I will carry along the requirement "for odd $n > 0$" when required, rather than the more cumbersome construction where one substitutes, "$2k-1$" (with k a positive integer) for n.

We are interested in the set point where the curve $dr/dt = \gamma r$ intersects our curve.

$$\gamma r = -\alpha r(r^2-1)^n \quad \text{for odd } n > 0. \tag{3.43}$$

This occurs at (solving for r and calling the limit cycle r_L)

$$r_L = \sqrt{1 - \left(\frac{\gamma}{\alpha}\right)^{1/n}} \quad \text{for odd } n > 0. \tag{3.44}$$

Equivalently, we can also ask questions about the zero crossing of the function $\varphi(r)$

$$\begin{aligned} \frac{dr}{dt} =& \varphi(r) \text{ where} \\ \varphi(r) =& -\alpha r(r^2-1)^n - \gamma r \end{aligned} \tag{3.45}$$

where the zero crossing is at the radius of the limit cycle, r_L.

3.6.1 Time Constant

We would like to know the slope of $\varphi(r)$ at the stable equilibrium point (the zero crossing), $\varphi(r_L) = 0$

$$\frac{d\varphi(r)}{dr} = -\alpha(r^2-1)^{n-1}(2nr^2 + r^2 - 1) - \gamma \tag{3.46}$$

$$\left.\frac{d\varphi(r)}{dr}\right|_{r_L} = 2n\gamma - 2n\gamma\left(\frac{\gamma}{\alpha}\right)^{\frac{-1}{n}} \quad \text{for odd } n > 0. \tag{3.47}$$

Setting this to $-1/\tau$ and solving we get:

$$\tau = \frac{\left(\frac{\gamma}{\alpha}\right)^{1/n}}{2n\gamma\left(1 - \left(\frac{\gamma}{\alpha}\right)^{1/n}\right)} \quad \text{for odd } n > 0. \tag{3.48}$$

3.6.2 Conversion Gain

Again start with:

$$r_L = \sqrt{1 - \left(\frac{\gamma}{\alpha}\right)^{1/n}} \quad \text{for odd } n > 0. \tag{3.49}$$

Let us get the partial conversion gain:

$$S_\gamma^{r_L} = \frac{\gamma}{r_L}\frac{\partial r_L}{\partial \gamma} = \frac{\left(\frac{\gamma}{\alpha}\right)^{1/n}}{2n\left(\left(\frac{\gamma}{\alpha}\right)^{1/n} - 1\right)} \quad \text{for odd } n > 0. \tag{3.50}$$

By inspection of Eq. (3.48) on page 67 and Eq. (3.50) we observe:

$$\tau = \frac{-S_\gamma^{r_L}}{\gamma} \tag{3.51}$$

At this point, the only thing we need is an expression for the slight perturbation of the resistance γ in the lab frame. The picture will be completed by using the chain rule for conversion gain:

$$\tau = \frac{-S_\gamma^{r_L} S_{R_{err}}^{\gamma}}{\gamma S_{R_{err}}^{\gamma}} = \frac{-G_C}{\gamma S_{R_{err}}^{\gamma}} \tag{3.52}$$

3.6.3 Inductance Version

As noted earlier, when the resistance R is on the same leg of the tank circuit as is the inductor L, we have

$$\gamma = \frac{(R - \widehat{R})}{2L} = \frac{-R_{err}}{2L} \tag{3.53}$$

From which it immediately follows that $S_{R_{err}}^{\gamma} = 1$. Therefore, by the chain rule: $G_C = S_\gamma^{r_L} S_{R_{err}}^{\gamma} = S_\gamma^{r_L}$.

Plugging into Eq. (3.52) we get:

$$\tau = \frac{-G_C}{\gamma} = \frac{2LG_C}{R_{err}} \tag{3.54}$$

Observe that, as the error R_{err} is driven to zero, the time constant becomes unbounded.

3.6.4 Capacitance Version

Now make a comparison with the original design of the marginal oscillator as per the Schiano group. If we laboriously repeated Section 3.5 on page 65, we would have obtained

$$\gamma = \frac{R_{err}}{2CR(R + R_{err})} \tag{3.55}$$

Implying that $S^{\gamma}_{R_{err}} = R/(R + R_{err})$.

Substituting as before, we arrive at

$$\tau = \frac{-2C(R + R_{err})^2}{R_{err}} G_C \tag{3.56}$$

So there is no advantage in this strategy either. Again, as the error R_{err} is driven to zero, the time constant becomes unbounded.

3.7 Analysis of a General Class of Functions

We are now in a position to prove one of the central points of this dissertation. The following theorem shows that under the very general constraints we have placed on the feedback function, there is no way to avoid having the conversion gain mathematically locked to the settling time. The reader is gently advised not to panic. What will be seen is that the marginal oscillator design does not require a large conversion gain to work correctly and quickly. Indeed the simulation results given in Section 3.9 on page 73 will show that the settling time to estimating the circuit's resistance is distinct from the settling time of the circuit and is acceptably fast.

Theorem 3.7.1 (Conversion Gain and Measurement Time Constant)**.** *Every function of the form $rf(r)$ where the inverse $f^{-1}(r)$ exists, and where the equation $\frac{dr}{dt} = rf(r) - \gamma r$ has a stable equilibrium point at $r = r_L(\gamma)$, satisfies the equation*

$$\tau = \frac{-S^{r_L}_{\gamma}}{\gamma}$$

Proof. Let

$$\frac{dr}{dt} = rf(r) - \gamma r \tag{3.57}$$

In the notation used by the Schiano group this is simply written as

$$\varphi(r) = rf(r) - \gamma r \tag{3.58}$$

The stable equilibrium radius r_L is at the zero crossing of the function $\varphi(r)$ by construction. Setting the function to zero, $\varphi(r) = 0$ allows us to solve for r_L.

$$\begin{aligned} \gamma r &= rf(r) \\ \gamma &= f(r) \\ f^{-1}(\gamma) &= r \\ r_L(\gamma) &= f^{-1}(\gamma) \end{aligned} \tag{3.59}$$

Now calculate the time constant

$$\frac{\partial}{\partial r}\varphi(r) = f(r) + r\frac{\partial}{\partial r}f(r) - \gamma \tag{3.60}$$

We evaluate this at the zero crossing r_L to obtain

$$\frac{\partial}{\partial r}\varphi(r)\bigg|_{r=r_L} = r\frac{\partial}{\partial r}f(r)\bigg|_{r=r_L} \tag{3.61}$$

Hence

$$\tau = \frac{-1}{r\frac{\partial}{\partial r}f(r)}\bigg|_{r=r_L} \tag{3.62}$$

This is where the notation gets delicate. Let us refine the expression for the time constant, τ, further. Notice that we are deeply interested in the term

$$r\frac{\partial}{\partial r}f(r)\bigg|_{r=r_L(\gamma)} \tag{3.63}$$

By use of the chain rule, we know that

$$\frac{\partial}{\partial \gamma}f(r(\gamma)) = \frac{\partial}{\partial r}f(r)\frac{\partial}{\partial \gamma}r(\gamma) \tag{3.64}$$

Rearranging terms in this expression yields

$$\frac{\partial}{\partial r}f(r) = \frac{\frac{\partial}{\partial \gamma}f(r(\gamma))}{\frac{\partial}{\partial \gamma}r(\gamma)} \tag{3.65}$$

Now taking Eq. (3.65) on page 70 and the limit cycle radius r_L given in Eq. (3.59) on page 70 and plugging into Eq. (3.62) on page 70, we obtain

$$\begin{aligned}\tau &= \left.\frac{-\frac{\partial}{\partial\gamma}r(\gamma)}{r(\gamma)\frac{\partial}{\partial\gamma}f(r(\gamma))}\right|_{r=r_L=f^{-1}(\gamma)} \\ &= \frac{-\frac{\partial}{\partial\gamma}f^{-1}(\gamma)}{f^{-1}(\gamma)\frac{\partial}{\partial\gamma}f(f^{-1}(\gamma))} \\ &= \frac{-1}{f^{-1}(\gamma)}\frac{\partial}{\partial\gamma}f^{-1}(\gamma)\end{aligned} \tag{3.66}$$

Now we obtain the conversion gain of the radius r_L with respect to γ

$$\begin{aligned}S_\gamma^{r_L} &= \frac{\gamma}{r_L(\gamma)}\frac{\partial r_L(\gamma)}{\partial\gamma} \\ &= \frac{\gamma}{f^{-1}(\gamma)}\frac{\partial f^{-1}(\gamma)}{\partial\gamma}\end{aligned} \tag{3.67}$$

And by comparison between Eq. (3.66) and Eq. (3.67), we notice

$$\tau = \frac{-S_\gamma^{r_L}}{\gamma} \tag{3.68}$$

□

Remark. The proof above holds trivially for all functions $\varphi(r) = F(r)$ where $g(r) = F(r)/r$ possesses an inverse $g^{-1}(r)$.

3.7.1 Consequences of the Theorem

As may be apparent to alert readers of the previous chapter, although it is good to have target amplitude of oscillations $r = 1$, there is no particular advantage to putting an inflection point at the targeted radius.

$$\varphi(r) = -\alpha r(r^2 - 1)^3 \tag{3.69}$$

Nor, for that matter, is there an advantage to replacing the exponent 3 with larger odd integers and seeing if they can be made to work. The cubic version is repeated in Fig. 3.2 on the next page. Such an inflection point was originally intended to push the conversion gain to large values.

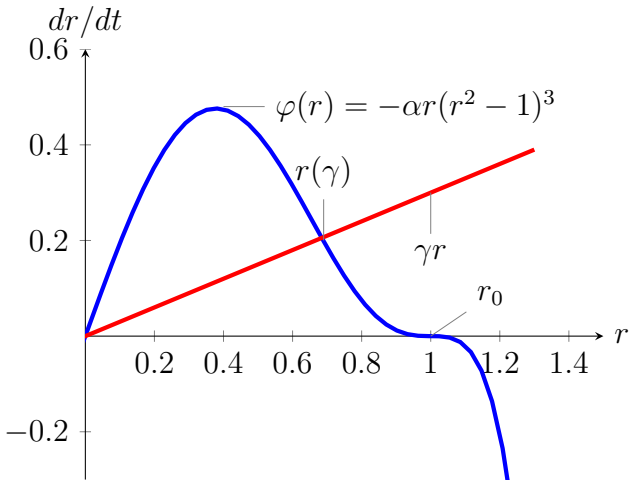


Figure 3.2: A cartoon showing the inflection point near the target limit cycle when using a cubic equation

However, this strategy is now not viable as it would only succeed in creating a larger range of radii where the settling time to the limit cycle would be unacceptably large.

3.8 Specific Values from Existing Lab Equipment

For the purposes of working towards verifying these equations in simulation, let us stick numbers on this. $L = 22.8\,\mu\text{F}$, $f = 3.0\,\text{MHz}$, $V_0 = 1.0\,\text{V}$ and the equation

$$\frac{L}{2} i_L(t)^2 + \frac{C}{2} v_0(t)^2 = E_T \tag{3.70}$$

allow us to solve for everything else.

At the chosen frequency we have

$$2\pi f = \frac{1}{\sqrt{LC}} \implies C = \frac{1}{L\omega^2} = 123.44\,\text{pF} \tag{3.71}$$

Using the notation $|v(t)|$ = amplitude of $v(t)$, the intercepts are:

$$\frac{C}{2}|v_0(t)|^2 = E_T \implies E_T = 61.72\,\text{pJ} \tag{3.72}$$

$$\frac{L}{2}|i_L(t)|^2 = E_T \implies |i_L(t)| = \sqrt{\frac{2E_T}{L}} = 2.3268\,\text{mA} \tag{3.73}$$

A summary of these results is listed in Table 3.2 on the following page for

easy reference.

Table 3.2: Constants for simulation

Name	Symbol and Value
Inductor	$L = 22.8\,\mu\text{H}$
Capacitor	$C = 123.44\,\text{pF}$
Angular Freq.	$\omega = 18.8497\,\text{Mrad}\,\text{s}^{-1}$
Maximum Current	$\lvert i_L(t)\rvert = 2.3268\,\text{mA}$
Maximum Voltage	$\lvert v_0(t)\rvert = 1.0\,\text{V}$
Total Power at Resonance	$E_T = 61.72\,\text{pJ}$

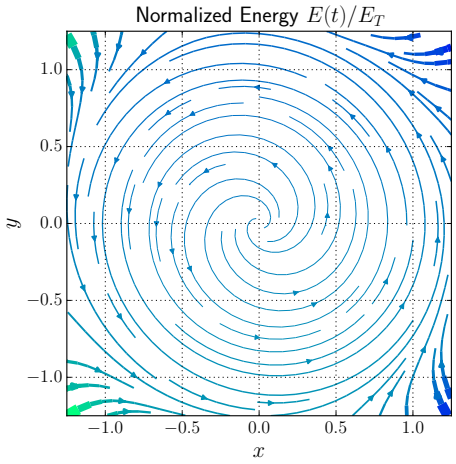
3.9 Simulation Results

The performance of the closed-loop system was studied using a computer simulation of the ODEs involved. As a starting point, the author was given access to a simulation written by the Schiano Group [25] in Matlab [29]. It was a fairly straightforward exercise to translate this code into Python [30] using the libraries Numpy [31] and Matplotlib [32]. Simulations were run both in normalized energy coordinates and in lab coordinates.

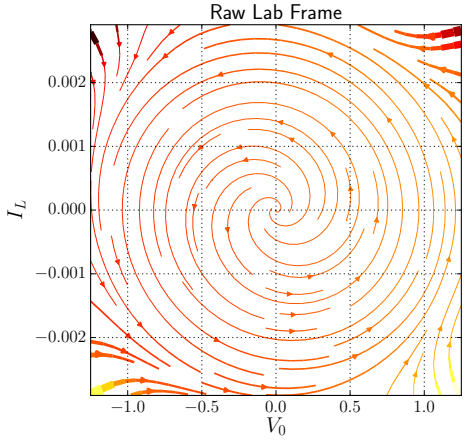
For the purposes of this write-up, the important observations are easier to explain in normalized energy coordinates unless otherwise noted.

3.9.1 Streamline Plots

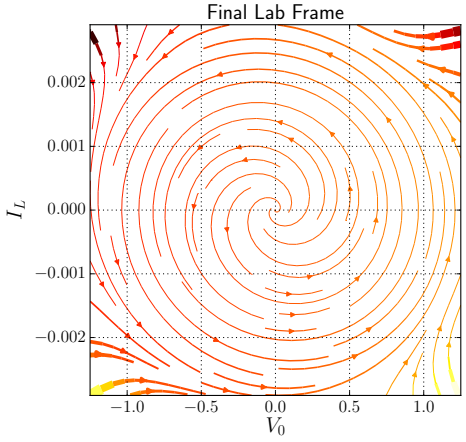
At the outset, the key equations Eq. (3.27), Eq. (3.32), and Eq. (3.33) were checked for any gross algebraic mistakes using streamline plot diagrams. The resulting streamline plots depicting the phase space in various coordinate systems appear in Fig. 3.3 on the next page. Respectively, they depict the ODE in: Normalized energy coordinates as given in Eq. (3.27), the simplified version of the lab coordinates given in Eq. (3.32) and the full lab coordinates as given in Eq. (3.33).



(a) Streamline plot in normal energy coordinates Eq. (3.27) on page 63.



(b) Streamline plot in lab coordinates as given in Eq. (3.32) on page 65.



(c) Streamline plot in lab coordinates as given in Eq. (3.33) on page 65.

Figure 3.3: Streamline plots of selected equations

3.9.2 Reduced Computation Time

As originally implemented in Matlab, the Schiano Group simulations on identical hardware were taking approximately 6 min computation time per 4 ms of simulated time. However, Matlab has known problems with making too many operating system calls and disc accesses while simulating ODEs. So it is not surprising that my initial Python simulations were about 12 times faster, yielding 30 s of computation time per 4 ms of simulated time.

3.9.3 Phase Noise

A cursory measurement of both Matlab and Python implemented simulations indicated that the results were accurate to no more than 4 significant digits in the best cases (about 1 part per 10,000), and much worse, exhibiting enormous sensitivity to how the *physical* parameters in the simulation were set. In stark contrast, the phenomena we wish to observe in the laboratory yield changes of resistance of about one part per million. Hence, extra effort was expended on understanding the details of the simulation model and software implementation. Several possible sources of error needed to be examined. For example, it became clear that the effect observed was very large compared to the error associated with relative and absolute tolerance limits given to the ODE solver (although these will be discussed later in Section 3.9.5 on page 78). The largest errors were traced to the use of a specific library routine entitled "peakdet", short for *peak detect*.

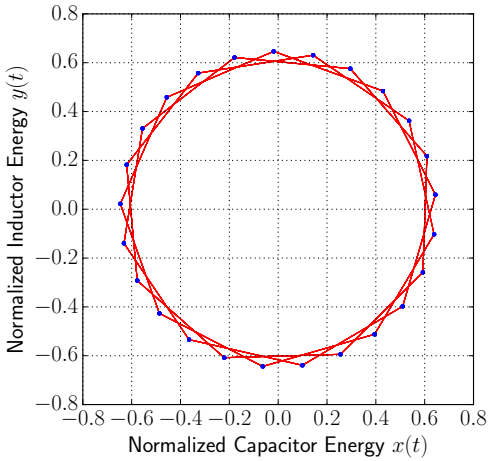
The error was induced by the simulator's estimate of the amplitude of the output voltage $|v_0(t)|$ by averaging the last hundred peaks of voltage in the simulation. This gives rise to a repeatable noise source based on the registration (typically, time varying) of the simulation time step to the phase of the simulated oscillation at the resonant frequency. Although the data at each time step landed with great precision on the limit cycle, the voltage component x can have peaks in this sampled signal that are less than the actual amplitude if the data points neatly miss those moments when the oscillation phase is precisely zero.

Surprisingly, the analysis of this problem dates back to the Greek mathematician Archimedes of Syracuse around 250 B.C. He attempted to estimate the value of the constant π ($= 3.14159\ldots$) by approximating a circle with an

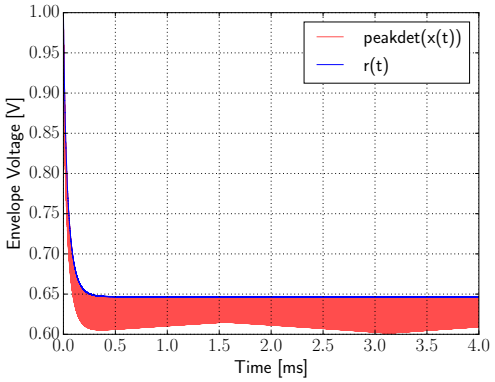
inscribed (and circumscribed) polygon of n sides and investigated how the estimate changed as n was pushed to increasingly higher values [33]. In our case, it is not difficult to show that if the chord between two adjacent vertices of a n sided regular polygon (similar to that shown in Fig. 3.4a) manages to straddle the x axis, then the projection onto the x axis can differ from the radius by as much as $error = 1 - \cos(\pi/n)$. A graph of this function in log-log coordinates is given in Fig. 3.4c. Unfortunately the convergence rate as the number of points in the polygon is pushed to higher values is miserably slow. It is interesting to note that an error of 1 part in 10^6 can only be achieved using a polygon of 2223 sides or more, and that 1 part in 10^9 would require a polygon of 70250 sides. As the resonant frequency in our simulation is at 3.0 MHz, this implies that the needed sample rates would be either 6.669 GHz or 156.17 THz respectively. Clearly, a much more clever strategy is required.

In this case the simulation can give the radius of the circle at each of the blue data points on the limit cycle (i.e. at each time step) described in the x, y plane of Fig. 3.4a. Given that the limit cycle is nearly circular, this yields a much more accurate estimate of the amplitude of the voltage signal $|v_0(t)| = \sqrt{\frac{2E_T}{C}}r(t)$. The graph shown in Fig. 3.4b is a representative example of the improvement achieved. Here the red line represents the voltage amplitude as estimated by use of the peak detector. This red curve gets badly smeared out into a large band as the phase relationship of the sample period changes with respect to the oscillation period. Barely discernible at the top of the red smear is a thin blue line depicting the voltage amplitude as estimated by calculating the radius at each data point. Clearly the blue line represents a much more reliable estimate of the voltage amplitude. (Indeed, as the relationship between the frequency of oscillation and the sample rate is known, the red curve in Fig. 3.4b is easily manipulated into many shapes including hypocycloids of various frequencies and amplitudes.)

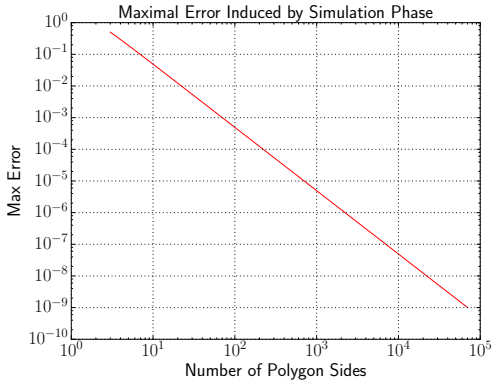
When this correction was made, the simulator agreed with theory to 1 part per 10^{12} plus a constant offset (see discussion in the next section), a marked improvement and more than sufficient for this application.



(a) Simulation's phase plane limit cycle



(b) Plot of peak detection vs. radius.



(c) Maximum error using a polygon of n sides per cycle

Figure 3.4: Plots showing phase noise induced by use of library routine "peakdet"

3.9.4 Predicted Versus Actual Limit Cycle Radius

By starting the simulation at known errors between the actual resistance R and the estimated resistance $\widehat{R}$, one observes the simulation relaxing to a limit cycle of a radius $r_L(\gamma, \alpha)$ as given in Eq. (3.44) on page 67. The displacement of the radius from $r(t) = 1$ can be substituted into the right-hand side of Eq. (3.39) on page 66 to yield a revised estimate of the resistance $\widehat{R}$. Therefore, one can start the simulation at any fixed error and measure the radius of the limit cycle. This result is shown in the graph depicted in Fig. 3.5 on the next page. It shows excellent agreement between the theoretical and actual limit cycle radius achieved.

Remark. In practice, the limit cycle radius r is mapped onto a revised estimate of the resistance $\widehat{R}$. However, the data depicted are generated by using the error $\widehat{R} - R$ as the independent variable and the limit cycle radius r is then dependent. It was deemed that this graph would be less confusing if presented in a manner consistent with its use rather than its generation, hence the choice of axes.

3.9.5 Differential Versus Additive Measurement Error

As depicted in Fig. 3.5 on the following page, the theory is in excellent agreement with the measured radius of the limit cycle. However, the graph is really quite misleading because we are attempting to make measurements that are accurate to 9 or more significant figures.

When simulation runs were made to measure the system's limit cycle, the crucial case was to compare the result of the limit cycle for a value of resistance R, versus the limit cycle for a value when the resistance was varied slightly to $R + \delta R = R(1 + 10^{-6})$. It was found, using an iterative algorithm (given in Section 3.9.6 on page 81), that within about 20 ms of simulated time, the simulator gave excellent *relative* values. In other words $\widehat{R}(R + \delta R) - \widehat{R}(R) = \delta R \pm 10^{-12}$; therefore, the differential error was approximately 10^{-12}. Somewhat more mysterious was a constant offset that would vary between 10^{-4} and 10^{-7} for reasons that were not initially apparent.

After considerable exploratory work, it was determined that the constant offset error was an *additive error* induced by the integration routine (and its associated parameters) used within the simulator. These additive errors were

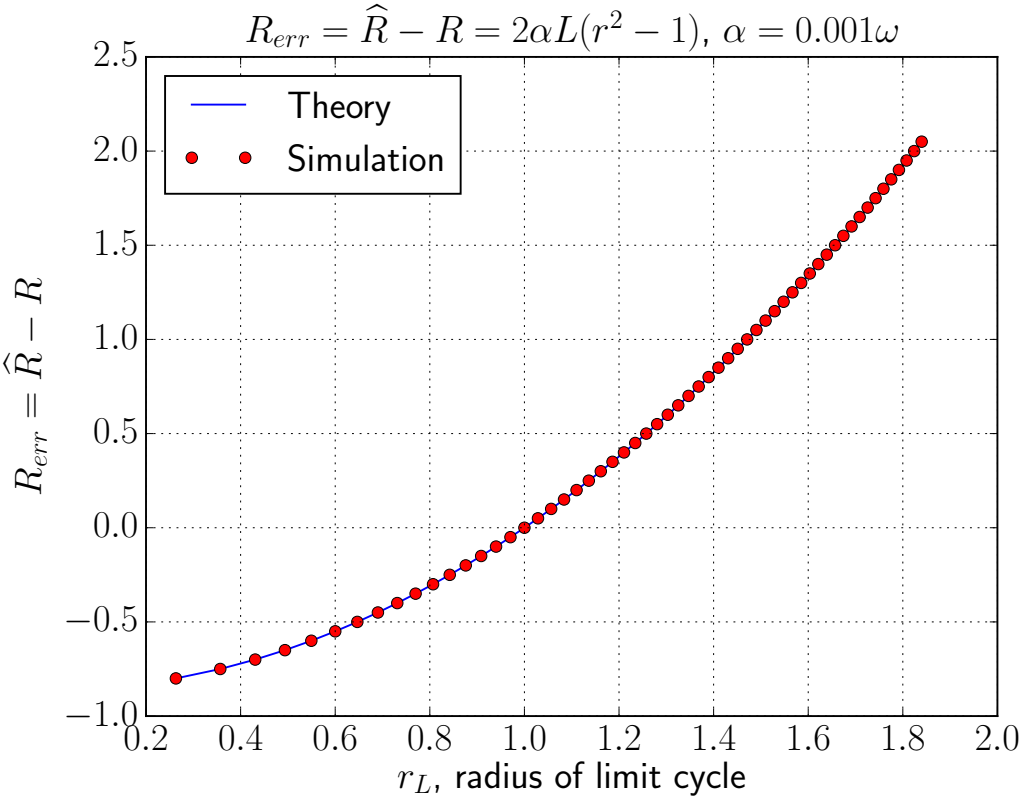


Figure 3.5: Plot of $R_{err} = \widehat{R} - R$ as a function of the limit cycle radius r as per Eq. (3.39) on page 66

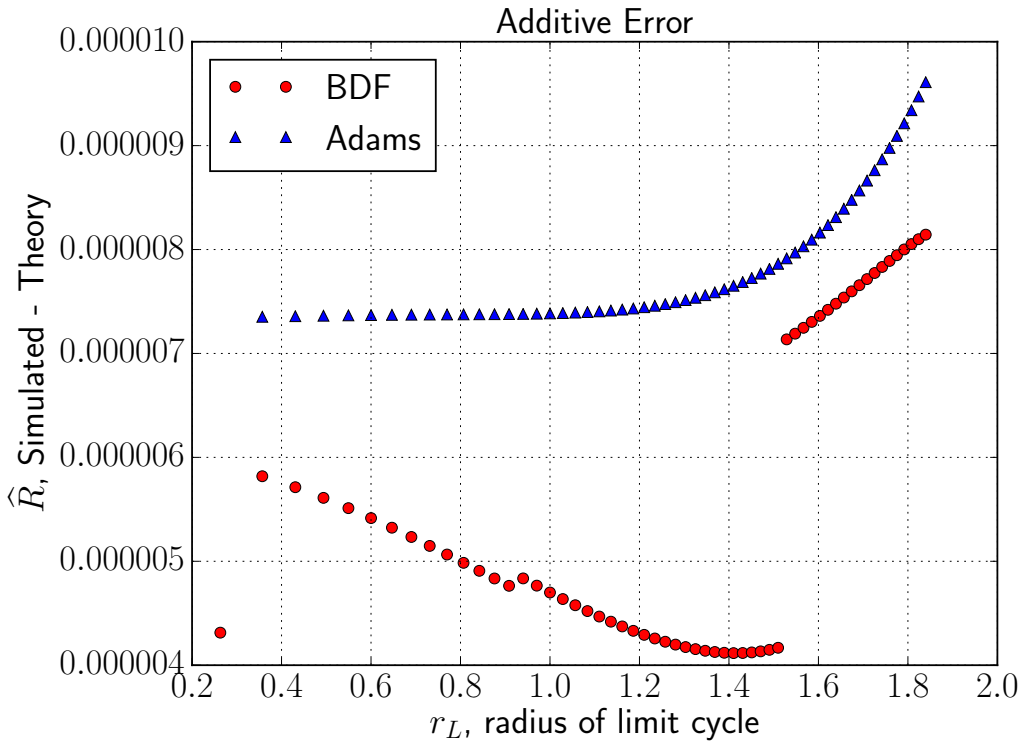


Figure 3.6: Plot of the simulated $\widehat{R}$ minus the theoretical $\widehat{R}$. The two data sets represent the effect of choosing a different integration routine.

completely repeatable from run to run. Therefore, the following two graphs were generated to indicate what was observed. In each of these graphs, rather than give a simple comparison between the theoretical and simulation curves as given in Fig. 3.5 on page 79, instead the difference is taken and the graph re-scaled to show the additive error in much greater detail. This is shown in Fig. 3.6 on page 80.

The only difference in the source code between the two graphs depicted in Fig. 3.6 on page 80 is the choice of solver. The two key lines from the source code are simply:

Adams

> solver = odespy.Vode(MO, adams_or_bdf = 'adams', order = 15, rtol= 1e-9, atol = 1e-10)

BDF

> solver = odespy.Vode(MO, adams_or_bdf = 'bdf', order = 15, rtol = 1e-9, atol = 1e-10)

The tolerances here were simply shrunk by factors of 10 until the ODE solver printed error messages stating that they had become too small.

For more complete coverage of the speed and accuracy of different Matlab solvers at different error tolerances, see the thesis of Tyson [25]. Please note that the thesis of Tyson never describes or diagnoses the phase noise as already documented in Section 3.9.3 on page 75. Instead, the physical parameters of the simulation are heuristically tuned to a low value of phase noise before analyzing the effect of changing the ODE solver's tolerances.

For purposes of this dissertation, the error induced by the ODE solver is so repeatable that using differential measurements reduces the error to well below the magnitude of the resonance phenomena we expect to see in laboratory data. For the moment, this result is more than adequate.

3.9.6 Iterative Algorithm

Consider the problem of attempting to conduct a CW-NMR experiment with a search coil of unknown resistance R. The following algorithm yields reasonable performance when implemented in simulation.

1. Set $\alpha = 0.0016$, and $\widehat{R} = 1.0$. (Prior experience with a coil of similar geometry and material composition may yield a good initial guess for $\widehat{R}$.)

2. Run the experiment and record the asymptotic value of the radius r_L after 1 ms of simulated time.

3. Update $\widehat{R}$ and α as follows:

$$R_{err} \Leftarrow 2\alpha\sqrt{L/C}(r^2 - 1) \tag{3.74}$$

$$\widehat{R}_{new} \Leftarrow \widehat{R}_{old} - R_{err} \tag{3.75}$$

   If $\alpha > (2 \times 10^5)R_{err}$ then

$$\alpha \Leftarrow \alpha/2 \tag{3.76}$$

   Else if $\alpha < R_{err}/2$ then

$$\alpha \Leftarrow 2\alpha \tag{3.77}$$

4. If α is below 10^{-12} end the run. Else, go back to step 2 and continue the run.

In simulation, even with initial guesses of the estimate $\widehat{R}$ that differ from the actual value of the resistance R by a factor of 10, the algorithm converges to the correct relative value to 1 part in 10^{12} in about 20 ms of simulated time. The value of α is manipulated to increase the conversion gain near the expected limit cycle. As α gets smaller, the conversion gain increases in magnitude until α begins to get smaller than γ.

The constant 2×10^5 in the inequality associated with Eq. (3.76) simply assures that alpha is much larger than the error of our resistance estimate; hence, it is time for the constant α to shrink into a range that drives up the conversion gain. If the constant α appears to be too small, it is important that we drive alpha high enough to initially capture a valid estimate of $\widehat{R}$, which happens at Eq. (3.77).

The halt condition given in step 4 of the algorithm is reached when the constant α has become so small that there is no point in pushing the algorithm to a higher conversion gain.

The Schiano group has settled on a convention that the conversion gain at the expected change of resistance at resonance be at least 8. This is

easily achieved by changing the update law given in Eq. (3.77) on page 82. Consider the expression for conversion gain given in Eq. (3.22) on page 62 and reiterated here.

$$S_{\gamma}^{r} = \frac{\gamma}{2\left(\gamma - \alpha\right)} \tag{3.78}$$

If the desire is to keep the magnitude of conversion gain bigger than some integer n, then the step size can be forced to shrink to include one or more steps in the critical region

$$\frac{2n}{2n+1} < \frac{\gamma}{\alpha} < \frac{2n}{2n-1} \tag{3.79}$$

3.9.7 Benefits of Normalized Energy Coordinates - Data Collapse

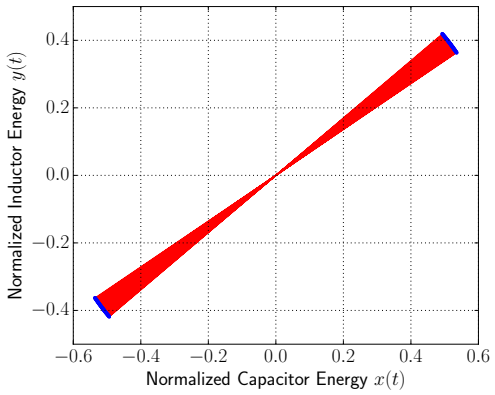
By arranging the feedback system to honor the system's natural proclivity to follow ellipses in the lab coordinate system and near-circles in normalized energy coordinates, we have demonstrated a similarity transformation into a dimensionless coordinate system which exhibits *data collapse*. The radius of every data sample in normalized energy coordinates is a snapshot giving the instantaneous energy divided by the asymptotic energy of the passive system components at resonance. This quantity now represents a complete characterization of the system's performance.

All NMR and NQR measurements require extraordinary effort to reduce noise. In the lab, many cycles of the system are box-car integrated to produce a noise-free picture of the system's response during a cycle (in pulsed spectroscopy, a pulse sequence; in continuous wave, one oscillation cycle). By contrast, we have now shown that the noise statistics can be collected with respect to radius (energy) and the box-car integration of a cycle collapses to averaging all data points of the radius independent of the phase relationship of the data sampler to the oscillator.

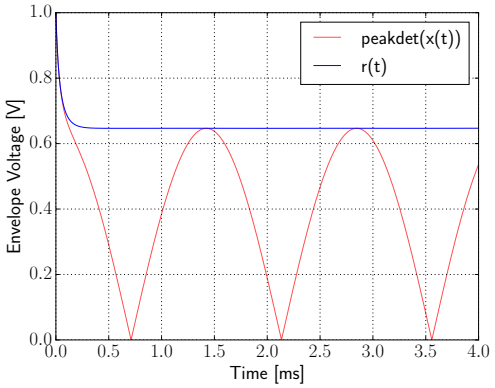
In this way, the number of samples required to achieve a fixed standard of noise reduction is greatly reduced. Minimally, it is reduced by the number of data samples taken per system oscillation. The graphs presented in Fig. 3.7 on the following page are representative of this fact.

In the graphs presented earlier in Fig. 3.4 on page 77 the system is oscillating at 3.0 MHz for 4.0 ms. Hence there are 12000 cycles depicted. We now present

a similar set of graphs in Fig. 3.7 that represent the system being grossly under-sampled at 8000 data points over the same interval. This is 1.5 cycles per sample. Yet notice that the estimate of $x(t)$ remains rock solid under these conditions where the peak detect algorithm gets hopelessly confused by the phase relationship between the sample frequency and oscillation frequency. It is hoped that when funding becomes available to pursue these results in laboratory conditions, the groups involved will be able to take advantage of this and related phenomena. Other natural consequences of this are also discussed below.



(a) Phase plane simulation at one sample per 1.5 cycles.



(b) Plot of peak detection vs. radius.

Figure 3.7: Data collapse. The system is easily characterized by radius, even at one data sample per 1.5 cycles.

3.9.8 Remarks on the Time Constants

One of the important consequences of this research is to realize the following. Each time the system relaxes to a limit cycle, it yields an independent estimate of resistance R given by the calculated update to $\widehat{R}$. The strong tendency in this work has been to slavishly follow the example of nearly every homework set given in courses on control: "Use this strategy to drive the error to zero." This standard admonition may turn out to be misguided in the present context. As each limit cycle yields a legitimate estimate of the resistance, the time constant associated with our estimate of resistance is wildly distinct from the relaxation time of the dynamical system to its limit cycle. This is a natural consequence of the algebra and associated approximations being

enormously simplified compared to the two-slope limiter.

Indeed the new system is now sufficiently distinct from the two-slope limiter that it will require a new approach to the lab procedures used. This is motivated by both the data collapse noted in the previous section, and the observation that each limit cycle (within a reasonable range) yields a legitimate independent measurement. Even when the possibly misguided strategy of 'force the error to zero' is used, as it is in Section 3.9.6 on page 81, the simulator can drive the differential error $R_{err} = \widehat{R} - R$ to less than $1\,\mathrm{p}\Omega$ in about 16 ms of simulated time. This is a slightly surprising result as one would think that the system is being driven into the precise range where its settling time is the slowest. Yet, the resulting differential accuracy is more than sufficient for, for example, CW-NQR.

CHAPTER 4

ANALYSIS OF PARASITIC RESISTANCE IN THE CW NQR SPECTROMETER

4.1 Summary

The marginal oscillator can measure a resistance with exceptional accuracy. In the previous chapters, we have been careful to make sure that the simulated system is accurate to at least one part in 10^9. So it is reasonable to ask, "Will the marginal oscillator be sensitive to parasitic resistance within its own circuitry?" In this chapter, we list all sources of parasitic resistance, and carefully determine if they have been analyzed previously. As it will turn out, the variable capacitor is the most significant, and last, unanalyzed portion. In practice, the variable capacitor is a parallel bank of relay-switched capacitors and two varactors. This exposition analyzes this parallel bank in three different ways. First, the time constants of the idealized circuit are compared to those from a second circuit with parasitic resistances. The resulting time constants are compared and found to be dissimilar by a factor of 10^3. Hence the time constants due to parasitic resistances, when the circuit is near resonance, are negligible. Second, the change in quality factor of the marginal oscillator circuit is analyzed with and without parasitic resistance and found to be very similar below 3 MHz. Above that frequency, one should consider replacing the varactors with a different technology. Third, we analyze the open-loop marginal oscillator circuit using Bode plots, and show that the transfer function at resonance is negligibly different.

4.2 Sources of Parasitic Resistance

The following list is based on information from the Schiano group at Penn State [34], and organized given the results contained in this chapter. The

measured sources of parasitic losses are, in decreasing order of magnitude for a CW-NQR, as follows:

1. Self-resistance of the probe coil
2. Losses in the varactor diodes for tuning of the resonant frequency
3. Equivalent series resistance of the fixed capacitors in the circuit
4. Trace resistances of the printed circuit board
5. Losses associated with mutual inductance, e.g., metallic shielding of the magnetic resonance instrument
6. Radiation losses
7. Resonance coupling to nuclei in the test sample

Previous research has summarized the effects of self-resistance of the probe coil, losses due to mutual inductance, and radiation losses. These losses are all characterized by analyzing their near equivalent effect on the value of the self-resistance of the probe coil R in Fig. 4.1 on the following page. The essence of the previous analysis is that parasitic resistances are smooth functions of frequency; therefore, tiny changes in the apparent self-resistance of the probe coil R associated with coupling to the nuclei are detected with a lock-in amplifier and yield a distinct and comparatively narrow Lorentzian peak as a function of frequency [35, 36].

For the purpose of completing the analysis, a repetition of the schematic of the marginal oscillator is given in Fig. 4.1 on the next page. We now turn our attention to the capacitor C in this diagram. In the lab, this is implemented as a switched capacitor bank. A simplified schematic of this circuit is given in Fig. 4.2 on page 89. Here the switches $s_3 ... s_9$ are used as rough tuning to get to the resonant frequency. The varactor C_1 is used for fine tuning of the spectrometer frequency and is therefore unswitched in normal use. The varactor C_2 is used for modulation as part of detection with a lock-in amplifier and is also unswitched in normal use.

If all the capacitors and varactors in the circuit were ideal capacitors, the resistances $R_1 ... R_9$ would be zero and the parallel capacitors would simply add to a single idealized value, C. However, the parasitic resistances are measurable. Data from the Schiano group summarizing their values

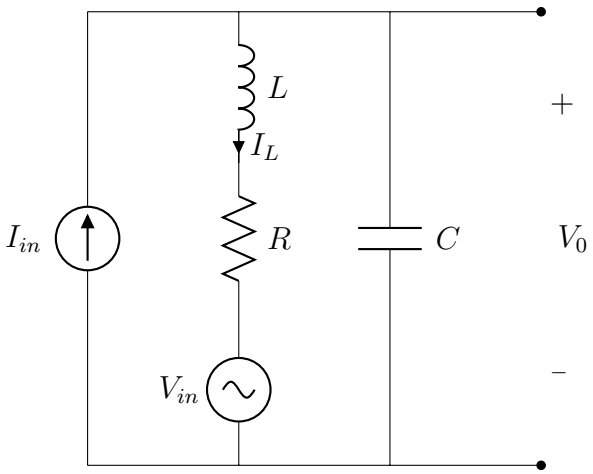


Figure 4.1: Open-loop CW circuit

are as given in Table 4.1 on page 90. It is interesting to note that those resistances that are measurable with the group's network analyzer agree with the calculated values based on the manufacturer's specifications to within 1%. Further analyzing the schematic using the capacitor bank in Fig. 4.2 on the following page yields a state-space equation with dimension equal to one state for the inductor L plus an additional state for each capacitor in the circuit.

Therefore, it is important to map out a solution strategy given that the number of state-space variables may become as large as 10. What is shown is that the eigenvalues of this large state-space equation are dominated by two complex conjugate poles that are relatively close to the origin compared to the other poles. All other poles are much farther from the origin, by about a factor of 10^3. Hence the system is correctly modeled using the simplified diagram as presented in Fig. 4.1. To confirm this, a sequence of graphs showing quality factor as a function of frequency and some representative Bode plots are exhibited as well.

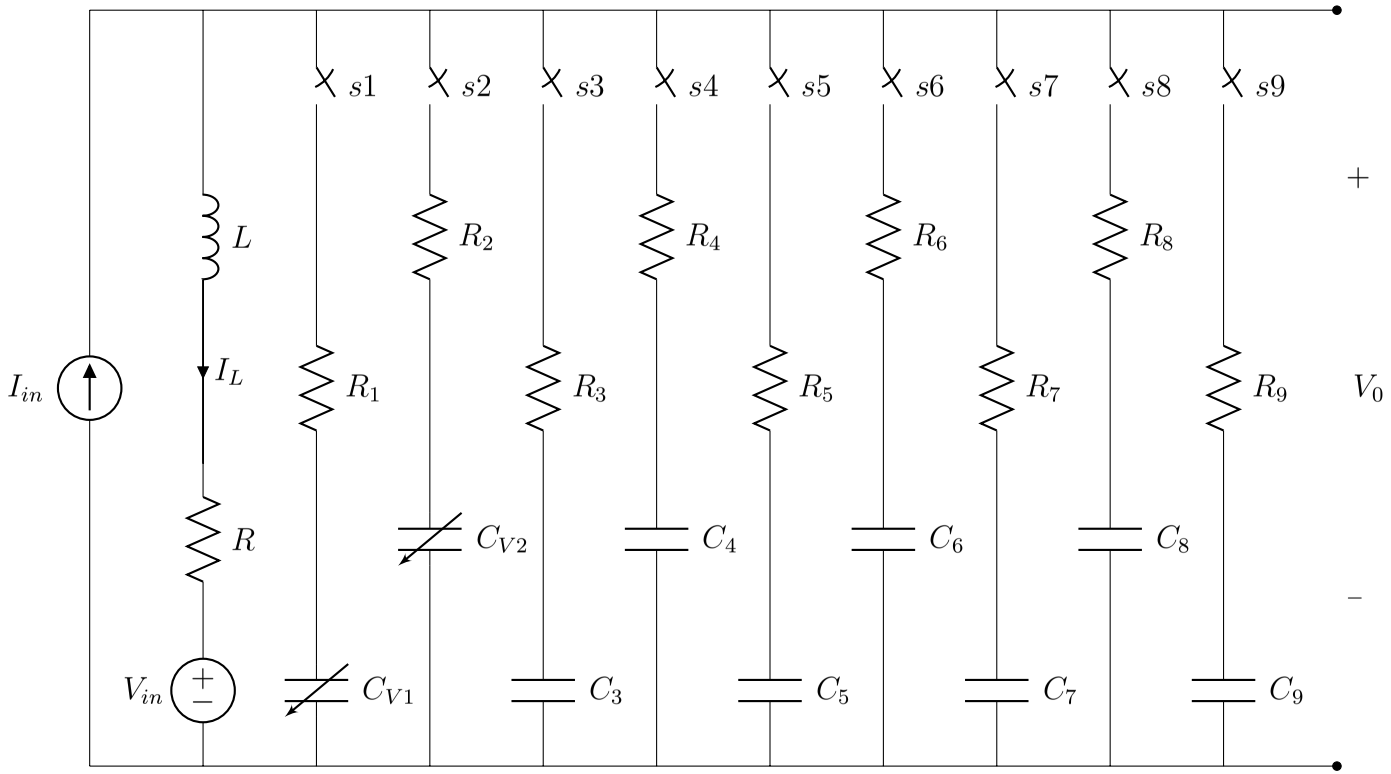


Figure 4.2: Circuit diagram of the switched capacitor bank. Resistances $R_1 ... R_9$ represent parasitic resistances.

Table 4.1: Calculated and measured values for parasitic resistance. '†' is calculated from manufacturer's specification of quality factor Q. 'M' is as measured, with Q as a dependent variable. When both are available, they agree to within 1%. The PCB trace resistance is 0.05 Ω, hence the difference between the ESR and the Total Resistance column. The inductance L is 22.8 µH. The resistance R is 1.5 Ω measured at 3.0 MHz

n	Capacitor	Quality Factor	Frequency	ESR	ESR Meas. or Spec.	Total
	C_n	Q	f	$R_{Cn} = \frac{1}{2\pi f Q C_n}$		R_n
	pF		MHz	Ω	Ω	Ω
9	470	1000	3.00	0.11	†M	0.16
8	240	1000	3.00	0.22	†	0.27
7	120	1000	3.00	0.44	†	0.49
6	62	1000	3.00	0.86	†	0.91
5	30	3000	3.00	0.59	†	0.64
4	15	3000	3.00	1.18	†	1.23
3	10	3000	3.00	1.77	†	1.82
2	39.5	141	3.00	9.55	M	9.6
1	29	202	3.00	9.05	M	9.1

4.3 Analysis of Parasitic Resistances Associated With the Switched Capacitor Bank

4.3.1 1st Analysis: Eigenvalues of the State-Space Model

The idealized problem in Fig. 4.1 on page 88 gives a state-space equation that has already been derived.

$$\begin{pmatrix} sI_L \\ sV_C \end{pmatrix} = \begin{pmatrix} -R/L & 1/L \\ -1/C & 0 \end{pmatrix} \begin{pmatrix} I_L \\ V_C \end{pmatrix} + \begin{pmatrix} -1/L & 0 \\ 0 & 1/C \end{pmatrix} \begin{pmatrix} V_{in} \\ I_{in} \end{pmatrix} \tag{4.1}$$

However, in the case of the capacitor bank, the parasitic resistances $R_1 ... R_9$ do not allow the capacitors to be added in parallel. The full state-space model has one state for the inductor plus one state for each capacitor switched into the circuit. Hence the full model has as many as 10 states. There is significant challenge associated with keeping the resulting algebraic expressions of the state-space model small enough to fit on standard paper. Hence, we shall need some clever definitions to keep the expressions short enough.

The most straightforward solution method uses substituted sources for each individual active component and then systematically solves by superposition, as given in Lathi on page 600 [37]. By solving these problems using the method from Lathi, the pattern of terms in the state matrix becomes clear as the number of capacitors increases.

First, consider resistors in parallel. Define the function $\mathbb{P}_i$

$$\mathbb{P}_i = \left(\sum_{\substack{1 \leq j \leq n \\ j \neq i \\ s_j = Closed}} \frac{1}{R_j} \right)^{-1} \tag{4.2}$$

The function $\mathbb{P}_i$ uses its domain i in an unusual manner. This function is defined for all integers, $i \in \mathbb{Z}$. When the argument i is not in the set of integers 1 through n, it is easy to recognize that, for example, the function $\mathbb{P}_0$ simply returns the equivalent resistance of n resistors in parallel, labeled $R_1 ... R_n$. When the argument i is within the set of integers 1 through n, the function returns the value of the resistors in parallel while dropping the i-th resistor.

For clarity, we shall use the argument $i = 0$, and hence the expression $\mathbb{P}_0$ when the equivalent resistance of the full complement of resistors is calculated in parallel. All other arguments i fed to $\mathbb{P}_i$ will be within the bounds $1 \leq i \leq n$ designating a specific resistor to be dropped. With this definition available, we are now in a position to write the general state-space description of the capacitor bank. With a bank of n capacitors (and their paired parasitic resistances) as depicted in Fig. 4.2 on page 89, the following is the state-space model.

$$\begin{pmatrix} sI_L \\ sV_{C1} \\ sV_{C2} \\ \vdots \\ sV_{Cn} \end{pmatrix} = \mathcal{A} \begin{pmatrix} I_L \\ V_{C1} \\ V_{C2} \\ \vdots \\ V_{Cn} \end{pmatrix} + \begin{pmatrix} \frac{\mathbb{P}_0}{L} \\ \frac{\mathbb{P}_0}{C_1 R_1} \\ \frac{\mathbb{P}_0}{C_2 R_2} \\ \vdots \\ \frac{\mathbb{P}_0}{C_n R_n} \end{pmatrix} I_{in} + \begin{pmatrix} \frac{-1}{L} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} V_{in} \tag{4.3}$$

where the state matrix $\mathcal{A}$ is defined as

$$\mathcal{A} = \begin{pmatrix} \frac{-(R+\mathbb{P}_0)}{L} & \frac{\mathbb{P}_1}{L(R_1+\mathbb{P}_1)} & \frac{\mathbb{P}_2}{L(R_2+\mathbb{P}_2)} & \cdots & \frac{\mathbb{P}_n}{L(R_n+\mathbb{P}_n)} \\ \frac{-\mathbb{P}_0}{C_1R_1} & \frac{-1}{C_1(R_1+\mathbb{P}_1)} & \frac{\mathbb{P}_2}{C_1(R_1(R_2+\mathbb{P}_2))} & \cdots & \frac{\mathbb{P}_n}{C_1(R_1(R_n+\mathbb{P}_n))} \\ \frac{-\mathbb{P}_0}{C_2R_2} & \frac{\mathbb{P}_1}{C_2(R_2(R_1+\mathbb{P}_1))} & \frac{-1}{C_2(R_2+\mathbb{P}_2)} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \frac{\mathbb{P}_n}{C_{n-1}(R_{n-1}(R_n+\mathbb{P}_n))} \\ \frac{-\mathbb{P}_0}{C_nR_n} & \frac{\mathbb{P}_1}{C_n(R_n(R_1+\mathbb{P}_1))} & \cdots & \frac{\mathbb{P}_{n-1}}{C_nR_n(R_{n-1}+\mathbb{P}_{n-1})} & \frac{-1}{C_n(R_n+\mathbb{P}_n)} \end{pmatrix} \tag{4.4}$$

To be utterly unambiguous about the definition of the terms appearing in state matrix $\mathcal{A}$, let us state them again as follows.

Consider a bank of n capacitors (with paired parasitic resistances) as depicted in Fig. 4.2 on page 89. In almost all mathematical texts, counting the rows and columns of a matrix is accomplished using integers starting at 1. Here it is far more convenient to use a zero-based counting system in this case. In this way, the i-th row or column of the matrix designates quantities associated with capacitor C_i or resistor R_i. And the zeroth row or column is associated with the inductor.

Before continuing, it is worth noting that the equations derived have been carried out with enough capacitors that the general solution for n capacitors can be discerned. Let us review the structure of the terms in Eq. (4.4).

Location	Location by index	Term
Upper left corner	$(0,0)$	$\frac{-(R+\mathbb{P}_0)}{L}$
Top row	$(0,1)...(0,j)...(0,n)$	$\frac{\mathbb{P}_j}{L(R_j+\mathbb{P}_j)}$
Left column	$(1,0)...(i,0)...(n,0)$	$\frac{-\mathbb{P}_0}{C_iR_i}$
Diagonal	$(1,1)...(i,i)...(n,n)$	$\frac{-1}{C_i(R_i+\mathbb{P}_i)}$
All others	(i,j) where $i \neq 0, j \neq 0, i \neq$ j	$\frac{\mathbb{P}_j}{C_iR_i(R_j+\mathbb{P}_j)}$

Substituting values into the state-matrix $\mathcal{A}$ will give rise to eigenvalues representing poles in the complex plane. It will be shown that all values induced by the parasitic resistances are larger than the single paired oscillatory poles by a factor of 10^3. Therefore the capacitor bank's performance is sufficiently close to a single ideal capacitor that there is effectively no difference.

All switch combinations have been analyzed. The best representative configuration is with all switches closed, where the poles are located at

$$\begin{aligned}
\lambda_1 &= -5.472\,358\,551\,05 \times 10^{10} \\
\lambda_2 &= -5.346\,500\,728\,81 \times 10^{10} \\
\lambda_3 &= -4.615\,034\,877\,14 \times 10^{10} \\
\lambda_4 &= -1.762\,418\,427\,21 \times 10^{10} \\
\lambda_5 &= -1.675\,204\,418\,46 \times 10^{10} \\
\lambda_6 &= -1.451\,115\,722\,49 \times 10^{10} \\
\lambda_7 &= -3.881\,355\,108\,69 \times 10^{10} \\
\lambda_8 &= -2.719\,903\,761\,1 \times 10^{9} \\
\lambda_9 &= -3.470\,171\,957\,85 \times 10^{4} \; -6.571\,834\,275\,44 \times 10^{6}\mathrm{j} \\
\lambda_{10} &= -3.470\,171\,957\,85 \times 10^{4} \; +6.571\,834\,275\,44 \times 10^{6}\mathrm{j}
\end{aligned}$$

Hence we find that the real poles induced by parasitic resistance occur a thousand times farther from the origin in the left half complex plane. Hence a state space of two dimensions captures the dynamics of the marginal oscillator and the analysis does not require a larger state space. In the simulation section (in Chapter 3) of this dissertation the closed-loop performance of the system with all poles intact was compared to that of the system containing only the dominant complex pole pair.

4.3.2 2nd Analysis: Quality Factor

Consider the schematic of a marginal oscillator given in Fig. 4.2 on page 89. Using superposition, the relationship between the input current I_{in} and output voltage V_0 is determined using KCL on the top rail.

$$I_{in} = \frac{V_0}{Z_L + Z_R} + \sum_{\substack{1 \le i \le n \\ s_i = Closed}} \frac{V_0}{Z_{Ci} + Z_{Ri}} = \frac{V_0}{sL + R} + \sum_{\substack{1 \le i \le n \\ s_i = Closed}} \frac{V_0}{(\frac{1}{sC_i} + R_i)} \quad (4.5)$$

The resulting transfer function is

$$\frac{V_0}{I_{in}} = \left(\frac{1}{sL + R} + \sum_{\substack{1 \le i \le n \\ s_i = Closed}} \frac{1}{(\frac{1}{sC_i} + R_i)} \right)^{-1} \quad (4.6)$$

Also, over the frequency range of interest in NQR, i.e., 0.5 MHz to 5 MHz,

the coil's self-resistance R is not a constant but is reasonably approximated by $R(\omega) = k\sqrt{\omega}$, where the constant k is determined by measuring the self-resistance $R(\omega_0)$ at a predetermined frequency ω_0. In this case, the Schiano Group measured a self-resistance of 1.5 Ω at a frequency of 3 MHz.

The transfer function for each switch combination has a distinct *quality factor*, Q, that is defined as the frequency of the peak divided by the bandwidth of the peak as measured by where the peak falls by 3 dB from its maximum value. Owing to the complexity of the resulting calculation, the mathematics is set up symbolically in the computer to obtain Eq. (4.6) on page 93. Then the component values for L, R, C_i, R_i from the physical circuit are substituted into the this expression, and the quality factor is determined numerically.

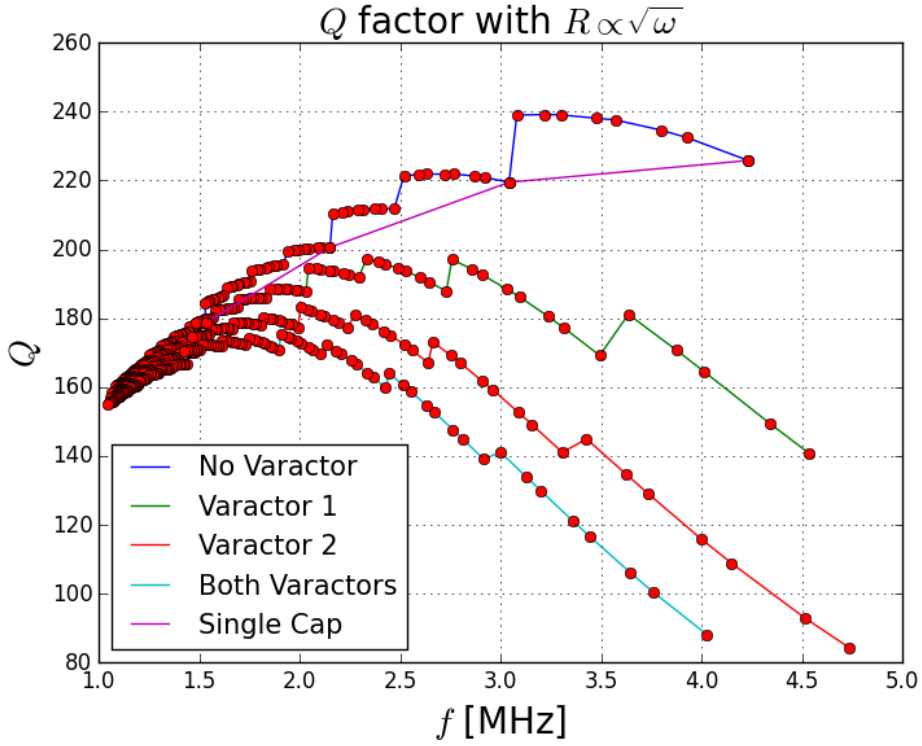


Figure 4.3: Quality factor Q as a function of resonant frequency using the capacitor bank over the nominal range of QR measurements

In the resulting graph shown in Fig. 4.3, there are several patterns that can be observed. Most of these observations are most easily explained when a detailed comparison is made between the graphs and quality factors of the individual capacitors in Table 4.1 on page 90 are noted. The varactors have a lower quality factor than the capacitors; the three capacitors with values

below 60 pF have a quality factor of 3000, while the four larger capacitor values have a quality factor of 1000.

As the two varactors have the lowest quality factor, the quality factor of the circuit as depicted in Fig. 4.3 on page 94 is dominated by their presence in the circuit. Hence the data is easily split into curves based on the varactor's respective switch state. Within each curve, there are arcs (followed by slight jumps) of up to 8 data points as the smallest three capacitors with highest Q values are put into the circuit in a binary sequence proceeding from right to left along the graph. In practice, the frequencies of interest are within a range of 1 MHz to 5 MHz, so only the switch combinations landing within this range are seen in Fig. 4.3 on page 94. The curve labeled as "Single Cap" corresponds to closing a single switch at a time.

In the laboratory, the detection limit appears to be a quality factor of no less than $Q = 100$ [34]. The smallest detectable change in the ratio of the quality factor using a method alternative to varactors would require that the quality factor increase by at least a multiple of 1.3 (i.e. larger than a 15% difference). This potentially does occur at frequencies above 3 MHz. Hence any experimental design should consider changing to a less lossy variable capacitor at frequencies above roughly 3 MHz.

4.3.3 3rd Analysis: Open-Loop Bode Plots

Previous work by the Schiano group has always captured parasitic resistances as adjustments of the estimated value of the self-resistance of the probe coil, R. The following analysis indicates that from the perspective of open-loop performance, this approximation, based on the understanding of the loss mechanisms and singular perturbation theory, is well founded. The Bode plots given in Fig. 4.4 and Fig. 4.5 compare the open-loop performance of two different circuits. The plot in green indicates the performance of the circuit with the full capacitor bank in Fig. 4.2 on page 89 with all parasitic resistances as given earlier. The resulting quality factor of this circuit is $Q = 140$. The second graph in blue uses the ideal circuit given in Fig. 4.1 on page 88 with self-resistance of the probe coil, R, adjusted so that the resulting quality factor matches. In this case the measured value is moved from 1.5 Ω to 3.067 Ω. The Bode plot in Fig. 4.4 on the next page indicates that, near

resonance, the two systems are indistinguishable from one another. The Bode plot in Fig. 4.5 on the following page shows that all effects from the extra zeros and poles used in analysis of the parasitic resistances yield changes in the Bode plots far from resonance.

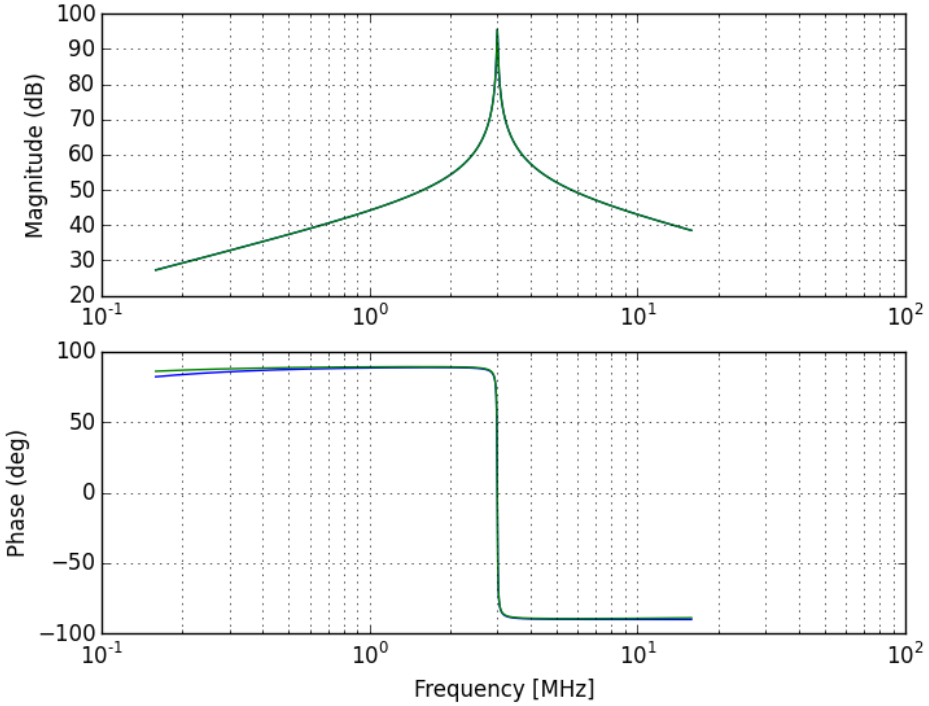


Figure 4.4: Transfer function of the second-order system with adjusted estimate of inductor resistance R to match quality factor Q. The second-order system is in blue, compared to the sixth-order system with parasitics in green.

4.4 Simulation

The results thus far in this chapter indicate that, using the usual linear design tools, the parasitic resistances in the capacitor bank are indistinguishable from the second-order system using a simplified circuit given in Fig. 4.1 on page 88 with adjustments made to the inductor self-resistance R.

Given that the feedback applied to the circuit will be nonlinear, these observations are best characterized as very promising, but not entirely conclusive.

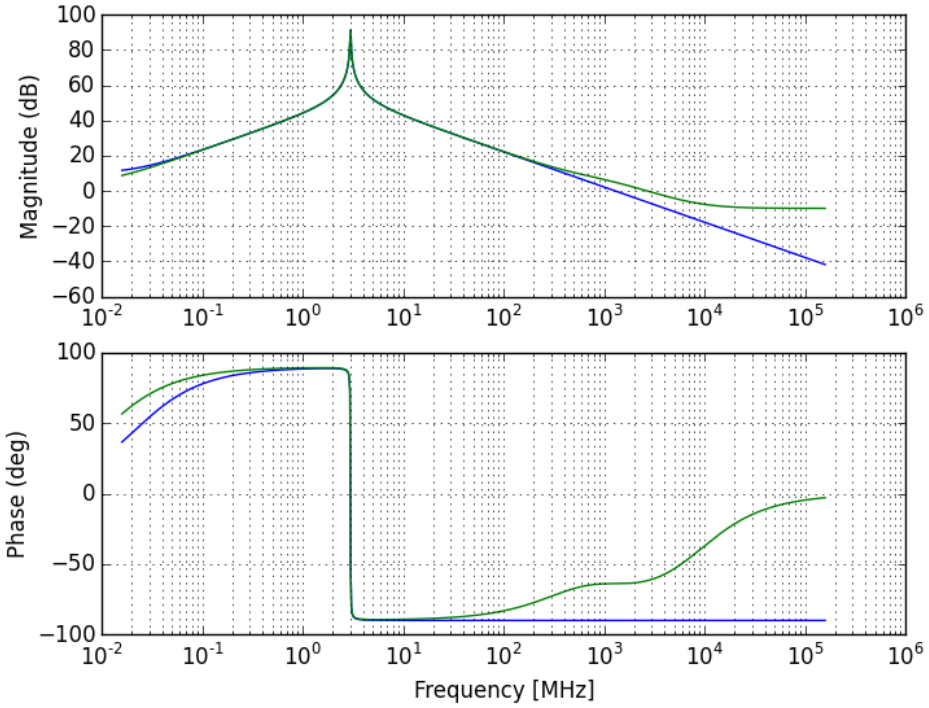


Figure 4.5: Transfer function of the second-order system with adjusted estimate of inductor resistance R to match quality factor Q. The second-order system is in blue, compared to the sixth-order system with parasitics in green. Note how the additional poles and zeros only have an effect far from resonance.

Therefore careful comparison of nonlinear, closed-loop system performance to demonstrate the effect of parasitic resistances is given here.

4.4.1 Estimation of Output Voltage

The simulation results given thus far are trivially adapted to simulations incorporating the larger state space needed. The only significant wrinkle is to realize that the output voltage no longer appears as a single state in the ODE (Eq. (4.3) on page 91). This makes monitoring the output voltage V_0 and applying feedback based on that voltage slightly more complex. Given previous calculations, there are two observations that will help.

First, in the usual sequence of our analysis it is easy to show that

$$I_L = \frac{V_0 - V_{in}}{sL + R} \tag{4.7}$$

Rearranging terms yields

$$V_0 = sLI_L + RI_L + V_{in} \tag{4.8}$$

Following our usual custom, this has to be put into time-domain before nonlinear feedback is applied.

$$v_0(t) = L\frac{d}{dt}i_L(t) + Ri_L(t) + v_{in}(t) \tag{4.9}$$

Second, reading the top line of the state equation Eq. (4.3) on page 91 and multiplying through by L yields

$$\begin{aligned} L\frac{d}{dt}i_L(t) = &-(R + \mathbb{P}_0)i_L(t) \\ &+ \sum_{\substack{1\le i\le n \\ s_i = Closed}} \frac{\mathbb{P}_i}{R_i + \mathbb{P}_i}v_i(t) \\ &+ \mathbb{P}_0 i_{in}(t) - v_{in}(t) \end{aligned} \tag{4.10}$$

Substituting Eq. (4.10) into Eq. (4.9) yields

$$v_0(t) = -\mathbb{P}_0 i_L(t) + \sum_{\substack{1\le i\le n \\ s_i = Closed}} \frac{\mathbb{P}_i}{R_i + \mathbb{P}_i}v_i(t) + \mathbb{P}_0 i_{in}(t) \tag{4.11}$$

We cannot take the last term in this expression for granted because we are applying feedback and the input current, $i_{in}(t)$, is as given earlier

$$i_{in}(t) = -Cv_0(t)\alpha\left(\frac{E(t)}{E_T} - 1\right) \tag{4.12}$$

Substituting Eq. (4.12) into Eq. (4.11) on page 98 and solving for the output voltage, $v_0(t)$, gives

$$v_0(t) = \frac{-\mathbb{P}_0 i_L(t) + \sum\limits_{\substack{1\le i\le n \\ s_i = Closed}} \frac{\mathbb{P}_i}{R_i + \mathbb{P}_i} v_i(t)}{1 + \mathbb{P}_0 C\alpha\left(\frac{E(t)}{E_T} - 1\right)} \tag{4.13}$$

Hence, for any system state, we have a valid estimate of the output voltage.

4.4.2 Equivalent Resistance

For the case outlined above in Section 4.3.3 on page 95 we find that the equivalent static resistance to make the quality factor match the open-loop Bode plots is $R = 3.067$. The simulation measures this resistance at $R = 3.06664$. As stated earlier, there is no point in any finer comparison as error in the numerical integration method dominates anything after 6 significant digits. Therefore, it appears that wrapping all error caused by parasitic resistance into an associated change in the value of static resistance R is appropriate.

CHAPTER 5

CONCLUSION

In this dissertation, the marginal oscillator was described with a state-space model. Two changes of the coordinate system allowed significant computational simplification and increased insight into the design and application of such oscillators. The resulting mathematics is exact and contains no approximations. Also, the mathematical analysis is much easier to pursue in a more general context of using digital electronics in the feedback loop.

The dissertation provides several results that would be much more complicated, or impossible, to achieve using previous methods. These include:

- Showing how to update the circuit model to one that is more physically intuitive
- Updating the circuit model to include a voltage source in addition to the usual current source
- Showing how to implement feedback that uses both the capacitor voltage and inductor current
- Showing how to implement a nonlinear feedback loop, and then redesigning it for easier implementation
- Proving a theorem linking conversion gain and settling time — this had previously only been an observation
- Eliminating phase noise by directly calculating the energy stored in the circuit
- Giving an automated algorithm for obtaining a marginal oscillator measurement in the absence of any estimate of the circuit's internal resistance

- Analyzing the effect of parasitic resistance in the capacitor/varactor switching network

Each of these advances can be used to enhance the performance of CW spectrometers for detecting and quantifying explosives. The results, as given, can markedly increase the speed and sensitivity of such a detector. The results given are also applicable to other problems that have been previously solved using marginal oscillators as given in Section 1.2 on page 4.

5.1 Technical Outcomes

5.1.1 Marginal Oscillator Design

A marginal oscillator is normally a tank circuit with output feedback applied to maximize the change in amplitude with respect to a change in the circuit's internal resistance. Nonlinear feedback is required to accentuate the change in amplitude while preserving the oscillation within a fixed range of amplitude. For historical reasons, marginal oscillators were based on nonlinear circuit elements (FETs and vacuum tubes), and the only correct detailed analysis was given by Viswanathan et al. [9]. What slowly becomes clear in the process of redesigning using state-space methods is that Viswanathan et al. provided a successful description of why the marginal oscillator worked; however, their description was not terribly useful in assisting with the design of new instrumentation.

This dissertation has presented a more complete analysis and redesign of the marginal oscillator based on state-space modeling and full-state feedback. By updating the analysis to modern standards, it is now possible to achieve new goals that were, heretofore, impossible.

The upgraded analysis demonstrates that moving to a more realistic and complex model poses no particular challenge. The new analysis was used to design a simple feedback loop and then redesign the feedback loop for easier implementation.

The upgraded analysis was also used to show that any parasitic resistance in the circuit is indistinguishable from a constant change in the estimate of the idealized tank-circuit's resistance.

The newer circuit model was transformed into a dimensionless coordinate system which induces data collapse. This collapse can be used immediately to improve performance in the presence of thermal noise by reducing the data collection requirements of previous box-car averaging methods. An equivalent description is that the redesign now allows us to measure signal amplitude directly at each data sample. Previous methods attempted to find the envelope of a sinusoidal signal and, hence, generated repeatable phase noise.

Previously, it had been conjectured that conversion-gain appears to be directly proportional to the settling time of the circuit. Under very loose constraints, this conjecture has been proved as a theorem.

An iterative scheme for measurement of the unknown resistance demonstrated that the increased settling time associated with large conversion gain has no impact on the time required to make a measurement. Indeed, as the new model uses more straightforward algebra, each time the system relaxes to a limit cycle, it yields a valid estimate of the unknown resistance. Therefore, the accuracy of the measurement does not require the system to be at a limit cycle with a poor time constant.

5.1.2 Future Applied Research

The redesigned marginal oscillator can be used in any of the previous applications as listed in Section 1.2 on page 4 including: detecting defects in irradiated silicon, measuring skin-depth of superconductors, characterizing thin films or curing of plastics, and measurement of capacitance or temperature. For the purpose of this dissertation, the application emphasized is detection and quantification of magnetic resonance transitions in solid materials, especially existing and newly developed explosives.

The marginal oscillator redesign has helped put the field of CW spectroscopy on a much more sound mathematical and, hence, scientific footing. By reducing the time required to detect and quantify an analyte, here are some of the applications that are now far more likely to be successful.

To summarize, CW spectroscopy excels at detection and identification of nitrogenous compounds. The molecular structure surrounding the nitrogen atoms changes the frequency of resonance. This means the technique can

discriminate between isomers of the same chemical formula. It can do so at much lower power and cost than pulsed techniques.

Much recent world-wide research has been directed at developing new munitions that are less sensitive to storage conditions and rough handling. Almost all of these compounds are nitrogen based and should be readily detected by CW-NQR equipment.

CW equipment uses far less energy and is, therefore, far more amenable to applications such as shoe scanners. As the radiation required is non-ionizing and sufficiently low power, exposure to a human body is low risk. Hence, shoe removal for such a scanner may be unnecessary.

Computational chemistry has advanced to the point where compounds of interest can be constructed *in silico* and analyzed for their likely magnetic transition frequencies. In part, this dissertation was motivated by discussions with F. C. Hill [38] regarding the feasibility of improving the development cycle of *a priori* prediction of magnetic transition frequencies followed by rapid lab validation of such predictions.

Finally, new nitrogen-bearing compounds, energetic or not, can be studied for their fate-and-transport either in the lab, or in the environment. Current research is directed at understanding if such compounds may adhere to surfaces in some preferred orientation. If so, the preferred orientation may cause the compound's magnetic transition frequencies to shift in a predictable manner. Laboratory-based experiments may be able to measure these characteristic changes. If there are preferred orientations of adhesion, then fate and transport in the environment may be far more predictable than using current heuristic and laboratory-based techniques.

5.1.3 Future Theoretical Research

This dissertation does raise certain points that deserve consideration for follow-up work.

1. Using the idea of sampling each data point in dimensionless coordinates, resolution of the data can now be improved using techniques originally developed for astronomy. Astronomers use the slightly smeared image of a star over a CCD array to obtain an estimate of the star's position at *sub-pixel accuracy*. Typical results yield a measurement with a

resolution approximately 1/12 of the pixel widths [39]. This is akin to obtaining an extra 3 to 4 bits resolution for a measurement without requiring any change to the pre-existing digital-to-analog converters. This is achieved by sampling over a longer span of limit cycles, and trades away some of the time efficiency gained.

2. If the strategy of driving the residual error signal to zero is not necessarily optimal, is there a viable alternative that optimizes speed and accuracy?

3. Are the feedback equations now sufficiently simple that they can be implemented in analog hardware?

4. The system simulations should now be upgraded to include the type of thermal noise observed in the laboratory setting to take advantage of the new noise-reducing schemes based on data collapse in normalized energy units.

APPENDIX A

BIOPROTECTION OF FACILITIES

A.1 Summary

The anthrax attacks of 2001 energized research directed toward reducing health consequences from airborne contaminants by augmenting current heating, ventilation, and air-conditioning (HVAC) systems. Even during peacetime, interest will continue in improving HVAC components to reduce biocontaminants associated with sick building syndrome. Current HVAC design uses numerical simulation methods of ordinary differential equations to predict approximate performance. We show that state-space, Laplace transform calculations actually solve the underlying differential equations and yield algebraic expressions that provide new insight. To sharpen the arguments in favor of this methodology, attention is restricted to improving existing HVAC systems to increase protection from an external release of hazardous particulates. By nearly eliminating the need for dynamical simulation, the resulting methods can be applied to far more complex HVAC designs with little additional computational effort. The new methods reduce the time required for computation by 3 orders of magnitude. These algebraic methods also can be extended to disparate technical problems including internal particulate release, gas masks, and designing new protective buildings.

A.2 Objective

The design of HVAC systems to resist biological contaminants presents a combinatorial explosion of different technical problems. These problems are far more numerous than can be addressed in a single article. To focus attention on the new methodology, the problems addressed here are limited

in several ways. The deeply interested reader will find it easy to expand the solutions presented here for similar problems in the resulting combinatorial problem space. The central concepts pointed out in this exposition will recur for most other scenarios, e.g., internal release of contaminants. The authors believe that many aspects of this approach are new. W. R. Ott previously made the connection to transform methods addressing smoke moving from room to room in the absence of an HVAC system [7]. Other related articles describing heating, cooling, and indoor air quality are available [40–42].

Our objective is to present a way to protect an existing, simple structure from external release of hazardous aerosolized particles, via changes to the HVAC system. As a metric, we choose the following: In the event of an external release of contaminants, is there a way to decrease by 80% the area in which buildings will have impacted occupants compared to buildings with current standard HVAC designs?

Again, these methods are trivially extended to changing HVAC design for protection against internal release of airborne contaminants. The results also are easily extended to designing critical infrastructure where far more expensive methods of air purification are available to protect building occupants [43, 44].

A.3 Approach

The approach begins with an extremely simple, idealized model structure and slowly adds complexity to gain additional insight into how such buildings can be protected against contaminants. Candidate methods required to redesign a structure's HVAC system and the likely cost of such improvements are then explored.

A.3.1 Model Simplification

To reduce the combinatorial explosion of possibilities, it is now important state what simplifications are being made. These simplifications will each be pessimistic in the sense that they will each increase the estimate of how much particulate matter will be inhaled by a building occupant.

First, we neglect transport terms describing the air ventilation ducts. Most

buildings are designed so that the HVAC system circulates air at about five changes per hour, so the dwell time within the ducts can be thought of as negligible. Second, we will use the *well mixed* assumption, where particulate matter is not allowed to be sequestered by anything other than the HVAC system filters. In other words, there is no deposition on building contents: floors, ducts, furniture, vents, etc.

Last, we will restrict our calculations to cover the scenario of *external release*, where a plume of particulate matter is generated externally to the building and taken into the building at the HVAC intake or exterior doors. This assumption both takes away and then adds complexity. The simplification is that particulate concentration in external air is a scalar quantity. The increased complexity is that the overall problem has to be solved in two steps: transport from the outdoor release point to the structure, and then transport from the HVAC intake and exterior doors to the building occupants. We begin with the building model first.

A.4 Modeling Aerosolized Particle Transport

The objective of this section is to derive algebraic solutions to the problem of describing the mathematical origins of *protection factor*. Protection factor is defined as the asymptotic ratio of outdoor to indoor air concentration of particulate matter when the outdoor air is held at a fixed contaminant concentration. Moreover, protection factor appears when calculating the dose imparted to building occupants exposed to more complex dynamics of outdoor air concentration. When analyzing the results of such complex dynamics, it will be shown that these dynamics can simply be integrated out. In other words, all questions regarding human exposure can be answered by suitable modifications to the state-space variables and determining how they behave asymptotically. If a later, detailed study of the dynamics is warranted, the exact solution to the ordinary differential equations involved is straightforward.

A.4.1 Transform Methods Required

State-space notation

Consider that an HVAC system can be approximated by an ordinary differential equation in the standard state-space form.

$$\begin{aligned} \tfrac{d}{dt}\vec{x}(t) &= \overleftrightarrow{F}\vec{x}(t) + \overleftrightarrow{G}\vec{u}(t) \\ \vec{y}(t) &= \overleftrightarrow{H}\vec{x}(t) \end{aligned} \tag{A.1}$$

For simplicity, this neglects transport lags and a number of other effects. For clarity, the vector notation arrows are dropped to keep clutter to a minimum.

Using a single-sided Laplace transform with initial conditions set to zero, these expressions appear in the following standard form.

$$\begin{aligned} sX(s) &= FX(s) + GU(s) \\ Y(s) &= HX(s) \end{aligned} \tag{A.2}$$

Solving for X in the first line and substituting into the second line, the solution in the transform space is also a standard result.

$$Y(s) = H(s\mathbb{I} - F)^{-1}GU(s) \tag{A.3}$$

The vector $X(s)$ will be organized to contain the state-space variables of interest. $U(s)$ will model the exogenous concentration of particulate matter produced either: externally and drawn in at the HVAC intake, or from internal release points within the structure.

The Final Value Theorem

The final value theorem is a shortcut that gives important information in the time domain by inspection of algebraic expressions in the transform space. Suppose there is a need to know the asymptotic value of a function $y(t)$ as $t \to \infty$, and only the Laplace transform $Y(s)$ is known.

$$\lim_{t\to\infty} y(t) = \lim_{s\to 0} sY(s)$$

The above equations are now used to gain an understanding of a bio-

protected facility.

A.4.2 Idealized Building With HVAC

Start with an idealized building as depicted in Fig. A.1. The following equations hold for a single interior space with no entryways. The point of this calculation is to yield an ideal case. All other calculations with a door or a vestibule (air-curtain fortified entryway) can later be compared to this ideal case.

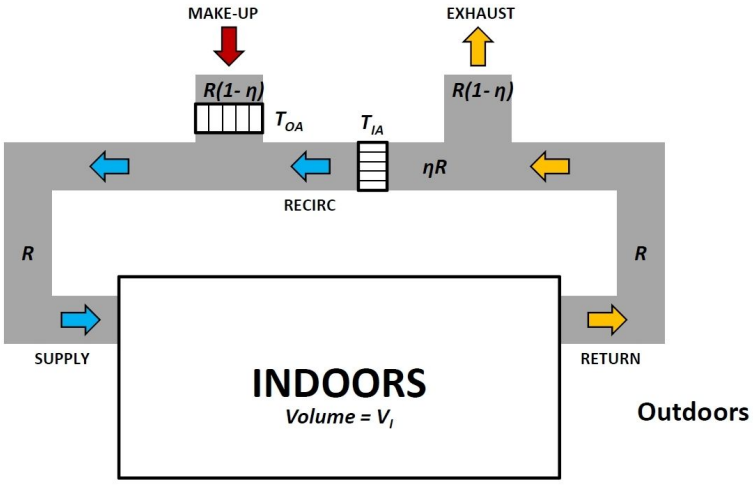


Figure A.1: Schematic of idealized building (no doors or windows or other leakage) with its HVAC system

For purposes of capturing the correct terms in our model, there are some traditional definitions that are more easily understood if expressed in an altered way. For instance, consider the notion of *filter efficiency* defined as the fraction of particles that become trapped in a filter. From the standpoint of ordinary differential equation (ODE) modeling, there is more conceptual clarity in using *filter transmittance*, defined as the fraction of particulate matter that goes through the filter without becoming trapped. Define this relation letting filter efficiency be 1 minus the filter transmittance. Let:

$R =$ Rate of ventilation in the building

$\eta =$ fraction of air recirculated, implies $(1 - \eta)$ is the fraction of "make-up air"

$T_{OA} =$ Transmittance of the outdoor air filter at a fixed particle size

$T_{IA} =$ Transmittance of the indoor air filter at a fixed particle size

$A_I =$ The number of particles inside the building

$L_I =$ The number of particles deposited in the lungs of a person placed inside the building

$\rho =$ The respiration rate of the human

$V_I =$ The volume of the building's interior

The ODE representation of Fig. A.1 on page 109 is then given in Eq. (A.4)

$$\begin{pmatrix} \dot{A}_I \\ \dot{L}_I \end{pmatrix} = \begin{pmatrix} a & 0 \\ b & 0 \end{pmatrix} \begin{pmatrix} A_I \\ L_I \end{pmatrix} + \begin{pmatrix} \gamma \\ 0 \end{pmatrix} u(t) \tag{A.4}$$

The constants in this equation, a, b, γ, will be expanded. For the moment it is known that:

a is a constant describing the processing of recirculating air

b is a constant summarizing rate of respiration of a person present inside the building

γ is a constant describing the number of particles per second introduced to the building from make-up air

At first, it might seem that the choice of number of particles (as opposed to concentration) is somewhat unusual. Having worked with these equations, we have found that there is significant advantage in our choice – when each quantity in the state space is measured in the same units, the constants can be readily inspected for dimensional consistency.

For the moment, several choices of H are easily constructed because the interesting quantities are directly related to the individual terms in the state-space vector. For example, if the calculation requires particle concentration in the room then $H = (1/V_I, 0)$. To calculate the number of particles deposited in the human's lungs, set $H = (0, 1)$.

Now expand the constants within the ODE:

$$a = -\frac{R}{V_I} + \frac{\eta R}{V_I} T_{IA}$$

$$\gamma = R(1-\eta)T_{OA}$$

$$b = \frac{\rho}{V_I}$$

These expanded constants are derived by careful reference to Fig. A.1 on page 109. Because the inhalation rate ρ is tiny compared to the flow rate R (as given in Appendix A.4.5 on page 116), it is not required to compose additional terms signifying that the person also exhales nearly particulate-free air.

Deposition of particles into the lungs of a human inside the building

Now construct $u(t)$ that captures arbitrarily complex dynamics of any particulate plume that has finite temporal duration or at least asymptotically goes to zero as $t \to \infty$. Let $\phi(t)$ possess a valid Laplace transform, and be positive over values of time larger than zero. Then let $u(t) = D\phi(t)$ so that $u(t)$ constitutes an input profile of outdoor air concentration of particulate matter where, over the course of time, D particles get into the intake vent of the HVAC system.

The asymptote describing particle deposition in the lungs of a human inside the building is expressed in Eq. (A.5)

$$\lim_{t\to\infty} L_I(t) = -Db\frac{\gamma}{a} = \rho D\frac{T_{OA}(1-\eta)}{1-\eta T_{IA}} \tag{A.5}$$

Notice that the right-most expression in Eq. (A.5) is a product of three terms. Respiration rate ρ, and number of bioactive particles drawn into the HVAC intake D, are exogenous to the building design. The third (last) term only contains variables that are set by the design of the HVAC system. In the next section, it is shown that this third term is the reciprocal of the protection factor.

Steady state under constant contamination - Protection Factor

Protection factor can be looked at directly by using its definition. Given that the outside of the building is held at a fixed concentration of particulates, find the asymptotic ratio of outside concentration to inside concentration. Here, let $u(t) = C$. In our coordinate system, A_I is the number of particles

inside the building; therefore, the concentration of particulate matter in the indoor air is A_I/V_I. This assumption, called the *well mixed assumption*, is common in such modeling efforts and expressed here as Eq. (A.6).

$$\lim_{t\to\infty} \frac{A_I(t)}{V_I} = \frac{-C\gamma}{V_I a} = C\frac{T_{OA}(1-\eta)}{1-\eta T_{IA}} \tag{A.6}$$

From this it is immediately shown that the expression $\frac{T_{OA}(1-\eta)}{1-\eta T_{IA}}$ is of central importance. It is also the $-\gamma/(aV_I)$ term of the middle expression in Eq. (A.5) on page 111.

Now solve for the protection factor as the ratio of outdoor concentration to indoor concentration:

$$\lim_{t\to\infty} C\left(\frac{A_I(t)}{V_I}\right)^{-1} = \frac{1-\eta T_{IA}}{T_{OA}(1-\eta)}$$

Protection Factor as a central metric

As can be easily appreciated by comparing the Eqs. (A.5) and (A.6) on page 111 and on the current page, the reciprocal protection factor is seen both in analysis of events of finite duration with complex dynamics and in those of infinite duration with no dynamics. Hence it is *the* centrally important metric of the quality of bio-protection.

A.4.3 Building With Leakage Due to a Door

Now the model building becomes more realistic by incorporating a door as depicted in Fig. A.2 on the following page. As in previous examples, the strategy is to integrate out any detailed dynamics that generate the leakage current α. For purposes of later analysis, we will need to know the number of door cycles per minute, and will assume that in each door cycle a specific volume of air will intermix between the adjacent air volumes, i.e., a fixed volume of outdoor air and an equal volume of indoor air will be mixed and exchanged. Including the leakage rate, Eq. (A.4) on page 110 is modified in the following way:

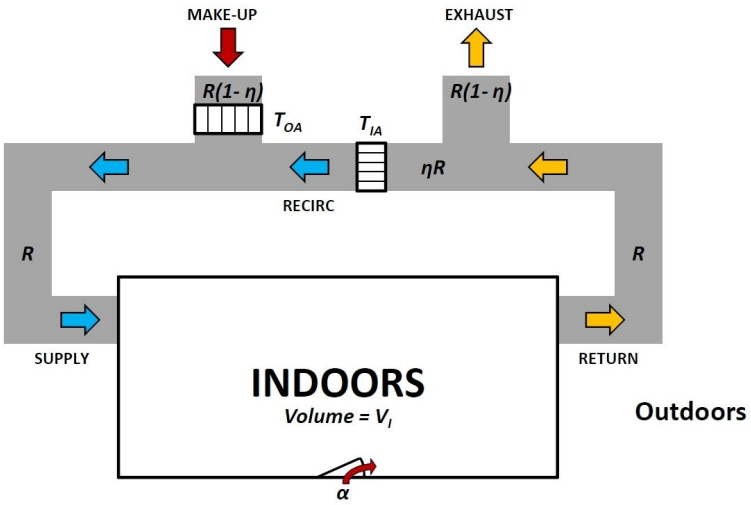


Figure A.2: Building plus HVAC system with outside door and associated leakage rate α

$$\begin{pmatrix} \dot{A}_I \\ \dot{L}_I \end{pmatrix} = \begin{pmatrix} a-d & 0 \\ b & 0 \end{pmatrix} \begin{pmatrix} A_I \\ L_I \end{pmatrix} + \begin{pmatrix} \gamma+\alpha \\ 0 \end{pmatrix} u(t)$$

Here all variables are defined as before with α representing the leakage rate from outdoor to indoor air through the door, and d being the inverse time per particle escaping from the indoors to outdoors through the door, hence, $d = \alpha/V_I$

So the transfer function to indoor air concentration becomes

$$\lim_{t\to\infty} \frac{A_I(t)}{V_I} = \frac{-(\gamma+\alpha)}{V_I(a-d)} = \frac{T_{OA}(1-\eta)+\frac{\alpha}{R}}{1-\eta T_{IA}+\alpha}$$

Comparing this result with Eq. (A.6) on page 112 one can observe that as $\alpha \to 0$ the protection factor is as before. Further, at sufficiently small leakage rates, $\alpha \ll R$ and $\alpha \ll 1-\eta T_{IA}$, one retains the high protection factor evidenced in the ideal case of Eq. (A.6) on page 112. These results are graphed in Appendix A.4.5 on page 116.

A.4.4 Facility With Leakage Due to a Vestibule

Here the situation is as depicted in Fig. A.3 on the next page, a building with a vestibule inside the outer entry door. The definitions of a, b, c, and γ have

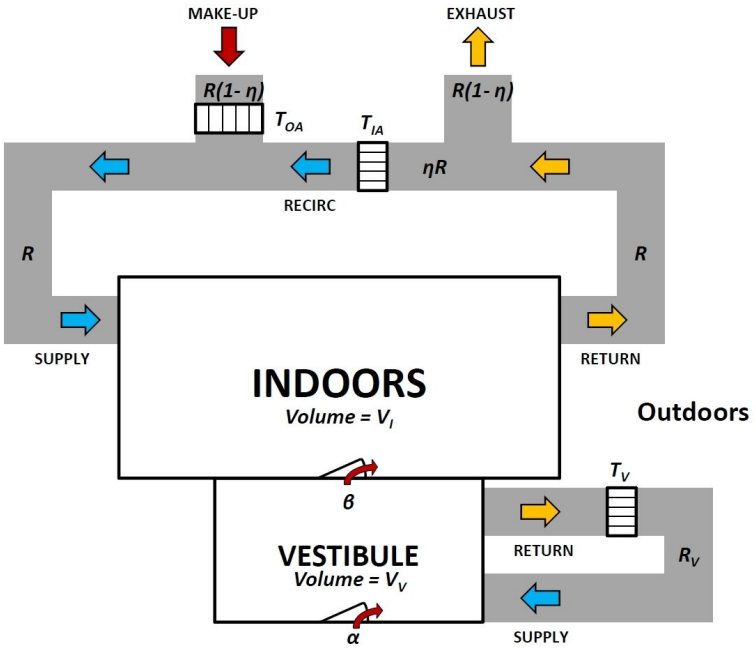


Figure A.3: Building plus HVAC system with entryway vestibule

not changed. Assume that the aggregate leakage current through the outside door α remains the same. The leakage current through the inner door of the vestibule is named β as we expect to slightly pressurize the building while leaving the vestibule at ambient outdoor air pressure. (Detailed justification for this pressurization scheme is given in reference [45] and discussed further in Appendix A.5 on page 121.) Hence, it is expected that $\beta < \alpha$.

Our ordinary differential equation now takes the form:

$$\begin{pmatrix} \dot{A} \\ \dot{L_I} \\ \dot{V} \end{pmatrix} = \begin{pmatrix} a-g & 0 & f \\ b & 0 & 0 \\ g & 0 & c-f-h \end{pmatrix} \begin{pmatrix} A \\ L_I \\ V \end{pmatrix} + \begin{pmatrix} \gamma \\ 0 \\ \alpha \end{pmatrix} u(t)$$

Again, the variables in the matrix describe the following quantities:

a describes the processing of recirculated air

g describes the leakage β taking air from the interior into the vestibule

c is a term analogous to a describing the HVAC system cleaning air from the vestibule

h is a term describing the leakage α exchanging air between the building

exterior and the vestibule

b is related to the rate of respiration of a person present inside the building

f is related to exchange rate from the vestibule to the interior

γ is a constant describing number of particles per second introduced to the room from make-up air

Analogously, H takes on different values depending on the quantity of interest: concentration in the room requires $H = (1/V_I, 0, 0)$, number of particles deposited in the lung requires $H = (0, 1, 0)$, and concentration of particles in the vestibule requires $H = (0, 0, 1/V_V)$.

By working analogously with the other cases one obtains an expression for the asymptotic concentration of particulates in the building interior when the outside air is at a fixed concentration:

$$\lim_{t\to\infty} \frac{A_I(t)}{V_I} = \frac{\gamma - \frac{f\alpha + f\gamma + h\gamma}{c}}{V_I(-a + g + \frac{-gh + af + ah}{c})}$$

where

$$a = -\frac{R}{V_I} + \frac{\eta R}{V_I} T_{IA} = \frac{-R}{V_I}(1 - \eta T_{IA})$$

$$\gamma = R(1 - \eta) T_{OA}$$

$$c = -\frac{R_V}{V_V} + \frac{R_V}{V_V} T_V = -\frac{R_V}{V_V}(1 - T_V)$$

$$b = \rho / V_I$$

$$f = \beta / V_V$$

$$g = \beta / V_I$$

$$h = \alpha / V_V$$

Notice that the terms are arranged for the indoor air concentration to show that the expression will go to the ideal condition of Eq. (A.6) on page 112 as α and β go to zero. In other words, the ideal protection factor is still obtained when the leakage rates are zero.

A.4.5 Resulting Graphs and Analysis

It is important to recognize that the method developed here has obtained closed-form solutions that describe both the protection factor and the amount of particulate matter that will accrete in the lungs of a building occupant. In this way the entire calculation can easily be extended to particles of all sizes given charts of the filter efficiency for the air filters T_{OA}, T_{IA}, T_V as a function of particle size as seen in Fig. A.6 on page 120. Further, these calculations apply to buildings of any size. These results are purely algebraic and can easily be obtained without having to run a simulation. If detailed analysis of the dynamics is later required, the Laplace transform has already been calculated, so the state variables can also be expressed exactly as a function of time without the approximations yielded by many runs of a numerical simulation.

The graphs from these three building configurations, with varying parameters of interest, are now exhibited and discussed. In Fig. A.4 on page 118 and Fig. A.5 on page 119 one can observe the results of these equations by using the values that are representative of a notional building approximately the size of a test-bed our group plans to use for verification and validation. The following is a list of our default values (all were carefully changed to SI units internal to the calculation):

$T_{OA} = T_{IA} = T_V \simeq 0.0038$ for HEPA filter and $\simeq 0.38$ for MERV-8 filter

$\eta = 0.8$

$\alpha = \beta = 1/30\,\mathrm{m}^3\,\mathrm{s}^{-1}$

$R = 1000\,\mathrm{ft}^3/\,\mathrm{min}$

$R_V = 200\,\mathrm{ft}^3/\,\mathrm{min}$

$V_V = 288\,\mathrm{ft}^3$

$V_I = 8000\,\mathrm{ft}^3$

$\rho = 14.5\,\mathrm{L}/\,\mathrm{min}$

For this analysis, let the door be opened and closed once per minute, and in each door cycle let $2\,\mathrm{m}^3$ of air intermix between the adjacent air volumes,

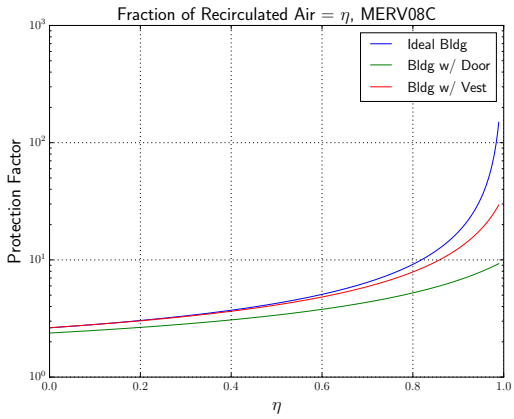
i.e., $1\,\mathrm{m}^3$ of outdoor air and $1\,\mathrm{m}^3$ of indoor air. Hence, the values given for constants α and β.

Figure A.4a on the next page shows the results for this notional building as the fraction of recirculated air is allowed to move over its full range from $\eta = 0$ to $\eta = 1$. The important thing to observe here is that as the idealized building with no doors or windows (blue curve) goes to completely using recirculated air, the protection factor will become arbitrarily large. Setting $\eta = 1$ is somewhat unrealistic as the building occupants would suffer from poor indoor air quality due to lack of any fresh air. Further, in many types of construction (particularly residential) the building's air leakage is the only source of fresh air, and hence cannot be shut off completely. However, the concept of boosting the circulation rate to high levels through a poor quality air filter and allowing η to be set to levels much closer to 1 might also seem to be a viable strategy depending on the lifetime economic costs. This strategy is discredited in Appendix A.6 on page 123.

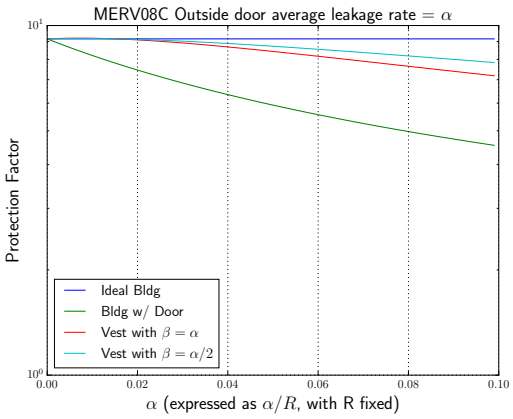
In Fig. A.4b on the next page one can observe the results of allowing the leakage rates α and β to range over small values (normalized to the ventilation rate R). Recall that it is expected that α and β are about $0.033\,\mathrm{m}^3/\,\mathrm{sec}$. In this graph one observes the reason a vestibule is important. The curves describing protection factor with a vestibule are nearly flat when α and β are small. In this way the building is protected far better than using a standard door, as seen from the green curve.

Figure A.4c on the following page shows the effect of increasing the vestibule air recirculation rate R_V over the range from 0 to the building's overall fan rate, R. Clearly, there may be an economic balancing act between the filter's transmittance (or efficiency) and the increased lifetime costs associated with electrical power and filter maintenance. In Appendix A.6 on page 123, however, it is shown that the electrical costs are negligible and that the filter maintenance cost is reasonable.

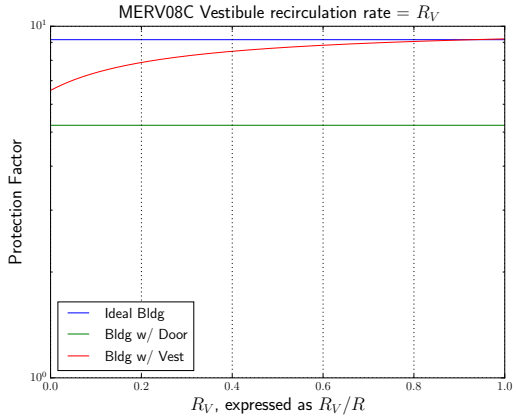
So far, all these observations relate to the use of MERV-8 filters as depicted in Fig. A.4 on the following page. By contrast, changing to HEPA filtration in Fig. A.5 on page 119 yields the same observations as the qualitative shape of the resulting graphs remains intact. Hence, these figures show that the observations made above are qualitatively generic and independent of the filter type.



(a) MERV-8 performance with varying recirculating air fraction η

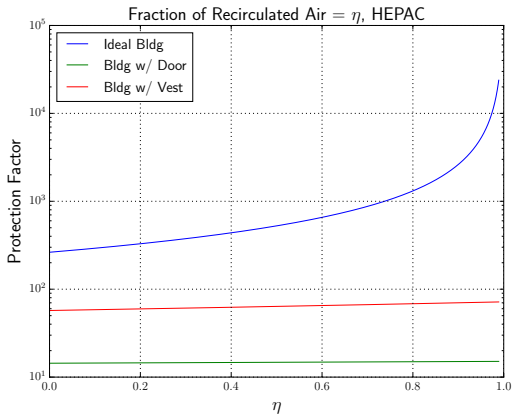


(b) MERV-8 performance as a function of door leakage rates α and β

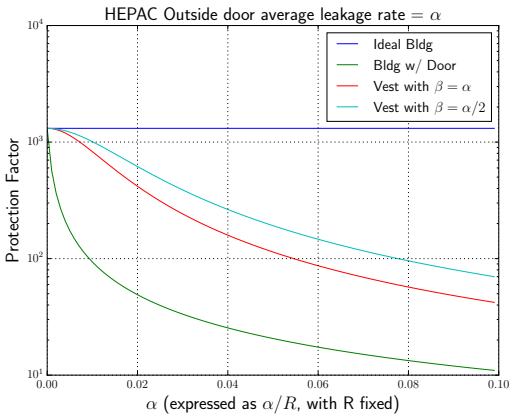


(c) MERV-8 performance as a function of air recirculation rate in the vestibule R_V

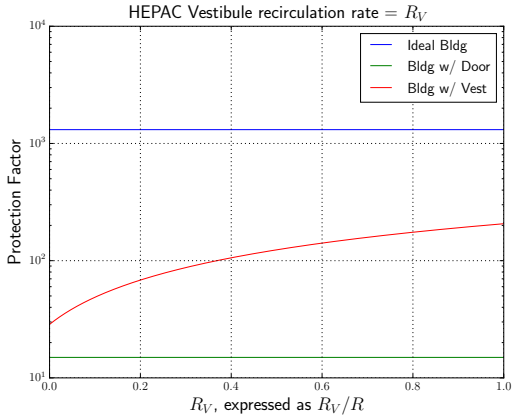
Figure A.4: Protection factors with MERV-8 filters



(a) HEPA performance with varying recirculating air fraction η



(b) HEPA performance as a function of door leakage rates α and β



(c) HEPA performance as a function of air recirculation rate in the vestibule R_V

Figure A.5: Protection factors with HEPA filters

A.4.6 Library of Filter Transmittances

All calculations to this point have been performed without numerical simulation. Hence, any desired quantity can be calculated trivially in software. This allows the capability to obtain more comprehensive results without significantly adding to the computational effort required. For a facility to be protected from particulates of varying sizes, it would be useful to expand each of our transmittance terms so that they are each functions of particle size. Using sources from the open literature [46, 47] we now have a primitive library of exactly this type. A typical entry is shown in Fig. A.6. After trying several styles of curve interpolation, it was decided that linear interpolation made the most sense. The resulting curve is quite acceptable for the current set of calculations.

It is also important to note that the filter library can easily be expanded to take into account that various filtration schemes may also vary in efficiency (and hence, transmittance) by face velocity and loading. The additional information could be easily incorporated into the filter library with negligible additional computational effort.

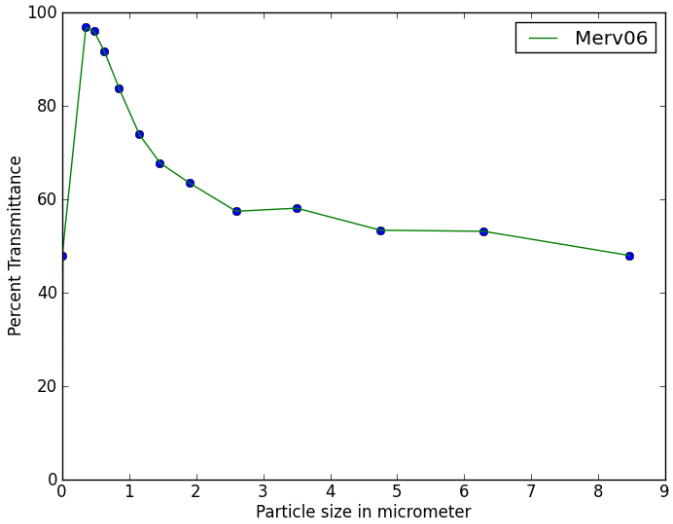


Figure A.6: Library entry showing filter transmittance and linear interpolation as a function of particle size for a MERV-6 filter

A.5 Plume Model

The question now arises, how many particles are available at the HVAC intake and exterior doors as a function of the release point of the particles, the placement of the building, and wind conditions? In Appendix A.4 on page 107, the calculations have been very successful in obtaining meaningful measures of building performance and human protection using ODEs to integrate out the detailed dynamics. Therefore, the same tactic has been used in our plume modeling work.

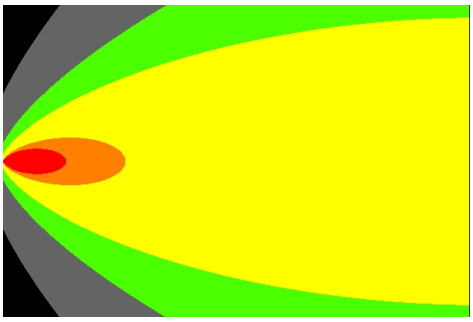
We implemented EPA Industrial Source Complex (ISC3) dispersion model as a dynamical simulation, and simply summed the particle concentrations available to the building's intake vents at each moment in time. This yields a library of diagrams for all stability classes (A - F). Hence for each stability class, there is a single static image summarizing how many particles are available to the building's HVAC system. As the HVAC system is a time-invariant linear system, and given that superposition holds, the resulting analysis summarizes how many particles would have been inhaled by a building occupant over the entire course of the simulation. Using the analogous method, the number of particles taken into the building's interior is easily calculated.

Figure A.7 on the following page shows a representative result from our simulations. In each picture, the release point is centered on the left-hand edge with the wind blowing to the right. The raw result of the simulation shows how many particles are available to be drawn into the building as a function of where the building sits on the picture. Then we can apply a protection factor consistent with the results of the previous section to calculate the number of particles deposited in the lungs of a building occupant.

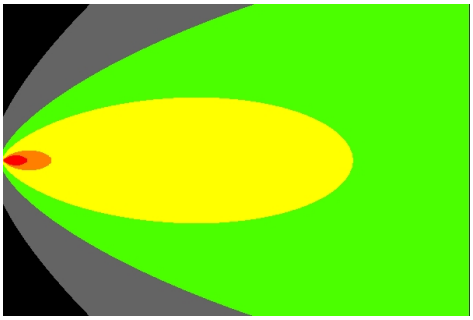
Using ISC3's stability class C as a baseline, the dynamics are integrated out of the simulation. Then a protection factor of 9 is applied (consistent with the MERV-8 results given above) to obtain Fig. A.7a on the next page. As MERV-8 filters are the most commonly used filters, this serves as a baseline to determine if a changed filtering scheme meets the goal of reducing by 80% the affected area from such a plume. Figure A.7b on the following page shows that such a scheme exists. Here all three MERV-8 filters are replaced with MERV-15 filters. The resulting graph does shrink the affected area by the required 80%. Hence, we have just shown that a viable solution exists without requiring full HEPA filtration. For completeness, Fig. A.7c on the next page

shows that if MERV-8 filters are replaced with HEPA filters, then the HVAC system far exceeds the requirements of the project by reducing the affected area by 99.94%.

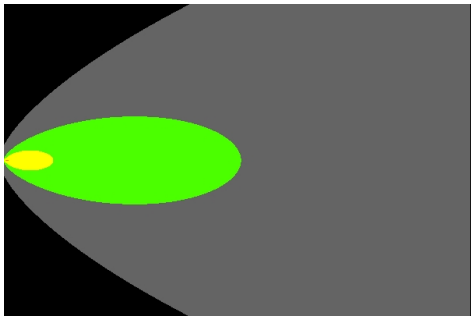
To re-state for clarity, Fig. A.7 is a measure of how many particles are available inside a structure for inhalation by a building occupant when the building's intake vent was situated at a particular point on the chart. Because the inhalation rate is only 14.5 L/ min, the total exposure would be multiplied by 0.0145 m^3/ min.



(a) Plume Stability Class C. MERV-8 filtration, PF=9, serves as a baseline against competing solutions.



(b) Plume Stability Class C. MERV-15 filtration, PF=45, shrinks the affected area by 80% compared to MERV-8 filtration



(c) Plume Stability Class C. HEPA filtration, PF=1400, shrinks the affected area by 99.94% compared to MERV-8

Figure A.7: Plume Models from ISC3 integrated over the complete life of the plume. Color coded: red > 143000 particles, orange > 45000 particles, yellow > 1400 particles, green > 90 particles, gray > 1 particle. Each diagram is 2 km vertically and 3 km horizontally.

An additional concern is whether the structure's walls have to be essentially resistant to penetration from particulate matter driven by wind. Recent results [45, 48] have shown that very modest over-pressure of a structure (5 Pa) leads to Peclet numbers that indicate there will be no penetration from

outdoor particulate flows in wind lower than 20 mph (8.94 m/s). Above such a wind speed, however, the effectiveness of outdoor air contamination is highly debatable. As an example, the ISC3 only covers wind speeds below 6 m/s (13.4 mph).

The analysis of protection factor is now quite complete. The calculations listed above allow us to try many strategies using any combination of air filters, recirculation rates, and recirculation fraction; they can carry out the calculations for any particle size without requiring a dynamical simulation. Further, the terms of the ODEs can be arranged to mimic any particular configuration of HVAC system within a building of interest.

A.6 Impact on Energy and Life-Cycle Cost

With these algebraic tools and the filter library, one can reduce by 80% the area in which building occupants would be affected, without requiring HEPA filtration. Further MERV-15 filtration implemented in current buildings is adequate. To enhance field acceptance, we now look for ways that the increased quality of air filtration can decrease life-cycle costs.

A.6.1 An Active Control Scheme to Reduce Energy Consumption

Here we present a quick calculation demonstrating that a MERV-14 filter is usable with active controls. As the goal is to achieve a protection factor of 45, it would be desirable to understand if controlling the recirculation fraction η might allow us to use a smaller MERV number and still achieve the required protection factor. In Fig. A.8 on the next page the given result shows how the protection factor would change with respect to recirculated air fraction η. For reference there is a flat line plotted at the expected goal of a protection factor of 45. The uppermost line indicates that current HEPA filtration far exceeds the goal over the entire range. MERV-16 and MERV-15 filters are both capable of meeting the goal at a recirculated air fraction $\eta = 0.8$. It is not unreasonable to expect that the Department of Defense (DOD) could arrange to have louvers placed within the building that would be capable of pushing recirculated air fraction η to much higher levels on demand. Such automated

adjustments are possible as part of commercial-off-the-shelf (COTS) "enthalpy switchover" now commonly used to save energy in buildings. In the worst case the DOD would require that this switchover equipment be able to go to higher values of η than the commercial sector requires.

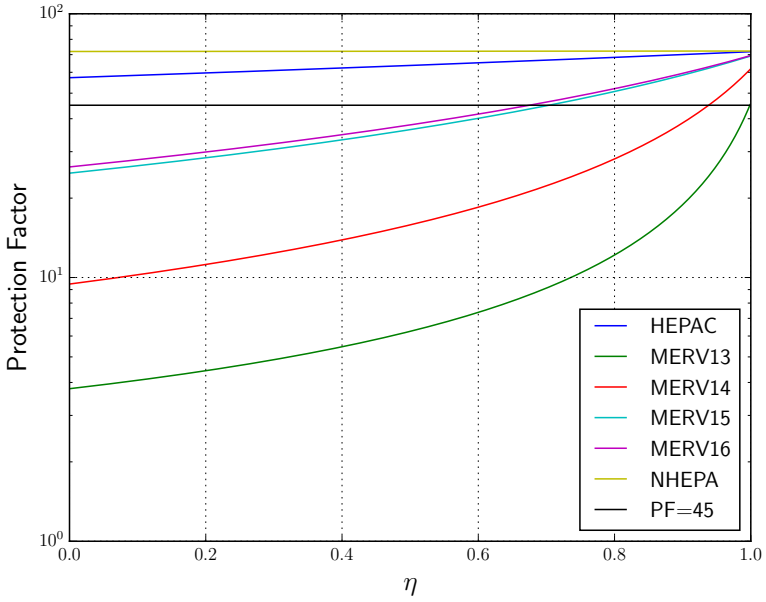


Figure A.8: Increasing the protection factor by changing the recirculation fraction. The line placed at a protection factor of 45 is for reference.

A.6.2 Is Any Control Scheme Required at All?

Surprisingly, it is premature to jump to the conclusion that MERV-15 air filters impose increased energy costs. In this theoretical exposition, the filter's efficiency and maximal flow rates are interpreted as independent variables. However, commercial filters are developed using different assumptions. The filters are produced and then their flow rate and efficiency are measured at a predetermined pressure drop. This fixed value of pressure drop keeps the system's blower motor at a specific spot on the fan curve. Notice that this means the pressure drop is the independent variable with both filter capacity (flow at fixed pressure drop) and filter efficiency as dependent variables. In the real world, whole families of filters subtending multiple MERV ratings all have the same pressure drop. Military specified HEPA filters are the exception. Mil-spec HEPA filters are designed for efficiency and compactness

with the pressure drop held as the dependent variable. Hence, among MERV rated filters, a change in filter efficiency may cause an increase in filter size, but does not impact energy efficiency at all.

Therefore, the greater efficiency of moving to a MERV-15 filter (from the more commonly used MERV-8 filter) merely imposes cost associated with using a physically larger filter, and such filters are a bit more expensive. A brief perusal of air-filter vendor web sites [49–51] reveals that the MERV-15 is likely to be about 2-4 inches thick compared to a MERV-8 that is typically 1 inch thick. In addition, MERV-8 16x25 filters appear to be about $4, and a MERV-15 filter of the same size is about $40.

Further, the active control scheme given in the previous section does not save energy, as moving from a MERV-14 to a MERV-15 filter within the same filter family typically does not change the pressure-drop and therefore is achieved at no change in energy cost.

A.6.3 Minimizing Cost Using System Operation

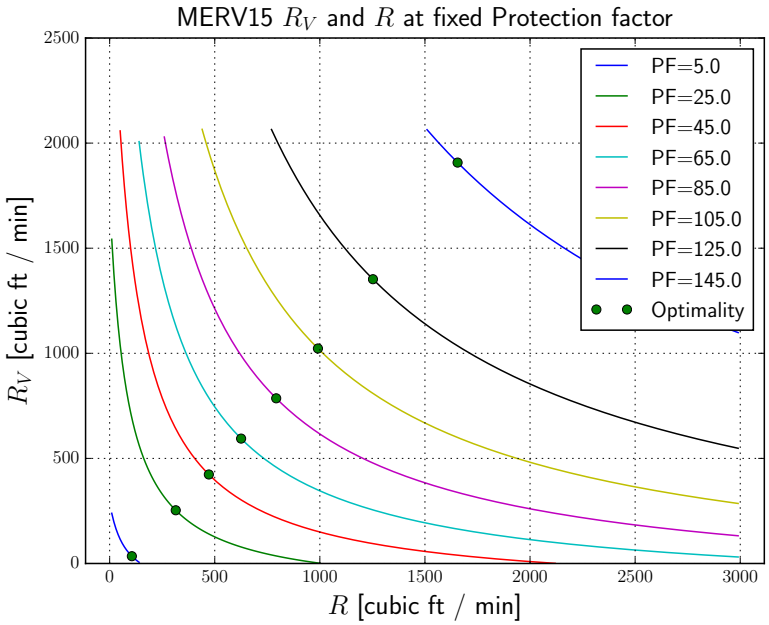


Figure A.9: Surface map of the protection factor as a function of flow rates R and R_V using a MERV-15 filter. The green dots indicate where one has achieved maximal protection factor as a function of total fan speed, $R + R_V$.

Therefore, to minimize energy consumption, it is desirable to minimize the

total rate of air blown through the HVAC system while achieving a fixed protection factor. The amount of air being blown is the interior flow plus the flow though the vestibule $R+R_V$. A graph that summarizes possible set points for the fan speeds R and R_V is now exhibited. Fig. A.9 on page 125 shows that the highest protection factor is achieved at a vestibule flow rate, R_V, that is, surprisingly, nearly equal to the flow rate for the interior of the structure, R. It is interesting to note that using the standard five changes of indoor air per hour on our notional building, the flow rate would be approximately $R =$ 1000 cfm. A protection factor of 45 is achievable when the vestibule flow rate is as low as 156 cfm. However, a protection factor of 105 is achievable if we are willing to push the vestibule flow rate to approximately 1000 cfm. The right and upper bounds of Fig. A.9 on page 125 occur at the maximal human comfort level air velocity of 0.3 m/s. So, although 1000 cfm in the vestibule seems to be an alarmingly large number, it is not beyond consideration. To save energy cost in an enclosed area that merely serves as a vestibule, it follows that the vestibule air should neither be heated nor cooled if possible.

A.7 Consequences to Building Design

Newer building designs can now be formulated with simple extensions to this analysis. In the previous section it was shown that the air circulation rate, R_V, at optimum, is nearly the air circulation rate for the rest of the structure, R. Further, the calculations yield results that are invariant with respect to the volume of the structure V_I and invariant with respect to the volume of the vestibule V_V. Therefore it is legitimate to assume that a building designer would take advantage of these findings to increase the floor-space of the vestibule until the air-flow-rate per unit area is similar to the structure's interior. At this point, the "vestibule" is no longer correctly labeled as such. It may be more accurately named a "buffer zone" and given its own fraction of recirculation, η_V. Clearly, an alert architect would then re-purpose such a large area for additional functionality, such as an atrium with solar-thermal mass to meet the structures' heating and cooling needs. Even if the buffer zone is now far more prominent in the building's design, we will continue to use the subscript V to avoid confusion.

After a moment's thought, the overall building diagram is quite similar to

that given in Fig. A.3 on page 114 with exhaust and make-up air implemented in the new buffer zone. This new diagram appears in Fig. A.10.

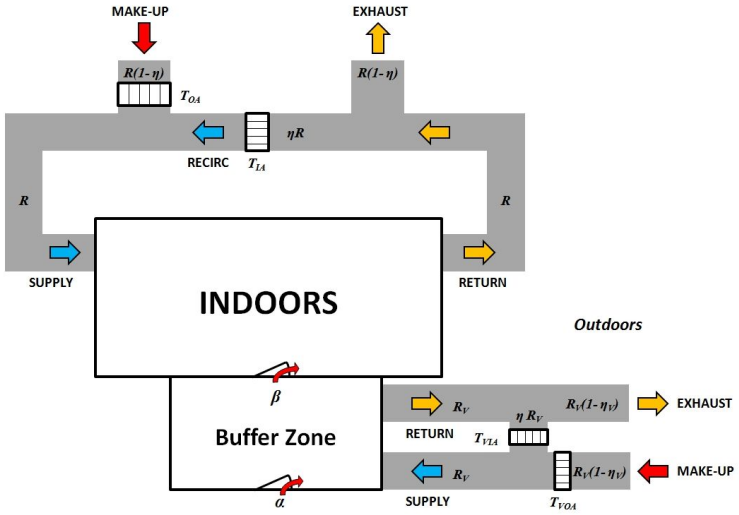


Figure A.10: Building plus HVAC system with buffer zone

The state-space model is now

$$\begin{pmatrix} \dot{A} \\ \dot{L}_I \\ \dot{V} \end{pmatrix} = \begin{pmatrix} a-g & 0 & f \\ b & 0 & 0 \\ g & 0 & c-f-h \end{pmatrix} \begin{pmatrix} A \\ L_I \\ V \end{pmatrix} + \begin{pmatrix} \gamma \\ 0 \\ \gamma_V + \alpha \end{pmatrix} u(t)$$

with the reciprocal protection factor

$$\lim_{t\to\infty} \frac{A_I(t)}{V_I} = \frac{\gamma - \frac{f(\gamma_V+\alpha)+f\gamma+h\gamma}{c}}{V_I(-a+g+\frac{-gh+af+ah}{c})}$$

where

$$a = -\frac{R}{V_I} + \frac{\eta R}{V_I} T_{IA} = \frac{-R}{V_I}(1-\eta T_{IA})$$

$$\gamma = R(1-\eta)T_{OA}$$

$$\gamma_V = R_V(1-\eta_V)T_{OV}$$

$$c = -\frac{R_V}{V_V} + \frac{R_V}{V_V}\eta_V T_V = -\frac{R_V}{V_V}(1-\eta_V T_V)$$

$$b = \rho/V_I$$

$$f = \beta/V_V$$

$$g = \beta / V_I$$

$$h = \alpha / V_V$$

All resulting graphs of the new design's performance are very close to indistinguishable from those of the building with vestibule given in Appendix A.4.5 on page 116. Therefore, the results of the calculation are dominated by the values of an average leakage current through the doorways α and β. The fresh air circulation, as given by the addition of vestibule recirculation fraction η_V, is a very minor effect. Therefore, some attention must be given to the estimates of door leakage rates α and β; however, their functional form radically changes the slope of the protection factor when they take on the small values envisioned in Appendix A.4.5, especially in Figs. A.4b and A.5b on page 118 and on page 119. Therefore, the result stands independent of the exact estimate.

A.8 Discussion

Using the approach described in this appendix for modeling infrastructure turns out to be an exceptionally flexible way to arrive at closed-form solutions summarizing how well any building can be protected from particulate matter.

As a very quick example to test the approach's flexibility, one can reasonably ask, Do these techniques relate well to residential housing? Using the description of the building with vestibule to describe a house with a front and back porch, a MERV-15 filter can be fitted to the recirculated air and porch air-handling unit. As the interior air of a house is being fully recirculated with fresh air exclusively provided by ambient building air leakage, one can quickly calculate the equivalent filter required to match this parasitic source of fresh air to keep the house at a protection factor of 45. In this case, the lowest quality filter is a MERV-8. Therefore, as long as the outer building envelope is sufficiently tight to serve as a pseudo MERV-8 filter, the required protection factor can be maintained. In recent work L. Meng [45] showed that typical residential building envelopes maintain an effective filter efficiency that is nearly the same as a MERV-14 air filter. Hence the MERV-15 filter will also provide bioprotection to residential housing.

All suggested changes to existing infrastructure in our approach harmonize

well with current emphases on DOD buildings, such as those requiring less energy for heating and cooling. Therefore, field acceptance will likely hinge on obtaining a product champion within the DOD for this method of determining building protection.

The central conclusions regarding the use of MERV-15 filters is hardly surprising, as these filters are normally associated with HVAC systems in hospitals. Therefore, the technology required for most structures will not be considered exotic. Further, this modest investment in protecting building occupants from mold, dust, allergens, and infectious diseases will likely pay for itself, considering the improved health and increased productivity of building occupants [52]. The current generation of MERV-15 filters, at full capacity, have pressure drops that are well matched to current electric fan efficiency curves. Therefore, the life-cycle cost is a relatively modest investment in adapting existing HVAC systems to accept the filters, and an increase in filter replacement costs of approximately $40 per filter every three to six months, depending on local practice. These costs are minimal compared to use of military M-98 filters that have one-fifth of the air capacity and cost $300 each.

The algebraic solutions for bioprotection of facilities outlined herein deserve a more comprehensive database of air-filter performance data. As an example, data could be obtained by using samples analyzed in the ERDC-CERL Bio-Tech Laboratory for scale test and evaluation, analogous to an ASHRAE 52.2 setup [53].

A.9 Conclusion

This article has shown a methodology for design and retrofit of structures to provide a high level of bioprotection. In the process, dynamical simulations have nearly been eliminated and substituted with algebraic solutions. These solutions provide greater insight into the design challenges involved. Using current simulation techniques, the graphs presented in this exposition would have required many hundreds of hours of computer time. Total time for all of the calculations presented is now less than 10 seconds.

From a mechanical engineering perspective, in the process of analyzing simple idealized structures, the following conclusions are demonstrated:

- State-space Laplace transforms reduce the problem from numerical simulation to algebra. Resulting code is thousands of times faster and will change point of use.
- MERV-15 suffices to reduce the area of affected buildings by about 80%.
- MERV-15 imposes minimal increase in electrical cost.
- MERV-15 filters cost $40, whereas MERV-8 cost $4.
- There is no advantage to active control of recirculation rate to achieve a lower MERV rated filter.
- At optimal recirculation rate, the vestibule is a candidate for being a much larger buffer zone with extra features such as passive solar mass, atrium, etc.
- A typical residential building envelope filters at an efficiency of approximately MERV-14; therefore, MERV-15 suffices for buildings that have no makeup air intake.

Therefore, suggested changes to a building retrofitted to a reasonable bioprotection standard are economically feasible. The exposition has also provided some insight into the likely changes in building design when bioprotection is considered.

APPENDIX B

BIOINSPIRED DESIGN OF EXOSKELETAL SENSORS

B.1 Summary

As part of a bio-inspired design process, the authors examine exoskeletal sensors found in insects, and their potential application to armor and hardened buildings. In this way, the outer hardening of a structure or vehicle would not limit the ability of occupants to arrive at an actionable picture of the outer environment. To this end, various sensor modalities employed by insects are compared and contrasted with their current human-engineered equivalents. In several sensing modalities, biosensors perform better, are smaller, and more energy efficient than human-engineered equivalents. We note that biological designs tend to employ nonlinear response to signal amplitude and respond with heightened sensitivity over a greater dynamic range of signals than human-engineered sensors. The insect biological sensors have structural and mechanical innovations that preserve the protective capacity of the exoskeleton.

B.2 Background

As early as the fourth century BCE, engineers have borrowed designs from nature to achieve specific technical goals [54]. Over the past few decades, bio-inspired engineering has become a catchphrase for engineers borrowing design ideas from nature in an attempt to replicate some of the remarkable properties exhibited by natural systems.

As a source of engineering designs, the popularity of bio-inspiration has accelerated in recent years due to scientific and technological advances and socio-cultural factors. In recent years, modeling and simulation software

targeted to visualization and analysis of biological mechanisms is now widely available and more commonly used by many engineering disciplines [55–57]. This allows engineers to appreciate subtleties of natural materials and how such materials may be used to advantage [58]. These are now modeled in sufficient detail that the designs can be both understood and adapted to engineering needs. From a socio-cultural perspective, renewed concerns regarding energy efficiency and sustainability have motivated engineering disciplines to reverse-engineer natural designs in an effort to obtain more sustainable systems. Focusing on the example of exoskeletal sensors allows exposition of how bio-inspired design does not result in single ideas for implementation; it yields an alternative way of thinking about practical engineering designs. When pursued in the light of bio-inspiration, the process of engineering development changes. Moreover, bio-inspired design forces the engineer to confront constructive uses of nonlinearity much more directly.

B.3 Overview of Natural versus Engineering Perspective

From an engineering standpoint, a designer wants to take advantage of the designs that are already pioneered by nature. Through evolution, these natural designs are developed: over many more generations with a new engineering iteration for each new mutation or variation, with more stringent engineering success criteria (survival to reproduction), and with an effectively larger aggregate budget and much longer deadlines than those designs allotted to human engineering development. Life on earth has been adapting and evolving to changing conditions for more than 3.8 billion years and hence represents a generous development program that engineers can draw from.

As an orientation, one can summarize the differences between natural design and engineering design from four perspectives: material selection, energy usage, common design methods and design goals. Evolutionary pressures generate countless design variations and restless innovation; however, the resultant designs have specific limits in energy consumption and material selection. In nature, material and energy are expensive, yet variable form is cheap. Human engineering traditionally works based on the opposite principles.

B.3.1 Material Selection

Animals and plants are constructed of raw materials that are readily available such as carbon, oxygen, hydrogen, nitrogen, phosphorous, calcium and silicon. The majority of the structural materials created are two types of polymers (proteins and polysaccharides), and two types of ceramics (calcium salts and silica), sometimes combined with a few metals. Many human-made materials are manufactured using high-temperature methods and elements from the periodic table that exist in low natural abundance. These materials are therefore energy intensive to refine and fabricate. Additionally, natural materials are often self-assembled and are hierarchically organized. Materials are "grown" from basic components up to complex structures interacting in systems; in essence, a synthesis or bottom-up approach. In contrast, human-engineered materials are usually manufactured by combining all components from the start to form a bulk material and then physically carving or chemically reacting to remove excess "waste" material, in a top-down approach.

B.3.2 Energy Usage

In nature, the vast majority of fabrication processes occur at ambient temperatures and pressures. Human engineering has, historically, made profligate use of cheap energy sources (high temperatures, high pressures and sometimes toxic chemicals) that are inherently less sustainable and efficient than natural forms of energy generation: e.g., using solar radiation to perform chemical reactions through photosynthesis.

B.3.3 Common Design Methods

In nature, many organisms' senses response varies logarithmically with respect to the magnitude of the sensory input. This observation is so often repeated that it is sometimes referred to as Weber's law [59]. By using logarithmic strength of response, nature as viewed from an engineering perspective achieves two ends.

First, this logarithmic response markedly increases the dynamic range of the senses. As an example, the light inside a well-lit closed structure is millions of times less intense than sunlight, yet it is common for organisms to see quite

well in both environments. Similarly, numerous biological senses including sight, taste, touch, hearing and smell respond over enormous dynamic ranges of stimuli. In biological sensing, it is not surprising to encounter animal sensors that correctly interpret stimuli that vary over six decades in dynamic range. This is to say that the ratio of the strongest to weakest stimuli commonly exceeds a factor of a million. In engineered systems, responding to such an enormous dynamic range is difficult to achieve.

Second, by responding to the strength of input stimuli logarithmically, biological systems are startlingly sensitive at low stimulus levels. Engineered systems, with their reliance on sensors that respond linearly, are usually far less sensitive and detection of weak signals is very difficult to achieve.

Historically, this disparity is easy to describe in terms of the design methods and the fundamental mathematical analysis tools available to a design engineer. The most often-used mathematical methods employ transform techniques that fundamentally require the assumption of linear response. In nature, systems adhere to designs that respond logarithmically. One way to notice this is to observe that the ratio of "just-noticeable difference" to the magnitude of the stimuli remains fixed over many decades of stimulus intensity, directly leading to the logarithmic response summarized by Weber's law [59]. Engineering tools for analysis of any form of nonlinearity are far more difficult to formulate and use than those which assume linearity.

B.3.4 Design Goals

Natural variations in form occur constantly due to inherent genetic variation. The central goal of natural design is simply stated, yet leads to complex variations in morphology. The goal is, simply: Survival to reproduction in a competitive, harsh, and variable environment. Evolution toward achieving this goal, however, is limited because species are only able to evolve starting from the most recent previous platform; each successive generation can only exist as a modification of what has come directly before it. Because of this, some biological solutions have evolved in a roundabout manner. In contrast to natural design goals, designs used in modern engineering are driven by human needs and desires. These goals can be summarized as "faster, better, cheaper, safer, ubiquitous, robust, and scalable." Additionally, engineered systems

can evolve just as biological systems, such as the design of car components slowly changing with each passing decade. However, engineered designs also have a significant advantage; they can start from a clean slate and take a novel approach to a design problem by combining many sources of design inspiration. In this way engineered systems can "cherry-pick" components and combine them into a very successful system. Table B.1 on the next page summarizes the differences in perspective.

So far, the focus has been on the differences between engineering and biological design that appear to be of manageable complexity. That is to say, we limit ourselves to forms of bio-inspiration in which the process of modifying the engineering perspective appears to be reasonably tractable using current theoretical understanding. Therefore, Table B.1 on the following page is far from an exhaustive list. Many features of hypothetical bio-inspired design are far from achievable through current science. For example, many biological systems are self-healing in a manner that is completely intractable using current engineering methods: several species of salamanders can regenerate legs, tail, jaws and eyes, but it is difficult to conceive of an automobile that can spontaneously regenerate a tire that has been entirely removed.

To sharpen and illustrate the arguments in Table B.1 on the next page, the focus of attention is on sensors for use in exoskeletal structures.

B.4 Exoskeletal Sensors

Consider the exoskeletal material used by invertebrate animals such as insects, called the cuticle. The cuticle functions as a protective barrier and also serves as a scaffold, giving the insect its shape and giving its internal structures support. The cuticle is made up of chitin, which forms fibers that are embedded in a protein matrix. Chitin is one of the most abundant biopolymers in nature. Depending on the chitin and protein binding to phenolics within the cuticle, the amount of water present, and the amount and location of metals in the structure, the chitin can be extremely stiff or very soft, even within the same animal. This is a manifestation of nature not having access to the profligate energy required to achieve the advanced material properties afforded by metallurgy or polymer science, yet creating materials able to accommodate many different environmental conditions and achieve many targeted end-uses.

Table B.1: Differences in perspective between engineered and bio-inspired designs

	Engineering Designs	Natural Designs
Material Selection	Material is cheap, abundant and diverse only because profligate amounts of processing energy are cheap and available	Material is expensive – uses abundantly available materials that can be processed at environmental temperatures and pressures.
Energy Usage	Energy for fabrication and operation is (historically) cheap and abundant	Dependence on near-environmental temperatures and pressures; efficient
Design Methods	Mathematical tools strongly limit interest to systems where the output magnitude is a linear function of the input magnitude.	By use of nonlinearity, most commonly logarithmic response, designs work well over many decades of stimulus magnitude, and can be far more sensitive at extremely low input magnitudes. Systems are therefore flexible and can accommodate variable situations and stimuli.
Design Goals	Variations in form are expensive: Standards are: faster, better, cheaper, safer, ubiquitous, robust, and scalable (larger or smaller). Approach can be "novel" drawing from many sources.	Variations in form are cheap. Standards are survival to reproduction in a competitive and/or harsh environment.

However, this natural structure does have its limits. Biological exoskeletons of land-dwelling organisms appear to be limited in size to several inches, likely due to how the strength-to-weight ratio scales with size. Exoskeletal man-made devices do not share this constraint and can be considerably larger and heavier.

From the standpoint of military engineering, the use of armor and hardened structures would seem to work against our ability to sense the battlefield or the built environment (e.g., the driver's optics in an armored vehicle, which are quite limiting). The very hardening used to protect also tends to impose limits on sensing equipment that needs access to the outer environment. This greatly limits the human ability to form a holistic and actionable environmental awareness when protected by an engineered exoskeletal structure. In contrast, we know that many insects are capable of very accurate and extensive sensing of their environment, despite having a tough and stiff composite as their exoskeleton. This is achieved by the many sensors, of different modalities, that are embedded in the cuticle. Many of these sensors are shaped to fit in blind holes in the cuticle. In this way, the sensor does not cause full penetrations in the cuticle that would weaken the exoskeletal structure.

Hence, exoskeletal creatures are a clear source of bio-inspiration in their ability to negotiate the environment, sometimes being exquisitely sensitive to specific sensing modalities, while retaining the protection of an exoskeleton. Therefore, it would be advantageous from an engineering perspective to understand the functionality of insect-based sensors.

Bio-inspiration for armored vehicles or hardened structures here is a directly motivated application. Having a hardened surface bristling with sensors only requiring blind holes of negligible depth in the surface would represent an enabling technology allowing ubiquitous sensing applications on the outer surface of armor and other hardened structures and buildings.

B.5 Boosting Sensitivity to Biological Levels

The sensitivity of biological sensors is notably higher than most engineering sensors. Insects exploit the nonlinear behavior of materials to increase the sensitivity of biosensors to stimuli [19]. For example, campaniform sensillum, a strain sensor found in insects, relies upon the nonlinear response of a

membrane structure to deflection forces to amplify the strain-sensing property of the microstructure [20]. As another example, the fire beetle uses a thermo-pneumatic transduction mechanism that consists of a cavity covered with a cuticular film [21]. A pyrophilous buprestid beetle uses a thermo-pneumatic transduction mechanism that consists of sensilla that are contained within dense spheres covered by a fairly thick cuticular dome [21]. Infrared (IR) radiation will cause the material within the sphere to expand, but expansion is resisted by the cuticle surrounding it. The pressure within the sphere will eventually deflect the dendrite portion of the neuron. The expansion of the material contained within the sphere is a linear function of IR light with a spectral distribution corresponding to a fire of about 700 °C, except near a small region centered near room temperature. Within this region, where the film deflection is a nonlinear function of temperature, the sensor has the greatest sensitivity to IR light [22, 23]. It is theorized that the overall sensor draws the insect toward a fire, while the nonlinearity at room temperature assures that the insect will only land on a branch that will not burn the insect. Studies of this sensor structure have resulted in biometric strain-sensing microstructures and micro-fabricated cavities for IR micro-imaging [55].

Although the idea of incorporating a nonlinear feedback element to improve measurement sensitivity is not new, the closest literature demonstrating how to boost sensitivity by incorporating nonlinear elements is the marginal oscillator, which is an instrument for revealing small changes in the losses of a harmonic oscillator. Roberts and Rollin credit Pound with developing the first marginal oscillator in the 1940s to observe nuclear magnetic resonance transitions within solids [8, 10, 11]. As such, it is a very early example of engineering design taking advantage of nonlinearity to boost sensitivity. It can be readily shown that a marginal oscillator can boost sensitivity to changes of resistance, capacitance or inductance.

This ability to measure tiny changes in resistance, capacitance or inductance has important ramifications in how engineering designs would be pursued in the absence of bio-inspiration. One can easily imagine pursuing a string of applications where a primary device has been fabricated to undergo a small but predictable change in resistance, capacitance or inductance. Embedding these devices in a marginal oscillator, gives rise to a new secondary device that possesses an increased sensitivity never anticipated by the designer of the primary device. For example, consider a thermistor (a primary device

that changes resistance based on temperature). A thermistor could be embedded in a marginal oscillator so that the oscillations would yield a very sensitive measurement of changes in the thermistor's resistance. The resulting secondary device could, in principle, measure tiny fluctuations in temperature never considered during the design of the original thermistor.

From an historic perspective, common resistive sensors include temperature, strain and magnetic flux; capacitive sensors include touch-pads, precision positioning, microphones and accelerometers; inductive sensors include metal detection and physical proximity. Clearly, each of these can be revisited by embedding the key variable resistive, capacitive or inductive element in the sensor circuit.

Therefore, in the engineering context of exoskeletal sensors, the principal candidates for further development are temperature sensing via small change in resistance; microphones and accelerometers via changes in capacitance; and physical proximity via inductance.

Another example of possible interest is micro-cantilever sensors for detection of chemical and biological analytes [60]. These sensors use a beam whose length is on the order of 100 µm. A custom polymer coating on the beam preferentially adsorbs the analyte of interest. The beam is excited at its resonant mode using either a piezoelectric or electrostatic actuator driven by an external oscillator. The adsorption of the target material is revealed by a decrease in the resonant mode frequency. A challenge lies in detecting the small change in resonant frequency. As will be shown, these detectors can be functionalized to detect a single analyte of interest. As they can be miniaturized, an array of such sensors could check for multiple analytes of interest. Although there is some similarity between these sensors and insect antennae, antennae work reliably over a much wider dynamic range of stimuli than current cantilever beam sensors.

B.6 Shifting to a Bioinspired Perspective

There are two distinct problems in perspective that occur when a bio-inspired design path is adopted. First, when one looks for naturally occurring sensor devices from an engineering perspective, the differences listed under "Design methods" in Table B.1 on page 136 cause immediate difficulties. Engineers rely

on lumped linear elements (resistance, capacitance and inductance) that have been analyzed fairly completely from a mathematical standpoint. Therefore, the engineering approach is to state that all nonlinear elements were implicitly designed by evolutionary selection. This approach is intellectually rigorous, and remains so even if a close study of evolutionary adaptation later shows that the linear elements were more difficult for nature to develop. In typical applications, this problem is the main barrier to adopting more bio-inspired methods.

The second problem occurs when one sees that nature has already arrived at exquisite designs for sensor modalities that are wildly dissimilar from human-engineered designs. As an example, insects do possess sensors that are finely tuned to specific pheromones. Hence, the enticement is to start at an uncomplicated easy-to-adapt natural design, and to begin to work forward from that point based on the morphological similarity to other biological sensors of interest.

As a first example, let us start with the campaniform sensillum, a biological sensor with two applications. First, it can detect the amount of strain experienced by the exoskeleton; second, it is used to monitor articulated pieces of an exoskeleton as they move with respect to one another, a process called proprioception (i.e., the ability to sense the body's position and posture).

B.7 The Entomology Design Perspective: Campaniform Sensillum

Insects make use of nonlinear materials to increase the sensitivity of biosensors to stimuli [19]. For example, the campaniform sensillum (Fig. B.1), a strain sensor found in insects, relies upon the nonlinear response of a membrane structure to deflection forces to amplify the strain-sensing property of the microstructure [20]. Mechanosensory organs in an insect's leg, such as the campaniform sensilla, contribute to locomotion such as walking or running, and its control.

For an engineer, the outer-fiber of any exoskeletal structure carries a large fraction of the mechanical load placed on that structure. Because of this, positioning any sensor in the outer fiber without compromising the mechanical strength of the structure is difficult. As an initial example, we show how

nature solves this problem of material selection and shape in the campaniform sensillum. The sensillum is a small hole within the cuticle under a cuticular cap or dome, seemingly a weakening feature. However, chitin fibers around the hole are oriented in such a manner that they can carry load more efficiently, and the presence of the hole therefore does not compromise the mechanical strength of the surrounding material. The sensory neuron is located under this hole and within the cuticular dome, contained at the distal end within a cuticular scolopale. Furthermore, the hole is oval-shaped, and when the cuticle is deformed the shape of the hole can give information about the direction and size of the strain on the cuticle. In combination, these fields of multiple campaniform sensilla in an insect's leg relay information to the animal's brain about bending, compression, tension and twisting.

Entomologists have studied a variety of insect mechanosensors with widely different sensory functions that are mechanically quite similar to the campaniform sensilla, including thermo- and hygro-receptors, IR receptors and chordotonal organs. The chordotonal organs are, in turn, present in: Johnston's organs (which monitor antenna deflections), subgenual organs (located inside of appendages for surface vibration detection) and tympanal organs (used in sound and vibration detection). Other types of receptors, which are discussed in Figs. B.1 to B.12 on pages 143–153, also contribute to the high success of insects as a group.

In contrast to the engineering approach, in which sensors for different modalities tend to be very different from one another, biologists have found that a wide variety of exoskeletal mechanosensors exist that, while exhibiting great functional variation, are morphologically only slight variants of the campaniform sensillum. In these figures, note the similarity of campaniform sensilla to: thermo- and hygro-receptors, IR receptors, trichoid sensilla and chordotonal sensilla. Figures B.1 to B.12 on pages 143–153 give short descriptions of various exoskeletal sensors, a rough engineering equivalent and a short description describing and contrasting the capabilities of the two sensors in the caption.

B.8 Discussion

The devices summarized in Figs. B.1 to B.12 on pages 143–153 are diverse in their sensing applications: temperature, humidity, strain, vibration, proprioception, chemoreception, aural and visual. It is interesting to note that some classes of engineering sensors have already gone well beyond their insect analogs: visual, thermal and aural. However, several classes of biosensors are still superior to anything available in modern human engineering for detecting strain, humidity, vibration and chemoreception. The evolutionary development path leads naturally from strain, to humidity, to vibration, with chemoreception being important but morphologically and technically quite dissimilar from the others. In the absence of bio-inspiration, there would be little motivation to guess, much less to follow, this particular development path, as the analogous human-engineered sensors are developmentally unrelated.

Also noteworthy is the use of cuticular material in the sensor design. In this way, the exoskeleton's protective capacity is preserved despite the sensors' presence in a blind hole located in the outer fiber. This is accomplished in two different ways. First, the microstructure of the cuticular material around the blind hole is slightly reinforced by changing the direction of the polymeric chains near the sensors' opening. Second, many of the sensors retain a cap made of cuticular material to keep out dust and moisture. In contrast, normal engineering practice would discourage placing blind holes in the outer fiber to preserve structural integrity and increase the load capacity.

B.9 Conclusion

The principal differences between bio-inspiration and a pure engineering approach are changes in material selection, energy usage and design goals. Nature is also more likely to take advantage of nonlinear phenomena where the sensor responds as a logarithm of the magnitude of the stimuli. In this way, natural sensors usually have a far larger dynamic range and far better sensitivity to small stimuli. Further, the novel use of cuticular material preserves the protective capacity of the exoskeleton despite sensor placement within the outer fiber of the exoskeleton. These observations point to novel

design strategies when designing hardened structures and armor. Therefore, it may be possible for such hardened structures and machines to harbor a large number of sensors while maintaining their protective capacity.

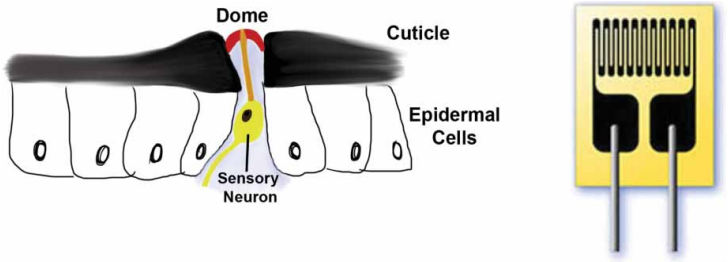


Figure B.1: Campaniform sensilla - shear forces within the cuticle deform the cuticular dome causing movement of the scolopale (a cuticular cap that covers the neuron) generating a receptor potential in the dendrite. Notice the far more elegant engineering solution of using a dome to suspend the sensor. The dome can keep out dirt and dust. The strain that is recorded by the sensilla can originate from a load on the cuticle due to an outside force or load, and also from forces applied to the cuticle by leg muscles during movement. The dome of the sensilla is approximately 7 µm in diameter, and is capable of measuring displacements of about 100 nm. At right is a stick-on strain gauge that is quite sophisticated by engineering standards, yet inferior to campaniform sensilla in both size and sensitivity. The gauge measures about 2 mm by 1.2 mm and is capable of measuring displacements of about 1 µm. The biological sensor is ~ 300 times smaller and ~ 10 times more sensitive.

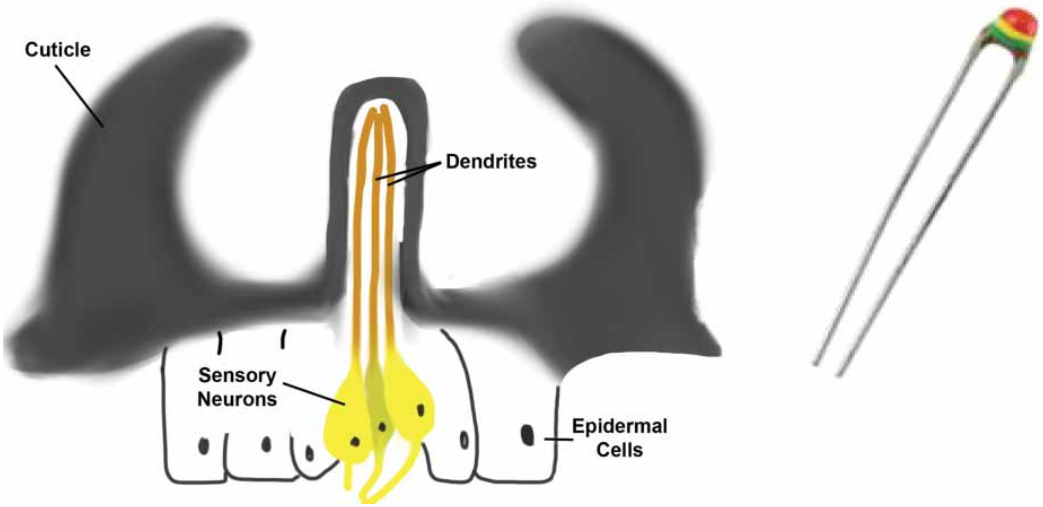


Figure B.2: Thermo- and hygro-receptors - notice that the fully enclosed cuticle allows continuity of the surface. In the most common arrangement, three neurons come together in a triad arrangement; one is sensitive to cold air, one to moist air and one to dry air. In other arrangements neurons sensitive to cold and warm temperature come together. The information relayed to, and integrated by, the central nervous system enables insects to discriminate changes in environmental conditions. The thermistor is likely superior in an exoskeleton, being capable of working over large ranges in temperature with excellent sensitivity. However, the biological hygrometer is superior to the human-engineered version. The best un-calibrated electronic hygrometers are only accurate to about $\pm 6\%$ relative humidity, whereas the biosensor gives an accuracy $<2\%$ relative humidity.

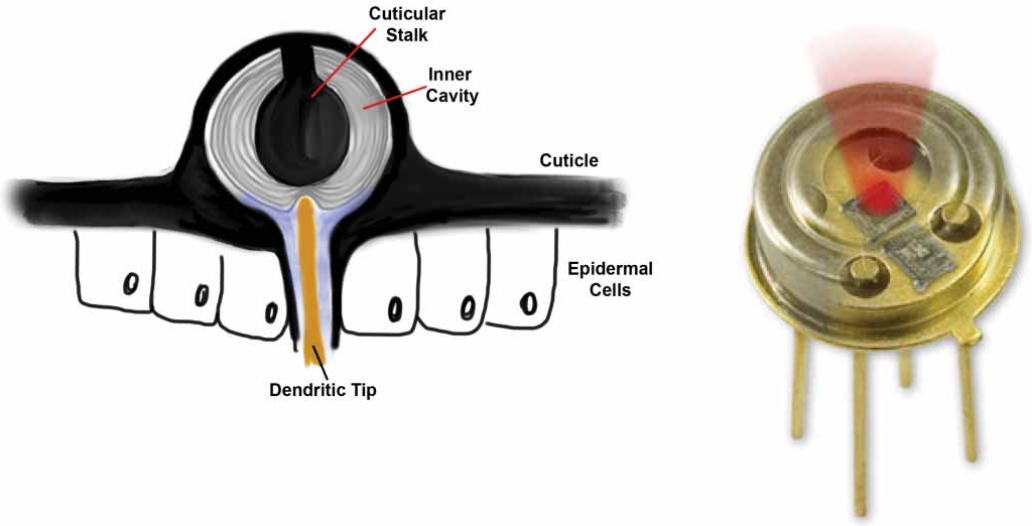


Figure B.3: IR receptors - the structure of this receptor as a modification to the campaniform sensillum should be readily apparent. Again, the entirely enclosed structure is advantageous in keeping out dirt and dust. The sensor relies on the cuticle's stiffness, as well the fluid-filled chamber surrounded by the cuticle. As the material in the interior sphere heats up, strain on the cuticle increases. But the cuticle is stiff enough that strain is deferred and the dendrite of the neuron is triggered. The engineered IR sensor is linear over its dynamic range, whereas the insect sensor uses nonlinear feedback to be far more sensitive to a narrow range temperature. It is thought that this complex design is associated with finding optimal spots for laying eggs on recently burned branches that have cooled to a fairly narrow temperature range.

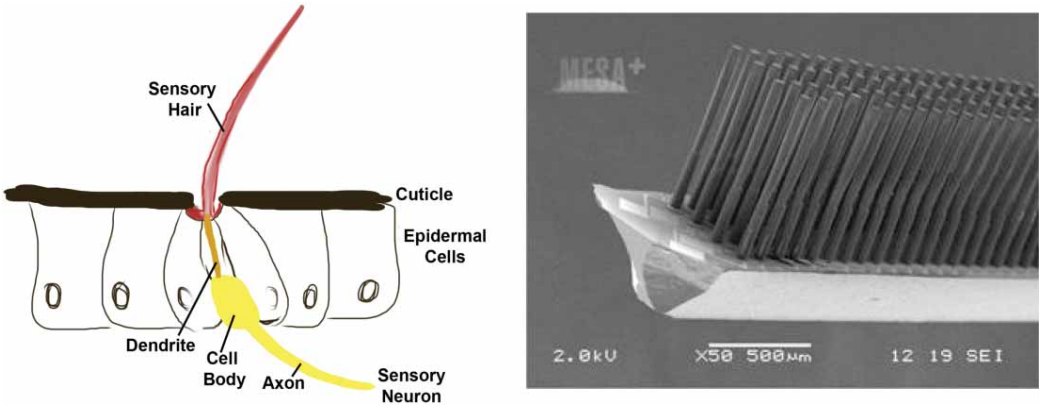


Figure B.4: Trichoid sensilla - single tactile hair. Hair receptors of touch (of external objects) and body position (when placed in contact with an articulated joint). Sensors respond to tactile input, but also to vibrations in the air or substrate. The sensor is relatively simple, but rapid transmission is essential for escape maneuvers. Hairs can be unidirectional or omnidirectional. Sensors may be grouped together in a hair plate - common at leg joints. The engineering equivalents are all recently developed bio-inspired designs. As these designs are in their infancy, it remains to be seen how their eventual performance will compare with the original biological design. Source: Krijnen et al. [61]

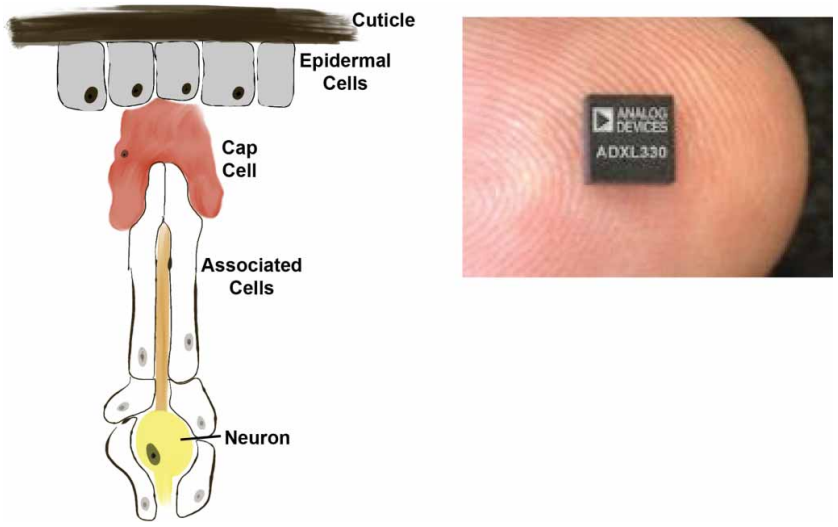


Figure B.5: Chordotonal sensilla (scolopidia) are anatomically more complex than trichoid sensilla. The sensory neuron is enclosed within parts of two or three other cells, including a scolopale cell and cap cell which occurs at, or just under, the cuticle. Any stress or strain at the cuticular surface is transmitted to the sensory neuron beneath. Chordotonal organs are often used to detect vibrations. Complex chordotonal organs contain many individual scolopidia. The scolopidia are individual structures that are typically combined in groups to form sense organs. In this article, chordotonal, subgenual, tympanal and Johnston's organs are all examples of this compound design. The biosensor is approximately 100 times more sensitive than its engineering analog. Insects use these sensors to detect the vibration of substrates - for example, to determine the contents of an insect egg case or to communicate with prospective mates.

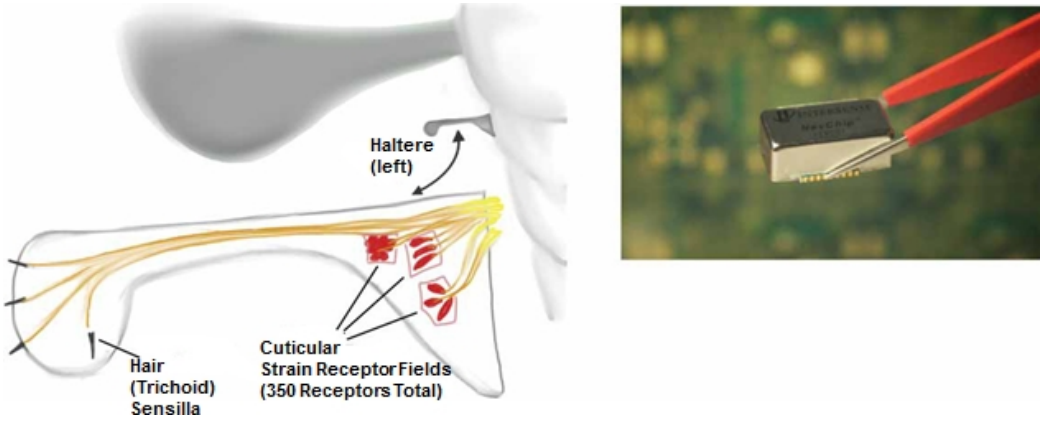


Figure B.6: Halteres are small external projections, derived from ancestral wings, on the bodies of true flies (Diptera), which have developed as a gyroscope and accelerometer combination for flying insects. The chip at right is an integrated gyroscope/accelerometer. The two sensors are likely close to one another in sensitivity, however, the engineering unit is still much larger at 21 mm × 13 mm × 9 mm where a haltere is typically 2 mm × 1 mm × 1 mm. Source: Intersense Incorporated, NavChipTM website, http://www.intersense.com

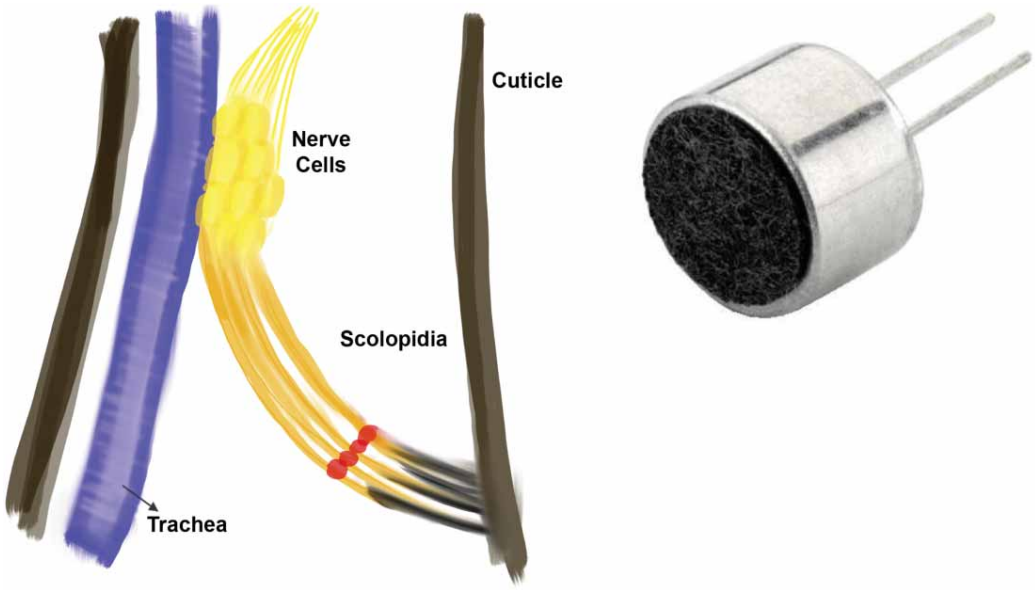


Figure B.7: The subgenual organ combines multiple scolopidia and is usually located within the insect leg. It is involved in proprioception and can detect vibration in the substrate that the insect is standing on. Scolopidia are closely associated with both the cuticle and with the tracheal (respiratory) system. The trachea is filled with air and probably aids in amplification of the vibratory signal. A modern engineering equivalent, the contact microphone, is typically about 2 mm across - whereas the insect organ tends to be 0.3 mm. The two sensors have comparable sensitivities.

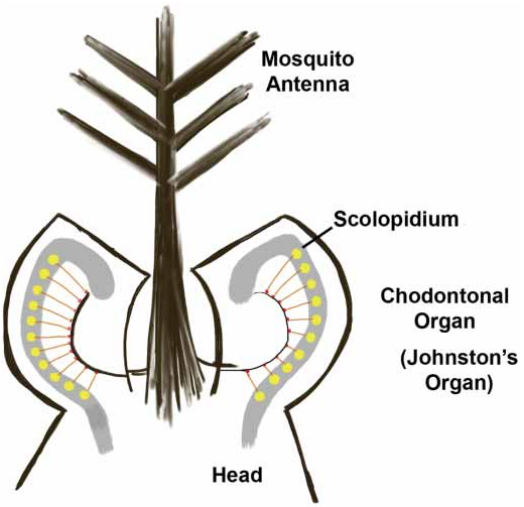


Figure B.8: Johnston's organ is a large and complex chordotonal organ located near the basis of the antennae of adult insects. The sensor monitors antenna movement, either on the part of the insect or due to wind, gravity, flight, sound and other forces. Currently, there is no particular engineering equivalent. These organs are remarkable in that they give a nonlinear response to stimuli that achieve amplitude compression, vastly increasing their dynamic range. In this way, insect songs can be detected at both long range and very short range [62].

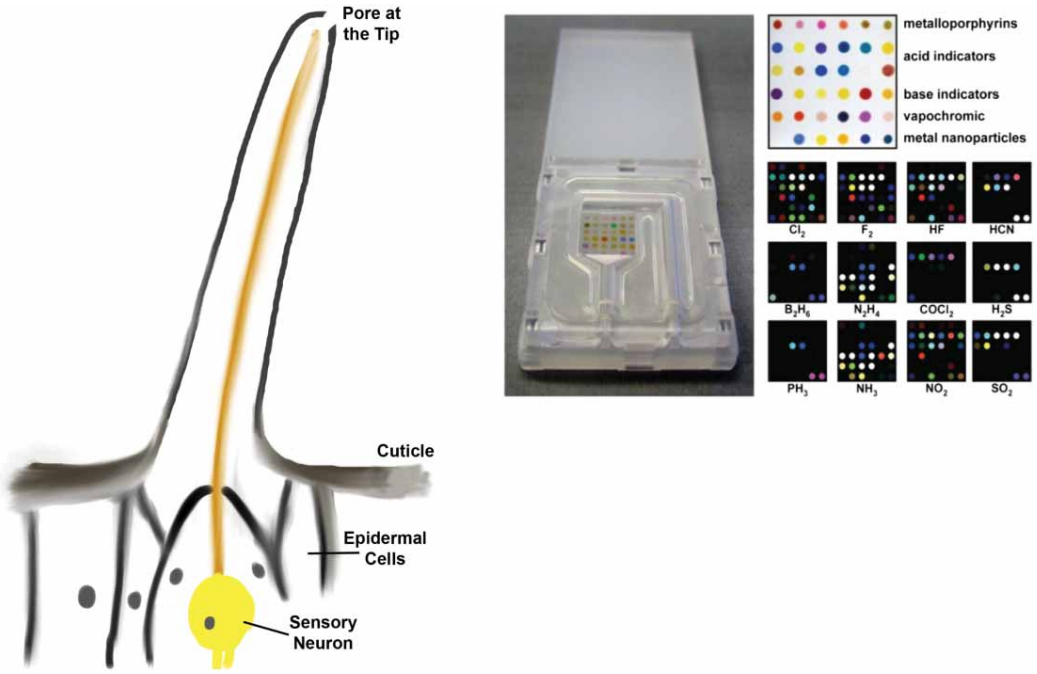


Figure B.9: Gustatory contact chemoreceptors are modified trichoid sensilla with one pore near the tip of the cuticular structure. These sensors detect chemical substances in solution. They can be hair, peg or dome shaped. These chemoreceptors are located on mouth parts, tarsi (insect feet) and female insects' ovipositors. A chemical enters the pore and is soluble in the lymph. The chemical will attach to specific receptors on the dendrites of their corresponding receptor neurons. The result is that each receptor detects a very specific subset of chemicals: only those that are relevant to the organism. In recent years, an engineering equivalent using metalloporphyrins has been developed that has similar functionality. Although their sensitivities are comparable, the natural system is far more compact: the engineered sensor is about $1\,\text{cm} \times 1\,\text{cm}$, while the natural chemoreceptor measures $15\,\mu\text{m}$. Additionally, the biological receptor can self-reset, whereas the engineered system is not yet capable of resetting. Photo Credit: Kenneth Suslick.

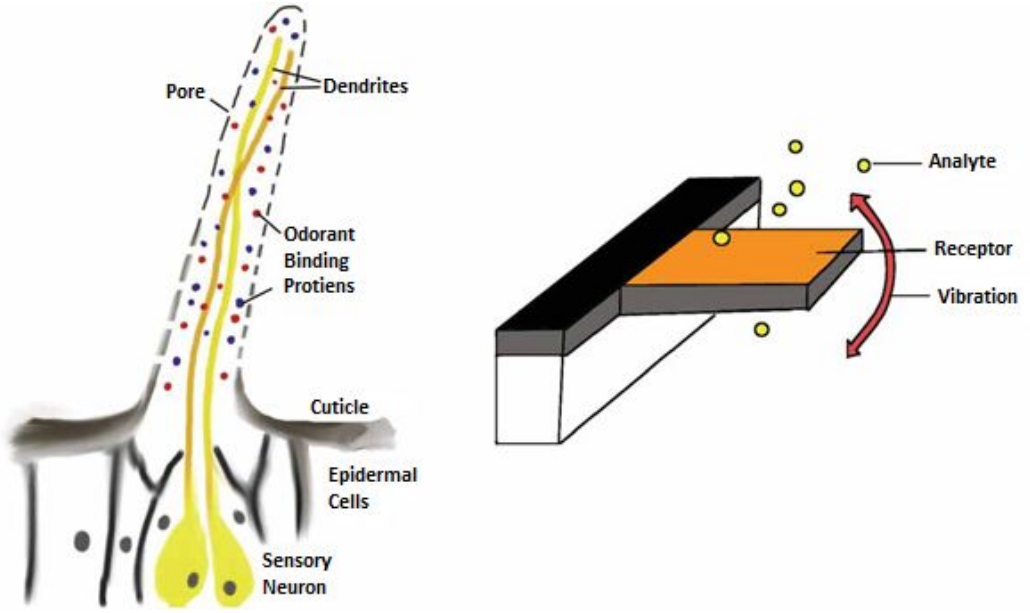


Figure B.10: Chemoreceptors - again, these sensors are based on the trichoid sensillum design, but with multiple pores. The biodesign is far more complex than any engineered sensor. The biosensor is capable of sensing multiple analytes, as there are heteroreceptors attached to the dendrites. The space around the dendrites is filled with fluid, binders for the analytes of interest and antagonists for the binders once they have been sensed by the dendrites. This structure allows the chemoreceptor to work continuously. The engineered sensors at the right are of three different styles for finding single analytes, and all are based on the concept of taking a single measurement with no capability to reset. Here bio-inspiration would lead to far more sophisticated mechanisms for achieving a reset of the individual receptors.

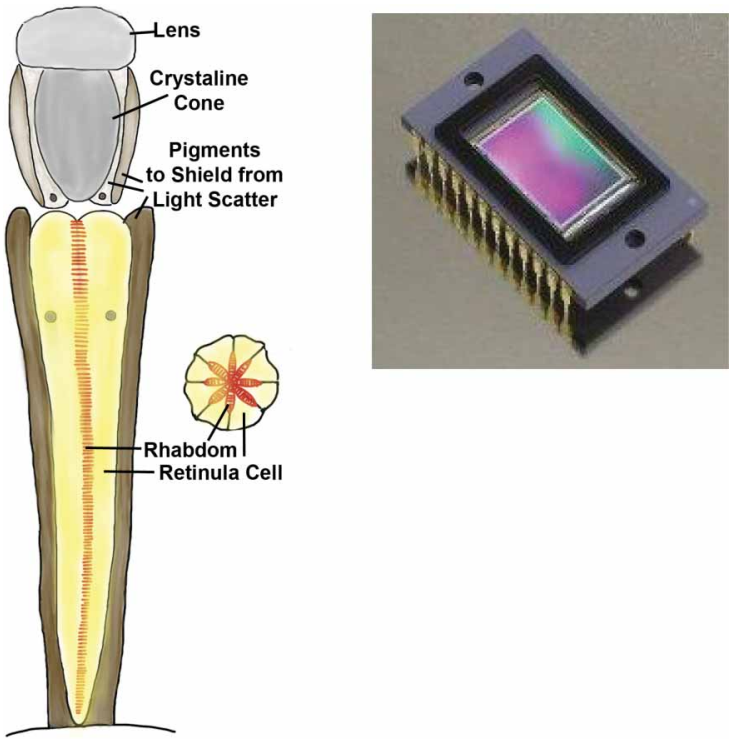


Figure B.11: The ommatidium, the basic small unit of an insect's compound eye, can be thought of as a single pixel in the insect's visual field. Ommatidia allow insects to detect color and to form visual images. They are especially sensitive motion detectors. Each ommatidium is composed of many cells, including the retinula cells containing the photosensitive pigments. Other functional cells contain pigments that shield light not coming from directly above the lens. The retinular cells send their axons directly to the brain. Illumination of the eye causes a photochemical reaction involving the pigment rhodopsin. This chemical change in the pigment causes an electrical activity in the axon. The engineering equivalent, a charged-coupled device (CCD) array, is likely superior for exoskeletal use. The sensors on a CCD surface are currently one sixth the size of corneal lenses, measuring 15 µm versus 100 µm. Further, having one lens occurring in isolation is likely to yield a more economical solution than multiple individual lenses.

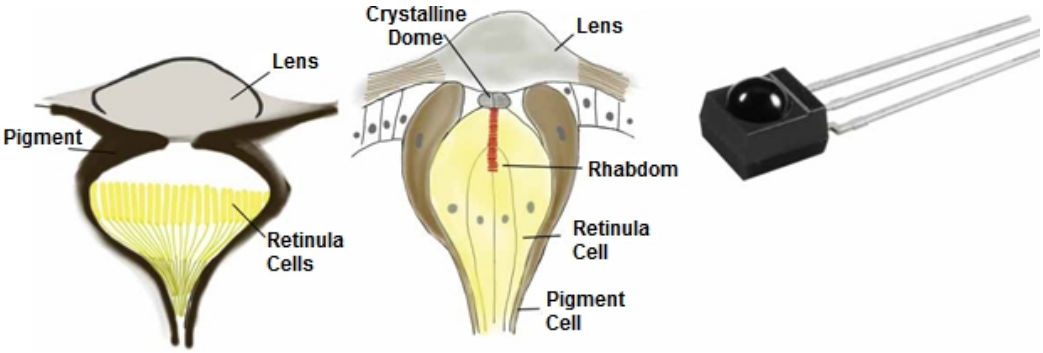


Figure B.12: Ocelli, another light-sensing structure, also have a corneal lens and retinular cells, similar to the ommatidia. However, they do not convey an image to the brain; instead they detect the presence or absence of light, the intensity of the illumination and the wavelength of the light. It is currently thought that this is related to sensing slow changes in lighting such as diurnal cycles. Immature insects also employ another light-sensitive organ, called stemmata (not pictured here). Stemmata, like ocelli, form only poorly resolved images, but they do give important visual information without a lot of energy investment in complicated structures. Again, the CCD likely yields a technologically superior solution as each sensor is one sixth the size of an ocellus. An example of a much larger common optical receiver is depicted here.

REFERENCES

[1] M. Ginsberg, J. Schiano, M. Kramer, and M. Alleyne, "A case study in bio-inspired engineering design: defense applications of exoskeletal sensors," *Defense & Security Analysis*, vol. 29, no. 2, pp. 156–169, 2013. [Online]. Available: http://dx.doi.org/10.1080/14751798.2013.787798

[2] N. Suga, *Dynamic Aspects of Neocortical Function*, ser. The Neurosciences Institute Publications Series. John Wiley & Sons Inc, January 1985, ch. The extent to which biosonar information is represented in the bat auditory cortex.

[3] N. Suga, "Biosonar and neural computation in bats," *Scientific American*, vol. 262, no. 6, pp. 60–68, June 1990.

[4] J. L. Schiano, "Enhancing electrostatic transducer sensitivity," 18 June 2013, private communication.

[5] M. D. Ginsberg and A. T. Bui, "Bioprotection of facilities," *Defense & Security Analysis*, vol. 31, no. 1, pp. 1–18, February 2015, on-line 25 February 2015.

[6] National Institute of Building Sciences, "Whole building design guide," July 2013. [Online]. Available: \url{http://www.wbdg.org}

[7] W. R. Ott, N. E. Klepeis, and P. Switzer, "Analytical solutions to compartmental indoor air quality models with application to environmental tobacco smoke concentrations measured in a house," *Journal of the Air & Waste Management Association*, vol. 53, pp. 918–936, August 2003.

[8] R. V. Pound and W. D. Knight, "A radio frequency spectrograph and simple magnetic-field meter," *The Review of Scientific Instruments*, vol. 21, pp. 219–225, 1950.

[9] T. L. Viswanathan, T. R. Viswanathan, and K. V. Sane, "Study of marginal oscillator behavior," *IEEE Transactions on Instrumentation and Measurement*, vol. IM24, no. 1, pp. 55–61, 1975.

[10] A. Roberts, "Two new methods for detecting nuclear radiofrequency resonance absorption," *The Review of Scientific Instruments*, vol. 18, no. 11, pp. 845–848, Nov. 1947. [Online]. Available: http://dx.doi.org/10.1063/1.1740859

[11] B. V. Rollin, "Nuclear paramagnetism," *Reports on Progress in Physics*, vol. 12, no. 1, p. 22, 1949. [Online]. Available: http://stacks.iop.org/0034-4885/12/i=1/a=302

[12] D. Alexiev, M. Reinhard, L. Mo, and A. Rosenfeld, "A transient conductance technique for characterisation of deep-level defects in highly irradiated detector-grade silicon," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 434, no. 1, pp. 103–113, Sep 1999. [Online]. Available: http://dx.doi.org/10.1016/S0168-9002(99)00441-6

[13] A. Gauzzi, J. Le Cochec, G. Lamura, B. J. Jönsson, V. A. Gasparov, F. R. Ladan, B. Plaçais, P. A. Probst, D. Pavuna, and J. Bok, "Very high resolution measurement of the penetration depth of superconductors by a novel single-coil inductance technique," *Review of Scientific Instruments*, vol. 71, no. 5, pp. 2147–2153, 2000. [Online]. Available: http://scitation.aip.org/content/aip/journal/rsi/71/5/10.1063/1.1150597

[14] A. Warnick, L. R. Anders, and T. E. Sharp, "A marginal oscillator detector for ion cyclotron resonance spectrometers," *Review of Scientific Instruments*, vol. 45, no. 7, pp. 929–935, 1974. [Online]. Available: http://scitation.aip.org/content/aip/journal/rsi/45/7/10.1063/1.1686771

[15] G. L. Miller, M. Soni, and R. L. Fenstermacher, "A technique for investigating the properties of surfaces, thin films, and interfaces by means of a mechanical marginal oscillator," *Journal of Applied Physics*, vol. 53, no. 2, pp. 979–983, 1982. [Online]. Available: http://scitation.aip.org/content/aip/journal/jap/53/2/10.1063/1.330577

[16] L. Thomas, "Marginal oscillator for acoustic monitoring of curing of plastics," July 1988, US Patent 4,758,803. [Online]. Available: http://www.google.com/patents/US4758803

[17] F. N. H. Robinson, "An accurate measuring instrument for capacitance above 0.01 microfarad," *J. Phys. E: Sci. Instrum.*, vol. 21, no. 3, pp. 257–258, Mar 1988. [Online]. Available: http://dx.doi.org/10.1088/0022-3735/21/3/003

[18] D. S. Betts, D. T. Edmonds, B. E. Keen, and P. W. Matthews, "A susceptibility thermometer for use at very low temperatures," *J. Sci. Instrum.*, vol. 41, no. 8, pp. 515–516, Aug 1964. [Online]. Available: http://dx.doi.org/10.1088/0950-7671/41/8/414

[19] J. K. Stroble, S. Watkins, and R. Stone, "Biology-inspired sensor design," *IEEE Potentials*, vol. 28, pp. 19–24, 2009.

[20] D. H. B. Wicaksono, J. F. V. Vincent, G. Pandraud, G. Craciun, and P. J. French, "Biomimetic strain-sensing microstructure for improved strain sensor: fabrication results and optical characterization," *Journal of Micromechanics and Microengineering*, vol. 15, no. 7, p. 72, 2005. [Online]. Available: http://stacks.iop.org/0960-1317/15/i=7/a=011

[21] H. Bleckmann, H. Schmitz, and G. von der Emde, "Nature as a model for technical sensors," *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, vol. 190, pp. 971–981, 2004, 10.1007/s00359-004-0563-y. [Online]. Available: http://dx.doi.org/10.1007/s00359-004-0563-y

[22] C. Jiang, M. E. McConney, S. Singamaneni, E. Merrick, Y. Chen, J. Zhao, L. Zhang, and V. V. Tsukruk, "Thermo-optical arrays of flexible nanoscale nanomembranes freely suspended over microfabricated cavities as ir microimagers," *Chem. Mater.*, vol. 18, no. 11, pp. 2632–2634, May 2006. [Online]. Available: http://dx.doi.org/10.1021/cm060416x

[23] M. E. McConney, K. D. Anderson, L. L. Brott, R. R. Naik, and V. V. Tsukruk, "Bioinspired material approaches to sensing," *Advanced Functional Materials*, vol. 19, pp. 2527–2544, 2009.

[24] N. Krylov and N. N. Bogoliubov, *Introduction to Non-linear Mechanics*, ser. Annals of Mathematics Studies. Princeton University Press, 1943. [Online]. Available: http://books.google.com/books?id=XvLHko4Ru3QC

[25] T. J. Tyson, "Marginal oscillator conversion gain: Prediction, simulation, and experimental measurements," M.S. thesis, The Pennsylvania State University, August 2011.

[26] A. Gelb and W. Vander Velde, *Multiple-Input Describing Functions and Nonlinear System Design*, ser. McGraw-Hill Electronic Sciences Series. McGraw-Hill, 1968. [Online]. Available: http://books.google.com/books?id=aak8AAAAIAAJ

[27] A. I. Mees and A. R. Bergen, "Describing functions revisited," *IEEE Transactions on Automatic Control*, vol. 20, pp. 473–478, 1975.

[28] K. K. Zhang, "Conversion gain and sensitivity in marginal oscillators: continuous and sampled-data negative resistance converters," M.S. thesis, The Pennsylvania State University, 2011.

[29] MATLAB, *Version 7.10.0 (R2010a)*. Natick, Massachusetts: The MathWorks Inc., 2010.

[30] G. van Rossum, "Python reference manual," CWI (Centre for Mathematics and Computer Science), Amsterdam, The Netherlands, The Netherlands, Tech. Rep., 1995.

[31] S. van der Walt, S. C. Colbert, and G. Varoquaux, "The numpy array: A structure for efficient numerical computation," *Computing in Science & Engineering*, vol. 13, no. 2, pp. 22–30, 2011. [Online]. Available: http://scitation.aip.org/content/aip/journal/cise/13/2/10.1109/MCSE.2011.37

[32] J. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, May 2007.

[33] T. L. Heath, Ed., *The Works of Archimedes; The Method of Archimedes.* Dover, 1953.

[34] J. L. Schiano, March 2015, private communication.

[35] D. I. Hoult and R. E. Richards, "The signal-to-noise ratio of the nuclear magnetic resonance experiment," *Journal of Magnetic Resonance*, vol. 24, no. 1, pp. 71–85, October 1976.

[36] H. D. W. Hill and R. E. Richards, "Limits of measurement in magnetic resonance," *Journal of Physics E: Scientific Instruments*, vol. 1, no. 10, p. 977, 1968. [Online]. Available: http://stacks.iop.org/0022-3735/1/i=10/a=202

[37] B. P. Lathi, *Linear Systems and Signals.* Berkeley-Cambridge Press, 1992.

[38] F. C. Hill, June 2013, private communication.

[39] B. Quine, V. Tarasyuk, H. Mebrahtu, and R. Hornsey, "Determining star-image location: A new sub-pixel interpolation technique to process image centroids," *Computer Physics Communications*, vol. 177, no. 9, pp. 700–706, November 2007.

[40] I. M. Budaiwi and M. S. Al-Homoud, "Effect of ventilation strategies on air contaminant concentrations and energyconsumption in buildings," *International Journal of Energy Research*, vol. 25, pp. 1073–1089, 2001.

[41] B. Tashtoush, M. Molhim, and M. Al-Rousan, "Dynamic model of an HVAC system for control analysis," *Energy*, vol. 30, no. 10, pp. 1729–1745, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360544204004761

[42] T. Marsik and R. Johnson, "Use of a state space model to study the effect of outdoor air quality on indoor air in Fairbanks, Alaska," *Indoor Built Environment*, vol. 16, no. 6, pp. 538–547, August 2007.

[43] D. M. Bailey, D. Chu, D. L. Herron, D. M. Schwenk, C. W. Sohn, and D. M. Underwood, "Protection of Department of Defense facilities from airborne CBR threats," ERDC/CERL, Tech. Rep. SR-04-25, 2004.

[44] W. A. Bryden, "Immune building program," in *Collective Protection Conference*, Joint Program Management Office for Collective Protection. National Defense Industrial Association, June 2005.

[45] L. Meng, "Modeling contaminant spread and mitigation in the indoor environment," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 2013.

[46] W. J. Kowalski, W. Bahnfleth, and T. S. Whittam, "Filtration of airborne microorganisms: Modeling and prediction," *ASHRAE Transactions: Research*, vol. 105, no. 2, pp. 4–17, 1999.

[47] W. J. Kowalski, W. P. Bahnfleth, and A. Musser, "Modeling immune building systems for bioterrorism defense," *Journal of Architectural Engineering*, vol. 9, no. 2, pp. 222–227, June 2003.

[48] L. Meng, K. Nawaz, J. He, A. M. Jacobi, A. J. Nelson, and M. D. Ginsberg, "Fast estimation of leakage area in a multizone testing facility," in *Proceedings International High Performance Buildings Conference*, vol. 3496, July 2012, pp. 1–11.

[49] Global Industrial, http://www.globalindustrial.com/, May 2013. [Online]. Available: http://www.globalindustrial.com/

[50] Purolator Air Filters, http://www.purolatorair.com/, May 2013. [Online]. Available: http://www.purolatorair.com/

[51] E-Spin, http://espintechnologies.com/, May 2013. [Online]. Available: http://espintechnologies.com/

[52] V. Loftness, V. Hartkopf, B. Gurtekin, and R. Hitchcock, "Linking energy to health and productivity in the built environment evaluating the cost-benefits of high performance building and community design for sustainability, health and productivity," in *Proceedings of the Greenbuild Conference*, 2003.

[53] A. Nelson, M. Page, M. Ginsberg, and M. Rood, "Bench-scale evaluation of energy efficient air purification technologies," in *Proceedings of the Air and Waste Management Association Annual Conference*, no. 2011-A-548-AWMA, 2011.

[54] J. Needham, *Science and Civilization in China, Volume 2. History of Scientific Thought.* Cambridge University Press, 1986, vol. 2.

[55] A. Aksimentiev and K. Schulten, "Extending molecular modeling methodology to study insertion of membrane nanopores," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 13, pp. 4337–4338, 2004. [Online]. Available: http://www.pnas.org/content/101/13/4337.short

[56] F. Khalili-Araghi, J. Gumbart, P.-C. Wen, M. Sotomayor, E. Tajkhorshid, and K. Schulten, "Molecular dynamics simulations of membrane channels and transporters," *Current Opinion in Structural Biology*, vol. 19, no. 2, pp. 128 – 137, 2009, theory and simulation / Macromolecular assemblages. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0959440X09000293

[57] E. H. Lee, J. Hsin, M. Sotomayor, G. Comellas, and K. Schulten, "Discovery through the computational microscope," *Structure*, vol. 17, no. 10, pp. 1295–1306, Oct. 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0969212609003323

[58] R. Carr, J. Comer, M. D. Ginsberg, and A. Aksimentiev, "Atoms-to-microns model for small solute transport through sticky nanochannels," *Lab Chip*, vol. 11, no. 22, pp. 3766–3773, 2011. [Online]. Available: http://dx.doi.org/10.1039/C1LC20697D

[59] W. A. Yost and D. W. Nielsen, *Fundamentals of Hearing*. Holt, Rinehart and Winston, 1977, chapter 8.

[60] I. Voiculescu, M. Zaghloul, R. McGill, E. Houser, and G. Fedder, "Electrostatically actuated resonant microcantilever beam in CMOS technology for the detection of chemical weapons," *IEEE Sensors Journal*, vol. 5, no. 4, pp. 641–647, Aug. 2005.

[61] G. J. M. Krijnen, M. Dijkstra, J. J. van Baar, S. S. Shankar, W. J. Kuipers, R. J. H. de Boer, D. Altpeter, T. S. J. Lammerink, and R. Wiegerink, "MEMS based hair flow-sensors as model systems for acoustic perception studies," *Nanotechnology*, vol. 17, no. 4, p. S84, 2006. [Online]. Available: http://stacks.iop.org/0957-4484/17/i=4/a=013

[62] M. C. Göpfert and D. Robert, "The mechanical basis of drosophila audition," *Journal of Experimental Biology*, vol. 205, no. 9, pp. 1199–1208, May 2002. [Online]. Available: http://jeb.biologists.org/content/205/9/1199.abstract