Sponsors

Los Alamos National Laboratory
The National Nuclear Security Agency (NNSA) Nuclear Weapons Program, the Educational Program Office and Department of Energy, Science Education and Nuclear Weapons Programs (Defense Programs)

The University of New Mexico
Department of Physics and Astronomy and Extended University Graduate and Upper Divisions Programs, and The National Science Foundation: Research Experience for Undergraduates (REU)

Contact
Lee A. Collins, Theoretical Division, Mail Stop B212, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, lac@lanl.gov

Publication
Sharon Mikkelson, Illustrator/Designer IRM-CAS/T-DO
Denise Sessions, Writer/Editor IRM-CAS/T-DO

September 2006/LA-UR-06-5939
Table of Contents

• Preface

• Class of 2006

• Mentors and Students

• Lectures

• Photos - Selected Scenes from Summer 2006

• Titles of Papers

• Student Papers
Preface

This year marked the sixteenth session of the Los Alamos Summer School, a joint educational project between the University of New Mexico and the Los Alamos National Laboratory. The National Science Foundation provides funding to the University side as a Research Experience for Undergraduates (REU) site, while National Nuclear Security Agency (NNSA) supports the Laboratory component through the Nuclear Weapons Program. The School recruits nationwide and primarily focuses on upper-level undergraduate students. Thirteen students, representing a diverse set of backgrounds and geographic locations, participated in this year’s program.

The School employs a dual track philosophy of lectures and student research projects to stimulate students to pursue a research career in the sciences. The lectures, an eclectic group of distinguished speakers from UNM, LANL, and other institutions, focused on the latest developments in many fields, giving students a broad view of the great diversity of scientific topics and research opportunities. On the other hand, the research project, mentored by a staff scientist for the full ten-week term of the School, concentrated upon a specific scientific problem and allowed the student to experience scientific investigation first-hand. All the lecturers and mentors, listed on the following pages, were volunteers. The lecture and mentor topics included atomic, molecular, optical, condensed-matter, and plasma physics, as well as astrophysics, biophysics, and chemistry.

This book, a collection of research reports from the students, reflects the emphasis we have placed on significant research experiences. Because of the brevity of a ten-week session, these papers are essentially descriptions of works in progress, rather than polished final documents. Nevertheless, we believe they demonstrate clearly the remarkable and proficient accomplishments of all the students.

LANL
Lee Collins
Norm Magee
James Colgan

UNM
Sally Seidel
Bittner, Jonathan M. 
Cardenas, Edna S. 
Cecil, Jeffrey J. 
Lewis, George H. 
Looker, Quinn M. 
Raach, Catherine (Kate) M. 
Saelim, Michael J. 
Sharpnack, Lewis (Lee) L. 
Soong, Ken 
Streufert, Benjamin D. 
Teplitskiy, Mikhail (Misha) A. 
Uzgil, Bade D. 
Warner, Jacob A. 

Yale University 
University of New Mexico 
Illinois Institute of Technology 
Columbia University 
University of Missouri-Rolla 
University of Minnesota 
Michigan State University 
Kent State University 
Cornell University 
University of Virginia 
Rice University 
University of California, Los Angeles 
University of Wisconsin
<table>
<thead>
<tr>
<th>Mentor</th>
<th>Student</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liu, Siming; T-6</td>
<td>Bittner, Jonathan M.</td>
</tr>
<tr>
<td>Kenyon, Garrett; P-21</td>
<td>Cardenas, Edna S.</td>
</tr>
<tr>
<td>Makaruk, Hanna E.; P-22</td>
<td>Cecil, Jeffrey J.</td>
</tr>
<tr>
<td>Conradson, Steve; MST-8</td>
<td>Lewis, George H.</td>
</tr>
<tr>
<td>MacDonald, Elizabeth A.; ISR-8</td>
<td>Looker, Quinn M.</td>
</tr>
<tr>
<td>Crooker, Scott A.; MPA-NHMFL</td>
<td>Raach, Catherine (Kate) M.</td>
</tr>
<tr>
<td>Hlavacek, William S.; T-CNLS</td>
<td>Saelim, Michael J.</td>
</tr>
<tr>
<td>Kunde, Gerd J.; P-25</td>
<td>Sharpnack, Lewis (Lee) L.</td>
</tr>
<tr>
<td>Ziock, Hans-Joachim; EES-6</td>
<td>Soong, Ken</td>
</tr>
<tr>
<td>Colgan, James; T-4</td>
<td>Streufert, Benjamin D.</td>
</tr>
<tr>
<td>Wang, Zhehui; P-24</td>
<td>Teplitskiy, Mikhail (Misha) A.</td>
</tr>
<tr>
<td>Johnson, N. Gulbahce; T-7</td>
<td>Uzgil, Bade D.</td>
</tr>
<tr>
<td>Andrews, Malcolm J.; CCS-2</td>
<td>Warner, Jacob A.</td>
</tr>
<tr>
<td>Author</td>
<td>Title</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Alexander Heger; T-6</td>
<td><em>I Evolution of Stars; II Creation of the Elements; III The First Stars in the Universe</em></td>
</tr>
<tr>
<td>Steve Conradson; MST-8</td>
<td><em>Collective Phenomena, Nanoscale Heterogeneity, and Aging Mechanisms in Delta-Retained PuGa Alloys</em></td>
</tr>
<tr>
<td>Malcolm Andrews; CCS-2</td>
<td><em>Three Core Lectures on Buoyancy Driven Turbulent Mixing</em></td>
</tr>
<tr>
<td>Jolante Van Wijk; EES-11</td>
<td><em>Mantle Plumes: A Journey to the Centre of the Earth?</em></td>
</tr>
<tr>
<td>Lakshman Prasad; ISR-2</td>
<td><em>Geometry of Shapes and Images</em></td>
</tr>
<tr>
<td>Haiyan Wang; MST-STC</td>
<td><em>Heterogeneous Interface in Epitaxial Thin Films</em></td>
</tr>
<tr>
<td>Steven Brumby</td>
<td><em>Space and Remote Sensing Sciences; machine learning/pattern recognition for a range of basic science and national security applications</em></td>
</tr>
<tr>
<td>Ed Fenimore; ISR-1</td>
<td><em>Gamma-ray Bursts</em></td>
</tr>
<tr>
<td>David Hobart; C-AAC</td>
<td><em>Radiochemistry: history, and fundamentals through the synthesis of man-made actinide elements and superheavy elements</em></td>
</tr>
<tr>
<td>Hans Ziock; EES-6</td>
<td><em>1) The Scale of the CO₂ Issue. 2) The Protocell Assembly Project. 3) Advanced Fossil Energy Concepts</em></td>
</tr>
<tr>
<td>Sigrid Close; ISR-2</td>
<td><em>Impact of Meteoroids on Earth’s Atmosphere</em></td>
</tr>
<tr>
<td>Kendra Vant; P-21</td>
<td><em>Building a Quantum Simulator From Trapped Ions</em></td>
</tr>
<tr>
<td>Ilya Nemenman; CCS-3</td>
<td><em>Biophysics of Cellular Regulatory Networks</em></td>
</tr>
<tr>
<td>Sally Seidel; UNM</td>
<td><em>Particle Physics</em></td>
</tr>
<tr>
<td>Mark Petersen; CCS-2</td>
<td><em>Computational Fluid Dynamics</em></td>
</tr>
<tr>
<td>Garret Kenyon; P-21</td>
<td><em>Computational Neuroscience</em></td>
</tr>
<tr>
<td>Wilbert Weijer; CCS-2</td>
<td><em>The Ocean Circulation and the Climate System: global warming or next ice age?</em></td>
</tr>
<tr>
<td>Vivien Zapf; MST-NHMFL</td>
<td><em>Quantum Magnetism and the National High Magnetic Field Laboratory</em></td>
</tr>
<tr>
<td>Martin Piltch; MST-6</td>
<td><em>Lasers and Quantum Electronics</em></td>
</tr>
<tr>
<td>Tim Thomas; UNM</td>
<td><em>Grid Computing for Scientists</em></td>
</tr>
<tr>
<td>Praveen Ramaprabhu; CNLS</td>
<td><em>New Directions in Fluids Research</em></td>
</tr>
<tr>
<td>Daniel James; University of Toronto</td>
<td><em>Neat stuff in Quantum Mechanics</em></td>
</tr>
<tr>
<td>Susan Atlas; UNM</td>
<td><em>Density Functional Theory</em></td>
</tr>
<tr>
<td>Gerd Kunde; P-25</td>
<td><em>The Quark-Gluon Plasma</em></td>
</tr>
<tr>
<td>Brian O’Shea; T-6</td>
<td><em>Astrophysics and Cosmology</em></td>
</tr>
<tr>
<td>Roberto Mancini</td>
<td><em>University of Nevada-Reno; Plasma Spectroscopy</em></td>
</tr>
<tr>
<td>Stephen Becker; X-2</td>
<td><em>Production of Heavy Transuranic Isotopes in Thermonuclear Explosions</em></td>
</tr>
<tr>
<td>Xinghang Zhang</td>
<td><em>Texas A&amp;M, Interface Induced Strengthening in Nanolayered Thin Films</em></td>
</tr>
<tr>
<td>Elizabeth MacDonald; ISR-1</td>
<td><em>Introduction to Experimental Space Science</em></td>
</tr>
</tbody>
</table>
2006 volume 8
Los Alamos Summer School Photographs
Flares and Electron Acceleration Near the Galactic Supermassive Black Hole
Jonathan M. Bittner

Computational Modeling of the Neural Network in the Retina of a Salamander
Edna S. Cardenas

Analysis of Proton Radiography Images of Shock Melted and Damaged Metals
Jeffrey J. Cecil

Covalency of Uranium-Ligand Bonding
George H. Lewis

Laboratory Calibration of a Space Plasma Spectrometer
Quinn M. Looker

Magneto-optical spectroscopy of semiconductor quantum dots containing single magnetic atoms
Kate M. Raach

TCR-mediated T cell activation by kinetic proofreading and intracomplex transphosphorylation
Michael J. Saelim

Separating the Z^0 from Heavy Quark Background in Relativistic Heavy Ion Collisions
Lewis L. Sharpnack

Exploring the Feasibility of Using Dynamic Light Scattering as a Method of Assessing Protocell Replication
Ken Soong

Evaluating Semi-Empirical Models for Electron Impact Double Ionization Cross Sections for Positive, Neutral, and Negative Ions
Benjamin D. Streufert

Are there drift waves in the Flowing Magnetized Plasma (FMP) experiment?
Mikhail A. Teplitskiy

Identification of a lethality core in the metabolic network of Escherichia coli
Bade D. Uzgil

Reproducing Numerical Simulations More Efficiently and More Rapidly for Turbulent Rayleigh-Taylor Instability
Jacob A. Warner

MPI: Exploring the Advantages, Limitations, and Applications
Jacob A. Warner
Flares and Electron Acceleration Near the Galactic Supermassive Black Hole

Jonathan M. Bittner, Siming Liu, Christopher L. Fryer, and Vahé Petrosian

ABSTRACT

There is compelling evidence that a black hole of about 4 million solar masses lies in the center of our Milky Way galaxy. Flares in the near infrared (NIR) and X-ray bands are routinely observed from that direction, specifically from the compact radio source called Sagittarius A*. It is believed that they are produced by hot plasmas within a few Schwarzschild radii of the black hole. Recently, a correlation was observed between spectral index and flux during an NIR flare, which provides a means to conduct a detailed investigation of the plasma heating and radiation processes. A code has been developed to study the electron distribution function under the influence of a turbulent magnetic field in a hot, collision-less plasma. The magnetic field, presumably generated through instabilities in the accretion flow, can both heat the plasma via resonant wave-particle coupling and cool the electrons via synchrotron radiation. These two effects have been modeled in a previously published differential equation, which has now been solved for a time-dependent solution using a finite difference method. The radiation associated with the electron acceleration was also calculated with numerical integration. This model was found to recover the observed correlation, and predicts simultaneous flare activity in the NIR and X-ray bands. The results can be compared with observation, which would also provide a test of the model and a measurement of the four basic model parameters, namely: magnetic field, source size, temperature, and number density.

Subject headings: acceleration of particles — black hole physics — Galaxy: center — plasmas — radiation mechanisms: thermal— turbulence

1Physics Department, Yale University, New Haven, CT 06520; jonathan.bittner@yale.edu
2Los Alamos National Laboratory, Los Alamos, NM 87545; liusm@lanl.gov
3Physics Department, The University of Arizona, Tucson, AZ 85721; fryer@lanl.gov
4Center for Space Science and Astrophysics, Department of Physics and Applied Physics, Stanford University, Stanford, CA 94305; vahe@astronomy.stanford.edu
1. Introduction

Several sources of evidence have led astrophysicists to infer that the compact radio source at the galactic center, Sagittarius A* (Sgr A*), is produced by a radiating plasma accretion disk surrounding a $M \sim 3 - 4 \times 10^6 M_\odot$ mass supermassive black hole (Schödel et al. 2002; Ghez et al. 2004). Near-infrared (NIR) and X-ray flares are also frequently observed from that region, on the order of twice/once per day, respectively, and in the NIR band a correlation between spectral index and flux has been observed (Gillessen et al. 2006). The exact cause of these flares and the associated correlation are not well understood, but what is known is that the radiation from the flares is produced by the radiative processes of accelerating electrons moving close to the speed of light, within a few Schwarzschild radii of the black hole. Specifically, synchrotron radiation produces NIR emissions and synchrotron self-Comptonization (SSC) produces X-ray flares (Baganoff et al. 2001, 2003; Genzel et al. 2003; Belanger et al. 2006; Eckart et al. 2004, 2006; Yusef-Zadeh et al. 2006a).

A model of the electron acceleration has already been developed to calculate this radiation, but it has not previously been shown whether or not a variation of the model can reproduce the observed correlation. Reproducing that correlation is the goal of this paper.

Understanding the details of the electron acceleration model is of some importance to understanding our results. We assume that acceleration of the electrons is caused by turbulent plasma waves propagating through an accretion torus. Some unknown instabilities cause these waves, which in turn causes an infalling of the plasma and therefore a release of gravitational energy (Balbus & Hawley 1991; Tagger & Melia 2006). Assuming there is no large-scale magnetic field, the turbulent field $B$ plays the role of both heating and cooling the electrons. The heating is presumably caused by a resonant wave-particle interaction, and the cooling is caused by radiation, which are possible simultaneously because their rates have different energy dependencies. Previous work by SL showed that the steady-state electron spectrum can be approximated by a relativistic Maxwellian distribution with a “temperature” $\gamma c m_e c^2$ depending on the plasma density $n$ and the coherent length of the magnetic field $B$, which should be comparable to the size of the flare region $R$. This implies that simultaneous observations of flux and spectral index in the NIR and X-ray bands (four total measurements) can be used to determine the four model parameters mentioned above ($B$, $n$, $\gamma c m_e c^2$, and $R$ ) and to test the model (Liu et al. 2006a, 2006b).

Several scenarios have been suggested to explain the observed correlation between flux and spectral index. Our results show that the dynamical processes of plasma heating and cooling are the most likely mechanism and that the observation requires that the magnetic field be anti-correlated with the electron temperature. Since the evolution of the magnetic field depends on the turbulence generation mechanism, which needs to be addressed via
MHD simulations, we calculate the evolution of the electron distribution function and the corresponding radiation spectrum for a cooling and a heating phase, where the magnetic field is set to be inverse proportional to the mean electron energy. The model naturally recovers the observed correlation between the NIR flux and spectral index and predicts the X-ray emission characteristics accompanying the NIR flares. Simultaneous flare observations in the NIR and X-ray bands can readily test the model and uncover the underlying physical processes producing these flares.

Demonstrating that our model explains the correlation constitutes a significant advance towards our understanding of the plasma physics close to the black hole. If a detailed understanding of the flares were coupled with general relativistic MHD and ray-tracing simulations of the black hole accretion and radiation transfer, it might be possible to constrain the spin of the black hole. This is a major goal of future research into this area.

It must be noted that this paper bears considerable resemblance to a paper currently submitted to Astrophysical Journal, with the same authors, under the title “Correlation between Flux and Spectral Index during Flares in Sagittarius A*”. That paper describes exactly the same results as this one. The version presented here is part of the Los Alamos Summer School journal of student papers and has been rewritten accordingly. The ApJ version of the paper is slightly more technical, contains a few additional results, and a format intended for that publication.

In § 2, we will describe kinetic equation for electron acceleration and its time-dependent solutions. In § 3, we apply our model to Sagittarius A* by calculating the radiation from a flaring plasma, given a magnetic field evolution. A detailed discussion of the results, limitations, and implications of our results are discussed in § 4.

2. Time Dependent Solutions of Electron Acceleration by Plasma Waves

The foundational work for our calculations has already been laid by previous studies of the electron acceleration in a turbulent magnetic field (Liu et al. 2006b). It has been shown that regardless of the details of the resonant wave-particle coupling, the evolution of the electron distribution function \( N(\gamma, t) \) can be described by the following kinetic equation:

\[
\frac{\partial N}{\partial t} = \frac{\partial}{\partial \gamma} \left[ \frac{\gamma^2}{\tau_{ac}} \frac{\partial N}{\partial \gamma} + \left( \frac{\gamma^2}{\tau_0} - \frac{2\gamma}{\tau_{ac}} \right) N \right],
\]

where the synchrotron cooling and acceleration times are given, respectively, by

\[
\tau_{syn}(\gamma) = \frac{\tau_0}{\gamma} \equiv \frac{9m_e^3c^5}{4e^4B^2\gamma},
\]

\[
\tau_{ac} \equiv 2\gamma^2/\langle \Delta \gamma \Delta \gamma/\Delta t \rangle = C_13Rc/v_A^2 = 12\pi C_1nm_p\gamma cR/B^2,
\]
the Alfvén velocity $v_A = B/(4\pi nm_p)^{1/2}$, $C_1$ is a dimensionless quantity of order 1, and we have assumed that there is no particle escape and therefore no return current in association with the electron acceleration. The size of the flaring region, the magnetic field and gas density are given respectively by $R$, $B$ and $n$. The scattering mean free path of the electrons, $C_1 R$, is comparable to the source size. $m_e$, $m_p$, $c$, $e$, and $\gamma$ are the electron mass, proton mass, speed of light, elemental charge unit and the Lorentz factor of the electron, respectively.

This differential equation always yields a steady state solution for a given $\tau_0$ or $\gamma_c$, which is simply $N(\gamma) = (\gamma^2/2\gamma_c^3) \exp(-\gamma/\gamma_c)$, with $\gamma_c = \tau_0/\tau_{ac} = 3m_e^3c^4/16\pi C_1 e^4m_p n R$. $\gamma_c$ is just the steady-state electron temperature. The time dependent evolution of this function is of more interest, however, so we used the semi-implicit Chang-Cooper method to solve the kinetic equation numerically from an initial condition of a relativistic Maxwellian distribution. The steady state solution provided a useful check of our results, which are presented below.

![Evolution of the electron distribution function $N(\gamma, t)$](image_url)

Fig. 1.— Evolution of the electron distribution function $N(\gamma, t)$. Left panel is a heating phase, with $\gamma_i = 10$ and $\gamma_i = 200$, right panel is a cooling phase with the temperatures interchanged. Time steps are represented by individual curves, and the arrow denotes the direction of evolution. The time steps are $4(i/14)^2\tau_{ac}$, where $i$ is an integer ranging from 0 to 14. The thick line represents a time step of $t = \tau_{ac}$.

One notable feature of this evolution is that particle distributions are already close to
the steady-state solution after the time step \( t = \tau_{ac} \). Because the synchrotron cooling (with a timescale \( \gamma_c \tau_{ac}/\gamma \)) dominates above \( \gamma_c \), the particle distribution evolves faster than in the heating phase (in unit of \( \tau_{ac} \)). We also note the pileup of electrons to a nearly monotonic distribution in the early section of this cooling phase, due to more rapid cooling of higher energy electrons. Later the distribution is broadened by the diffusion term in the kinetic equation (the first term on the right hand side of eq. [1]).

We used this result in turn to create a correlation between \( N(\gamma, t) \) and electron spectral index, \( \beta \). This cannot be compared with observation directly, but nevertheless provides a check to see if our results are reasonable.

![Fig. 2.— Correlations between \( N \) and \( \beta = d \ln N/d \ln \gamma \) at \( \gamma = 50 \) for the two runs in 1. The solid line is for the heating phase. The dashed line is for the cooling phase. Note that the correlation is quite different for the two phases.](image)

In reading this plot, one may consider the time \( t \) as a parametric variable which traces out the correlation. The rising of \( \beta \) to above 2 is due to the pileup of electrons in the early cooling phase. The correlations are quite different for the two phases and energies. For a given magnetic field, these correlations mimic similar correlations in the synchrotron radiation, which we investigate in § 3.

To characterize the evolution of \( N(\gamma, t) \) in general, we next consider the evolution of the mean energy \( < \gamma > (t) = \int \gamma N(\gamma, t) d\gamma / \int N(\gamma, t) d\gamma \). Because the heating of electrons
by turbulence depends on the derivative of \( N(\gamma, t) \) with respect to \( \gamma \) via the diffusion term, the evolution of \( \langle \gamma \rangle \) can be very complicated. Fortunately for any smooth function \( N(\gamma) \) one may approximate this derivative term with a heating term, i.e. the total energy change rate
\[
\dot{\langle \gamma \rangle} \simeq -\frac{\langle \gamma \rangle^2}{\tau_{ac}\gamma_c} + \frac{(2 + a) \langle \gamma \rangle}{\tau_{ac}},
\]
where a “\( \cdot \)” indicates a derivative with respect to time and \( a \) is a parameter to be determined by the steady state solution. Then we have
\[
\langle \gamma \rangle_t = \frac{(2 + a)\gamma_c\gamma_i}{\gamma_i - (\gamma_i - \gamma_c) \exp[-(2 + a)t/\tau_{ac}]}.
\]
In the steady state, \( \langle \gamma \rangle_t = 3\gamma_c \), therefore \( a = 1 \).

Fig. 3.— Left: Evolution of the mean energy of the electrons in the heating phase. The dashed line indicates the theoretical result given by equation (6) with \( a = 1.6 \). Right: Same as the left panel but for the cooling phase with \( a = 1.0 \).

However the above approximation is not good enough for the intermediate steps of the heating phase when the particles have a broad distribution (Fig. 1) and the heating by the diffusion term is more efficient. To fit the numerical results we set the “\( a \)” in the numerator of equation (5) equal to one to be consistent with the steady state solution and leave the
Fig. 4. — *Left:* The ratio of the mean energy of the electrons to the theoretical value as a function of time for several initial and final temperatures in the heating phase. The solid lines have an initial temperature of 10 and final temperatures of 200, 500, 1000. The dashed lines have a final temperature of 200 and initial temperatures of 20, 50, and 100. The relative error (the difference between the numerical and theoretical results divided by the theoretical values) increases with the increase of the ratio of the final to initial temperature (the dynamical range). *Right:* Same as the left panel but for the cooling phase. The line with the greatest relative error corresponds to a cooling from 1000 to 10. The other lines have an initial temperature of 200 and final temperatures of 10, 20, 50, and 100. The relative error also increases with the dynamical range.

"a" in the denominator as a free parameter:

\[
< \gamma > \gamma = \frac{3 \gamma_c \gamma_i}{\gamma_i - (\gamma_i - \gamma_c) \exp[-(2 + a)t/\tau_{ac}]},
\]

We find that \( a = 1.6 \) and 1.0 give good fits to the \(< \gamma >\) evolutions of the heating and cooling phases, respectively. Figure 3 (previous page) shows the evolution of \(< \gamma >\) and the corresponding fits for the heating (left) and cooling (right) phases. Figure 4 above gives the ratio of \(< \gamma >\) to the fit values for several initial and final temperatures. The relative error is within 20% for the heating phase and within 10% for the cooling. The error also increases with the increase of the dynamical range, which is the ratio of the initial temperature to steady state temperature for the cooling phase and vice versa for the heating phase.
Equation (6) therefore gives a good description of the evolution of $<\gamma>$ under the influence of a turbulent magnetic field.

We note that equation (6) can be generalized to address the heating and cooling of collisionless plasma by turbulent plasma waves. The heating time $\tau_{ac}$ (see eq.[3]) is usually a function of time. Taking into account the effects discussed in the previous paragraph, equation (5) leads to

$$<\gamma>_t = \frac{3\gamma_c \gamma_i}{\gamma_i - (\gamma_i - \gamma_c) \exp \left[ -(2 + a) \int dt/\tau_{ac(t)} \right]}.$$  \hfill (7)

MHD simulations can give the time evolution of $B$, its coherent length, and $n$. Equation (7) can then be used in these simulations to address the heating and cooling of electrons. In §3 we study the SSC emission of these electrons during flares in Sagittarius A*.

3. Synchrotron Emission and SSC

Of course, the only way to observe any correlation or spectrum observationally is by measuring the radiation produced by the accelerating electrons. This can be performed with a few simple equations taken from radiative astrophysics texts.

If we assume the particle distribution is isotropic with respect to the turbulent magnetic field, the synchrotron flux density and emission coefficient at frequency $\nu$ are given, respectively, by (Pacholczyk 1970)

$$F_\nu(\nu) = \frac{4\pi R^3}{3D^2} \mathcal{E}_\nu, \quad \mathcal{E}_\nu(\nu) = \frac{\sqrt{3}e^3}{4\pi m_ec^2} B n \int_0^\infty \int_0^1 d\gamma d\mu(1 - \mu^2)^{1/2} N(\gamma) F(x),$$  \hfill (8)

where

$$x = \frac{\nu}{\nu_c} \equiv \frac{4\pi m_ec \nu}{3eB(1 - \mu^2)^{1/2}\gamma^2},$$  \hfill (9)

$$F(x) = x\int_x^\infty K_{5/3}(z)dz \simeq 2.1495x^{1/3}e^{-2x} + 1.348x^{1/2}e^{-x} - 0.249x^{-1.63},$$  \hfill (10)

$D$, $\mu$, and $K_{5/3}$ are the distance to the Galactic Center, the cosine of the angle between the magnetic field and line of sight, and the corresponding Bessel function, respectively, and the approximation for $F(x)$ is accurate within 8%. We therefore have the spectral index in a given narrow frequency range $\alpha \equiv d\ln(\nu F_\nu)/d\ln \nu$.

As shown by Liu et al. (2006b), most of the synchrotron radiation is emitted in the optically thin region for flares in Sagittarius A*. For a uniform spherically symmetric source,
the self-Comptonization flux density is then given by (Blumenthal & Gould 1970)

\[ F_X(\nu) \simeq \frac{\pi e^4 \nu n R}{6m_e^2 c^4} \int_0^\infty d\gamma \frac{N(\gamma)}{\gamma^4} \int_{\nu/4\gamma^2}^\infty d\nu' \frac{F_{\nu'}(\nu')}{\nu'^3} \left( 2\nu \ln \frac{\nu}{4\gamma^2 \nu'} + \nu + 4\gamma^2 \nu' - \frac{\nu^2}{2\gamma^2 \nu'} \right), \tag{11} \]

Similarly one can define the X-ray spectral index \( \alpha_X \equiv d \ln (\nu F_X)/d \ln \nu \). These constitute the fundamental set of integrals we use to compute emitted radiation from the electron spectrum.

### 3.1. Flux and Spectral Index Correlations with a Constant Magnetic Field

For each time step of our electron evolution, we compute a variant of integral giving the synchrotron emission:

\[ \epsilon_\nu(\nu, t) \equiv \frac{F_\nu(\nu, t)\sqrt{3}m_ec^2D^2}{e^2 BnR^3} = \int_0^\infty d\gamma \int_0^1 d\mu (1 - \mu^2)^{1/2}N(\gamma, t)F(x) \tag{12} \]

The results of this integration are displayed below in Figure 5.

Fig. 5.— Evolution of \( \epsilon_\nu \) for the two runs shown in Figure 1. The time steps are the same as in Figure 1 and the magnetic field is \( B = 100 \) G. The left and right panels correspond to the heating and cooling phases, respectively.
$\epsilon_{\nu}(\nu)$ is less dramatic because the synchrotron spectrum is dominated by emission from the more energetic electrons, and the radiation spectra are very similar to thermal synchrotron spectra. However, the difference between the heating and cooling phases is obvious.

Two predicted correlations between spectral index and flux are shown in Figure 6 below. The left panel corresponds to $\nu = 1.4 \times 10^{14}$ Hz and the right to $1.4 \times 10^{13}$ Hz. The dotted lines indicate the observed correlations with different background subtraction methods (Gillessen et al. 2006). The cooling phase fits the observations marginally, while the heating phase predicts a correlation much weaker than is observed.

![Figure 6](image)

Fig. 6.— Correlation between the normalized flux density $\epsilon_{\nu}$ and the spectral index $\alpha$ at $\nu = 1.4 \times 10^{14}$ Hz (left) and $1.4 \times 10^{13}$ Hz (right) for the two runs in Figure 5. The solid and dashed lines are for the heating and cooling phases, respectively. The dotted lines give the observational results, with various background subtraction methods.

Since the shape of electron spectrum evolution is relatively independent of the initial and final temperatures, this correlation also does not change dramatically with respect to these parameters. Therefore, we are forced to conclude that a plasma with a constant magnetic field cannot explain the observed correlation, except marginally in the cooling phase.
3.2. Flux and Spectral Index Correlations with a Variable Magnetic Field

In order to improve our predictions of the observed correlation in the NIR band, we chose to vary the magnetic field $B$ over the course of the flare. We used an adiabatic approximation to investigate, before running any simulations, what would be a prudent way to vary the magnetic field. Figure 7 shows the correlation predicted in the adiabatic case, for a variety of functional dependencies. In the right panel, the solid line gives a better fit to the observations than all models discussed above. This line was created with a magnetic field inversely proportional to field. We note, however, that for $\alpha$ increasing from -4 to 1, the temperature of the plasma has to increase by about three orders of magnitude (Liu et al. 2006b).

Fig. 7.— Left: Correlation between $\epsilon_\nu$ and $\alpha$ for electrons with a relativistic Maxwellian distribution. In this case, the magnetic field is constant. The dotted lines indicate the observed results, which clearly lie below the theoretical curve. Right: Correlations with $B = 1/T$ (solid line), and $T^{-4/3}$ (dashed line), which are more in agreement with observation.

The time-dependent case does not always have a well defined temperature, however, since the relativistic Maxwellian distribution is not maintained exactly during the evolution. Choosing a magnetic field inverse proportional to $< \gamma >$ during the evolution of the flare, is a close approximation to this effect, and the one used in the analysis that follows. It should also be noted that varying the magnetic field in a time-dependent case changes the
relevant time scale of the system, $\tau_{ac}$. This makes the numerical methods more difficult, and in practice required using an adaptive time step. Another issue is that, compared with the heating and cooling processes studied in § 2, one more parameter is needed to obtain the solutions: the initial magnetic field. We chose this parameter such that it would create a range of magnetic fields that encompassed a region we believe to physically reasonable, from $B = 10$ to $B = 100$.

Regardless of these difficulties, our results are shown below in Figure 8:

![Figure 8](image-url)

Fig. 8.— *Left:* Correlation between the flux $\epsilon_\nu B$ and spectral index $\alpha$ at $\nu = 1.4 \times 10^{14}$ Hz for the heating (solid line) and cooling (dashed line) phases, where the magnetic field $B$ is chosen to be inverse proportional to the mean energy of the electrons $< \gamma >$. The initial magnetic field, the initial and final temperatures are 4 Gauss, 2000 and 20 and 200 Gauss, 10 and 1000 for the cooling and heating phases, respectively. The dotted lines indicate the observational results. The scale of the heating phase flux is indicated on the upper axis. *Right:* Associated magnetic field evolution for the heating (solid) and cooling (dashed) phases.

A relatively large temperature change is chosen to produce significant variation of the synchrotron spectral index. The dashed line in the right panel gives the evolution of the magnetic field. Although the initial magnetic field is relatively low, the steady state magnetic field is 400 Gauss for the cooling phase, which is much higher than the typical value for the quiescent state.
Now, we see that the cooling phase agrees with observation and the heating phase agrees with the correlation marginally: this is a significant improvement over the result with a constant magnetic field, enough so that feel comfortable saying that these results agree with observation on the whole.

Selecting a given frequency allows us to pick out predicted light curves for an NIR flare. They are plotted below:

![Light curves for the flux (left) and spectral index (right) at 1.4 × 10^{14} Hz, for a cooling phase (dashed) and heating phase (solid). The scale of the heating phase flux is indicated on the right and top axis. Points are full curves plotted in Figure 10.](image)

The dashed lines in Figure 9 show the light curves of the flux (left) and spectral index (right) for the cooling phase. This comparison gives us a more visual picture of the correlation between these two quantities. Note that the correlation between the flux and spectral index mostly occurs between 130 to 160 minutes (again in the cooling phase). Before 130 minutes, the system evolves slowly due to the relatively lower magnetic field (< 20 Gauss). Because the magnetic field in the accretion torus in Sagittarius A* is about a few tens of Gauss in the quiescent state, the evolution before 100 minutes is likely irrelevant to flares in Sagittarius A* since the magnetic field is below 10 Gauss during this period. From 100 to 170 minutes the model predicted correlation is quite in line with observations. The model not only produces the correct correlation slope, but also recovers the typical variation timescale of a few tens
of minutes. After 170 minutes, the system reaches a steady state.

The dips near 160 mins in the flux and spectral index light curves are related to the turn around of the spectral index and flux correlation before reaching the steady-state. These are caused by the sharp cutoff (sharper than an exponential cutoff) of the electron spectra in the cooling phase, which makes the spectra softer than that of a relativistic Maxwellian distribution. To produce a flare state with $\alpha = -0.5$ and $F_{\nu} = 5 \text{ mJy}$, the total number of electrons involved in the flare $4\pi n R^3/3 = 2.6 \times 10^{41}$, giving rise to $R = 5.5 \times 10^{10} C_1^{1/2} \text{ cm}$ and $n = 3.7 \times 10^8 C_1^{-3/2} \text{ cm}^{-3}$. These values agree with results of previous studies (Liu et al. 2006a, 2006b).

As stated previously, the heating phase fits the observed correlation marginally (the solid line in the left panel of Figure 8). The heating phase proceeds much faster than the cooling phase due to the relatively high initial magnetic field and final temperature. The flux reaches the peak value within $\sim 6$ minutes, after which the spectral index saturates near 0.2, and both the magnetic field and flux decrease slowly. To produce even harder spectra, one has to increase the value of $B\gamma_c^2$ at the steady state dramatically (Liu et al. 2006b), making the heating phase proceed on an even shorter timescale ($\propto 1/B^2\gamma_c$). A very weak initial magnetic field ($< 1$ Gauss) and a very high temperature ($\gg 1000$) are required to produce hard NIR spectra and a flux rising time of a few minutes. To produce a flare state with $\alpha = -0.5$ and $F_{\nu} = 5 \text{ mJy}$, the total number of electrons involved in the flare $4\pi n R^3/3 = 4.7 \times 10^{42}$, which is more than ten times larger than that in the cooling phase. This leads to $R = 1.6 \times 10^{12} C_1^{1/2} \text{ cm}$ and $n = 2.5 \times 10^5 C_1^{-3/2} \text{ cm}^{-3}$.

These results are promising, so we proceed to calculate our predicted X-ray emissions from our best fit model: cooling with an inversely proportional magnetic field.

### 3.3. Flux and Spectral Index Correlation in the X-ray Band

In analogy to equation (12), one can define a dimensionless SSC flux density spectrum, which is the quantity we actually calculated:

$$
\epsilon_X(\nu) \equiv \frac{6\sqrt{3}m_e^3c^6D^2F_X(\nu)}{\pi e^7n^2R^4B} \approx \nu \int_0^\infty \frac{d\gamma N(\gamma)}{\gamma^4} \int_{\frac{\nu}{\gamma^2}}^\infty \frac{d\nu'}{\nu'^3} \frac{\epsilon_X(\nu')}{\nu'^3} \left(2\nu \ln \frac{\nu}{4\gamma^2\nu'} + \nu + 4\gamma^2\nu' - \frac{\nu^2}{4\gamma^2}\right).
$$

The spectrum of the Comptonization component and its evolution should be quite different for the two phases, and may be used to distinguish between them. Figure 10 shows the evolution of the radiation spectrum during the cooling (left) and heating (right) phases, with the NIR spectrum overplotted with the newly calculated X-ray spectrum. Besides indicating
the initial and steady-state spectra, we highlight the spectral evolution when the correlation between NIR flux and spectral index are prominent, corresponding to the shaded periods in Figure 9, where the spectra are indicated by filled circles. For the model parameters chosen in §3.2, much more high energy emission is produced via SSC in the cooling phase. This is because the electron column depth $nR$ in the cooling phase is 50 times that in the heating phase while their synchrotron luminosities are comparable.

![Graph showing spectral evolution](image)

Fig. 10.— Evolution of the SSC spectrum $B\nu\epsilon_X$ during the cooling (left) and heating (right) phases shown in Figure 8. Results from §3.2 are overplotted on the left side of each figure. Besides the initial and steady-state spectra, we show the spectral evolution for the cooling and heating phases when the correlation is produced. These spectra are indicated by the solid circles in Figure 9 and Figure 11.

In analogy to the previous section, these spectra can be used to generate predicted light curves that can be compared with observation. These light curves are shown in Figure 11, below. As expected, higher energy emission precedes (lags) lower energy emission slightly during the cooling (heating) phase because they are produced by more energetic electrons. However the difference in the peak times is within 10 minutes and may not be distinguished with the capacities of current instruments. The time delay between the NIR and X-ray peaks is also within 30 minutes. We therefore expect a good correlation between the NIR and X-ray flux densities. The right panel shows the corresponding correlation between $\epsilon_X$ and the X-ray spectral index $\alpha_X$. This correlation gives a correlation between the X-ray flux
density and spectral index, which can be tested with observations. Interestingly the X-ray flux density peaks near $\alpha_x \simeq 0.6 - 0.7$, which explains the hardness of X-ray spectra.

![Graph of flux and spectral index over time](image)

Fig. 11.— *Left:* Light curve of the flux and spectral index at $0.5 \times 10^{18}$ Hz (thin lines, corresponding to $\sim 2.1$ keV) and $2.0 \times 10^{18}$ Hz (thick lines) during the cooling (dashed lines) and heating (solid lines) phases. The scales for the heating phase are indicated on the right and top axes. *Right:* The corresponding correlation between the X-ray flux and spectral index. The scale of the heating phase is indicated on the top axis.

4. Conclusion and Discussion

Flares from Sgr A* are believed to be produced within a few Schwarzschild radii of the galactic supermassive black hole. The observations and dynamics of these flares provide an interesting topic of study, because they offer insights into the plasma physics of the accretion disk and the space time near the black hole. Electron acceleration seems to be the most natural explanation for the observed correlation between spectral index and flux. A time-dependent approach is necessary to recover all the information in the kinetic equation, and also helps address the similarity between the flaring timescale, synchrotron cooling timescale, and the dynamical time at the last-stable-orbit. Although there are already attempts to carry out this study analytically (Becker et al. 2006; Schlickeiser 1984), numerical solutions are required when specific events are considered.
This paper addressed electron acceleration by using the Chang-Cooper method to solve the kinetic equation for a turbulent magnetic field. The theory (and kinetic equation itself) of this turbulent magnetic field were developed previously by Liu et al, and has four basic model parameters: the magnetic field and its coherent length, gas density and the initial distribution of electrons. The electrons gain energy from plasma waves and lose energy via radiation, and the energy dependencies of these rates are quite different. For the optically thin collisionless plasma in Sagittarius A*, we have shown that synchrotron cooling dominates and that relativistic Maxwellian distributions usually give the correct description of the electron spectra. We derived a formula for the temperature/average energy dependence during the evolution, and it proved to be accurate within +/- 5-10 percent for our examples.

It is not trivial to explain the observed correlation between spectral index and flux in the NIR band. Electron acceleration with a constant magnetic field was unable to produce a strong enough correlation in the heating phase, and could only marginally agree with observations in the heating phase. Doppler effects and adiabatic processes, which have been suggested previously as explanations for the correlation (Gillessen et al. 2006; Yusef-Zadeh et al. 2006b), cannot reproduce the observed results without a very unusual electron spectrum.

We argue that the electron heating and cooling of a flare with a magnetic field anti-correlated to electron “temperature” (or mean energy) provides the most reasonable explanation for the observations. This reduces the flux density at high temperatures (and hard, which strengthens the correlation and brings it closer to observation. It’s not precisely clear what mechanism could cause such an anti-correlation. For the decrease in magnetic field in the heating phase, it’s possible to imagine some sort of magnetic reconnection phenomenon transferring the magnetic field energy into the electrons as it heats. However, even a handwaving explanation of the increase in magnetic field during a cooling phase is difficult. That said, since the processes that cause the magnetic field are not well understood, it’s not easy to rule out such an increase.

Other explanations of the correlation are still possible. While Doppler boosting and adiabatic processes have been ruled out, there may be a complicated source structure to the flaring region. This structure could have been inferred from observation in order to reproduce the correlation, but that would add many additional parameters to the model. For a more detailed discussion, see the other version of this paper.

In future studies, other radiative processes can be added to refine our model, such as self-absorption and inverse-Compton cooling. In the long-term, our results could be combined with magnetohyrdodynamic simulations and relativistic ray-tracing code to form a more detailed investigation of the space-time near the black hole and possibly constrain its spin, which would be a major achievement.
This work was carried out under the auspices of the National Nuclear Security Administration of the U.S. Department of Energy at Los Alamos National Laboratory under Contract No. DE-AC52-06NA25396. This research was partially supported by NSF grant ATM-0312344, NASA grants NAG5-12111, NAG5 11918-1 (at Stanford), and NSF grant PHY99-07949 (at KITP at UCSB). SL would like to thank Justin R. Pelzer for developing part of the code. JB would like to thank Norman H. Magee for his support, the Los Alamos Summer School for providing such a wonderful educational opportunity, and SL for his guidance and patience.
REFERENCES


Blumenthal, G. R., & Gould, R. J. 1970, Rev. of Mod. Phys. 42, 237


-----------------------------------------------------------------------------------------------------------------------------------

This preprint was prepared with the AAS L\TeX macros v5.2.
Werblin et al. modeled part of the visual system consisting of the retina in a salamander. By understanding the Werblin model and developing a similar model in the NEURON environment an extension of the model can be developed to a retina in a primate. The hope is to create a model that can be used in the design of a retinal prosthesis. This paper discusses the Werblin model and its implementation into the NEURON simulating environment.

I. Introduction

The retina is a sensory receptor at the rear of the eye that is made up of millions of neurons. Neurons are cells that can transmit and receive electrical signals. These electrical signals are called synapses and they form pathways between neurons creating neural circuits. Neurons can be characterized by their distinct functions. The retina is comprised of five types of neurons: photoreceptors, horizontal cells, bipolar cells, amacrine cells and ganglia. Photoreceptors (rods and cones) respond to light and transmit signals to both the horizontal and bipolar cells. Once the horizontal cells receive the signal, they respond synaptically to the photoreceptors sending the signal to the bipolar cells and laterally with neighboring cells. The bipolar cells then signal both the to amacrine and ganglion cells. They also signal within their own neural network. Finally, input signals to the amacrine cells propagate to the ganglion cells, which are the final signal output of the retina. The retina can be fully imaged and therefore studied as a fully intact system while other sections of the central nervous system including the brain are generally studied in sections.

Scientists use models and simulations to study real world interactions. It is fairly easy to change parameters in a simulation. By changing parameters, a better understanding of how these parameter changes affect the real system can be studied. Neuroscientists believe that by developing an accurate model of the retina, an increased understanding of our visual system can be obtained. In this paper, we examine a computational model of the visual function of the salamander retina developed by Teeters, Jacobs and Werblin (1996), which will further be described as the Werblin model. The retina of a salamander was modeled because it is in essence a simpler system then that of a primate.

Our goals of the project are to a) understand the Werblin model b) understand NEURON and implement the Werblin model into the NEURON environment and c) extend the Werblin model to a 3-dimensional model d) develop a model for the retina in a primate and e) study simulation results to aid in experimental research ongoing in retinal prosthetics. This paper focuses on the first two ideas, understanding the Werblin model and NEURON and also implementing that model into the NEURON environment.
II. Understanding The Werblin Model

The Werblin model describes interactions between neurons and networks of neurons comprising the retina of a salamander. The structure of the model was designed on a PIPE (Pipelined Image Processing Engine). The PIPE processes, in parallel, 2-dimensional image arrays of 256 x 256 elements. Neurons were laid out in rectangular grids or images, which represent an area of 16 µm². Each image is representative of a neural layer (of differing class or function) within the retina. The model works by changing both the conduction and potential properties of each class of neurons as a response to an initial light input, which propagates as a synapse throughout the neural network. Each grid compartment within the image could either represent a neuron or part of a neuron and this compartment stores a value for the net membrane voltage. The model in essence captures the conduction a voltage properties image by image over time and within neuron types. Light, at variable intensities, is input to the cones. The cones respond to this light with an excitatory synapse. The synapse creates transient voltage values that travel the length of the neural circuit. This temporal affect is not homogeneous from one class of neuron to another. At each time step and within a particular neuron type, net values for both the sustained and transient voltages are computed. At time t = 1, the transformed image (rectangular grid) contains the modified properties due to a spatially diffused signal and a change in the calculated value for the transient effects. Integrated synapses were calculated using a time step of 1/15 second. Only the cell body or the soma is incorporated into the model therefore the neuron cell is made up of only one compartment. The model essentially sets an initial voltage value for the photoreceptors, modifies this potential due to excitation and feedback then passes the voltage value along, which in turn gets further modified by different types of neurons. The ending result is a final output voltage set in the ganglia.

Werblin developed two models, one for what he calls the Outer Retina Model and the other the Inner Retina Model. Both models contain properties that were observed from experimental data or that were derived under characteristic conditions.

The Outer Retina Model simulates responses by the photoreceptors (cones only, representing a light activated state), horizontal cells and on-type bipolar cells. Each cone contains a conductance value that includes the value for a variable light intensity, the response from the horizontal cell and the resting conductance. These values also have associated potential values. The response is passed in the form of a voltage value from the cone through a filter (representing the horizontal cells) and directly to the bipolar cells. The filter modifies the synapse and sends signals back to the cone and to the bipolar cells. The bipolar cells receive both conductance values and potentials from the filter, the cone and their own intrinsic resting properties. The summation of this final voltage value from the bipolar cells is maintained for the Inner Retina Model.

The Inner Retina Model begins with the stored voltage value received from the bipolar cells (the ending value from the Outer Retina Model). The bipolar cells contain both a constant and variable (changing in time) voltage value. The constant or sustained value is passed to the narrow field amacrine cells, which inhibit
neighboring bipolar cells. This creates a variable voltage emanating from the neighboring bipolars. The network consists of bipolars with sustained and transient voltage values. The transient along with the sustained potential is summed via spatial convolution through the amacrine cells to the ganglion cells. The ganglion cells have both conductance and potential values for that coming from the amacrine and bipolar cells and resting values.

A. Properties

The Werblin model provides properties for each neuron type. The Table 1 lists the properties and their respective values.

<table>
<thead>
<tr>
<th>Neuron Type</th>
<th>Photoreceptors (cones)</th>
<th>Bipolar Cells</th>
<th>Ganglion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resting Conduction [nSiemens]</td>
<td>4</td>
<td>0.128</td>
<td>2</td>
</tr>
<tr>
<td>Resting Potential [mV]</td>
<td>-50</td>
<td>-40</td>
<td>-24</td>
</tr>
<tr>
<td>Feedback Synapse [mV]</td>
<td>-67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Light Activated Potential [mV]</td>
<td>-8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reversal Potentials [mV]</td>
<td>(horizontal) -17</td>
<td>(bipolars) 28</td>
<td></td>
</tr>
<tr>
<td>Time Constant [ms]</td>
<td>55</td>
<td>55</td>
<td>55</td>
</tr>
</tbody>
</table>

Properties of differing class of neurons. Note, both the horizontal and amacrine cells act as filters.

B. Spatial Convolutions

In the model, spatial convolution serves two functions; it represents differing class dendritic fan-in or fan-out (due to unmodeled dendrites) as well as coupling within the same class of neurons. Werblin uses spatial convolutions to modify voltage values using weighted sums. This simulates a lateral fan-in affect from the dendrites to the soma or fan-out to the axons. Werblin uses the fan-in effect so that a model can be built using only the soma or cell body. Adding more compartments for the dendrites and axons only complicates the model. The weighted values are in some cases homogeneous and others non-homogeneous. Convolution kernels are 3 x 3 and 9 x 9 matrices. The 3 x 3 matrix (Av3) has is a uniform weighted kernel. There are five 9 x 9 convolution kernels used in the model. One is homogeneous while the other four are non-homogeneous. The values in the kernels are used to increase or decrease the resolution or contrast in the image. Table 2 describes the weighted values of each cell type for the 9 x 9 convolution kernels. These weighted values are duplicated in each row making up the entire 9 x 9 sub grid.
Table 2

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Mask Name</th>
<th>Mask Values (Center to edge)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontal</td>
<td>Av9</td>
<td>1   1   1   1   1</td>
</tr>
<tr>
<td>Bipolar</td>
<td>Av9b</td>
<td>1   0.597 0.356 0.213 0.127</td>
</tr>
<tr>
<td>Narrow-field amacrine</td>
<td>Av9n</td>
<td>1   0.950 0.805 0.639 0.463</td>
</tr>
<tr>
<td>Wide-field amacrine</td>
<td>Av9w</td>
<td>1   0.925 0.823 0.697 0.549</td>
</tr>
<tr>
<td>Ganglion</td>
<td>Av9g</td>
<td>1   0.908 0.685 0.427 0.220</td>
</tr>
</tbody>
</table>

9 x 9 convolution kernels.

Convolution is a way to take input and produce a weighted and sometimes normalized output. The weighted output is calculated by applying a convolution kernel or matrix to a set of input values. The kernel is used as a one to one mapping function in which the center value of the convolution kernel is mapped with the value being computed in the image matrix. The image matrix is calculated by mapping and multiplying the weighted values in the kernel to the matrix element value that it is mapped to. All weighted values are summed and the new final value is placed in the center cell of the image matrix. If the final value is normalized, it is then divided by the sum of the values in the convolution kernel.

Werblin uses squeeze and zoom operations to increase or decrease the size of the field within the image. The zoom operation increases the resolution while the squeeze operation blurs the edges decreasing the resolution. What this effectively means is that the length over which the distance from the center is measured changes. More (squeeze operation) or less (zoom operation) cells are mapped from the initial image to the convolution kernel. If more cells are used then in some further operation, the zoom function must be applied making the final image the same size as the initial and vice versa, the squeeze operation must be used in response to a zoom operation.

We apply the zoom operation to the image matrix by repeating values in the corners of the chosen matrix. These values are repeated at each adjacent cell to the corner cells. In the squeeze operation, we take the values at the edge of the chosen matrix within the image matrix and average the values replacing the corner cells in the image matrix with this average value. The squeeze operation blurs or smoothes the edges, which corresponds to a decrease in the number of cells included in the sample.

Weighting in the model represents how surrounding neurons respond to a synapse that has occurred in a local neuron. While the squeeze and zoom operations increase or decrease the receptive field.

C. Model Equations

To fully describe the pattern of activity in the neural circuit, Werblin computes synaptic functions in one of two ways: by using a modification to the Goldman equation (Equ. 1) or by using a look-up table.
Where $V$ is the voltage, $G$ is the conductance, $E_S$ equals the potential and $r$ stands for the resting value. The look-up table is used when the synaptic function requires only one input and the Goldman equation is used when two inputs are required. There are sixteen model equations, which consist of functions that compute final voltage values in each class of cells in response to the initial light input. The functions may contain the Goldman equation with various inputs and other calculations that modify voltage values. The equations show the interactions between the neurons represented by spatial convolutions and the squeeze and zoom operations. They also describe the temporal dependence and synaptic propagation.

**IV. Understanding NEURON and Implementing the Werblin Model**

To implement the Werblin model we wanted to design a model that used NEURON’s rich set of features while trying to maintain the computational integrity of Werblin’s experimentally supported model. It was important to address three main areas: implementing each class of neurons and their properties, connecting neurons within classes and between classes, and implementing synaptic behaviors.

**A. Neurons and their Properties**

Neurons were modeled with one compartment representing the cell body or soma to match Werblin’s description of the geometry of the neuron. When defining neuron objects, NEURON requires the user to define the geometry of the cell body. We defined the diameter and length as the same value to obtain a spherical cell. Werblin gives us a value for the time constant, this value along with the specific lipid membrane capacitance ($1$ microF/cm$^2$), is used to determine the radius of the cell. The radius is computed using Equ. 2.

$$R = \sqrt{\frac{\tau G}{4 \pi C_s}}$$

Where $R$ is the radius, $\tau$ the time constant, $G$ equals the conductance and $C$ is the specific capacitance.

**B. Neural Networks and Their Connections**

We are implementing the Werblin model on a standard PC and thus are not using a parallel computer. Although, many of the simulations have not been performed, we don’t anticipate seeing a difference in Werblin’s results as a function of not using a
parallel computer. To mimic the images or 256 x 256 rectangular grids in our model, we have used 2-dimensional lists. A list in NEURON is an object that holds or contains other objects. In this case the list holds individual neurons per list element. To build the 256 x 256 image of neurons, we have created 256 lists and within each list object, we created 256 more lists. The neuron objects were placed within these inner lists. There are several differences to note between the Werblin model and our NEURON model. One is that each element within our model grid houses a full individual neuron. In Werblin’s model, grid cells could house either a full or partial neuron. We also decided to make each layer of neurons its own 2-dimensional list instead of transforming or modifying one 2-dimensional list when necessary. This was done so that we could maintain property values for classes of neurons for further calculations. In the Werblin model, the horizontal cells were not explicitly modeled instead a filter was used to represent these cells. In our model all five classes of neurons were modeled in their own 256 x 256 list.

In NEURON, connections between neurons are programmed by how synapses are sent throughout the system. Spatial convolutions as well as squeeze and zoom operations are modeled by sending weighted synapses from multiple neurons to particular neurons. These synapse values are summed through convolution kernels and squeeze and zoom sub samples. The connections are made using sets of for loops.

C. Approach

We are using a conduction model within the NEURON environment, which means conduction values are changed in each class of neurons. In addition to a change in conduction in response to a synapse, each class of neuron has a leak current. A leak current decays exponentially to the cells equilibrium potential. A leak current is physically modeled with a capacitor in series with a resistor in an electrical circuit. This situation is represented by a time dependent differential equation. To simulate the leak current an initial current must be applied to each neuron. This current is applied with a current clamp, which is set for a particular length of time through each neuron. We represent an electrical synapse by using point processes. Point processes can be represented by an ohmic connection between neighboring cells in the neural network. Unlike the leak current which affects only a local neuron, an electrical synapse affects neighboring neurons. Neighboring neurons could be within the same class of neurons and/or be a part of a neighboring neural layer. In our design, we will start the simulation with one unconnected neuron. At this point we plan to apply the leak current as well as any point process that that particular class of neuron requires. Finally, we will apply the point processes between neurons in effect connecting them as the synapse propagates along the neural network. The connections between neurons are made through spatial convolutions and by using the squeeze and zoom operations described earlier. It is important to note that the description of the squeeze and zoom operations is our interpretation of those used in the Werblin model. One way to test our interpretation would be by matching our results to Werblin’s. If our results cannot be matched, we will need to modify or fine-tune the operations. NEURON supports a high level language, NMODL (NEURON Model Description Language). NMODL extends the functionality of NEURON by allowing the user to describe chemical and mathematical equations in familiar forms. NMODL then takes these equations and upon being compiled, turns them
into C code. By using NMODL we can easily describe physical functions using differential equations. We have used NMODL to describe the leak currents as well as synapses.

The leak current is inserted into the neuron object under its properties. To simulate the affects of the leak current, we needed to apply a current through the cell. NEURON supplies a current clamp in its library of mechanisms that allows us to apply this current. In doing so, we set a delay of 200 ms and a duration time of 100 ms as well as a current amplitude of 1 nA. We set the delay to 200 ms, which is approximately 4 time constants; this is done so that the current can equilibrate at the cells resting membrane potential. We have applied a current clamp along with the leak current to the center of the soma of a cone cell. Figure 1 shows the results of the leak current.

![Graph](image)

The graph shows an exponential decline starting at the point where the current is suspended. The voltage decays to the equilibrium potential.

V. Future Work

A lot of work needs to be accomplished to fully implement the Werblin model. We have developed a plan of the work necessary to complete this stage of the project. Our next step is to match leak current results in the cone to those of Werblin’s. Once that is complete, we need to apply a current to the cone neurons due to the light input. We will use an exponential current to simulate this light current. After we simulate the light current we will simulate a synapse modifying an alpha synapse function prewritten with NMODL. At this point, we will need to apply leak currents to each type of neuron changing the properties to fit the particular types. From here, we will need to connect the neurons using spatial convolution and squeeze and zoom operations through the synapses. It will be necessary along the way to match our results to those of the Werblin model.

VI. Conclusion

A retina is modeled by Werblin et al. and is used to study the visual function of a
salamander. By understanding the visual function of a salamander, neuroscientists are on a path to determining the visual function of the retina in a primate. It is the hope that these studies will eventually be used to develop a retinal prosthesis. It was necessary to understand the Werblin model in order to implement it into the NEURON simulation environment. NEURON is an environment developed for neuroscientists to aid in the development of simulating neurons and neural networks. The Werblin model uses image arrays with 256 x 256 elements, in which each element has a set voltage value. The voltage values are then modified in response to electrical signals. Werblin’s model uses spatial convolutions along with squeeze and zoom operations to simulate connections between neurons. In NEURON, we have used 2-dimensional lists containing neuron objects to represent the neural network. We have used a conduction model within NEURON to define changes in response to electrical signals. To fully implement the Werblin model, it will be necessary to apply electrical circuits to all types of neurons. Future work also consists of making connections between neurons through synaptic propagation.

VII. Acknowledgements

I would like to thank the Los Alamos Summer School and the University of New Mexico REU organizers for their help and for giving me the opportunity to be apart of this program. I also would like to thank Mark Flynn and John George for their help interpreting the Werblin model. I would especially like to thank my mentor Garrett Kenyon for his patience and for giving me such an interesting and challenging project to work on.

Analysis of Proton Radiography Images of Shock Melted and Damaged Metals

Jeffrey Cecil\textsuperscript{1,3}
Hanna Makaruk\textsuperscript{2}

\textsuperscript{1}T-4/LASS, Los Alamos National Laboratory, Los Alamos, NM
\textsuperscript{2}Mentor; P-22, Los Alamos National Laboratory, Los Alamos, NM
\textsuperscript{3}Illinois Institute of Technology, Chicago, IL

Abstract

This paper discusses the analysis of proton radiography (pRad) images of shockwave metal melting. The experiments investigated the physical properties of metals under extreme pressure and temperature conditions. Aluminum, copper, and steel coupons were melted and/or damaged by the shockwave of a diverging high explosive. Multiple experiments were performed in identical conditions, with various coupon thicknesses and metal types. pRad was used to record each explosion and the resulting images were then analyzed by software written in Mathematica. The program used a level set type method to extract numerical data. Data obtained from a series of images captured during an individual experiment was used to calculate the velocities of metal fragments. The measured velocities of the top surfaces of the explosions agreed very well with the data collected by the Velocity Interferometer System for Any Reflector (VISAR). This technique additionally allowed the velocities of occluded surfaces to be measured even though they could not be seen by VISAR. Also, velocities calculated for identical experiments almost perfectly agreed with one another. This result showed that in these experiments both the velocities and shapes of the major fragments of the metal coupons are highly deterministic functions of coupon thickness and metal type. These findings are being used to verify the predictions of hydrocode modeling that is currently being developed.

Introduction

New techniques in pRad image analysis make it possible to gather more data than ever before about metal melting and damage that occur during explosions. This data includes velocity and contour evolutions and are needed not only to better understand experiments but also to compare with hydrocode modeling. In this manner, it is possible to choose the most accurate hydrocode.

What follows is a description of the pRad technique and the experiments that required pRad image analysis. Methodology and results are also discussed and later reviewed in the conclusion.
Proton Radiography (pRad)

pRad is an imaging technique that uses a proton beam in place of conventional X-rays [1, 4, 5]. High energy protons are focused by magnetic lenses and captured by a series of cameras after penetrating the sample under consideration. A schematic of the pRad apparatus is shown in Figure 1.

![Schematic of the pRad apparatus](image)

**Figure 1. pRad facility located at Los Alamos National Laboratory [4].**

The attenuation lengths of high energy protons are much longer than those of high energy X-rays and, as a result, pRad provides many advantages over conventional radiography. pRad detects small changes in density that cannot be seen with X-rays. Longer attenuation lengths also mean that more protons than X-rays can penetrate a given sample.

A proton beam is adjusted separately for each experiment. The intensity of a typical proton beam is so high that billions of protons can pass through a sample over extremely short time intervals. The proton source at Los Alamos National Laboratory is an 800 MeV beam produced at the Los Alamos Neutron Science Center (LANSCE). Traveling at approximately 84% of the speed of light, four billion protons pass through the target in 50 ns. This produces very crisp images and eliminates “snow” that would otherwise be present if too few protons were used. Since X-rays do not penetrate metal coupons as well as protons, an X-ray beam would need to be very powerful and hard (consisting of high energy photons) in order to achieve comparable results. However, since such X-ray beams are associated with frequent artifacts and a high incidence of scattering, it is difficult to use them effectively.
**Experimental Setup**

This discussion of the experimental setup and measurement methods is described in a previous work [6].

![Diagram of experimental setup](image)

**Figure 2. Identical configuration for all experiments.**

A 0.50 in. thick, 2 in. diameter PBX9501 High Explosive (HE) is placed in contact with metal coupons of different types and thicknesses. The HE is ignited using an SE-1 point detonator located along the symmetry axis of the experiment. This type of detonation creates symmetric shockwaves and produces axial symmetry in the system. Therefore, it is possible to recreate nearly all information from an experiment using only pRad images since they are geometrically parallel projections of a beam perpendicular to the symmetry axis of the system.

The VISAR velocimetry probe [2, 3] is placed directly above the coupon. VISAR measures the velocity of the top surface of the coupon by emitting 532 nm laser light and then calculating the Doppler shift of the reflected light from its interference with the original beam. During the actual experiments, the package is mounted upside down from how it is shown. However, the pRad images themselves are analyzed in a configuration consistent with Figure 2.

**Special Considerations**

While pRad does provide many advantages over X-ray radiography for the type of experiment considered here, there is no flawless imaging technique. There are pRad
specific problems that must be accounted for when analyzing images of this type, such as small color gradients. The most significant problem that occurs with contour detection is known as limbing and it appears on the edges between different densities. Due to this artifact, a contour may either vanish or appear as a double contour when detected by a gradient based method. In these situations, standard (i.e. gradient) based edge detection methods cannot be used to detect contours with sufficient precision. To avoid this problem, a level set method was developed so that a specific density threshold could be chosen for each image. This method allows information to be extracted about many physical features of the system that would otherwise not be detected.

**Method**

**Interactive Contour Detection**

In order to identify a density contour from an image, a level set method [7] must be applied in a systematic way. First, a specific region of an image must be chosen for analysis. Noise is then removed from the selected portion using a heat-equation based algorithm so that an accurate contour can be produced. If not enough denoising is applied to an image, the contour produced from that image will contain extraneous information that does not represent the surface under consideration. However, if there is too much denoising, real features of the surface may not appear in the contour because they have been smoothed out of the image. For these reasons, very little automatic denoising is applied to an image and the majority of the process is left to users experienced with the technique.

Once an image has been properly denoised, a density threshold must be chosen. This threshold takes the form of a number between zero and 255, the range of possible pixel colors contained in an image. The pixel value is correlated with the thickness and density of the metal; higher values represent higher densities or thicknesses of a fragment. Once a threshold is chosen, all pixels with a lower value (less dense regions) are converted to black while all pixels with a higher value (denser regions) are converted to white. The border between the two regions is subsequently eroded one pixel deep. The resulting image is then subtracted from the previous black and white image to produce a one pixel thick contour.

The choice of threshold is critical in determining the final appearance of the contour. Figures 3, 4, and 5 display this effect. For this particular image, a higher threshold resulted in the appearance of more internal detail while the lower threshold merely generated an outline of the structure. It is also important to note that the correct threshold often varies from image to image, even for images that are part of the same experiment. Therefore, it is necessary to closely monitor this parameter for each individual image.
Velocity Calculations

After extracting the proper contour from each image for a given experiment, it is possible to calculate the velocity of the surface being examined. Every contour is saved as a set of two dimensional numerical vectors that represent the physical location of each point of the contour. Sets of vectors from consecutive contours are then subtracted to determine...
the vertical displacement (Δy) between two points with the same horizontal coordinate. Since the timing between consecutive images (Δt) is known precisely, the vertical velocity of each point can be calculated by finding Δy/Δt.

Once velocities are calculated from each set of consecutive contours, a set of velocity curves is generated which displays velocity as a function of horizontal coordinate. This data is then averaged to determine a single value of the velocity for each point along the contour. See Figure 10 in the Results section for an example of a velocity plot.

Velocity data obtained in this manner has been validated by several independent methods. First of all, the data is clearly symmetric about the symmetry axis of the system even though no knowledge of this symmetry is actually used to calculate the velocities. Also, the top surface velocities measured by VISAR agree very well with the velocities calculated using this technique. One benefit of this method of velocity calculation is that it has the ability to calculate the velocities of occluded surfaces that cannot be measured by VISAR. It is important to gather as much velocity data as possible because not only is it used to further understand the dynamics of the system, but it is also used for comparison with hydrocode modeling.

Results

This section presents typical images from the three considered metals. pRad images show clearly that under identical experimental conditions, coupons of different metals are damaged and melted in different ways. On the other hand, when experiments are repeated with the same metal type and coupon thickness, almost all of the features are identical. Only small, non-axially symmetric structures are not repeatable.

Glossary of Terms

Several terms are used in this section to describe different regions of a given image. A visual glossary of terms has been provided as a reference.

Aluminum Image

![Aluminum Image](image)

Figure 6. Surfaces of interest apparent in aluminum experiments.
Figure 7. Surfaces of interest apparent in copper experiments.

Figure 8. A surface unique to copper experiments using 0.50 in. coupons.
Data from six experiments with aluminum was analyzed. Four of the experiments (numbers 84, 85, 117, and 118) used 0.25 in. coupons while the other two (numbers 119 and 120) used 0.50 in. coupons. The velocities of the top surfaces of the 0.25 in. coupons varied by less than 6%, with a range from 1870 m/s to 1970 m/s. The data was also highly consistent with experimental results, differing from VISAR measurements by as little as 1%. The velocity of the top surface in experiment 119 was 1350 m/s and in experiment 120 it was 1390 m/s, a difference of less than 3%. There was no VISAR data available for experiment 119. However the result obtained for experiment 120 differed from the VISAR measurement by less than 1%. Table 1 shows both the calculated and experimentally measured velocities of the top surfaces for each experiment involving aluminum.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Top Surface Velocity (m/s)</th>
<th>VISAR (m/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>84</td>
<td>1870</td>
<td>1970</td>
</tr>
<tr>
<td>85</td>
<td>1970</td>
<td>1950</td>
</tr>
<tr>
<td>117</td>
<td>1930</td>
<td>1950</td>
</tr>
<tr>
<td>118</td>
<td>1950</td>
<td>1900</td>
</tr>
<tr>
<td>119</td>
<td>1350</td>
<td>Not Available</td>
</tr>
<tr>
<td>120</td>
<td>1390</td>
<td>1380</td>
</tr>
</tbody>
</table>

Table 1. Calculated and experimentally measured velocities of the top surface for all aluminum experiments.

There were two sets of aluminum experiments where the grain orientation of the metal was varied for a given coupon thickness. Experiments 117 and 119 used coupons with a “road” grain orientation while experiments 118 and 120 used coupons with a “plate”
grain orientation. However, there were no major differences in the velocity data between experiments 117 and 118 or between the velocity data for experiments 119 and 120. Since the coupon thicknesses were held constant across each pair of experiments and only the grain orientations of the coupons were varied, it is likely that grain orientation does not influence the velocity of aluminum coupons.

An interesting feature of the velocity plots of the top surfaces found in experiments 84 and 85 is their parabolic nature. In neither of the other two experiments involving aluminum coupons of this thickness did the velocity of the top surface vary significantly with the radial coordinate. The velocity profile for experiment 85 is shown in Figure 10. It displays velocity in units of 100 m/s versus horizontal position in units of 0.1 mm. The axes labels are identical for all subsequent velocity plots in this paper. The time order of the contours is denoted by their colors and is consistent for all other plots involving time.

![Figure 10. Velocity plot of the top surface for experiment 85.](image)

In addition to velocity, the time evolution of the physical shape of the top surface of each aluminum coupon was also recorded. For each experiment, it was observed that the curvature of the top surface increased over time. The most dramatic increases in curvature occurred in experiments 84 and 85. Figure 11 shows the evolution of the shape of the top surface with time for experiment 85. The vertical and horizontal axes represent vertical and horizontal position in units of 0.1 mm, respectively. These labels are the same for all contour plots throughout the paper. In this plot the contours were overlapped at a specific point in order to exemplify the change in curvature over time. As a result, only the horizontal position and differences in vertical position for a specific curve are relevant. The actual physical locations of the top surfaces are not depicted in this figure.
For each of the six aluminum experiments, a distinct middle surface also emerged. The velocity of this surface was calculated for each experiment and in the experiments that used a coupon thickness of 0.25 in., the velocities ranged from 1470 m/s to 1580 m/s. This is only a spread of approximately 7%. The experiments that used 0.50 in. coupons had middle surface velocities of 1000 m/s and 1010 m/s. An interesting feature that was observed from measurements of both types of experiments was a fairly constant difference between the top and middle surface velocities. For all experiments except 85, this difference was between 330 m/s and 390 m/s. Table 2 shows the top and middle surface velocities of the aluminum experiments.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Top Surface Velocity (m/s)</th>
<th>Middle Surface Velocity (m/s)</th>
<th>Velocity Difference (m/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>84</td>
<td>1870</td>
<td>1540</td>
<td>330</td>
</tr>
<tr>
<td>85</td>
<td>1970</td>
<td>1470</td>
<td>500</td>
</tr>
<tr>
<td>117</td>
<td>1930</td>
<td>1580</td>
<td>350</td>
</tr>
<tr>
<td>118</td>
<td>1950</td>
<td>1570</td>
<td>380</td>
</tr>
<tr>
<td>119</td>
<td>1350</td>
<td>1010</td>
<td>340</td>
</tr>
<tr>
<td>120</td>
<td>1390</td>
<td>1000</td>
<td>390</td>
</tr>
</tbody>
</table>

Table 2. Calculated velocities of the top and middle surfaces for all aluminum experiments.

Velocity data for the bottom surface of the aluminum coupons was difficult to obtain due to a lack of contrast in this region in many of the pRad images. Measurements were only completed for experiments 84, 119, and 120. The velocities measured from these three experiments were very similar. The velocity of the bottom surface of the 0.25 in. coupon was 560 m/s and those of the 0.50 in. coupons were 460 m/s and 540 m/s. More data would be required to determine if velocities in this range are a deterministic property of this surface or if these results are merely random.

As a final summary of the velocities of the three primary surfaces found in experiments involving aluminum, Figure 12 shows the calculated velocity of each surface for each coupon thickness. It is interesting to note that as the coupon thickness increases, the velocity of a particular surface decreases.
Copper

Four experiments were performed using copper. Three of the experiments (numbers 82, 83, and 121) used 0.25 in. coupons while the other (number 122) used a 0.50 in. coupon. The velocities of the top surfaces of the 0.25 in. coupons varied by approximately 7%, with a range from 980 m/s to 1050 m/s. The data was also highly consistent with experimental results, differing from VISAR measurements by less than 1% in some cases. For the 0.50 in. coupon, the velocity of the top surface was measured to be 695 m/s, a less than 4% difference from the VISAR measurement of 720 m/s. Table 3 shows both the calculated and experimentally measured velocities of the top surfaces for each copper experiment.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Top Surface Velocity (m/s)</th>
<th>VISAR (m/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>82</td>
<td>1050</td>
<td>1040</td>
</tr>
<tr>
<td>83</td>
<td>1000</td>
<td>1040</td>
</tr>
<tr>
<td>121</td>
<td>980</td>
<td>1030</td>
</tr>
<tr>
<td>122</td>
<td>695</td>
<td>720</td>
</tr>
</tbody>
</table>

Table 3. Calculated and experimentally measured velocities of the top surface for all copper experiments.

In each of the three copper experiments involving a 0.25 in. coupon, a distinct internal bubble structure was visible. However, the velocity of the top of the bubble was only
measurable for experiments 82 and 121. The velocities were fairly similar: 725 m/s and 680 m/s for experiments 82 and 121, respectively. A particularly interesting feature was noticed in the velocity plot of the bubble top for experiment 121, shown in Figure 13.

![Velocity plot of the bubble top surface for experiment 121.](image)

There are visible, consistent fluctuations of the velocity curves which are not present in any of the other velocity plots of this surface. However, more data is necessary to determine the nature of these results. The effect seems to be physical and may illustrate waves on a melted surface or other types of oscillations.

A contour plot of the bubble top surface was also created for experiment 121. Unlike the previous contour plot that overlapped the curves at a specific point, this plot shows the actual physical positions of the surface over time. As the bubble top surface changes position with time, there are only minor changes in the shape of the contour. This feature suggests that the copper comprising the bubble remains solid over the course of the experiment. Figure 14 shows the contour plot of the bubble top surface.

![Contour evolution of the bubble top surface for experiment 121.](image)
In experiment 122, a trapezoidal structure replaced the bubble that was present in the experiments with 0.25 in. coupons. The velocity of the top surface of the trapezoid was calculated to be 350 m/s, but since only one experiment was completed with a 0.50 in. coupon there are no other measurements available for comparison. The shape evolution of the trapezoid top surface was also recorded and, similar to the bubble top surface in experiment 121, the contours only varied slightly with time. Once again, this is an indication that the structure is made of solid copper. The contour plot of the trapezoid top surface is shown in Figure 15. The contours have been overlapped and do not represent physical position with time.

![Figure 15. Contour evolution of the trapezoid top surface for experiment 122.](image)

The velocity data for the bottom surface of the copper coupons was fairly consistent. The velocities ranged from 190 m/s to 300 m/s and are shown by experiment in Table 4.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Bottom Surface Velocity (m/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>82</td>
<td>300</td>
</tr>
<tr>
<td>83</td>
<td>200</td>
</tr>
<tr>
<td>121</td>
<td>310</td>
</tr>
<tr>
<td>122</td>
<td>190</td>
</tr>
</tbody>
</table>

Table 4. Calculated velocities of the bottom surface for all copper experiments.

The most interesting feature of the bottom surface of the 0.50 in. copper coupon was the shape evolution. Not only did the curvature of the surface change with time, but there were also growing instabilities in the surface that were especially apparent in the later stages of evolution. The changing nature of the shape of the surface suggests that this portion of the copper coupon may have been liquefied as a result of the shockwave. The contour plot for this experiment is shown in Figure 16. The actual physical positions of the contours are shown over time.
Figure 17 shows the calculated velocities of the top surface, bubble top surface, and bottom surface for each copper coupon thickness. It is interesting to note that as the coupon thickness increases, the velocity of a particular surface decreases. This is the same result that is encountered in the aluminum experiments.

![Velocity Measurements for Cu](image)

Figure 17. Summary of the calculated velocities of selected surfaces for all copper experiments. VISAR data is given when available.
Steel

Two experiments were performed using steel. Experiment 147 used a 0.50 in. coupon while experiment 148 used a 0.25 in. coupon. The calculated velocities of the top surfaces of these coupons were highly consistent with experimental data collected by VISAR. Table 5 shows both the calculated and experimentally measured velocities of the top surface for each steel experiment.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Top Surface Velocity (m/s)</th>
<th>VISAR (m/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>147</td>
<td>560</td>
<td>600</td>
</tr>
<tr>
<td>148</td>
<td>800</td>
<td>920</td>
</tr>
</tbody>
</table>

Table 5. Calculated and experimentally measured velocities of the top surface for all steel experiments.

In both steel experiments a trapezoidal structure was apparent, similar to that found in the experiment using the 0.50 in. copper coupon. The measured velocities of the bottom surface of this structure were 410 m/s and 730 m/s for the 0.50 in. and 0.25 in. coupons, respectively. The contour plots of the trapezoid bottom surface for both experiments reveal further interesting information about the structure and are shown in Figures 18 and 19. In these plots the contours have been overlapped and thus do not show physical position with time. In neither experiment does the shape of the trapezoid bottom change appreciably over time, suggesting that this portion of the coupon remains in the solid phase for both coupon thicknesses.

Figure 18. Contour evolution of the trapezoid bottom surface for experiment 147.
Another major surface of interest in the steel experiments was the bottom of the coupon. The velocity of the bottom surface for the 0.50 in. coupon was 225 m/s while it was 340 m/s for the 0.25 in. coupon. These results, along with other calculated velocities and VISAR data, are summarized in Figure 20. As with both aluminum and copper, the velocity of a particular surface is inversely proportional to coupon thickness.

![Velocity Measurements for Steel](image)

**Figure 19.** Contour evolution of the trapezoid bottom surface for experiment 148.

**Figure 20.** Summary of the calculated velocities of selected surfaces for all steel experiments. VISAR data is given when available.
Conclusions

The work presented here is primarily a collection of data extracted from pRad images using a very precise image analysis method. This method was not only able to detect many features that were difficult to observe on raw images, but it also produced quantitative results regarding their behavior. The precise nature of the data obtained from image processing makes it very suitable for comparison with hydrocode modeling. Eulerian and Lagrangian hydrocodes based on different physical models simulate shockwave metal melting experiments and generate velocity values as well as synthetic images of the explosions. Since the velocities and shape evolutions of important surfaces are accurately known from image processing, the validity of each hydrocode can be determined. Once a model is selected and its parameters are adjusted in such a way that it functions properly, it can be used to ascertain the values of currently immeasurable quantities such as the temperature and pressure of the system. The model will also be able to simulate experiments that have not actually been completed yet or are impossible to complete for various reasons.

The analysis of proton radiography images of the shockwave metal melting and damage of aluminum, copper, and steel has provided precise data about the physical properties of these metals under extreme temperature and pressure conditions. Distinct surfaces emerged for each metal type and thickness and they were analyzed to determine velocity and shape evolution. Experiments using the same metal type and coupon thickness were replicable in all major respects. In many instances, the contour evolution of a particular surface offered clues about the phase of that material. A constant shape indicated a solid phase while significant curvature changes and growing instabilities pointed to a liquid phase.

The method of analysis made it possible to measure the velocity not only of the top surface of the coupons, but also of occluded surfaces not measurable by the VISAR probe. When VISAR velocity data was available, it agreed well with the results obtained using the analytic method. The ability of this image processing procedure to confirm experimental results and even reach beyond the capabilities of current instrumentation clearly establishes it as a useful technique.

Acknowledgements

Many thanks to the organizers of the Los Alamos Summer School: James Colgan, Lee Collins, Norm Magee, and Sally Seidel. Special mention is owed to Norm, a man of unlimited kindness and generosity who ensured a wonderful time for all. Finally, I would like to thank my peers in the Los Alamos Summer School for making this summer an unforgettable experience.
References


Covalency of Uranium-Ligand Bonding
G.H. Lewis
August 8th, 2006

Abstract

The covalency of actinides was studied using an XAFS technique. Ligand-Metal bonding of UCl$_6$$^-$ was analyzed using XANES techniques to determine how outer orbitals of the metal interact and form covalent bonds. The energy absorption signature at the Chlorine K edge demonstrated a strong affinity for 6f orbitals to participate in covalent bonding. More detailed analysis on the gathered data will determine the exact extent to which covalency is possible.

Introduction

The exact nature of ligand bonding in actinide compounds is a topic of debate and many of the complexities associated with actinide bonding are not well understood. Historically, actinides were viewed as elements similar to the lanthanides whose bonds were characterized as largely ionic. It was believed that the outer orbitals such as the 5f were too contracted by electrostatic forces to play a major role in covalent bonding. This belief was supported by early spectroscopic analyses$^2$ and the ionic nature of actinides remained a generally accepted perspective. However, there are many cases of actinide compounds which show covalent behavior, and therefore the perspective that actinide compounds bond exclusively through ionic means has come under question.$^1$ It has so far proved totally impractical if not impossible to obtain through purely analytical means the nature of outer actinide orbitals and the types of bonds that they can form, and therefore evidence of covalent bonding must be obtained experimentally. To understand the exact nature of actinide ligand bonding, one must obtain detailed information about the 5f and 6d orbital states. This type of information can be obtained using a proven technique known as XAFS.

Theory

X-Ray Absorption Fine Structure is a technique that has had great success in obtaining information about, among other things, the chemical properties of atoms and structured arrangement of molecules. Specifically, XAFS techniques can be used to study the covalence of a particular atom in a relatively simple way. XAFS techniques take advantage of the fact that there exists a relationship between the coefficient of a particular atom within a compound to absorb photons as a function of energy and many characteristics of that atom. The data gathered in an XAFS experiment plots an atom’s
absorption coefficient of photons as a function of the energy of the photons.

There occurs a dramatic leap in the absorption of light at a particular energy due to the quantum mechanical nature of photons and electronic states. The excitation of electrons due to electric fields (here, in the form of photons) will only occur if the electrons are given exactly enough energy to enter a higher state. Below this threshold energy, electrons will not be able to absorb photons and will remain in their original state, and only once electrons are given enough energy to make stimulated transitions into a new final state will light be absorbed. This will cause a sudden rise in coefficient of absorption of light, which is known as the absorption edge due to the resulting steepness of the peak. The absorption edge associated with ejecting atoms in the n = 1 bound state is known by convention as the K edge and the edge caused by ejection of electrons in the n = 2 state is known as the L edge.

The greater technique of XAFS can be divided into two different classes: X-Ray Absorption Near-Edge Structure (XANES) and Extended X-Ray Absorption Fine Structure (EXAFS). An absorption edge is divided into two energy regions based on the final electron states that result from the stimulated emission of electrons at a certain energy. The energies region before and including the actual absorption edge are known as the XANES region and the study of energies after the absorption are known as the EXAFS region. The details of these two regions will be discussed later.

The derivation of the relevant equations and the demonstration that light absorption is related to many properties of the materials in question comes from fundamental atomic physics. XAFS experiments use light, usually in the x-ray range, to stimulate core electron ejection. A photon must have energy greater than the binding energy of a core electron in order to be able to eject that electron from its shell. The associated momentum of the ejected energy can be obtained from the (non-relativistic) equation

\[ P^2/2m = h\nu - E_0, \]

where \( P \) is the momentum of the ejected photoelectron, \( \nu \) is the frequency of the incoming photon, and \( E_0 \) is the biding energy of the electron (before it was ejected).
When a core electron is ejected, it must be treated as a wave that can either jump to a higher excited state within its own atom or enter the continuum and interact with nearby atoms and molecules. In the former case where the electron is only given enough energy to jump to a higher state within its own atom, known as the XANES region, one can obtain very detailed information concerning an atom’s electron shell availability and its ability to form covalent bonds. In the latter case, known as the EXAFS region, the electron fully escapes the core atom and enters a final state that is described as a superposition of outgoing and backscattered waves. In this region, one can obtain information about the molecular structure and the positions of nearby atoms. In both cases, the relevant can be obtained from a quantity as simple as the absorption coefficient of light at different energies by the atom in question. We will here only describe in detail the XANES results obtained from this experiment.

The x-ray photon entering the core atom can be viewed semi classically as an electromagnetic perturbation. If we assert that the electromagnetic wave is polarized in the z direction, we can describe our incoming light as a dipole perturbation in the form of

\[ eE_0 z \cos(\omega t), \]

where \( \omega \) is the angular frequency of the incoming light. This perturbation can have two differing effects on core electrons once they are ejected.

In XANES, the final atomic orbital states available to an excited photoelectron are determined by selection rules and by the Pauli Exclusion Principle, meaning an electron can only transition into empty and dipole transition allowed states. The available states for a photoelectron are the empty outer states, which is important because these states are the most important in the formation of covalent bonds. Thus, one application of XANES is the study of the covalence of atoms and the orbitals that these atoms use to form bonds.

The probability of a transition of a core electron to an available outer electron state can be described using Fermi’s Golden Rule, which gives the probability of transition from an initial state to a final state as

\[ W = \frac{2\pi}{\hbar} \left( eE_0/2\right)^2 \langle \psi_l | \hat{H}_0 | \phi_f \rangle^2 \rho(E_f). \]

Thus, the probability of transition between two states is non zero only if the matrix element \( \langle \psi_l | \hat{H}_0 | \phi_f \rangle \) between the sates is non zero for a give perturbation Hamiltonian \( \hat{H}_0 \) causing the transition. The patterns of allowed transition states form the selection rules for a system. For the K-edge excitations, which involves the ejection of the \( n = 1 \) s state core electron, an electric dipole perturbation as described above can only lead to transition to available \( p \) states (though quadruple perturbations can lead to \( d \) transitions).

The majority of information gathered from XANES experiments comes in the form of peaks in the absorption of light at certain energies. The XAFS graphs are in general dominated by one large peak, known for historical reasons as the white line, which represents the threshold energy at which core electrons are ejected away from the core.
atom and into the continuum. However, in many cases, depending on the outer orbitals of a particular atom, there exist what are known as pre edge features. These come in the form of peaks occurring before the main white line peak. Often smaller and sometimes blurred into background, these peaks are of crucial importance. They occur at the energies when electrons are given enough energy to be ejected from their core states but not enough energy to fully escape the atom. These electrons enter bound states among the outer orbitals, and if the state is involved in covalent bonding, it can be represented as a linear combination of orbital states from the ligand and states from the central metal atom. The relative influence of each is described by a mixing coefficient, and one can represent the state as

$$\Psi = c_L \phi_L + c_m \phi_m,$$

(4)

where $\phi_L$ and $\phi_m$ are the ligand and metal orbital states, respectively, and $c_L$ and $c_m$ are the mixing coefficients. An example of how two electron orbital states may mix is shown below.

The intensity of ligand pre edge peaks are related to the mixing coefficient for a particular orbital state and therefore the availability and electrical accessibility of that particular state. For this reason, determining these mixing coefficients is crucial to understanding the ways that a certain atom can form covalent bonds. The coefficients tell directly the relative contributions of the orbital states involved in the final state as a linear combination. The integral of a pre edge peak above background determines the coefficient for a particular state in the linear combination of possible outer bound states for quasi-ejected photoelectrons and therefore the relative strength of a state for forming covalent bonds.
Experimental Setup

The experiment was performed at the Stanford Synchrotron Radiation Laboratory at Stanford University. All experiments were performed using the 54-pole wiggler beam at the 6-2 hutch, which specializes in low energy studies. The samples were specifically chosen as to have orbital and structural geometries that are very easy to work with and perform calculations on. The samples used were UCl$_6^{n-}$ with $n$ being 1, 2, or 3 (i.e. U(V), U(IV), and U(III), respectively).

These compounds were preferred for examining the ligand bonds due to their simple octahedral geometry. The ring accelerator at Stanford provides, through the angular acceleration of charged particles in an electric field, light in the form of synchrotron radiation. This radiation comes as “white” light and must be filtered for a particular energy of light to be selected. Through the use of a double crystal monochromator and a Ni-coated harmonic rejection mirror, which uses the reflection of the light at certain angles to select certain energies of light, one can choose and dynamically vary the energy of light entering a sample.

The monochromatic synchrotron light was first passed through an initial ion chamber which serves to measure the initial intensity of the light. This was measured through the stimulation of ions, which allows a current to pass across a voltage within the ion chamber. The intensity of this current is related to the amount of ions created within the chamber and therefore the intensity of the light passing through the chamber and creating ions. After passing through the first ion chamber, the light was incident upon the sample, which was set at 45° relative to the beam. As light passes through the sample, its intensity is reduced in accordance with the equations above and the outgoing fluorescence is measured as data. The fluorescence was measured using an N$_2$-filled Lytle detector. The entire sample compartment was purged with N$_2$ to maintain the integrity of the samples and the purity of the air for the beam. The system was initially calibrated using a Cs$_2$CuCl$_4$ sample that acted as a control. This calibration was performed by fixing the maximum of the first pre edge feature at 2820.20 eV, which is an experimentally well-known value.

The method of holding the samples themselves was constructed with the consideration of reducing background noise and minimizing sources of error in mind. A major source of error in XAFS experiments is self-absorption within the sample itself. Therefore, it was necessary that the sample be dilute enough so that self-absorption be minimized but must exist in large enough numbers to obtain a signal that can actually be detected. The
samples were ground into a fine powder and spread over a layer of Kapton tape at several different concentrations. Samples were measured at several different concentration in order to monitor for and minimize self-absorption. The samples were held together on aluminum slides, which were capable of holding several different type of samples. These samples slides were mounted on top of a positioner to ensure alignment with the x-ray beam and the ion chamber.

Data Analysis

The data obtained in XAFS experiments comes in the form of normalized beam absorption by the sample in terms of energy of the beam. In analysis of both the XANES and EXAFS regions, the resulting data must be represented in a functional form through curve fittings. Often, for both theoretical and practical reasons, the curves are best fitted with the use of sums of arctangent, Gaussian functions, Voigt functions, or combination of any of the above. Background comes in the form of random noise and steady non peak slopes. The latter background slopes can be eliminated by subtracting off polynomial function fitted to this background. The entire curve is normalized such that the edge following the white line peak sits at a value of 1. The resulting data is then curve fitted to obtain a function that almost exactly approximates the XAFS peaks. It is necessary to represent the absorption and pre edge peaks as functions in order to perform the necessary mathematical analysis and obtain the desired experimental results. In order to obtain mixing coefficients for the XANES analysis, one must integrate peaks to find their relative strength.

Results

In order to fully understand the results of the data obtained through experiments, a good deal of analysis must first be performed. However, one can infer the nature of these final results by closely examining the raw data of an experiment. In the XANES case, this is possible because the main source of information, the pre edge peaks, are easy to visually pick out and analyze in an informal way.
The figure above plots the three different oxidation states of UCl$_6$ that were run in the experiment. The x-axis plots the energy, in eV, of x-ray light being shined on the sample. The y-axis plots the normalized absorbance coefficient of the light. The y-axis is normalized such that the edge directly after the large white line peak occurs at a value of 1. Other than the normalization of the y axis, the data plotted above is raw and direct from experiment. Its background has not been subtracted away through the use of curve fittings. The energy positions of pre edge features for each compound are plotted below.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Pre-edge1</th>
<th>Pre-edge2</th>
<th>rising edge</th>
<th>$\Delta_0$ (eV)</th>
<th>$\Delta_0$ (cm$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCl$_6$$^{3+}$</td>
<td>2822.2</td>
<td>2824.0</td>
<td>2825.9</td>
<td>1.8</td>
<td>14,515</td>
</tr>
<tr>
<td>UCl$_6$$^{2-}$</td>
<td>2822.2</td>
<td>2823.9</td>
<td>2825.6</td>
<td>1.7</td>
<td>13,709</td>
</tr>
<tr>
<td>UCl$_6$$^{3-}$</td>
<td>-</td>
<td>2823.9</td>
<td>2825.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These energy value positions of peaks and other features were obtained by analyzing the derivatives of curve fits for each sample.

From this simple plot, a wealth of chemical information can be obtained. The two distinct pre edge peaks can be identified due to their energy positions. The pre edge
features seen here represent two different split levels of the 6d electron orbital. In other words, the absorption peaks because the core 1s electrons are able to be stimulated to higher energy levels and end up in one of the two levels of the 6d orbital (the orbital is split into two different levels due to the presence of a ligand field). The first peak represents the triply degenerate $t_{2g}$ state, which leads to $\pi$ bonding, and the second peak represents the doubly degenerate $e_g$ state, which leads to $\sigma$ bonding. It is clear that outer 6d orbital plays an important role in covalent bonding. As the oxidation state increases, the intensity of the pre edge peak and therefore the relative intensity of the 6d states decreases. This is consistent with the increasing size of the outer orbitals. A larger orbital decreases the ability to form covalent bonds, and therefore the decreasing strength of pre edge features fits with the sizes of the orbitals in question of the oxidation states, which are 2.51 Angstroms for U(V), 2.63 Angstroms for U(IV), and 2.81 Angstroms for U(III). The exact nature strength of the covalency of this orbital will be determined after more rigorous analysis is performed. Further analysis should yield the exact mixing coefficients for the covalent bond state.

Many more subtle features can be discerned from the XAFS plot. The energy position of the rising edge, which is the position of the change in concavity that takes place just before the white line features, is shifted toward higher energies as the oxidation state decreases. This means that for lower oxidation states, it takes less energy to initially eject the chlorine core electron. The energy of the core 1s electron is shifted toward a deeper binding energy, and from this one can infer that there is a greater charge donation from chlorine’s outer electrons to the uranium metal.

There are two clear pre edge features for each sample, and the relative intensity of those pre edge peaks drop in intensity as the oxidation state decreases. Since the strength of the pre edge feature is the

The initial XAFS data has demonstrated that many initial assumptions about the nature of actinide covalent bonds were mistaken. Contrary to intuition, the somewhat shallow orbitals, such as the 6p, 5f, and, 6d have relatively strong contributions toward covalent bonding.

**Future Work**

The future of related experiments is unfortunately limited by the practicality of working with higher actinides. The high radioactivity of elements such as plutonium makes performing typical XAFS experiments a much more difficult task. It is necessary when working with such elements that the samples involved be fully enclosed and sealed with ultra high vacuum. Being forced to use uhv chambers, of course, effects the feasibility of these type of experiments. However, it indeed seems possible that suitable systems utilizing uhv technology can be constructed and will be able to perform under experimental conditions. Further experiments on higher actinides would serve to
supplement the data obtained from these uranium studies and would greatly expand the understanding of atomic chemistry and the nature of actinide and high atomic number element covalent bonding.

References

1. Strittmatter, RJ; Bursten, BE. American Chemical Society, 1991, 113, 552–559


Laboratory Calibration of a Space Plasma Spectrometer

Quinn Looker, University of Missouri-Rolla
Elizabeth MacDonald, Los Alamos National Laboratory

Abstract

Though space is a near vacuum, there is a myriad of particles propagating through the space environment. With a steadily increasing armada of satellites and space vehicles, all containing sensitive electronics, space weather has become an important factor that must be monitored. A great number of particles stream from the Sun, and as variations in this solar wind couple into the Earth’s magnetosphere they can have many significant effects on artificial satellites and even on the Earth’s surface. The Advanced Miniaturized Plasma Spectrometer (AMPS) was designed to detect magnetospheric plasma particles in order to determine their flux and its variation with time and space. AMPS utilizes an electrostatic analyzer and a solid state detector to provide information about the charge, mass, and energy of incident particles while maintaining a very lightweight and compact instrument profile. In order to quantify the detector’s response to space plasma particles, it must be properly calibrated. Instrument effects due to the solid state detector response, electronics noise, carbon foil, and detector geometry must be characterized and accounted for. This presentation is a description of the methods used to test and calibrate the AMPS solid state detector subsystem.

Introduction

Though there is seemingly little occupying the space around Earth, the planet’s influence extends far beyond the point where the atmosphere fizzes away. The Earth’s magnetic field extends many times further than its radius. This field interacts with the magnetized solar wind as it constantly streams toward Earth. Hydrogen and helium ions, of which the solar wind is mostly comprised, become trapped in the Earth’s magnetosphere and follow the magnetic field lines around the Earth. Also, hydrogen and oxygen ions rise from the ionosphere into the space environment and join the solar wind particles in complex orbits around Earth. The orbit of charged particles around the planet
creates a ring current as well as radiation belts. These particles can also precipitate into the atmosphere, creating the auroras.

Besides interesting effects of this plasma, there are also many practical reasons to study the distribution and transport of these particles. The electrons that orbit the Earth inevitably collide with artificial satellites, quickly building up a surface charge on the outer hull that can disrupt measurements. Also, electrons can penetrate into the spacecraft and potentially cause arcs that could damage sensitive electronic equipment. Oxygen is thought to cause significantly more damage than hydrogen when it collides with spacecraft materials, such as solar panels, which can significantly reduce the power availability and lifetime of the satellite.

A system of Los Alamos National Laboratory (LANL) satellites is already in place to monitor ion concentration in geosynchronous orbit, but all detectors thus far have only served to detect the presence of ions and electrons. No mass-specific properties of the particles can be distinguished directly since previous plasma spectrometers utilize channel electron multipliers (CEM) in order to detect incoming ions. A CEM is essentially a small tube coated on the interior with glass that is doped to emit many secondary electrons following impact. When an incoming ion or electron hits the wall of the CEM, it causes secondary electrons to bounce off, and these in turn hit the wall and cause more electrons to come off in a cascade. This results in a measurable current after a gain of about a million and the presence of a particle can be detected. However, the CEM itself contributes no information besides the presence of a particle.

Because of the relative abundance, all ions thus far have been generally assumed to be hydrogen. A new plasma spectrometer called the Advanced Miniature Plasma Spectrometer (AMPS) is being put into place to measure the composition of this plasma using a solid state detector (SSD). In this way AMPS will be able to trace the concentration of oxygen ions, and therefore track the motion of ionospheric ions through the space environment and measure their contribution to dynamic processes in the magnetosphere. Oxygen only comes from the Earth’s atmosphere, so this distinguishes it from solar wind particles.
AMPS

AMPS was designed to be a next-generation plasma spectrometer that could perform that function while remaining very lightweight, compact, and power-conserving. The device itself is not more than a few inches across in any direction and mounts as part of the SABRS (Space and Atmospheric Burst Reporting System) Validation Experiment (SAVE). AMPS uses a typical hemispherical electrostatic analyzer (ESA) to simultaneously detect ions and electrons. Figure 1 shows the ESA and its three nested hemispheres, where the outer two are grounded and the middle has an applied varying voltage. This creates an electric field radially away from the middle hemisphere, allowing ions to take the inner track and electrons to take the outer track. The balance between electric and centripetal forces allows the device to select energy per charge of incoming particles. Using a single high voltage power supply, the middle hemisphere is logarithmically stepped from 0 to 5 kV to allow detection of particles from 2eV/q to 40 keV/q.

![Figure 1. A cross-section schematic of the electrostatic analyzer.](image)

AMPS has five CEMs and one solid state detector arranged opposite the apertures to detect the individual ions and electrons that traverse the ESA.

The Solid State Detector

The most innovative detection device in AMPS is the solid state detector. This is what allows AMPS to complete its challenging mission to differentiate particle species. The solid state detector is a subsystem in itself, including a plastic tower that contains the
The detector is preceded by a thin carbon foil for several reasons. AMPS is designed to detect low energy particles down to 2 eV, which is impossible for a silicon detector. The foil is biased to -10 kV so the lowest energy particles gain energy to rise above the detection threshold for the SSD. Another purpose for the carbon foil is to neutralize the incoming ions. The positively charged ions pick up electrons from the carbon to become neutral, and continue to the detector with nearly all of their initial energy and direction preserved. Though only about 75% of the ions are neutralized in this way, the remaining ions are swept away by the other arm of the metal holder.

Lastly, the foil serves to partially scatter the incoming particles. This is important because of an effect called channeling, to be discussed shortly.

The operation of the solid state detector is heavily dependent on energy deposition. The detector itself actually picks up the amount of deposited energy, but energy can be deposited in different ways. An initial portion of particle energy is lost in what is called the dead layer of the detector. Due to the manufacturing process for a solid state detector, a portion on the face of the detector cannot contribute to the resulting pulse out of the detector. Since AMPS is designed to detect low energy particles, it is important to have a very thin dead layer so little energy will be lost in this fashion. The
AMPS solid state detector has a dead layer about 80 Å thick, which is excellent relative to the 500 µm total detector thickness. Beyond the dead layer is the active layer, where energy deposited has the possibility of being detected. However, to determine an ion’s deposited energy, the particle must be completely stopped in this active layer or it leaves the back of the detector with some unknown amount of energy. For the energy range of particles used in our experiments, nearly all of the ions were stopped in less than a micrometer.

In any region of the detector, energy is lost in two forms. A reverse biased solid state detector detects incoming particles by the creation of electron/hole pairs in the silicon crystal. This happens only during electronic stopping, meaning the incoming ion interacts with the silicon electrons and transfers energy to them. The ionization energy for the outer electron in silicon is 3.6 eV, so every 3.6 eV of ion energy deposited in electronic stopping creates an electron/hole pair. Nuclear stopping is when an ion interacts with silicon nuclei and transfers some of its energy. The silicon nucleus is either set into vibration or knocked out of the lattice completely. Ion energy lost in this manner cannot be detected since it does not create an electron/hole pair.

Another effect related to nuclear stopping is channeling, where some ions are allowed to travel between more nuclei than average. This is a result of the face-centered cubic structure of the silicon crystal, which has a <100> facing directly toward the detector surface, meaning 0° is a channeling axis. There are certain “channels” along which an ion can travel and encounter fewer nuclei than the average density of the silicon, meaning less nuclear stopping.

The silicon detector is a p-n junction with p on front and n on back. The front of the detector is biased to -200 V and the back is grounded, so this creates an electric field across the detector. When an electron is knocked from a silicon atom it is quickly forced to the back where it is picked up as a pulse. However, there is some leakage current associated with this bias, thereby limiting the detection efficiency and lower limit of particle energy that can be detected.
Laboratory Calibration

The laboratory calibration was a simple set of experiments that involved shooting a beam of energetic ions at the solid state detector to test its response and operation. The tests were performed on a spare copy of the solid state detector subsystem, as the flight instrument had already been sent to Florida for impending launch. The SSD and pre-amplification electronics were placed in a vacuum chamber, which was relieved of air. An ion beam was directed into the chamber toward the detector. The beam apparatus was a slightly limiting factor because of the way it produced ions. Essentially air was let into a chamber where it was subjected to microwave radiation to ionize the atoms. While an effective method, it can only produce singly charged ions and not the He$^{++}$ that comes from the solar wind. These positively charged ions are then put through a region at negative potential determined by a high voltage power supply. In this manner the ions can be accelerated to energies up to 60 keV. We used beam energies from 8-60 keV for hydrogen or about 20-60 keV for oxygen. The beam proceeds immediately to a chamber containing an adjustable magnet, where the beam is curved in the magnetic field and different ion masses can be selected by adjusting the field strength. Directly after this chamber are four perpendicular metal plates attached to four different power supplies. Adjusting the voltage on these plates steers the beam and adjusting their signal frequency can make the beam narrower or wider. The end result is a monoenergetic beam of a single ion species that is adjustable in many ways.

This beam is fired at the solid state detector and the signal is collected after pre-amplification by a shaping amplifier. This allows the signal to be properly shaped into a peak, and then it goes to an external lab amplifier. After a gain of 50, the pulse then travels to a multi-channel analyzer (MCA) to be converted to a digital signal that is sorted into bins. The result is similar to a histogram, where a set of channels count all pulses within a certain voltage range. The multi-channel data set was taken using the MAESTRO program and analyzed and plotted using IDL.

Noise

Noise became a major issue in the lab measurements on the solid state detector. AMPS is designed to detect low energy particles, but the lab is a much noisier
environment than space, so noise obscured many of the low energy particles. Extraneous noise signals affect every part of the calibration process from the beam to the detector and all the way through the amplification electronics. Besides tuning the beam and leakage current from the solid state detector, some noise enters the vacuum chamber and is included in pre-amplification. However, the vacuum chamber makes a very good shield since it entirely encloses the SSD in a metal surface. Though we had favorable shielding, the multi-channel digital signal did not show any noise whatsoever, and the peaks were unexpectedly asymmetric. The left edge of the peak was very straight with a very high slope, making it appear that the lower end of the peaks were being cut off. The laboratory equipment was investigated for possible causes.

The main problem with the signal cutoff was the Ortec Ethernim 919E multi-channel analyzer. Its automatic noise rejection system had no manual controls and was found to be the source of the cutoff by using a different MCA. Our MCA used a system that assumed all signals were positive and all noise was symmetric about ground. It set a maximum threshold at the peak of the most negative signal, then inverted this threshold to the positive side and cut out everything below. For some unknown reason, our noise was initially more negative than positive, so this increased our positive threshold for noise and cut out part of our real peaks we wanted to observe. When this was found, we corrected by using a positive DC offset on the signal into the MCA to lower the positive noise rejection threshold. This allowed our entire ion peak through the MCA along with the noise. An example is shown in figure 3 of the typical noise and peak signal with the MCA threshold system.
Figure 3. Typical analog signal and noise threshold.

This solved the noise rejection problem, but revealed another problem involving a drifting DC offset. The external amplifier was not reliable enough to keep a constant offset, and as it shifted it moved the actual peaks and changed the noise threshold significantly. Figure 4 shows an example under controlled conditions of how increasing DC offset shifts the ion peak by that amount, but also lowers the noise threshold so the noise peak expands to the left and grows.

![Graph showing 60 keV H with Varying DC offset](image)

Figure 4. A typical ion signal with varying DC offset.

Since this was discovered after many measurements had already been taken, it is possible that much of the error or varying background was introduced into our data by the drifting DC offset, of which we were ignorant.

The problem with seeing the entire ion peak was resolved, but in doing this we allowed noise peaks into all measurements. In the lower energy runs, the peak begins to blend with the noise and becomes indistinguishable. We wanted to observe these peaks since AMPS is designed to detect lower energy particles, so noise subtraction became necessary. By taking separate runs with no ion beam, we recorded the noise signature at various intervals and subtracted this from the beam data. In theory this leaves only the
ion peak, but in practice this was somewhat inaccurate because of shifting noise levels and the drifting DC offset.

**Analysis**

The actual beam measurements took perhaps two weeks, while analysis of the data has lasted for a month and still continues. All of the signals that resulted were arrays of 4096 elements (covering 12 volts), so it was necessary to convert them to a useful format and put them in some form that could be easily analyzed. Plots are the obvious solution to this problem since most effects can be clearly visualized, and the IDL programming language was used to convert and create these graphs in a versatile fashion.

The simplest method of testing the solid state detector response was the detector by itself with no carbon foil. This way we could directly observe the response when a certain particle energy was shot into the detector. Figure 5 shows an example of particles with the same energy shot into the detector and the different responses.

![Graph showing different ion species with the same energy](image)

**Figure 5.** An example of different ion species with the same energy.

The most important aspect of figure 5 is the clearly defined peaks, especially between oxygen and hydrogen. These are the two species we want to distinguish the most, and their peak channels are well-separated because of the much greater amount of nuclear
stopping occurring with oxygen when compared with hydrogen. Since more energy is deposited in this non-detectable form, less energy is picked up in the SSD. This validates the capabilities of AMPS to separately measure oxygen and hydrogen populations at geosynchronous orbit. However, the oxygen peak is noticeably very wide compared with hydrogen and helium. This is a serious detriment to energy resolution, and is caused by channeling. The oxygen peak without channeling is most likely to be at the left edge of the peak seen in figure 5, but some ions are penetrating much further into the detector. They encounter fewer nuclei and therefore have less nuclear stopping. Because this is most significant for oxygen, it is much more obvious than in the hydrogen or helium peaks.

The next experiment was to add the carbon foil to the front of the detector, but to eliminate uncertainty the foil was not yet biased to -10 kV. Instead it was grounded and placed on the solid state detector and the same experiments were performed as before. An example at the same energy is shown below.

![Figure 6. Example of ion species with the same energy and the carbon foil added.](image)

The most significant feature of figure 6 is the much narrower oxygen peak, found at the left side of the previous wide peak as expected. This can be explained by ions being slightly scattered by the carbon foil, so they are mostly off the channeling axis and little channeling occurs. Another important effect is the loss of energy in the foil. All of the
peaks are consistently shifted to the left when compared with the previous plot, reflecting this energy loss.

The next step was to put voltage on the foil. If the foil works properly, then it should add the appropriate amount of energy corresponding to the bias on the foil and the peak should mimic that of ions with higher energy. Included in figure 6 is a dotted line showing an experiment where 50 keV hydrogen atoms were used with -10 kV on the foil. In theory, this should be the same peak as a 60 keV beam with grounded foil. As seen in the plot, the lines do not match exactly, but they are close enough to be within expected error due to noise ambiguity. This experiment was to validate that the foil was adding the correct amount of energy and had no problems carrying –10 kV. While it may seem trivial, it was important to test, because if it had not worked the device would need serious revision.

The previous figures were simply examples of 60 keV hydrogen, but the effects observed are important for all species at all energies. Figure 7 shows the peak channel number for all energies and species with foil and no foil runs. A few of the lower energy runs, specifically hydrogen below 12 keV and helium and oxygen below 20 keV, are not shown because the noise effects became significant. The ion species are clearly separated and the foil measurements are consistently in lower peak channels than the non-foil measurements, allowing the energy loss to be measured.
Figure 7. Peak channels of all energies of ion species.
The hydrogen and helium peaks seem to follow straight lines, while the oxygen is less well-behaved due to channeling and noise issues. We also had to monitor the energy resolution over the entire range of particle energies to ensure it was sufficiently low to differentiate between them. We measured the energy resolution using the full width half maximum (FWHM) of the peaks for all energies and species taken. While it is expected the energy resolution should decrease at higher energy, if it gets so that the peaks are so wide they blend together then the data are useless. Shown below are all of the collected FWHM values.

![Full Width Half Max vs. Particle Energy](image)

Figure 8. FWHM of all energies and ion species.
The hydrogen and helium stay fairly low even up to 60 keV, but the oxygen steadily increases with energy. However, the peaks were still narrow enough to separate the peaks for different energy.

Now that we knew how the SSD system reacted at different energies, we needed to find out how much of that energy was being lost in non-detectable forms, namely the dead layer and nuclear stopping. In order to make these predictions, we used the SRIM program (Stopping and Range of Ions in Matter, Ziegler). This computer program uses typical tables for the rate of energy loss per unit distance traveled in the detector (dE/dx)
due to nuclear and electronic stopping. When these tables were applied to the energies used in the lab, we calculated the following results.

Figures 9 & 10. Predicted energy loss in dead layer and to nuclear stopping.

It is apparent from the plots that little energy is lost to the dead layer, and it is fairly consistent. Nuclear stopping takes much more energy for oxygen and provides the stark contrast observed between hydrogen and oxygen peaks. A combined table was created using incident particle energy and the predicted amount of energy that would be detected. This was done by subtracting the predicted nuclear and dead layer energy from the incident energy.

Figure 11. Incident particle energy and predicted deposited energy.

If all energy were deposited as electronic stopping, the lines would follow the dotted black line shown with a slope of one. Hydrogen follows this line quite well, thought
slightly lower. Oxygen, however, is significantly below the line because of significant nuclear stopping.

Finally this prediction was combined with our experimental data. The observed peak channel at a given incident energy was compared with the predicted deposited energy at that incident energy. The comparison is shown in figure 12 with the plus signs.

![Figure 12. Peak channel number and predicted deposited energy.](image)

The values for each species should all follow the same straight line since a given amount of energy deposited should give the same peak channel. Hydrogen and helium follow the line very well, yet oxygen strays from the line significantly. This called for an investigation into the accuracy of the SRIM predictions.

As we knew from the beginning, the SRIM predictions would not be perfectly accurate. The primary reason is they assume constant dE/dx, which is simply not true. We chose the dE/dx value from the table that corresponded to the incident particle energy, but dE/dx does not remain constant since the particle continually loses energy as it travels through the detector. This is certainly a subject for future work, but a simple test was performed to determine if this was the most significant source of error. We incorporated previous experimental results taken to directly measure nuclear stopping (Funsten et al, “Fundamental Limits to Detection of Low-Energy Ions using Silicon Solid-State Detectors”). The plus points on figure 12 are the SRIM predictions and the X
points are the experimental predictions. The new set shows much better agreement with the straight line, which leads to a reasonable conclusion that the SRIM predictions were off in the case of oxygen. The straight line relationship of figure 12 is important because it allows us to quantify all of the laboratory results from MCA channels to an absolute deposited energy.

**Conclusions and Future Work**

The most important result from the calibrations is that we have a clear method of separating ion species detected in geosynchronous orbit. When a particle goes through the SSD, we know its energy because of the ESA. We get a response centered around a certain peak channel, which tells us the amount of deposited energy. This deposited energy compared with the incident energy tells us how much energy was lost in non-detectable forms, and this energy loss is dependent on mass. Therefore, we can measure different particle species.

The lab measurements for the SSD calibration are complete, but there is still work that can be done to improve the results. No tests were performed with the hemispheres. Also, the SRIM calculations can be improved by adjusting dE/dx as energy is lost. Channeling is another important effect that can be corrected. It is not particularly applicable to AMPS because the flight instrument has already been sent to Florida and cannot be adjusted for this effect. However, future instruments could benefit by having the SSD turned by about 20 degrees to keep all particles off the channeling axis and improve energy resolution.
Acknowledgements

I would like to thank Elizabeth MacDonald for all of her help as a mentor. She did an excellent job, especially in helping me revise my poster and paper. Also, Paul Janzen and Ron Harper helped me greatly with learning things in the lab.
Another thanks goes to Lee Collins, James Colgan, and Sally Seidel for running the summer school and organizing this great opportunity for us.
And finally, Norm Magee was incredibly generous with the entirety of our group. We can’t even begin to thank him enough. What a guy.
Magneto-optical spectroscopy of semiconductor quantum dots containing single magnetic atoms

Kate Raach
LASS, University of New Mexico, Los Alamos, NM 87544
T-4, Los Alamos National Laboratory, Los Alamos, NM 87545

Mentor: Scott Crooker
MST- NHMFL, Los Alamos National Laboratory, Los Alamos, NM 87545

We report on the photoluminescence (PL) of magnetically (Mn) doped ZnSe/CdSe nanocrystal quantum dots subject to high pulsed magnetic fields. With increasing magnetic field the low temperature (1.5 K) PL becomes circularly polarized, indicating the presence of spin polarized excitons. The PL does not, however, show especially large magneto-optical effects, suggesting that the Mn atoms may not reside in the cores of the nanocrystal quantum dots.

Introduction

A nanocrystal quantum dot (NQD) is a roughly spherical semiconductor crystal, the diameter of which ranges in size from 3 to 10 nm. It is large enough to have a well defined interior and exterior but is small enough to exhibit quantum mechanical effects due to the confinement imposed by the crystal boundaries. Among these effects are discrete, atomic-like densities of states and size-tunable band gap energies across the visible spectrum. These qualities make NQDs attractive for use in biological applications such as biotagging and in optical applications such as diode lasers and amplifiers. NQDs are also a potential candidate for quantum computing qubits. In order for this to be done a single spin must be isolated and its magnetization manipulated, detected, and read out. To measure the magnetization of a single Mn atom is no easy task. Nuclear magnetic resonance instruments need a minimum of about $10^{12}$ polarized nuclear spins to detect the magnetization. Electron spin resonance spectrometers can detect magnetization to about $10^7 \mu_B$, where $\mu_B$, the Bohr magneton, is the magnetic moment of a single electron. Among the most advanced detectors is the cryogenic dc SQUID magnetometer which can detect to about $10^3 \mu_B$. An alternative to directly measuring the magnetization is to infer it by exploiting the optical properties resulting from coupling in magnetically doped semiconductors, called “diluted magnetic semiconductors” (DMS). Since the 1970s, DMS, such as $\text{Zn}_{1-x}\text{Mn}_x\text{Se}$ have been well-studied. It is known that the spin orientations of the magnetic atoms in DMS dramatically alter the radiative transitions of the semiconductor. In bulk material, however, it is not possible to constrain free electrons and holes to a volume about a single Mn atom. NQDs are ideal for this. Modern chemistry makes it possible to dope NQDs with a single manganese atom and it is well established that experimentally a single NQD can be isolated and its PL studied. A Mn atom has spin $5/2$, and so 6 possible spin projections. The PL of a single Mn doped NQD should, due to the spin-exchange interaction between the electron-hole pairs and the Mn spin, exhibit 6 peaks, one for each allowed Mn spin projection. The intensity...
of each peak should then be proportional to the probability of occupying that spin state. Before performing single dot studies of the Mn doped NQDs, we need to know the Mn has actually been incorporated into the interior of the NQDs. Here, we investigate the spectra of an ensemble of Mn doped NQDs in an effort to detect the presence of the Mn. In particular we’re looking for enhanced magneto-optical properties such as seen in bulk DMS. Namely, we’re looking for a giant Zeeman effect, with a g-Lande factor much greater than unity, and a shift in the peak PL intensity energy with an applied magnetic field.

**Background**

Figure 1 shows the density of states in semiconductors as the dimensionality is decreased: bulk (3D), quantum wells (2D), quantum wires (1D), and quantum dots (“0D”). Density of states versus energy is a measure of how many electrons may occupy certain energy levels. In a bulk semiconductor, there are so many possible states that the density of states is essentially a continuous function. As each spatial dimension is restricted, the density of states develops more quantized features. The 0D quantum dot has only discrete allowed electron energies, much like a single atom.¹

![Figure 1. The density of states as a function of energy for a bulk semiconductor (3D), a quantum well (2D), a quantum wire (1D), and a quantum dot (“0D”), from Reference [1].](image)

Figure 2 shows the band structure of a semiconductor. In bulk semiconductors, there are bands of allowed energies and thus also bands of forbidden energies. The highest fully occupied band is called the valence band and the next unoccupied band is called the conduction band. The minimum energy input required for an electron to move from the valence band to the conduction band is called the band gap energy, \( E_g \). To excite an electron is to supply it with enough energy to jump from the valence to the conduction band. The absence of an electron is called a hole. This electron-hole pair is called an exciton, bound by the Coloumb interaction. After some time, picoseconds to microseconds, the electron will relax back to the valence band and emit a photon with energy equal to that of the gap. This process of the electron falling back to the hole and emitting a photon of energy \( E_g \) is called radiative recombination. Photoluminescence is the name for the light emitted by radiative recombination when the initial electron excitation energy was provided by a photon.
Excitons in a non-magnetic NQD have two optically allowed, spin degenerate states, $J = \pm 1$. Much like a magnetic field lifts the spin degeneracy of electron spin states, so it does for the exciton spin states. This is called Zeeman splitting. The energy difference between the two states, $J = +1$ and $J = -1$, is $E_Z = g_{ex} \mu_B B$, where $g_{ex}$ is the exciton Lande $g$-factor, $\mu_B$ is the Bohr magneton (58 μeV/T), and $B$ is the magnetic field in Tesla.

Excitons in a magnetic NQD also have two optically allowed, spin degenerate states, $J = \pm 1$. In addition to the Zeeman splitting, there is an exchange interaction between the exciton spin and the spin of the magnetic atom. As the magnetic field the NQD is subject to is increased, the additional exchange interaction ought to increase the energy difference between the spin states. The energy difference will be of the form $E_{split} = g^* \mu_B B$, where $g^* > g_{ex}$ is an effective exciton Lande $g$-factor.

Due to angular momentum selection rules, right circularly polarized light, $\sigma^+$, couples to the $J = +1$ state and left circularly polarized light, $\sigma^-$, couples to the $J = -1$ state. We excite the electrons of our NQD sample with a laser and measure the intensity of the left circularly polarized and right circularly polarized PL.
Experimental Set-up

Figure 3. a. Cartoon of a Mn doped ZnSe/CdSe nanocrystal quantum dot. b. TEM (transmission electron microscopy) picture of a quantum dot.

A cartoon depiction of an NQD is shown in Figure 3a and an actual image, taken using transmission electron microscopy, is shown in Figure 3b. The core of our NQD is cadmium selenide with, statistically speaking, a single manganese atom. The shell is zinc selenide. The sample is a colloidal suspension of NQDs in toluene.

Figure 4. a. Schematic diagram of setup. b. Photograph of a 50 Tesla pulsed magnet outside of the cryogen bath.

Figure 4a shows our experimental set-up. We use a helium-cadmium laser with a wavelength of 442 nm to excite our NQD samples. The laser light is coupled into a 600 μm diameter optical fiber and sent through a probe onto which a thin film circular polarizer and the sample are mounted. The circular polarizer is made up of a linear polarizer and a quarter wave plate.
The tip of the probe is centered in the National High Magnetic Field Laboratory’s 50 Tesla pulsed magnet, see Figure 4b. Figure 5 shows a characteristic profile of the magnetic field as a function of time. The peak field is 44 T. The generated pulse lasts approximately 50 ms; 9 ms to peak and about 40 ms to return to zero field. The PL returns through the fiber optic cable and is directed into a spectrometer. A liquid nitrogen cooled CCD camera records the spectra at a rate of 660 Hz. Roughly 35 field dependent spectra can be obtained in one shot of the magnet.

![Figure 5. Pulsed magnet's field as a function of time.](image)

The emission is measured at temperatures of 1.5, 70 and 78 K. For the measurements taken at 1.5 K the probe is immersed in a bath of pumped liquid helium, which is surrounded by a vat of liquid nitrogen. For the measurements at 70 and 78 K, the liquid helium is not used.
Results/Discussion

Figure 6 displays the intensity of the PL emitted by the dots at 1.5 K. With increased magnetic field the PL does become circularly polarized. Figure 6a is a measurement of the left circularly polarized light, $\sigma^-$, and Figure 6b is a measurement of the right circularly polarized light, $\sigma^+$. The intensity of $\sigma^-$ continually increases with field while the intensity of $\sigma^+$ decreases and remains constant for fields greater than 15 T. The appearance of this circular polarization with magnetic field indicates the presence of spin-polarized excitons. The PL is predominately left circularly polarized, indicating more excitons in the $J = -1$ spin state.

At zero field the peak PL intensity is at an energy of 2.08 eV, or a wavelength of 596 nm. As the magnetic field is increased, the energy corresponding to the peak intensity does not shift. Based on the radiative transition shift observed in bulk DMS, if there were many Mn atoms present in the NQDs, with increased magnetic field we would expect a shift in the peak PL intensity energy. The absence of this shift suggests the Mn is not in the core of the NQD. In the fabrication of the NQD, the Mn may have just been deposited on the shell or may not have incorporated into the dot at all.
The intensity spectrum for each field is integrated and then the data is normalized such that the zero field integrated intensity is equal to one. Figure 7 shows this normalized PL intensity as a function of magnetic field at 1.5 K and 70/78K. The degree of polarization is significantly less at higher temperatures because the thermal energy, $k_B T$, becomes comparable to the Zeeman energy, $E_Z$.

Above a magnetic field of 10 T, the intensity of the $\sigma^+$ PL remains at 0.75 of the original intensity. On the other hand, the intensity of the $\sigma^-$ PL increases with field until about 35 T where it plateaus at 1.7 the original intensity. Based on the behavior of bulk DMS, if Mn were present in the NQD, we would expect to see the right hand polarization $\sigma^+$ go to zero and the left hand polarization $\sigma^-$ to steadily increase. The light would be completely circularly polarized.
The degree of circular polarization, as a function of magnetic field, is

\[ P(\vec{B}) = \frac{I_{\sigma^-} - I_{\sigma^+}}{I_{\sigma^-} + I_{\sigma^+}}, \]

where \( I_{\sigma^-} \) is the integrated intensity for \( \sigma^- \) and \( I_{\sigma^+} \) is the integrated intensity for \( \sigma^+ \). This is calculated using the data from Figure 7 for both \( T = 1.6 \) K and \( T = 70/78 \) K. It is plotted in Figure 8. The degree of circular polarization saturates at less than 100 \% because \( I_{\sigma^+} \) saturates at a value greater than zero. If Mn causes the right circular polarization to go to zero at high fields, we would expect the degree of polarization to approach 100 \% at high fields.

**Conclusion**

The emission of circularly polarized light from the NQD gives evidence of spin polarized excitons. It did not appear, however, that there was Mn in the core of the NQD. With an increasing magnetic field, we expected to see a shift in the energy corresponding to peak intensity of the PL. We also expected the right circularly polarized component of the emission to go to zero and the left circularly polarized component to continue increasing. This would have resulted in saturation of the degree of circular polarization at 100 \% instead of 40 \%. The absence of these effects in the PL suggests the Mn is not present in the core of the NQD. Further refinement in the fabrication of Mn doped NQDs will hopefully make it possible to detect, via magneto-optical spectroscopy, the Mn in an ensemble of doped NQDs.

**Acknowledgements**

I would like to thank Scott Crooker, my mentor, and Madalina Furis, Scott’s post-doc, not only for putting up with me the entire summer but also for welcoming me into the lab and letting me do research with them. I would like to thank Norm Magee, Lee Collins, Daniel James and Sally Seidel, all of whom helped make it possible and pleasurable for me to be a part of LASS this summer.
References

TCR-mediated T cell activation by kinetic proofreading and intracomplex transphosphorylation

Michael J. Saelim∗ and William S. Hlavacek† (mentor)

August 11, 2006

Abstract

There has been much recent interest in examining the processes behind T cell activation upon contact with infected antigen-presenting cells. Qi-Jing Li et al. (2004) proposed a model for the early signal transduction steps, including involvement of both CD4 and endogenous peptide-MHC complexes, which seems to explain the T cell’s ability to discriminate between agonist and endogenous peptide. We examined this model by using BioNetGen2 to generate a network of chemical species that was evaluated both by ODE and stochastic calculations under various initial conditions. Here, we raise some concerns about the model’s behavior. Furthermore, we sought an alternative model to explain the first steps of T cell activation based on a kinetic proofreading framework and signal amplification by transphosphorylation inside the signaling complex. While this model has not yet been fully studied, we present our current findings on its applicability to modeling T cell behavior.

Introduction

While it was once thought that only foreign, agonist peptide-major histocompatibility complexes (pMHCs) on antigen-presenting cells (APCs) are involved in triggering T cell activation and immune system response, recent studies have shown the importance of endogenous pMHCs in lowering the threshold for the number of agonist pMHCs needed. CD4 co-receptor molecules on T cells have also been implicated in this sensitization to agonist peptide [1]. Evidence has also been revealed showing that CD4 dimers can dimerize MHC molecules [2, 3] and that Lck, a Src-family kinase, binds to the cytoplasmic tail of CD4 [4]. In addition, agonist and endogenous pMHCs have been found to differ in the half lives of their interactions with T cell receptors (TCRs) [5].

The curiosity of T cell activation stems from its ability to not only discriminate between endogenous and agonist pMHCs, but also to form immunological synapses when only a few agonist pMHCs are present while triggering as many as 1500 TCRs in the process.

∗Michigan State University, through the Los Alamos Summer School
†Theoretical Biology and Biophysics Group (T-10), Los Alamos National Laboratory
[6, 7]. So far, it is only known that some process involving these pieces results in the phosphorylative activation of ZAP-70, a Syk-family kinase, which goes on to activate further steps that conclude in immunological synapse formation [8]. The challenge is thus to create a suitable model to explain the first molecular interactions that trigger T cells.

Qi-Jing Li et al. have developed a model for these first steps of T cell activation in which agonist pMHC-TCR binding triggers the formation of agonist-endogenous pMHC heterodimers mediated by CD4 [6]. Although the model seems to generate both discrimination and activation at low amounts of agonist pMHC, our analysis with BioNetGen2, a program simulating biochemical networks, presents some interesting and concerning results, which we enumerate here.

### The Li et al. Model

#### Theory

This model, further detailed in [6], is dependent on a series of complex formation steps to supply both kinetic proofreading and signal amplification. A free pMHC binds to a free TCR, and this pMHC-TCR complex is recruited by a permanently bound CD4 dimer-Lck complex, with CD4 binding to the pMHC and Lck binding to the TCR. At this time, Lck is free to phosphorylate the single immunoreceptor tyrosine-based activation motif (ITAM) on the TCR, while CD4 can recruit a second pMHC. Li et al. suggest that because of this five-member signaling complex, any TCR that binds to the second pMHC will be spatially localized close to a Lck molecule, which can speedily bind to and phosphorylate the ITAM on the second TCR. However, if at any time the original pMHC-TCR bond breaks, the complex completely dissociates. Because agonist pMHC have longer half-lives than their endogenous counterparts, kinetic proofreading allows complexes to form if they are started by agonist pMHC-TCR binding, but discourages complexes started by endogenous pMHC-TCR binding, since the bond will not last long enough. Furthermore, signal amplification is achieved because the spatial localization of TCR and Lck allows the TCR ITAM to be phosphorylated before it dissociates from endogenous pMHC. Thus, signal

Figure 1: This graphical representation of the main kinetic proofreading process in the Li et al. model shows the steps of complex formation: (a) agonist pMHC-TCR binding, (b) pMHC-CD4 binding and TCR-Lck binding, (c) recruitment of an endogenous pMHC, and (d) serial triggering of TCRs by the signaling complex.
amplification results from the serial triggering of TCRs from signaling complexes. Figure 1 shows the main process of this model.

From the simulation of this model, calcium traces are then generated with the following equations, based on a model proposed by J.W. Putney [9].

\[ \frac{dZ_{ER}}{dt} = -k_{ER}H(\sum_i P_i - P_0)Z_{ER} \]  

(1)

\[ \frac{d[Ca^{+2}]}{dt} = k_{ER}H(\sum_i P_i - P_0)Z_{ER} - \mu[Ca^{+2}] + k_1\frac{[Ca^{+2}]}{[Ca^{+2}] + K_{Ca}}(\sum_i P_i + K_P) \]  

(2)

where \( Z_{ER} \) is the endoplasmic reticulum calcium concentration, \([Ca^{+2}]\) is the cytosolic calcium concentration, \( \sum_i P_i \) is the number of phosphorylated TCR, and \( H() \) is the Heaviside function. Li et al. assume that calcium levels are sustained if \([Ca^{+2}]\) stays above 0.001 M between \( t = 360 \) s and 600 s. Exact numbers for the parameters are given in [6].

**BioNetGen2 Analysis**

From the supplemental materials and methods supplied by the authors, we were able to encode the rules into BioNetGen2 and generate the chemical network. The conversion was not difficult because both their model and BioNetGen2 use a rule-based approach to modeling biochemical systems. A copy of the reactions from the Li et al. supplementary materials is given in Appendix A, and a sample input (.bngl) file is shown in Appendix B. The model to produce the calcium traces was also encoded in Java, and entails an Euler method approach to numerically evaluating the calcium output. The code for this program is shown in Appendix C.

Because BioNetGen2 currently does not have a formal implementation of a zeroth-order reaction rate for TCR dephosphorylation, an approximation was used involving the saturated rate model, which implements Michaelis-Menten kinetics. For a rate input of \( \text{Sat}(\alpha, \beta) \), the rate law is

\[ \frac{d[P]}{dt} = \frac{\alpha \cdot [R]}{\beta + [R]} \]  

(3)
where $[P]$ is the product concentration and $[R]$ is the reactant concentration. The approximation is implemented by setting $\alpha$ to be the zeroth-order rate and $\beta << [R]$, so that $\frac{d[P]}{dt} \rightarrow \alpha$. This implementation is shown in Appendix B.

Results

We were able to qualitatively reproduce the \textit{in silico} results of the Li et al. model with our implementation for BioNetGen2 and our calcium trace program, by running the BioNetGen2 calculations stochastically. Example traces for phosphorylated TCR and calcium concentrations are shown in Figure 2. Interestingly, the network generation feature of BioNetGen2 presented 16 more dephosphorylation reactions than the number listed in the supplemental materials of [6], though it is unknown which dephosphorylation reactions were not counted in their model.

We also ran into a problem running the calculations using the ODE solver instead of the stochastic solver, which caused the phosphorylated TCR levels to level off very quickly around a value of 15 molecules. We believe that this is an artefact of a zeroth-order dephosphorylation rate acting on a phosphorylated TCR level that is less than 1 molecule at the early timesteps, but have not yet attempted to conclusively prove so.

Violation of Detailed Balance

One concern that was brought up is that detailed balance may be violated. For any chemical system where there are two different paths from one state to another, such as in Figure 3(a), the product of the equilibrium constants

Figure 3: Complex formation in the transphosphorylation model, which shows detailed balance. (b) One route of complex formation in the Li et al. model, which does not.
Figure 4: (a) The phosphorylated TCR levels and (b) the calcium trace for the Li et al. model with 30 agonist, an endogenous dissociation rate of 170 s$^{-1}$, and a first-order dephosphorylation rate of 0.994 s$^{-1}$. The trial fails to reach T cell activation.

in one direction must equal the product in the other direction if the system is not energy-driven. In other words, for Figure 3(a),

$$\frac{k_1 f}{k_1 r} \left( \frac{k_2 f}{k_2 r} \right) = \left( \frac{k_3 f}{k_3 r} \right) \left( \frac{k_4 f}{k_4 r} \right)$$

(4)

if no energy is being expended in the system.

The Li et al. model is a developed version of the mathematical framework for TCR activation by kinetic proofreading first laid down by T.W. McKeithan, where after pMHC-TCR binding, the complex goes through a series of activation steps eventually culminating in a signaling complex, but always in danger of pMHC-TCR dissociation returning the complex to a completely inactive state [10]. The Li et al. model attempts to accomplish this kinetic proofreading through complex formation steps.

However, as shown in Figure 3(b), the model violates detailed balance. There are two ways to completely dissociate pMHC-TCR-CD4 to pMHC. One way involves piecemeal dissociation where CD4 dissociates first, and then TCR; the other involves a shotgun dissociation resulting from the automatic destruction of the pMHC-CD4 bond when the pMHC-TCR bond is broken. By the principle of detailed balance, if the model is not energy-driven,

$$\left( \frac{k_1 f}{k_2 r} \right) \left( \frac{k_3 f}{k_5 r} \right) = 0 \quad \Rightarrow \quad k_1 = 0 \text{ or } k_3 = 0$$

(5)

which says either pMHC-TCR bonding or pMHC-CD4 bonding does not happen, contradicting the model. Thus, model must be energy-driven—like the McKeithan model—but Li et al. do not give a mechanistic basis for any energy-driven reactions in the model.

**Zeroth-order Dephosphorylation Rate**

Our second concern regards the low dephosphorylation rate used in this model. In the supplemental materials, it is set to $2.2 \times 10^{-9}$ Ms$^{-1}$, which, for a specified volume of $7.5 \times 10^{-16}$ L, corresponds to 0.994 molecules·s$^{-1}$. The justification for a low, zeroth-order dephosphorylation rate is unclear, as it allows the phosphorylation of TCRs to run away, unchecked, once levels are substantially greater than 1 molecule. The use of a low zeroth-order dephosphorylation rate implies that phosphatases are excluded from the area and otherwise saturated, but since there
has been no evidence that phosphatase exclud-

tion occurs in the cell contact region before the

immunological synapse forms, that does not

seem to be a solution to this problem.

The model was tested with a first-order de-

phosphorylation rate of 0.994 s$^{-1}$, and failed
to phosphorylate enough TCRs to generate a
calcium signal above the cutoff concentration
of 0.001 M, as shown in Figure 4. Since the
T cell activation system has shown many simi-
larities to the Fc$\epsilon$RI system, it is reasonable to
expect the actual dephosphorylation rate to be
near that in the model for Fc$\epsilon$RI described by
Faeder et al. 20 s$^{-1}$ [11]. But, if the T cell
model fails to activate at a first-order dephos-
phorylation rate of 0.994 s$^{-1}$, it is highly un-
likely that it will activate at an approximately
20-fold higher rate.

### Linear Phosphorylation Signal

Implementing the model without dephospho-
ylation yielded a linear growth curve of phos-
phorylated TCR, as shown in Figure 5(a). This
led to the conjecture that it was possible to split
the equation for the production rate of phos-
phorylated TCR into an entirely concentra-
tion-independent zeroth-order rate law:

$$ \frac{dT_{CRP}}{dt} = k_{no\, dephos} - k_{-P} \tag{6} $$

where $T_{CRP}$ is the number of phosphor-
ylated TCR, $k_{no\, dephos}$ is the zeroth-order phos-
phorylated TCR production rate without de-
phosphorylation, and $k_{-P}$ is the zeroth-order
dephosphorylation rate. Running the model
without phosphorylation, $k_{no\, dephos}$ was ob-
tained by taking an approximate slope of the
stochastically-generated phosphorylated TCR
curve. $k_{-P}$ was then subtracted from this
amount, and a stochastically-generated line, as
in Figure 5(b), with the resulting slope was fed
into the calcium program. This generated the
calcium trace in Figure 5(c), which behaves
very similarly to the calcium outputs of the full
model.

This process was done for agonist pMHC
amounts from 0 to 40, in increments of 10,
and for endogenous pMHC dissociation rates
from 0 to 250 s$^{-1}$, in increments of 50 s$^{-1}$.
This produced the same results as found by Li
et al. with the graph looking similar to Fig-
ure S3(a) in the supplemental materials of [6].
Since the model governing the level of phos-
phorylated TCR can be simulated by a linear
curve, it seems that the crux of the model rests
on the two calcium equations, and not on the
kinetic proofreading model.
Krogsgaard et al. (2005)

A recent paper by Krogsgaard et al. raises doubts about the model’s validity [12]. Experimentally, either agonist or endogenous pMHCs were mutated to be unable to bind to CD4, and then soluble agonist/endogenous pMHC homodimers and heterodimers were created by artificial crosslinks. When only endogenous pMHC were mutated and unable to bind to CD4, no effect was observed; when only agonist pMHC were mutated, activation was impaired. Krogsgaard et al. propose that only agonist pMHC-CD4 binding occurs, and that agonist pMHC-TCR binding stimulates TCR-CD4 interactions to pull TCR to regions where endogenous pMHC are waiting.

While there is ample evidence that TCR-CD4 interactions occur [13], the Krogsgaard et al. model implies that such interactions are stronger and more important than pMHC-TCR and pMHC-CD4 interactions. Because the relative importance of TCR-CD4 reactions has not been assessed, it is impossible to evaluate the accuracy of this model and its contradiction to the Li et al. model. Also, it is possible that phosphorylated CD3ε-Lck binding may be responsible for these results [14]. When endogenous pMHC is mutated, TCR bound to it can still be phosphorylated by the CD3ε-Lck complex attached to its neighboring agonist pMHC; when agonist pMHC is mutated, the CD3ε bound to its neighboring endogenous pMHC cannot get phosphorylated because of the short endogenous half-life.

Transphosphorylation Model

Goals of the Model

Kinetic proofreading has worked very well in providing a mathematical framework for peptide discrimination solely on the basis of interaction half-life, and a model for the early steps of T cell activation will probably incorporate it in some sense. Kinetic proofreading steps should be energy-driven to avoid complications with detailed balance, as McKeithan suggested [10]. Some serial triggering should also be incorporated so that amplification can occur. The system should not be able to activate with absence or very low levels of agonist pMHC, nor should it be able to activate when endogenous pMHC is not present except at high levels of agonist. Finally, the system should show cooperativity, where TCR phosphorylated while bound to endogenous pMHC is far higher when agonist pMHC is present than when it is not.

Theory

While Li et al. propose a kinetic proofreading model based on complex formation, we propose a kinetic proofreading model based on transphosphorylation of TCR/CD3 subunits. The α and β subunits of TCR bind to the peptide-MHC complex, while the CD3γε and δε heterodimers and the ζζ homodimer participate in further steps [15]. Each γ, δ, and ε subunit contains one ITAM, and each ζ chain contains three ITAMs [16]. However, for simplicity, the current form of this model only includes a single ε subunit and a single ζ chain, each having only one ITAM, on every TCR/CD3 complex.

Since the model does not rely on complex formation for its kinetic proofreading, the constituent parts—pMHC, TCR/CD3, and CD4 dimer-Lck complex—can associate in any order, and dissociation of any bond does not destroy the complex, making the reaction paths akin to Figure 3(a), where detailed balance is obeyed. Also, Lck can only bind to either CD4 [4] or singly-phosphorylated CD3ε ITAM (pCD3ε) [14], but not both at the same time. ZAP-70 can only bind to doubly-
Figure 6: (a) The main process of the transphosphorylation model begins with the formation of the pMHC-TCR/CD3-CD4 dimer-Lck complex. (b) CD4-associated Lck then phosphorylates CD3ε and ζ and ZAP-70 binds to the phosphorylated ζ chain. (c) Lck binds to phosphorylated CD3ε and gets phosphorylated by the CD4-associated Lck, while a second pMHC-TCR/CD3 pair binds. (d) The second TCR/CD3 is quickly brought to full activation by the transphosphorylation of the CD3ε-associated Lck.

phosphorylated CD3ζ ITAM (pCD3ζ). We assume that TCR downregulation, which has been used to measure TCR activation [7], is linear to ZAP-70 binding to pCD3ζ, and thus use ZAP-70 binding to assess signal strength. While ZAP-70 also gets phosphorylated by Lck [8], this was not included in the model because of over-complexity.

When a pMHC-TCR/CD3-CD4 dimer-Lck complex forms, the CD4-associated Lck is able to singly-phosphorylate the CD3ε ITAM and doubly-phosphorylate the ζ ITAM at a low phosphorylation rate. Once the ζ chain is doubly-phosphorylated, free ZAP-70 can bind to it, its SH2 domain protecting the ITAM from dephosphorylation [17]. Free Lck acts similarly to pCD3ε. However, the pCD3ε-associated Lck is then phosphorylated by the CD4-associated Lck, enabling it to phosphorylate all other ITAMs and pCD3ε-associated Lcks in the complex at a higher rate. However, the system is constantly in danger of pMHC-TCR dissociation, halting the process. A high dephosphorylation rate for uncomplexed TCR/CD3 subunits and Lck returns the molecule to its totally inactivated state once Lck and ZAP-70 unbind and release their protection from dephosphorylation, thus generating kinetic proofreading.

When another pMHC-TCR pair joins the existing complex, it can get quickly transphosphorylated by the pCD3ε-associated phosphorylated Lck (pLck). If one pMHC in the complex is agonist and the other is endogenous, the transphosphorylation results in signal amplification, because the agonist pMHC-bound CD3 will quickly gain pCD3ε-associated pLck, which will quickly activate any TCR subunit or Lck associated with endogenous pMHC in the complex, resulting in signal amplification. The main transphosphorylation process is shown in Figure 6, and the model is shown encoded for BioNetGen2 in Appendix D.
Figure 7: CD3ζ-bound ZAP-70 traces from the transphosphorylation model for an endogenous dissociation rate of 170 s\(^{-1}\) and (a) 0 agonist pMHCs, and (b) 24 agonist pMHCs.

Parameters

The molecules in this model interact in a cylindrical cell-contact region with a radius of 5 \(\mu m\) and a height of 1.2 nm, to make a reaction volume of 9.425 \(\times 10^{-17}\) L. Initial molecular amounts of endogenous pMHC, TCR/CD3, and CD4 dimer-Lck were taken from Li et al. and converted to have the same concentration in the smaller reaction volume. We assumed that agonist pMHC has a long enough half-life such that when one enters the contact region, it stays there due to the constant reactions with TCR/CD3 inside the reaction volume. While a compartmentalized model would be a preferred implementation of this, BioNetGen2 cannot currently run compartmentalized models, so it was assumed that all of the agonist pMHC on the APC is and stays contained in the reaction volume.

Some parameters followed choices of dissociation equilibrium constant (\(K_D\)). For example, while the agonist pMHC-TCR \(K_D\) was set to 1 \(\mu M\), the endogenous pMHC-TCR \(K_D\) was set to 125 \(\mu M\). pMHC-CD4 rate constants were governed by a \(K_D\) of 3.2 \(\mu M\). Also, since the T cell activation model is similar to the FcεRI model, where Lck is analogous to Lyn and ZAP-70 is analogous to Syk, the decrease in the \(\zeta\) phosphorylation rate compared that of \(\epsilon\) was assumed to be 30-fold in order to bring it in line with the 30-fold difference between the phosphorylation rates of \(\gamma\) and \(\beta\) subunits of FcεRI in Faeder et al. [11].

Except for values that had been already decided, as enumerated above, most of the values were chosen randomly because there was not enough time to do a full sensitivity analysis of the model and try many combinations of parameters. However, a few key characteristics about the parameters were evident. pMHC-CD4 binding, while at a set equilibrium constant, should have association and dissociation rates fast enough such that serial triggering can occur, but slow enough such that amplification by transphosphorylation can complete. While the \(\epsilon\) and \(\zeta\) subunit rates of phosphorylation by CD4-associated Lck must be low enough to produce kinetic proofreading, the rates of phosphorylation by pCD3\(\epsilon\)-associated pLck must be high enough for transphosphorylation to occur rapidly. Also, association and dissociation rates for Lck and ZAP-70 must be placed so that they can protect phosphorylation of subunits inside signaling complexes, and dephos-
Figure 8: When endogenous pMHC are taken away, not even 24 pMHC can activate the cell, showing that amplification by transphosphorylation is both present and necessary in the transphosphorylation model.

phosphorylation rates outside of the complex must be higher than those inside the complex so that TCR/CD3 can be reset to a totally non-signaling state once they leave the complex—otherwise, a TCR/CD3 could activate by jumping between signaling complexes and inching its way toward activation at each stop, defeating kinetic proofreading.

Results

We are able to get discrimination between numbers of agonist with this model. Shown in Figure 7 are runs for 0 and 24 agonist pMHCs. Because levels go so high, we know that there must be evidence of serial triggering of TCRs bound to endogenous pMHC, since there are at most 24 agonist pMHC-TCR-CD4 signaling complexes. Also, amplification by transphosphorylation is in effect, because when endogenous pMHC is taken away, there is not enough serial triggering by 24 agonist pMHC to bring the cell to activation, as shown in Figure 8.

Concerns and future research

One concern that we have for the model is the high dephosphorylation rate for ε and ζ sub-units outside of the complex. As explained before, a high dephosphorylation rate for units outside the complex is essential; however, we are currently attempting to decrease the $10^{10}$ s$^{-1}$ dephosphorylation rate because such a high rate is not present in any known physical system. ZAP-70 has also been found to bind to doubly-phosphorylated CD3ε [14]. Implementing this possibility generates Figure 9, where discrimination is still present but the behavior of the ζ-bound ZAP-70 levels seems to change slightly. Further study will examine the nature of this perturbation.

Because it is not entirely known how CD3ζ-bound ZAP-70 levels correlate with TCR downregulation and T cell activation, we cannot say with any certainty whether these levels show evidence of T cell activation or not. However, we can show that the transphosphorylation model generates discrimination between agonist and endogenous peptide and uses amplification by transphosphorylation to achieve a much higher ZAP-70 signal with agonist pMHC than without. As the field of immunology becomes more quantitative, we hope that more evidence guide us toward a completely successful, discriminating, and mechanistic model for the early steps of T cell activation.
Figure 9: Implementing double phosphorylation of CD3ε and ZAP-70 binding to CD3ε still shows discrimination between (a) 0 agonist pMHC and (b) 24 agonist pMHC, but the behavior of the CD3ζ-bound ZAP-70 levels seems to differ from the regular model in the case of 24 agonist. This implementation used the same phosphorylation and dephosphorylation rates for both tyrosines on ε, and the same association and dissociation rates for ZAP-70-ε binding and ZAP-70-ζ binding.

Acknowledgments

M.J.S. would like to thank Dr. Bill Hlavacek for all of his help, guidance, and introduction to the worlds of immunology and biological systems modeling, Dr. Jim Faeder for the BioNetGen2 program and help understanding its code and input syntax, Dr. Byron Goldstein for insights and discussions about TCR/CD3 structure, and Dr. Norm Magee for being there and caring for our program—whether we need paper towels or a volleyball court. This Research Experience for Undergraduates program is funded by the National Science Foundation.

A Reactions of the Li et al. model

The following page is a numbered version of Figure S2 from the Li et al. supplemental materials [6]. The numbers correspond to the reaction rules in the BioNetGen2 implementation of the model, seen in Appendix B. Some reactions listed in the figure are split into two reaction rules for BioNetGen2 to accommodate for the different dissociation rates of agonist and endogenous pMHC-TCR interactions.
Initial TCR-pMHC binding (4)
\[ M \rightarrow \frac{k_1}{M} + \frac{k_2}{T} \]

Tetrameric binding events (4)
\[ C \rightarrow \frac{k_3}{M} + \frac{k_5}{T} \]
\[ C \rightarrow \frac{k_4}{M} + \frac{k_5}{L-T} \]
\[ C \rightarrow \frac{k_6}{M} + \frac{k_8}{L-T} \]

Pentameric binding events (8)
\[ M \rightarrow \frac{k_3}{C-M} + \frac{k_5}{L-T} \]
\[ M \rightarrow \frac{k_3}{C-M} + \frac{k_5}{L-T} \]
\[ M \rightarrow \frac{k_7}{C-M} + \frac{k_8}{L-T} \]

Hexameric binding events (16)
\[ M \rightarrow \frac{k_1}{C-M} + \frac{k_2^*}{T} \]
\[ M \rightarrow \frac{k_3}{C-M} + \frac{k_5}{L-T} \]
\[ M \rightarrow \frac{k_3}{C-M} + \frac{k_5}{L-T} \]
\[ M \rightarrow \frac{k_3}{C-M} + \frac{k_5}{L-T} \]
\[ M \rightarrow \frac{k_4}{C-M} + \frac{k_5}{L-T} \]
\[ M \rightarrow \frac{k_4}{C-M} + \frac{k_5}{L-T} \]
\[ M \rightarrow \frac{k_7}{C-M} + \frac{k_8}{L-T} \]

Tetramer dissociation due to MT separation (4)
\[ C \rightarrow \frac{k_2^*}{M} + \frac{k_2^*}{C-M} \]
\[ C \rightarrow \frac{k_2^*}{M} + \frac{k_2^*}{C-M} \]
\[ C \rightarrow \frac{k_2^*}{M} + \frac{k_2^*}{C-M} \]

Pentamer dissociation due to MT separation (8)
\[ M \rightarrow \frac{k_2^*}{M-C-M} + \frac{k_2^*}{M-C-M} \]
\[ M \rightarrow \frac{k_2^*}{M-C-M} + \frac{k_2^*}{M-C-M} \]
\[ M \rightarrow \frac{k_2^*}{M-C-M} + \frac{k_2^*}{M-C-M} \]

Hexamer dissociation due to MT separation (16)
\[ M \rightarrow \frac{k_2^*}{M-C-M} + \frac{k_2^*}{M-C-M} \]
\[ M \rightarrow \frac{k_2^*}{M-C-M} + \frac{k_2^*}{M-C-M} \]
\[ M \rightarrow \frac{k_2^*}{M-C-M} + \frac{k_2^*}{M-C-M} \]
B Sample .bngl file for the Li et al. model

This is a sample input file for BioNetGen2 for a trial run with 30 agonist pMHC and an endogenous dissociation rate of 170 s$^{-1}$, calculated stochastically. ODE simulations are also possible. All parameter values are given as molecular concentrations in the file, so second-order and zeroth-order rates had to be converted to molecules$^{-1}$·s$^{-1}$ and molecules·s$^{-1}$, respectively, for a volume of 7.5×10$^{-16}$ L. First-order rates needed no conversion, and are in units of s$^{-1}$. Original molar rates are stated in the comments for second-order and zeroth-order reactions. In the Li et al. supplemental materials, all reaction rates are taken from the Li et al. supplementary notes, and the reaction rules correspond the numbers in Appendix A [6].

\[ \text{This is a rendering of a model of the early signal transduction pathway for T cell activation proposed by Qi-Jing Li et al. in Nature Immunology (2004) for the BioNetGen2 program.} \]

\[ \text{Rendering written by Michael Saelim, Michigan State University, working at the Los Alamos Summer School program under Dr. Bill Hlavacek, summer 2006.} \]

\[ \text{begin parameters} \]
\[ 1 \text{ k1 1.993e-6 #pMHC-TCR association rate (900 M-1 s-1)} \]
\[ 2 \text{ k2a 0.054 #agonist pMHC-TCR dissociation rate (s-1)} \]
\[ 3 \text{ k2e 170 #endogenous pMHC-TCR dissociation rate (s-1)} \]
\[ 4 \text{ k3 1.993e-6 #intracomplex pMHC-CD4 association rate (9000 M-1 s-1)} \]
\[ 5 \text{ k4 700 #intracomplex pMHC-CD4 association rate (s-1)} \]
\[ 6 \text{ k5 0.020 #pMHC-CD4 dissociation rate (s-1)} \]
\[ 7 \text{ k6 9.963e-6 #intercomplex TCR-Lck association rate (4500 M-1 s-1)} \]
\[ 8 \text{ k7 350 #intracomplex TCR-Lck association rate (s-1)} \]
\[ 9 \text{ k8 0.020 #TCR-Lck dissociation rate (s-1)} \]
\[ 10 \text{ k9 4 #rate of TCR phosphorylation by Lck (s-1)} \]
\[ 11 \text{ k10a 0.994 #rate of TCR dephosphorylation (2.2e-9 M s-1)} \]
\[ 12 \text{ k10b 1e-5 #extremely low rate for the Sat approximation} \]
\[ 13 \text{ Ma0 30 #number of agonist pMHC on the APC} \]
\[ 14 \text{ Me0 3e4 #number of endogenous pMHC on the APC} \]
\[ 15 \text{ Tu0 1e5 #initial number of unphosphorylated TCR} \]
\[ 16 \text{ Tp0 0 #initial number of phosphorylated TCR} \]
\[ 17 \text{ Cl0 1e4 #number of CD4 dimer-Lck complexes} \]
\[ \text{end parameters} \]

\[ \text{begin molecule types} \]
\[ 1 \text{ MHC(p^ag,T,C) #p=peptide type, T=TCR binding site, C=CD4 binding site} \]
\[ 2 \text{ TCR(M,L,ITAM^U) #M=pMHC binding site, L=Lck binding site} \]
\[ 3 \text{ CD4(L,M,M) #L=Lck binding site, M=pMHC binding site} \]
\[ 4 \text{ LCK(C,T,T) #C=CD4 binding site, T=TCR binding site} \]
\[ \text{end molecule types} \]

\[ \text{begin species} \]
\[ 1 \text{ MHC(p^ag,T,C) Ma0 #ag=agonist/antigen} \]
\[ 2 \text{ MHC(p^en,T,C) Me0 #en=endogenous} \]
\[ 3 \text{ TCR(M,L,ITAM^U) Tu0 #unphosphorylated TCR} \]
\[ 4 \text{ TCR(M,L,ITAM^P) Tp0 #phosphorylated TCR} \]
\[ 5 \text{ CD4(L1,L1,M,M).LCK(C11,T,T) C10 #CD4 dimer-Lck complex} \]
\[ \text{end species} \]

\[ \text{begin reaction rules} \]
\[ \text{#Initial TCR-pMHC binding} \]
\[ 1 \text{ MHC(p^ag,I,T,C) + TCR(M,L) <-> \} \]
\[ 2 \text{ MHC(p^ag,T11,L,C).TCR(M11,L) k1, k2a} \]
\[ 3 \text{ MHC(p^en,I,T,C) + TCR(M,L) <-> \} \]

92
MHC(p\textsuperscript{en}, T!1, C) . TCR(M!1, L) \rightleftharpoons k_1, k_2e

## Tetrameric binding events
3 CD4(M, M) . LCK(T, T) + MHC(T!1, C) . TCR(M!1, L) \rightleftharpoons \n4 CD4(M, M!2) . LCK(T, T) . MHC(T!1, C). TCR(M!1, L) \rightarrow k_3, k_5
5 CD4(M, M) . LCK(T, T!1) . MHC(T!2, C) . TCR(M1!1, L!2) \rightarrow k_6, k_8
6 CD4(M, M!1) . LCK(T, T!3) . MHC(T1!2, C!1) . TCR(M!2, L!1) \rightarrow k_7, k_8

## Pentameric binding events
7 MHC(T, C) + CD4(M, M!1) . LCK(T, T). MHC(T!2, C!1) . TCR(M!2, L) \rightarrow k_3, k_5
8 MHC(T, C!4) . CD4(M, M!1) . LCK(T, T!3) . MHC(T!4, C!2) . TCR(M!4, L!3) \rightarrow k_3, k_5
9 MHC(T, C!1) . CD4(M, M!2) . LCK(T, T). MHC(T!3, C!1) . TCR(M!3, L) \rightarrow k_7, k_8
10 MHC(T, C!4) . CD4(M, M!2) . LCK(T, T!3) . MHC(T!4, C!2) . TCR(M!4, L!3) \rightarrow k_4, k_5

## Hexameric binding events
11 MHC(p\textsuperscript{ag}, T, C!1) . CD4(M, M!2) . LCK(T, T) + TCR(M, L) \rightarrow MHC(p\textsuperscript{ag}, T!4, C!1) . TCR(M!4, L) . CD4(M!, M!2) . LCK(T, T) . MHC(T!3, C!2) . TCR(M!3, L) \rightarrow k_1, k_2a
12 MHC(p\textsuperscript{en}, T, C!1) . CD4(M, M!2) . LCK(T, T) + TCR(M, L) \rightarrow MHC(p\textsuperscript{en}, T!4, C!1) . TCR(M!4, L) . CD4(M!, M!2) . LCK(T, T) . MHC(T!3, C!2) . TCR(M!3, L) \rightarrow k_1, k_2e
13 MHC(T, C!2) . CD4(M!2, M!3) . LCK(T!3, T!4) . MHC(T!5, C!3) . TCR(M!5, L!4) \rightarrow k_7, k_8
14 MHC(T, C!2) . CD4(M!2, M!3) . LCK(T!3, T!4) . MHC(T!5, C!3) . TCR(M!5, L!4) \rightarrow k_7, k_8

## Tetramer dissociation due to MT separation
24 CD4(M, M) . LCK(T, T) . MHC(p\textsuperscript{ag}, T!1, C!1) . TCR(M!1, L) \rightarrow \n25 CD4(M, M) . LCK(T, T) . MHC(p\textsuperscript{en}, T!1, C!1) . TCR(M!1, L) \rightarrow \n26 CD4(M, M) . LCK(T, T!1) . MHC(p\textsuperscript{ag}, T!2, C!1) . TCR(M!2, L!1) \rightarrow \n27 CD4(M, M) . LCK(T, T!1) . MHC(p\textsuperscript{en}, T!2, C!1) . TCR(M!2, L!1) \rightarrow \n
## Pentamer dissociation due to MT separation
30 MHC(T, C!1) . CD4(M, M!2) . LCK(T, T) . MHC(p\textsuperscript{ag}, T!3, C!2) . TCR(M!3, L) \rightarrow \n31 MHC(T, C!1) . CD4(M, M!2) . LCK(T, T) . MHC(p\textsuperscript{en}, T!3, C!2) . TCR(M!3, L) \rightarrow \n32 MHC(T, C!1) . CD4(M, M!2) . LCK(T, T!3) . MHC(T!4, C!2) . TCR(M!4, L!3) \rightarrow \n
MHC(T,C) + CD4(M,M).LCK(T,T) + MHC(p^ag,T,C) + TCR(M,L) k2a
MHC(T,C) + CD4(M,M).LCK(T,T) + MHC(p^en,T,C) + TCR(M,L) k2e

## Hexamer dissociation due to MT separation

34 MHC(T,C!)1.CD4(M!1,M!2).LCK(T,T!)3.MHC(p^en,T,C!4).TCR(M!5,L!4) -> MHC(T,C) + CD4(M,M).LCK(T,T) + MHC(p^en,T,C) + TCR(M,L) k2e
35 MHC(T,C!1).TCR(M!1,L!3).CD4(M!2,M!3).LCK(T,T!)4.MHC(p^ag,T,C!5).TCR(M!6,L!5) -> MHC(T,C) + CD4(M,M).LCK(T,T) + MHC(p^en,T,C) + TCR(M,L) k2e
36 MHC(T,C!1).TCR(M!1,L!3).CD4(M!2,M!3).LCK(T,T!)4.MHC(p^ag,T,C!5).TCR(M!6,L!5) -> MHC(T,C) + CD4(M,M).LCK(T,T) + MHC(p^en,T,C) + TCR(M,L) k2e
37 MHC(T,C!1).TCR(M!1,L!3).CD4(M!2,M!3).LCK(T,T!)4.MHC(p^ag,T,C!5).TCR(M!6,L!5) -> MHC(T,C) + CD4(M,M).LCK(T,T) + MHC(p^en,T,C) + TCR(M,L) k2e
38 MHC(p^ag,T,C!2).TCR(M!1,L!3).CD4(M!2,M!3).LCK(T,T!)4.MHC(p^en,T,C!5).TCR(M!6,L!5) -> MHC(T,C) + CD4(M,M).LCK(T,T) + MHC(p^en,T,C) + TCR(M,L) k2e
39 MHC(p^en,T,C!2).TCR(M!1,L!3).CD4(M!2,M!3).LCK(T,T!)4.MHC(p^ag,T,C!5).TCR(M!6,L!5) -> MHC(T,C) + CD4(M,M).LCK(T,T) + MHC(p^en,T,C) + TCR(M,L) k2e
40 MHC(p^ag,T,C!2).TCR(M!1,L!3).CD4(M!2,M!3).LCK(T,T!)4.MHC(p^en,T,C!5).TCR(M!6,L!5) -> MHC(T,C) + CD4(M,M).LCK(T,T) + MHC(p^en,T,C) + TCR(M,L) k2e

## If unphosphorylated TCR is bound to pMHC and LCK, phosphorylate TCR

48 TCR(M!,T!,ITAM^U) -> TCR(M!,T!,ITAM^P) k9

## If TCR is phosphorylated, dephosphorylate TCR

49 TCR(ITAM^P) -> TCR(ITAM^U) Sat(k10a,k10b) ## Sat approximation of zeroth-order end reaction rules

begin observables
Molecules TCRp TCR(ITAM^P) # Number of phosphorylated ITAMs is observed
end observables

generate_network({overwrite=>1}); # Network generation step
simulate_ssa({t_end=>600,n_steps=>6000}); # 600 second stochastic simulation with 0.1s timesteps
C  Source code for the calcium model implementation

This Java code numerically integrates the calcium model equations using an Euler method. Since there is no feedback from the calcium model to the T cell receptor model, this can be kept separate from the BioNetGen2 evaluation of phosphorylated TCR. Also, since \( Z_{ER} \) values in equation 1 do not depend on equation 2, equation 1 can be evaluated before 2. Methods \( \text{dZ/dt} \) and \( \text{dCa/dt} \) are implementations of the two equations, and the main method implements the Euler method integration.

```java
import java.lang.*;
import java.io.*;
import java.util.*;

class Calcium {
    private static final double Z0 = 1.0;
    private static final double Ca0 = 0.0;
    private static final double kER = 0.02;
    private static final double P0 = 5.0;
    private static final double mu = 10.0;
    private static final double k1 = 0.02;
    private static final double kCa = 0.0006;
    private static final double kP = 100.0;

    public static void main(String[] args) throws FileNotFoundException, IOException {
        //Reading the input file
        Vector<Double> TCRp = new Vector<Double>(7000,1000);
        double Dt; //Time step
        char[] Dtbuff = new char[15]; //Time step buffer
        BufferedReader infile = 
        new BufferedReader(new InputStreamReader(new FileInputStream(args[0])));
        infile.readLine(); //Get rid of the first, useless line
        TCRp.addElement(Double.parseDouble(infile.readLine())); //Read the first TCRp entry
        infile.read(Dtbuff,0,15); //Read the time step length
        infile.skip(3); //Skip over the spaces
        TCRp.addElement(Double.parseDouble(infile.readLine())); //Read the second TCRp entry
        Dt = Double.parseDouble(new String(Dtbuff));

        while(infile.ready()) {
            infile.skip(17); //Skip over the time entry and blank space
            TCRp.addElement(Double.parseDouble(infile.readLine())); //Read the TCRp entry
        }
        infile.close();
    }
}
```

This program takes a .gdat file from BioNetGen2 containing a time series of phosphorylated TCR amounts (in molecules) as command-line input and outputs the calculated calcium signal as a .cal file using the calcium model described in the methods section of Qi-Jing Li et al., 2004.

//The model is calculated numerically with Euler’s method.

//Written by Michael Saelim, Michigan State University, working at Los Alamos National Laboratory under Dr. Bill Hlavacek, Summer 2006.
//Computing Z values (ER calcium concentration)
Vector<Double> Z = new Vector<Double>(TCRp.size(),1000);
double Z0;

for(int i=0; i<TCRp.size()-1; i++) { //Euler’s method
    Zi = Z.get(i) + Dt*dZdt(Z.get(i), TCRp.get(i));
    if(Zi < 0) { //Cannot have a negative ER calcium concentration
        Zi = 0;
    }
    Z.addElement(Zi);
}

//Computing Ca values (Cytosolic calcium concentration)
Vector<Double> Ca = new Vector<Double>(TCRp.size(),1000);
double Ca0;

for(int i=0; i<TCRp.size()-1; i++) { //Euler’s method
    Cai = Ca.get(i) + Dt*dCadt(Ca.get(i), TCRp.get(i), Z.get(i));
    if(Cai < 0) { //Cannot have a negative cytosolic calcium concentration
        Cai = 0;
    }
    Ca.addElement(Cai);
}

//Finding the name of the .cal file
String calname = new String();
calname = args[0].substring(0,args[0].length()-5) + " .cal";

//Outputting the calcium data
PrintWriter outfile = new PrintWriter(new BufferedWriter(new FileWriter(calname)));

outfile.println("# time calcium");
for(int i=0; i<Ca.size(); i++) {
    outfile.println(" " + i*Dt + " " + Ca.get(i));
}
outfile.close();
return;
}

public static double dZdt(double Z, double P) {
    //dZ/dt = -kER * H(P-P0) * Z
    //where Z = calcium stored in ER and P = # of TCRp

    //Implement the Heaviside function as an if-else block:
    if(P-P0 < 0) {
        return 0;
    }
    else if(P-P0 == 0) {
        return -kER * Z / 2; //H(0) = 1/2
    }
    else {
        return -kER * Z;
    }
}

public static double dCadt(double Ca, double P, double Z) {
\[ \frac{dCa}{dt} = -\frac{dZ}{dt} - \mu Ca + k1 \cdot \frac{[Ca/(Ca+kCa)] \cdot [P/(P+kP)]}{Ca+kCa} \]

where \( Ca \) = cytosolic calcium, \( Z \) = calcium stored in ER
and \( P \) = \# of TCRp

//Implement the Heaviside function as an if-else block:

double firstterm;
if(P-P0 < 0) {
    firstterm = 0.0;
}
else if(P-P0 == 0) {
    firstterm = kER * Z / 2; //H(0) = 1/2
}
else {
    firstterm = kER * Z;
}

return firstterm - \mu Ca + k1 \cdot \frac{[Ca/(Ca+kCa)] \cdot [P/(P+kP)]}{Ca+kCa} \]
D Sample .bngl file for the transphosphorylation model

This BioNetGen2 input file implements the transphosphorylation model for a system with 24 agonist pMHC and an endogenous pMHC-TCR dissociation rate of 170 s$^{-1}$, calculated with ODEs. As with the implementation of the Li et al. model, all parameter values are given as molecular concentrations in the file, so second-order and zeroth-order rates had to be converted to molecules$^{-1}\cdot$s$^{-1}$ and molecules$\cdot$s$^{-1}$, respectively, for a volume of $9.425 \times 10^{-17}$ L. First-order rates needed no conversion, and are in units of s$^{-1}$. Original molar rates are stated in the comments for second-order and zeroth-order reactions. A table of parameters in molar units with explanations on their choices is included in Appendix E.

---

begin parameters
1 Kaf 3.524e-4 #agonist pMHC-TCR association rate (2e4 M$^{-1}$ s$^{-1}$)
2 Kar 0.02 #agonist pMHC-TCR dissociation rate (s$^{-1}$)
3 Kef 2.396e-2 #endogenous pMHC-TCR association rate (1.36 e6 M$^{-1}$ s$^{-1}$)
4 Ker 170 #endogenous pMHC-TCR dissociation rate (s$^{-1}$)
5 Kmcf 1.762e-1 #pMHC-CD4 association rate (1e7 M$^{-1}$ s$^{-1}$)
6 Kmcr 32 #pMHC-CD4 dissociation rate (s$^{-1}$)
7 Kep_lo 0.1 #rate of epsilon phosphorylation by CD4-associated Lck (s$^{-1}$)
8 Kep_hi 300 #rate of epsilon phosphorylation by pCD3e-associated pLck (s$^{-1}$)
9 Ked 0.5 #rate of epsilon dephosphorylation for CD3s in complex (s$^{-1}$)
10 Kxed 1e10 #rate of epsilon dephosphorylation for CD3s out of complex (s$^{-1}$)
11 Klef 0.1 #Lck-pCD3e association rate (5.68e6 M$^{-1}$ s$^{-1}$)
12 Kler 0.5 #Lck-pCD3e dissociation rate (s$^{-1}$)
13 Klp_lo 3 #rate of Lck phosphorylation by CD4-associated Lck (s$^{-1}$)
14 Klp_hi 300 #rate of Lck phosphorylation by pCD3e-associated pLck (s$^{-1}$)
15 Kld 0.1 #rate of Lck dephosphorylation for CD3s in complex (s$^{-1}$)
16 Kxld 1e10 #rate of Lck dephosphorylation for CD3s out of complex (s$^{-1}$)
17 Kzp_lo 3.333e-3 #rate of zeta phosphorylation by CD4-associated Lck (s$^{-1}$)
18 Kzp_hi 10 #rate of zeta phosphorylation by pCD3e-associated pLck (s$^{-1}$)
19 Kzd 0.25 #rate of zeta dephosphorylation for CD3s in complex (s$^{-1}$)
20 Kxzd 1e9 #rate of zeta dephosphorylation for CD3s out of complex (s$^{-1}$)
21 Kzzf 1 #ZAP70-CD3z association rate (5.68e7 M$^{-1}$ s$^{-1}$)
22 Kzzr 0.1 #ZAP70-CD3z dissociation rate (s$^{-1}$)
23 Ma0 24 #number of agonist pMHC
24 Me0 3770 #number of endogenous pMHC
25 T0 1.257e4 #initial number of totally unphosphorylated TCR/CD3
26 C0 1.257e3 #number of CD4 dimer-Lck complexes
27 L0 1e4 #number of free Lck molecules
28 Z0 1e4 #number of ZAP70 molecules
end parameters

begin molecule types
1 MHC("p\textasciitilde ag\textasciitilde en,C")
2 TCR(ab,e⁻U⁺P,z⁻U⁺P) # epsilon and zeta can be unphosphorylated or phosphorylated
3 CD4(M,M,L) # Representative of a CD4 homodimer
4 LCK(UD,SH2,Y394⁻U⁺P) # Lck contains a Unique Domain, SH2 domain, and tyrosine 394
5 ZAP70(TSH2) # ZAP70 has Tandem SH2 domains

end molecule types

begin species
1 MHC(p⁻ag,C) Ma0
2 MHC(p⁻en,C) Me0
3 TCR(ab,e⁻U⁺z⁻U) T0
4 CD4(M,M,L!1).LCK(UD!1,SH2,Y394⁻U) C0 # Lck permanently associated
5 LCK(UD,SH2,Y394⁻U) L0 # Free Lck
6 ZAP70(TSH2) Z0

end species

begin reaction rules

## pMHC-TCR binding
1 MHC(p⁻ag) + TCR(ab) <-> MHC(p⁻ag!1).TCR(ab!1) Kaf,Kar
2 MHC(p⁻en) + TCR(ab) <-> MHC(p⁻en!1).TCR(ab!1) Kef,Ker

## pMHC-CD4 binding
3 MHC(C) + CD4(M) <-> MHC(C!1).CD4(M!1) Kmcf,Kmcr

## CD4-associated Lck transphosphorylation of CD3 epsilon
4 TCR(e⁻U⁺).LCK(UD↑+) --> TCR(e⁺⁺).LCK(UD↑+) Kep_lo

## CD3e-associated pLck transphosphorylation of CD3 epsilon
5 TCR(e⁻U⁺).LCK(SH2↑+,Y394⁻P) --> TCR(e⁺⁺).LCK(SH2↑+,Y394⁻P) Kep_hi

## Complexed unbound CD3 epsilon dephosphorylation
6 TCR(ab!+,e⁻P) --> TCR(ab!+,e⁻U) Ked

## Uncomplexed unbound CD3 epsilon dephosphorylation
7 TCR(ab,e⁻P) --> TCR(ab,e⁻U) Kxed

## Lck binding to phosphorylated CD3 epsilon
8 TCR(e⁻P⁺) + LCK(UD,SH2) <-> TCR(e⁻P⁺!1).LCK(UD,SH2!1) Klef,Kler

## CD4-associated Lck transphosphorylation of Lck
9 LCK(SH2!+,Y394⁻U⁺).LCK(UD↑+) --> LCK(SH2!+,Y394⁻P⁺).LCK(UD↑+) Klp_lo

## CD3e-associated pLck transphosphorylation of Lck

## Complexed Lck dephosphorylation
11 TCR(ab!+,e⁻P⁺!1).LCK(SH2!1,Y394⁻P⁺) --> TCR(ab!+,e⁻P⁺!1).LCK(SH2!1,Y394⁻U⁺) Kld

## Uncomplexed Lck dephosphorylation
12 TCR(ab,e⁻P⁺!1).LCK(UD,SH2) --> TCR(ab,e⁻P⁺).LCK(UD,SH2)!1) Kxld

## CD4-associated Lck transphosphorylation of CD3 zeta
13 TCR(z⁻U⁺).LCK(UD↑+) --> TCR(z⁺⁺).LCK(UD↑+) Kzp_lo

## CD3e-associated pLck transphosphorylation of CD3 zeta
14 TCR(z⁻U⁺).LCK(SH2↑+,Y394⁻P⁺) --> TCR(z⁺⁺).LCK(SH2↑+,Y394⁻P⁺) Kzp_hi

## Complexed unbound CD3 zeta dephosphorylation
15 TCR(z⁺⁺).LCK(SH2!+,Y394⁻U⁺) --> TCR(z⁺⁺).LCK(SH2!+,Y394⁻U⁺) Kzp_lo

## Uncomplexed unbound CD3 zeta dephosphorylation
16 TCR(z⁺⁺).LCK(SH2!+,Y394⁻P⁺) --> TCR(z⁺⁺).LCK(SH2!+,Y394⁻P⁺) Kzp_hi

## ZAP70 binding to complexed phosphorylated CD3 zeta
17 TCR(z⁺⁺) + ZAP70(TSH2) <-> TCR(z⁺⁺!1).ZAP70(TSH2!1) Kzzf,Kzzr

end reaction rules

begin observables
Molecules Zap ZAP70(TSH2) # Observing total number of bound ZAP70

end observables

generate_network({overwrite=>1});
simulate_ode({t_end=>600,n_steps=>6000}); # ODE simulation to 600s, timestep of 0.1s
### E Parameters for the sample transphosphorylation model

These parameters are used in the current version of the transphosphorylation model. pMHC dissociation rates were determined in advance and speculated to be characteristic of agonist and endogenous off rates. Blank rationales indicate that the value was arbitrarily chosen because it made the model work.

Table 1: Model parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Rationale/Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_a$</td>
<td>$2 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$</td>
<td>Agonist $K_D = 1 \mu\text{M}, k_{off} = 0.02 \text{s}^{-1}$</td>
</tr>
<tr>
<td>$K_a$</td>
<td>0.02 s$^{-1}$</td>
<td>Preset (near to value used by Li et al. [6])</td>
</tr>
<tr>
<td>$K_e$</td>
<td>$1.36 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$</td>
<td>Endogenous $K_D = 125 \mu\text{M}, k_{off} = 170 \text{s}^{-1}$</td>
</tr>
<tr>
<td>$K_r$</td>
<td>170 s$^{-1}$</td>
<td>Preset (value used by Li et al. [6])</td>
</tr>
<tr>
<td>$K_{mcf}$</td>
<td>$10^7 \text{ M}^{-1}\text{s}^{-1}$</td>
<td>$K_D = 3.2\mu\text{M},$ Cammarota et al. [2]</td>
</tr>
<tr>
<td>$K_{mcr}$</td>
<td>32 s$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>$K_{ep,lo}$</td>
<td>0.1 s$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>$K_{ep,hi}$</td>
<td>300 s$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>$K_{ed}$</td>
<td>0.5 s$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>$K_{xed}$</td>
<td>$10^{10} \text{ s}^{-1}$</td>
<td>Fast dephosphorylation outside complex</td>
</tr>
<tr>
<td>$K_{le}$</td>
<td>$5.68 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$</td>
<td></td>
</tr>
<tr>
<td>$K_{ler}$</td>
<td>0.5 s$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>$K_{lp,lo}$</td>
<td>3 s$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>$K_{lp,hi}$</td>
<td>300 s$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>$K_{ld}$</td>
<td>0.1 s$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>$K_{xld}$</td>
<td>$10^{10} \text{ s}^{-1}$</td>
<td></td>
</tr>
<tr>
<td>$K_{zp,lo}$</td>
<td>$3.333 \times 10^{-3} \text{ s}^{-1}$</td>
<td>Zeta phosphorylation is 1/30 the value of epsilon phosphorylation to be similar to treatment of double phosphorylation in Faeder et al. [11]</td>
</tr>
<tr>
<td>$K_{zp,hi}$</td>
<td>10 s$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>$K_{zd}$</td>
<td>0.25 s$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>$K_{xzd}$</td>
<td>$5 \times 10^9 \text{ s}^{-1}$</td>
<td></td>
</tr>
<tr>
<td>$K_{zf}$</td>
<td>$5.68 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$</td>
<td></td>
</tr>
<tr>
<td>$K_{zz}$</td>
<td>0.1 s$^{-1}$</td>
<td></td>
</tr>
</tbody>
</table>
References


Separating the $Z^0$ from Heavy Quark Background in Relativistic Heavy Ion Collisions

Lewis L. Sharpnack
Kent State University
P-25 Subatomic Physics, Los Alamos National Labs
Mentor Gerd J. Kunde
August 10, 2006

Abstract

The Compact Muon Solenoid (CMS) particle detector at the Large Hadron Collider (LHC) is scheduled to begin experimental operations in late 2007. [1], [2] With a center of mass energy of 11 TeV, Lead-Lead (Pb-Pb) collisions at the LHC will produce quark gluon plasma (QGP), as well as numerous particle jets. A special class of these jets are $Z^0$ tagged jets. When a $Z^0$ and a parton (quark or gluon) are produced back to back during the collision of partons in the QGP, the produced parton interacts with the plasma through the strong force and loses energy, but the $Z^0$, which lacks the color charge responsible for the strong interaction, passes through the QGP without interacting. When the parton created during the collision exits the QGP, it causes a jet of hadrons to form. In order to determine the initial energy and momentum of the jet forming partons created during these collisions, $Z^0$ vector bosons coming out of the plasma on the opposite side of the jet are examined. Since direct observation of the $Z^0$ is impossible due to its short lifetime, we measure the decay products of the $Z^0$ to reconstruct the particle’s energy prior to exiting the QGP. Some of the most useful decay products of the $Z^0$ to study are muons produced as a particle antiparticle pair, also called a dimuon. Dimuons are also copiously produced from decays of other particles, most notably, decays of D and B mesons originating from heavy quark pairs. During event reconstruction these muons can mimic muons from the $Z^0$’s and constitute the background for our measurements. Using various simulations, the dimuon signals from the $Z^0$ decays and the D and B decays have been modeled. Based on the simulated data, a series of limits on accepted values for such physical quantities as opening angle and momentum asymmetry will be established to maximize the signal to background ratio for $Z^0$ particles in the detector.

Introduction

Relativistic heavy ion collisions contain some of the most complicated physical interactions that have been observed in the controlled conditions of a laboratory. Due to their enormous complexity, many tools need to be developed to simplify and reduce the amount of data that one looks at from such a collision. A lead on lead collision at the LHC when the accelerator is fully operational will produce thousands of particles. (By convention, any time a collision at LHC is referred to it will mean a Pb-Pb collision.) All of these particles provide us with information about the collision, but not necessarily with
a better understanding of the fundamental physics that occurred during the collision. This project focused on a special rare event that can occur at LHC energies.

The high energies of LHC require strong forces to produce detectable changes in the momenta and energies of particles. Color charge is responsible for the strong nuclear interaction that binds quarks into protons and neutrons and binds those protons and neutrons into the nuclei of atoms. The quark gluon plasma is a special plasma in which the constituent particles have a color charge rather than an electric charge, as they do in more normal plasmas such as solar plasma or lightning. The quarks carry partial electric charges, but the nuclear force is several orders of magnitude stronger than the electric force and it is the driving force for interactions in the QGP. Another feature unique to the QGP is that the energies are high enough that quarks and gluons are deconfined from hadrons (protons, neutrons, etc.) During a collision at the LHC, the energy density is high enough to form the QGP. In the QGP, many particles can be formed, including quarks and $Z^0$ vector bosons. $Z^0$ tagged jets can form when partons collide in the QGP with high energy forming another parton and a $Z^0$. Initially, the $Z^0$ and the parton have the same momentum. As the parton formed during this collision travels through the QGP, it loses some energy and momentum during strong interactions with the partons in the plasma. This plasma cools rapidly and the QGP only lasts for $10^{-23}$ seconds until it cools to energies where the quarks and gluons cannot exist as free particles. As the parton formed along with the $Z^0$ exits the QGP, it combines with other quarks and begins forming quark anti-quark pairs out of the vacuum. These quarks group together forming the hadrons that we detect in the lab as a particle jet. The total momentum of the particles in the jet is lower than the momentum of the $Z^0$, since the parton that formed the jet lost energy during interactions with the QGP. Consequently, jet suppression is a useful tool for studying the QGP because it probes the strength of the interactions occurring therein.

The challenge inherent to studying jet suppression is determining how much energy and momentum the parton that formed the jet had before they interacted with the plasma. This project studied one possible way of determining the jet forming parton’s initial energy and momentum. At the high energies available to the LHC, heavy particles, such as the $Z^0$ (Mass 91.1876 GeV/c²) are formed frequently. [3] Occasionally, the $Z^0$ is formed anti-parallel to a particle jet. The $Z^0$ is a vector boson and a carrier for the weak force. It carries no color charge and is thus ideally suited to probe the energy of the hard scattered jet forming parton because it will not interact with the QGP and retains its initial momentum and energy after interacting with the jet. The $Z^0$ has a short lifetime (mean life $3*10^{-25}$ seconds). [4] Even at relativistic velocities, it decays within a nanometer of the vertex or center of the collision. One of its main decay channels is into a dimuon (muon and antimuon pair). These muons don’t interact with the QGP because they don’t carry color charge. Regardless of whether the $Z^0$ decays in the QGP or just outside it, the dimuon produced during the decay carries the initial energy and momentum of the $Z^0$.

Since the $Z^0$ cannot be detected directly, we pair all the muons produced during collisions and reconstruct the energy and momentum of the $Z^0$. At this point we begin to introduce background into our data. Many process, other than the decay of the $Z^0$ that occur during relativistic heavy ion collisions, can produce muons. The most troublesome are the decays of heavy quarks which produce muons. Sometimes, the muons coming
from such decays can mimic the energy and momentum of a dimuon produced by a $Z^0$
decay, i.e. they have the same invariant mass as the $Z^0$.

Two methods are being explored to reduce this heavy quark background. The
first involves examining the point of origin of the muons. Mesons containing heavy
quarks, such as the B and D mesons, have longer half lives than the $Z^0$, and thus they
decay away from the vertex of the collision. By reconstructing their point of origin, it is
possible to determine whether the muons came from the vertex, as they would if they
were produced by the decay of a $Z^0$, or from outside the vertex, as they would if they
were produced by the decay of a B or D meson. The other method for separating the
background from the signal was the focus of this project. Since the muons are produced
in different ways between the background and signal, we expect that there would be some
subtle differences in certain kinematic variables between the signal and background. The
goal of this project was to find these differences and exploit them.

For this project, various programs were used to simulate the background and
signal. The main three were Pythia, MNR and a “Particle Gun” [Appendix A] Pythia is
an event simulator, which generates proton on proton collision data based on cross
sections for events determined in previous experiments. [5] MNR is a Quantum Chromo
Dynamics simulator, which we use to generate our background. The “Particle Gun”
program is used to simulate the decay of a single particle into a particle pair where the
masses of the decaying particle and its products are known and $P_t$, $\phi$, and $\theta$ of the $Z^0$
and the muons in the rest frame of the $Z^0$ are generated at random using realistic distributions.
For this programming, a few languages were used to write the code for the simulations.
Most of the code was written using ROOT, which is an object oriented programming
language developed for studying accelerator physics. [6] ROOT runs through a linux
environment and is based on C and C++. Unlike C and C++, ROOT has the advantage of
using an interpreter, which allows code to be written into macros and run without waiting
for a compiler.

The first part of this project centered on understanding the data from the particle
gun to make sure that it matched theoretical limits on quantities such as Opening Angle
between the muons that were determined using special relativity and Lorentz equations.
The particle gun program was then used to generate decay data for thousands of
simulated $Z^0$ particles. The simulated variables were plotted out into various histograms.
Additionally, signal cut efficiency plots were created. Using the data and histograms in
conjunction with the background data to be created in the near future, it will be possible
to develop signal and background cutting schemes to improve the signal to background
ratio.

Discussion

In heavy ion colliders, the lab reference convention has the axis the particle beam
travels on as the z-axis and the plane perpendicular to that axis as the xy plane, which is
referred to as the transverse plane. The kinematic variables used for the project were the
following (See Figure 1):

• $\eta$, pseudorapidity = -$\ln(\tan(\theta/2))$
• $\theta$ = Angle relative to the Beam Axis
• $\phi$ angle = Angle of the particle’s path in the XY Plane
• Opening Angle = Angle between Muons
• Transverse Momentum (PT) = Momentum in the XY Plane
• Momentum Asymmetry = |P1 - P2|/(P1 + P2)

The “Particle Gun” generates this data in the following way: The program generates an initial Z^0 with a random Transverse Momentum between 0 and 250 GeV/c, a random Phi angle between 0 and 2π, and a random Eta from -2.5 to 2.5. Next, the program decays the Z^0 into a pair of muons in the rest frame of the Z^0 with fixed momenta and a random orientation of the decay in the rest frame of the Z^0. The muons are always created back to back, but the dimuon in the rest frame of the Z^0 has a completely random orientation. The vectors describing the momentum of the muons are then pushed into the lab frame using Lorentz equations through a method from ROOT called boost, which pushes the two muons using the momentum and orientation of the Z^0 vector in the lab frame.

Using the simulated data from the “Particle Gun,” plots were created for variables such as Asymmetry and Opening Angle versus Total and Transverse Momentum of the Z^0. The following plots will be used to illustrate how the plots were created, how they were scrutinized for correctness, what they mean, and how they can be used to develop techniques for background suppression.

This first plot (Figure 2) shows the Opening Angle of the particles as a function of the Total Momentum of the Z^0. The primary features to notice in this plot are the curve below the scatter plot and the fact that very few particles are produced at high momentum with large Opening Angles. The line below the scatter plot is a theoretical lower bound for the Opening Angle between the muons produced by the Z^0. The line is calculated using conservation of momentum and energy. The total energy of the Z^0 must be equal to the total energy of the muons. Since the muons are less massive than the Z^0, they carry the extra energy as momentum. The case where the Opening Angle is minimized occurs when the muons are produced in the rest frame of the Z^0 perpendicular to the path that the Z^0 is traveling on in the lab frame. Similar calculations can be found in numerous undergraduate physics texts and on the web. [7] The important result of these calculations is the following equation describing the lower bound on the Opening Angle
versus $Z^0$ Total Momentum: Opening Angle Min = $2 \cdot \arctan(\sqrt{((M_Z)^2-2(M_\mu)^2)/P_z})$, where $M_Z$ is the $Z^0$ Mass, $M_\mu$ is the muon Mass, and $P_z$ is the $Z^0$ Total Momentum. Similarly one can determine the upper bound on the Opening Angle, using the conservation of momentum and energy. It turns out that, regardless of the Total Momentum of the $Z^0$, the particles can always be produced at an Opening Angle of $\pi$ radians, as long as they decay parallel to the direction of travel of the $Z^0$. Thus, we have data from the particle gun that agrees with analytic calculations. We were also able to use simulated events from Pythia, the collision event generator, and compare it with the particle gun data. Pythia is widely used and accepted in the particle physics community as a way to produce realistic collision simulations. After comparing the particle gun data with Pythia, the two were found to be in agreement, and we were satisfied that the particle gun created valid data.

The next step was to begin generating plots using the particle gun data. Several different types of plots were used to graph the data. The Figure 1 was a scatter plot. In a scatter plot, every dot corresponds to an event. The Opening Angle scatter plot shows that at different $Z^0$ Total Momenta, the Opening Angle had the values plotted. In the case of a 2D histogram, which are the colorful sets of plots that can be seen in Figure 3, the $y$ coordinate corresponds to a given value of one of the non-momentum variables and the $x$ coordinate is either $Z^0$ Total or Transverse Momentum. In this case the variables are plotted against the Total Momentum of the $Z^0$. These 2D histograms are formed when the code ROOT takes a data set and breaks it up into bins in $x$ and $y$. Each bin has a range of $x$ and $y$ values that it can contain and the plotting program places events in the bins based on that range. The program finishes by counting the number of entries in the bin and assigns the bin its color based on the number of entries. In this case the redder the color, the more points are in the bin. The scales for the color coding and how many events each color corresponds to can be seen on the right-hand side of the graph. The other type of plot that we used was a profile histogram. A profile histogram shows the underlying structure of the scatter plots. A profile histogram (Figure 4) separates the plot into bins along the $x$ axis (in this case 100 bins) in each bin the $y$-values of the points are averaged and a point is plotted with the $x$ and $y$ coordinates being the average $x$ and $y$ values in the given bin. A representative set of plots can be seen on the right. (Figure 4) The three variables that we expected to see the greatest difference in between signal and
background were the difference of the Phi Angles of the two particles, the Opening Angle between the muons, and the Momentum Asymmetry. Additionally, we looked at the differences in the Pseudorapidities of the two muons produced. We noticed immediately that the differences in Pseudorapidities and Phi Angles were symmetric about zero, so we chose to look at the absolute value of the difference in Pseudorapidity and Delta Phi. In this set of plots one can see that although the Momentum Asymmetry ranges from 0 to 1 in the scatter plot, it becomes flat with a value of .5 for the higher Total Momentum regions. One can also see that although Delta Phi ranges from 0 to π radians in the scatter plot it averages out to approximately 1 radian with a linearly decrease in the y average for the high Total Momentum regions. This information, while not immediately useful, may provide us with additional methods for separating background in the future, based upon the structure of the signal and background.

The next step in the process of looking at the background was to separate the signal data in the same manner that the background data is separated. In MNR, we go the background in various Transverse Momentum (P_T) bins. They were set to P_T =10, 20, 40, and 60 GeV/c. In order to compare the background with the signal from the “Particle Gun,” we had to generate the data and plots with fixed P_T at the same ranges as MNR. Essentially, we did the following: We took slices of the plots of kinematic variables at PT = 10, 20, etc in GeV/c (see Figure 5). We then plotted them in 1D histograms (see Figure 6), where the number of particles was on the y-axis and the value of the variable on the x-axis. Those plots can be seen below the Open Angle vs. P_T plot. Noteworthy features of the opening angle vs. P_T plots are that the distribution becomes more spread out as P_T increases and that the peak moves toward lower Opening Angles with higher P_T.

The final part of the analysis that needed to be completed was to use these 1D histograms to develop signal cut efficiency plots. A signal cut efficiency measures how much of the signal remains if you cut out signal events with a kinematic variable below the cut. For example, take the Opening Angle 1d histogram with P_T of the Z^0= 10 GeV/c. (Figure 7) To find the total signal, we integrate the histogram from an angle of 0 to π radians. We find the area under the curve, which is equal to the number of events generated in that particular P_T region. To get signal cut efficiencies we set our cut point to values from 0 to π in increments of π/100. We then
integrate the area under the histogram curve from our cut to \( \pi \). This is the number of signal events that remain after the cut has been applied. The signal cut efficiency is the ratio of the number of signal events remaining after the cut divided by the total number of events that were simulated. This ratio is 1 when 100% of the signal is preserved and 0 when all of the signal events are lost. For the background, we apply the same cuts and find the same ratios to find the background cut efficiency. As was mentioned earlier, the heavy invariant mass background has not been generated yet. As a result, background cut efficiencies and a background suppression scheme cannot be created for the \( Z^0 \). To illustrate what the cut efficiencies might look like and how a background suppression scheme is developed, the low invariant mass region, which is relevant to a virtual photon created in the same manner as the \( Z^0 \), has been plotted in the same plot as the \( Z^0 \) signal cut efficiency for \( \text{PT} = 10 \text{ GeV/c} \). (See Figure 8) The green crosses are the signal cut efficiencies at sequential opening angles and the red circles are the background at the same points. The goal in background suppression is to maximize the signal, while eliminating as much background as possible. If we were to cut out all events, where the Opening Angle between muons was below 1.5 radians, we would keep almost 100% of the signal (indicated by the cut efficiency being 1 for the signal at a cut of 1.5 radians) and we would eliminate 80% of the background. This means that we suppress the background by a factor of five. Unfortunately, this is only for the low invariant mass portion of the background. When the data for the high invariant mass background is available, it will probably more closely mimic the signal and the cuts will not be nearly as efficient as the one shown above.

Since one of the main goals of this project was to study the signal cut efficiency for the \( Z^0 \) signal, the plots of the cut efficiency using the different variables are in Appendix B. The most noteworthy change in the plots is that the signal cut efficiencies for Momentum Asymmetry, Opening Angle, and Delta Phi become flatter at higher \( \text{PT} \). Delta Eta remains virtually unchanged across the \( \text{PT} \) regions and, since it closely follows the background, Delta Eta probably won’t be useful for creating background suppressing cuts. Since the other three variables’ cut efficiencies flatten out at higher \( \text{PT} \), this indicates that the signal is spread over a broader range in those variables at higher \( \text{PT} \). If one looks back at the 1D histograms for the Opening Angle, they are wider and flatter at the higher \( \text{PT} \) ranges. This means that signal is cut out at lower Opening Angles. Since the distributions are also flatter, the signal cut efficiency becomes more linear. Unfortunately, if the low mass background is any indication, the background also flattens out in the Delta Phi, Opening Angle, and Asymmetry plots. Thus for higher Transverse Momentum cases it may be impossible to suppress the background with these simple cutting schemes.

Conclusions

As stated at the beginning, this project is a work in progress. The results so far are the signal cut efficiency plots, which can be seen in Appendix B. Until background data are available from MNR, no conclusive cutting schemes can be developed to...
suppress the heavy quark background. This work is continuing and will be completed some time in August or September to be presented at the Conference Experience for Undergraduates in October of this year. We can currently conclude that the signal cutting efficiencies will probably not be very efficient for the high invariant mass background if work by one of my colleagues on the virtual photon is any indication of what will occur with the $Z^0$. However, in the lower PT regions, where we expect many of the particles to be produced, the cuts will be an effective way of suppressing the background. This understanding of the $Z^0$ and the background associated with it will allow us to use it as an effective way of measuring jet energies and momenta when they are produced in Heavy Ion Collisions. Using these jets as probes, we will be able to further our understanding of the Quark Gluon Plasma.

References


Acknowledgements:
Thanks to my mentor Gerd J. Kunde
Thanks to post-docs Camelia Mironov and Paul Constantin
Thanks to the Kunde Crew: Miroslav Mihaylov, Maria A. Castro, Alberto Canabal-Rey, and Eric Schneider
Thanks to the students of Los Alamos Summer School 2006
Final thanks to Norm Magee, Sally Seidel and Lee Collins
Appendix A

The Particle Gun Code with modifications
Originally Written using ROOT code by Paul Constantin
Modified by Lewis L. Sharpnack to include additional variables, a realistic $Z^0$ mass distribution and $P_T = 10, 20, 40,$ or $60 \text{ GeV/c}$

```cpp
#include "TRandom.h"
const double pi = TMath::Pi();

void ZoDecay2(const int NumPart=200000, 
               const double m1=0.105, const double m2=0.105, 
               const double PtLow=1., const double PtHigh=200., const double EtaRange=3.0){
    TFile *fout = new TFile("ZoDecay2.root","RECREATE");
    TNtuple *ntu = new TNtuple("ntu","ZoDecay2 Ntuple",
    float var[20] = {0.};
    gRandom->SetSeed();
    // Loop over number of parent particles:
    for(int i=0; i<NumPart; i++){
        //Randomly Generate Masses Following the Z0 Gaussian Mass Distribution from Pythia
        double M = gRandom->Gaus(91.1602,1.85248);
        // Decay at rest the parent Z0:
        TLorentzVector child1(0.,0.,0.,0.);
        TLorentzVector child2(0.,0.,0.,0.);
        DecayAtRest(M,m1,m2,&child1,&child2);
        // Generate the parent kinematics:
        // for now, the simplest case - flat in all variables with PT binned to 10,20,40,or 60 GeV/c
        float PT;
        float randomizer  = gRandom->Uniform(0,1);
        if(randomizer<.25){
            PT=10.;
        }else if((randomizer>=.25)&&(randomizer<.5)){
            PT=20.;
        }else if((randomizer>=.5)&&(randomizer<.75)){
            PT=40.;
        }else if((randomizer>=.75)&&(randomizer<2)){
            PT=60.;
        }
        float PHI = 2*pi*gRandom->Uniform(0,1);
        float ETA = gRandom->Uniform(-EtaRange,EtaRange);
        TLorentzVector Parent(0.,0.,0.,0.);
        Parent.SetPtEtaPhiM(PT, ETA, PHI, M);
        var[0] = M;
        var[1] = Parent.P();
        var[2] = PT;
        var[3] = PHI;
        var[4] = ETA;
        // Lorentz boost the children:
        TVector3 b = Parent.BoostVector();
        child1.Boost(b);
        child2.Boost(b);
    }
}
```
// Get the children kinematics:
float pt1 = child1.Pt();
float phi1 = child1.Phi();
float eta1 = child1.Eta();
float mom1 = child1.P();
float pt2 = child2.Pt();
float phi2 = child2.Phi();
float eta2 = child2.Eta();
float mom2 = child2.P();
float E1 = child1.Energy();
float E2 = child2.Energy();
var[6] = E1;
var[7] = mom1;
var[8] = pt1;
var[9] = phi1;
var[10] = eta1;
var[12] = mom2;
var[13] = pt2;
var[14] = phi2;
var[15] = eta2;
var[16] = child1.Angle(child2.Vect());
//var[17] = phi1-phi2;
//dot product based method for finding the differences in the phi angles of the muons
float x1=child1.X();
float y1=child1.Y();
float x2=child2.X();
float y2=child2.Y();
float mag1=sqrt(x1*x1+y1*y1);
float mag2=sqrt(x2*x2+y2*y2);
x1/=mag1;
y1/=mag1;
x2/=mag2;
y2/=mag2;
float dotprod=x1*x2+y1*y2;
float dphi = TMath::ACos(dotprod);
//var[17] = dphi;
var[18] = eta1-eta2;
//momentum asymmetry
var[19] = fabs(mom1-mom2)/(mom1+mom2);
//energy asymmetry, not used in later codes used earlier to see if there were differences between momentum and energy asymmetries
//var[19] = fabs(E1-E2)/(E1+E2);
ntu->Fill(var);
}

fout->Write();
fout->Close();
}

// method to decay a particle of mass M into two particles of masses m1 and m2
// fills the TLorentzVectors of the childs in the rest frame of the parent
void DecayAtRest(float mpar, float mch1, float mch2,
TLorentzVector *child1, TLorentzVector *child2){

// Get the children kinematics:
float pt1 = child1.Pt();
float phi1 = child1.Phi();
float eta1 = child1.Eta();
float mom1 = child1.P();
float pt2 = child2.Pt();
float phi2 = child2.Phi();
float eta2 = child2.Eta();
float mom2 = child2.P();
float E1 = child1.Energy();
float E2 = child2.Energy();
var[6] = E1;
var[7] = mom1;
var[8] = pt1;
var[9] = phi1;
var[10] = eta1;
var[12] = mom2;
var[13] = pt2;
var[14] = phi2;
var[15] = eta2;
var[16] = child1.Angle(child2.Vect());
//var[17] = phi1-phi2;
//dot product based method for finding the differences in the phi angles of the muons
float x1=child1.X();
float y1=child1.Y();
float x2=child2.X();
float y2=child2.Y();
float mag1=sqrt(x1*x1+y1*y1);
float mag2=sqrt(x2*x2+y2*y2);
x1/=mag1;
y1/=mag1;
x2/=mag2;
y2/=mag2;
float dotprod=x1*x2+y1*y2;
float dphi = TMath::ACos(dotprod);
//var[17] = dphi;
var[18] = eta1-eta2;
//momentum asymmetry
var[19] = fabs(mom1-mom2)/(mom1+mom2);
//energy asymmetry, not used in later codes used earlier to see if there were differences between momentum and energy asymmetries
//var[19] = fabs(E1-E2)/(E1+E2);
ntu->Fill(var);
}

fout->Write();
fout->Close();
}
// energy-momentum conservation:
float sum = mch1+mch2;
float diff = mch1-mch2;
float e1 = (mpar*mpar-mch2*mch2+mch1*mch1)/mpar/2.;
float e2 = (mpar*mpar-mch1*mch1+mch2*mch2)/mpar/2.;
float p  = sqrt((mpar*mpar-sum*sum)*(mpar*mpar-diff*diff))/mpar/2.;

// momentum components with random orientation:
float phi   = 2*pi*gRandom->Uniform(0,1);
float theta = TMath::ACos(1-2*gRandom->Uniform(0,1));
float px = p*TMath::Sin(theta)*TMath::Cos(phi);
float py = p*TMath::Sin(theta)*TMath::Sin(phi);
float pz = p*TMath::Cos(theta);
child1->SetPxPyPzE( px,  py,  pz,  e1);
child2->SetPxPyPzE(-px, -py, -pz,  e2);
}

This code generates and stores the data in an ntuple. The original code for this
particle gun was written by Paul Constantin, and modified by Lewis Sharpnack to be
more accurate to the decay of the $Z^0$.

{This macro needs the preexisting file /root_v5/root/macros/ZoDecay2.root
//This macro creates 4 histograms in a single canvas
//The plots are opening angle histograms with PT bins of 10, 20, 40, 60

c1 = new TCanvas("c1","Realistic Zo mass opening angle plots",50,10,900,900);
c1->SetGrid();
c1->SetFillColor(11);
c1->GetFrame()->SetFillColor(11);
c1->GetFrame()->SetBorderSize(6);
c1->GetFrame()->SetBorderMode(-1);
gStyle->SetPalette(1);
pad1 = new TPad("pad1","PT cut >10",0.05,0.50,0.45,0.95,21);
pad2 = new TPad("pad2","PT cut >20",0.05,0.05,0.45,0.45,21);
pad3 = new TPad("pad3","PT cut >40",0.50,0.50,0.95,0.95,21);
pad4 = new TPad("pad4","PT cut >60",0.50,0.05,0.95,0.45,21);
pad1->Draw();
pad2->Draw();
pad3->Draw();
pad4->Draw();
TFile f("ZoDecay2.root");

//Draws the PT>10 histogram in pad1
pad1->cd();
pad1->GetFrame()->SetFillColor(42);
pad1->GetFrame()->SetBorderMode(-1);
pad1->GetFrame()->SetBorderSize(5);
TH1D *Hist10 = new TH1D("Hist10","Opening Angle with PT==10;Opening Angle
(Radians);Number of Entries",100,0.,3.2);
ntu->Draw("open>>Hist10","PT==10","n");
pad1->SetLogy(1);
pad1->SetGridx();
This code uses the ntuple to generate 1d histograms for the opening angle that were shown earlier in the paper.
Appendix B

Signal Cutting Efficiencies for PT = 10 GeV/c

![Graphs showing signal cutting efficiencies for PT = 10 GeV/c.](image)

Signal Cutting Efficiencies for PT = 20 GeV/c

![Graphs showing signal cutting efficiencies for PT = 20 GeV/c.](image)
Signal Cutting Efficiencies for PT = 40 GeV/c

Signal Cutting Efficiencies for PT = 60 GeV/c
Exploring the Feasibility of Using Dynamic Light Scattering as a Method of Assessing Protocell Replication.

Ken Soong
Pierre-Alain Monnard, Hans Ziock

I. Abstract
Dynamic light scattering is a technique used to determine the size distributions of particles in a solution. These particles may be as small as a few nanometers or as large as a few micrometers. In this paper we have evaluated the possibility of using the single angle dynamic light scattering method to determine the composition of a solution consisting of protocells and their precursor material. We have found that dynamic light scattering methods can effectively determine the size of a homogeneous population of particles. However, for populations with two particle sizes, the single angle dynamic light scattering method becomes significantly less effective at determining the size distributions of the constituent particles.

II. Introduction
Protocell Assembly is a Los Alamos National Laboratory LDRD-DR sponsored project, which seeks to assemble a minimal self-replicating molecular machine. This project seeks to develop the underpinning science for the assembly of functional protocells, i.e., simple self-reproducing nanosystems that can perform useful tasks. These proposed protoorganisms are only about 5 nm in size. In their simplest form, these protoorganisms consist of a micelle which acts as the container, a light driven metabolism, and a peptide nucleic acid (PNA) based genetic system, whose functions are all very tightly coupled. The container consists of amphiphilic fatty acid molecules that self-assemble into a micelle. The hydrophobic interior of the micelle provides an alternative thermodynamic environment from the aqueous exterior and acts as a sticking point for the photosensitizer, fatty acid precursors (food), and the genetic material. The metabolism involves the photoexcitation of an electron in the photosensitizer which is stabilized by the donation of an electron from one of the PNA bases. The excited electron is in turn used to cleave a fatty acid precursor to yield another fatty acid molecule, thereby allowing the container to grow until it reaches an unstable size and divide. In this paper we explore the feasibility of using the dynamic light scattering method to follow the growth and division of these protocell containers.

III. Dynamic Light Scattering Theory
Dynamic light scattering is a technique used to determine the size of particles in solution. When a beam of light passes through a solution, the particulate components scatter some of the light in all directions. If the light is coherent and monochromatic, in our case a laser, it is possible to detect time-dependent fluctuations in the scattered intensity using an appropriate photomultiplier. These time-dependent fluctuations arise from the fact that the particles in solution undergo random Brownian motion, causing the distance between particles to be constantly varying. These moving particles scatter light and create constructive and destructive interference which is detected by the photomultiplier. Analysis of the time dependence of the fluctuations can yield the
diffusion coefficient, from which we can infer, via the Stokes-Einstein equations, the hydrodynamic radius.

The waves scattered by particles in the solution are not all constructive or destructive and the phase of these waves must be accounted for. We will use a particle located at the origin as our reference. The path of the wave scattered by a particle located at radius vector \( r \) from the origin, differs from the path traveled by a wave scattered by a particle at the origin by the segments 1 and 2, with lengths \( l_1 \) and \( l_2 \) respectively (Figure 1). The phase difference is then,

\[
\Delta \phi = k(l_1 + l_2), \quad \text{where } k = \frac{2\pi n}{\lambda}.
\]  

Figure 1. A schematic diagram of the paths traveled by two waves of light as it passes through a section of the sample.

We can see that segment \( l_1 \) is the projection of \( r \) onto the wave vector of the incident beam \( k_0 \), or otherwise \( l_1 = -r \cdot k_0/k \). Similarly, \( l_2 = -r \cdot k/k \). From equation (1) we see that \( \Delta \phi = r \cdot (k_0 - k) = r \cdot q \), where \( q = (k_0 - k) \). The vector \( q \) is known as the scattering vector, which is a fundamental characteristic of any scattering process. The length of the vector \( q \) is given by

\[
q \equiv |q| = \frac{4\pi n}{\lambda} \sin \frac{\theta}{2} \quad (2)
\]

where \( n \) is the index of refraction of the medium, \( \lambda \) is the wavelength of light, and \( \theta \) is the scattering angle. The combined effects of the destructive interference reduces the intensity of the light scattered by a factor of \( |\alpha|^2 \), where \( \alpha \) is the average value of the phase factors \( \exp(iq \cdot r) \) for all particles.

The idea behind the dynamic light scattering technique is to measure the temporal correlations in the fluctuations of the scattered light intensity and from this, to reconstruct the physical characteristics of the scatterers. Although the photodetector signal in dynamic light scattering looks like random noise, there is information contained in the correlation function of this random noise. The correlation function of the photocurrent \( (i) \), is defined as,

\[
G^{(2)}(\tau) = \langle i(t) \cdot i(t + \tau) \rangle \quad (3)
\]

and the correlation function of the electromagnetic field \( (E) \) as,
\[ G^{(1)}(\tau) = \langle E(t) \cdot E^*(t + \tau) \rangle \]  

(4)

In the above formula, the angular brackets denote an average over time \( t \), which is necessary to extract information from the random fluctuations in the intensity of scattered light. The instrument we used determined the correlation functions digitally.

Due to the fact that the scattered light is the sum of waves scattered by many independent particles and therefore displays Gaussian statistics, we can find a relationship between the intensity correlation function and the field correlation function, namely

\[ G^{(2)}(\tau) = I_0^2 (1 + \gamma |g^{(1)}(\tau)|^2). \]  

(5)

As explained earlier, the diffusion of the solute is a product of Brownian motion. This diffusion is quantitatively characterized by the diffusion coefficient \( D \). Further analysis of the process of light scattering by Brownian particles leads to the following expression for the correlation function of the scattered light as a function of the diffusion coefficient:

\[ |g^{(1)}(\tau)| = \exp(-Dq^2\tau). \]  

(6)

If we now combine equations (5) and (6) we see that from measurements of the intensity correlation we can obtain the diffusion coefficients of the scattering particles. For spherical particles, the relationship between its radius \( R \) and its diffusion coefficient \( D \) is given by the Stokes-Einstein equation:

\[ D = \frac{k_B T}{6\pi \eta R} \]  

(7)

where \( k_B \) is the Boltzmann constant, \( T \) is the absolute temperature, and \( \eta \) is the viscosity of the solution. For non-spherical particles, we approximate it with a hydrodynamic radius \( R_h^{app} \), defined as:

\[ R_h^{app} = \frac{k_B T}{6\pi \eta D^{app}} \]  

(8)

where \( D^{app} \) is the diffusion coefficient measured in the dynamic light scattering experiment.

IV. Method

To study the feasibility of using dynamic light scattering as a method of analyzing protocell dynamics, we calibrated the single angle dynamic light scattering technique on particles of known size. For our standards we used Latex Microsphere Suspensions of diameter 30nm, 60nm, and 200nm. These standards contain microspheres made of polystyrene, which are packaged in aqueous suspensions at 10% solids by weight. These microspheres have a refractive index of 1.59 at 589nm and a density of 1.05g/cm³.

We began the calibration by analyzing the size distribution of each individual standard in isolation. In order to avoid saturating the response from the photodetector and to reduce the effects of the multiple scattering of light, we diluted the standard suspensions in filtered water. The standards were diluted until the photon count rates were less than 1 million counts per second. The 30nm standard was diluted to 1 part in 300, the 60nm standard was diluted to 1 part in 2000, and the 200nm solution was diluted to 1 part in 84000. These diluted solutions were then placed in the CoolBatch 90T System for analysis. For each of the diluted standards, dynamic light scattering measurements were made at 20°C and a minimum of three trials were performed.
After completing the analysis of the homogeneous standard suspensions, we proceeded to analyze systems comprised of a mixture of two pure standards. Since all of the standard suspensions were 10% solids by weight and had the same density, we can easily calculate the required proportions to get an equal number of particles per unit volume for each standard in a solution. To calculate the average volume containing a single microsphere particle of diameter \(d\), we carried out the following derivation:

\[
\begin{align*}
\text{(a)} & \quad \text{Volume of 1 particle: } \frac{4}{3} \pi \left(\frac{d}{2}\right)^3 \\
\text{(b)} & \quad \text{Mass of 1 particle: } 1.05 \text{g/cm}^3 \cdot \text{(a)} \\
\text{(c)} & \quad \text{Mass of starting solution containing 1 particle: } 10 \cdot \text{(b)} \text{ (i.e. 10\% by weight solids)} \\
\text{(d)} & \quad \text{Mass of water in the solution volume containing 1 particle: } \text{(c)} - \text{(b)} \\
\text{(e)} & \quad \text{Volume of suspension containing 1 particle: } \text{(d)} + \text{(a)}, \text{ (assuming a density of 1 for pure water) or equivalently: } 10.45 \cdot (\text{Volume of particle})
\end{align*}
\]

From this it is apparent that the ratio of suspensions required to have an equal amount of particles is simply the ratio of the diameters cubed.

Once the number of particles per unit volume was calculated, we varied the concentrations of the mixtures. For each combination of standard suspension pairs we made a series of solutions with the ratio of small particles to large particles being 5/95, 10/90, 15/85, 20/80, 50/50, 80/20, 85/15, 90/10, and 95/5. Each of these solutions was then diluted to the point where the average intensity reading from the photodetector was less than 1 million counts per second and then placed in the CoolBatch 90T System for analysis. Again a minimum of three trials were performed on each solution.

V. Results
The data from the dynamic light scattering analysis of the solutions with only a single sized particle showed us what we expected. The 30nm standard solution displayed a Gaussian distribution with an average diameter of 33.4nm. The 60nm standard solution displayed a Gaussian distribution with an average diameter of 59.6nm. And the 200nm standard solution displayed a Gaussian distribution with an average diameter of 202nm (see Figure 2).
Figure 2. Histograms of particle size vs. their relative abundance, as determined by the dynamic light scattering technique. The parameter D represents the average diameter and 100% indicates that all of the data collected is displayed in the histogram. Histograms of the 30nm standard (top), the 60nm standard (mid), and the 200nm standard (bottom) are displayed.

Additionally, when the dynamic light scattering technique was applied to two separately prepared samples of phosphatidylcholine (POPC) vesicles, the results were exceedingly similar (see figure 3).

Figure 3. Histograms of particle size vs. relative abundance for a two separately prepared samples of POPC vesicles with pyrene hexanoic acid. Each of these samples was extruded through a 1 micron filter before they were analyzed.
Furthermore, once POPC vesicles, of average size 550nm, were extruded through a 100nm filter, dynamic light scattering analysis displayed a size distribution centered on 102nm. This agrees well with our expectations since the process of extruding through a 100nm filter seeks to remove particles larger than 100nm from the solution, leaving only particles smaller than 100nm.

![Histogram of particle size vs. relative abundance for POPC vesicles. This histogram was plotted from data obtained by dynamic light scattering. These vesicles initially averaged about 550nm but have been extruded through a 100nm filter. As expected the distribution of the particles are located around 100nm.](image)

The dynamic light scattering analysis of the standard suspension mixtures on the other hand, did not produce such agreeable results. The mixture of the 30nm and the 60nm standards produced a distribution centered approximately on 58nm, regardless of the relative concentration of the particles (see Figure 5). Furthermore both the mixtures of the 30nm standard with the 200nm standard and the mixtures of the 60nm standard with the 200nm standard gave an average particle diameter centered approximately on 180nm, again regardless of the relative concentrations (see Figure 6).

![A plot of the average diameter of the particles, as determined by dynamic light scattering, as a function of the 30nm particles present (relative to 60nm particles). The black line is a trend line of the observed data, and the red line represents the ideal trend, assuming a linear relationship between the concentration and the average diameter.](image)
Figure 6. A plot of the average diameter of the particles, as determined by dynamic light scattering, as a function of the 60nm particles present (relative to 200nm particles). The black line is a trend line of the observed data, and the red line represents the ideal trend, assuming a linear relationship between the concentration and the average diameter.

In order to compensate for the fact that the intensity of the scattered light from the larger particles may completely overwhelm the light scattered from the smaller particles we employed the molecular weight normalization function. The molecular weight normalization function assumes the power law dependency of the molecular weight on the size of the particle as defined by: $MW = kR_h^\alpha$. We used $\alpha=3$, which is the typical $\alpha$ for globular proteins. The molecular weight normalization function attempts to apply a greater weighing on particles of smaller radii, in order to compensate for the greater quantity of light scattered by larger particles. Using molecular weight normalization we obtained concentration normalized plots, as opposed to intensity normalization plots. Our normalization function was calculated and applied to the data by the PrecisionDeconolve software.

The results of normalizing the plots did slightly improve the correlation between the concentration and the average diameter, but overall it is apparent that molecular weight normalization did not create a clear relationship between the concentrations of the particles and the average diameter of the aggregate (see Figure 7 and Figure 8).
30 nm and 60 nm Solution

\[ y = -0.0966x + 54.302 \]
\[ R^2 = 0.1731 \]

Figure 7. A plot of the average diameter of the particles, after a molecular weight normalization function has been applied, as a function of the 30nm particles present (relative to 60nm particles). The black line is a trend line of the observed data, and the red line represents the ideal trend, assuming a linear relationship between the concentration and the average diameter.

60 nm to 200 nm Solution

\[ y = -0.738x + 112.45 \]
\[ R^2 = 0.492 \]

Figure 8. A plot of the average diameter of the particles, after a molecular weight normalization function, has been applied as a function of the 60nm particles present (relative to 200nm particles). The black line is a trend line of the observed data, and the red line represents the ideal trend, assuming a linear relationship between the concentration and the average diameter.

Molecular weight normalization did, however, allow us to extract two distinct populations in some of the data of the 30nm and 200nm mixtures and of the 60nm and
200nm mixtures. Unfortunately, the area under the size distribution curves did not correspond to the relative concentrations of the solutions (see Figure 9 and Table 1).

Figure 9. Histograms of particle size vs. relative abundance for a solution containing 20% 30nm particles and 80% 200nm particles. Top: The histogram before the application of a molecular weight normalization function. Bottom: The histogram after the application of a molecular weight normalization function. After normalization, two distinct peaks can be seen one around 44.4nm and another around 183nm.

Table 1. A table of the data from the histogram of particle size vs. relative abundance, after molecular weight normalization. The first two columns indicate the actual composition of the solution, and the remaining columns contain data from the dynamic light scattering measurements.

<table>
<thead>
<tr>
<th>Percent 30nm particles in the solution</th>
<th>Percent 200nm particles in the solution</th>
<th>Average size of the smaller particle detected</th>
<th>Fraction of the observed that is the smaller particle</th>
<th>Average size of the larger particle detected</th>
<th>Fraction of the observed that is the larger particle</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>95</td>
<td>19.2</td>
<td>96</td>
<td>178</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>90</td>
<td>17.3</td>
<td>96</td>
<td>176</td>
<td>4</td>
</tr>
<tr>
<td>15</td>
<td>85</td>
<td>16</td>
<td>63</td>
<td>174</td>
<td>37</td>
</tr>
<tr>
<td>20</td>
<td>80</td>
<td>23</td>
<td>83</td>
<td>173</td>
<td>17</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>20.8</td>
<td>90</td>
<td>174</td>
<td>10</td>
</tr>
<tr>
<td>80</td>
<td>20</td>
<td>16</td>
<td>74</td>
<td>177</td>
<td>26</td>
</tr>
<tr>
<td>85</td>
<td>15</td>
<td>68.8</td>
<td>55</td>
<td>179</td>
<td>45</td>
</tr>
<tr>
<td>90</td>
<td>10</td>
<td>33.2</td>
<td>71</td>
<td>151</td>
<td>29</td>
</tr>
<tr>
<td>95</td>
<td>5</td>
<td>39.8</td>
<td>91</td>
<td>191</td>
<td>9</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>22.8</td>
<td>100</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Continuation of Table 1.

<table>
<thead>
<tr>
<th>Percent 60nm particles in the solution</th>
<th>Percent 200nm particles in the solution</th>
<th>Average size of the smaller particle detected</th>
<th>Fraction of the observed that is the smaller particle</th>
<th>Average size of the larger particle detected</th>
<th>Fraction of the observed that is the larger particle</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
<td>143</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>95</td>
<td>39.8</td>
<td>91</td>
<td>183</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>90</td>
<td>47.2</td>
<td>76</td>
<td>155</td>
<td>24</td>
</tr>
<tr>
<td>15</td>
<td>85</td>
<td>64</td>
<td>91</td>
<td>190</td>
<td>9</td>
</tr>
<tr>
<td>20</td>
<td>80</td>
<td>47.8</td>
<td>52</td>
<td>161</td>
<td>48</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>39.8</td>
<td>32</td>
<td>178</td>
<td>68</td>
</tr>
<tr>
<td>80</td>
<td>20</td>
<td>23</td>
<td>95</td>
<td>173</td>
<td>5</td>
</tr>
<tr>
<td>85</td>
<td>15</td>
<td>34.6</td>
<td>79</td>
<td>178</td>
<td>21</td>
</tr>
<tr>
<td>90</td>
<td>10</td>
<td>21.8</td>
<td>92</td>
<td>171</td>
<td>8</td>
</tr>
<tr>
<td>95</td>
<td>5</td>
<td>23.8</td>
<td>92</td>
<td>172</td>
<td>8</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>45</td>
<td>100</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

VI. Discussion

For a homogeneous particle population the dynamic light scattering technique seemed to have no trouble calculating an accurate average diameter. However once multiple particle sizes are present the technique becomes inaccurate and the calculations of average diameter become very misleading. Even at only a 5% concentration of 200nm particles, the contributions from the 30nm particles seems to have little effect on the average diameter. The possible problem may lie in the relatively short time scale in which the trials were conducted. Since larger particles move much slower than smaller particles, once a large particle moves into the field of view it may linger for a large extent of the trial. The effects of the time scale of the trial on the readings can easily be tested by increasing the integration time of the photodetector before an average intensity is recorded. This however, may end up smoothening out the shape of the size distribution to the point where only one widely spread peak is displayed. Alternatively, a large number of trials can be performed and the average of all trials can give a more robust average diameter.

Another possible source of error lies in the arrangement of the large and small particles. Large particles scatter more light and may hide or greatly reduce the effects of the scattering by the small particles. Further, large particles may shield small particles from the laser, effectively rendering the small particles invisible. These effects may cause the instrument’s correlator to miss the small particles, or worst to incorrectly infer a particles size. A possible resolution to this problem may be to dilute the solutions, so the chances of such interactions occur less frequently. This however, increases the effects of both background noise and impurities in the solution.

Another apparent problem is the molecular weight normalization factor, which when applied to the standards in isolation give values which are far below what is expected (see Table 1, specifically the 0% and 100% values). This flaw in the normalization function arises from the fact that each particle in the standard does not have the exact diameter published by the manufacturer, but instead the standard contains
particles which have a Gaussian distribution centered on the published diameter. The application of the normalization function increases the weight of the slightly smaller particles to the point where the distribution is no longer Gaussian, instead the distribution peaks near the smallest particle detected and decreases with size thereafter. Since the normalization function can be altered after the data are recorded, one can tinker with the values for \( k \) and \( \alpha \) to try to obtain results which match the properties of the solution being tested.

In future studies, one may try to calibrate the measurement parameters, such as the correlation run time, and integration time, as well as adjust the molecular weight normalization function in order to obtain results that better match the properties of the sample being tested. These parameters can be calibrated for a single known solution and the same parameters can be applied to a series of known samples to determine whether these parameters are robust over the range of possible concentrations.

VII. Conclusion

The data from the analysis of a single homogeneous population of particles show us that the dynamic light scattering technique can accurately and consistently measure the average diameter of a single particle population. In this way, the dynamic light scattering technique can be used to analyze micelles and vesicles with one population size, or samples which have been extruded through filters smaller than the particle itself. Conversely, for samples which have heterogeneous population sizes, such as a dynamic system of protocells, our current results suggest that single angle dynamic light scattering will not provide an accurate analysis of the sample. The average diameter as calculated by the single angle dynamic light scattering technique drastically under represents the smaller particles in the solution. Although the application of the molecular weight normalization function does provide us with some information about the populations present, it does not provide us with an accurate estimate of the relative concentrations of each particle population.

VIII. Acknowledgements

I would like to acknowledge the people responsible for this project. First and foremost, I would like to acknowledge my mentors Hans Ziock and Pierre-Alain Monnard, without their guidance and instruction this project would not have been possible. I would also like to acknowledge Norman Magee for his dedication, time, and effort which go above and beyond his duties. And I would like to acknowledge Lee Collins, James Colgan, and Sally Seidel, for their efforts in administering the Los Alamos Summer School program.

IX. References

Evaluating Semi-Empirical Models for Electron Impact Double Ionization Cross Sections for Positive, Neutral, and Negative Ions

B. Streufert¹, J. Colgan
Los Alamos National Laboratory, Los Alamos, NM 87545

Abstract. Accurate electron impact double ionization (DI) cross sections are very difficult to calculate for almost all atoms and ions, therefore semi-empirical modeling is of increasing interest. In this paper, we test three semi-empirical models for electron impact DI cross sections, reporting on the strengths and weaknesses of each model. A background on semi-empirical modeling of DI cross sections is presented, providing a basis for analysis of the three models we chose. The models vary in input parameters and method. The ions of interest represent a variety of cases to test the models, encompassing the neutral atom helium, singly ionized light ions, highly charged argon and krypton ions, and the negative ions C⁻ and O⁻. We present results by comparing our models to experimental results on electron impact DI cross sections of each of the ions. Finally, we discuss possible applications to plasma kinetics modeling.

¹ Present address: University of Virginia, Charlottesville, VA 22192
1. Introduction

Double ionization of ions can involve an impact of an electron of sufficient energy to ionize two electrons, creating 3 electrons in the exit channel:

$$e^- + A^{q+} \rightarrow e^- + A^{(q+2)^+} + 2e^-.$$  

The cross section of this process, called electron impact double ionization (DI), is negligible at all but high impact electron energies. Even then, the ionization of ions is dominated by single ionization (SI), the process whereby an impact electron ejects one bound electron from an ion, rather than the less likely DI. DI is generally approximately 1000 times less significant in the concurrent energy regimes.

The process of double and even more multiple ionization can become significant in a plasma, where electron energies can reach those high enough to cause double ionization of ions. The charge state distribution of a plasma is an important characteristic in plasma physics and is certainly affected by multiple ionization processes. Fisher et al. (1995) note that multiple ionization processes may contribute significantly to consecutive single ionization events. In the case of contributing to charge state distributions, it is important to know the DI cross section over the range of energies of plasma electrons, from orders of eV to MeV. Evidence of DI has purportedly been seen in high energy laser impact on metal targets (Mancini 2006), creating interest in testing the effect of DI on less extreme systems.

The processes involved remove two electrons from an ion. The simplest process—so-called direct DI—ionizes two outer shell electrons, as the impact electron knocks out two others (1). More exotic processes—so-called indirect DI—occur when the incoming electron impacts an inner shell electron. The collision can ionize or simply excite the inner shell electron, leaving the ion in a highly excited state. The ion is said to be in an autoionizing state, which then decays by emitting a number of electrons. Khouilid et al. (2001) describes this ionization-autoionization (IA) as a single ionization process in the ion after loss of an inner shell electron. Further, resonant processes can cause DI. In this type of process, the impact electron excites an inner shell electron, thereby losing most of its energy. The electron is captured by the ion, creating a multiply-excited autoionizing state which can emit up to 3 electrons (Belenger et al. 1997). Each process, to qualify as DI, must have an impact electron of sufficient energy to create an intermediate state which then emits 3 electrons. In this study, however, the focus is only on direct processes involving outer shell electrons.

The theoretical background on DI of ions is limited. The complexity of considering four particles in the exit channel interacting via long-range Coulomb potentials makes computation very difficult (Belenger et al. 1997). Calculations of DI cross sections of any accuracy are usually prohibitively time consuming because of this complexity. The solution to this problem is predicting the DI cross section over a range of energies using semi-empirical modeling. A range
of models exist, varying in number of inputs and accuracy against experimental data. One application of a model for DI is in a plasma kinetics calculation, which to the authors' best knowledge has not been done. The models proposed here could eventually be used in the Los Alamos multi-purpose ionization code (GIPPER) (Archer et al. 2002) code for plasmas of interest.

In this paper we will study three models over a range of energies and ions against experimental data from previous studies of DI. The ions were chosen because of their interest as applicable to certain plasmas, including highly ionized noble gases, ions to be isoelectronic with aluminum at certain ion stages, and ions that would round out a variety of situations to test the models. Also, in the short time frame of the study, the availability of experimental data guided our selection. The impact electron energy ranges for these experiments vary from about 10 eV to about 6000 eV. We judge the models on the ability to reproduce accurately the collected experimental data and hope to make recommendations of the best models for different ions and charge states.

2. Theory

Computation of cross sections for electron impact DI can be difficult due to the four bodies in the exit channel. As such, modeling the cross sections for DI processes relies heavily on experimental data. Several well-known relationships of parameters in single ionization (SI), however, provide useful predictions of cross sections over a wide variety of energies, completing the semi-empirical models studied here. While this study focuses primarily on double ionization, many models tackle multiple ionization in general, including triple and higher order ionization (Fisher et al. 1995, Belenger et al. 1997, Shevelko et al. 2005, Shevelko et al. 2006).

In direct DI, much of the theoretical work in semi-empirical models derives fromanalogs in SI models. The classical Thomson scaling law, for instance, models SI over a wide variety of energies and reproduces high quality models in DI:

$$\sigma_{\text{max}} = \frac{B}{I^2}. \quad (2)$$

The parameter $B$ represents a scaling factor and $I$ the energy of the ionization process in consideration. Belenger et al. (1999) studied the behavior of fluorine-like ions and found that $B$ depends strongly on ion charge state for a given isoelectronic structure, producing better results for weak charge states. In exploration of a selection of different models, most were found to employ the Thomson scaling law (2). A notable exception is recent work of Shevelko et al. (2005) that studied DI and found better agreement with cross section proportional to $I^{-3}$.

Most semi-empirical models employ dependence on the number of target electrons to better agree with a wide variety of charge states over a given energy. Zambra et al. (1994)
provide a simple scaling law they call the ‘electron pair formation law,’ to include the number of pairs that could be targeted for DI with $n_e$ electrons in the shell:

$$C_2^n = \frac{n_e(n_e - 1)}{2}.$$  

(3)

This study considered the value of $n_e$ to be the number of electrons in the outermost subshell, dependent on both principal and angular momentum quantum number, such as the 2$p$ or 3$d$ subshells. The scaling factor then gives the number of different pairs—the number of permutations—of electrons that could be ejected for DI. When multiplied by the cross section of a single occurrence of DI from that shell, the model gives the total predicted cross section. Models can incorporate other scaling factors, almost exclusively derived from empirical methods in the direct DI case, to better approximate experimental data.

The final element of the semi-empirical model for DI is the dependence on impact electron energy. In the low energy region, near the threshold for DI for a given ion, Wannier (1955) predicted a model based on the number of electrons ejected $w$ and the electron energy in threshold units, $x = E_{\text{impact}}/I_2$:

$$\sigma \propto (x - 1)^w.$$  

(4)

For the double ionization models considered here, the value of $w$ is approximately 2. Fisher et al. (1995) verified this value of 2, examining 18 experimental DI cross sections that give values of $w = 2 \pm 0.5$.

In the high energy region, the Bethe-Born approximation gives for an electron energy $x$ and DI energy $I_2$:

$$\sigma \propto \frac{1}{I_2^2 x} \ln(x + b).$$  

(5)

Belenger et al. (1999) note the difference between SI and DI application of the approximation: in SI the logarithmic term dominates while the $x^{-1}$ term can dominate in some DI cross sections. The models discussed below provided the focus of the study and incorporate many of the various empirical and semi-empirical methods discussed above.

2.1. Zambra et al. (1994)

In a study of singly charged light ions, Zambra et al. (1994) proposed a model of the DI cross section over a range of energies $x$, given here in threshold units:
The model employs the classical Thomson scaling law (2), $x^l$ energy dependence as in the Bethe-Born approximation (5), and a scaling factor given by $C$, where as shown above (3) is a function of the number of electrons in the target subshells. The parameters $D$ and $2\kappa$ are fitting parameters, enumerated by Zambra et al. (1994) as $4.7 \times 10^{-15}$ cm$^2$ eV$^2 \pm 25\%$ and 2.5, respectively, for the purpose of modeling the singly charged light ions in that study.

For the purposes of the present study, one fitting parameter was not sufficient. Indeed, one focus of this study was to provide a general report on the values of $D$ for various ions. To provide information on the fitting value $D$ for a number of ions, we modified the model of Zambra et al. (1994). The model incorporated a set of experimental data on each ion, notably the cross section maximum and associated energy. The model then calculated, with a given $C$, $I_2$, and $x_{\text{max}}$ the value for $\kappa$ to make the value of the cross section a local maximum at that energy. Then, with a known $\sigma_{\text{max}}$ we fit $D$ to the experimental data. Combining the fits for $\kappa$ and $D$, the model would match exactly the maximum of the cross section plot of each experiment. While this method provides good agreement with the data, the obvious drawback is lack of predictability with new ions or ones without experimental data with which to fit the model.

2.2 Belenger et al. (1997)

In a study surveying a variety of neutral atoms, positive ions in a variety of charge states, and negative ions, Belenger et al. (1997) gave a model for electron impact DI:

$$\sigma_n = \frac{D C_n^l}{I_2^l x} \left(1 - \frac{1}{x}\right)^{2\kappa}.$$  \hspace{1cm} (6)

where $u$ is a modified version of a threshold unit with $E$ as incident electron energy. This model incorporates features that make it applicable to ionization of $n$ electrons ($n \geq 3$), but has been applied to cases of $n = 2$ (Belenger et al. 1997). In this study, $n = 2$, and the parameters $a$ and $b$ are given by Belenger et al. (1997) as 14.0 and 1.08 respectively. The values for $I_2$, the DI energy; $N$, the number of total electrons in the ion; and $c$, the value 1.0 for neutral atoms and 0.75 for ions vary from case to case. The model incorporates the Bethe-Born approximation (5), the classical Thomson scaling law (2), fitting parameters derived empirically ($a$ and $b$), and a scaling factor based on the number of electrons in the ion. A major strength of this model from a theoretical point of view is its predictive power. No inputs are necessary from experimental data, and the model is applicable to a wide variety of ions and to multiple ionization.
2.3 Fisher et al. (1995)

In a study surveying a variety of ions over a variety of multiple ionizations, Fisher et al. (1995) gave two models for electron impact DI, each valid for separate ion states. For ions without any M-shell electrons, to be referenced hereafter as scaling law 1,

\[
\sigma = \frac{4\pi a_0^2}{17^{m-1}} \left( \frac{I_H}{I_2} \right)^2 \zeta^d \ln x \left( 1 - 2e^{-(\ln 2)x} \right) .
\]  

(8)

For ions with M-shell electrons, to be referenced hereafter as scaling law 2,

\[
\sigma = \frac{1.6 \pi a_0^2}{17^{m-1}} \left( \frac{I_H}{I_2} \right)^2 \zeta^t \ln x \left( 1 - 2e^{-(\ln 2)x} \right) .
\]  

(9)

In both models, \(a_0\) is the Bohr radius, \(I_H\) is the ionization energy of hydrogen, \(m\) is the number of electrons to be ionized, and \(\zeta\) is a scaling factor.

In the study conducted by Fisher et al. (1995), the authors used a complicated formula for determining the values for \(\zeta\). The model differentiated energy regimes for the two scaling laws, noting that above a certain threshold, multi-step processes become more significant in calculations of DI cross sections. Scaling law 1 is used below that threshold, with scaling factor \(\zeta^d\). Above the threshold, scaling law 2 gives a better prediction with scaling factor \(\zeta = \zeta^d + \zeta^t\), where \(\zeta^t\) gives the scaling factor for all of the multi-step processes in an ion with more electrons and therefore more highly probable multi-step processes. Fisher et al. (1995) demonstrated the scaling factor depends on target electrons to the general case of \(m\) ejected electrons from a shell with \(\xi_0\) electrons:

\[
\zeta^d = \frac{\xi_0!}{m!(\xi_0 - m)!} \quad \xi_0 \geq m .
\]  

(10)

In particular, if the number of electrons to be ejected is greater than the number of electrons in that shell, \(m > \xi_0\), it is shown that the scaling factor \(\zeta^d\) incorporates the ejection of \(\xi_0\) electrons from the outer shell and \(m' = (m - \xi_0)\) electrons from the next outermost shell:

\[
\zeta^d = \frac{\xi_s!}{m'!(\xi_s - m')!} \quad \xi_0 < m .
\]  

(11)

The authors found \(\zeta^t\) in a similar way, but with stipulations that were beyond the scope of the present study. Instead, the present study considers both scaling laws—over the energy ranges
given by each experiment—with scaling factors as calculated by (10, 11). In addition, the present study uses subshells such as 2p rather than whole shells such as \( n = 2 \) in calculating the scaling factors. The authors of this study felt that since often the same shell can have subshells with significantly different ionization potentials, it was more physically accurate to consider electron pairs from defined subshells. The factor of \( 4/17^n \) and \( 1.6/17^n \), derived empirically by Fisher et al. (1995) to give a further scale, gave the authors of the present study confidence in the ability to redefine the scaling factor. Future studies could modify that parameter, while the present study attempts to verify the applicability of each model, albeit using modified parameters.

With respect to the other models and previous attempts, the models given by Fisher et al. (1995) incorporate the Bethe-Born approximation at large energies (5), the classical Thomson scaling law (2), fitting parameters derived empirically, and a scaling law that takes into account the numbers of electrons in the active subshell(s).

3. Analysis

Numerous experiments of DI cross sections on a variety of atoms and ions allowed testability of the various models in this study. Of particular interest were noble gas ions of varying ionicities and an assortment of different ionicities of other ions to test the validity of the models over all possible cases. Neutral atoms, singly charged ions, highly charged ions, and even negative ions provided test cases for the models in this study. Not only did each set of experimental data test our models for accuracy in cross section, they also tested the models over a wide energy range. In this study, we focused mainly on the direct processes for DI, ignoring when possible the obvious indirect processes measured by experiment in calculating the models.

The inputs for the models involved several parameters. For the models in this study, the parameters were the double ionization energy of a given two electrons, the occupation numbers of the subshells containing those electrons, the charge of the ion in question, and the ion’s atomic number. In addition, the model by Zambra et al. (1994) took in experimental data to produce different \( D \) values for each case. The Los Alamos Atomic Structure Codes (CATS) (Abdallah et al. 1988) gave energies for the initial and final configurations, allowing calculation of double ionization energies for the pair of electrons in the same way single ionization energies are calculated. The energies listed do not take into account any configuration interaction, as they are configuration averages; the purpose of this study was more a general examination of the models available and utmost accuracy was not of highest concern. The final input was impact electron energy, which was calculated using the same 21 threshold energies as used in GIPPER (Archer et al. 2002). Our models plotted each of these 21 energies in threshold units, from 1.01 to 1000.0.

The input parameters listed above were enough to produce the models by Belenger et al. (1997) and Zambra et al. (1994), but not yet sufficient to reproduce the model by Fisher et al.
(1995). Each of the former only considered the most likely process of DI within an atom, ionizing the most loosely bound pair of electrons. By using the models by Fisher et al. (1995) with different pairs of electrons that could be doubly ionized at energies slightly higher than the most likely pair, the results showed increased accuracy to experiment. For example, with the ion \( \text{Kr}^{14+} \), the electron configuration is given by [Ar] 3d\(^4\), where [Ar] represents the electronic structure of neutral argon. The models in this study take inputs for ionizing 2 3d electrons directly, leaving \( \text{Kr}^{16+} \) with [Ar] 3d\(^2\). This process has the lowest DI threshold. Ionizing a 3d and a 2p electron, however, takes only slightly more energy and provides a significant contribution to the total DI cross section at higher energies. For this reason, this study incorporated the three lowest energy DI processes into the Fisher models, which calculated the sum of each process at each impact electron energy. Below the threshold for a particular reaction, the cross section for the reaction was approximately zero and contributed a negligible amount to the total cross section. Above the threshold, the cross section can contribute quite significantly, as will be shown below for the different ions and atoms.

### 3.1 Neutral Helium

Neutral helium, with an atomic structure of \( 1s^2 \), is the simplest of all targets for DI. Each model only considers one pair of electrons to be ionized, and direct processes of ionization dominate the cross section. Figure 1 shows the total cross section of electron impact ionization of neutral helium as compared with the various models in this study.

All of the models in this study reproduce the experimental data with some error, but the maximum deviation is only by a factor of 2. The model by Belenger et al. (1997) poorly reproduces the data of the experiment throughout the energy range, and the model by Zambra et al. (1994) fits much of the data to within experimental error up to and including the cross section peak. At higher energies, the model deviates from experiment more significantly. The models by Fisher et al. (1995) underestimate the cross section on the whole, but the first scaling law reproduces the data accurately throughout the energy range. The second scaling law does not reproduce the data as accurately, to be expected for an ion without M-shell electrons. A notable coincidence is the locations in energy of the cross section maximum and the maximum predictions of the models by Belenger et al. (1997) and Fisher et al. (1995). Since the model by Zambra et al. (1994) takes in the experimental data to fit the model, it should reproduce the data most accurately; that the model overestimates the cross section at higher energies is significant.
3.2. Singly Ionized Light Ions

The experiments of Zambra et al. (1994) on C⁺, N⁺, O⁺, and Ne⁺ provided a good test of the models against singly-charged systems for different atomic number of possible interest. The range of atomic numbers below neon, for instance, could be used to approximate different ion stages of neon or aluminum, if there exists a relationship between isoelectronic structure and DI cross section. Figure 2 shows the total cross section of electron impact ionization of four singly ionized ions as compared with the various models in this study. Table 1 shows the configurations of these ions and the ionization energies of different pairs of electrons.

As the table shows, the ionization of the most loosely bound electrons is not the only process probable over the energy ranges of the experiments. The addition of two other pairs has the effect of increasing the cross sections predicted by the models of Fisher et al. (1995). The scaling factors are also different for each pair as well, with the scaling factor for the (2p, 2s) ionization of nitrogen being 4 rather than 1 for the (2p, 2p) pair, thereby making the cross section addition to the total cross section more significant. The models of Belenger et al. (1997) and
Table 1. Ionization energies of singly charged ions. Energies from CATS (Abdallah et al. 1988).

<table>
<thead>
<tr>
<th>Ion</th>
<th>Configuration</th>
<th>Ejection Pair</th>
<th>DI Energy (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C⁺</td>
<td>1s²2s²2p¹</td>
<td>(2p, 2s)</td>
<td>71.577</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2s, 2s)</td>
<td>79.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2p, 1s)</td>
<td>365.105</td>
</tr>
<tr>
<td>N⁺</td>
<td>1s²2s²2p²</td>
<td>(2p, 2p)</td>
<td>77.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2p, 2s)</td>
<td>84.405</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2s, 2s)</td>
<td>97.86</td>
</tr>
<tr>
<td>O⁺</td>
<td>1s²2s²2p¹</td>
<td>(2p, 2p)</td>
<td>90.315</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2p, 2s)</td>
<td>97.732</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2s, 2s)</td>
<td>117.708</td>
</tr>
<tr>
<td>Ne⁺</td>
<td>1s²2s²2p³</td>
<td>(2p, 2p)</td>
<td>103.682</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2p, 2s)</td>
<td>126.901</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2s, 2s)</td>
<td>165.215</td>
</tr>
</tbody>
</table>
Zambra et al. (1994) are unaffected by these additions, as they are not incorporated in the models.

A notable feature of the C$^+$ graph is the shoulder located at ~320 eV. Zambra et al. (1994) notes that the ionization-autoionization threshold is located at 323 eV, nearly coinciding with the direct cross section maximum. As this study is mainly concerned with the direct DI processes, we eliminated the points above the shoulder from consideration for the model by Zambra et al. (1994) so as not to skew the model. This study did not separate the direct and indirect process energy regimes in all cases as will be seen, but in this case the picture proves significant, as the models of Zambra et al. (1994) and Fisher et al. (1995) reproduce the direct portion of the total cross section accurately.

As for the overall behavior of the models with the different ions, most of the time the models approximated the data to about 30%. The exception is the model of Belenger et al. (1997) for C$^+$, where the model overestimated the data by a factor of 3. The model did quite well in reproducing the energy of the cross section maximum of all the ions, even though the magnitude of the cross section at all energies was less accurate. A better calibration of the scaling factors $a$ and $b$ may improve the prediction. The model of Zambra et al. (1994) approximated the behavior from the threshold to the maximum cross section to almost within experimental error, but over the different ions was less accurate at higher energies, possibly due to the (1-1/x)$^2$ dependence. The model by Fisher et al. (1995) appeared to reproduce the data well at most energies, but underestimated the cross section near the total cross section maximum for N$^+$ and O$^+$, calling into question the empirically derived scaling factors as they apply to the regime for approximately half filled active subshells.

### 3.3. Highly ionized argon ions

Experiments by Zhang et al. (2002) on argon ions Ar$q^+$ ($q = 5$-$11$) provide a test of the models against highly charged ions of noble gases, often used as quenchers in magnetic fusion plasmas. Figure 3 shows the total cross section of electron impact ionization of Ar$q^+$ ($q = 7$-$10$) as compared with the various models in this study. Table 2 shows the configurations of these ions and the ionization energies of different pairs of electrons. The configurations and electron pairs show most of the ionizations containing an electron from the filled or nearly filled 2p subshell. In addition, it is notable that the best agreement for all the graphs comes from Ar$^7^+$, which has the lowest DI threshold because of a single 3s electron.

The models on the whole do quite a poor job reproducing the experimental data. The model by Zambra et al. (1994) seems to be the most accurate for Ar$^7^+$ and Ar$^8^+$ but does not model close to the behavior of the experimental data for Ar$^9^+$ or Ar$^{10^+}$. Eliminating the energy range where it appears that indirect processes dominate would possibly help this model, but the data, the experimenters note, does not show precisely the thresholds for different processes.
Table 2. Ionization energies of highly ionized argon ions. Energies from CATS (Abdallah et al. 1988).

<table>
<thead>
<tr>
<th>Ion</th>
<th>Configuration</th>
<th>Ejection Pair</th>
<th>DI Energy (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Ar}^{7+}$</td>
<td>$1s^22s^22p^6$</td>
<td>$(3s, 2p)$</td>
<td>565.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(3s, 2s)$</td>
<td>643.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(2p, 2p)$</td>
<td>844.86</td>
</tr>
<tr>
<td>$\text{Ar}^{8+}$</td>
<td>$1s^22s^22p^6$</td>
<td>$(2p, 2p)$</td>
<td>901.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(2p, 2s)$</td>
<td>968.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(2s, 2s)$</td>
<td>1056.16</td>
</tr>
<tr>
<td>$\text{Ar}^{9+}$</td>
<td>$1s^22s^22p^6$</td>
<td>$(2p, 2p)$</td>
<td>1019.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(2p, 2s)$</td>
<td>1075.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(2s, 2s)$</td>
<td>1164.19</td>
</tr>
<tr>
<td>$\text{Ar}^{10+}$</td>
<td>$1s^22s^22p^4$</td>
<td>$(2p, 2p)$</td>
<td>1159.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(2p, 2s)$</td>
<td>1186.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$(2s, 2s)$</td>
<td>1276.49</td>
</tr>
</tbody>
</table>
Zhang et al. (2002) note that the precision of the data is poor because the cross section measurement registered in the very low $10^{-21} - 10^{-20}$ cm$^2$ range. For Ar$^{7+}$, Zhang et al. (2002) note that the most loosely bound pair of electrons, $(3s, 2p)$, does not appear to play any role in the direct DI. The first threshold Zhang et al. (2002) find is nearer to the energy of the $(3s, 2s)$ pair. Further, Zhang et al. (2002) note the discrepancy between their results and the previous experimental efforts. While the authors of this study expected to see a large indirect process contribution to the total cross section at high energies such as in highly ionized neon or krypton, none was present.

All of the models appear to overestimate the energy of the cross section maximum, as well as overestimating the maximum cross section value by a factor of up to 5 in the case of Fisher et al. (1995) scaling law 1 for Ar$^{8+}$ and Ar$^{9+}$ and up to about 4 for the case of Fisher et al. (1995) scaling law 1 for Ar$^{10+}$. Scaling law 2 does better in general but still overestimates the cross section significantly. The present study modified the models of Fisher et al. (1995) as described above in section 2.3, and therefore the data might be better reproduced with a recalibration of the empirically derived scaling factors. The model of Belenger et al. (1997) appears to approximate the data the best, especially for regions near the threshold for DI. The drawback is in the lesser accuracy for the energy of the cross section maximum, where the model was poorest of the group. The model of Zambra et al. (1994) appears to do the best, where if the fitting parameter was fixed for the isonuclear charge sequence, the approximations would probably still reproduce the data more accurately than the other models.

### 3.4. Highly ionized krypton ions

Experiments of Khouilid et al. (2001) provide a test of highly charged ions of greater atomic number. Krypton could become this highly charged from acting as a quencher in a magnetic fusion plasma, therefore it is of practical importance to approximate the double ionization cross section of such ions under high energy electron impact. Figure 4 shows the total cross section of electron impact DI of Kr$^{q+}$ ($q = 14-17$) as compared with the various models in this study. Table 3 shows the configurations of these ions and ionization energies of different pairs of electrons.

Highly ionized krypton in the range of Kr$^{q+}$ ($q = 14-17$) has the atomic structure of argon with more electrons bound in the $3d$ orbital, rather than filling up the $3s$ and $3p$ orbitals. In all of the ion stages studied, Khouilid et al. (2001) note a dominance of indirect processes. In Kr$^{14+}$, for example, the theoretical ionization-autoionization (IA) threshold occurs at 2142 eV, at higher energies than the direct part—the focus of this study—but low enough energy to dominate the cross section over the energy range considered (1000-6000 eV). This experimental fact limits the range of interest from about 1000-2000 eV, where the direct DI dominates. At higher
Figure 4. Electron impact DI cross sections of Kr\textsuperscript{q+} (q = 14-17).

Table 3. Ionization energies of highly ionized krypton ions. Energies from CATS (Abdallah et al. 1988).

<table>
<thead>
<tr>
<th>Ion</th>
<th>Configuration</th>
<th>Ejection Pair</th>
<th>DI Energy (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kr\textsuperscript{14+}</td>
<td>[Ne]2s\textsuperscript{2}2p\textsuperscript{3}3d\textsuperscript{1}</td>
<td>(3d, 3d)</td>
<td>1032.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3d, 2p)</td>
<td>1125.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3d, 2s)</td>
<td>1212.45</td>
</tr>
<tr>
<td>Kr\textsuperscript{15+}</td>
<td>[Ne]2s\textsuperscript{2}2p\textsuperscript{3}3d\textsuperscript{2}</td>
<td>(3d, 3d)</td>
<td>1132.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3d, 2p)</td>
<td>1221.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3d, 2s)</td>
<td>1310.18</td>
</tr>
<tr>
<td>Kr\textsuperscript{16+}</td>
<td>[Ne]2s\textsuperscript{2}2p\textsuperscript{3}3d\textsuperscript{3}</td>
<td>(3d, 3d)</td>
<td>1232.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3d, 2p)</td>
<td>1249.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3d, 2s)</td>
<td>1411.06</td>
</tr>
<tr>
<td>Kr\textsuperscript{17+}</td>
<td>[Ne]2s\textsuperscript{2}2p\textsuperscript{3}3d\textsuperscript{4}</td>
<td>(3d, 2p)</td>
<td>1426.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3d, 2s)</td>
<td>1511.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2p, 2p)</td>
<td>1513.00</td>
</tr>
</tbody>
</table>
energies, Khouilid et al. (2001) note that the indirect processes can dominate by contributing up to 90% of the DI cross section.

In the range of energy for direct DI, the models reproduce the experimental data to a high degree of accuracy. The model of Belenger et al. (1997) does well at lower energies, but tends to overestimate the cross section nearer to the IA thresholds. Also, the model gets more accurate as the charge state increases or, in other words, when the number of target electrons decreases. As it is impossible to see where on the energy scale the direct DI cross section would be at a maximum, no evaluation of the models is possible for that parameter. The model by Zambra et al. (1994) takes in the experimental data as a parameter, and as the behavior of the direct DI cross section is unknown at energies beyond the IA thresholds, the model scales to the maximum. An evaluation of the model shows good agreement to the indirect processes, proving possible application of the model to other cross section processes outside the scope of this study.

The models of Fisher et al. (1995) provide an interesting demonstration of the necessity of including multiple ionization processes in the same plot of total cross section. With Kr$^{14+}$ and Kr$^{15+}$, where ordinarily only 2 3d electrons would be considered, the inclusion of the other pairs listed in table 3 is necessary to reproduce the experimental data. Law 2 (9) gives better agreement than law 1 (8) for both cases considering all active subshells 3d, 2p, and 2s. Law 2 (9) should give better agreement as stated by Fisher et al. (1995), since Kr$^{14+}$ and Kr$^{15+}$ contain M-shell electrons. Here the difference clearly shows by inclusion of all active subshells. Without such inclusion, law 1 (8) would overestimate the contribution of the one process for the ion and therefore be more accurate as compared with experiment. Rather, law 2 (9) clearly dominates accurate reproduction of experimental data in all cases of krypton ions studied here after inclusion of all active subshells.

3.5. Negative Ions

Although DI of certain ions and atoms are not of direct interest for application in plasma research, the inclusion of a variety of atoms and ions provides an opportunity to test the applicability of various models across a wide range of cases. The experiments of Belenger et al. (1999) provide an extension into negative ions. Figure 5 shows the total cross section of electron impact DI of C$^-$ and O$^-$ as compared with the various models in this study.

The models do not reproduce the experimental data accurately for these negative ions. Near the threshold energy for C$^-$, the model of Belenger et al. (1997) recreates the experimental data the best, while the models of Fisher et al. (1995) underestimate the values. Near the energy of the cross section maximum, all models except for that of Zambra et al. (1994) underestimate the cross section by 60-80%. The location of energy for the maximum cross section seems to be reproduced well by all models. For O$, the energy of the cross section maximum is predicted at least a factor of 2 lower for all models except Zambra et al. (1994). Law 2 of Fisher et al. (1995)
seems to approximate the data best at lower energies, with law 1 and the model of Belenger et al. (1997) overestimating the cross section at low energies. All models except Zambra et al. (1994) underestimate the cross section at high energies. The model of Zambra et al. (1994) clearly does the best reproduction of data over all energies. A fixed value of $D$, the fitting parameter, would probably still give good agreement with experiment and may allow accurate predictions for other negative ions.

The experimenters Belenger et al. (1999) note several characteristics of their experiments that help explain the low predictions of the models. The authors note that both carbon and oxygen have autoionization states with configurations $1s^22s2p^3$ and $1s^22s2p^5$, respectively. The states arise when an electron ionizes a $2s$ electron, occurring only a few eV above the DDI thresholds for each ion. The authors acknowledge the impossibility of discerning the relative contributions of direct and indirect DDI past the threshold, effectively over their entire experiment (Belenger et al. 1999). The inability to discern between processes makes it difficult to draw conclusions on the strengths of the models studied here, except to say that to provide more accurate reproduction of experimental data, models should incorporate not only direct but indirect processes as well.

### 3.6. Overall Results

The model of Zambra et al. (1994) incorporated a scaling factor that allowed examination of different ions over a rigid model of energy dependence for ionization cross section. The values for $D$ in (6) should give insight on trends of cross section maximum over a variety of ions. Table 4 shows the empirically derived values from this study.

The $D$-values show definite trends within ionization regimes. For singly ionized cases and for helium, the values for $D$ appear to be within the same order of magnitude, near to $D \sim 1.3 \times 10^{-14}$ cm$^2$eV$^2$. This suggests that neutral and singly ionized ions with lower atomic number, that is, lower charge in the nucleus, could share similar DI cross sections. For the highly ionized argon, Ar$^{7+}$ ($q = 7-11$), all of the values are in quite good agreement, with $D \sim 5.3 \times 10^{-15}$ cm$^2$eV$^2$. 
Table 4. D-values for ions considered in this study. Experimental data from Ne\(^{5+}\) and Ne\(^{6+}\) from Duponchelle et al. (1997).

<table>
<thead>
<tr>
<th>Ion</th>
<th>D – value ( (\text{cm}^2\text{eV}^2) )</th>
<th>Ion</th>
<th>D – value ( (\text{cm}^2\text{eV}^2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>He</td>
<td>7.29E-15</td>
<td>Ar(^{5+})</td>
<td>5.36E-13</td>
</tr>
<tr>
<td>C(^{+})</td>
<td>5.92E-15</td>
<td>Ar(^{7+})</td>
<td>6.97E-15</td>
</tr>
<tr>
<td>N(^{+})</td>
<td>2.60E-14</td>
<td>Ar(^{8+})</td>
<td>1.44E-15</td>
</tr>
<tr>
<td>O(^{+})</td>
<td>1.74E-14</td>
<td>Ar(^{8+})</td>
<td>3.76E-15</td>
</tr>
<tr>
<td>Ne(^{+})</td>
<td>9.22E-15</td>
<td>Ar(^{10+})</td>
<td>5.75E-15</td>
</tr>
<tr>
<td>C(^{-})</td>
<td>5.65E-14</td>
<td>Ar(^{11+})</td>
<td>8.64E-15</td>
</tr>
<tr>
<td>O(^{-})</td>
<td>2.45E-14</td>
<td>Kr(^{14+})</td>
<td>5.60E-14</td>
</tr>
<tr>
<td>Ne(^{5+})</td>
<td>4.38E-14</td>
<td>Kr(^{15+})</td>
<td>1.80E-13</td>
</tr>
<tr>
<td>Ne(^{6+})</td>
<td>1.04E-13</td>
<td>Kr(^{16+})</td>
<td>4.73E-13</td>
</tr>
<tr>
<td>Ar(^{5+})</td>
<td>5.67E-13</td>
<td>Kr(^{17+})</td>
<td>8.25E-14</td>
</tr>
</tbody>
</table>

Throughout an isonuclear sequence, then, the fitting parameter \( D \) might be expected to stay within the same order of magnitude, just as it does with highly ionized krypton.

The model by Belenger et al. (1997) was made to be as general as possible, handling multiple ionizations and incorporating much of the previous theoretical work on double ionization. Because of its generality, however, most of the reproductions of experimental data were fairly accurate but many were only within a factor of 2 and in some cases could not predict the energy region of the cross section maximum with acceptable accuracy. For all ions considered, the model did very well near the threshold for double ionization, but often overestimated the cross sections nearer to the maximum cross section, as in neutral helium. In the high energy region, the model also tended to overestimate the DI cross section. For the ions considered, the model by Belenger et al. (1997) provided considerable accuracy in predicting the direct DI cross section for the krypton isonuclear sequence. For the singly ionized atoms, the predictions varied somewhat inconsistently near the maximum cross section, possibly due to calibration of the scaling factors.

The models by Fisher et al. (1995) appeared to be the most accurate under the widest variety of ions. For atoms without M-shell electrons, law 1 (8) reproduced experimental data more accurately than law 2 (9), as expected. Especially notable is the predictions for atoms without significant indirect process contributions such as helium and the singly ionized cases. In the case of krypton ions, however, law 2 (9) reproduced the data below the IA threshold more accurately, as expected. Inclusion of multiple electron pairs proved notable in that case. For the negative ions, both models were unable to reproduce experimental results, however. As the
experimenters noted, the direct and indirect processes were effectively impossible to distinguish, so meaningful evaluation is difficult (Belenger et al. 1999).

4. Conclusion

In this paper, the authors tested several semi-empirical models for double ionization over a variety of ions and a wide range of energies. The lack of solid, theoretically accurate models without empirically derived parameters makes reproduction of experimental data difficult for certain situations. Nevertheless, we tested three semi-empirical models on experimental data obtained from a search of the available data. The model by Zambra et al. (1994) included a fit parameter from scanning the experimental data and fitting the maximum predicted cross section to the experimental cross section. This led to very accurate predictions and possible application to those cases where general models fail, such as with negative ions or when there is otherwise a lack of prediction for indirect processes. An obvious drawback is the necessary inclusion of a large assortment of $D$-values to fit the model to more general cases. The more general models proved effective, with the models of Fisher et al. (1995) proving more accurate over a variety of cases.

The present study provided an analysis of possible models of double ionization cross sections to eventually be placed in the GIPPER code (Archer et al. 2002), eventually providing input information for plasma kinetics modeling. Electron impact DI is a process that occurs in certain plasmas when an impact electron has enough energy to doubly ionize an ion, but the principle of detailed balance demands that the reverse must be considered as well, where three electrons impact an ion and two recombine. Further research into the DI process will involve study of this reverse process of DI, four-body recombination.

In plasma kinetics modeling, solutions of the collisional rate matrix help determine the population of ions at a given temperature and density. The collisional rate matrix in plasma kinetics modeling incorporates the principle of detailed balance in all processes to ensure accurate ion populations. DI and four-body recombination improve the plasma kinetics modeling by refining the collisional rate matrix. The valuable information on plasma spectroscopy contained in the more accurate ion population data that comes out of the plasma kinetics modeling may help physicists in all areas of plasma research.

Acknowledgements

Many thanks to the National Science Foundation and the University of New Mexico for making the Research Experience for Undergraduates program possible. Thanks to Norm Magee, Lee Collins, James Colgan, and Sally Seidel for making the Los Alamos Summer School such a shining success. It is a pleasure to acknowledge Roberto Mancini for fruitful discussions and
Chris Fontes for collaboration on integrating double ionization into GIPPER. I would like to express a deep personal thanks to Norm Magee for his role as provider and mentor outside the lab. His energy and enthusiasm made the Summer School especially memorable.

References

Archer B J, Clark R E H, Fontes C J and Zhang H L 2002 Los Alamos Memorandum, LA-UR-02-1526
Mancini, Roberto 2006 (private communication)
Wannier G H 1955 Phys. Rev. 100 1180
Are there drift waves in the Flowing Magnetized Plasma (FMP) experiment?

Plasmas are the fourth state of matter. In contrast to solid, liquid and gas, plasmas consist of freely moving charged particles, electrons and ions that, just like water and air, can support waves. Furthermore, due to the charged particle motion coupling to electric and magnetic fields, waves in plasmas are much richer and complex.

The Flowing Magnetized Plasma (FMP) experiment at the Plasma Physics Group (P-24) generates a multi-millisecond, rotating plasma and much data describing this plasma has accumulated, but not fully understood. In particular, internal magnetic fields, floating potential, density and electron temperature show low frequency (< ion gyrofrequency) oscillations (waves) whose mechanism and origin have not yet been identified. IDL software is used to analyze experimental data, primarily by Fourier (spectral) analysis. The results are compared to the simplest linear theory of drift waves. However, more analysis must be done to rule out any other (non-drift wave) mechanisms of the oscillations.

Introduction

One of the main thrusts in plasma research today, the reason why the International Thermonuclear Experimental Reactor (ITER), a 10 billion Euro international experiment, is under construction, is fusion energy. And one of the biggest problems in making fusion reactors a reality is confinement, that is keeping plasma particles inside their non-material (magnetic) container for a sufficient amount of time for them to fuse. In addition, the particles are very hot (energetic) so the material container can be damaged if the particles are allowed to contact the walls at large rates. The problem has been a bit of a surprise [1], because one “classically” expects a much lower rate of particle escape. In particular, particles in a magnetic field \( B \) gyrate about a center, called a guiding center, due to the Lorentz force, which is perpendicular to a particle’s velocity and \( B \) and, thus, always bends the particle’s trajectory. This radius of gyration, or Larmor radius, is quite small for ions of relatively low energy in a strong \( B \), and smaller yet for electrons. The Lorentz force doesn’t affect particle movement along the magnetic field lines but forces gyrations in the perpendicular plane. The resulting trajectory is a tight spiral about a field line of \( B \); particles are “glued” to the lines. Collisions can knock particles from one line to another and this diffusion was predicted to be the only mechanism for particle loss. Actual rates of particle escape far exceeded the diffusion prediction by orders of magnitude and were termed “anomalous transport” [2]. Because of the low cross-sections of particles to fuse,
it’s desirable to keep particles contained and colliding as often as possible and so a flurry of research focused upon solving the anomalous transport problem. Up to the 1960’s, plasma theory made the assumption of constant density. Of course, any plasma in a container would have its density approach zero at the edge, or the plasma would not be confined! The addition of a density gradient to theory gave rise to a new type of instability: drift waves. And since there is always a density gradient, this instability is also termed *universal instability*. (However, non-linear effects can make the instability growth negligible or level out at some amplitude. Otherwise, all confined plasmas would have this as the dominating instability, which is not the case). Reviews of drift wave research progress have been compiled by [3], [4], [5]. The FMP plasma exhibits low-frequency oscillations. Drift waves have been extensively studied and identified in many experiments (see reviews above) and they, too, have low frequency oscillations. It is the goal of this paper to examine drift waves as a candidate for the observation.

In the rest of the paper, we present, first, a review of the discrete Fourier transform (DFT) and a description of resistive drift waves (linear theory). Next, a description of the FMP apparatus is followed by some results and analysis. The paper ends with a concluding paragraph.

Theory

1. DFT

The one-dimensional Fourier transform is the well-known analysis tool used to transfer data in time domain into frequency space. The usual formula of the transform of the continuous signal \( f(t) \) is

\[
F(\omega) = \int_{-\infty}^{\infty} f(t) \exp(-i\omega t) \, dt \quad (1)
\]

where \( F(\omega) \) is the amplitude of a mode of angular frequency \( \omega \) and is defined for all \( \omega \). One regains the function \( f(t) \) by the inverse Fourier transform

\[
f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) \exp(i\omega t) \, d\omega \quad (2)
\]

Laboratory measurements are usually a discrete sequence of samples of \( f, \{ f(nT) \} \), where \( n \) runs from 0 to \( N \), the total number of samples taken and \( T \) is the time between samples, known as sampling time. The adaptation of the Fourier transform (1) to such a discrete set of values is known as the Discrete Fourier Transform, DFT, and is given by

\[
F_D(\omega_k) = \sum_{n=0}^{N-1} f(nT) \exp(-i\omega_k nT) \quad (3)
\]

where the subscript \( D \) denotes the discrete transform, defined only for the \( N \) values of \( \omega_k \) (see below for which these are). The function \( F_D(\omega_k) \) is closely related to the coefficients \( a_k \) in the Fourier series of \( f(t) \) (periodic in \( L \)),

\[
f(t) = \sum_{k=-\infty}^{\infty} a_k \exp(i\omega_k t/L) \quad .
\]

Thus, we can interpret \( F_D(\omega_k) \) as the “amount” of a certain mode in \( f(t) \) and, indeed, \( |F_D(\omega_k)|^2 \) is the power of the mode. This is why spectral analysis is often referred to as power
analysis, for one can judge from the values of $F_D(\omega_k)$ how power in the signal $f(t)$ is distributed between various modes.

We expect the transform $F_D$ of a pure sine wave of frequency $\omega$, for instance, to have a peak at $\omega$ and be zero everywhere else. But when we enter the discrete, laboratory world, what happens to the discrete transform of a sine wave of angular frequency $\omega$ when $\omega$ is not one of the $\omega_k$ that $F_D$ is defined for? Answer: all other values $F_D(\omega_k)$ will assume non-zero values! The set $\{F_D(\omega_k)\}$ is called the spectrum of $f(nT)$ and the phenomenon just described is termed “spectral leakage,” because the actual frequency of the mode “leaks” (gives non-zero values) into the other, discrete frequencies. Clearly, spectral leakage is undesirable and should be minimized; we want a mode of certain frequency, even if not one of the $\omega_k$, to yield finite $F_D$ values only in those discrete frequencies closest to it on the spectrum. In other words, if we find the DFT of a sine wave of frequency 10.1 Hz but $F_D$’s are defined only for frequencies that are multiples of 10 Hz, we want to see a large peak at 10 Hz and possibly a small one at 11 Hz, but not one at 100 Hz! A solution is to use a weighting function $w(t)$ called a “window” by which we multiply the data $f(nT)$ inside of (3). The transform $F_w(\omega)$ of $f(nT)w(nT)$ is a convolution of the transforms of the two separate functions (given below for the continuous case),

$$F_w(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) W(\omega - x) \exp(-i\omega x) dx \quad (4)$$

where $W$ is the transform of $w(t)$. We can see from (4) that if $W$ is even 0 and vanishes quickly as $|x - \omega_k|$ increases from 0, then the finite components of $F_D(\omega_k)$ that are far from (large) the actual frequency $\omega$ (which we assume we know) will be suppressed by the product with $W$. For use in the DFT, these windows (functions) are sampled in the same way as the signal $f$ and have the following properties:

$$w(nT) = 0 \text{ for } n < 0, n > N$$
$$w(nT) = w([N - n]T), \text{ for } n < N/2$$
$$w(NT) = 0$$

There are many different windows and many ways of measuring their performance, such as equivalent noise bandwidth, side-lobe fall off, etc. [6]. We use the Hanning window because it’s easily implemented and has low side-lobe levels [7].

The lowest frequency that can be recognized with the DFT is $f_{\text{base}} = 1/NT$ and the highest, called the Nyquist frequency is $f_{\text{Ny}} = 1/(2T)$ [6]. The DFT ($F_D$) is defined on the $N$ values of $\omega_k = 2\pi k * f_{\text{base}}$, where $k$ is 0, ..., $N-1$: however the meaningful $F_D$ values are only the first $N/2$ multiples, $k = 0$, ..., $N/2-1$, of the base frequency; the remaining $N/2$ points in the transform are noise. This is because the $N$ signal points are real numbers, whereas $F_D(\omega_k)$ is, in general, complex (i.e. takes 2 real numbers to specify). Thus, the $N/2$ complex amplitudes exhaust the $N$ degrees of freedom [7]. Lastly, the Fast Fourier Transform (FFT) is simply an algorithm for computing the DFT, implemented in most software packages. The words DFT and FFT are used interchangeably.
2. Resistive Drift Waves

Plasmas are usually described in 3 ways: by the trajectories of individual particles – the particle picture, by the evolution of a velocity distribution function – kinetic picture, and by the fluid picture, in which a fluid of electrons is mixed with fluids of ions and neutrals. Each picture has its own advantages and domains of application; we will consider the two-fluid picture, a special case of magneto-hydrodynamics (MHD).

A term that arises in fluid dynamics is the pressure force, \(-\nabla \cdot \overrightarrow{P}\), where \(\overrightarrow{P}\) is the pressure tensor. If we neglect viscosity (off-diagonal elements in the \(\overrightarrow{P}\)), and assume isotropic pressure, the force reduces to \(-\nabla p\) [Chen Propagation].

First, a semantic clarification: instabilities are behaviors in plasma parameters that grow in amplitude, whereas waves are oscillations of constant amplitude some parameters (The growth is indefinite only mathematically. Once the oscillations of some quantity \(G\) become comparable to the unperturbed \(G\), linear theory, used in this paper, breaks down and non-linear effects limit growth.) Waves and oscillations are related: theories mostly use the complex exponential to represent waves and when the angular frequency \(\omega\) has \(\text{Im}(\omega) > 0\),

\[
\exp(i[kx - \omega t]) = \exp(\text{Im}(\omega)t)\exp(i[kx - \text{Re}(\omega)t])
\]

the first factor on the right is a real number that grows exponentially. This means the behavior is unstable. Conversely, if the imaginary part is negative, the behavior is exponentially damped. Because of the relationship, we will use the terms interchangeably.

Drift instabilities/waves arise because of plasma inhomogeneity and come in many flavors. Gradients in density or temperature lead to the variety, each with its own dispersion relation (\(\omega = \omega(k)\)). Indeed, [8] lists 11 different drift instabilities, and that’s only in the electrostatic (\(B\) approximately constant) regime. We concern ourselves with resistive drift waves.

In particular, if there’s a density perturbation, an electric field would cause charge separation, and if the charge separation enhances the perturbed electric field, the instability grows. In such a situation \(\partial E / \partial t \neq 0\), and the finite-Larmor-radius effect [9] sets in, altering \(v_E = E_x B / B^2\) (called \(E_x B\) drift) for the different species (ions and electrons). All these motions are of course related and discussed below.

We will follow Chen’s [9] and Zweben’s [10] presentations. In what follows, quantities labeled with a naught denote equilibrium (unperturbed) quantities, those labeled with a sub-1 are perturbations, having the oscillatory dependence \(\exp(i[k \cdot r - \omega t])\), and subscript-i stands for ions, subscript-e for electrons. The typical geometry for the discussion of such instabilities is a cylindrical plasma with \(B_0\) pointing along the cylinder (\(\hat{z}\) axis and density gradient \(\nabla n_0 = n r\) in the radial direction (See Fig. 1. In Fig. 1(a), the variation of the constant density surface is exaggerated, at least at the onset of the instability. The perturbation is initially very small and grows due to the reinforcing mechanism.)
Drift waves have a small but finite $k_{||}$, that is $k \cdot B_0 \neq 0$, where $k$ is the wavevector and $\lambda = 2\pi / |k|$; this means the waves propagate along the field and the constant density surfaces therefore look like twisted flutes. We now assume $K_B T = \text{constant}$, where $K_B$ is Boltzmann’s constant, so that $\nabla p = \nabla (n_0 KT) = KT \nabla n_0$. This assumption must be closely adhered by analyzing regions of the plasma where $\text{grad} \ n / n \gg \text{grad} \ T / T$ [12]. This is to avoid ambiguity in wave mechanism, since a temperature gradient can also give rise to waves [8].

Since we allowed a finite $k_{||}$, electrons can flow along $B$ and through collisions, establish a thermodynamic equilibrium. This allows us to invoke the Boltzmann relation $n_i / n_0 = e\phi / KT_e$, i.e. a density gradient for electrons creates a potential gradient and, thereby, an $E_1(t)$. We assume that a density gradient has formed within the plasma. The pressure (density) difference forces thee particles into the area of lower density, but the particles evacuate at different speeds because of inertia and collisions, and, thus, create charge separation (and electric field). The collective electric field $E_1(t)$ will be nearly perpendicular to $B_0$. This collective field will produce an $v_1(t) = E_1(t) \times B_0$ drift. Without resistivity, this disturbance simply propagates in the $\hat{y}$ direction (Fig. 1(b)), with oscillations in $\dot{x}$ in (b) ($\hat{\theta}$ and $\hat{r}$ in (a)), but no growth and no radial particle/heat transport. The particle/heat flux formulas [2] are non-zero when there’s a phase shift between the density and potential. This causes $v_1(\phi(t))$ to point radially out after the density wave has already shifted the plasma density in the same direction. Thus, the wave is unstable. The lag of potential behind density arises because of the polarization drift and the finite-Larmor-radius effect, which causes the $E_1(t) \times B$ drift to differ for electrons and ions.
The drift instability is strongly dependent upon a parameter \( \_ \) which is the ratio of particle pressure to magnetic pressure \( \_ << 1 \). One interpretation of \( \_ \) [11] is that it gives the fractional change in an applied field \( \mathbf{B} \) due to diamagnetic particles. In other words, in the limit \( \_ --> 0 \), plasma particles don’t perturb the applied field, which can therefore be taken as constant throughout. This leads to great simplifications in theory and is the case in many laboratory plasmas [2].

3. Identification of Drift Waves

There are several characteristics of drift waves that we should look for in our plasma, as is pointed out by [10], [12] and [13].

1. Oscillations frequencies \( \omega \) of plasma quantities should be much smaller than \( \Omega_i \), the ion gyro-frequency. 2. There must be a finite \( k \| \) and the phase velocity \( v_{\|ph} = \omega / k_\| \) must be \( v_{\|ph} \leq v_{\|ph} \leq v_{\|ph} \) in order for the instability to occur [9, p. 197], where subscript-th means thermal velocity. Phase velocity can be measured by placing two probes along the same magnetic line and using correlation methods [7, 13]. This is also important because \( k \) and \( k_i \) will determine the collisionality regime of the plasma (collisionless or collisional), as a collisional plasma should have \( v_{ei} >> k \| v_e \). 3. We can check to see that the phase of potential \( \phi \) lags behind the density 4. The biggest red flag for identification is the azimuthal phase velocity, since we know that drift waves propagate at about the diamagnetic drift velocity [4]. Here, one must be sure to correct the experimental frequency for the Doppler shift, \( \omega = \omega_{exp} - v_{rot, edge} / r_0 \); in FMP, plasma rotates at about 5km/sec at the edge and the radius is 0.75m, resulting in a correction of a few KHz. 5. The effect of magnetic field, plasma temperature, ion mass and density on mode stabilization. 6. Theory also predicts the dispersion relation \( \omega = \omega(k) \) and thus the growth \( \text{Im}(\omega) \), and both of these can be compared with experiment.

Linear theory for a cylindrical geometry in the collisionless regime can be found in [11], collisional in [10]

Methods

The basics of the FMP setup can be found in [14]. In overview, FMP is a cylindrical vacuum chamber 4.5m in length and 1.5 in diameter with a long coaxial plasma gun on the axis (see Fig. 2 below). This gun ionizes the gas inside when a voltage up to 900V is applied between the inner electrode and its outer wall.
Magnetic field is measured with B-dot probes, density with a triple probe, ion and plasma velocity with a Mach probe. An built-in IDL Hanning window [7] was used for Fourier analysis. Oscillating plasma potential $\phi$ is estimated with the plasma floating potential $\phi_F$ and density oscillations are obtained from the ion saturation current and plasma electron temperature.

Results

Table 1 below shows an approximate calculation of typical FMP frequencies for $B = 100$G, ionization state of 1 for ions ($n_e = n_i$), and density $n \sim 10^{19}$ particles/m$^3$.

<table>
<thead>
<tr>
<th>Phenomenon</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electron gyrofrequency</td>
<td>280 MHz</td>
</tr>
<tr>
<td>Ion gyrofrequency</td>
<td>200 KHz</td>
</tr>
<tr>
<td>Electron plasma frequency</td>
<td>90 GHz</td>
</tr>
<tr>
<td>Ion plasma frequency</td>
<td>300 MHz</td>
</tr>
</tbody>
</table>

Table 1. Typical FMP frequencies

An example of the $B_t$ FFT spectrum, for FMP plasma, is shown in Fig. 3 below.
We see from this spectrum several things. First, several modes exist on the order of 2-15 KHz. Thus, the frequencies are \(<<\) ion gyration frequency and therefore satisfy that characteristic of drift waves. Secondly, the fact that several modes exist simultaneously and implies there is a likelihood of mode-mode interaction, which requires a more sophisticated analysis. Thirdly, the very fact that B_z oscillates means that the standard assumption of a static, ambient B_z is not true, begging for a more sophisticated theory [6]. In other words, we cannot make the low-beta approximation in the FMP plasma. Disturbance in the \(z\) direction indicates that there’s a finite \(k_\parallel\), which is a requirement for drift waves.

By partitioning the signal into several time intervals of equal length (“windows”) and performing FFT on each window separately, we can trace the evolution of a frequency of interest. If we take the average of a parameter (plasma density or total plasma current, for instance) over the time period of each window, we can plot the amplitude of this frequency as a function of the average parameter, getting as many points on this graph as there are windows. This method can be used to identify whether a certain parameter suppresses or promotes a frequency. For instance, in Fig. 4(a) below, we plot the spectrum of each window, one above another, arranged in increasing value of current (average value for that window); in Fig. 4(b) we plot the amplitude of frequency 7.6 KHz vs current.
Fig. 4. Evolution of mode with frequency 7.6KHz. In (a), the whole spectrum for a “window” is plotted and (b) follows the progress of just one frequency vs gun current.

Conclusion

We present here the first two drift wave characteristics can be identified in the FMP plasma. The frequency of oscillations is in the proper range (a few KHz). Secondly, there exists a finite wave number in the axial direction as can be seen by oscillations in this direction. Furthermore, the machinery (and data) to analyze phase velocities and wavevectors in the all three directions exists and should be analyzed, and compared with appropriate theory. The difficulty is to find such an appropriate theory since the sizable $B_z$ oscillations constitute major complications that most theories choose not to address, in part because many laboratory plasmas do have a low-beta regime. If drift waves cannot be positively identified, there are other candidates, such as Alfvén waves, which also have low frequency oscillations, [9, p. 124] that would then be considered.

Acknowledgments

I’d like to thank Jeff Wang for the long conversations leading to much of the material above.

References


Identification of a lethality core in the metabolic network of

Escherichia coli

Bade Uzgil$^{1,2}$

1 Department of Physics and Astronomy, University of California, Los Angeles
2 Center for Nonlinear Studies, Los Alamos National Laboratory

Prediction of the functionality of genes in metabolic networks is an outstanding puzzle in modern biology that is well suited for study in the inter-disciplinary arena of systems biology research. Techniques that utilize linear and quadratic programming, namely Flux Balance Analysis and Minimization of Metabolic Adjustment, to calculate growth rates in simple organisms like Escherichia coli have proven useful in identifying genes that alter the viability of the cell. After constructing an in silico model of E. coli and applying various internal or external constraints (i.e. genetic mutations or environmental pressures), cellular behavior has been studied in comparison to experiments. Here we investigate genetic mutations that are lethal in various media and construct a core metabolism based on these essential genes. The identification of the lethality core has potential applications to the development of advanced antibiotics designed to target enzyme production in specific metabolic pathways crucial to cell survival across a range of environments.

1. Introduction

Open questions in modern biology often lend themselves to be addressed in arenas that blur the demarcations between life, physical, mathematical, and computational sciences. An example of such a longstanding puzzle that may be better clarified as it traverses freely between these seemingly disparate fields is the prediction of the functionality of genes. This problem has received much attention since progress in genetic sequencing has led to the increasingly complete annotation of the genomes for organisms of varying complexity$^1$, which has subsequently paved the way for analyzing the role genes play within larger biological systems like the network of biochemical reactions in metabolism.

In this study, we examine gene functionality in the perturbed metabolic network of the single-celled, prokaryotic bacterium E. coli. Point deletions, or gene knockouts (KOs), are a class of mutations we apply that are known to be potentially lethal to a cell in a specific medium, e.g. dapA in rich Luria-Bertani media$^2$, if, for instance, network plasticity is not sufficient to cope with the induced disturbance. Further qualities that may signal the presence of a lethal gene may include whether there exists relatively high connectedness of the gene product to other proteins in the network$^3$, or if the lethal gene inactivates a reaction that produces or uses metabolites that are poorly connected with the rest of the network$^4$. It is possible for a gene KO to exhibit conditional lethality, i.e., to prohibit growth of the mutant in certain media but also permit growth in different surroundings. The KO of aceA is an example of such a gene that is essential in one
media, namely acetate, but is nonessential for growth in glucose or glycerol. The present study makes use of this conditionality as a line of investigation toward better understanding gene functionality by searching for a set of genes that are unconditionally lethal, and defining a metabolic core consisting strictly of the members in this set. In this case of global lethality, the knocked-out organism displays no growth in all tested media, and the interpretation of this set as an identifier of reactions essential for cell survival independent of environment has applicability to design of antibiotics that look to unfailingly kill the cell.

Related efforts to search for a single connected group of reactions across numerous growth conditions have resulted in the identification of an activity reaction core. This core is defined to contain the reactions in the network that remain active after exposing the in silico organism to different environments. In comparison, the core in this study is based on reactions that correspond to globally lethal genes. We chose this characterization because activity of metabolic reactions is not necessarily an indicator of viability of the cell, yet it may be advantageous from a pharmacological perspective to understand when the bacteria live or die with respect to inhibition of particular enzymes coded for by these lethal genes.

2. Methods

The deletion phenotype can be measured experimentally, though cost effectiveness and freedom in choice of KOs are driving factors toward constructing computational models that predict fluxes in metabolic reactions once the cell’s genotype has been altered. Analytic and numerical calculations that model changes in metabolite concentrations with differential equations are hampered by limited knowledge of most kinetic parameters. In this study, Flux Balance Analysis (FBA) and Minimization of Metabolic Adjustment (MOMA) are exploited as alternate constraint-based modeling approaches to predicting cell behavior. These methods make use of the relatively well-understood topology of the network and stoichiometry of the biochemical reactions that comprise the network in order to evade need for kinetic parameter fitting.

Flux Balance Analysis and Minimization of Metabolic Adjustment

FBA begins its translation of the complexity of the living organism into the language of computational modeling by constructing a set of linear equations to describe movement of metabolites through the network, where input metabolites (e.g., ATP, glucose) are indicated by positive coefficients and output metabolites (ADP, biomass) are denoted with negative coefficients (“Step 1” in Figure 1). These mass balance equations also distinguish exchange fluxes (b1 through b4 in Figure 1) that carry metabolites across the cell membrane from internal fluxes (v1 through v5) in the cytoplasm. Once the balance equations have been assembled, the coefficients can be extracted to serve as entries of the m x n stoichiometric matrix, S (“Step 2”). Next, the matrix S, flux vector v, and the assumption of steady state metabolite concentrations present us with a simplified means of expressing the mass balance equations in terms of the dot product between S and v, namely, $S \cdot v = 0$. 

159
Fig. 1. The steps involved in performing FBA. The first step involves taking the mapped metabolic network and writing mass balance equations to describe internal and exchange fluxes passing through the system. Dashed contour represents the system boundary. “Step 2” illustrates how the mass balance equations are used to construct the stoichiometric matrix $S$, which is then used in conjunction with the steady state assumption to solve for the solutions to a set of linear homogeneous balance equations (“Step 3”) previously established. With nine fluxes defined in the system, the third step symbolizes a nine-dimensional coordinate system where the feasible space of solutions exists. The final step exemplifies how constraints are applied in order to reduce the space of feasible solutions.

(It should be noted that the assumption that metabolite concentrations are time-independent is a valid one when compared to the time scale of gene regulation.) As $m$ number of metabolites is typically less than $n$ number of reaction fluxes in the network, the system is underdetermined, and the kernel of $S$ presents a feasible space of fluxes that solve the system of mass balance equations (“Step 3”). To reduce the feasible space, reaction fluxes are constrained to account for such features as reversibility of reactions, or to match experimentally the true measured flux distribution of the cell. In further efforts to reduce the set of possible solutions, linear programming is applied to locate a point in the feasible space that represents the successful optimization of an ordained objective function. In the case of FBA, linear programming is employed to single out the flux that maximizes cell growth, i.e.
production of biomass ("Step 4"). Optimization of cell growth in *E. coli* is an experimentally verified objective for the cell⁹.

While FBA has been proven a successful means of predicting the behavior of *E. coli* after a change in carbon source or genetic material, there are instances where FBA incorrectly predicts cell survival in a media and KO combination where, experimentally, growth rate goes to zero⁵.

By relaxing the requirement on optimizing cell growth, MOMA provides an alternative method of predicting the mutant phenotype. This method preserves many of the same principles in the FBA scheme, but allows the cell to operate in a suboptimal regime immediately after the KO is induced. The wild type flux as an optimal growth rate becomes a point in flux space closest to the suboptimal rate, since MOMA locates the suboptimal rate $v_{\text{MOMA}}$ by minimizing Euclidean distance $D = [(v_{\text{wild-type}} - v_{\text{MOMA}})]^{1/2}$ between the optimal growth rate of the wild type, $v_{\text{wild-type}}$, cell to the reduced flux space of the mutant⁹. Interestingly, it is presently unclear why minimization of the Euclidean metric successfully predicts performance of the mutated *E. coli*.

Comparing recent findings in the literature that utilize MOMA to identify lethal genes with results based on FBA suggests that MOMA correlates more strongly with experiment than FBA; there are cases where MOMA predicts zero cell growth when FBA predicted nonzero growth, and the cell does not live in reality¹¹.

**Construction of the lethal core after identification of globally lethal genes**

The metabolic network of *E. coli* strain K-12 MG1655 used in this study contained 720 fluxes and 436 metabolites, and was obtained from the iJE660 database⁵.

Global lethality was tested in 12 media. Throughout each test, the amount of nitrogen and oxygen provided to the cell was held fixed at 100 mmol and 20 mmol, respectively, while the carbon source was varied. In 11 of the cases, the *in silico E. coli* was fed 10 mmol of a single carbon source, and the output fluxes for each source were permitted to range from 0 to 1000 mmol. In the twelfth case, the *E. coli* was subjected to perform in a rich media containing 10 mmol each of the 11 previous sources. To be considered lethal, the knockout of a gene was required to have diminished the growth rate to less than $10^{-2}$ mmol g⁻¹ hr⁻¹. On the occasion where there was disagreement in MOMA and FBA results regarding lethality, the MOMA prediction was considered more accurate on the basis of MOMA’s history with experimental verification.

Upon identifying the collection of globally lethal genes, the lethal core was assembled by isolating reactions that were inactivated by their corresponding KOs in the globally lethal set. The metabolites present in these reactions served as nodes in the network, and connectedness between the nodes was determined by linking input metabolites to output metabolites. For an undirected graph constructed primarily to distinguish the degree to which the network was interconnected, specifically, whether the network was comprised of a single component rather than several components, it was not necessary to take into account the reversibility of some reactions nor to obtain a directed graph at this stage in the study.
3. Results and Discussion

After performing KOs across the 12 media, we found, on average, each media contained 220 ± 3 lethal genes, with the highest and lowest occurrences of lethality found in ethanol media (wild-type growth rate: 0.36 mmol) and the rich media (wild-type growth rate: 2.90 mmol), respectively. The number of lethal genes per each of the eleven media is displayed in Fig. 2. It is not surprising that the richest media exhibits the lowest number of lethal KOs because in such circumstances resources do not severely restrict the perturbed network, as there are many available pathways for the cell to utilize in case the knocked out gene debilitates part of a redundant pathway. Such robustness would not necessarily be expected in poor growth conditions, where various pathways can shut down due to limited resources and, consequently, the KO may have more impact in prohibiting cell growth.

The number of globally lethal genes across all media was slightly less than the average number of lethal genes found in any one medium. We identified 211 such genes with global lethality.

The reactions inactivated by their respective KOs can be mapped onto the metabolic network of *E. coli* in order to locate pathways heavily targeted by globally lethal genes. In Fig. 3 it can be seen that a number of the components of central metabolism, including glycolysis I, are unable to function after knocking out the globally lethal genes.

With the set of globally lethal genes identified, the lethality core was constructed for the *in silico E. coli*, and it was found that this network based on essential genes contains only a single component (Fig. 4).

![Fig. 2. Number of lethal genes identified per single carbon source tested.](image-url)
Fig. 3. Location of reactions knocked out by globally lethal genes with respect to metabolic pathways in *E. coli* K-12 MG1655 metabolism. Red shaded regions represent pathways where reactions were inactivated. Metabolic map provided by Ecocyc.org.

Fig. 4. The lethality core constructed as an undirected network. Notable features include its single-component structure and high density of edges out of nodes such as ATP or NADP.
To explore whether the lethal core constructed from 12 sampled media can be generalized to represent a core constructed from a larger set of sampled growth conditions, the change in the core size versus the number of sampled media is examined (Fig. 5). The relative size of the core is seen to change slightly from the case when one media is sampled to the case when all twelve are sampled, dropping in size from a core that contains 41% of all reactions in the metabolic network, to 39%.

Studying the lethality core in anaerobic growth conditions is a natural extension of the findings here, as *E. coli* are able to survive deprived of oxygen, and we have searched for the set of genes that are lethal under all conditions the cell may be exposed to. After repeating the procedure described above with the same set of 12 tested media, we report preliminary results that suggest the core undergoes no significant change with respect to size or connectedness as the cell grows in anaerobic surroundings.

Practical applications of the identification of a lethality core may be naturally extended to the design of antibiotics. As designers of advanced drugs look for means of reliably killing bacteria, reactions found in the metabolic core constructed here point toward a method of exploiting the set of globally lethal genes to assure a cell dies regardless of the particular environment it happens to find itself in at the time when the antibiotic becomes effective. Enzymes that catalyze reactions encountered in the metabolic pathways tetrahydrafolate biosynthesis and cell peptidoglycan biosynthesis are already being targeted by antibiotics tailored to inhibit their production14, and both of these pathways are present in the lethality core constructed here. It should be noted, however, that the strain studied in the present analysis is non pathogenic, though the concept of constructing a core based on essential genes can be readily extended to the pathogenic strain, as well as other organisms in general.

### 4. Conclusion

Gene functionality can be explored in the setting of metabolic networks, particularly by applying gene KOs and observing the effect on growth rate during the cell’s reaction to the perturbation. We have identified 211 globally lethal genes across 12 media, and have constructed a wholly connected lethality core from this set of essential genes. It is foreseeable to use this core as a tool for the design of advanced antibiotics that inhibit enzyme production in particular metabolic pathways of the living bacteria.
Potential questions posed by these findings include whether the core is present in other bacteria or increasingly complex organisms. The probability of finding a randomly connected core must also be addressed.

Acknowledgements

BU is grateful to Natali Gulbahce for guidance in conducting this research, and to Eivind Almaas and Adilson Motter for helpful discussions.

References

Reproducing Numerical Simulations More Efficiently and More Rapidly for Turbulent Rayleigh-Taylor Instability

J. A. Warner¹, M. J. Andrews²

1. Los Alamos National Laboratory, LASS, University of Wisconsin-Madison, Department of Physics
2. Texas A & M University, College Station, Texas 77843-3123

Abstract: Using multimode, one dimensional numerical simulations, turbulent Rayleigh-Taylor instability is investigated. The initial perturbations are given through a random number generator with the root mean square being kept the same as previous numerical simulations. Initially the smaller wavelengths dominated the modes, but as they reached terminal velocity, the larger wavelengths tended to overcome them in the end of the mixing. We also show that lower values for the alpha growth constant in experiments are due to the initial perturbations that are usually not prevalent in most numerical simulations, in addition to a molecular mixing variable (σ). We reproduce other numerical simulations and their output data much more quickly, as in a matter of seconds, rather than hours.

I. Introduction

Rayleigh-Taylor (RT) instability occurs when a dense fluid is either accelerated or pushed by gravity down toward a less dense fluid.¹ The two fluids start mixing at the interface. The paths into the less dense fluid by the more dense fluid are called “spikes” while the paths into the more dense fluid by the less dense are called “bubbles.” The initial perturbations at the interface increase exponentially until the amplitudes of the bubbles become large relative to its wavelength.² The rate at which these fluids mix is affected by the Atwood number A: \[(p_h - p_l)/(p_h + p_l)\]. The amplitude of the bubbles rising into the denser fluid grow as: \[h_b = \alpha_b A g t^2\]. As the smaller wavelengths reach a terminal velocity that is proportional to their wavelength, the larger wavelengths continue to grow and eventually dominate because their terminal velocity will be greater.

II. Numerical Simulation Configuration

The numerical simulation is done with one code. The goal was to reproduce the seven different numerical simulations given in the Dimonte et al. paper.³ While some of these simulations took a full day to complete, we were able to complete our multi-mode, one dimensional simulation in only a few minutes. Using a random number generator and a target root mean square value for all of the amplitudes of 3E10-4, we created initial amplitudes for each of the modes 16-32 and 32-64. Each of these modes was then fed into the main part of the program and grew according to the classical growth rate. The main part of the program updated the amplitudes of each mode and took the maximum amplitude of the modes and set that as the height. Each time step also modified the Atwood number according to: \[A = 2\sigma A_0\]. The initial conditions for \(\sigma\), the molecular mixing variable, is set as: \(\sigma^2 = 1/4\). Every time step the \(\sigma\) was modified according to it’s derivative: \[\dot{\sigma} = (1/4)*v/h-(1+C_m)*\sigma^2*v/h\]. \(C_m\) is initially set to 2.

III. Results
As the modes and their amplitudes are fed into the main part of the numerical simulation, the amplitudes are updated with each time step. Every step the maximum amplitude of the modes is set as the height. Whichever mode had the largest amplitude at that time step was also outputted. As the simulation continued, it showed a pattern of the smaller wavelengths having the greatest amplitudes initially, while the larger wavelengths tended to dominate eventually. This agrees with the theory mentioned in the introduction.

Several problems came about from the simulation. The first problem came from our random number generator. The amplitudes that we were creating were “random” and varied, but unfortunately every time the program ran, it created this same list. Therefore, it was necessary to change the seed for the number generator every time the program was initiated in order to create a new random list of numbers.

The second problem that came about was that the mode number 28 seemed to be a particularly special mode. When the modes 16 to 32 or 16 to 64 were run, the program always outputted that the mode with the largest amplitude was 28 and then it collapsed down to the lowest mode, such as 16. This “collapse” agrees with the theory because it shows that the lower mode numbers, which means the higher wavelengths, tend to dominate the process toward the end of the process. We still cannot explain why the highest mode ran, say 32, would not dominate at first. Any mode number higher than 28 never tended to have the highest amplitude in our simulation.

IV. Conclusion

Our model has had a modest level of success. It manages to incorporate random initial perturbations, something others fail to do. The Atwood number and the molecular mixing variable also vary accordingly in the model. It uses multimode mixing with all these variables and manages to process all the information in only a matter of minutes. The limitations are the fact that it is only one dimensional, whereas other models are two dimensional and three dimensional. If this model is improved upon and mode maximums are fixed so that it falls in amplitude from the highest mode to the lowest, then this model will be a very fast and efficient simulation of Rayleigh-Taylor turbulent mixing.
MPI: Exploring the Advantages, Limitations, and Applications

J. A. Warner¹, M. J. Andrews²

¹ Los Alamos National Laboratory, LASS, University of Wisconsin- Madison, Dept. of Physics
² Texas A & M University, College Station, Texas 77843-3121

Abstract: In recent years, the need for a standard in parallel computing has come to surface because of the various different languages and software used on different systems. MPI, or Message-Passing Interface, has become the most widely used standard for parallel computing. MPI's foundation of libraries with common functions allows the machines to talk to one another in a simple way of message passing. We have created a small program that displays the advantages and disadvantages of MPI and parallel computing. Our example program shows that it reduces the total process time from 5.35 seconds on one processor to 0.83 seconds when run on 10 processors. This 85 percent decrease in computing time opens many new options to scientists working on numerical simulations. We also take a look at the limitations of parallel processing and how in some instances it may even slow down the performance of a program.

I. Introduction

The drive for more computing power has been fed with a voracious appetite ever since the first computer came to creation. The idea of a supercomputer arose and companies such as IBM have continuously managed to increase the performance of these massive mainframes at an unimaginable pace. Still, few research centers and universities can afford these superior machines. By hooking up several (or several hundred) machines together, research scientists have been able to create efficient and relatively cheap processing centers for their numerical simulations and time-intensive programs. In the early 1990's, it became clear that there was a need for a common language or standard so that all of the different machines that were grouped together to run the same program could communicate effectively¹. In 1994, this standard was introduced as MPI², creating a simple, yet flexible foundation for the parallel computing community. If one machine wants to send data to another, it sends a “message” which the other computer receives. By breaking the sending of data down into simple messages that can be passed back and forth from various processes, MPI creates a very efficient way of managing data.
II. The Idea Behind the Array Program

The program that we have created uses several different functions of MPI. The idea of the program is to show the possible applications of MPI. It creates several three-dimensional arrays that simulate the x, y, and z coordinates of a 256x256x512 grid. According to the user's needs, this array may contain the velocity, gradient, density, etc., at each point in the grid. While it may take one processor a large amount of time to do calculations for each point in this grid, we show the advantages of parallel computing. It divides the array up into equal blocks of data. These blocks are divided up among the processors available, allowing them to each do their calculations with only that portion of the array. When the processors are each done with their blocks, they wait until the others have completed their jobs as well. Then the 0 process takes its three dimensional array and slices it up into different “k-planes.” It takes each one of these k-planes and forms a two dimensional array out of it. Then it sends this two dimensional array to an output file called array.dat. Once it is done, it starts receiving one dimensional arrays from the others processes and converts them into two dimensional arrays to be outputted. It repeats this process until every plane of the three dimensional array has been outputted to the array.dat file and then they all wait until the next loop of work is needed to be done.

III. Step-By-Step Explanation of the Program

The following is the actual code and comments made for each step of the program:

PROGRAM array

c Written by Jacob Warner 7/28/06

c Program is only meant to be run on a number of processors that

c is a factor of 512 so the block of k-planes can be divided evenly.
c For the lepton workstation, that means only 1,2,4,8,or 16 processors.

c The command line (on lepton) to run this program was "mpirun -np 16 array"
c The -np meant that the number after it is the number of processors you
c want to use. The array is simply the name of the program.

c This includes the MPI library where all the calls come from
 INCLUDE '/usr/local/mpich/include/mpif.h'

c Creates the internal error integer, something that only Fortran requires
integer ierr

c The rank of the current process, the starting k plane for that process,
c the ending k plane for that process, the number of k planes per process,
c and the total number of processes working.
  integer my_rank, start, endnum, blocksize, p

c Creates 3 allocatable arrays, which means that you can specify the
c dimensions of them later in the program. Also creates a temp 2D array
c on each process as well as a 1D array on each process named small
REAL, DIMENSION (:,:,:), ALLOCATABLE::a, b, c
REAL, DIMENSION (1:256,1:256)::temp
REAL, DIMENSION (65536)::small

c Starts up MPI. No MPI functions can be called until this one has been called.
call MPI_INIT(ierr)

c Find out the rank of the current process
c The MPI_COMM_WORLD is the "group" that contains all the processes
c For this program, we only have the global group of processes
c The my_rank variable takes the rank that the function returns and stores
c it as my_rank. Each process calls this function and stores a different
c value for my_rank. The ranking system goes from 0 to p-1, where p is the
c total number of processes. The ierr variable at the end is always needed
for every MPI call in Fortran.
call MPI_Comm_rank(MPI_COMM_WORLD, my_rank, ierr)

c Find out the number of processes
c The total number of processes will be stored as the variable p on each process.
call MPI_Comm_size(MPI_COMM_WORLD, p, ierr)

c Divides the k part of the array into a number of blocks.
  blocksize=512/p

c Each process uses its rank to figure out what k plane it will start at and end at.
  start = my_rank*blocksize + 1
  WRITE(*,*) "Process ",my_rank," starting at ",start
  endnum = (my_rank+1)*blocksize
  WRITE(*,*) " Ending at ", endnum

  Each process creates the 3 Dimensional arrays, but has a different
  section of k-planes.
  ALLOCATE (a(0:257,0:257,start:endnum))
  ALLOCATE (b(0:257,0:257,start:endnum))
  ALLOCATE (c(0:257,0:257,start:endnum))

  Each process does its own portion of the array.
DO 100 k=start, endnum
  DO 100 j=1,256
    DO 100 i=1,256
      a(i,j,k)=i+j+k
      b(i,j,k)=i-j+k
      c(i,j,k)=i-j-k
  100 CONTINUE

DO 300 k=start, endnum
  DO 250 j=1,256
    DO 200 i=1,256
      a(i,j,k)=b(i,j,k) + c(i,j+1,k) + c(i-1,j,k)
    200 CONTINUE
  250 CONTINUE
300 CONTINUE

c Every process calls this function. No process goes any further until
c every process has called it.
  CALL MPI_BARRIER(MPI_COMM_WORLD,ierr)

c If it is process 0, it takes all of its data and sends it to an output file.
c After that it starts receiving 1D arrays from all of the other processes.
c Then it reshapes them to 2D arrays and outputs it to a file.
IF(my_rank==0)THEN
  OPEN(UNIT=1,FILE="array.dat",STATUS="REPLACE")

  DO 440 k=start,endnum
    temp=a(1:256,1:256,k)
    WRITE(1,430) temp
  430 FORMAT(16(F8.2,X))
  WRITE(1,435)
  435 FORMAT(/)
  WRITE(*,*) "Done outputting k-plane=",k," to array.dat"

440 CONTINUE

DO 600 j=1,p-1

c For some reason, by this point in the program, the process 0 has forgotten
c how many processes are being used as well as the blocksize.
call MPI_COMM_SIZE(MPI_COMM_WORLD,p,ierr)
blocksize=512/p

DO 500 k=1,blocksize
call MPI_RECV(small,65536,MPI_REAL,j,1,& MPI_COMM_WORLD,ierr)
temp=RESHAPE(small,(/256,256/))
WRITE(1,450) temp
450 FORMAT(16(F8.2,X))
WRITE(1,475) FORMAT(/)
475 FORMAT(/)
500 CONTINUE

call MPI_BARRIER(MPI_COMM_WORLD,ierr)
600 CONTINUE

CLOSE(1)

ENDIF

c If it is not the 0 process, each process will slice their k planes up and
c reshape them to a 1D array. Then they send this to the process 0. These
c sending and receiving functions of MPI are blocking. That is, no process
c will either send or receive until it has a digital "handshake" with the
c other process in the communication.
IF(my_rank.ne.0)THEN

DO 425 ord=1,p-1
IF(my_rank==ord)THEN
start = my_rank*blocksize + 1
endnum = (my_rank+1)*blocksize
DO 400 k=start,endnum
  temp=a(1:256,1:256,k)
small=RESHAPE(temp,(/65536/))
call MPI_SEND(small,65536,MPI_REAL,0,1,& MPI_COMM_WORLD,ierr)
WRITE(*,*) "Done outputting k-plane= ",k," to array.dat"
400 CONTINUE
IV. Some Limitations of MPI

Sending and receiving messages takes up much more time in a program than simply having the machines do their own work. This is why it is important to reduce the number of messages to as few as possible when writing an MPI program. If each machine had to send each of the other machines all of their data, it would increase the process time immensely. While one machine is receiving data from another, it could not send its own data to yet another. With a three dimensional array that gets large in size, this can clearly create a “speed bump” for efficiency. The slow part of this program is the output file. While the actual work takes only seconds to process, the output to the array.dat file takes around five minutes. This shows it is very beneficial to only output the data at the end, instead of between loops.

V. Applications of MPI

In general, any numerical simulation that needs to work with a lot of data and needs to make a number of calculations would greatly benefit from MPI. One current research project that we are investigating is Rayleigh-Taylor turbulent mixing. Experimental results cannot be reproduced very precisely using computer simulations. By taking in more factors into the calculations and also initial perturbations that are only present in experimental tests, it is predicted that numerical models will be able to attain greater precision in their results. Current programs either require massive amounts of
computing power in order to run, or simply are too rough to get good results. Using MPI, a researcher could parallelize his or her program and decrease the run-time of the program down to realistic levels. Whatever the application, it is obvious that MPI will help remove many obstacles to researchers and their ventures into parallel computing.

VI. Conclusion

We have shown that MPI has a wide variety of applications in the scientific community. Through the creation and execution of our sample program, we have shown that massive amounts of data can be divided amongst a number of machines in order to decrease the total run-time of the program. In addition, the limitations and applications have been shown and discussed. We hope it is a recognizable fact that MPI is a very advantageous standard in parallel computing.

References