The Board of Trustees of the Center for Excellence in Education

The Center for Excellence in Education
140 Park Street S.E., Second Floor
Vienna, VA 22180-4627

A Nonprofit Organization Under U.S. Code 501 (c) (3)

Research Science Institute 2003
Faculty and Staff

RSI Director		
Dr. Carey Inouye		Iolani School; CEE
Director of Administration		
Ms. Maite P. Ballestero		CEE
Assistant Director		
Mr. Daniel Haspel	RSI '98	Bain & Company
Technical Assistant		
Mr. Christopher C. Mihelich	RSI '97	Princeton University
Professors		
Ms. Amy Bany	RSI '90E	Yale University
Mr. Adrian Campbell	RSI '92	Bayer
Dr. Charles Doran	RSI '87	Columbia University
Mr. Lance Rhoades		University of Washington
Dr. Dennis Ugolini	RSI '90W	Caltech
Academic Staff		
Mr. Jeremy England	RSI '98	Harvard University
Ms. Yi-Ching Ong	RSI '98	Harvard University
Mr. Matthew Paschke	RSI '91	Center for Excellence in Education
Dr. John Rickert		Rose-Hulman Institute of Technology
Dr. Jenny Sendova		Bulgarian Academy of Sciences
Mr. Sapan Shah	RSI '97	Case Western Reserve University
Teaching Assistants		
Mr. Edward Boas	RSI '94	Stanford University
Mr. Gabriel Carroll	RSI '00	Harvard University
Mr. Josh Gottlieb	RSI '02	MIT
Ms. Sanghamitra Sen	RSI '00	MIT
Mr. Balint Veto	RSI '98	MIT
Student Counselors		
Mr. Arrak Bhattacharyya	RSI '01	Midland, MI; MIT
Mr. Robert Bryant	RSI '02	Essex Junction, VT
Mr. Douglas McClure	RSI '01	Rye, NY; Harvard University
Mr. Malcolm Murdock	RSI '00	Stanford University
Ms. Jamie Rubin	RSI '02	Ft. Myers, FL
Ms. Emma Schmidgall	RSI '02	Golden Valley, MN
Mr. Vivek Venkatachalam	RSI '01	Berkeley Heights, NJ; MIT

Preface

The *Research Science Institute Student Reports 2003* demonstrate the work of 79 academically talented high-school students who participated in the 2003 Research Science Institute (RSI) sponsored by the Center for Excellence in Education in collaboration with the Massachusetts Institute of Technology in Cambridge, Massachusetts, from 22 June through 2 August.

This unique six-week summer program combines theoretical classroom instruction in mathematics and science with one-on-one research experience with scientists from the Massachusetts Institute of Technology, Harvard University, Boston University, Northeastern University, private corporations, and research organizations in and around Boston, Massachusetts. It is the only U.S. program of its kind offered at no cost to participants. The scientists and their respective organizations are acknowledged in the headings of the reports.

These reports are the culmination of students' research in mathematics, physics, chemistry, biology, computer science, and engineering. The faculty of the Institute, together with a team of outside experts, selected five outstanding written reports for special recognition. A panel of visiting educators and scientists also chose five projects to receive awards for outstanding oral presentation. The winning written reports are printed first in their entirety, followed by the abstracts for the reports that won the oral-presentation awards, and finally by the remaining student abstracts.

The Institute has drawn students from all over the U.S. and from 45 other countries. In 2003 the 79 students represented 33 states of the U.S. and the District of Columbia, as well as the countries of Bulgaria, Egypt, Germany, India, Israel, Lebanon, Morocco, Pakistan, Poland, Russia, Singapore, Turkey, and the United Kingdom. The students were all selected having completed coursework comparable to or exceeding that of the third year of high school. The U.S. students averaged 78 of 80 points on the mathematics section and 75 of 80 points on the verbal sections of the Preliminary Scholastic Aptitude Test (PSAT), representing the 99^{th} percentile of all high-school students taking the PSAT.

The inestimable value of the Institute to scientifically talented students is evident from the following praises from participants of RSI 2003:

> *I adopted a more analytical and critical approach to research, and learned to ask questions about anything that I was unsure of.*
>
> *The step-by-step process of making a good research report was invaluable and will help me later on in my scientific endeavors.*
>
> *For the sciences, especially biology, along with physics and chemistry, I have gained a deeper respect.*
>
> *I have reawakened my love for physics.*
>
> *I have never learned so much in such a short period of time.*
>
> *I strongly believe my mentor has been one of the greatest people that I've met.*
>
> *I love my counselor, my tutor, my friends ... and the social life!*
>
> *The lab people are the most wonderful thinkers and debaters I have ever had the pleasure of conversing with.*

The editor is honored to have taken part in the Institute as student and as staff member and is pleased to present to you the excellent work of his successors.

Christopher C. Mihelich, RSI 1997, *Editor*

Table of Contents

The Effects of Natural Selection and Mutation on Inbreeding Effective Size

Can Cenik

under the direction of
Fangyun Yang
Department of Mathematics
Massachusetts Institute of Technology

Abstract

In this paper Ryman's effective-size model is extended by integrating basic natural-selection and mutation models into it. In addition, a novel natural-selection model is developed for a sequence of generations that has fluctuating numbers of individuals in each generation. A modified Poisson distribution is used to model different productivities of individuals and therefore natural selection. Finally, we review the related literature to propose future work on the subject.

1 Introduction

When genes are randomly assorted from one generation to the next in finite populations, genetic diversity is lost. This phenomenon is known as genetic drift. For selectively neutral genes, the dynamics of this process depend on the effective population size. The effective size for a given genetic parameter is defined as the size of an ideal population that will show the same genetic behaviour (Wright [6]). Depending on the parameter considered, different effective sizes, such as variance effective size, eigenvalue effective size, and inbreeding effective size, can be derived. This paper focuses on inbreeding effective population size (Wright [1]).

Genetic polymorphism within a population is a prerequisite for adaptive change, so it is crucial to understand the factors that promote the retention of genetic variability. The importance of preserving levels of genetic variation within populations has long been recognized for influencing short-term population viability as well as for maintaining a population's adaptability to environmental changes (Frankham [14]). In an isolated population, mutation rate and effective population size determine the level of genetic variation (Nunney [11]). Inbreeding causes a loss in the number of heterozygotes and in genetic diversity. For this reason, inbreeding effective size is an important issue in conservation biology. A better understanding of the interaction between natural selection, mutation, gene flow (Hoole et al. [19], Tremblay et al. [20]), and effective population size will help us to design better conservation strategies (Lewandowski et al. [21]) for organisms ranging from long-lived trees of rainforests (Shapcott [13]) to rare cattle breeds of commercial interest (Jordana [16]). Even the controversial issues of conservation strategies like consequences of lineage mixing (Berger [18]) and effects of habitat alteration and fragmentation (Scribner [17]) can be understood better when we have more knowledge about effective population size.

Effective-size estimates in discrete and overlapping generations are very different. The Wright-Fisher model is derived for nonoverlapping generations, whereas Pollak [22, 23] and Wang et al. [24] suggested models for overlapping generations. Ryman et al. [3] derived a model that can be used for organisms having a combination of discrete and overlapping models. This model can be used for a time series only, but it generalizes the harmonic and arithmetic model under different demographic assumptions. This paper uses the same approach as Ryman's model.

A great deal of work has been done in order to incorporate natural selection into effective-size theories (Caballero [5]). Gillespie [8], Araki et al. [7] and Tachida [9] used effective-size models in order

to test the neutral theory of molecular evolution, and Burrows (1984) [4] used phenotypic selection and optimum-index selection models to understand the effects of inbreeding on unrelated families. All of this effort was based on the Wright-Fisher allelic model.

In this paper, the definition of effective size suggested by Ryman's [3] model is used as a basis for the incorporation of natural selection and mutation to develop a more complete model. Ryman et al. [3] defined effective size as the inverse of the probability that two homologous genes in two randomly chosen individuals of one generation were derived from the same individual in the previous generation. This model takes into account different productivities in different reproduction periods in a single generation. This paper uses two different approaches to model natural selection. First, a simple single-locus natural-selection model is used to investigate the effects of mutation-selection equilibrium on Ryman's model, and equations are derived for calculating effective size after these processes. Second, a modified Poisson distribution is used to model different productivities of individuals, and novel equations modeling natural selection are derived.

2 Theory

2.1 Extension of Ryman's Model

First we consider a monoecious haploid organism. We assume that generations are nonoverlapping and population size is stable over generations. In this situation let p be the frequency of the dominant form of the allele. There are N groups, and N_i parents in the i^{th} group. x_i is the probability of an individual coming from the i^{th} group, and P is the probability that two homologous genes in two randomly chosen individuals of one generation were derived from the same individual in the previous generation. In this case P may be expressed by the equation

$$P = \sum_{i=1}^{N} p^3 \frac{x_i^2}{N_i} + \sum_{i=1}^{N} (1-p)^3 \frac{x_i^2}{N_i}. \tag{1}$$

From this equation[1] it is possible to define the effective population size N_e by the equation

$$\frac{1}{N_e} \equiv (p^3 + (1-p)^3) \sum_{i=1}^{N} \frac{x_i^2}{N_i}. \tag{3}$$

Next the one-locus natural-selection model was incorporated into the haploid model. In order to do this, W_p and $W_{(1-p)}$ are defined as the relative fitnesses of alleles p and $(1-p)$. Relative fitnesses are normalized weight factors of reproductive success associated with each allele. This yields

$$\begin{aligned} P = & \\ & \sum_{i=1}^{N} \left(\frac{pW_p}{pW_p + (1-p)W_{(1-p)}} \right)^2 x_i^2 p^2 \\ & \qquad \times \sum_{i=1}^{N_i} \left(\frac{1}{pN_i} \right)^2 \\ & + \sum_{i=1}^{N} \left(\frac{(1-p)W_{(1-p)}}{pW_p + (1-p)W_{(1-p)}} \right)^2 x_i^2 (1-p)^2 \\ & \qquad \times \sum_{i=1}^{N_i} \left(\frac{1}{(1-p)N_i} \right)^2. \end{aligned} \tag{4}$$

It is possible to consider the effects of mutation on Ryman's effective-size model in the same manner. Letting u be the rate of forward mutation and v be

[1] It is also possible to derive a similar equation for randomly mating diploid organisms. Let k be an arbitary positive integer such that each individual produces gametes proportional to k. We assumed that there are equal numbers of females and males. Let N denote the number of females and males; because each individual has two copies of the same gene i, it can be stated that each gender produces $2Nk$ gametes. All individuals in the next generation will take half of their genes from the mother and half from the father. For simplicity, it was assumed $k = 1$, making the probability of two randomly chosen individuals coming from the same parent

$$P = 2\left[\frac{1}{4} p^2 \left(\frac{1}{2Np} \right)^2 + \frac{1}{4} (1-p)^2 \left(\frac{1}{2N(1-p)} \right)^2 \right]. \tag{2}$$

the rate of reverse mutation, we would have

$$P = \sum_{i=1}^{N} \left(\frac{(p + v(1-p) - up) - v(1-p)}{(p + v(1-p) - up)} \right)^2 x_i^2 p \times \sum_{i=1}^{N_i} \left(\frac{1}{pN_i} \right)^2 + \sum_{i=1}^{N} \left(\frac{((1-p) + up - v(1-p)) - up}{(1-p) + up - v(1-p)} \right)^2 x_i^2 (1-p) \times \sum_{i=1}^{N_i} \left(\frac{1}{(1-p)N_i} \right)^2. \quad (5)$$

Equations (4) and (5) can be straightforwardly joined together to obtain a model that takes both mutation-selection equilibrium and inbreeding effective size into account.

All of the above equations are concerned with two consecutive generations. The next step is to generalize these models for an infinite number of generations. We assume that in every generation there are k groups and the progeny of the s^{th} group in the t^{th} generation constitute the s^{th} group in the $(t+1)^{\text{th}}$ generation. Each group has either N_1 or N_2 individuals in any given generation, and X_s is the number of individuals in the s^{th} group. Therefore a two-valued stationary Markov chain (see Appendix B) is defined such that its transition functions are

$$P(X_s(t+1) = X_s(t) | X_s(t) = N_j) = 1 - q_{ij} \quad (6)$$

and

$$P(X_s(t+1) \neq X_s(t) | X_s(t) = N_j) = q_{ij}, \quad (7)$$

and

$$P(N_s(0) = N_j) = p_{sj}^{(0)} \quad (8)$$

is the stationary distribution of the Markov chain (see Appendix A). Here $s \in \{1, 2, \ldots, k\}$, $j \in \{1, 2\}$, $0 < q_{ij} < 1$, and

$$(p_{s1}^{(0)}, p_{s2}^{(0)}) = \left(\frac{q_{s2}}{q_{s1} + q_{s2}}, \frac{q_{s1}}{q_{s1} + q_{s2}} \right). \quad (9)$$

E is the expectation calculated over the stationary distribution of the Markov chain. Letting $\pi_2(t)$ be the probability that two randomly chosen individuals from the population in generation t have the same parent in the initial state (Ewens) [12], we obtain

$$1 - \pi_2(t) = E\left[\prod_{g=0}^{t-1} \left(1 - \sum_{s=1}^{k} \frac{N_s(g)}{\sum_{s=1}^{k} N_s(g)} \frac{1}{N_s(g-1)} \right) \right] \times (1 - \pi_2(0)). \quad (10)$$

The inbreeding effective size can now be defined by

$$1 - \frac{1}{N_e} = \lim_{t \to \infty} \left(E \prod_{g=0}^{t-1} \left(1 - \sum_{s=1}^{k} \frac{N_s(g)}{\sum_{s=1}^{k} N_s(g)} \frac{1}{N_s(g-1)} \right) \right)^{1/t}. \quad (11)$$

We have therefore extended Ryman's effective-population-size model to include the effects of natural selection and mutation. The first equations ((1) to (5)) take into account the allelic frequencies and some basic mutation and selection models for haploid organisms, but they are derived for just two generations. This model is different from the classical Wright-Fisher model because it allows different productivities within a generation, so this is very useful in predicting the effective population sizes of semelparous organisms like salmon (Waples [2]). Since the model for infinitely many generations (equations (6) to (11)) is a two-valued stationary Markov-chain, computer simulations of these equations can be used to estimate the possible consequences of different productivities on the genetic behaviour of different populations. This will help us develop better conservation strategies when the fluctuations due to different productivities are important.

2.2 A New Natural-Selection Model

In this next section we suggest a new natural-selection model and integrate it into Ryman's effective size model through creation of diversity in productivities within a generation. This diversity will result in a fluctuation of population size (Iizuka [10]) in time, and we first consider this effect. In this model, population size can be any nonnegative integer, and the productivity within the generation is uniform and is Poisson-distributed with parameter λ. Multiplication of productivity and population size in the $(t-1)^{\text{th}}$ generation gives the population size in

the t^{th} generation. We have $N_1 = cN_0$ and

$$P(N_1 = cN_0) = x_i, \tag{12}$$

$$P(N_2 = mN_0) = \sum_{j_1 j_2 = m} P(N_2 = j_1 N_1 | N_1 = j_2 N_0). \tag{13}$$

As these two events are independent it is possible to derive

$$\sum_{j_1 j_2 = m} P(N_2 = j_1 N_1) P(N_1 = j_2 N_0) = \sum_{j_1 j_2 = m} x_{j_1} x_{j_2}. \tag{14}$$

It is also possible to obtain by induction that

$$P(N_t = mN_0) = P(N_t = j_1 N_{t-1}) \dots P(N_1 = j_k N_0) = \sum_{j_1 \dots j_k = m} x_{j_1} x_{j_2} \dots x_{j_k}. \tag{15}$$

The next step is considering different productivities within a generation. A new random variable G_i, which is a subgroup generated by the i^{th} individual in the 0^{th} generation, is defined. The probability that $G_i = n$ is $\lambda^n e^{-\lambda}/n!$. So the probability that two randomly chosen individuals come from the same parent in the previous generation is

$$P = E \cdot \frac{\sum_{i=1}^{N_0} G_i^2}{(G_1 + G_2 + \dots + G_{N_0})^2}. \tag{16}$$

After two generations every subgroup in G_1 generates new subgroups of random variables. Theoretically, after this kind of branching it is possible to sum all probabilities and find the same expectation. To make the model more realistic, it is possible to define an upper bound M [15] such that it is not possible for any single individual to have more progeny than M. With this upper bound set, $\sum_{x=0}^{M} \lambda^n e^{-\lambda}/n!$ will be smaller than one. To normalize the distribution, every single term from 0 to M can be divided by $\sum_{x=0}^{M} \lambda^n e^{-\lambda}/n!$.

Equations (12) to (14) suggest a novel model of natural selection. The approach is the same as Ryman et al. [3], but it is suggested that every individual can be treated as a subgroup having a specific probability distribution of having different numbers of progeny, instead of uncertain different productivities within a generation. The model proposed is based on differences in each individual's reproduction rate. This takes every individual as one phenotype that natural selection acts on and does not deal with any allelic frequencies. Our model assumes that every individual has some probability of having 0 to M progeny. The probabilities for having different numbers of progeny form a modified Poisson distribution. Our model is completely governed by a stochastic selection process. It is also possible to find the effective population size (equation (15)) after stochastic selection, but calculations become very complicated when generation numbers increase.

3 Discussion

The genetic structures of populations change according to specific factors such as mutation, migration, natural selection, inbreeding, random sampling of alleles due to finite number of individuals, and nonrandom mating. In order to understand the course and mechanisms of evolution, an understanding of the dynamics of population genetics is essential. All existing models consider a relatively small portion of the factors mentioned above.

Both models proposed in this paper are extensions of Ryman's effective-size model. After the incorporation of single-locus natural selection, and basic mutation models, a new stochastic natural-selection model based on individuals' difference in reproductivity was suggested, and equation (15) is derived for calculating effective size after this process. This model makes it possible to develop better conservation strategies for organisms that have the characteristics of both overlapping and discrete generation models. In a broader sense, this model is a more complete attempt to understand the genetic structure of populations and, therefore, the course of evolution.

Effective population size is a useful way to study different factors acting simultaneously on a population because it considers the divergence times from the ancestral genes and allows us to study the effects of various factors on this process. The approach used in this paper is to fix a starting generation and to figure out the probabilities of divergence from the same individual in this generation. It is also possible to investigate the distributions of times to coalescence as in Polanski et al. [25] without fixing a parental generation. However Polanski et al.'s paper does not take natural selection and mutation into account.

The models in equations (1) to (5) are for autosomal characters that are determined by single genes. The model can be extended for polygenic characters and sex-linked loci by similar approaches used in Burger [26] and Nomura [27] respectively.

Our model is limited in the sense that it does not take some factors into account. It is a future problem to investigate the properties of our effective-size model (equations (12) to (15)) with the effects of stable mutation rate (Ishii et al. [28]), age-structure (Pollak [29]), spatial subdivision (Wakeley [31]), and/or nongenetic inheritance of fertility (Sibert et al. [30]).

4 Acknowledgments

I would like to thank my mentor Fangyun Yang for her help throughout this program, and Prof. Hartley Rogers for making this mentorship possible. I am especially grateful to Jeremy England for his patience and excellent suggestions, and Matt Paschke for his ongoing assistance. I would also like to thank all the other staff members, especially my tutor, Sapan Shah, for helping me organize my ideas and for reading over many drafts of my paper. I am grateful for the support of the Center for Excellence in Education and the Research Science Institute.

References

[1] Wright, S.: Evolution and the Genetics of Populations. Vol. 2, The Theory of Gene Frequencies. *The University of Chicago Press*, 1969: 11(C).

[2] Waples, R.S.: Effective Size of Fluctuating Salmon Populations. *Genetics*, 2002: 161, 783–791.

[3] Ryman, N. & Laikre, L.: Effects of Supportive Breeding on the Genetically Effective Population Size. *Conservation Biology*, 1991: 5, 325–329.

[4] Burrows, P.M.: Inbreeding under Selection from Unrelated Families. *Biometrics*, 1984: 40, 357–366.

[5] Santiago, E. & Caballero, A.: Effective Size of Populations Under Selection. *Genetics*, 1995: 139, 1013–1030.

[6] Wright, S.: Evolution in mendelian populations. *Genetics*, 1931: 16, 97–159.

[7] Araki, A. & Tachida, H.: Bottleneck effect on evolutionary rate in the nearly neutral mutation model. *Genetics*, 1997: 147, 907–914.

[8] Gillespie, J.H.: The Neutral Theory in an infinite population. *Gene*, 2000: 261, 11–18.

[9] Tachida, H.: DNA evolution under weak selection. *Gene*, 2000: 261, 3–9.

[10] Iizuka, M.: The effective size of fluctuating populations. *Theoretical Population Biology*, 2001: 59, 281–286.

[11] Nunney, L.: The effective size of annual plant populations: The interaction of a seed bank with fluctuating population size in maintaining genetic variation. *The American Naturalist*, 2002: 160, 195–204.

[12] Ewens, W.: *Mathematical Population Genetics Springer-Verlag*, New York, 1979.

[13] Shapcott, A.: Population genetics of the long-lived Huon pine *Lagarostrobos franklinii*: An endemic Tasmanian temperate rainforest tree. *Biological Conservation*, 1997: 80, 169–179.

[14] Frankham, R.: Inbreeding and extinction: a threshold effect. *Conservation Biology*, 1995: 9, 792–799.

[15] Watson, H.W. & Galton, F.: On the probability of the extinction of families. *The Journal of the Anthropological Institude of Great Britain and Ireland*, 1875: 4, 138–144.

[16] Jordana, J. & Piedrafita, J. : The *Bruna dels Pirineus* (Pyrenean brown breed): a genetic study of a rare cattle breed in Catalonia (Spain).

[17] Scribner, K.T. & Arntzen, J.W. & Cruddace, N. & Oldham, R.S. & Burke, T.: Environmental correlates of toad abundance and population genetic diversity. *Biological Conservation*, 2001: 98, 201–210.

[18] Berger, J. & Cunningham, C.: Multiple bottlenecks, allopatric lineages and badlands bison *Bos bison*: Consequences of lineage mixing. *Biological Conservation*, 1995: 71, 13–23.

[19] Hoole, J.C. & Joyce, D.A. & Pullin, A.S.: Estimates of gene flow between populations of the swallowtail butterfly, *Papilio machaon* in Broadland, UK and implications for conservation. *Biological Conservation*, 1999: 89, 293–299.

[20] Tremblay, R.L. & Ackerman, J.D.: Gene flow and effective population size in *Lepanthes* (Orchidaceae): a case for genetic drift. *Biological Journal of the Linnean Society*, 2001: 72, 47–62.

[21] Lewandowski, A. & Burczyk, J. & Mejnartowicz, L.: Genetic structure of English yew (*Taxus baccata* L.) in the Wierzchlas Reserve: implications for genetic conservation. *Forest Ecology and Management*, 1995: 73, 221–227.

[22] Pollak, E.: Effective population numbers and mean time to extinction in dioecious populations with overlapping generations. *Mathematical Biosciences*, 1980: 52, 1–25.

[23] Pollak, E.: Effective population numbers and mean time to extinction in monoecious populations with overlapping generations. *Mathematical Biosciences*, 1979: 46, 87–106.

[24] Wang, Y. & Pollak, E.: The effective number of a population that varies cyclically in size. II. Overlapping generations. *Mathematical Biosciences*, 2002: 179, 161–181.

[25] Polanski, A. & Bobrowski, A. & Kimmel, M.: A note on distributions of times to coalescence, under time-dependent population size. *Theoretical Population Biology*, 2003: 63, 33–40.

[26] Burger, R.: Predictions of the dynamics of a polygenic character under directional selection. *Journal of Theoretical Biology*, 1993: 162, 487–513.

[27] Nomura, T. : Effective population size for a sex-linked locus in populations under selection. *Mathematical Biosciences*, 1997: 142, 79–89.

[28] Ishii, K. & Matsuda, H. & Iwasa, Y. & Sasaki, A.: Evolutionary stable mutation rate in a periodically changing environment. *Genetics*, 1989: 121, 163–174.

[29] Pollak, E.: The effective population size of some age-structured populations. *Mathematical Biosciences*, 2000: 168, 39–56.

[30] Sibert, A. & Austerlitz, F. & Heyer, E.: Wright-Fisher revisited: the case of fertility correlation. *Theoretical Population Biology*, 2002: 62, 181–197.

[31] Wakeley, J. : The coalescent in an island model of population subdivision with variation among demes. *Theoretical Population Biology*, 2001: 59, 133–144.

A Appendix

Let X_s be the random variable denoting the number of population at s^{th} generation, and let the state space φ be $\{N_1, N_2\}$. Let

$$P(X_s(t+1) = X_s(t) | X_s(t) = N_j) = 1 - q_{ij}, \quad (17)$$

$$P(X_s(t+1) \neq X_s(t) | X_s(t) = N_j) = q_{ij}. \quad (18)$$

According to the above description

$$P(X_s(t+1) = N_2 | X_s(t) = N_1) = q_{i1}, \quad (19)$$

$$P(X_s(t+1) = N_1 | X_s(t) = N_2) = q_{i2}, \quad (20)$$

and

$$P(X_s(0) = N_1) = p_{s1}^{(0)}. \quad (21)$$

It follows immediately that

$$P(X_s(t+1) = N_1 | X_s(t) = N_1) = 1 - q_{i1}, \quad (22)$$

$$P(X_s(t+1) = N_2 | X_s(t) = N_2) = 1 - q_{i2}, \quad (23)$$

and that the probability $p_{s2}^{(0)}$ of initially having N_2 individuals is given by

$$p_{s2}^{(0)} = P(X_0 = N_2) = 1 - p_{s1}^{(0)}. \quad (24)$$

From this information, we can compute $P(X_s = N_1)$ and $P(X_s = N_2)$. We observe that

$$\begin{aligned} &P(X_s(t+1) = N_1) = \\ &\quad P(X_s = N_1 \wedge X_s(t+1) = N_1) \\ &\qquad + P(X_s = N_2 \wedge X_s(t+1) = N_1) \\ &= P(X_s = N_1) P(X_s(t+1) = N_1 | X_s = N_1) \\ &\qquad + P(X_s = N_2) P(X_s(t+1) = N_1 | X_s = N_2) \\ &= (1 - q_{i1}) P(X_s = N_1) + q_{i2} P(X_s = N_2) \\ &= (1 - q_{i1}) P(X_s = N_1) + q_{i2}(1 - P(X_s = N_1)) \\ &= (1 - q_{i1} - q_{i2}) P(X_s = N_1) + q_{i2}. \end{aligned} \quad (25)$$

Now $P(X_0 = N_1) = p_{s1}^{(0)}$, so

$$P(X_1 = N_1) = (1 - q_{i1} - q_{i2}) p_{s1}^{(0)} + q_{i2} \quad (26)$$

and

$$\begin{aligned} &P(X_2 = N_1) = (1 - q_{i1} - q_{i2}) P(X_1 = N_1) + q_{i2} \\ &= (1 - q_{i1} - q_{i2})^2 p_{s1}^{(0)} + q_{i2}[1 + (1 - q_{i1} - q_{i2})] \end{aligned} \quad (27)$$

It is easily seen by repeating this procedure s times that

$$P(X_s = N_1) = (1 - q_{i1} - q_{i2})^s p_{s1}^{(0)} + q_{i2} \sum_{j=0}^{s-1} (1 - q_{i1} - q_{i2})^j. \quad (28)$$

As indicated in the paper let $0 < q_{ij} < 1$. Then by the formula for the sum of a finite geometric progression,

$$\sum_{j=0}^{s-1} (1 - q_{i1} - q_{i2})^j = \frac{1 - (1 - q_{i1} - q_{i2})^s}{q_{i1} + q_{i2}}. \quad (29)$$

Then we conclude from (28) that

$$P(X_n = N_1) = \frac{q_{i2}}{q_{i1} + q_{i2}} + (1 - q_{i1} - q_{i2})^s \left(p_{s1}^{(0)} - \frac{q_{i2}}{q_{i1} + q_{i2}} \right) \quad (30)$$

and consequently that

$$P(X_n = N_2) = \frac{q_{i1}}{q_{i1} + q_{i2}} + (1 - q_{i1} - q_{i2})^s \left(p_{s2}^{(0)} - \frac{q_{i1}}{q_{i1} + q_{i2}} \right). \quad (31)$$

As $0 < q_{i1} + q_{i2} < 2$, which implies that $|1 - q_{i1} - q_{i2}| < 1$, we can let $s \to \infty$ in (30) and (31) and conclude that

$$\lim_{s \to \infty} P(X_s = N_1) = \frac{q_{i2}}{q_{i1} + q_{i2}} \quad (32)$$

and

$$\lim_{s \to \infty} P(X_s = N_2) = \frac{q_{i1}}{q_{i1} + q_{i2}}. \quad (33)$$

It is clear from the previous three equations that we should have

$$(p_{s1}^{(0)}, p_{s2}^{(0)}) = \left(\frac{q_{s2}}{q_{s1} + q_{s2}}, \frac{q_{s1}}{q_{s1} + q_{s2}} \right) \quad (34)$$

as starting distributions for $P(X_s = N_1)$ and $P(X_s = N_1)$ to be independent of s.

B Appendix

A stochastic process can be defined as any collection of random variables X_t, $t \in T$, defined on a common probability space, where T is a subset of $(-\infty, \infty)$ and is thought of as the time-parameter set. The process is called a *continuous parameter process* if T is an interval having positive length and a *discrete parameter process* if T is a subset of the integers. If the random variables X_t all take on values from the fixed set φ, then φ is called the *state space* of the process.

Many systems of theoretical and applied interest have the property that given the present state, the past states have no influence on the future. This property is called the *Markov property*, and systems having this property are called *Markov chains*. The Markov property is defined precisely by the requirement that

$$P(X_{n+1} = x_{n+1} | X_0 = x_0, \ldots, X_n = x_n) = P(X_{n+1} = x_{n+1} | X_n = x_n) \quad (35)$$

for every choice of the nonnegative integer n and the numbers $x_0, \ldots, x_{n+1}$, each in φ. The conditional probabilities $P(X_{n+1} = y | X_n = x)$ are called the *transition probabilities* of the chain. If the transition probabilities are independent of n, they are called *stationary transition probabilities*.

A function $P(x, y)$ for $x, y \in \varphi$ is a *transition function* if it satisfies

$$P(x, y) \geq 0 \quad (36)$$

and

$$\sum_y P(x, y) = 1. \quad (37)$$

A function $p_s^{(0)}$ for $s \in \varphi$ is an *initial distribution* if it satisfies

$$p_s^{(0)} = P(X_0 = s), \quad (38)$$

$$p_s^{(0)} \geq 0 \quad \text{for all } s \in \varphi, \quad (39)$$

and

$$\sum_s p_s^{(0)} = 1. \quad (40)$$

Zero-sum problems in finite groups

Vesselin Dimitrov

under the direction of
Mr. Pavlo Pylyavskyy
Massachusetts Institute of Technology

Abstract

We develop new methods for investigating problems of zero-sum type in general finite groups. We establish a new bound on Davenport's constant for abelian groups that asymptotically improves the previously known bounds. We use tools from representation theory to study properties of zero-sum sequences through nilpotent ideals of group algebras. A new relationship between zero-sum problems and multidimensional covers of $\mathbb{Z}$ is also established.

1 Introduction

One of the most important problems in combinatorial group theory is that of characterizing the zero-sum sequences in general finite groups. This is a fundamental problem with applications well beyond group theory and additive number theory. A prime example is the proof [1] of the infinitude of the Carmichael numbers, where some knowledge of the zero-sum sequences in the group of units of the ring $\mathbb{Z}_n$ is essential. We refer the reader to [7] for a summary of the classical group-theoretic results and applications in this direction; for a discussion of the combinatorial aspects of the problem, see Caro's survey paper [6]. The goal of the present paper is to improve the previously best general upper bounds for certain *Ramsey-type* numbers concerning zero-sum sequences in finite abelian groups, as well as to introduce and apply a new approach based on the theory of group representations.

Let $G = \mathbb{Z}_{d_1} \oplus \mathbb{Z}_{d_2} \oplus \cdots \oplus \mathbb{Z}_{d_r}$ be a multiplicatively written finite abelian group, where $d_i | d_{i+1}$ for $1 \leq i < r$, and $d_1 > 1$. Define the quantity $d(G) = 1 + \sum_{j=1}^{r}(d_r - 1)$, and recall that r is the *rank* of G, $|G| = d_1 \dots d_r$ is the *order* of G, and $\exp G = d_r$ is the *exponent* of G.

A sequence of elements of G is called *zero-sum*[1] if the product of its elements is the identity element of G (which will usually be denoted by 1). We say that a zero-sum sequence is *minimal* if it does not contain proper nonempty zero-sum subsequences.

In 1966 Davenport [8] posed the problem of finding the maximal possible length of a minimal zero-sum sequence of G. Throughout the paper, we denote this quantity by $D(G)$, and call it the *Davenport constant* of G. Note that, alternatively, $D(G)$ is the minimal $s \in \mathbb{N}$ such that for any s (not necessarily distinct) elements $g_1, g_2, \dots, g_s$ of G there exists a nonempty set $I \subseteq \{1, 2, \dots, s\}$ such that $\prod_{j \in I} g_j = 1$.

There are numerous results showing the great importance of Davenport's constant in various areas of both number theory and group theory. Davenport's initial motivation for studying the properties of these numbers was based on his discovery [8] that if K is an algebraic number field and G is the ideal class group of K, then $D(G)$ is the maximal number of prime ideals (counting multiplicities) in the decomposition of an irreducible integer in K. More recent papers reveal deeper applications of $D(G)$ in the factorization theory of *Krull domains*; see [7] for such topics, as well as for further references of applications in graph theory and Ramsey theory. Yet another important application is given by Mariusz Skalba [21], who uses a slight generalization of the Davenport constant in his investigations on the prime factorization of numbers that are uniquely represented by certain quadratic forms.

Note that the inequality $D(G) \geq d(G)$ holds trivially, since if $\alpha_1, \dots, \alpha_r$ is a basis of G, then every sequence consisting of $d_i - 1$ copies of α_i for $1 \leq i \leq r$ does not contain zero-sum subsequences, but has length $d(G) - 1$. It was further conjectured (by Olson [17]) that $D(G) = d(G)$ for all abelian groups G, but this was proved false by several re-

[1] Although we are using multiplicative notation.

searchers [11, 14]. For easy reference, we include a list of known groups G with $D(G) = d(G)$ (see Appendix B). However, the conjecture holds under certain circumstances, the most famous being when G is a p-group [17] (that is, $|G|$ is a power of a prime) or G has rank 1 or 2 [18]. The exact evaluation in the general case is an open problem that seems to be extremely difficult. General upper bounds for $D(G)$ are therefore very important. One of the best current bounds, taken from [15], is

$$D(G) \le \exp G\Big(1 + \log \frac{|G|}{\exp G}\Big). \tag{1}$$

Our goal is to develop two new methods for obtaining better upper bounds.

2 Statement of Results

The paper consists of two parts. In the first part we obtain a result that improves (1) in almost all cases. In the second part we analyze certain nilpotent ideals of the group algebras $\mathbb{F}G$ to extract information about minimal zero-sum sequences of maximal length. For this purpose we use tools from representation theory.

2.1 Notations

Let us first introduce some notation that we will use consistently throughout the paper. In general, we consider G an arbitrary finite group, not necessarily abelian. If, in addition, G is abelian, we shall always write it as $G = \mathbb{Z}_{d_1} \oplus \cdots \oplus \mathbb{Z}_{d_r}$.

Define $\mathcal{F}(G)$ to be the free abelian semigroup over G. We can identify sequences of G with the elements of $\mathcal{F}(G)$. The multiplication of elements $a_1, \ldots, a_s$ will be denoted by $a_1 \times \ldots \times a_s = \coprod_{i=1}^{s} a_i$, to avoid confusion with multiplication in G. Also, if $S \in \mathcal{F}(G)$, let $\prod(S)$ be the usual product[2] of the elements of S. Each sequence of $\mathcal{F}(G)$ that can be obtained from $S \in \mathcal{F}(G)$ by successive replacements of elements of S with their product will be called a *minor* of S.

For a set $S \subseteq \mathbb{N}$, we define the *generalized Davenport constant* $D^{(S)}(G)$ to be the maximal length of a zero-sum sequence that contains no proper zero-sum subsequences having length in S. With $S = \mathbb{N}$ this reduces to the classical Davenport constant. If $S = \{1, 2, \ldots, m\}$, we denote $D^{(S)}(G)$ by $D^{(m)}(G)$. Note that $D^{(S)}(G)$ is finite if and only if S contains an element divisible by $\exp G$.

[2] By *usual product* we mean the group operation of G.

2.2 Results

If G is abelian, define $h(G) = D(G)/d(G)$ as a measure of the validity of Olson's conjecture [17] (which is precisely $h(G) = 1$). It remains an open problem whether $h(G)$ is bounded from above.[3] As a step towards the solution of this problem, we prove in Section 3 that $h(G)$ is bounded within the class of the groups of fixed rank r. For this purpose, we combine our Proposition 3.1.1 with a deep result of Alon and Dubiner [4]. This will enable us to give the following very good general estimate that stands as the best current general bound for groups with $d_1 \ge 3$.

Theorem 1 *Let $r \in \mathbb{N}$, $r \ge 2$. Then for all abelian groups G of rank r we have the inequality*

$$h(G) < \big(500r \ln r\big)^r. \tag{2}$$

Very little is known about Davenport's constant in nonabelian groups. To the best of our knowledge, the only significant result in this direction is a deep theorem of Olson [19] that asserts that $D^{\{|G|\}}(G) = 2|G| - 1$ for all finite groups G. In Section 3.3 we establish the following result.

Theorem 2 *Let G be a soluble group and C be a cyclic group, both of order n. Then the inequality*

$$D(G \oplus C) \le 2n - 1$$

holds.

This result is important because it gives some new information about Davenport's constant for soluble groups. It generalizes the result $D(\mathbb{Z}_n^2) = 2n-1$, and gives a simple proof of Olson's inequality $D^{\{|G|\}}(G) \le 2|G| - 1$ in the case of soluble groups.

In Section 4.1 we prove the following generalization of Olson's p-group theorem:

Theorem 3 *Let p be a prime and G be a p-group. Suppose that $m_1, \ldots, m_s \in \mathbb{N}$ are such that $\sum_{j=1}^{s} m_j \ge d(G)$. Consider arbitrary $g_1, \ldots, g_s \in G$. Then the equation*

$$g_1^{x_1} \ldots g_s^{x_s} = 1$$

has a nontrivial solution $x_1, \ldots, x_s$ satisfying $0 \le x_i \le m_i$ for all $i \in \{1, \ldots, s\}$. In particular, by setting each $m_i = 1$ and $s = d(G)$, we have $h(G) = 1$.

[3] In the known examples [14] of infinite families of groups G with $h(G) > 1$ the value of $h(G)$ is asymptotic to 1.

In Section 4.2 we introduce a new method using group characters. We reduce our conjecture $h(\mathbb{Z}_n^r) = 1$ to a simple problem involving covering systems. Furthermore, we establish an intimate relationship between zero-sums in G and covers of $\mathbb{Z}$, thus reducing the upper estimation of $D(G)$ to the construction of certain multidimensional covers of $\mathbb{Z}$. We pose a problem that seems to be of fundamental importance in both the estimation of Davenport's constant and the characterization of multidimensional equimodular covering systems. We note that a deep unification of *zero-sum problems*, *subset sums*, and *covers of* $\mathbb{Z}$ (the three main areas of additive number theory, each initiated by P. Erdős) was very recently established in [23]. However, the relationship we propose is independent and conceptually different.

3 Some upper estimates for Davenport's constant

3.1 The Multiplicative Machinery

In this section we set up the machinery that will enable us to prove our bound for Davenport's constant. We employ the generalized Davenport constant $D^{(S)}(G)$ as a tool.

Proposition 3.1.1 *Let G be any finite group and H be a normal subgroup of G. Then for all $m \in \mathbb{N}$, we have*

$$D(G) \le D^{(m)}(G/H) + m\big(D(H) - 1\big). \tag{3}$$

Moreover, under the additional assumption that $D^{(m)}(G/H) \le m + D(G/H)$, the inequality can be strengthened to

$$D(G) \le D(G/H) + m\big(D(H) - 1\big). \tag{4}$$

Proof. Since $H \lhd G$, there exists an endomorphism $\phi \in \mathrm{End}(G)$ with $H = \ker\phi$. Let $k = D(\ker\phi)$ and $l = D^{(m)}(G/\ker\phi)$, and suppose that we are given a sequence S in $\mathcal{F}(G)$ of length $s = l + (k-1)m$. The definition of l implies that S contains pairwise disjoint subsequences $S_1, \ldots, S_k \in \mathcal{F}(G)$, each having length at most m, such that $\prod(S_i) \in \ker\phi$ for all i. Since there are k such products $\prod(S_i)$, we have that $\coprod\big(\prod(S_1)\ldots\prod(S_k)\big)$ has a zero-sum subsequence in $\mathcal{F}(\ker\phi)$. Since it is also a minor of S, it follows that S has a zero-sum subsequence. But S was arbitrarily chosen, so $D(G) \le l + (k-1)m$. For the strengthening, note that under the additional assumption $l \le D(G/H) + m$, we would still be able to pick k pairwise nonintersecting subsequences of a sequence of length $D(G/H) + (k-1)m$. This completes the proof. □

3.2 Proof of Theorem 1.

The following deep result was proved by Alon and Dubiner [4]; we state it as a lemma.

Lemma 3.2.1 (Alon-Dubiner [4]) *Suppose that $r \ge 2$. Then*

$$D^{(\{n\})}(\mathbb{Z}_n^r) \le c(r)n \tag{5}$$

for all $n \in \mathbb{N}$, where the constants $c(r)$ are defined recursively by $c(1) = 2$ and

$$c(r) = 256(r\log_2 r + 5)c(r-1) + (r+1) \quad \textit{for } d \ge 2.$$

We shall combine this result with our Proposition 3.1.1 to prove Theorem 1.

We will first prove that

$$D(G) < c(r)s(G), \tag{6}$$

where by $s(G)$ we denote the quantity $\sum_{j=1}^{r} d_j$.

We induct on r, the rank of G. For $r = 1$ the assertion is trivial. Fix $r \ge 2$ and suppose that the assertion holds for smaller values of r. Let

$$H = \mathbb{Z}_{d_2/d_1} \oplus \cdots \oplus \mathbb{Z}_{d_r/d_1}.$$

Then $H < G$ and $G/H = \mathbb{Z}_{d_1}^r$. But H has rank at most $r-1$. Hence, we may apply the induction hypothesis to H to obtain the inequality

$$D(H) < c(r-1)s(H) = c(r-1)\left(\frac{s(G)}{d_1} - 1\right). \tag{7}$$

On the other hand, Lemma 3.2.1 gives

$$D^{(d_1)}(\mathbb{Z}_{d_1}^r) \le c(r)d_1. \tag{8}$$

If we now combine (7) and (8) with Proposition 3.1.1 and use that $c(r-1) < c(r)$, we get that

$$\begin{aligned} D(G) &\le D^{(d_1)}(\mathbb{Z}_{d_1}^r) + d_1(D(H) - 1) \\ &\le c(r)d_1 + c(r-1)(s(G) - d_1) \\ &< c(r)s(G), \end{aligned}$$

which completes the induction step, proving (6).

Next, note that $s(G) \geq 2r$, so $r - 1 < \frac{1}{2}s(G)$. Since $d(G) = s(G) - r + 1$, it follows that $d(G) > \frac{1}{2}s(G)$. Combining this with (6), we obtain

$$h(G) = \frac{D(G)}{d(G)} < 2\frac{D(G)}{s(G)} < 2c(r).$$

Now the assertion of the theorem may easily be deduced from (6) and the recurrence relation for the numbers $c(r)$ by using some (rough) estimates of the Alon-Dubiner constants. We skip this routine step. □

3.3 A bound in soluble groups: a discussion of Theorem 2

In the case when G is an abelian group, we can apply Proposition 3.1.1 to obtain an upper bound of $D^{(\{n\})}(G)$. The main result of this section is the proof of Theorem 2. We shall first demonstrate the importance of our result by deducing the following special case of a deep theorem of Olson [19].

Corollary 3.3.1 (Olson [19]) *If G is soluble and of finite order n, then $D^{(\{n\})}(G) \leq 2n - 1$.*

Proof. Embed G into $F = G \oplus C$, where C is the cyclic group of order $|G|$. Choose a generator $x \in C$; then x has order n and $x \in F$, but $x \notin G$. Consider a sequence $S = \{g_1, \ldots, g_{2n-1}\}$ of $2n-1$ elements of G, and consider the sequence $xg_1, \ldots, xg_{2n-1}$ in F. By Theorem 2, $D(F) \leq 2n - 1$. Hence, there exist some indices $i_1, \ldots, i_k$ satisfying $(xg_1)(xg_2)\ldots(xg_k) = 1$. Since x commutes with each g_i, we have $x^k g_1 \ldots g_k = 1$. Now it follows that $x^k = 1$ and $g_1 \ldots g_k = 1$. But x has order n, and hence $n|k$, which, in view of $k \leq 2n - 1$, leads to $k = n$. We therefore obtain the existence of exactly n indices with $g_1 \ldots g_n = 1$. Since S was chosen arbitrarily, the conclusion follows. □

Taking $G = \mathbb{Z}_n$, we recover the theorem of Erdős, Ginzburg and Ziv.

We now proceed to the proof of Theorem 2. We begin with the following famous lemma; we include the short proof for the sake of completeness.

Lemma 3.3.1 (Olson, [18]) *If p is a prime number, then $D^{(p)}(\mathbb{Z}_p^2) \leq 3p - 2$.*

Proof. Let $E = \mathbb{Z}_p^2$ and $F = \mathbb{Z}_p^3$. By Olson's theorem on p-groups, we have $D(E) = 2p - 1$ and $D(F) = 3p - 2$. Embed E into F. Let $x \in F, x \notin E$. Now if $g_1, \ldots, g_{3p-2} \in E$, some subsequence of $xg_1, \ldots, xg_{3p-2}$ has product 1. Since x has order p, this subsequence must contain either p or $2p$ elements. If it contains exactly p elements, there is nothing to show. So assume that it contains exactly $2p$ elements. Without loss of generality, we may consider these elements to be $g_1, \ldots, g_{2p}$. Now $g_1 \ldots g_{2p} = 1$. But $D(E) = 2p-1 < 2p$, so there is a proper subsequence of $g_1, \ldots, g_{2p}$, say $g_1, \ldots, g_k$, satisfying $g_1 \cdots g_k = 1$. If $k \leq p$, we can choose these elements. If $k > p$, we choose the elements $g_{k+1}, \ldots, g_{2p}$ whose product is also 1. This completes the proof. □

Proof of Theorem 2. Since G is soluble, there exists a series

$$\langle 1 \rangle = G_0 \lhd G_1 \lhd \cdots \lhd G_d = G$$

such that each factor group $C_i = G_i/G_{i-1}$ is cyclic of prime order.

We proceed by induction on d. If $d = 1$, then G is of the form $\mathbb{Z}_p$ with p prime, and the assertion follows from Olson's p-group theorem. So assume that $d \geq 2$ and that the assertion is true for smaller values of d. Let $G' = G_{d-1}$ and $C' = C/C_d$. Then $G' \lhd G$ and $C' \lhd C$, and $|G'| = |C'|$ with C' a cyclic group. Note that C/C' is a cyclic group of prime order; we shall write it in the form $\mathbb{Z}_p$ with p prime. We can invoke the induction hypothesis to obtain the inequality

$$D(G' \oplus C') \leq 2\frac{n}{p} - 1. \tag{9}$$

On the other hand, we know that $D(\mathbb{Z}_p^2) = 2p-1$ and $D^{(p)}(\mathbb{Z}_p^2) = 3p-2$, so we can apply the second part of Proposition 3.1.1 with $H \mapsto G \oplus C/\mathbb{Z}_p^2$, $G \mapsto G \oplus C$, and $m = p$ to get

$$D(G \oplus C) \leq D(\mathbb{Z}_p^2) + p(D(G' \oplus C') - 1). \tag{10}$$

Combining (9) and (10), we finally obtain

$$D(G \oplus C) \leq 2p - 1 + p\left(2\frac{n}{p} - 2\right) = 2n - 1,$$

as desired. This completes the induction and concludes the proof. □

4 Zero-sum sequences through group algebras

Let $\mathbb{F}$ be a field and G be a finite group. Consider the *group algebra* $\mathbb{A} = \mathbb{F}G$. For $a, b \in \mathbb{A}$ with $a =$

$\sum_{g\in G} a_g g$, $b = \sum_{g\in G} b_g g$, denote by $a \circ b$ the element $\sum_{g\in G} a_g b_g g \in \mathbb{A}$.

The idea of extracting information about zero-sum sequences from nilpotent ideals in group algebras dates back to Olson. In his classic paper [17], he proved that $h(G) = 1$ holds for p-groups by showing that

$$(1 - g_1) \cdot (1 - g_2) \cdot \ldots \cdot (1 - g_s) = 0$$

holds in $\mathbb{F}_p G$ provided that G is a p-group and $s \geq d(G)$. Here we view 1 as an element of $\mathbb{F}$.

Olson's p-group theorem would then follow, in view of the following easy lemma that is crucial for our discussion.

Proposition 4.0.1 *Fix $g_1, \ldots, g_s \in G$, and suppose that there exist s nonzero elements $a_1, \ldots, a_s \in \mathbb{F}^*$ such that*

$$(a_1 - g_1) \cdot (a_2 - g_2) \cdot \ldots \cdot (a_s - g_s) = 0$$

Then there exist indices $i_1 < i_2 < \cdots < i_k$ such that $g_{i_1} g_{i_2} \cdots g_{i_k} = 1$.

Proof. Assume the contrary, and write

$$J = (a_1 - g_1) \cdot (a_2 - g_2) \cdot \ldots \cdot (a_s - g_s)$$

in normal form as $\sum_{g\in G} k_g g$, where $k_g \in \mathbb{F}$. By our assumption, $k_1 = a_1 \ldots a_s \neq 0$, and this contradicts $J = 0$, completing the proof. □

In the special case when G is abelian, further techniques are available. For abelian G, let $\widehat{G}$ be the *dual group* of G. Any character $\chi : G \to \mathbb{C}^*$ can be extended to a *ring homomorphism* of $\mathbb{C}G$ by letting $\chi\big(\sum_{g\in G} k_g g\big) = \sum_{g\in G} k_g \chi(g)$.

The following proposition is a fundamental result of representation theory and is crucial for the method.

Proposition 4.0.2 *Let $\alpha \in \mathbb{C}G$. Then, $\alpha = 0$ if and only if $\chi(\alpha) = 0$ for all characters $\chi \in \widehat{G}$.*

Proof. If $\alpha = 0$, then $\chi(\alpha) = 0$ holds trivially for all characters χ. The nontrivial implication is to show that if $\chi(\alpha) = 0$ for all characters χ, then $\alpha = 0$.

Let $\alpha = \sum_{g\in G} k_g g$. We need to show that $k_g = 0$ for all $g \in G$. Pick an arbitrary such g. Put $n = |G|$.

Choose an arbitrary character χ. By assumption

$$\sum_{x\in G} k_x \chi(x) = 0.$$

Multiplying this equation by $\chi(g)^{-1} = \chi(g^{-1})$, we obtain

$$\sum_{x\in G\setminus\{g\}} k_x \chi(x)\chi(g^{-1}) = -k_g.$$

Now, summation over all characters $\chi \in \widehat{G}$ gives, in view of the second orthogonality relation for (abelian) group characters,

$$\begin{aligned} -nk_g &= \sum_{\chi\in\widehat{G}} \sum_{x\neq g} k_x \chi(x)\chi(g^{-1}) \\ &= \sum_{x\neq g} k_x \sum_{\chi\in\widehat{G}} \chi(x)\chi(g^{-1}) = 0, \end{aligned}$$

and we thus conclude that $k_g = 0$. But g was chosen arbitrarily, hence $\alpha = 0$. The proof is complete. □

4.1 The ideal of nilpotent elements of a group algebra

In this section we prove Theorem 3. Let p be a prime, and consider the group algebra $\mathbb{A} = \mathbb{F}_p G$ of G over $\mathbb{F}_p$. Here G is a finite abelian p-group. Let U be a fixed set of generators of G. For $H \subseteq G$, denote by $U_H \subseteq U$ the set of elements of U that appear in the decomposition of some element of H.

Let

$$\mathcal{S}_k(H) = \{(1 - h_1) \cdots (1 - h_k) |\, h_1, \ldots, h_k \in H\},$$

and let $I_H^{(k)}$ be the ideal generated by the elements of $\mathcal{S}_k(U_H) \subseteq \mathbb{A}$.

Proposition 4.1.1 *For all $k \in \mathbb{N}$ and $H \subseteq G$ we have $\mathcal{S}_k(H) \subseteq I_H^{(k)}$.*

Proof. It follows from the fact that $1 - xy$ lies in the ideal generated by $1 - x$ and $1 - y$, which, in turn, is implied by the identity $1 - xy = (1 - x) + x(1 - y)$. □

We now proceed to the

Proof of Theorem 3. We shall first prove that if $t \geq d(G)$, then $\mathcal{S}_t(G) \equiv \{0\}$. By Proposition 4.1.1, $\mathcal{S}_t(G)$ is a subset of the ideal generated by the elements of $\mathcal{S}_t(U_G)$, so it is sufficient to prove that $\mathcal{S}_t(U_G) \equiv \{0\}$. That is, we have to prove that the polynomial

$$f(x_1, \ldots, x_t) = (1 - x_1)(1 - x_2) \cdots (1 - x_t)$$

vanishes in $\mathbb{F}_p G$ whenever all of its variables take values in U. But in this case, by the pigeonhole principle, there exists an index i for which at least p^{e_i} variables are set to α_i, and the claim follows from

$$(1 - \alpha_i)^{p^{e_i}} = 1 + (-1)^p \alpha_i^{p^{e_i}} = 1 + (-1)^p = 0.$$

It now follows that

$$J := (1-g_1)^{m_1}(1-g_2)^{m_2}\cdots(1-g_s)^{m_s} = 0$$

for all $g_1, \ldots, g_s \in G$, since $J \in \mathcal{S}_t(G)$ for

$$t = \sum_{j=1}^{s} m_j \geq d(G).$$

Note that, on the other hand, the expression

$$(1-g_1)^{m_1}\cdots(1-g_s)^{m_s}$$

is a linear combination of terms of the form $g_1^{x_1}\ldots g_s^{x_s}$ with $0 \leq x_i \leq m_i$, $1 \leq i \leq s$. If our equation does not have a solution, then the coefficient of 1 is $1 \neq 0$, which contradicts $f(g_1, \ldots, g_s) = 0$. Hence, our equation does have a solution. □

4.2 A new approach to zero-sum problems

In this section we propose a new method for investigating the general problem of bounding $D^{(S)}(G)$ using representation theory.[4]

4.2.1 The Main Theorem

We reduce the problem of estimating $D^{(S)}(G)$ to a simpler problem involving *covering systems*. We first need to give some definitions concerning covers. In what follows, $\mathbb{F}$ is an arbitrary algebraically closed field of characteristic 0, and we consider 1 as an element of $\mathbb{F}$. Also, if $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n)$ are vectors, by $x \cdot y$ we denote the usual *dot product* $x \cdot y = x_1y_1 + x_2y_2 + \cdots + x_ny_n$.

Definition 4.2.1.1 *Let $M = (v_{i,j})$ $(1 \leq i \leq m;$ $1 \leq j \leq n)$ be an $m \times n$ matrix with entries in $\mathbb{F}^k$. A* cover *of M is a set $\mathcal{P} = \{w_1, \ldots, w_n\}$ of vectors in $\mathbb{F}^k$ such that for each $i \in \{1, \ldots, m\}$ there exists a $j \in \{1, \ldots, n\}$ with $v_{i,j} \cdot w_j \in \{0, 1\}$. We shall say that $\mathcal{P}$ is a proper cover if $v_{i,j} \cdot w_j = 1$ for all i and the chosen j.*

Given this definition, we can now define *covers of finite sets.*

Definition 4.2.1.2 *Let $S \subseteq \mathbb{F}^k$ be a finite set of elements of $\mathbb{F}^k$. We say that S admits an (m, n, k)-cover if all $m \times n$ matrices M with entries in S have a proper cover.*

[4]See Appendix A for basic definitions and results.

Since $\mathbb{F}$ is an algebraically closed field, it contains a primitive n^{th} root of unity. Let U_n be the set of all n^{th} roots of unity.

In this section, we prove the following result.

Theorem 4 *If U_n admits an (n^r, l, k)-cover, then $lk \geq D^{(l)}(\mathbb{Z}_n^r)$.*

We will first briefly discuss the importance of this theorem that is hidden behind the formulation. We reduced the problem of bounding the generalized Davenport constant of abelian groups[5] from above to that of constructing covers of U_n. As we shall see in Section 4.4, the special case $k = 1$ can be reformulated in terms of classical covers of $\mathbb{Z}$, another area of additive number theory initiated by Paul Erdős. This relationship is itself worth attention, since the first unification of the covers and zero-sum problems was given only very recently [23], and yet here is another relationship.

4.3 Proof of the main theorem

Fix $n, r \in \mathbb{N}$ Let $\mathbb{A}$ be the group algebra $\mathbb{F}G$ of G over the field $\mathbb{F}$.

The idea is to use tools from representation theory in analyzing products of the form

$$J = (1-w_1)\cdot(1-w_2)\cdot\ldots\cdot(1-w_n)$$

for variable $w_1, \ldots, w_n \in \mathbb{A}$ that fulfil certain conditions.

Proof of Theorem 4. Since $G = \mathbb{Z}_n^r$ is abelian, we have $\chi(g) \in U_n$ for each character $\chi \in \widehat{G}$ and each $g \in G$. We can therefore view the character table of G as an $n^r \times n^r$ matrix with elements in U_n. We need to prove that, given kl elements $g_1, \ldots, g_{kl} \in G = \mathbb{Z}_n^r$, there exist indices $i_1, \ldots, i_s$ with $g_{i_1}\ldots g_{i_s} = 1$. Consider the columns of the character table of G that correspond to the elements $g_1, \ldots, g_{kl}$. They form an $n^r \times kl$ matrix T. Construct an $n^r \times l$ matrix $\widehat{T}$ with elements in U_n^k, obtained from T by replacing, in each row, the elements with indices $jl+1, \ldots, jl+k$ $(0 \leq j \leq k-1)$ by a k-tuple that will be regarded as an entry of $\widehat{T}$.

Since U_n admits an (n^r, l, k)-cover, $\widehat{T}$ has a cover, say $\mathcal{P} = \{v_1 \ldots, v_l\}$; let $v_i = (v_i^1, \ldots, v_i^k) \in U_n^k$.

[5]An upper bound of $\mathbb{Z}_n^r$ automatically produces an upper bound of any abelian group of rank r and exponent n.

We now define

$$w_i = \sum_{j=(i-1)l+1}^{(i-1)l+k} v_i^j g_j.$$

Since $\mathcal{P}$ is a proper cover of $\widehat{T}$, we have that, for each character $\chi \in \widehat{G}$, there exists an index $i = i(\chi)$ such that $\chi(w_i) = 1$.

Now, consider the element

$$J = (1 - w_1) \cdot (1 - w_2) \cdot \ldots \cdot (1 - w_l) \tag{11}$$

of the group algebra $\mathbb{A}$. We claim that $J = 0$. Indeed, by Proposition 4.0.2 it suffices to show that $\chi(J) = 0$ for all irreducible characters χ. But, as shown above, there always exists an index $i = i(\chi)$ with $1-\chi(w_i) = 0$. Hence, the claim will follow at once if we show that

$$\chi(J) = \prod_{i=1}^{l}(1 - \chi(w_i)), \tag{12}$$

which holds trivially since χ is a group homomorphism. Now, the assertion follows from a slight modification of Proposition 4.0.1. □

4.4 Multidimensional covering systems of $\mathbb{Z}$ and Olson's conjecture for groups of the form $\mathbb{Z}_n^r$

As we saw in Section 4.2, zero-sum problems are intimately related to covering systems. In this section, we show that these covering systems can be further reduced to the classical multidimensional covers of $\mathbb{Z}$.

Recall the definition of the classical multidimensional covers of $\mathbb{Z}$.

Definition 4.4.0.3 *A system of congruences*

$$a_{i0} + \sum_{j=1}^{r} a_{ij}x_j \equiv 0 \pmod{n_i} \tag{13}$$

for $1 \le i \le m$ covers a set $S \subseteq \mathbb{Z}^k$ if every vector $(x_1, \ldots, x_k) \in S$ satisfies at least one of the congruences (13). If $S = \mathbb{Z}^k$, the system (13) is called a covering system.

4.4.1 The case of $k = 1$

We can identify each system of the form (13) with a triple $S = (M, R; V)$, where $M = \left(a_{ij}\right)$ $(1 \le i \le m,$ $1 \le j \le r)$ is an $m \times r$ matrix, $R = (a_{10}, \ldots, a_{m0})$, and $V = (n_1, \ldots, n_m)$ is the sequence of moduli. We shall call M the *matrix* of the covering system S, and R the *residue vector* of S.

We will consider covering systems with equal moduli, i.e., with each $n_i = n$. Let us fix n and write (M, R) for the covering system $(M, R; V)$, $V = (n, \ldots, n)$.

Call a finite-dimensional matrix M (over $\mathbb{Z}_n$) a *cover* if it is the matrix of some covering system.

In the fundamental problem of the classification of the equimodular multidimensional covering systems, it is natural to ask for a description of all matrices that are covers.

We pose the following very natural problem:

Problem. For fixed $n, r \in \mathbb{N}$, what is the minimal integer $m = c(n, r)$ such that all $m \times r$ matrices over $\mathbb{Z}_n$ are covers?

The importance of the problem of estimating $c(n, r)$ is twofold. Firstly, this is a very simple and naturally arising question whose solution will be a key advance in the problem of classifying the multidimensional covering systems of equal moduli. Secondly, and most importantly for our purposes, an estimate for $c(m, k)$ would automatically produce an estimate for $D(\mathbb{Z}_m^k)$:

Theorem 5 *For all $n, r \in \mathbb{N}$, the inequality $D(\mathbb{Z}_n^r) \le c(n, r)$ holds.*

Proof. Choose $\mathbb{F} = \mathbb{C}$, and let $G = \mathbb{Z}_n^r$. We shall slightly modify the proof of Theorem 4. Recall that it will suffice to prove that if $s = c(n, r)$, then, for any fixed $g_1, \ldots, g_s \in G$, there exist $a_1, \ldots, a_s \in \mathbb{C}$ such that for each character $\chi \in \widehat{G}$ there is an index $i = i(\chi)$ with $a_i = \chi(g_i)$.

Let us choose a basis $\alpha_1, \ldots, \alpha_r$ of G and fix a primitive n^{th} root of unity ζ. Then each character χ is of the form

$$\chi\left(\alpha_1^{b_1} \ldots \alpha_r^{b_r}\right) = \zeta^{b_1x_1 + \cdots + b_rx_r}$$

for some $x_1, \ldots, x_r \in \mathbb{Z}_n$. Conversely, each such function is a group character.

Of course, we will look for a_i of the form $a_i = \zeta^{c_i}$ with $c_i \in \mathbb{Z}_n$. Define an $s \times r$ matrix $M = \left(a_{ij}\right)$ over $\mathbb{Z}_n$ by $g_i = \alpha_1^{a_{i1}} \cdots \alpha_r^{a_{ir}}$. Since $s = c(n, r)$, M is a cover, and hence it has a residue vector R such that $S = (M, R)$ is a covering system. If $R = (d_1, \ldots, d_r)$, we simply choose $a_i = \zeta^{-c_i}$.

Since S is a covering system, we know that, for each vector $(x_1, \ldots x_r)$, there exists an index i with

$$\sum_{1 \le j \le r} a_{ij}x_j \equiv -c_i \pmod{n}. \tag{14}$$

Now let χ be any character, and let $(x_1, \ldots, x_r)$ be the corresponding n-tuple. By (14), there exists an i for which

$$\chi\big(\alpha_1^{a_{i1}} \ldots \alpha_r^{a_{ir}}\big) = \zeta^{-d_i} = a_i.$$

This is exactly what we needed to show, and the conclusion follows. □

We conjecture that the expected exact evaluation holds, i.e., that $c(n, r) = r(n-1) + 1$. This would imply Olson's conjecture that $h(\mathbb{Z}_n^r) = 1$.

5 Conclusion

We have developed two general methods for bounding Davenport's constant of finite soluble groups. One of these methods led to an improvement of the current best general upper bound for abelian groups of fixed rank and order. Moreover, this bound implies a slightly weaker version of the asymptotic analogue of Olson's conjecture, and the investigation of the analogue itself is a topic for further studies. We also intend to investigate the upper estimation of Davenport's constant for nonabelian groups through the apparatus of representation theory.

6 Acknowledgments

I would like to thank Mr. Pavlo Pylyavskyy from the Massachusetts Institute of Technology for directing the research. My deep gratitude to Prof. Hartley Rogers for finding such a mentor for me. My special thanks to my tutor Dr. John Rickert for his ongoing assistance during the program, and to Dr. Jenny Sendova for her encouragement and help. I am greatly indebted to Mr. Gabriel Carroll for his expert, detailed analysis of my work and his numerous helpful suggestions and remarks that greatly improved the paper. Thanks go to Chris Mihelich and Sanghamitra Sen for their technical help and on-time e-mails. I am also indebted to the Center for Excellence in Education for providing the opportunity to do this research.

References

[1] Alfred, W.R., A. Granville, C. Pomerance: There are infinitely many Carmichael numbers, *Annals of Mathematics* **2** 139 (1994), no. 3, 703–722.

[2] Alon, N.: Combinatorial Nullstellensatz, *Combin. Probab. Comput.* **8** (1999), 7–29.

[3] Alon, N., M. Dubiner: Zero-sum sets of prescribed size, *Combinatorics, Paul Erdős is Eighty*, Janos Bolyai Math. Soc., Budapest, 1993, 33–50.

[4] Alon, N., M. Dubiner: A lattice point problem and additive number theory *Combinatorica* **15** (1995), 301–309.

[5] Baker, Roger C., Wolfgang M. Schmidt: Diophantine Problems in Variables Restricted to the Values 0 and 1, *Journal of Number Theory*, **12** (1980), 460–486.

[6] Caro, Yair: Zero-sum problems — A survey, *Discrete Mathematics* **152** (1996), 93–113.

[7] Chapman, Scott T.: On the Davenport Constant, the Cross Number, and Their Application in Factorization Theory, *Zero-dimensional commutative rings*, Lecture notes in Pure and Applied Math., vol. 171, 167–190.

[8] Davenport, H.: Proceedings of the Midwestern Conference on Group Theory and Number Theory. *Ohio State University*, April 1966.

[9] Doerk, Klaus, Trevor Hawkes: Finite Soluble Groups, New York: de Gruyter, 1992.

[10] Erdős P., A. Ginzburg, A. Ziv: Theorem in additive number theory, *Bull. Research Council Israel* **10** (1961), 41–43.

[11] Geroldinger, Alfred: On Davenport's Constant, *Journal of Combinatorial Theory*, Series A **61**, 1992, 147–152.

[12] Hill, Victor E.: Groups and Characters *Chapman & Hall, 2000*

[13] Kemnitz, A.: On a lattice point problem, *Ars Combin.* **16** (1983), 151–160.

[14] Mazur, Marcin: A note on the growth of Davenport's constant, *Manuscripta Mathematica* **74**, (1992), 229–235

[15] Meshulam, Roy: An uncertainty inequality and zero subsums, *Discrete Mathematics* **84** (1990) 197–200.

[16] Nathanson, Melvyn B.: Additive Number Theory — *Inverse Problems and the Geometry of Sumsets*, *Graduate Texts in Mathematics*, Springer-Verlag, 1996/

[17] Olson, John E.: A Combinatorial Problem in Finite Abelian Groups, I, *Journal of Number Theory*, **1** 1969, 8–10.

[18] Olson, John E.: A combinatorial Problem in Finite Abelian Groups, II, *Journal of Number Theory*, **1** 1969, 195–199.

[19] Olson, John E.: On a combinatorial problem of Erdős, Ginzburg and Ziv, *Journal of Number Theory*, **8** 1976, 52–57.

[20] Schmidt, Wolfgang M.: Equations over Finite Fields, An Elementary Approach, *Lecture Notes in Mathematics*, Springer-Verlag, 1976.

[21] Skalba, Mariusz: On numbers with unique representation by a binary quadratic form, *Acta Arithmetica*, **64** (1993), 59–68.

[22] Skalba, Mariusz: On the Relative Davenport Constant, *European Journal of Combinatorics*, **19** (1998), 221–225.

[23] Sun, Zhi-Wei: Unification of zero-sum problems, subset sums and covers of $\mathbb{Z}$, *Electronic Research Announcements of the American Mathematical Society*, **9** (2003), 51–60.

A Group characters

A.1 Group representations

Definition 1 *Let G be a finite group and $\mathbb{F}$ be a field. Consider a finite-dimensional linear space V over $\mathbb{F}$. A* group representation of G over F *is a homomorphism $\phi : G \to GL(V)$. The dimension* $\dim(V)$ *of V is called the* degree *of the representation ϕ.*

Definition 2 (Equivalence of representations) *Two representations ϕ, ψ of a finite group G over $\mathbb{F}$ with corresponding linear spaces V, V' are* equivalent *if there exists a bijection $\tau : GL(V') \to GL(V)$ such that $\psi(g) = \tau^{-1}\phi(g)\tau$ for all $g \in G$. In particular, equivalent representations have the same degrees.*

Definition 3 (Subrepresentations) *If ϕ is a representation of G corresponding to the linear space V and ρ is a representation corresponding to a subspace V' of V, then ρ is called a* subrepresentation *of ϕ.*

Definition 4 (Irreducible representations) *A representation ϕ of G of degree d is* reducible *if it has a subrepresentation of degree d' satisfying $0 < d' < d$. Otherwise ϕ is* irreducible.

Denote by

$$\widehat{G} = \mathrm{Hom}(G, \mathbb{F}^*)$$

the set of equivalence classes of irreducible representations of G. Then $|\widehat{G}|$ equals the number of equivalence classes of G. In the special case when G is abelian, $\widehat{G}$ is a group called the *dual group* of G.

A.2 Characters

If ϕ is a representation of the finite group G over the field $\mathbb{F}$, its *character* is defined as

$$\chi(g) = \mathrm{tr}(\rho(g)), \qquad g \in G,$$

where tr denotes the *trace* of the matrix $\rho(x)$ (that is, the sum of the elements of its main diagonal). A character χ is said to be *irreducible* if it is obtained from an irreducible representation. For a representation $\alpha \in \widehat{G}$, denote by $\chi^{(\alpha)}$ the character of α.

Basic properties of characters:

1. Equivalent representations have the same character.
2. Each character is a *class function* on G, i.e., it is constant on conjugacy classes:
$$\chi(y^{-1}xy) = \chi(x) \quad \text{for all } x, y \in G.$$
3. If ϕ, ψ are representations of G with characters χ_1, χ_2, then the characters of $\phi \oplus \psi$ and the *tensor product representation* $\phi \otimes \psi$ are $\chi_1 + \chi_2$ and $\chi_1\chi_2$, respectively.

Characters are mainly considered over *algebraically closed fields* $\mathbb{F}$. For simplicity, from now on we consider characters over the field of complex numbers $\mathbb{C}$.

Theorem A.1 *The finite group G is abelian if and only if all irreducible characters of G have degree 1.*

Theorem A.2 *Every character χ is of the form*

$$\chi = \sum_{\alpha \in \widehat{G}} n_\alpha \chi^{(\alpha)}$$

for nonnegative integers n_α. Also, every such sum is the character of some representation.

A.3 Characters of Abelian groups

In the case when G is abelian, the irreducible characters of G are easily described. The irreducible characters in abelian groups always have degree 1 and their number is $|G|$.

Theorem A.3 (Orthogonality relations) *Let G be a finite abelian group of order n. Then:*

1. *For any characters $\chi, \psi \in \widehat{G}$,*

$$\frac{1}{n}\sum_{x\in G}\chi(x)\psi(x^{-1}) = \begin{cases} 1 & \text{if } \chi = \psi, \\ 0 & \text{if } \chi \neq \psi. \end{cases}$$

2. *For any $x, y \in G$,*

$$\frac{1}{n}\sum_{\chi\in\widehat{G}}\chi(x)\chi(y^{-1}) = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{if } x \neq y. \end{cases}$$

B The known groups for which $h(G) = 1$

Here is a list of the known parametric families of abelian groups G satisfying the Olson conjecture; the list is taken from [14].

Olson's conjecture $h(G) = 1$ holds if:

1. G is a p-group [17].
2. $\text{rank}(G) \leq 2$ [18].
3. $G = \mathbb{Z}_{2p^n} \oplus \mathbb{Z}_{2p} \oplus \mathbb{Z}_{2p}$ for p prime (Emde Boas).
4. $G = H \oplus \mathbb{Z}_{mp^n}$ where H is a p-group and $d(H) \leq p^n$ (Emde Boas).
5. $G = \mathbb{Z}_2 \oplus \mathbb{Z}_{2na} \oplus \mathbb{Z}_{2nb}$ where $n = 2^t 3^u 5^v 7^w$ and either $a = 1$ or $a = p^r$, $b = p^s$ with p prime (Emde Boas).
6. $G = \mathbb{Z}_2^3 \oplus \mathbb{Z}_{2m}$ with odd m (Baayen).
7. $G = \mathbb{Z}_2^3 \oplus \mathbb{Z}_{6m}$ with $3 \nmid m$ (Emde Boas, Kruyswijk).
8. $G = \mathbb{Z}_{3\cdot 2^n} \oplus \mathbb{Z}_{3\cdot 2^m} \oplus \mathbb{Z}_{3\cdot 2^s}$ (Emde Boas, Kruyswijk).
9. $G = \mathbb{Z}_3 \oplus \mathbb{Z}_{6na} \oplus \mathbb{Z}_{6nb}$ with n, a, b as in case 5 (Emde Boas).

An Improved Quantum Algorithm for Searching an Ordered List

Brian Jacokes

under the direction of
Dr. Andrew Landahl
Massachusetts Institute of Technology

Abstract

The best known quantum algorithm for searching an ordered list of N items is recursive and requires $3\log_{52} N \approx 0.526\log_2 N$ queries. We find an improved base case for this recursion, yielding an algorithm that requires only $4\log_{434} N \approx 0.457\log_2 N$ queries. This algorithm is more than twice as fast as the best possible classical algorithm, which requires $\log_2 N$ queries.

1 Introduction

A common computing problem is to find a target element in a list of size N, returning the index of that element as the output. The goal of algorithms for this problem is to find the target element as quickly as possible. One measurement of speed for such algorithms is how many queries the algorithm makes, where a query is an operation which gathers information on elements in the list.

The class of searching problems is divided into the problems of searching unordered and ordered lists. In both problems, querying an element of the list results in obtaining a one-bit answer. However, in searching unordered lists, a query only returns whether the element queried is the target element or not, while in searching ordered lists, a query returns whether or not the element queried is after the target element in the list. Thus, extra information is returned by queries to an ordered list, resulting in faster algorithms than those for unordered lists.

Although searching an unordered list can be performed in a provably optimal $O(N)$ queries on classical computers, it was shown in [1] that quantum computers can perform the same computation in $O(\sqrt{N})$ queries, which was shown in [2] to be the quantum lower bound for the problem. Searching an ordered list can be performed classically in $\log_2 N$ queries, which is also provably optimal. Quantum algorithms have been found which solve the same problem in both $\log_3 N \approx 0.631\log_2 N$ queries [3] and $3\log_{52} N \approx 0.526\log_2 N$ queries [4]. A quantum lower bound for searching an ordered list was shown in [3] to be $\frac{1}{\pi}(\ln N - 1) \approx 0.221\log_2 N$ queries, suggesting that a better quantum algorithm could exist for the problem.

In this paper, we build on methods presented in [4], which were used to obtain the previous upper bound of $0.526\log_2 N$ queries for searching an ordered list. These methods include a series of equivalences, the last of which is to find a function of multiple variables that obeys certain properties. We define a cost function whose minimization will result in finding such a function and then present methods for minimizing this cost function. Using our methods, we find a new algorithm that solves the problem in $4\log_{434} N \approx 0.457\log_2 N$ queries, providing an improvement over the previous upper bound.

2 Background

Although quantum computing is still in its infancy, quantum algorithms have provided speedups over classical algorithms for many problems. This speedup is due in part to the ability of quantum systems to be in superpositions of multiple states, a phenomenon known as working in *quantum parallel* [5].

2.1 Quantum computation

Classical computers encode information using bits, which may take on values of either 0 or 1. Thus, any system of n classical bits represents exactly one state: a string of 0s and 1s chosen from among 2^n possibilities. However, if we were instead using quantum bits, or *qubits*, then the state of this n-qubit system, denoted $|\Psi\rangle$, would be given by a superposition of the 2^n classical states, each having a complex coefficient a_i:

$$|\Psi\rangle = \sum_{i=0}^{2^n-1} a_i|i\rangle, \tag{1}$$

with

$$\sum_{i=0}^{2^n-1} |a_i|^2 = 1. \tag{2}$$

In this notation, $|i\rangle$ is the classical state with a string of N 0s and 1s corresponding to the binary representation of the number i. The 2^n states in this sum are mutually orthonormal, so the *inner product*, or overlap, of these states is defined to be $\langle i|j\rangle = \delta_{ij}$. Therefore, the set of states $\{|0\rangle, |1\rangle, \ldots, |2^n - 1\rangle\}$ forms a basis for the state of the n qubits, which we call the *computational basis* [6].

Although a system of qubits can store information on multiple states at once through the a_is, we can only obtain information about the the state $|\Psi\rangle$ by performing a measurement on the system. For the purpose of this paper, we will only deal with measurements that project $|\Psi\rangle$ onto the computational basis, although a projection onto any set of 2^n orthonormal states may be performed by applying the correct operations before and after the measurement. When a measurement is performed, the probability of obtaining the state $|i\rangle$ from the measurement is $|\langle\Psi|i\rangle|^2 = |a_i|^2$. Since every such measurement results in obtaining exactly one of the computational basis states, this directly implies (2). However, any measurement also collapses the state of the system: if the state of the system is measured as $|i\rangle$, then after the measurement, the new state of the system is $|i\rangle$, and the old state $|\Psi\rangle$ is lost [5, 6].

Quantum gates are unitary operators that take a state $|\Psi\rangle$ to a new state $U|\Psi\rangle$. An operator U is unitary if and only if it preserves inner products, meaning that

$$\langle U\Psi|U\phi\rangle = \langle\Psi|\phi\rangle. \tag{3}$$

Thus, these operators may also be seen as rotations in 2^n-space. It follows from (3) that unitary operators are invertible and therefore preserve all of the information from the original state, which is a requirement for quantum operators that are not measurements. Quantum algorithms use unitary operators to transform an initial state into a final state whose measurement will yield a useful result to high or unit probability [5, 6].

2.2 Ordered Search

The ordered search problem is defined as follows:

> Ordered Search: *Given a ordered list of N elements, and an element that appears exactly once in the list, find the index w of this element.*

A simple way to characterize the complexity of this problem is to ask how many times one must query the list to determine w. Thus, one may ignore how the list is actually stored, and instead consider a model of the problem in which a black box, called an *oracle*, tells one whether or not a position is after w. In classical computing, the oracle is a function defined as

$$f_w(i) = \begin{cases} -1 & \text{if } i < w; \\ 1 & \text{if } i \geq w. \end{cases} \tag{4}$$

In quantum computing, this oracle is elevated to an operator F_w such that

$$F_w|i\rangle = f_w(i)|i\rangle \tag{5}$$

for any computational basis state $|i\rangle$. The goal of an algorithm that uses the oracle model is not to construct the oracle, but rather to minimize the number of queries to the oracle once it is given [4].

Classical computers can solve the ordered search problem in $\log_2 N$ queries by the well-known binary-search algorithm. In this algorithm, a query is made to f_w at the center of the list, dividing the space to be searched by 2, and the process is repeated. Quantum computers, working in quantum parallel, can use F_w to operate on a superposition of states. This may be used to improve on the $\log_2 N$ queries required by the binary search.

There are many applications in which ordered search is used, the most prominent of which is database search. The ordered-search problem has also been shown to have implications in the related problems of inserting elements into and testing for element distinction in ordered lists [3]. Because of how often this problem appears, it is of interest to

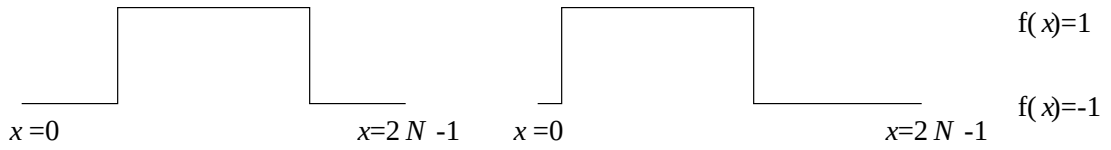


Figure 1: Two possible oracles. Each is equivalent to the other if shifted cyclically, demonstrating that they are translates of each other. The jumps in each are located at $x = w$, and the drops are located at $x = w+N$.

find an improved algorithm for it, despite the fact that only constant-factor improvements are possible. It is also of theoretical interest, as methods for finding ordered-search algorithms may carry over to improved algorithms for other problems.

2.3 Translation-Invariant Algorithms

In [4], Farhi et al. give a method for finding quantum algorithms for the ordered-search problem and find exact algorithms which solve the problem for 6 elements in 2 queries and for 52 elements in 3 queries. One may use these base cases to create a recursive solution to the general ordered-search problem. Using the algorithm for ordered search of 52 elements in 3 queries, one may repeatedly divide a list of arbitrary size into 52 equal groups and perform the algorithm on the first element of each group, dividing the search space by 52. By repeatedly performing this process, the ordered search problem can thus be solved in $3\log_{52} N \approx 0.526\log_2 N$ queries. Note that the base-case algorithms must be exact, because if they fail with some probability $P \neq 0$, then the general algorithm will succeed only with probability $(1 - P^{\log_{52} N})$, which goes to zero for N large.

The method used for finding such algorithms considers a slightly different oracle, which is given by

$$F'_w|x\rangle = \begin{cases} F_w|x\rangle & \text{if } 0 \leq x \leq N-1, \\ -F_w|x\rangle & \text{if } N \leq x \leq 2N-1. \end{cases} \tag{6}$$

Clearly, this oracle contains no more information than the one in (5) and can be easily constructed from the first. Using the oracle operator, all k-query algorithms which start in a state $|s\rangle$ will end in a state

$$V_k F'_w V_{k-1} F'_w \dots V_1 F'_w |s\rangle, \tag{7}$$

where the V_ℓ are unitary operators applied by the algorithm. Since it is the goal of our algorithm to determine the value of w, we must choose the V_ℓ such that the ending states for (7) are mutually orthogonal for $w = 0, 1, \dots, N-1$, as this is the only way to distinguish the oracles from each other with unit probability. Thus, the task is to find operators V_ℓ which will satisfy this orthogonality condition.

The key idea in [4], which motivates the definition of the oracle in (6), is to restrict the search to operators V_ℓ which are *translationally invariant*. Consider a translation operator T which is defined as

$$T|x\rangle = \begin{cases} |x+1\rangle & \text{if } 0 \leq x \leq 2N-2, \\ |0\rangle & \text{if } x = 2N-1. \end{cases} \tag{8}$$

We define an operator A to be translationally invariant if $TA|\Psi\rangle = AT|\Psi\rangle$, or in other words, if $TAT^{-1} = A$. Note that the oracles are translates of each other because $TF'_wT^{-1} = F'_{w+1}$ (Figure 1). One of the advantages of restricting attention to translationally invariant algorithms is that if one demands that a successful algorithm start in the state

$$|s\rangle = \frac{1}{\sqrt{2N}} \sum_{x=0}^{2N-1} |x\rangle \tag{9}$$

and end in the state

$$|w+\rangle = \frac{1}{\sqrt{2}}(|w\rangle + |w+N\rangle) \quad \text{if } k \text{ is odd}, \tag{10}$$
$$|w-\rangle = \frac{1}{\sqrt{2}}(|w\rangle - |w+N\rangle) \quad \text{if } k \text{ is even},$$

then $T|s\rangle = |s\rangle$ and $T^w|0\pm\rangle = |w\pm\rangle$, so that if the algorithm is successful for $w = 0$, namely

$$|0\pm\rangle = V_k F'_0 V_{k-1} F'_0 \dots V_1 F'_0 |s\rangle, \tag{11}$$

then the algorithm holds for any w using the same V_ℓ:

$$|w\pm\rangle = V_k F'_w V_{k-1} F'_w \dots V_1 F'_w |s\rangle. \tag{12}$$

Each translationally invariant operator V_ℓ must be diagonal in the momentum basis

$$|\mathbf{p}\rangle = \frac{1}{\sqrt{2N}} \sum_{x=0}^{2N-1} e^{i\mathbf{p}x\pi/N}|x\rangle \tag{13}$$

because

$$T|\mathbf{p}\rangle = e^{i\alpha}|\mathbf{p}\rangle \tag{14}$$

for some α. Hence, writing the state of the system after ℓ steps as

$$|\psi_\ell\rangle = \begin{cases} |s\rangle & \text{if } \ell = 0, \\ V_\ell F'_w |\psi_{\ell-1}\rangle & \text{if } \ell > 0, \end{cases} \tag{15}$$

one finds that

$$|\langle \mathbf{p}|\psi_\ell\rangle| = |\langle \mathbf{p}|V_\ell F'_w|\psi_{\ell-1}\rangle|. \tag{16}$$

By (14), $|\langle \mathbf{p}|V_\ell|\Psi\rangle| = |\langle \mathbf{p}|\Psi\rangle|$ for any $|\Psi\rangle$, so (16) is equivalent to

$$|\langle \mathbf{p}|\psi_\ell\rangle| = |\langle \mathbf{p}|F'_w|\psi_{\ell-1}\rangle|. \tag{17}$$

In [4], (17) is expanded in terms of each computational basis vector $|x\rangle$, so that it takes the form

$$P_\ell(z) = P_{\ell-1}(z) \quad \text{at } |z| = 1 \tag{18}$$

for a set of polynomials $P_\ell(z)$ of degree $N-1$. In other words, finding polynomials P_ℓ satisfying (18) corresponds to finding an exact quantum algorithm for ordered search.

To simplify this search, Farhi et al. also define a new set of polynomials $Q_\ell(z) = P_\ell(z)[P_\ell(\frac{1}{z^*})]^*$, which decompose on $|z| = 1$ as

$$Q_\ell(e^{i\theta}) = 1 + A_\ell(\theta) + B_\ell(\theta), \tag{19}$$

where

$$A_\ell(\theta) = \sum_{r=1}^{N-1} a_{l,r} \cos r\theta, \quad a_{l,r} = a_{l,N-r}; \tag{20}$$

$$B_\ell(\theta) = \sum_{r=1}^{N-1} b_{l,r} \cos r\theta, \quad b_{l,r} = -b_{l,N-r}. \tag{21}$$

The conditions in (17) in terms of the polynomials Q_ℓ take the form

$$Q_\ell(z) \geq 0 \quad \text{on } |z| = 1 \tag{22}$$

and

$$\begin{aligned} B_0(\theta) &= B_1(\theta), \\ A_1(\theta) &= A_2(\theta), \\ B_2(\theta) &= B_3(\theta), \\ &\vdots \end{aligned} \tag{23}$$

while the starting and ending values of the algorithm in (9) and (10) have become

$$\begin{aligned} A_0(\theta) &= \sum_{r=1}^{N-1} \cos r\theta, \\ B_0(\theta) &= \sum_{r=1}^{N-1} \left(1 - \frac{2r}{N}\right) \cos r\theta, \end{aligned} \tag{24}$$

and

$$A_k(\theta) = B_k(\theta) = 0. \tag{25}$$

Thus, finding a k-query quantum algorithm for an ordered search of N elements is equivalent to finding functions A_ℓ and B_ℓ that satisfy the starting and ending conditions in (24) and (25) and the equality condition in (23) [4].

3 Improving the Algorithm

Using the equalities in (23), we may rewrite the set of inequalities in (22) as a set of inequalities for A_ℓ and B_ℓ by substituting $A_\ell = A_{\ell+1}$ for ℓ even, and $B_\ell = B_{\ell+1}$ for ℓ odd:

$$\begin{aligned} 1 + A_1(\theta) + B_0(\theta) &\geq 0 \\ 1 + A_1(\theta) + B_2(\theta) &\geq 0 \\ 1 + A_3(\theta) + B_2(\theta) &\geq 0 \\ &\vdots \end{aligned} \tag{26}$$

Since either A_ℓ or B_ℓ (but not both) is needed for each value of ℓ, we may define a new set of functions $\mathbf{C}$ as

$$C_\ell(\theta) = \begin{cases} A_\ell(\theta) & \text{if } \ell \text{ is odd}, \\ B_\ell(\theta) & \text{if } \ell \text{ is even}, \end{cases} \tag{27}$$

with the inequalities in (26) becoming

$$1 + C_\ell(\theta) + C_{\ell+1}(\theta) \geq 0 \quad (0 \leq \ell \leq k-2). \tag{28}$$

Each function C_ℓ has approximately $\frac{N}{2}$ free variables because of the equalities in (20) and (21), so the total dimension of the search space is approximately $\frac{N}{2}(k-1)$. As N must grow exponentially with k to achieve an improvement over the best known algorithms, we must refine our search in order to find this set of functions for values of $k > 3$.

3.1 Defining a Cost Function

We define a cost function whose minimization will result in finding a solution $\mathbf{C}$ to the conditions in

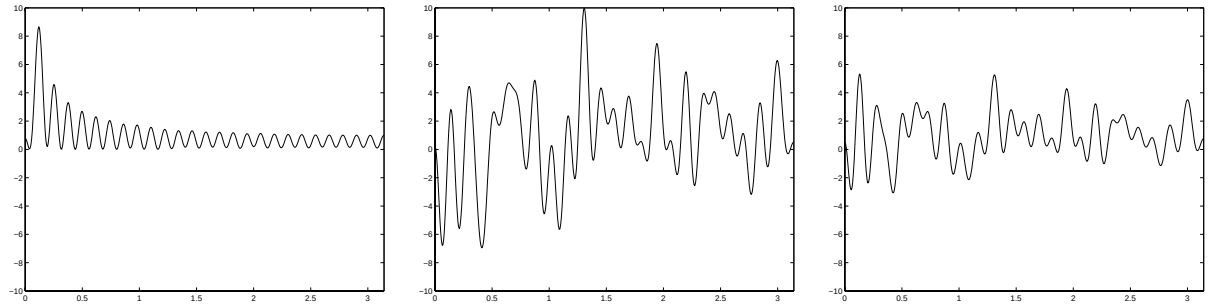


Figure 2: Plots of $f_\ell(\theta)$ for $\mathbf{C}^*$, $\mathbf{C}$, and $\frac{1}{2}(\mathbf{C}^* + \mathbf{C})$ vs. θ (left to right), with $k = 3$ queries and $N = 53$ elements ($\epsilon = \frac{1}{2}$). It is evident from the graphs that the average of the random set $\mathbf{C}$ and the solution set $\mathbf{C}^*$ has a smaller negative area than $\mathbf{C}$, and thus a lower cost function. This supports our observation that moving towards C^* in the search space always decreases the cost function.

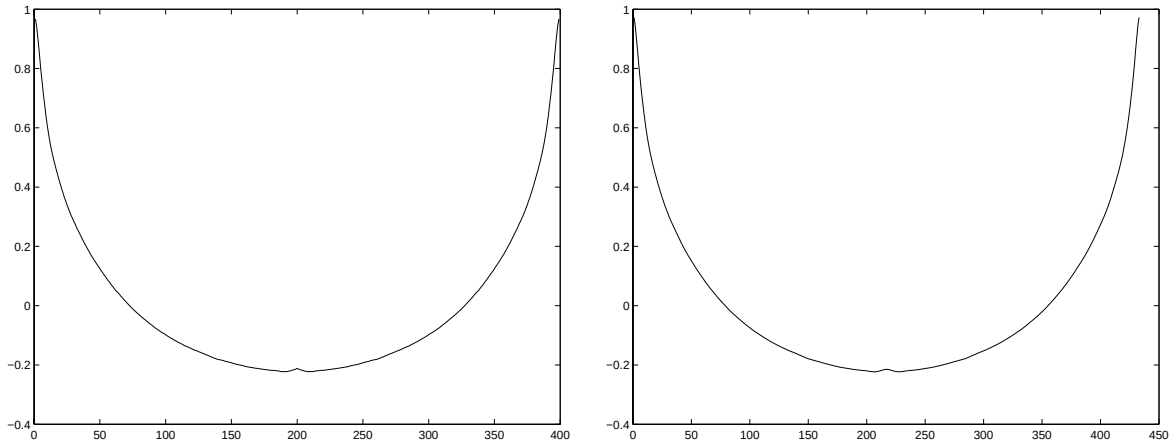


Figure 3: Plots of the coefficients $c_{1,r}$ of C_1 vs. r for $k = 4$ queries and $N = 400$ and $N = 434$ elements (left to right). Despite a difference in the number of coefficients, the curves for the two are nearly identical, with a minimum value of approximately -0.2 near the middle of the curve, and a maximum value of approximately 1 at either end.

(28). Let $f_\ell(\theta) = 1 + C_\ell(\theta) + C_{\ell-1}(\theta)$. We define our cost function as twice the total negative area under the curves $f_0, f_1, ..., f_{k-2}$ from 0 to π, or

$$\text{cost}(\mathbf{C}) = \sum_{\ell=0}^{k-2} \int_{\theta=0}^{\pi} (|f_\ell(\theta)| - f_\ell(\theta))\, d\theta. \qquad (29)$$

If each $f_\ell(\theta)$ is everywhere nonnegative, then cost($\mathbf{C}$) is zero. Thus, if the cost is minimized to 0 for a set of functions C, then (22) holds, and these functions will provide an algorithm for the ordered-search problem.

An important observation about the search space in which we are looking for $\mathbf{C}$ is that if there exists a set of functions $\mathbf{C}^*$ for which (28) holds, then there are no local minima for $\mathbf{C}$. We may see this by considering some $\mathbf{C}$ for which cost($\mathbf{C}$) is positive. If we move in the search space towards the solution set $\mathbf{C}^*$, taking $\mathbf{C}$ to $(1 - \epsilon)\mathbf{C} + \epsilon\mathbf{C}^*$ with $0 < \epsilon \leq 1$, we will decrease this cost function (Figure 2). This is clear because moving towards $\mathbf{C}^*$ will add a nonnegative value to $C_\ell(\theta)$ for every ℓ and θ. Note that the change in the cost function when moving towards $\mathbf{C}^*$ will always be nonzero, as $C^*_\ell(\theta)$ is zero at only finitely many points for each ℓ. Thus, from any $\mathbf{C}$, if a solution exists for N elements and k queries, we may find it by moving in a direction of decreasing cost through the search space until the cost function has been minimized to zero. Additionally, if a local minimum for the cost function has been found and the cost is not equal to zero, we may conclude that

there is no solution for k queries and N elements.

3.2 Minimizing the Cost Function

We found that standard methods such as genetic algorithms, simulated annealing, and the simplex method performed poorly in finding solutions to (28) [7]. None of these methods utilize knowledge about the absence of local minima in their search, nor do they utilize our explicit formula for the cost function in (29) to find a direction which will provide a large decrease in cost. We found that zero-temperature annealing performed better than these, as it used the former fact, but it still performed poorly because the direction of movement is still random. The only method we found that performed well was gradient descent with line minimization [7].

Our implementation of this method starts with a set of functions $\mathbf{C}$ satisfying (20) and (21) and the boundary conditions (24) and (25). It then calculates the gradient $\nabla\text{cost}(\mathbf{C})$ from (29), ensuring that the pairs of coefficients $a_{\ell,r}$ and $a_{\ell,N-r}$, and $b_{\ell,r}$ and $b_{\ell,N-r}$, are each treated as a single variable to respect the symmetries in (20) and (21).

Once we have found the gradient, we reduce the problem to a one-dimensional search along this vector to find a new $\mathbf{C}'$ for which $\text{cost}(\mathbf{C}') < \text{cost}(\mathbf{C})$. We perform this minimization by a process known as bracketing the minimum. Although this method is only useful for finding a local minimum along the line, we have already concluded that there are no local minima in the search space if a solution $\mathbf{C}^*$ exists. Thus, as long as this method decreases the cost function, it does not matter whether we move to a local or global minimum along the line of the gradient. The advantage of bracketing the minimum is that it is quick and will always return a local minimum. Our exact implementation of the line-minimization algorithm is as follows:

1. Begin with three sets of functions that lie along the gradient $\nabla\text{cost}(\mathbf{C})$ from $\mathbf{C}$.
2. Choose a fourth set of functions between the outermost two.
3. Compute the cost of all four sets of functions.
4. Repeat with the three sets of functions which have the smallest costs.

Choosing our initial point to be $\mathbf{C}$ and two functions along the gradient that are sufficiently far away from $\mathbf{C}$, we will in this way obtain a local minimum along the gradient.

Finally, we discovered a heuristic which speeds up this method even further. The solutions for fixed k and nearby N seem to be very closely related. In fact, the coefficients $c_{l,r}$ in the solution $\mathbf{C}^*$ appear to converge to a fixed curve as N becomes large (see Figure 3). Using the solution curves of a given k and N to generate an initial guess for $\mathbf{C}$ for a larger value of N speeds up the search substantially over what would be obtained if a random set of functions were chosen.

3.3 Results

We have found solutions $\mathbf{C}^*$ with $k = 3$ queries for lists of up to size $N = 53$ elements, and with $k = 4$ queries for lists of up to size $N = 434$ elements. The latter of these can be converted into a recursive quantum algorithm that solves the ordered-search problem in $4\log_{434} N \approx 0.457\log_2 N$ queries, providing a speedup over the quantum algorithms found in [3] and [4]. Neither of these are necessarily optimal solutions for their respective values of k, and better solutions are likely to follow as there is more time to run the program (for $k = 4$ queries and $N = 434$ elements, our search takes approximately 1.5 hours on a 1.80 GHz Pentium 4 machine).

The graphs of the coefficients $c_{l,r}$ for the solution functions follow patterns which suggest that a closed form for these functions may exist (see Figures 4 and 5). Such a solution would be helpful in establishing optimal solutions for each k and, in doing so, would likely find even better quantum algorithms for the problem. The graphs of the cost function $f_\ell(\theta)$ also suggest the problem has further structure and could provide additional clues as to a closed form for the coefficients for solutions (see Figures 6 and 7).

4 Conclusion

We have discovered a quantum algorithm that can search an ordered list of N items by making only $4\log_{434} N \approx 0.457\log_2 N$ queries, improving on the previous best algorithm, which requires $0.526\log_2 N$ queries [4]. We have developed a cost function containing no local minima whose global minimum corresponds to a translationally invariant recursive quantum algorithm for this problem, and have presented the possibility of a universal curve for minima of this

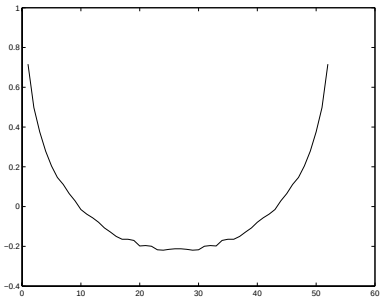


Figure 4: A plot of the coefficients $c_{l,r}$ of C_1 vs. r for $k = 3$ queries and $N = 53$ elements.

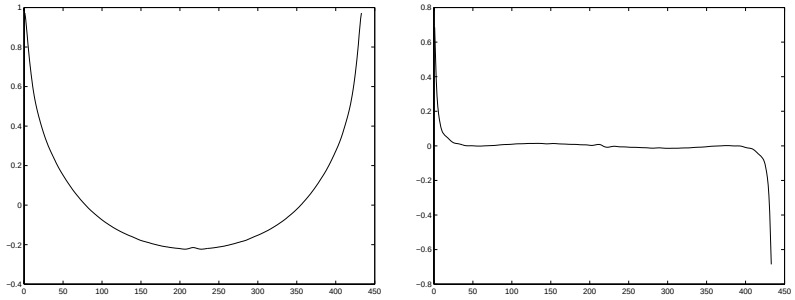


Figure 5: Plots of the coefficients $c_{l,r}$ of C_1 and C_2 vs. r (left to right) for $k = 4$ queries and $N = 434$ elements.

cost function. We believe further improvement is possible, although perhaps not quite as far as the best lower bound of $0.22 \log_2 N$ queries for this problem [3].

5 Acknowledgements

I would like to thank Dr. Andrew Landahl for his great help and support in my research, as well as Dr. Edward Farhi, Dr. Jeffrey Goldstone, and Dr. Sam Gutmann for useful discussions and suggestions. I would also like to thank those that have helped me to revise my paper, including Jeremy England and Ben Rahn. Finally, I would like to thank the Center for Excellence in Education for giving me the opportunity to do this research.

References

[1] Grover, Lov K. A fast quantum mechanical algorithm for database search. *Proceedings of the 28th Annual ACM Symposium on the Theory of Computing* (1996). 212–219.

[2] Boyer, M., Brassard, G., Høyer, P., and Tapp, A. Tight Bounds on Quantum Searching. *Fortsch. Phys.* **46** (1998). 493–506.

[3] Høyer, P., Neerbek, J., and Shi, Y. Quantum Complexities of Ordered Searching, Sorting, and Element Distinctness. *Algorithmica.* **34** (2002). 429–448.

[4] Farhi, E., Goldstone, J., Gutmann, S., and Sipser, M. Invariant Quantum Algorithms for Insertion into an Ordered List. MIT CTP #2815 (1999).

[5] Preskill, John. *Lecture Notes for Physics 229: Quantum Information and Computation* (1998). `http://www.theory.caltech.edu/people/preskill/ph229/`

[6] Nielsen, M., and Chuang, I. *Quantum Computation and Quantum Information* (2000). Cambridge, UK. Cambridge University Press.

[7] Flannery, B., Press, W., Teukolsky, S., and Vetterling, W. *Numerical Recipes in C* (1988). Cambridge, UK. Cambridge University Press.

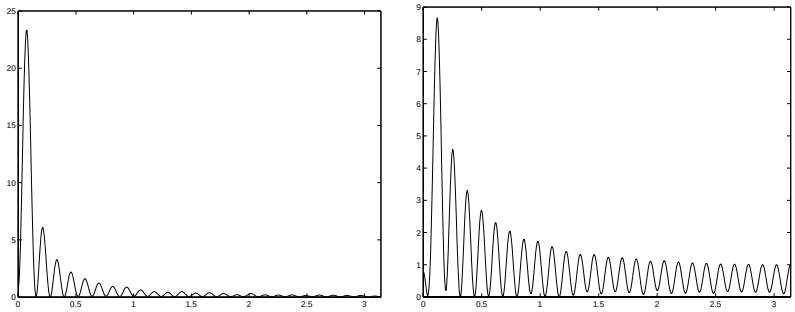


Figure 6: Plots of $f_1(\theta)$ and $f_2(\theta)$ vs. θ (left to right) for $k = 3$ queries and $N = 53$ elements.

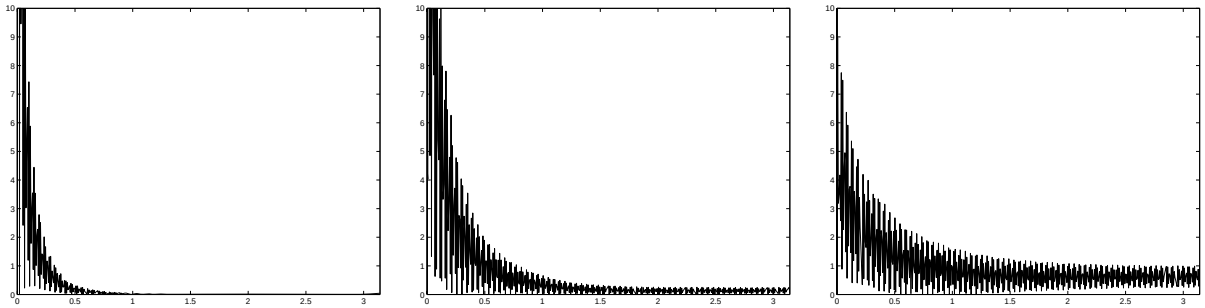


Figure 7: Plots of $f_1(\theta)$, $f_2(\theta)$, and $f_3(\theta)$ vs. θ (left to right) for $k = 4$ queries and $N = 434$ elements.

Neural Basis of Inhibition: A Study of Antisaccades Using fMRI and MEG

Bobby Xu

under the direction of
Dr. Dara Manoach
Massachusetts General Hospital

Abstract

Inhibition is a cognitive ability that allows humans to respond flexibly rather than reflexively to events. The purpose of this investigation is to elucidate the neural basis of inhibition in healthy humans. Combining magnetoencephalography and functional MRI, two unique brain-imaging instruments, allows both key brain regions and the timing of neuronal processes involved with inhibition to be resolved. The study found strong evidence that the dorsolateral prefrontal cortex is an important component in programming inhibitory signals.

1 Introduction

1.1 The Antisaccade Task

The analysis of eye movements has long been used as an important diagnostic tool in neurology. Its use is usually neglected in the diagnosis of brain disorders and only recently has it become clearer that a variety of neurological disorders are associated with an inability to inhibit saccades.[1] The ability to suppress reflexive responses and to generate voluntary motor commands is crucial for everyday life because it frees the organism from a stimulus-driven behavior in favor of the achievement of internal goals. It is possible to examine this ability in one oculomotor task by presenting a visual stimulus at one side and asking the subject to look to the opposite side. A deficit in the inhibition of reflexive responses will result in a high number of saccades towards the visual stimulus, prosaccades (Figure 1a), and a low number of saccades to the opposite side, antisaccades (Figure 1b). This task was introduced by Peter Hallett in 1978 as a "novel task" and is now called the antisaccade task [12].

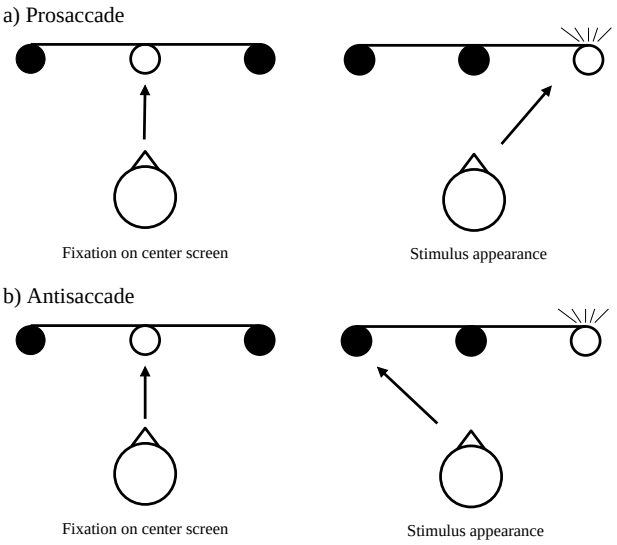


Figure 1: Two types of saccades. (a) Prosaccades are reflexive and directed *towards* the target. (b) Antisaccades are directed *away* from the target.

In the last 10 years, a large number of clinical studies have been conducted ranging from studies with patients with discrete lesions to those with psychiatric disorders. Functional imaging techniques have been applied to discover the brain areas involved in the generation of antisaccades. Basic research has characterized the properties of antisaccades and the conditions which lead to errors in this task. The findings obtained so far suggest that the antisaccade task

[1] A saccade is a rapid eye movement that brings the point of gaze to the image of interest.

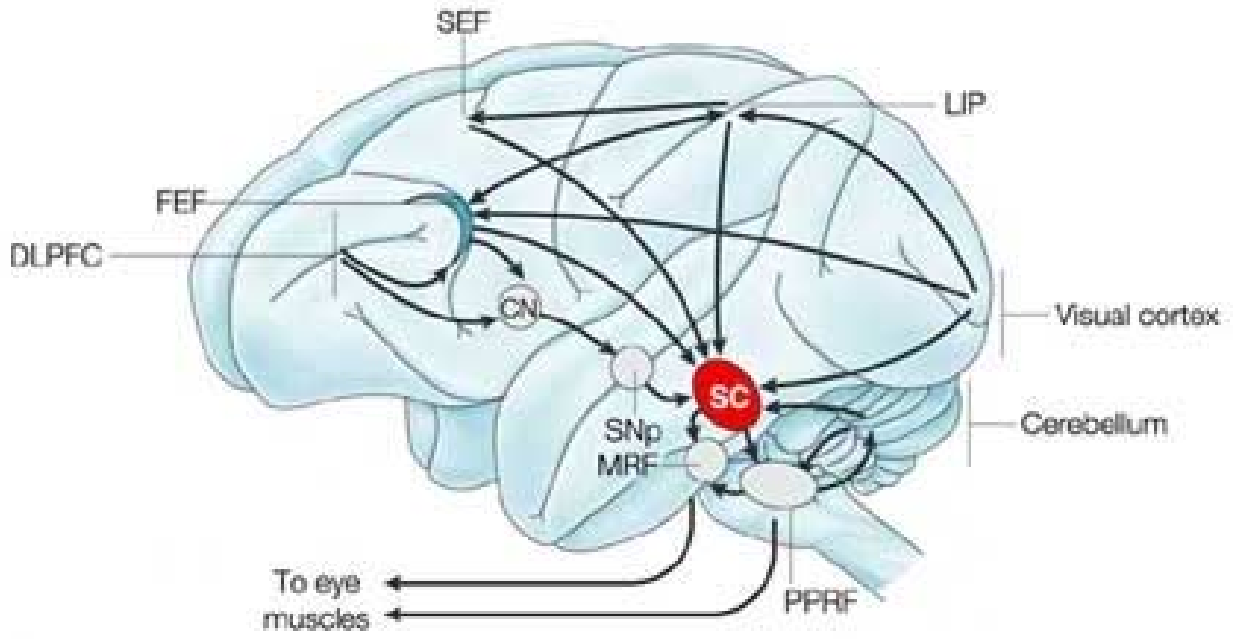


Figure 2: Structures believed to be involved in the generation of a saccade. A simplified schematic of neural activation is shown, with activity beginning in the visual cortex. CN, caudate nucleus; DLPFC, dorsolateral prefrontal cortex; FEF, frontal eye field; LIP, lateral intraparietal cortex; MRF, medullary reticular formation; PPRF, paramedian pontine reticular formation; SEF, supplementary eye field; SC, superior colliculus; SNp, substantia nigra pars reticulata. Image obtained from [10].

can—with certain restrictions—be applied as a diagnostic tool for diseases affecting cortical and subcortical structures [8]. Deficient inhibition is characteristic of disorders such as schizophrenia, bipolar disorder, unipolar depression and obsessive compulsive disorder. Understanding the neural circuitry involved in the generation of antisaccades will contribute to a more complete understanding of inhibition deficits and lead to rational interventions.

1.2 Studying the Neural Basis of Antisaccades

Previous studies of both human and nonhuman primates using an array of neuroimaging techniques have mapped out the processes involved in the generation of prosaccades (Figure 2). They converge in identifying the posterior parietal cortex (PPC) and frontal and supplementary eye fields (FEF, SEF) as being important components [1, 9, 18]. Use of temporal data has determined PPC activation to precede activation in frontal regions [19]. While the neural basis of normal saccades has been elucidated, understanding of antisaccades is less complete. This study aims to discover the key brain regions, neural circuits, and timing of neuronal processes that characterize antisaccades and inhibition.

Unfortunately, isolating a single neurocognitive function for study has proven to be difficult in the past. Efforts are often confounded because the tasks employed to study such functions involve more than one neurocognitive function [2]. In this study, saccadic inhibition is investigated using a setup from a recently completed project that isolates saccadic inhibition for study [13].

Subjects of this study were asked to perform a large number of both antisaccades and prosaccades. During performance of the task, neural activity will be recorded using functional MRI, which provides high spatial resolution, and magnetoencephalography (MEG) or electroencephalography (EEG), which provide high temporal resolution. The larger goal of using this data for antisaccade study is divided into three separate aims.

Aim 1 of this study is to determine whether antisaccades are associated with increased activation in the dorsolateral prefrontal cortex (DLPFC). Several studies examining regions of increased cerebral blood flow associated with the performance of antisaccades have yielded different results. McDowell et al. and Doricchi et al. demonstrated increased activation in the DLPFC as well as several other locations [6, 14]. However, Paus et al. found significantly greater activation only in the anterior cingulate cortex and the posterior parietal cortex during the performance of an antisaccade task compared to the activation during a prosaccade task [17]. In contrast, O'Driscoll et al. observed increased activation in the FEF, the supplementary motor area, the thalamus, the putamen, the superior parietal lobe and area 17 during antisaccades compared with prosaccades [16]. This study attempted to resolve this issue by comparing neural activation during antisaccades to activation during prosaccades by using fMRI data.

Aim 2 of this study focuses on finding the importance of DLPFC activation in the generation of correct antisaccades. If the DLPFC is a crucial component in inhibition, then error antisaccades should be associated with its failure, indicating that correct antisaccades can be generated only when that region is active. This aim was studied by comparing areas of activity, determined by fMRI, during performance of correct antisaccade trials to activity during error antisaccades. Because of the great number of saccadic trials each subject must perform, error antisaccades are common. This study used an event-related fMRI method that allowed error and correct antisaccade trials to be studied separately and compared [5].

Aim 3 of this study is to determine when inhibition occurs relative to the onset of the saccade. Evidence is abundant suggesting that the timing of neuronal processes across regions may be key to understanding neurocognitive functions [11]. Because the temporal resolution of fMRI is insufficient to study this aim, MEG data will be used. If the DLPFC is determined by fMRI to be activated or deactivated in error antisaccades, the change must occur before the onset of the actual saccade. MEG can find whether this is the case and also when exactly the signal is sent. A recent EEG study found evidence of frontal brain activity specific to correct antisaccade performance 160–60 ms before onset of the saccade [3]. Other studies have found that frontal inhibition occurs in the last 100 ms prior to the generation of a correct antisaccade [7].

2 Methods and Materials

2.1 Subjects

Twelve healthy right-handed subjects, seven females and five males, participated in the study after providing informed consent. Their ages ranged from from 19 to 39 (average age: 23.2 yr). Subjects were free of neurological or psychiatric illness and were screened to exclude substance abuse and depression in the past six months. Sociodemographic and neurocognitive information such as verbal IQ and socioeconomic status was collected from each subject. Subjects were compensated for participating in the study.

2.2 Visual Stimulus

The visual stimulus consisted initially of a white fixation ring of diameter 1.0° (measured in terms of degrees in the visual field) centered on a dark background (see Figure 3). The fixation ring was flanked by two 0.7° white dots placed 10° right and left of center. At the start of a trial, the central fixation ring was replaced by one of two prompts. For half of the subjects, a yellow "O" 4.5° in diameter was the cue for a prosaccade, and a blue "X" spanning 4.5° was the cue for an antisaccade. For the other half, the "O" was the cue for an antisaccade and the "X" prompted a prosaccade. Prompts lasted 300 ms before being replaced by the central fixation ring. The fixation ring disappeared after 1700 ms, and a similar ring appeared around one of the two peripheral dots, the side randomly determined. This signaled the subject to make their saccade quickly and accurately. The white ring remained to the side for 1000 ms during which time subjects can fixate the target. The white ring then returned to the center for 1000 ms before the start of the next trial. One complete trial lasted 4000 ms. Each run consisted of 26–46 of each trial type.

Antisaccade and prosaccade trials were randomly intermixed with the fixation baseline condition, which occurred about every fifth trial. The display during fixation was the same as the display during the cue-to-target interval. 10–17 fixation trials were present in each run.

2.3 Saccadic Measurements

The ISCAN® fMRI Remote Eye Tracking Labratory was used to record saccades during fMRI scanning. An ISCAN imaging video camera captured images of

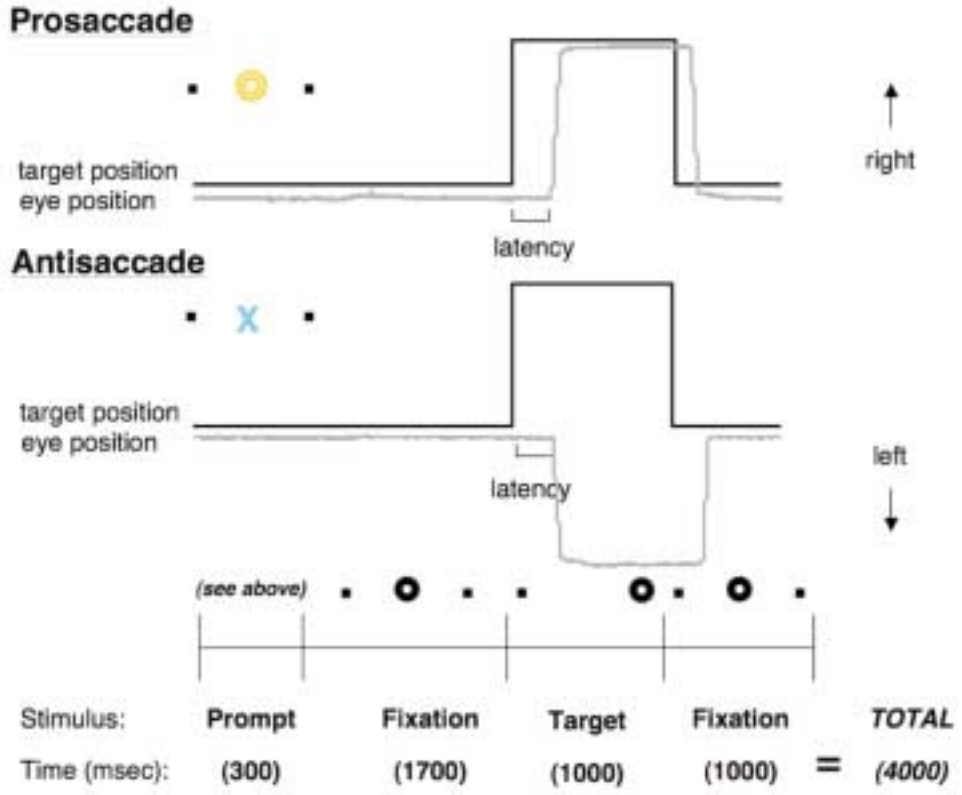


Figure 3: Trial illustration. Progress over time is from left to right. Top lines show horizontal position traces of targets (black smooth lines) and eyes (grey irregular lines) for a correct prosaccade (top) and antisaccade (below). Rightward motion is shown as up, by convention.

the eye during performance of the saccadic tasks. The eye images were sent to ISCAN's RK-726PCI high-resolution tracker located outside the fMRI room. The point where the eye was looking was determined by ISCAN's eye-tracking and calibration processors.

During MEG scanning, the direction of eye movement was recorded using electrooculargram (EOG) sensors, placed above and below one eye and on the left and right of the head. EOG is very sensitive to the detection of saccadic onset.

The raw data obtained directly from ISCAN and EOG took the form of a series of waveforms indicating the direction of eye movement. The primary purpose of recording saccades is to identify correct and error trials. When a subject looked in the wrong direction indicated by the cue, the trial was scored as error, and the corresponding MEG or fMRI data for that trial was marked as brain activity associated with an incorrect response. Because data obtained from the instruments was often noisy, scoring was performed manually. In addition, MEG trials in which the subject blinked were discarded because blinks caused strong neural interference.

2.4 fMRI Procedures

The fMRI experiment consisted of a total of six runs of 5 minutes 22 seconds each. Each run had approximately 70 saccadic trials, split evenly between antisaccades and prosaccades. The six runs generated a total of 200 trials of each type per subject. The first two trials of each run were excluded from analysis. The total experiment time was approximately 40 minutes, including short rests between each run. The visual task was generated by a Macintosh computer using the Vision Shell libraries (MicroML, St. Hyachinthe, Quebec) and projected onto a mirror while the subject lay inside the fMRI instrument.

Structural images of the brain were first taken us-

ing anatomical MRI. Structural images are detailed anatomical pictures of a subject's brain but provide no information regarding activation. A 3 Tesla Allegra MR scanner (Siemens Medical System, Inselin, NJ) was used to take 2 high-resolution 3D rf-spoiled gradient echo T1 weighted scans. The procedure produced 128 1 x 1.35mm in-plane slices, each 1.3mm thick.

Anatomical data was used primarily to create 3D reconstructions of the brain on which MEG and functional data could be mapped. The cortical surface was inflated to make activity buried within sulci more visible. The reconstructions were also morphed into a sphere for intersubject registration, which matched together corresponding areas of activation for comparison across several subjects.

Functional MRI was used to determine which parts of the brain were activated by different stimuli. Functional scans were performed using the same 3 Tesla MR scanner. They were collected using blood oxygen level dependent (BOLD) contrast and a gradient echo T2 weighted sequence to measure variations in blood flow and oxygenation.

Functional scans were processed by the FreeSurfer Functional Analysis STream (FS-FAST) to produce statistical maps data for overlay on the structural data. Individual hemodynamic estimates were averaged together for the group using a fixed-effects model, and statistical maps were created at every timepoint. Statistical contrasts were created for antisaccade versus prosaccade trials and antisaccade error versus antisaccade correct trials.

2.5 MEG procedures

The MEG task was identical to the fMRI experiment, except that eight runs were used instead of six to increase statistical power. Stimuli were generated by the same model Macintosh computer and projected onto a screen placed in front of the subject.

MEG data were acquired simultaneously in a magnetically and electrically shielded room. MEG signals were recorded from the entire head using a 306-channel SQUID Neuromag Vectorview. The signals were recorded continuously with 600 Hz sampling rate and were minimally filtered.

MEG data was analyzed using the MNE software package. The software was used to find minimum norm estimate (MNE) solutions to the MEG inverse problem. Solutions to the inverse problems are approximations of the areas of spatial activation in the brain. MNE solutions at each timepoint were overlaid on a 3D inflated brain to produce a movie of brain activation.

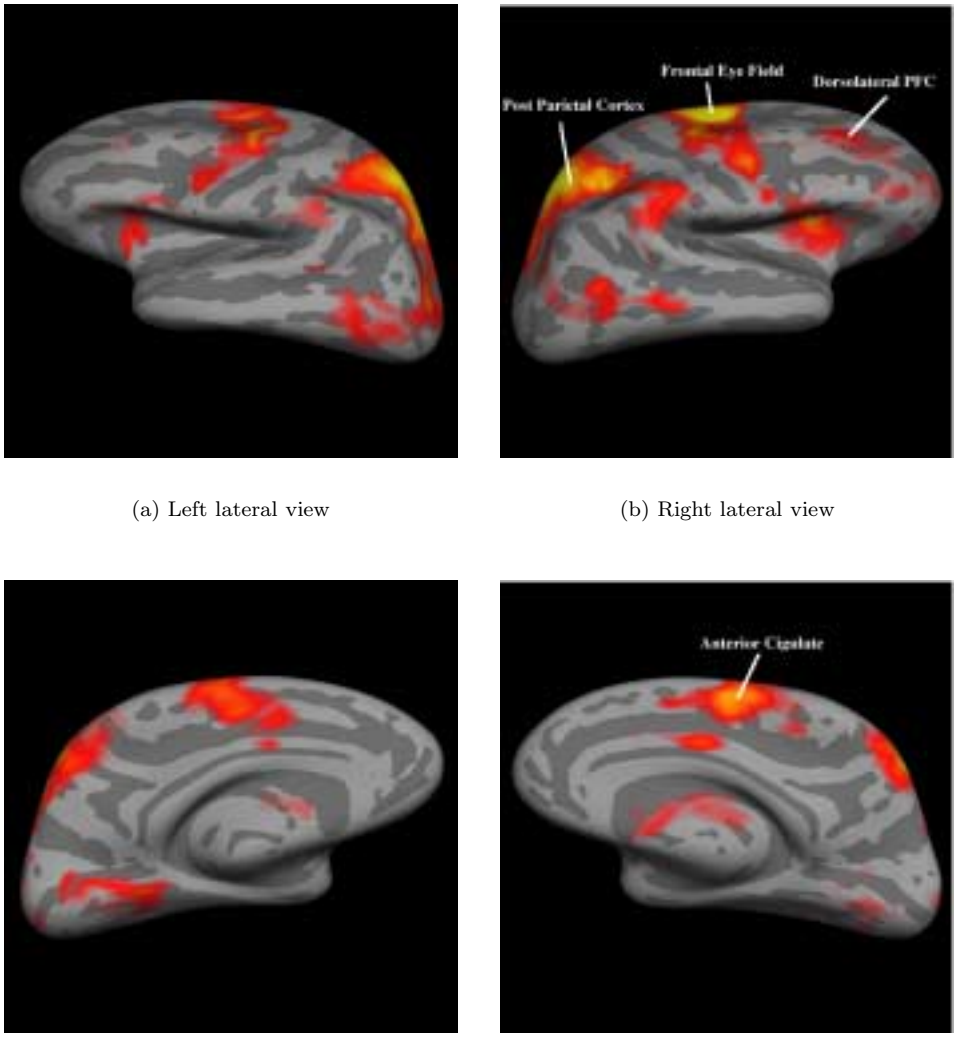
3 Analysis 1: Antisaccade vs. Prosaccade

3.1 Results

Figures 4a–4d show the contrast in neural activation between antisaccades and prosaccades based on BOLD fMRI data. Spherical spatial normalization was used for this figure. The activation maps use functional data averaged across all 12 subjects using fixed-effects analysis, making it more statistically powerful than analysis of any one individual. Voxels, or brain points, that are activated or deactivated to a statistically significant degree are mapped onto an inflated cortical model. Areas that show greater activation in antisaccades compared to prosaccades are red or yellow and areas of lesser activation are blue and white. The maps show activation averaged across all timepoints.

Examination of the lateral activation maps (Figures 4a and 4b) reveals increased activity in a distributed cortical and subcortical network of brain areas during antisaccades. The strongest activation occurs in the postparietal cortex (PPC) and the frontal eye field (FEF). Activation in these regions is roughly symmetrical on the left and right hemispheres of the brain. Notably, the dorsolateral prefrontal cortex also exhibits increased activation in antisaccades with activity substantially stronger on the right hemisphere. The lateral maps show additional activation in the insular, occipital, and temporal regions. The medial view (Figures 4c–4d) displays activation in the anterior cingulate cortex on both hemispheres.

Figures 5a–5d show timecourse maps of neural activation in the lateral view, starting at 2 seconds after the prompt ("O" or "X") first appears. Images were taken every two seconds, but do not accurately reflect brain activation at that exact time due to the latency of hemodynamic response. The brain remains relatively stagnant until the map at 4 seconds, when activation is similar to that in Figure 4b: strong activation in the PPC and FEF and some activation in the DLPFC. Brain activity in those regions remains mostly intact at 6 seconds. Traces of activity in the FEF, PPC, DLPFC, and middle temporal gyrus remain at 8 seconds, though most brain activity has diminished.



(a) Left lateral view

(b) Right lateral view

(c) Left medial view

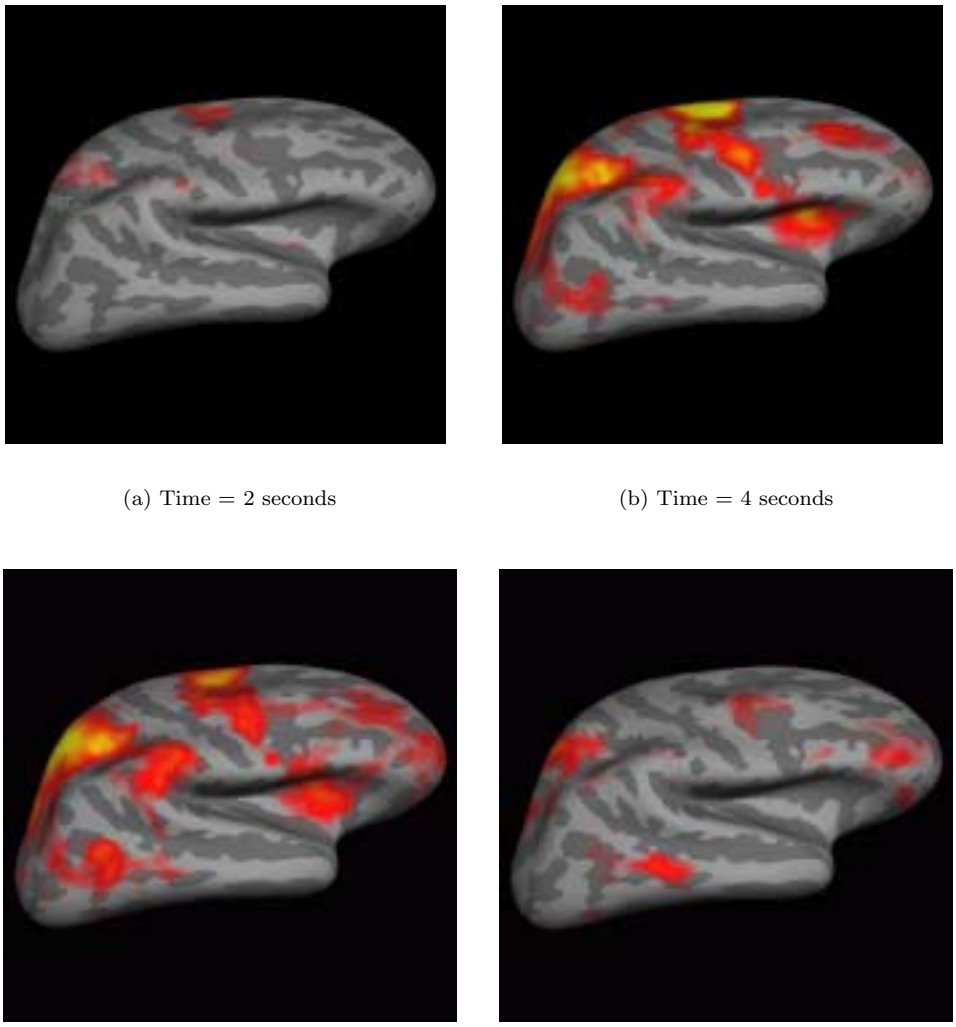
(d) Right medial view

Figure 4: Activation maps of Analysis 1: Antisaccade versus prosaccade. Lateral and medial surfaces on an inflated cortical model are displayed. Sulci are the darker stripes and gyri are the lighter stripes.

3.2 Discussion

The activation shown in Figures 5a and 5b support the hypothesis that DLPFC activation is more present in antisaccades than in prosaccades. A natural extension to this conclusion is that dorsolateral activation plays a role in saccadic inhibition and perhaps inhibition in general. The results presented are consistent with findings by McDowell et al. [14] and Doricchi et al. [6] supporting the conjecture that prefrontal activation is involved in suppressing a reflexive response in favor a deliberate behavior. The fact that DLPFC activation is greater on the right hemisphere can be explained by a study showing that the right hemisphere is dominant for spatial attention, including saccades [15].

The results are consistent with literature indicating activation in areas outside the DLPFC. The FEF and PPC in particular are regions most commonly associated with saccadic inhibition and reported in



(a) Time = 2 seconds

(b) Time = 4 seconds

(c) Time = 6 seconds

(d) Time = 8 seconds

Figure 5: Timecourse maps of Analysis 1: Antisaccade versus prosaccade. Lateral surfaces on an inflated cortical model displayed from 2 seconds to 8 seconds after the prompt. Sulci are the darker stripes and gyri are the lighter stripes.

other neuroimaging studies [6, 16, 17, 18]. Activation for these two regions was strongest in this study. The FEF, which controls premotor response to stimuli; the PPC, which integrates sensory information in cognition; and the DLPFC, traditionally supposed to suppress reflexive responses, all play logical roles in antisaccade generation. Other regions shown to be active such as the insula, temporal gyrus, and occipital lobe have functions that do not appear to affect antisaccade generation and their activation was not consistently found in other studies.

The timecourse images bring further support to the hypothesis that the prefrontal cortex plays a significant role in controlling antisaccades. Peak activation was shown in Figure 5b, 4 seconds after the initial prompt. The target (which the subject follows or looks away from) appears at 2 seconds, and considering an approximate 2-second hemodynamic

lag,[2] the activity shown in Figure 5 occurs around when the saccade is programmed. This gives strong support to the assertion that the DLPFC is involved in antisaccade generation. If principal dorsolateral activation occurred at 2 or 8 seconds, then it would not be possible that the DLPFC is involved in programming the inhibitory response because of timing. However, the temporal resolution of fMRI is quite poor and more complete data with regard to timing was collected from MEG (see Analysis 3).

However, whether each of these regions is absolutely crucial to inhibition cannot be determined by fMRI and MEG alone. These noninvasive studies can only bring support to associations between brain regions and their function in saccadic inhibition. The absolute importance of PPC, DLPFC, and FEF can only be determined by lesion studies, where those areas are completely impaired by lesions.

Subj. #	Correct ASs	Error ASs	Error rate
1	200	6	2.82%
2	209	4	1.88%
3	204	9	4.23%
4	180	33	15.49%
5	191	22	10.33%
6	210	3	1.41%
7	205	8	3.76%
8	202	11	5.16%
9	204	9	4.23%
10	196	17	7.98%
11	209	4	1.88%
12	209	4	1.88%
TOTAL	1320	130	5.09%

Table 1: Antisaccade error rates for 12 subjects.

4 Analysis 2: Antisaccade error vs. Antisaccade correct

4.1 Results

Figures 6a-6d are activation maps comparing brain activity associated with correct antisaccades and activity during error antisaccades. Areas of greater activation in correct trials are red and yellow and areas of deactivation are blue and white. The figures show activation across all timepoints. As in Analysis 1, the contrast was mapped based on BOLD fMRI data and averaged across all 12 subjects using fixed-effects analysis. Event-related fMRI was used to separate the functional data of correct and incorrect trials.

Table 1 shows the antisaccade error rate of each subject. While most error rates fall under 4%, a few individuals, notably 4 and 5, made significantly more errors. The cumulative error rate is 5.09%, falling within the 5%–7% antisaccade error rate reported in [8]. The total numbers of correct antisaccades generated (1374) and error antisaccades generated (130) are great enough for statistically valid analyses.

Activation and deactivation are more subtle and more scattered than in Analysis 1. The lateral view (Figures 6a and 6b) shows deactivation in the DLPFC and FEF in both the left and right hemispheres. Deactivation is also present in the superior temporal gyrus on both hemispheres. Activation does occur along the frontal eye field in both hemispheres, especially on the left. The medial surface (Figures 6c and 6d) displays decreased activity in the anterior cingulate cortex in correct trials versus error trials, with stronger deactivation occurring on the right.

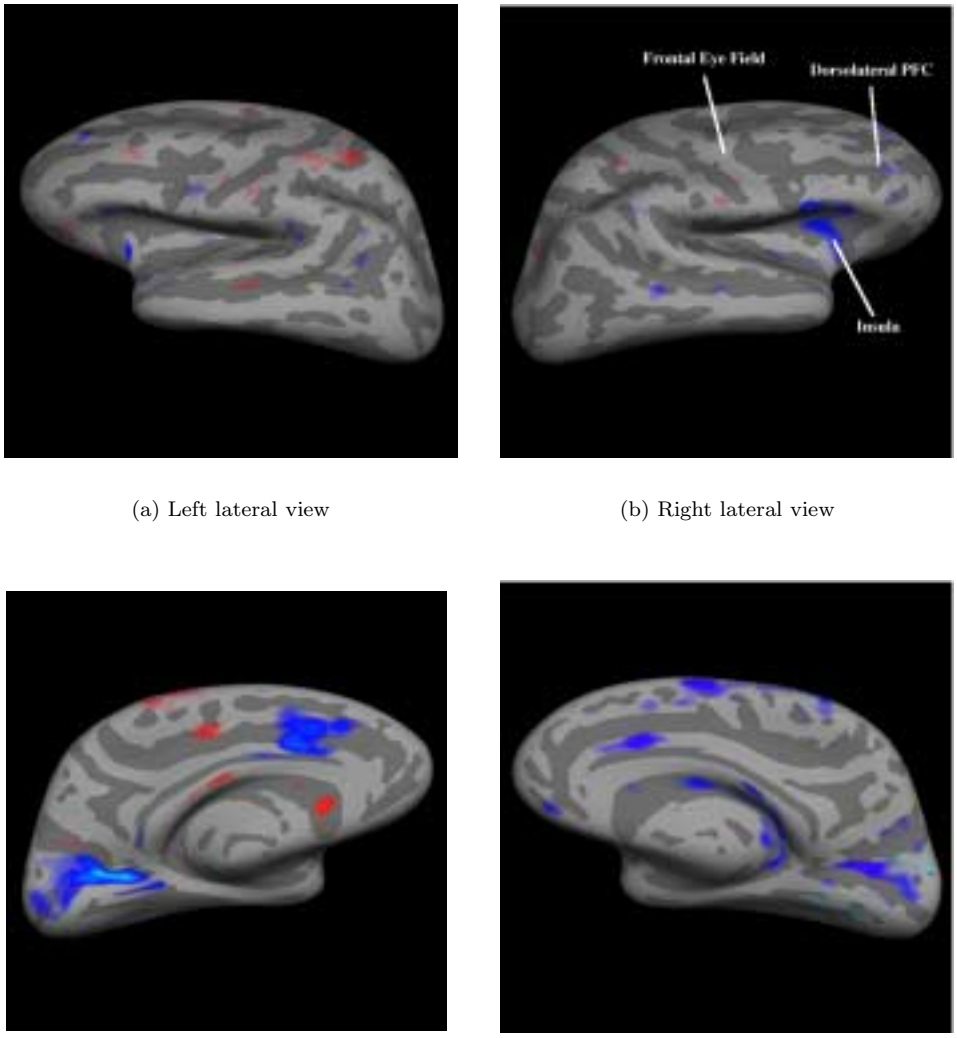
Figures 6a–6b are right lateral views of activation between 2 seconds and 8 seconds after the prompt is displayed. Scattered activity occurs between 2 seconds and 4 seconds in several regions of the brain. The strongest deactivation occurs at 6 seconds in the insula. Deactivation of DLPFC occurs at 6 seconds and 8 seconds after the prompt. FEF activation is present at 2 seconds and 6 seconds.

4.2 Discussion

The results of Analysis 2 do not support the hypothesis that the DLPFC is associated with correct antisaccade generation. In fact, Figure 6 suggests the contrary—that DLPFC is more active in *error* rather than correct trials. However, the timecourse images (Figure 7) show that deactivation of the DLPFC occurred at 6 seconds and 8 seconds, much too late for the activity to have an impact in programming the antisaccade. At the same time, the data does not show activation of the DLPFC to occur when the inhibitory signal is sent. The poor temporal resolution of fMRI is not sufficient to resolve exactly when regional brain activity occurs; it provides only a rough estimate. The exact timing of neural process was studied using MEG (see Analysis 3).

Activation is displayed in the FEF, another region

[2] Hemodynamic lag is the time between which a neural response actually occurs to when it impacts bloodflow and is measured by fMRI.



(a) Left lateral view

(b) Right lateral view

(c) Left medial view

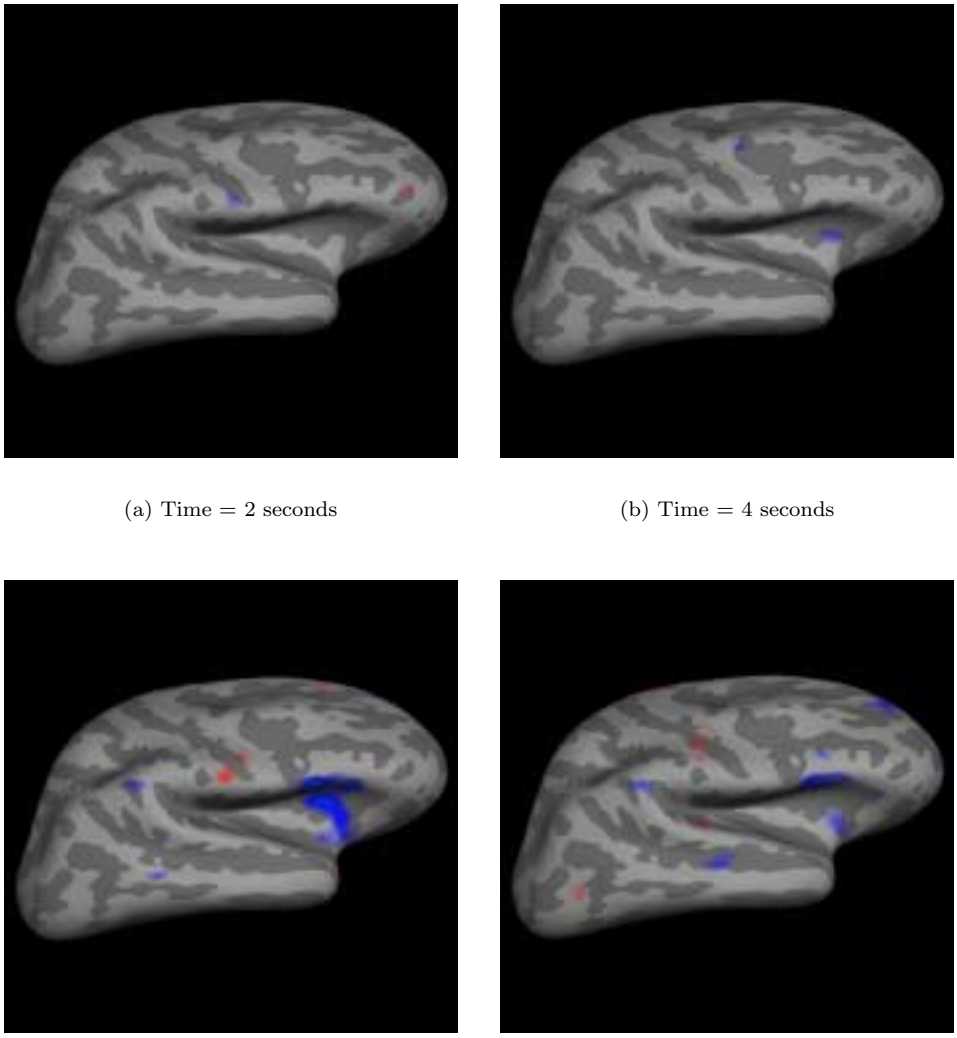
(d) Right medial view

Figure 6: Activation maps of Analysis 2: Antisaccade correct versus antisaccade error. Lateral and medial surfaces on an inflated cortical model are displayed. Sulci are the darker stripes and gyri are the lighter stripes.

that studies have shown to be involved with antisaccades. The results presented do support conclusions made by O'Driscoll et al. [16] asserting that the FEF is the component that inhibits reflexive prosaccades in the antisaccade task. The timecourse data from this study does not actively support O'Driscoll's conclusions, as FEF activation occurs too early or too late to impact the inhibitory response. Again, additional temporal data was collected in Analysis 3.

Importantly, error antisaccades are more complex than correct antisaccade. While presumably only one pathway exists to generate a correct antisaccade, there are several reasons a subject may make an error. Factors that contribute to the failure in generating an inhibitory response may reasonably be associated with distinct neural processes.

Another confounding issue is the diversity of antisaccade errors. The simplest error occurs when the



(a) Time = 2 seconds

(b) Time = 4 seconds

(c) Time = 6 seconds

(d) Time = 8 seconds

Figure 7: Timecourse maps of Analysis 2: Antisaccade correct versus antisaccade error. Lateral surfaces on an inflated cortical model displayed from 2 seconds to 8 seconds after the prompt. Sulci are the darker stripes and gyri are the lighter stripes.

subject looks in the wrong direction. But often subjects look initially towards the wrong direction, realize their error, and then attempt to correct it by looking towards the correct direction; this response is known as a self-correct. In other instances, subjects do not respond at all to the stimuli and remain fixated on the center.

The most common type of error, the self-correcting error, is also the most complicated. After making the initial error, the subject must first realize the error and then attempt to correct it. All of these steps occur in a short timespan, so fMRI is unable to separate these distinct processes. Conclusions drawn from this information may explain the increase in dorsolateral activation during error trials. In self-correcting errors, the subject still programs a correct antisaccade in the end, so presumably DLPFC activation would still occur. Activity due to other neural process that

accompany the realization of the error may cause the increase in activity among error antisaccades. Realization is not a component to correct antisaccade generation.

Because error trials of each type (no response, no self-correction, and self-correcting) are not numerous enough, functional data for all types were averaged together for comparison with correct trials. Isolating of each error type and analyzing them separately may yield different results. But given the low rate of no response and non-self-correcting errors, the number of trials that need to be performed would be impractical.

5 Analysis 3: MEG Study of Temporal Activation

5.1 Results

Figure 8 shows frames from a movie of brain activation during antisaccade performance obtained from MEG data. The movies reflect the neural activation of a single individual. One image was obtained every 15 ms between 1815 ms–1950 ms after the initial prompt ("O" or "X"). Target appearance occurred at 2000 ms. The sources of activation determined from the raw MEG data were plotted onto an inflated surface. Areas of activation are marked as yellow.

The timeframe used for the images was locked to the onset of DLPFC activation. Activation was shown to begin at approximately 1800 ms and lasted until 2035 ms (not shown).

5.2 Discussion

The information obtained regarding the timing of neural processes strengthened further the hypothesis that the DLPFC is important in programming an antisaccade. DLPFC activation was shown to occur approximately 200 ms before appearance of the target until 2035 ms, 35 ms after the target appears. During this interval before the actual eye movement, the subject makes preparations to react to the target, programming an antisaccade or prosaccade depending on the cue given. The MEG data shows that dorsolateral activation occurs right during that programming timeframe. While not absolute confirmation that the DLPFC is important to inhibition, this is nonetheless strong support.

The 200 ms presaccade timeframe for dorsolateral activity is similar to that in the study in [3], which found frontal activity specific to antisaccades 160–60 ms before the actual saccadic movement. The 100 ms of frontal inhibition prior to the antisaccade found in [7] is significantly less.

6 Conclusion

Strong evidence was found in this study supporting the hypothesis that the dorsolateral prefrontal cortex is an important component in saccadic inhibition. Both fMRI and MEG were used to provide high spatial and high temporal detail regarding the processes behind inhibition. Results from the fMRI phase of study show that the DLPFC is indeed more active during antisaccades than in prosaccades. Temporal data from MEG found that DLPFC activation occurs at the same time as when the inhibitory signal is sent. This suggests that dorsolateral activation is associated with the inhibitory response.

7 Acknowledgments

I thank my mentor, Dr. Dara Manoach, for her patient guidance and advice throughout my project. I would also like to extend my gratitude to Matt Cain, who walked me step by step to where I am now. I would also like to thank my tutor, Sapan Shah, for his help in organizing my paper and reading my drafts.

I am also grateful to the Center for Excellence in Education, Research Science Institute, the Massachusetts Institute of Technology, and all others that helped make this life-changing opportunity possible.

References

[1] Broerse, A., Crawford, T.J. & der Boer, J.A. Parsing cognition in schizophrenia using saccadic eye movements: a selective overview. *Neuropsychologia* **39**, 742–756 (2001).

[2] Chapman, L.J. & Chapman J.P. *Disordered thought in schizophrenia* (Appleton-Century-Crofts, New York, 1978).

[3] Clementz, B.A., McDowell, J.E. & Stewart, S.E. Timing and magnitude of frontal activity differentiates refixation and anti-saccade performance. *Neuroreport* **12**, 1863–1868 (2001).

[4] Dale, A.M., et al. Dynamic statistical parametric mapping: combining fMRI and MEG for

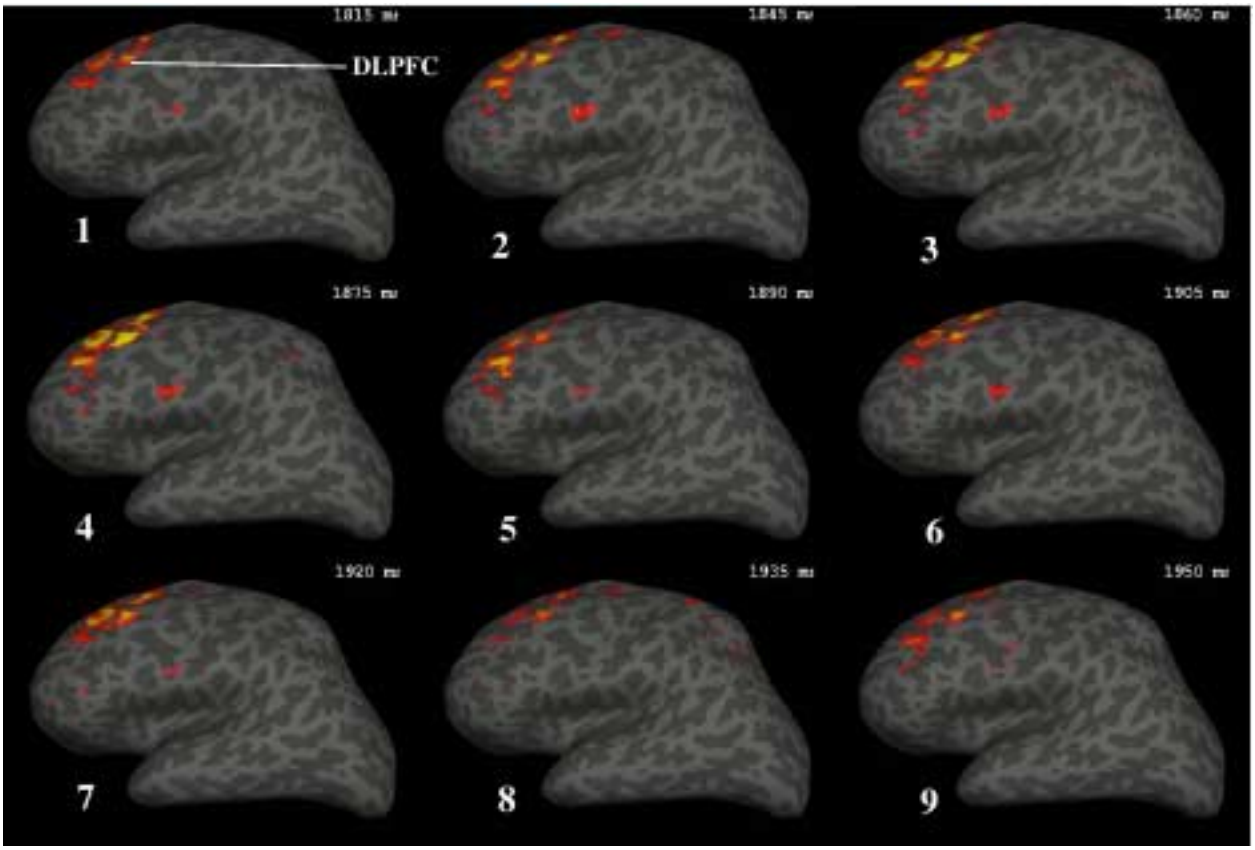


Figure 8: Frames from an MEG movie of activation in antisaccades from 1815ms to 1950ms. The left lateral hemisphere is shown. Sulci are the darker stripes and gyri are the lighter stripes.

high-resolution imaging of cortical activity. *Neuron* **26**, 55–67 (2000).

[5] Dale, A.M. & Buckner, R.L. Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping* **5**, 329–340 (1997).

[6] Doricchi, F., et al. Neural control of fast-regular saccades and antisaccades: an investigation using positron emission tomography. *Experimental Brain Research* **116**, 50–62 (1997).

[7] Evdokimidis, I., et al. Cortical potentials with antisaccades. *Electroencephalography and Clinical Neurophysiology* **98**, 377–384 (1996).

[8] Everling, S. & Fischer, B. The antisaccade: a review of basic research and clinical studies. *Neuropsychologia* **9**, 885–889 (1998).

[9] Everling, S. & Munoz, D.P. Neuronal correlates for preparatory set associated with pro-saccades and anti-saccades in the primate frontal eye field. *Journal of Neuroscience* **20**, 387–400 (2000).

[10] Fecteau, J.H. & Munoz, D.P. Exploring the consequences of the previous trial. *Nature Reviews Neuorscience* **4**, 1–9 (2003).

[11] Green, M.F. & Nuechterlein, K.H. Cortical oscillations and schizophrenia: timing is of the essence. *Archives of General Psychiatry* **56**, 1007–1008 (1999).

[12] Hallet, P. Primary and secondary saccades to goals defined by instructions. *Vision Research* **18**, 1279–1296 (1978).

[13] Manoach, D.S., et al. Schizophrenic Subjects Show Deficient Inhibition but Intact Task

Switching on Saccadic Tasks. *Biological Psychiatry* **51**, 816–826 (2002).

[14] McDowell, J.E., et al. Neural Correlates of Refixation Saccades and Antisaccades in Normal and Schizophrenia Subjects. *Biological Psychiatry* **51** 216–223 (2002).

[15] Mesulam, M.M. Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annual Neurol.* **28**, 597–613 (1990).

[16] O'Driscoll, G.A., et al. Differences in cerebral activation during smooth pursuit and saccadic eye movements using positron-emission tomography. *Biological Psychiatry* **44**, 685–689 (1998).

[17] Paus, T., et al. Role of the human anterior cingulate cortex in the control of oculomotor, manual, and speech responses, a positron emission tomography study. *Journal of Neuphysiology* **70**, 453-469 (1993).

[18] Sweeney, J.A., et al. Positron Emission Tomography Study of Voluntary Saccadic Eye Movements and Spatial Working Memory. *Journal of Neurophysiology* **75**, 454–468.

[19] Terao, Y., et al. Visualization of the informational flow through human oculomotor cortical regions by transcranial magnetic stimulation. *Journal of Neurophysiology.* **80**, 936–946 (1998).

Quantitative Trait Loci Modulating Corpus Callosum Size in the Mouse Brain

Ning Zhou

under the direction of
Dr. Glenn D. Rosen
BIDMC, Harvard Medical School

Abstract

The corpus callosum is the brain's primary pathway for communication between the left and right hemispheres, and abnormal function and anatomy of the corpus callosum have been associated with a variety of disorders in humans and animals. In this study, we mapped quantitative trait loci (QTLs) modulating corpus-callosum size in mice. By using mouse brain sections and stereological techniques, we estimated the midsagittal corpus-callosum size in 191 BXD recombinant inbred mice and their parent strains (C57BL/6J and DBA/2J). Analysis showed midsagittal corpus-callosum size to be a heritable trait ($h^2 = .42$). We detected a suggestive QTL modulating corpus-callosum size near the marker S10Gnf071.990 on chromosome 10 at approximately 73 Mb, in close proximity to a previously defined QTL modulating striatal volume. We also found correlations between corpus-callosum size and other behavioral and anatomical phenotypes. Overall, our results implicate a genetic role in modulation of corpus-callosum size.

1 Introduction

The corpus callosum is composed of bundles of nerve fibers that connect the neurons of the left and right hemispheres of the brain, thereby enabling interhemispheric communication. This commissure has been noted as being the most important pathway for the transfer of sensory, motor, and higher-order information between the two cerebral cortices of the brain, thereby coordinating interhemispheric integration of perceptual, mnemonic, learned, and volitional information [43]. Split-brain patients, who have their corpora callosa severed in an attempt to alleviate intractible epilepsy, provide a direct illustration of this. After their surgery, the two hemispheres of the brain cannot communicate directly with each other, and therefore response to stimuli directed to one hemisphere might differ from that directed at the other. For example, split-brain patients can name objects shown in the right visual field, which projects to the verbal left hemisphere. In contrast, they are unable to name objects shown in the left visual field (which projects to the nonverbal right hemisphere) but can identify them with their left hand [43]. It is interesting to note, however, that in patients with damaged, partially removed, or completely removed corpora callosa, behavior and everyday functions remain largely intact, and such patients show few outward signs of their condition [43]. Furthermore, it should be noted that lateralization, or specialization of brain functions by the two hemispheres, varies widely among individual brains. A function normally associated with one hemisphere may be controlled by the other hemisphere in any particular individual [1].

In humans, the corpus callosum consists of over 200 million nerve fibers aligned in a broad, thick band running from rostral to caudal and medial to lateral [1]. Nerve fibers mostly project to corresponding cortical areas (homotopic callosal connections), while some project to noncorresponding areas in the contralateral hemisphere (heterotopic callosal connections) [43]. Typically, the morphology of the corpus callosum is examined in the midsagittal plane and is divided into four parts: the rostrum, body, genu, and splenium (see Figure 1). Investigators have measured the area [8, 13], length [8, 13], and shape [5, 44] of the midsagittal corpus callosum, and have found that these parameters are associated with a variety of disorders. Studies have found that

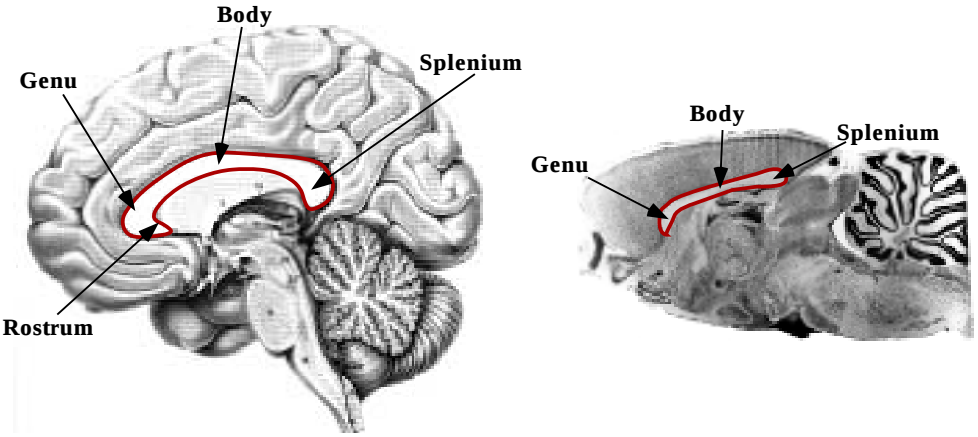


Figure 1: The corpus callosum in a human brain (left) and a mouse brain (right). Rostral is to the left and caudal is to the right.

patients with attention-deficit hyperactivity disorder have significantly smaller rostra than normal individuals [19], and callosal development and connectivity occurs abnormally in schizophrenics, as the corpus callosum develops bidirectionally [30] and causes confusion over interhemispheric transmissions [11]. Dyslexics have a smaller callosal genu [22] and larger splenium and total callosal area when compared to nondyslexics [16, 35]. Finally, patients suffering from Alzheimer's disease have significantly reduced callosal bodies [44].

The midsagittal area of the corpus callosum has been shown to be a reliable predictor of a number of biological traits. It is clear that certain factors, including sex and handedness, seem to account for some variation in corpus-callosum size. Women appear to have larger corpus-callosum to forebrain size ratios than men [25], and increased strength of handedness seems to suggest increased lateralization [25], although much debate remains about the role of these two factors and others in accounting for variations in corpus-callosum size [7].

Within animal species, the morphology and function of the corpus callosum are very similar to those in the human. The primary role of the corpus callosum in mammals appears to be interhemispheric communication and coordination of responses, and much variation lies in the structure of the corpus callosum. Surgical interruption of corpus-callosum function in animals has established the role of the corpus callosum in interhemispheric communication. For example, severing parts of the corpus callosum inhibits transfer of visual information in cats [34] and monkeys [4], of depth perception in cats [38], and of other sensory and cognitive functions [32]. Furthermore, as in the human, animal corpora callosa contain unevenly distributed nerve fibers in different callosal regions [23, 29], and most connections are homotopic, while a smaller number are heterotopic [24, 42].

Investigators have reported that strains of mice differ in midsagittal corpus-callosum area, suggesting a genetic role, and that genetic factors could underlie pathological and normal variation in corpus-callosum size [2, 26]. Wahlsten et al. looked at 21 strains and found a twofold difference in corpus-callosum size between strains of BTBR crossed with BALB/cWah1, 129P1/Rej, and 9XCA/Wah mice [47]. Rosen et al. found a difference in corpus-callosum size of a similar dimension with NZB/BINJ mice having larger corpora callosa than DBA/2J mice [41]. Le Roy et al. identified quantitative trait loci (QTL) linked to midsagittal corpus-callosum area in an F2 cross between NZB/BINJ and C57BL/6By strains and discovered QTLs on chromosomes 1 and 4 and a significant interaction between these two loci [31].

Recent research has found that changes in the size of brain structures in mouse such as the striatum [40] and cerebral ventricles [51] appear to be genetically modulated. In the latter study, recombinant inbred (RI) strains were used, which provide significant advantages in mapping precision when compared to the study of F2 mice [49]. Recombinant inbred strains are isogenic strains of mice derived from two parents through many generations of inbreeding. After suffi-

cient matings, a set of strains are created such that they have different homozygous alleles from the two parents at different locations in the genome. This allows for many possible phenotypic variations to be examined. In this experiment, we wished to confirm and extend the QTL findings of Le Roy [31] using alternate genetic and bioinformatic techniques. The objective of this study, therefore, is to map quantitative trait loci responsible for variation in the size of the corpus callosum in the mouse brain and to identify possible genes that could modulate corpus-callosum size. Specifically, we measured the midsagittal area of the corpus callosum in 34 BXD RI lines and their parent strains and mapped this phenotype against genome-wide markers.

2 Materials and Methods

2.1 Protocol

The midsagittal area of the corpus callosum was estimated from coronally cut sections of mouse brains from 191 mice representing 34 strains of recombinant inbred (RI) mice and their parent strains (C57BL/6J and DBA/2J). These traits were then mapped to 720 genome-wide polymorphic markers. Midsagittal corpus-callosum area was also correlated against approximately 430 known behavioral, anatomic, and electrophysiologic phenotypes as well as the expression levels of over 12,000 RNA transcripts using WebQTL (`http://webqtl.org`).

2.2 Subjects

We used 191 mice from 34 BXD RI strains as well as the two parent strains. There were, on average, about five mice per RI strain. All data for this study were obtained from the histological brain sections of the Mouse Brain Library (`http://mbl.org`). Mice were obtained from the Jackson Laboratory (Bar Harbor, ME) and prepared at BIDMC, Boston, MA, as detailed previously [40]. All procedures were approved by that institution's animal care and use committee and conform to NIH guidelines for the humane treatment of animals. Briefly, subjects were deeply anesthetized with Avertin (0.8 mL i.p.) and transcardially perfused with saline, followed by fixative (glutaraldehyde/paraformaldehyde), and their brains were removed and weighed. After variable postfixation times, the brains were embedded in 12% celloidin and sliced in the coronal plane at 30 μm. Two series of every tenth section, representing a one-in-five sample of the brain, were stained with cresyl violet and mounted onto glass slides. These series were offset by five sections wih the result being that we sampled every fifth section throughout the length of the corpus callosum.

2.3 Midsagittal area of the corpus callosum

The midsagittal area of the corpus callosum in each coronally sectioned mouse brain was estimated by measuring the dorsal-ventral length of the structure at the midsagittal region using StereoInvestigator (Microbrightfield Inc., Colchester, VT; see Figure 2). All measurements were made under 10X magnification with a Nikon Eclipse E800 microscope (Nikon Inc.). Wherever the corpus callosum appeared on a section, a reference line was drawn along the dorsal-ventral axis of the brain section by beginning a contour line at a point superior to the corpus callosum and ending the line at a point inferior to its ventral border. The length of the corpus callosum lying on that line was measured to within $\pm$ 2.5 μm for each section on which the corpus callosum appeared (5–11 sections per slide) and all measurements were recorded. The midsagittal area of the corpus callosum was then estimated for each slide using Cavalieri's rule, a method of estimating area and volume from serial section measurements [39]:

$$A = d\sum_{i=1}^{n} y_i - t y_{\mathrm{MAX}}, \tag{1}$$

where A is the estimated area, d is the distance between sections, t is the section thickness, y_i is the midsagittal corpus-callosum measurement, and y_{MAX} is the maximum measurement. In the cases where two reliable estimates of midsagittal area were available for each mouse brain ($N = 160$), they were averaged. In cases where there were missing sections from one of the slides ($N = 31$), the missing slide was removed from the analysis.

2.4 Analysis and genetic mapping

Intraobserver reliability was calculated by remeasuring ten randomly selected slides while blinded to their identity and calculating a Pearson product–moment correlation coefficient. The morphometric data were analyzed using ANOVA and multiple-regression techniques (JMP, SAS Institute, Cary, NC). QTL anal-

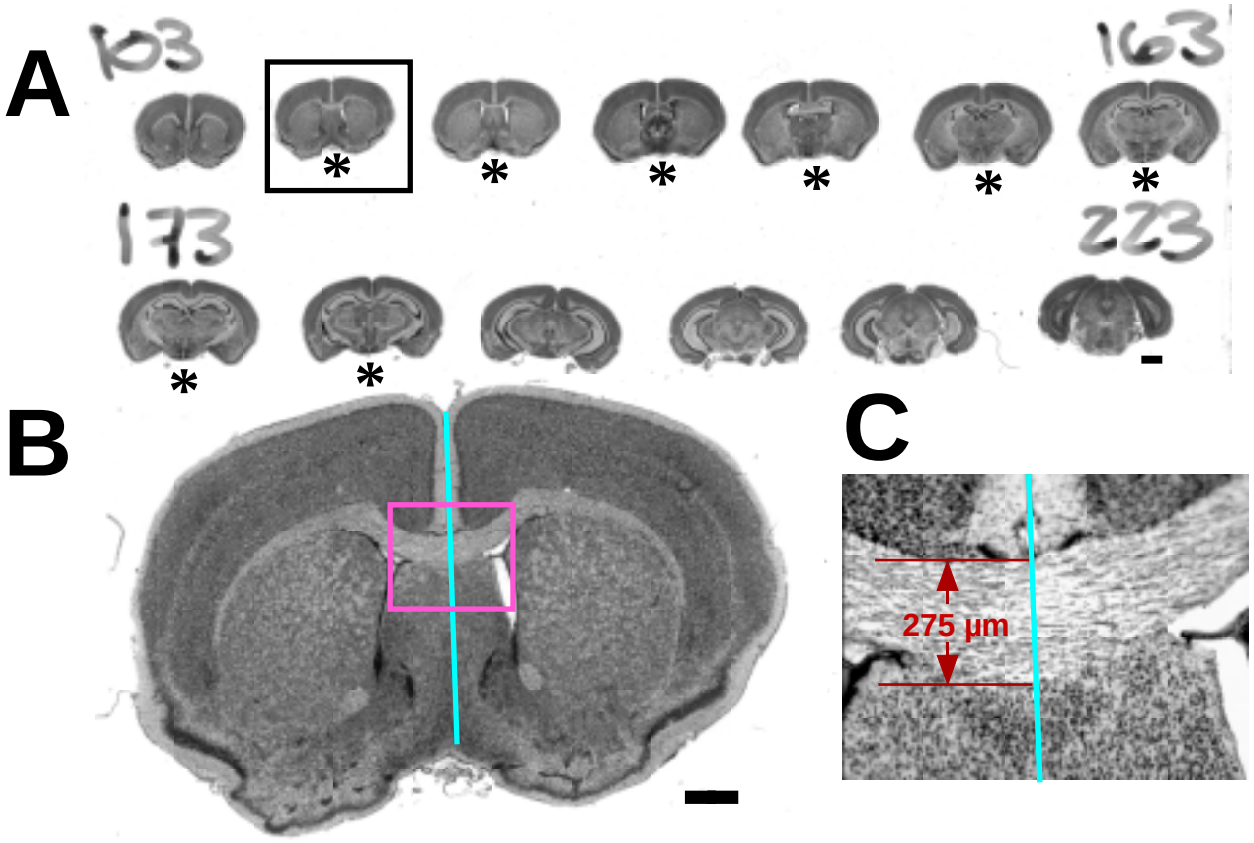


Figure 2: Panel A. Sections of coronally cut mouse brains on a microscope slide. Sections 113 through 183 (asterisked) contain visible corpora callosa. Scale = 1 mm. Panel B. One mouse brain section from Panel A. The corpus callosum is visible and runs from medial to lateral. The blue line is a reference line drawn along the dorsal-ventral axis. Scale = 500 μm. Panel C. The corpus callosum from rectangular area of Panel B. The midsagittal corpus-callosal length along the reference line is measured in μm.

ysis was carried out with WebQTL using a set of 720 markers distributed across all chromosomes in the BXD RI set. A likelihood ratio statistic (LRS) for linkage was computed, and significance was determined by permutation tests and bootstrap analysis [6, 46]. Using WebQTL, we correlated our anatomic phenotypes to a database of 430 published phenotypes of BXD RI sets (Published Phenotypes Database, WebQTL). Trait data was also compared using WebQTL to a gene expression database of over 12,000 forebrain RNA transcripts (UTHSC Brain mRNA U74Av2 Database).

2.5 Determining candidate genes

Genes or QTLs within three cM of any suggestive or significant QTL found in this study were examined using a bioinformatics database (Mouse Genome Informatics, `http://www.jax.org`). Any gene or locus having been connected to activity in the brain was examined, the literature on that gene reviewed, and its significance analyzed subjectively.

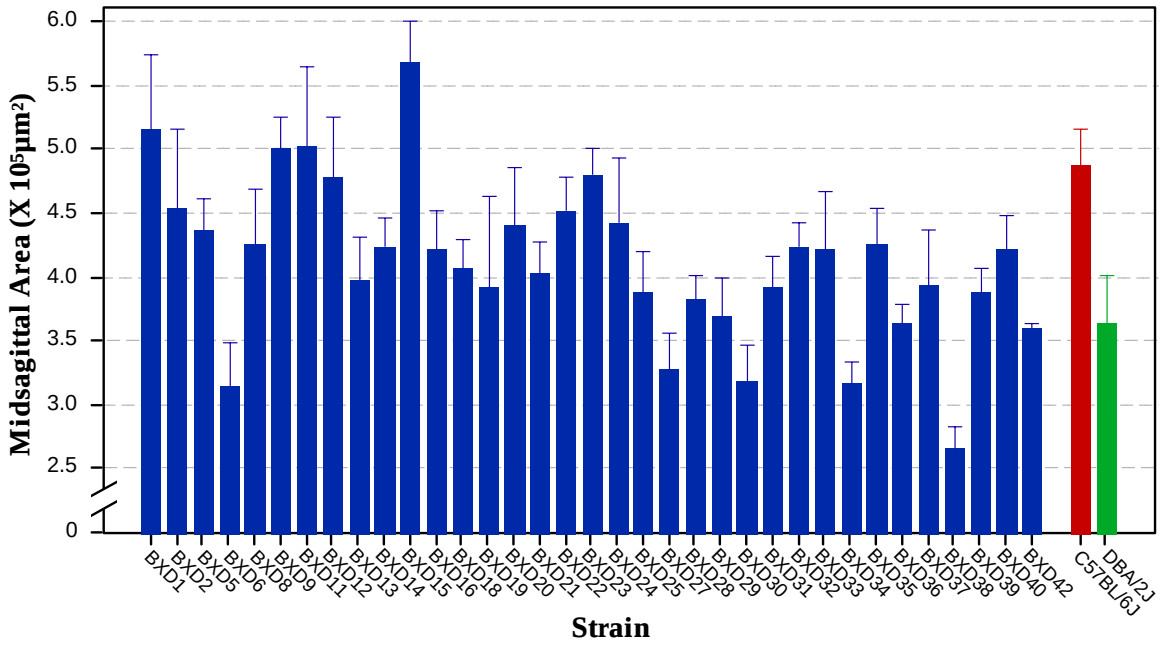


Figure 3: Histogram of mean midsagittal corpus-callosum area of BXD RI strains. Parent strains are on the right.

3 Results

3.1 Summary and reliability of measurements

To assess the reliability of the data, the midsagittal corpus-callosum areas of 10 randomly selected slides were blindly remeasured by the same observer. The test-retest reliability coefficient for midsagittal corpus-callosum area was 0.99. The average percentage difference between original and remeasured midsagittal areas was $\pm$ 0.6% μm^2.

The measured midsagittal corpus callosum areas of 191 subjects representing 34 strains and the two parent strains were included in the data analysis. 103 subjects were male and 88 were female. Ages ranged from 31 days to 493 days with an average of 93.4 $\pm$ 6.3 days and a median of 60 days. The midsagittal areas of the corpora callosa ranged from $2.26 * 10^5$ μm^2 to $7.57 * 10^5$ μm^2. Strain averages (mean $\pm$ SEM) ranged from $2.66 \pm 0.30 * 10^5$ μm^2 (BXD38) to $5.68 \pm 0.44 * 10^5$ μm^2 (BXD15; see Figure 3). The grand mean of midsagittal area of all 34 strains was $4.16 \pm 0.07 * 10^5$ μm^2 and the median was $4.08 * 10^5$ μm^2.

3.2 Heritability of midsagittal corpus-callosum size

ANOVA with strain as the independent variable and the midsagittal corpus-callosum area as the dependent variable indicated a significant effect of strain ($F_{33,138} = 2.75$, $P < .001$). Regression analysis using age, sex, and brain weight as predictors of midsagittal area of the corpus callosum was significant ($F_{3,168} = 23.6$, $P < .001$), with both age (t = 3.44, P < .001) and brain weight (t=5.92, P < .001) contributing significantly to the regression. Because these variables significantly predict midsagittal area of the corpus callosum, we extracted the residuals from this equation to begin our genetic mapping. This serves to remove the effects of age, brain weight, and sex from the phenotype. Strain remained a significant predictor of midsagittal corpus-callosum area ($F_{1,32} = 8.69$, $P < .01$) when using residuals as the dependent measure. The average within-strain and between-strain variances were calculated, yielding a heritability factor (h^2) of .42 [21].

3.3 Interval mapping shows a suggestive QTL

Linkage was computed with the residual midsagittal area of the corpus callosum (RMSACC) as the phenotype using WebQTL, which has over 720 polymorphic markers throughout the genome. WebQTL computes a likelihood ratio statistic (LRS) score, which reflects the chance that the phenotype is related to a specific genetic marker. This LRS score is mathematically identical to the more common logarithmic odd (LOD) score, with the LRS equal to LOD/4.6.

Using WebQTL to compute linkage with RMSACC as the phenotype, we found a suggestive QTL near marker S10Gnf071.990 (LRS = 12.4) on chromosome 10 at 73.28 Mb. Composite interval mapping (controlling for this QTL interval) revealed no additional significant QTL. ANOVA using genotypes ("B" or "D") of the strains at the chromosome 10 marker as the independent variable and midsagittal corpus-callosum size as the dependent variable showed a significant effect of genotype ($F_{1,32} = 8.70$, $P < .01$). Strains with a "B" allele in this interval had larger residuals and thus larger midsaggital corpus callosum sizes than those with a "D" allele (mean RMSACC $\pm$ SEM $= 2.48 \pm 1.04 * 10^5$ μm^2 vs. $-2.11 \pm 1.17 * 10^5$ μm^2, respectively).

3.4 Analysis of correlations with published phenotypes and gene expression levels

RMSACC values were compared to all values in the Published Phenotypes database using WebQTL and correlations were computed. Several interesting correlations were found (see Table 1). In particular, RMSACC seems to be negatively correlated with total hippocampal granule-cell growth in a study involving introducing adult BXD RI mice to a maze ($r = -.68$, $P < .02$; see Figure 5) [28]. There were no significant correlations between RMSACC and RNA transcript levels.

3.5 Possible candidate genes

One candidate QTL, Bsc10a (brain size control 10a), was found on chromosome 10 at 40 cM from the centromere, the exact area of the suggestive QTL we found. A previous study implicated this gene in modulation of mouse striatum volume and brain weight [40].

4 Discussion

Using measurements of midsagittal corpus-callosum area from 34 BXD RI strains of mice and their parent strains, we found that corpus-callosum size is a heritable trait, and mapped a suggestive QTL on chromosome 10 for corpus-callosum size. Further examination showed a strong negative correlation between corpus-callosum size and growth of adult hippocampal granule cells in BXD RI strains, as well as the presence of a QTL, Bsc10a, previously shown to modulate striatal volume and brain size independently, at the locus we found.

4.1 Variation in corpus-callosum size

There is much variation in corpus-callosum size in both humans and animals. The size of the corpus callosum is dependent on a variety of factors, such as the size of surrounding brain structures and diameters of nerve fibers, that are in part modulated genetically. Furthermore, we found age and brain weight to account for a significant portion of normal variation in corpus-callosum area. Variation can also be caused by environmental factors or genetic heritability or an interaction between both genes and environment. It should be kept in mind that genetic factors account for only a portion of the variability in corpus-callosum size and that the determination of heritability of corpus-callosum size and the mapping of QTLs assume that we are mapping a trait that could be indirectly influenced by other nongenetic traits.

4.2 Reliability of estimation of midsagittal corpus-callosum area

Measuring midsagittal corpus-callosum area has been accomplished mostly by examination of brain sections cut in the sagittal plane. In our experiment we did not have access to brains cut in this plane and were therefore forced to estimate midsagittal corpus-callosum area from coronal sections, which could lead to some possible sources of error. Nevertheless, reliability was very high (.99) and is unlikely to contribute much variation to the data. Similarly, Cavalieri's rule, the method used to approximate midsagittal area, is an accurate method of estimation for our purposes [39]. Although judgement of the experimenter in measurements under microscopy and use of Cavalieri's rule may result in some discrepancies from the actual midsagittal corpus-callosum area, the consistency of the measurements provides reliability

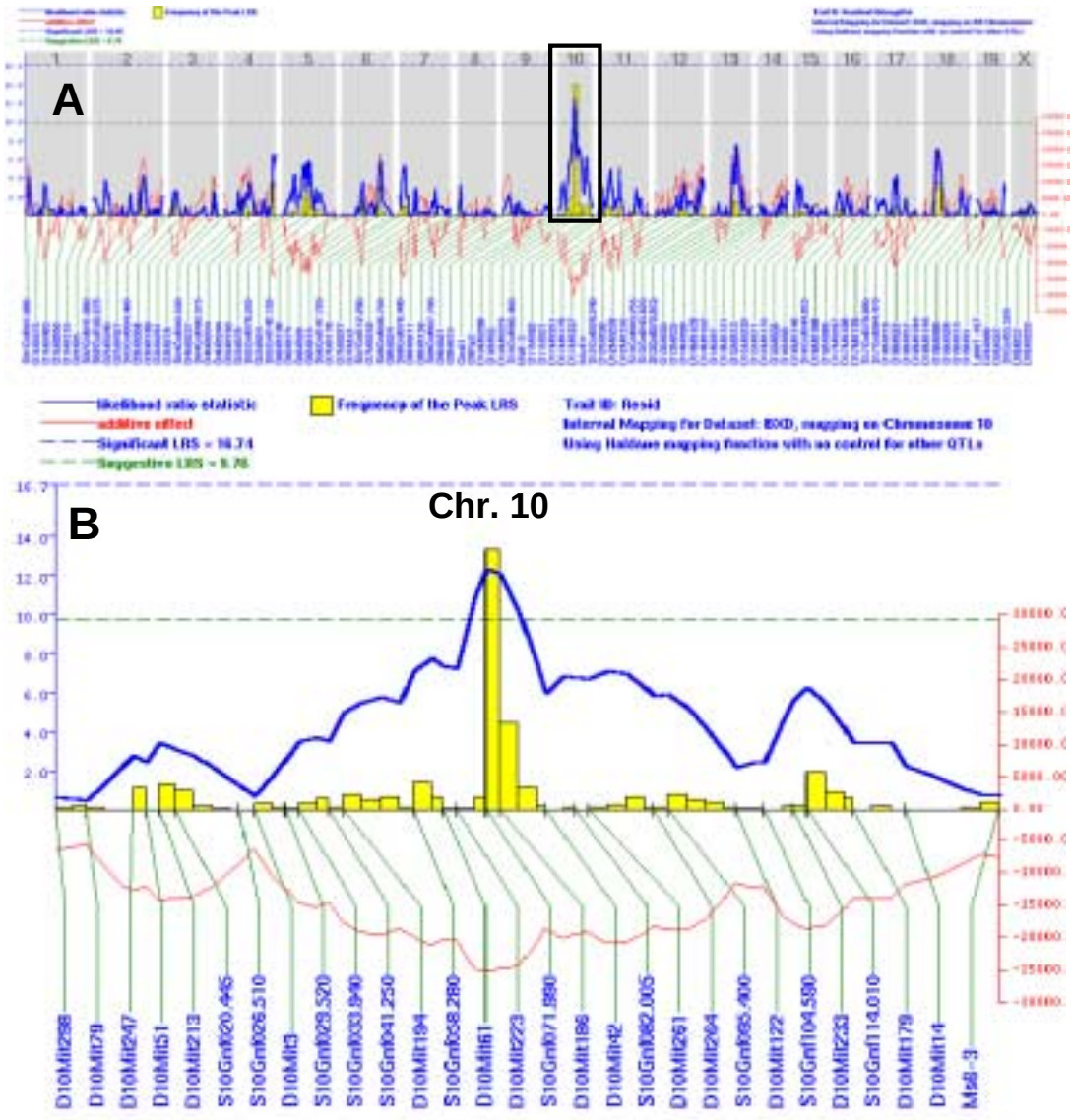


Figure 4: Panel A. LRS scores and bootstrap analysis for RMSACC across the entire genome. The solid blue line represents the LRS scores as determined by marker regression at each of the markers along the x-axis. The yellow histogram represents the results of the bootstrap analysis. The red line is the additive effect. Panel B. An expanded view of chromosome 10 from Panel A containing the suggestive QTL. The dotted line at LRS = 9.8 demonstrates level for suggestive QTL.

in intrastrain and interstrain comparisons. Furthermore, the ratio of midsagittal corpus-callosum areas of the parent strains in this study is very similar to the ratio of the same parent strains computed from another study [48]. We are therefore confident that our data represent a reliable and valid estimate of midsagittal corpus-callosum area.

#	Phenotype	Ref	Correlation	# Strains	P Value
1	Total Hippocampus Granule Cell Number	[28]	-.6924	11	.01840
2	% Freezing in Response to Auditory Stimulus US in Contextualized Fear Conditioning Paradigm	[36]	.4825	21	.0226
3	Saline Open Field Activity	[9]	-.4935	29	.0258
4	Saccharin Consumption	[3]	-.4831	20	.0298
5	Change in Body Temperature due to 4 mg/kg Methamphetamine	[20]	.4227	25	.0344
6	Ethanol Induced Conditioned Place Preference – Percent Time on Drug Paired Floor During First 30 Minutes of Test	[10]	.4670	20	.0369
7	Maximal Threshold to Ethanol-Induced Ataxia	[18]	.4171	25	.0372
8	Mean Life Span, Longevity in Days	[12]	.4396	22	.0398
9	Open-Field Activity Following Saline Injection	[37]	-.4526	20	.0442

Table 1: Correlations and P Values of RMSACC trait data versus values in the Published Phenotypes Database. The nine significant results are shown.

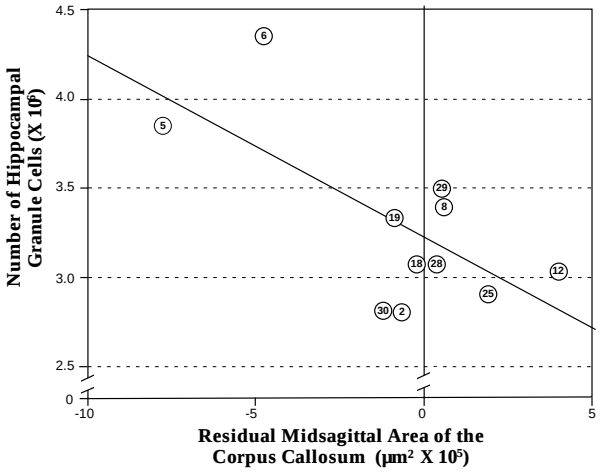


Figure 5: Scatter plot of RMSACC trait data versus hippocampal granule cell number in BXD mice. There is a high negative correlation between the two variables.

4.3 Heritability of Corpus-Callosum Size

We found that corpus-callosum size is a heritable trait in mice. This is consistent with other studies that have demonstrated interstrain differences in corpus-callosum size [31, 41]. Corpus-callosum size differs markedly in the BXD strains studied, including twofold differences (see strain BXD15 vs. BXD38) between strains. ANOVA confirmed that strain is a major determinant of corpus-callosum size in the BXD RI set and contributes significantly to variation in this structure. We estimated that narrow-sense heritability is approximately .42. This is relatively average for a morphometric trait and is similar to calculated heritabilities for ventricle size [51] and olfactory bulb size [50] in BXD RI strains.

4.4 Mapping showed suggestive QTL on chromomsome 10

Linkage analysis showed a suggestive corpus-callosum size QTL on chromosome 10. Examination of the ad-

ditive effect in the vicinity of the locus on chromosome 10 shows that the presence of a "B" allele from the C57BL/6J parents appeared to increase midsagittal corpus-callosum area. Examination of the parent strains for the BXD RI set confirm that DBA/2J mice have small corpora callosa, whereas C57BL/6J mice have appreciably larger corpora callosa. These results strongly suggest that the C57BL/6J parent contributes to an increase in corpus-callosum size. The fact that the "B" allele accounts for most of the additive effect to the trait value at chromosome 10 and the LRS score of 12.4 indicate that it can be likely that there is a gene at or near the QTL on chromosome 10 that modulates corpus-callosum size.

In 1998, Le Roy et al. also studied QTLs implicated in midsagittal corpus-callosum area and found in F2 generation mice from C57BL/By and NZB/BINJ parent strains two loci on chromosomes 1 and 4 that interacted significantly, but no QTL on chromosome 10 [31]. Our study did not find any suggestive or significant QTLs on chromosomes 1 and 4 and thus we cannot confirm the results of the prior study. This may be due to the fact that we used different parent strains; studied RI, not F2, mice; performed analysis with different sets of genetic markers; and used different methods of analysis—or it may be due purely to experimental variation.

4.5 Correlation with adult hippocampal granule cell growth

We considered the RMSACC as a phenotype and compared it to all values in the Published Phenotypes database to see whether there was a linkage or correlation with any other phenotype(s). The only notable correlation between between corpus-callosum size and a phenotype was with total hippocampal granule-cell number from a study that measured the number of granule-cell counts in BXD RI mice after they completed several trials of a water maze [28]. It has been suggested that hippocampal granule cells are generated in adulthood for the purpose of memory of and adaptation to new knowledge and new levels of complexity [17, 27]. Interestingly, the corpus callosum has been implicated in studies involving memory, specifically interhemispheric transfer of memory information [14, 15]. The high negative correlation between corpus-callosum size and growth of hippocampal granule cells could indicate that that increased numbers of fibers in the corpus callosum may offset some of the need to grow new adult granule cells in order to enhance memory of and adaptation to new knowledge or skill. However, one must always be careful of overinterpreting correlational data. A more direct test of this correlation could be further examined by monitoring hippocampal granule-cell growth versus original corpus-callosum size after adult mice of BXD RI strains are introduced to an intellectual task that requires memory and use of new knowledge.

4.6 Bsc10a as a candidate QTL

The only previously reported locus implicated with brain anatomy or function between 37 and 43 cM on chromosome 10 was Bsc10a (brain size control 10a) at 40.0 cM from the centromere. In a previous study, Bsc10a was found to have a significant independent effect on brain size as well as a significant effect on striatal brain volume [40]. In our study, marker regression analysis was performed using only RMSACC scores after brain weight had been factored out, but the QTL on chromosome 10, which lies in close proximity to Bsc10a, still registered up as a suggestive loci. Thus Bsc10a may have an independent effect on corpus-callosum size, especially since Bsc10a is the only QTL in proximity to our QTL. It is possible that Bsc10a modulates corpus-callosum size in addition to brain volume, since Bsc10a has already been implicated in the size of another brain structure, the striatum.

5 Conclusion

The corpus callosum is the most important and largest commisure connecting the two brain hemispheres, and is crucial in organizing interhemispheric communication involving sensory, motor, and higher-order information. While changes in the corpus-callosum structure and function have been recognized to be associated with several clinical manifestations and behavioral abnormalities, etiologies for the changes in corpus-callosum structure and function are not known. In this study, we examined the genetic modulation of normal variation in the midsagittal area of the corpus callosum. Our findings confirmed that corpus-callosum size is a highly heritable trait in mice. Furthermore, our QTL analysis identified a suggestive QTL on chromosome 10 where Bsc10a, a gene implicated with modulation of brain size and striatal volume, lies. Moreover, we found a correlation between corpus-callosum size and growth of adult hippocampal granule cells. These findings

support our hypothesis that variation in midsagittal area of the corpus callosum may be modulated by genes associated with brain size and function.

QTL analysis of small sets of RI strains is somewhat conditional. To expand on this study, statistical confidence could be improved by increasing precision of phenotypes, using larger RI sets, or generating RIX progeny [45], a set of RI F1 hybrids whose genetics allow further refinement of position and effects of QTLs.

This study has contributed significant new knowledge to the search for genetic modulation of corpus-callosum size. Hopefully, through new research building on the methods and results of this study, scientists will be able to isolate the primary genetic contributors to variation in mouse corpus-callosum size and extrapolate that knowledge to humans. Thus, the significance of this study is to eventually promote genetic measures to prevent diseases of or related to the corpus callosum through medical technologies such as gene therapy and stem-cell transplantation.

6 Acknowledgments

I would like to acknowledge my mentor, Glenn D. Rosen, Ph.D. (BIDMC, Harvard Medical School), for designing my project, providing me with the knowledge and equipment I needed to complete my research, and being an excellent mentor to me and always providing me with his full support.

I would also like to thank Robert Williams, Jintao Wang, and Kenneth Manly for creating and maintaining WebQTL and providing me with its use for my data analysis.

Furthermore, I would like to thank the Center for Excellence in Education for organizing and running the Research Science Institute, the program which provided me with the opportunity to do my research and supported my work.

Finally, I would like to thank my tutor, Sapan Shah, for directing my work and assisting in editing my paper.

References

[1] Aboitiz, Francisco et al.: Corpus Callosum Morphology in Relation to Cerebral Asymmetries in the Postmortem Human. *The Parallel Brain*, 2000, Zaidel, Eran, and Marco Iacoboni, eds. London, MIT Press, pp. 33–46.

[2] Barkovitch, A.J. and D. Norman.: Anomalies of the Corpus Callosum: Correlation with Further Anomalies of the Brain. *American Journal of Radiology*, 1988; 151:171–179.

[3] Belknap, J.K. et al.: Single-locus Control of Saccharin Intake in BXD/Ty Recombinant Inbred (RI) Mice: Some Methodological Implications for RI Strain Analysis. *Behavioral Genetics*, 1992; 22(1):81–100.

[4] Berlucchi, G.: Commissurotomy Studies in Animals. *Handbook of Neurophychology*, 1990, Vol. 4, Pompeiano, O., and C. Ajmone-Marsan, eds. Amsterdam: Elsevier, pp. 9–47.

[5] Bookstein, Fred L.: Morphometrics for Callosal Shape Studies. *The Parallel Brain*, 2000, Zaidel, Eran, and Marco Iacoboni, eds. London, MIT Press, pp. 75–92.

[6] Churchill, G.A. and R.W. Deorge: Empirical Threshold Values for Quantitative Trait Mapping. *Genetics*, 1994; 138:963–971.

[7] Cowell, Patricia E.: Size Differences in the Callosum: Analysis Beyond the Main Effects. *The Parallel Brain*, 2000, Zaidel, Eran, and Marco Iacoboni, eds. London, MIT Press, pp. 64–69.

[8] Cowell, Patricia E. et al.: Human Corpus Callosum: A Stable Mathematical Model of Regional Neuroanatomy. *Brain Cognition*, 1994; 25:52–66.

[9] Crabbe, J.C. et al.: Polygenic and Single-gene Determination of Responses to Ethanol in BXD/Ty Recombinant Inbred Mouse Strains. *Neurobehavioral Toxicological Teratology*, 1983; 5(2):181–187.

[10] Cunningham, C.L.: Localization of Genes Influencing Ethanol-induced Conditioned Place Preference and Locomotor Activity in BXD Recombinant Inbred Mice. *Psychopharmacology*, 1995; 120(1):28–41.

[11] David, A.S.: Tachistoscopic Tests of Colour Naming and Matching in Schizophrenia: Evidence for Posterior Callosum Disfunction? *Psychological Medicine*, 1987; 17 (3):621–630.

[12] de Haan, G. and G. Van Zant.: Genetic Analysis of Hemopoietic Cell Cycling in Mice Suggests its Involvement in Organismal Life Span. *FASEB*, 1999; 13(6):707–713.

[13] Denenberg, V.H. et al.: A Factor Analysis of the Human's Corpus Callosum. *Brain Research*, 1991; 548:126–132.

[14] Doty, R.W. et al.: Forebrain Commisures and Visual Memory. *Behavioral Brain Research*, 1988; 29:267–280.

[15] Doty, R.W. et al.: Hemispheric Distribution of Memory Traces. *The Memory System of the Brain*, 1994, Delacour, J. ed. Singapore: World Scientific, pp. 636-656.

[16] Duara, R.A. et al.: Neuroanatomic Differences Between Dyslexic and Normal Readers on Magnetic Resonance Imaging Scans. *Archives of Neurology*, 1991; 48:410–416.

[17] Duman, R.S. et al.: Neuronal Plasticity and Survival in Mood Disorders. *Biological Psychiatry*, 2000; 48:732–739.

[18] Gallaher, E.J. et al.: Identification of Genetic Markers for Initial Sensitivity and Rapid Tolerance to Ethanol-induced Ataxia Using Quantitative Trait Locus Analysis in BXD Recombinant Inbred Mice. *Journal of Pharmacological Experimental Therapies* 1996; 277(2):604–612.

[19] Geidd, Jay N. et al.: Quantitative Morphology of the Corpus Callosum in Attention Defecit Hyperactivity Disorder. *American Journal of Psychiatry*, 1994; 151:665–669.

[20] Grisel, J.E. et al.: Quantitative Trait Loci Affecting Methamphetamine Responses in BXD Recombinant Inbred Mouse Strains. *Journal of Neuroscience*, 1997;17(2):745–754.

[21] Hegmann, J.P., and B. Possidente: Estimating Genetic Correlations from Inbred Strains. *Behavioral Genetics*, 1981; 11:103–114.

[22] Hynd, G.W. et al.: Dyslexia and Corpus Callosum Morphology. *Archives of Neurology*, 1995; 52:32–38.

[23] Innocenti, Giorgio M.: General Organization of Callosal Connections in the Cerebral Cortex. *Cerebral Cortex*, 1986, Vol. 5, Jones, E.G., and A. Peters, eds. New York: Plenum, pp. 291–353.

[24] Innocenti, Giorgio M. and Raymond Bressoud: Callosal Axons and Their Development. *The Parallel Brain*, 2000, Zaidel, Eran, and Marco Iacoboni, eds. London, MIT Press, pp. 11–26.

[25] Jancke et al.: Brain Size: A Possible Source of Interindividual Variability in Corpus Callosum Morphology. *The Parallel Brain*, 2000, Zaidel, Eran, and Marco Iacoboni, eds. London, MIT Press, pp. 51–63.

[26] Jeret, J.S. et al.: Clinicopathological Findings Associated with Ageneis of the Corpus Callosum. *Brain Development*, 1987; 9:255–264.

[27] Kempermann, G.: Why New Neurons? Possible Functions for Adult Hippocampal Neurogenesis. *Journal of Neuroscience*, 2002; 22:635–638.

[28] Kempermann, G. and F.H. Gage: Genetic Determinants of Adult Hippocampal Neurogenesis Correlate with Aquisition, but not Probe Trial Performance, in the Water Maze Task. *European Journal of Neuroscience*, 2002; 16:129–136.

[29] Kennedy, H. et al.: Callosal Pathways and Their Compliance to General Rules Governing the Organization of Corticocortical Connectivity. *Vision and Visual Dysfunction*, Vol. 3.: *Neuroanatomy of the Visual Pathways and Their Development*, 1991, Dreher, B., and S. Robinson, eds. London: Macmillan, pp. 324–359.

[30] Kier, E.L. and C.L. Truwit: The Normal and Abnormal Genu of the Corpus Callosum: An Evolutionary, Embryologic, Anatomic, and MR Analysis. *American Journal of Neuroradiology*, 1996; 17:1631–1641.

[31] Le Roy, Isabelle et al.: Quantitative Trait Loci Implicated in Corpus Callosum Midsaggital Area in Mice. *Brain Research*, 1998; 811:173–176.

[32] Lassonde, M., and M.A. Jeeves.: Callosal Agenesis: A Natural Split-Brain? *Advances in Behavioral Biology*, 1994. New York: Plenum, pp. 275–284.

[33] Manly, K.F. et al.: Map Manager QTX, Cross-Platform Software for Genetic Mapping. *Mamm Genome*, 2001; 12:930–932.

[34] Myers, R.E. Functions of the Corpus Callosum in Interocular Transfer. *Brain*, 1956; 79:358–373.

[35] Njiokiktjien, C.L. de Sonneville, and J. Vaal: Callosal Size in Children with Learning Disabilities. *Behavioral Brain Research*, 1994; 64:213–218.

[36] Owen, E.H. et al.: Identification of Quantitative Trait Loci Involved in Contextual and Auditory-cued Fear Conditioning in BXD Recombinant Inbred Strains. *Behavioral Neuroscience*, 1997; 2:292–300.

[37] Plomin, R. et al.: Use of Recombinant Inbred Strains to Detect Quantitative Trait Loci Associated with Behavior. *Behavioral Genetics*, 1991; 21(2):99–116.

[38] Rivest, J. et al.: Interhemispheric Depth Judgement. *Neurophsychologia*, 1994; 32 (1):69–76.

[39] Rosen, Glenn D. and Jason D. Harry: Brain Volume Estimation from Serial Section Measurements: A Comparison of Methodologies. *Journal of Neuroscience Methods*, 1990; 35:115–124.

[40] Rosen, Glenn. D. and R.W. Williams.: Complex Trait Analysis of the Mouse Striatum: Independent QTLs Modulate Volume and Neuron Number. *BMC Neuroscience*, 2001; 2:5.

[41] Rosen, Glenn D. et al.: The Midsagittal Area of the Corpus Callosum and Total Neocortical Volume Differ in Three Inbred Strains of Mice. *Experimental Neurology*, 1990; 107:271–276.

[42] Rosenquist, A.C., Connections of Visual Cortical Areas in the Cat. *Cerebral Cortex*, *The Visual Cortex*, 1985, Vol. 3, A. Jones, D.G., and A. Peters, eds. New York: Plenum, pp. 81–117.

[43] Springer, Sally P. and Georg Deutsch: *Left Brain Right Brain*, Freeman & Co., 1985, New York.

[44] Thompson, Paul M. et al.: Mapping Structural Alterations of the Corpus Callosum During Brain Development and Degeneration. *The Parallel Brain*, 2000, Zaidel, Eran, and Marco Iacoboni, eds. London, MIT Press, pp. 93–130.

[45] Threadgill, D.W. et al.: Genetic Dissection of Complex and Quantitative Traits: From Fantasy to Reality Via a Community Effort. *Mamm Genome*, 2002; 13:175–178.

[46] Visscher, P.M. et al.: Confidence Intervals in QTL Mapping by Bootstrapping. *Genetics*, 1996; 143:1013–1020.

[47] Wahlsten, Douglas et al.: Genetic and Developmental Defects of the Mouse Corpus Callosum. *Experimentia*, 1989; 45:828–838.

[48] Wahlsten, Douglas et al.: Survey of 21 Inbred Mouse Strains in Two Laboratories Reveals that BTBR T/+ *tf/tf* Has Severely Reduced Hippocampal Commisure and Absent Corpus Callosum. *Brain Research*, 2003; 971:47–54.

[49] Williams, R.W. et al.: The Genetic Structure of Recombinant Inbred Mice: High-resolution Consensus Maps for Complez Trait Analysis. *Genome Biology*, 2001a; 2:Research0046.

[50] Williams, R.W. et al.: Genetic Dissection of the Olfactory Bulbs of Mice: QTLs on Four Chromosomes Modulate Bulb Size. *Behavioral Genetics*, 2001b; 31:61–77.

[51] Zygourakis, Corinna C. and Glenn D. Rosen.: Quantitative Trait Loci Modulate Ventricular Size in the Mouse Brain. *The Journal of Comparative Neurology*, 2003; 461:362–369.

Abstracts of Award-Winning Oral Presentations

Replacement of the α-3 helix in Zα_{ADAR1} with amino acids from Zα_{DLM1} and SARS and binding of chimera proteins to Z-DNA

Yong He Chong
Raffles JC
Singapore

Under the direction of
Dr. Ky Lowenhaupt
MIT

Zα is the Z-DNA binding domain of ADAR1 and adopts a helix-turn-helix-motif structure of which the α-3 helix is the specific region which binds to Z-DNA. Z-DNA binding is responsible for the pathogenicity of certain viruses like vaccinia. This experiment aims to replace the α-3 helix of Zα_{ADAR1} with 17 amino acids that have a similar sequence from DLM1, another Z-DNA binding protein, and the SARS virus. The Z-DNA binding activity of the chimera proteins were then checked using circular dichroism. Results showed that the DLM1 chimera protein was successfully produced and bound specifically to Z-DNA, even in low concentrations, suggesting that it is possible to replace the α-3 helix of Zα_{ADAR1} while maintaining the stability and function of the protein. However, the DNA sequence coding for the amino acids from SARS could not be inserted into the Zα_{ADAR1} gene and hence no chimera protein was produced. Further studies must be carried out to determine the exact structure of the DLM1 chimera protein, and an assay can be developed to substitute similar α-3 helixes from viral proteins into Zα_{ADAR1} to test for its Z-DNA binding ability.

Semiconducting Phase-Change Materials

Michael Coulter
Oregon Episcopal School
Portland, OR

Under the direction of
Dr. Jagadeesh Moodera
Francis Bitter Magnet Laboratory

Germanium-antimony-tellurium (GST) has been studied extensively for its applications as an erasable data storage material. GST is a phase-change material that can be changed from crystalline to amorphous and back very quickly. The two phases have different optical properties and so a laser can easily distinguish between a crystalline and an amorphous spot. A recent study has shown that the resistance of GST also can be easily changed, and when combined with probe technology, GST could store much more information than with optical change. The purpose of this project was to investigate the electrical properties of crystalline and amorphous GST. GST was grown as a thin film in a high-vacuum evaporation system. Both crystalline and amorphous films were grown. The optical properties were tested and x-ray diffraction and atomic force microscopy were used to determine the phase of the thin films that were tested. The results show that there is a difference in resistivity of several orders of magnitude between the crystalline and the amorphous phases. These results are very promising for applications in probe technology and should be investigated further. Future experiments should include testing GST with probe technology to determine what current is needed for writing and reading and how small a spot can be created.

Wireless Sensor Networks and Smart Materials

Jonas Ketterle
Brookline HS
Brookline, MA

Under the direction of
Prof. Ruaidhri M. O'Connor
MIT

This research examined advanced wireless sensor network technology and its applications. First, a model structure in the form of a vertical cantilever was instrumented with a wireless sensor device. The sensor measured the acceleration at the tip of the cantilever during free oscillation. Then, the results were used to determine the natural damping of the system and characterize the performance of the sensor network. Secondly, a piezoelectric damping system was developed to reduce acceleration at the tip of the cantilever. Due to the large displacement of the cantilever, increased damping was unsuccessful because the piezoelectric wafers are more sensitive to small amplitudes and high-frequency excitation. However, examination of the power dissipation of the external resistor circuit led to the conclusion that the piezoelectric wafers can be used to harvest power for wireless sensor networks.

On the Properties of Protection Systems in Finite Random Networks

Jie Tang
Bellevue HS
Bellevue, WA

Under the direction of
Dr. Muriel Médard
MIT

The topic of protection and recovery systems in near-infinite networks is one that has received a great deal of attention. This paper examines finite, random networks and presents a modified breadth-first search algorithm for finding protection cycles in such networks. It then compares how certain properties such as average cycle lengths compare to theoretical predictions for infinitely large networks. The average protection-cycle lengths in small networks of up to two thousand nodes exhibit an unexpected downward trend as the number of nodes increases, though it still fits the established model.

Investigations of the Number Derivative

Linda Westrick
MLW Governor's School
Mechanicsville, VA

Under the direction of
Mr. Pavlo Pylyavskyy
MIT

This paper explores the *number derivative*, n', which is defined in terms of the prime factorization of a positive integer n. We find an explicit formula and bounds for the function and investigate how the prime factorizations of n and n' are related. We then extend the function (originally defined only on $\mathbb{Z}^+ \cup \{0\}$) to the rational numbers and discuss its limits and continuity. Finally, we analyze several equations involving the derivative and seek the conditions under which they have a solution.

Abstracts of Other Reports

Nuclear Power for Deep Space Applications

Hachem Alaoui Soce
Casablanca American Sch
Casablanca

Under the direction of
Prof. Michael J. Driscoll
MIT

This paper proposes a design for a spacecraft engine that could power manned and unmanned deep space missions and gather data about the outer part of the solar system. The craft will be launched using a conventional chemical propulsion system. Once at a safe distance from Earth, a fission reactor will be activated, and will serve as the main power source. The thermal energy produced by the reactor will be converted to electricity by means of a Stirling engine. The electricity will then be used to power a nuclear electric propulsion system that uses a liquid-xenon propellant stored in cooled tanks near the spacecraft's nozzle. The rest of the energy generated will be used to power instruments on board the spacecraft as well as any life support. Excess heat will be discarded through black-body radiation using a lightweight treated graphite radiator. Finally, a depleted-uranium shield will be used to protect the payload from harmful radiation. Using nuclear power significantly increases the choice of scientific instruments that can be used on board, thus increasing the potential knowledge that could be acquired as a result of these missions.

Optimization of a Sound Module Compatible with the Lego™ Programmable Brick

Hans Anderson
Bear Creek HS
Evergreen, CO

Under the direction of
Mr. John Galinato
Build-it-Yourself

We built a sound module for the website, the Constructopedia, an online database of functional building modules that enable a broad range of kids to build creative and engaging robotics projects. Our sound module had to be price-effective as well as easy for kids aged 8–14 to build, to interface with the RCX microcontroller using only one of its 3 ports, and to have higher volume and sound quality than the competition. In 6 weeks we were able to accomplish this goal of a simple sound module.

Variations on the Knight Problem

Baris Aygoren
Izmir Fen Lisesi
Izmir, Turkey

Under the direction of
Mr. Huadong Pang
MIT

We discuss the problem of minimizing the number of moves a knight requires to move between chessboard squares. The case of the infinite chessboard is considered and a formula is found. The case of a finite board is discussed as possible future work.

Covering lattice points with circles

Samer Bazzi
Rafic Hariri HS
Saida, Lebanon

Under the direction of
Mr. Michael Ching
MIT

We study the circles of smallest radius containing n lattice points under any translation in the plane. We consider the special cases $1 \leq n \leq 5$ in search of a general approach and find an upper and a lower bound for the general solution.

Molecular and Anatomical Characterization of Two New Zebrafish Mutations

Leslie Beh
Raffles JC
Singapore

Under the direction of
Dr. Craig Albertson
The Forsyth Institute

The zebrafish (*Danio rerio*) is a useful model organism that has been widely used in developmental studies. It has also become an increasingly important tool for studying the mechanisms behind craniofacial development. Recently, a large-scale ENU mutagenesis screen has been carried out, producing a large number of new zebrafish mutants. In this project, two new mutations in the zebrafish: *forsyth flathead* (*ffh*) and *ghost* (*gho*) were characterized on both the anatomical and the molecular level. Anatomical characterization revealed that mutants with the

less severe *ffh* mutation had less severe anatomical defects, including a fused mandible, deformed palatoquadrates, misarticulation of the palatoquadrates with the hyosymplectic, inverted ceratohyals, and deformed and missing ceratobranchial arches. Mutants with the more severe *gho* mutations exhibited a massive reduction in cartilage and hindbrain tissue, leaving only the notochord, basicranium and a mandible of reduced size in cleared and stained mutants. Molecular characterization showed reduced levels of chm1 expression and dlx2 expression in *ffh* mutants and reduced levels of pax2a, chm1, and possibly dlx2 expression in *gho* mutants. However, the reduced chm1 expression in the pectorial fins in both mutations did not prove that differentiation of cranial neural crest cells was affected because the chondrocytes that gave rise to the pectorial fins are derived from the mesenchyme, and not from the neural crest. Hence, the migration of neural crest cells is affected in *ffh* mutants while the induction and possibly the migration of neural crest cells are affected in the *gho* mutants.

Theory of Polarization Transfer in a One-Dimensional Array of Quantum Dots

Victor V. Borzov
I.V. Kurchatov School 1189
Moscow, Russia

Under the direction of
Dr. Dmitry Novikov
MIT

The possibility of polarization transfer through a one-dimensional array of quantum dots is proposed in connection with near-field optical microscopy. Such an array could serve as an optical probe that could allow one to resolve an optical excitation on a scale much smaller than the optical wavelength.

Polarization transfer through the quantum-dot array is studied via mapping this problem onto a chain of coupled harmonic oscillators. We consider the cases of both the finite and infinite array and study the effect of boundary conditions in the finite system on the effectiveness of the polarization transfer.

Reminiscence in Rotary Pursuit

Kyla Bye-Nagel
Hanau HS
Grundau-Lieblos, Germany

Under the direction of
Daniel Press, M.D.
BIDMC/Harvard Med

The consolidation theory is the accepted theory for explaining reminiscence. Reminiscence is learning a skill while resting. The part of this theory that deals with reminiscence, called action decrement, says that while the brain is consolidating learning, the actual performance of the task that is being consolidated is smaller than the skill the brain has to do the task. As soon as the brain is done consolidating, the performance increases dramatically. This study attempted, by trying to disrupt the consolidation process, to prove or disprove the consolidation theory with respect to reminiscence. This study was, however, unable to conclude whether the consolidation theory accurately describes how reminiscence takes place.

The Effects of Natural Selection and Mutation on Inbreeding Effective Size

Can Cenik
Home-schooled
Ankara, Turkey

Under the direction of
Fangyun Yang
MIT

In this paper Ryman's effective-size model is extended by integrating basic natural-selection and mutation models into it. In addition a novel natural-selection model is developed for a sequence of generations that has fluctuating numbers of individuals in each generation. A modified Poisson distribution is used to model different productivities of individuals and therefore natural selection. Finally, we review the related literature to propose future work on the subject.

On Removing All Edges From the Complete Graph

Yuyin Chen
Cranbrook Upper School
West Bloomfield, MI

Under the direction of
Ms. Fu Liu
MIT

We seek to determine the minimum number of steps required to remove all edges from a complete graph, given that at most w edges may be removed in

one step and only one edge per component may be removed in one step. We calculate upper bounds for the cases $w = 2, 3, 4, 5$, and 2^k, and we give a method to calculate upper bounds for all w. In addition, we present a lower bound, for all w, on the number of steps required to remove all edges from a complete graph.

Hydrophobic Water-methane Interactions

Ann Chi
Terre Haute South Vigo HS
Terre Haute, IN

Under the direction of
Dr. Nicola Marzari
MIT

Understanding the hydrophobic interaction is key to understanding how many biological processes occur. Previous work by various researchers has attempted to identify the causes of the hydrophobic interaction, but no consensus has been reached. This work is the first to employ *ab-initio* molecular dynamics to obtain statistical ensembles for the calculation of radial distribution functions for a water-methane system. Comparison of the radial distribution functions for bulk water and a 63-water-1-methane system shows that the addition of a methane molecule weakens the structure of the first solvation shell of water. Further simulation should be performed to obtain the permanence of water molecules in the solvation shells in order to estimate the strength of the hydrophobic interaction.

Feasibility Study of Colloidal Quasicrystals

Cynthia Chi
Clements HS
Sugar Land, TX

Under the direction of
Dr. Rebecca Christianson
Harvard

Quasicrystals are highly ordered, aperiodic structures that have been produced in a number of alloy systems. They generally exhibit fivefold symmetry, which was previously thought to be impossible in ordered structures. Due to their unusual structure and geometry, quasicrystals exhibit unique patterns of diffraction and reflection. Colloid systems have long been used as models for study for atomic systems, and colloidal quasicrystals, if discovered, could be used in various fields, such as photonics. Quasicrystals in alloys were studied in order to determine structural properties of the general quasicrystalline structure. Two-component systems were deemed too physically difficult to create. One-component and three-component colloidal systems were then created to model alloy quasicrystalline systems and observed for any possible formation of quasicrystals. Although no quasicrystals were produced, confocal microscopy and analysis of the distribution of interparticle separations showed imperfect crystallization with no long-range order and phase separation of a binary crystal.

Developing a Novel Mechanism for Rapid Cloning of Tumor Suppressor Genes via Polymorphic Analysis

Jason G.L. Chu
Charter School of Wilmington
Hockessin, DE

Under the direction of
Dr. Mia MacCollin
Harvard Medical School/MGH

We conjecture that truncating mutations causing tumor suppressor gene (TSG) syndromes may be detected via polymorphic analysis of unequal expression. This hypothesis is based on the postulates that TSG mutations are generally truncating mutations and that truncating mutations result in unstable RNA products.

This novel screening mechanism is tested by analyzing two genes of high interest for the schwannomatosis (SCH) mutation via two polymorphic regions—D22S1167 (designated OE1) and *loc*150271 (OE2). OE1 contains a GT-GT repeat length polymorphism, and OE2 contains three single nucleotide polymorphisms (SNPs). OE1 is determined not to contain the germline mutation for SCH, while no informative polymorphisms were discovered for OE2, and thus no conclusions could be drawn.

As a positive control for the screening method, neurofibromatosis-2 (NF2) patients known to be highly informative for NFP3 were also tested for unequal expression, and the results used as a demonstration of the assay for mutation.

The Performance Characterization and Optimization of Polymer Electrolyte Membrane Fuel Cell Systems

Karen L. Chu
State College Area HS
State College, PA

Under the direction of
Prof. Yang Shao-Horn
MIT

PEMFCs convert hydrogen gas fuel into electricity

through an electrochemical reaction. The efficiency of a single polymer electrolyte membrane fuel cell (PEMFC) unit and the optimum performance conditions for such a unit were determined. Potentiodynamic and open circuit voltage (OCV) tests were conducted on the fuel cell unit with manipulations of the operational temperature, fuel gas pressure, and membrane composition. In addition, the efficiency of production of hydrogen fuel by electrolysis of water was obtained in order to study the efficiency of fuel production as well as fuel cell operation. It was found that the fuel cell unit operated best at 75°C, at higher pressures, with thinner membranes, and with higher catalyst loading.

Rendering Terrain for StarLogo in 3D Using OpenGL

Robert Crowell
White Station HS
Memphis, TN

Under the direction of
Dr. Eric Klopfer
MIT

StarLogo, an implementation of the Logo programming language, was rewritten to exist in a three-dimensional environment. However, this new environment is of limited value unless an easy and powerful means of editing the 3D landscape is implemented. Therefore, we built a terrain editor for StarLogo, and this paper will discuss its features and design choices, as well as the aspects of the program that can be improved.

Cellular Uptake and Trafficking of Dextran-Stabilized Colloidal Gold: Fluorescence and Transmission Electron Microscopy Studies

Maria Elena De Obaldia
Moorestown HS
Moorestown, NJ

Under the direction of
Dr. Mansoor Amiji
NEU Bouvé Coll of Health Sci

A diagnostic tool was developed to visualize cellular uptake and trafficking of colloidal gold nanoparticles. Colloidal gold was stabilized with dextran, labeled with pH-sensitive fluorescent probes, and visualized using fluorescent microscopy. Transmission electron microscopy (TEM) was used to observe the trafficking of gold nanoparticles. Internalization and trafficking through the endosomal pathways to the nucleus were found to occur within 0.5 hours. The rapid and efficient uptake of dextran-stabilized colloidal gold nanoparticles suggest that they could be used as drug or gene delivery vectors.

Towards DNA Recognition Properties of the Gyrase A C-Terminal Domain in *E. Coli* K12

Logan Dean
Quantico HS
Quantico, VA

Under the direction of
Ms. Daina Graybosch
Harvard

Currently a project is underway to determine the protein structure of the unique type II topoisomerase, gyrase. The protein is targeted by a new field of antibiotics and it is therefore pivotal to find the structure. The structure will be found by compiling the nuclear magnetic resonance reading of several fragments. It is necessary that these fragments have the least possible residues to increase the efficiency of the project. The GyrA C-terminus fragment has been sucessfully solved using X-ray crystallography at Verdine lab. The protein fragment gyrase A C-terminus has been successfully created with less cloned and expressed with less residues. The created fragment will allow researchers to fully understand the structure of the wrapping of DNA with gyrase, which occurs on the C-Terminus of gyrase A. It is the DNA-wrapping aspect of DNA that makes gyrase unique and offers possible molecule-binding sites for future antibiotics.

Zero-sum problems in finite groups

Vesselin Dimitrov
NHSMS
Sofia, Bulgaria

Under the direction of
Mr. Pavlo Pylyavskyy
MIT

We develop new methods for investigating problems of zero-sum type in general finite groups. We establish a new bound on Davenport's constant for abelian groups that asymptotically improves the previously known bounds. We use tools from representation theory to study properties of zero-sum sequences through nilpotent ideals of group algebras. A new relationship between zero-sum problems and multi-dimensional covers of $\mathbb{Z}$ is also established.

Calibration and Testing of Flight Data Recorders for Use in Unmanned Aerial Vehicles

Lyric Doshi
MLW Governor's School
Richmond, VA

Under the direction of
Col. Peter W. Young
MIT

The parameters that flight data recorders (FDRs) log during a flight are essential to recreating and analyzing the flight afterwards. FDRs are particularly useful on autonomous unmanned aerial vehicles (UAVs) because they provide a means for experimenters to learn what happened to and on the airplane throughout the flight. FDRs were calibrated and tested, and they produced accurate feedback. The airspeed tests highlighted how sensitive the airspeed sensor's accuracy is to the design of the pitot-static tubes. Data were collected to relate airspeed to engine rpm, and then engine rpm to thrust. Finally, the relationship between airspeed and thrust was determined for airplanes bearing various payloads that simulated equipment necessary for autonomous flight. It was found that heavy airplanes can maintain a higher airspeed than light airplanes at a given engine rpm. Tested and calibrated, the FDRs may be used to obtain comprehensive and accurate data in experiments.

Testing and Calibration of Flight Data Recording Systems on Unmanned Air Vehicles

Timothy Ebdon
6^{th} Form College, Farnborough
Farnborough, UK

Under the direction of
Col. Pete Young
MIT

Knowledge of flight parameters is vital for autonomous unmanned air vehicles (UAVs). This is a challenge as the UAVs diminish in size and power. It is still vital to know what is happening to the aircraft to learn about its performance and the performance of other internal systems. As the vehicle is unmanned, it is impossible to take direct measurements, and so a system that measures and records the data is required. This system needs to be of a known accuracy and precision. This paper describes the calibration, testing, and verification of a Real Design Ltd. flight data recorder (FDR). The FDR was then installed on a trainer model aircraft, flown, and the data analysed to learn more about the flight characteristics of the aircraft.

A Discrete Model For Option Pricing

Joumana El Jamal
Lycée Abdel Kader
Beirut, Lebanon

Under the direction of
Mr. Igor Pavlovsky
MIT

In this paper, we explain discrete tools of probability theory and apply them to answer some questions in mathematical finance. We discuss a modern approach to pricing of financial securities: construction of replicating portfolios. Using this technique, we derive a price formula for European options on binary trees. In addition, some known relationships between European and American options are also explained, such as the put-call parity satisfied by European put and call options: $c + Ke^{-rT} = p + S$, and the inequality $S - K < C - P < S - Ke^{-rT}$, satisfied by American put and call options.

A Network Creation Game Incorporating Traffic-Weighted Variables

Julie Finkelstein
Marlborough School
Los Angeles, CA

Under the direction of
Ms. Nicole Immorlica
MIT

The Internet's disorganization is a consequence of its rapid growth, with little opportunity for coordination to ensure efficiency. A ratio known as the *price of anarchy* compares the *social cost*, measured in hardware costs and quality of service, of a given network, such as the Internet, to that of an optimally designed network. To date, a constant lower bound has been established for the price of anarchy, while the upper bound, although postulated to be a constant, has not been proven. This paper presents a series of proofs that partially prove that the upper bound is a constant. Additionally, it explores a more realistic model of the Internet which takes into account the traffic within the network.

Carbon Nanotubes Grown on Sapphire Substrate

David Galkowski
Union Endicott HS
Endicott, NY

Under the direction of
Dr. M.S. Dresselhaus
MIT

The α constant of a substrate is used in the formula $\omega_{\mathrm{RBM}} = \alpha/d_t$ to determine the diameter, d_t, of a carbon nanotube based on the frequency of the radial breathing mode, ω_{RBM}, observed in the tube's Raman spectra. The α constant for sapphire substrate has been determined to be 253 $\mathrm{cm}^{-1}\,\mathrm{nm}$ by resonance Raman spectroscopy analysis on samples grown using the chemical vapor deposition method. Data were fit to the Kataura plot with precedence given to tubes with high-intensity resonance in the Raman spectra. The generated fit mapped the data onto the Kataura plot with an error of ± 4 $\mathrm{cm}^{-1}\mathrm{nm}$.

Retina Cognin Mediates Aggregation via Free Cell Surface Sulfhydryls

Shantanu K. Gaur
Bethel Park HS
Bethel Park, PA

Under the direction of
Dr. Robert Hausman
Boston University

Retina cognin (R-cognin) is a tissue-specific, 50-KDa protein that mediates the aggregation of retinal cells in chick embryos. R-cognin is related to protein disulfide isomerase (PDI) and is thought to regulate signalling between retinal cells via thioreductase activity. R-cognin's cell surface receptor has yet to be identified. To elucidate the structure of R-cognin's receptor and to define R-cognin's binding mechanism, the disulfide exchange inhibitor 5-5'-dithio-bis(2-nitrobenzoic acid) (DTNB) was used. It was determined here that the addition of DTNB to trypsinized cells blocked R-cognin binding, demonstrating that R-cognin requires free sulfhydryls on the cell surface to bind effectively. Through multiple assays, it was also determined that DTNB binds irreversibly to the cell surface and is able to block aggregation in an *in vitro* system.

Defining *fbp1*'s Control of Glucose Metabolism Through Characterization of the *git2* Domain in *Schizosaccharomyces pombe*

Delbert A. Green II
LSMSA
Opelousas, LA

Under the direction of
Dr. Charles Hoffman
Boston College

The gene *fbp1*, encoding fructose-1,6-bisphosphatase, is transcriptionally regulated by an overabundance of glucose. Ten *git* genes are known to be required for this repression. One, *git2*, encodes adenylate cyclase and, therefore, plays a crucial role in the cAMP pathway. This study sought to characterize this *git2* domain, specifically with respect to intragenic complementation within adenylate cyclase. It is demonstrated that overexpression of the wild type or mutant *git2* domain does not confer a *git* mutant phenotype in the presence of 5FOA, but does so in the absence of leucine. Furthermore, it is determined that certain *git2* alleles lower *fbp1* transcription. Discussed is preliminary research into the mechanisms by which *git2-7* alters transcription of *fbp1*. With further elucidation of the *git2* domain, substantial information can be gathered regarding adenylate cyclase function and cAMP signaling, and, therefore, applied to general cellular signaling and transduction.

On sliding straight polyominoes

Lewis Hahn
Hopkins School
Woodbridge, CT

Under the direction of
Ms. Fu Liu
MIT

In this paper, we attempt to analyze when it is possible to uncover certain squares on a straight polyomino-tiled grid by sliding straight polyominoes into squares that are uncovered. In particular, we were interested in when it is possible to uncover regularly spaced squares, and found a criterion for this property in certain cases.

Study of Different Catalyst Combinations for the Large Scale Synthesis of Carbon Nanotubes

Qilei Hang
Allegany HS
Cumberland, MD

Under the direction of
Dr. DeZhi Wang
Boston College

This paper found an optimum carbon nanotube growth catalyst ratio for chemical vapor deposition (CVD). Chemical vapor deposition is the process of growing carbon nanotubes by decomposing acetylene gas at a high temperature. The catalyst consists of three main solid components: $FeSO_4{\cdot}7H_2O$, $(NH_4)_6Mo_7O_{24}{\cdot}4H_2O$ and Al_2O_3. The optimum ratio of these components was found to be 1.0 : 0.10 : 1.37. This paper has also found that $NiNO_3{\cdot}6H_2O$ and $CoSO_4{\cdot}7H_2O$ when added to $FeSO_4{\cdot}7H_2O$ produce more carbon nanotubes than other catalyst combinations. A morphological analysis of the produced CNTs has been performed.

Electrostatics in the Adsorption and Desorption of Polyampholyte Chains

Jonathan Hanover
West Geauga HS
Chesterland, OH

Under the direction of
Dr. Udayan Mohanty
Boston College

In this paper, we develop a quantitative, theoretical model for the processes of adsorption and desorption of polyampholyte chains to a charged surface in the presence of multivalent ions. The electric field produced by the surface is described within the framework of the Gouy-Chapman approximation to the nonlinear Poisson-Boltzmann equation. Since divalent counterions near the charged surface create a strongly correlated two-dimensional liquid, we argue for the necessity of a renormalization of the Gouy-Chapman length, with which we calculate the effects of various concentrations of salt on the conformation of the polymer. The polyampholyte density profile for multilayer adsorption is determined for a theta solvent by balancing the monomer repulsion with polarization-induced attraction. At high salt we show that the thermal energy of the system equals the attractive energy of the chains to the surface, noting that this would lead to complete desorption of the chains.

Dynamics of Interfacial Ion Transfer Reactions

Michael Hell
Leibniz-Gymnasium
Altdorf, Germany

Under the direction of
Dr. Udayan Mohanty
Boston College

Many important chemical reactions involve ions diffusing across a liquid-liquid interface. To better understand these reactions, we derive an expression for the charge transfer coefficient α, evaluate the overall transfer rate k_r, and compare the results to experimental values. In the present paper a general mechanism of ion transfer reactions is proposed, based on a statistical-mechanical approach that accounts for the affect of capillary waves on the process of diffusion across the interfacial layer. From this mechanism an estimation of the transfer rate k_r is made, and we discuss an activation barrier for the transfer on the order of up to 11 kJ/mol. A possible combination of ion and electron transfer reactions is outlined, which would have a great impact on various fields in chemistry.

Ion Channel and Pump Function in Planarian Regeneration

Tim Hsiau
Hickman HS
Columbia, MO

Under the direction of
Dr. Michael Levin
The Forsyth Insitute

Regeneration is the process in which lost tissues are recovered through differentiation of stem cells kept in storage. A pharmacological screen in the organism planaria identified several channels as requisite for normal regeneration. Inhibition of *eag* potassium channels was shown to disrupt the initiation of regeneration, whereas inhibition of the H^+/K^+-ATPase pump and the minK voltage-dependent potassium channel caused eye defects in regenerates. Together, these results suggest an important role for potassium signalling and potassium-induced electrical potentials in regeneration.

On an extension of countable smooth dense homogeneity

Juliusz Jablecki Under the direction of
LO no. 3 im. Adama Mickiewicza Deniss Čebikins
Wroclaw, Poland MIT

The paper is devoted to the study of topological spaces for which there exist diffeomorphisms mapping a countable dense set onto a countable dense set. We prove that such diffeomorphism can be arbitrarily *close* to a given diffeomorphism, where closeness is measured by a given approximating function. We show that this result holds for compact manifolds, and we give an argument why it cannot hold for manifolds with boundary. We also extend the result for diffeomorphisms to flows induced by continuously differentiable vector fields.

An Improved Quantum Algorithm for Searching an Ordered List

Brian Jacokes Under the direction of
TJHSST Dr. Andrew Landahl
McLean, VA MIT

The best known quantum algorithm for searching an ordered list of N items is recursive and requires $3\log_{52} N \approx 0.526\log_2 N$ queries. We find an improved base case for this recursion, yielding an algorithm that requires only $4\log_{434} N \approx 0.457\log_2 N$ queries. This algorithm is more than twice as fast as the best possible classical algorithm, which requires $\log_2 N$ queries.

Minimal Surfaces

Risa Kawai Under the direction of
Alabama School of Math & Sci Mr. Mike Hill
Birmingham, AL MIT

Soap bubbles form shapes in a distinct way because they use the least amount of energy as possible. These shapes take the form of minimal surfaces. This paper focuses on minimal surfaces that have ellipses as their cross sections in the xy-plane. The minimal surfaces vary depending on the initial condition of the ellipses.

Restitution of epithelial cells

Farah Khachab Under the direction of
Rafic Hariri HS Dr. Susan J. Hagen
Tyre, Lebanon BIDMC

This study was conducted to study restitution after injury in the gastric mucosa and to determine the role of monocarboxylate transporter (MCT1). For this purpose, whole bullfrog stomach was mounted in Ussing chambers. Injury was done with NaCl and restitution was evaluated in the presence of CHC, an antagonist of MCT1. Restitution was evaluated by recovery of transepithelial cells resistance (TER). Restitution was significantly inhibited by CHC, which suggests that MCT1 is essential for restitution.

Full-Featured Virtual World Games Framework using RSIp, a Novel Command Protocol

Todor Kolev Under the direction of
NHSMS Prof. Eric Klopfer
Sofia, Bulgaria MIT

This paper introduces a new high-level object-oriented framework providing all the necessary capabilities for creating virtual-world games.

The existing tools for creating games in virtual-world environments provide only the basic low-level API functions for interacting with the virtual world. The advantages of this newly developed framework is that it offers the full features necessary for a virtual-world game implementation, rather than just the tools for creating the basic 3D interaction.

A novel remote procedure calling protocol called RSIp (Rapid Secure Invoke Protocol) was created to enable the realization of the framework. RSIp is specially designed to optimize the speed of the web-to-process RPC communication. The protocol provides extended security and session-support facilities which make it especially efficient when the web system's sole purpose is to provide interface to another application.

Based on this framework, a virtual-world game called *Environmental Detectives* was developed to be used for education research.

Semantic Analysis of Song Lyrics

Andrew Kositsky
Ferndale HS; WWU
Lummi Island, WA

Under the direction of
Dr. Beth Logan
Hewlett-Packard Labs

This paper proposes a solution to the problem of automatically indexing vast quantities of music by implementing a novel approach of using probabilistic latent semantic analysis on song lyrics. Hofmann's is the basis for our analysis of text-based documents. The results of our experiment were compared against ground truth provided by an online survey looking at different artists and the acoustical model of determining song similarity of Logan and Salomon.

Stochastic Progress-Based Soft Real-Time Scheduling

Nathan La Porte
Georgetown Day School
Washington, DC

Under the direction of
Dr. Martin Rinard
MIT LCS

In this paper, we present a soft real-time adaptation of the earliest deadline first (EDF) scheduling algorithm, called the ProgressScheduler, that guarantees consistent latency for microprocessor threads by taking into account their execution times and progress. This algorithm was designed for automatic target recognition (ATR), with the goal of making the output more consistent. It avoids EDF's overload failure by adjusting deadlines dynamically, and provides consistent latency by giving more CPU to slow-performing threads, and less CPU to faster-performing threads. This algorithm is appropriate for any application which uses a number of identical threads to process data in a pipeline.

Body Sway Technology: Studies on Data Correlations to Identify Elderly People Prone to Falling

Serene Lee Min Wai
Raffles JC
Singapore

Under the direction of
Prof. David Garelick
Northeastern University

To identify elderly subjects prone to falling, a sound wave assessment device has been developed to measure human body sway. In this study, the correlation between the root mean squared velocity (Vrms) of human subjects and their likelihood of falling was investigated. Vrms values were found to be inconclusive in differentiating between 'fallers' and 'nonfallers'. Swaying motion in both the anteroposterior and mediolateral directions was investigated, and the average ratio of the Vrms values in both directions was found to be nearly constant. It was also concluded that age does not affect the chances of falling. In this study, a novel equation similar to that of acceleration was introduced. By comparing the maximum 'acceleration' values in the human sujects tested, a threshold value was determined to distinguish 'fallers' from 'nonfallers' more effectively. Another method involving moments was also investigated. This was found to be fairly effective in accounting for a higher fall probability in 'fallers', though not as effective as the sonar method.

Discrimination of Spatial Relations and Contour Differences in Faces: Effects of Inversion

Connie Leung
Palos Verdes Peninsula HS
Palos Verdes Estates, CA

Under the direction of
Dr. Jason Barton
BIDMC

Discrimination of changes in eye position, eye size, mouth position, mouth size, forehead size, and chin shape were studied to determine the effects of inversion on these areas. Face inversion greatly impaired discrimination of mouth position and, to a lesser degree, mouth size. Eye shape and size, in addition to contour changes, showed fewer recognition penalties for inversion. These results suggest that the dominant effect of face inversion is decreased discrimination in less salient facial regions. Although the chin and mouth are both located in the lower half of the face, discrimination of chin shape suffered little when faces were inverted. Thus, facial contours play a less vital role in perception, implying that people rely more heavily on facial features.

Approximating the Composite Light Curves Due to the Lensing of Large Source Fields

Jason Li
Great Neck North HS
Great Neck, NY

Under the direction of
Dr. Rosanne Di Stefano
Harvard-Smithsonian Ctr

Gravitational microlensing occurs when light from a

distant source is deflected by the gravitational pull of an intervening mass. The object that deflects the source light is known as a lens. Microlensing phenomena that involve single or binary lenses and sources are quite well understood, but the microlensing of large source fields has received only limited attention. The purpose of this research is to determine the nature of the light curves associated with the microlensing of large source fields. The results of this research show that a combination of (1) monitoring only the few sources closest to the lens in the plane of the sky at each time step and (2) considering density fluctuations can very accurately predict the composite light curve due to all the sources within a specified region. This is especially true at higher source densities because of the lower fluctuations in the baseline luminosity at these higher densities. A reduced chi-squared goodness-of-fit test was used to verify the extent to which approximated light curves matched up with composite light curves.

Thermoelectric Auxiliary Power Supply for Gas-cooled Fast Reactor Decay Heat Removal

Liu Hongwan
Raffles JC
Singapore

Under the direction of
Prof. Michael Driscoll
MIT

Decay heat removal has long been a problem in nuclear power plants. Modern power plants use combustion engines to move coolant through the core to remove the heat. In this paper, a more reliable thermoelectric generator is successfully designed to replace the combustion engine in a power failure or scram situation in the power plant, and the dimensions of the generator are found to fit compactly within an annulus between the inner and outer shell.

An Implementation of the Karp-Zhang Parallel Branch-and-Bound Algorithm

Tiankai Liu
Phillips Exeter
Saratoga, CA

Under the direction of
Prof. Charles E. Leiserson
MIT

This paper studies an implementation of the Karp-Zhang parallel branch-and-bound algorithm on a shared memory machine. By employing it to solve a solitaire card puzzle, empirical data on the speedup of the algorithm is (going to be) obtained.

Interaction with Public Displays through Bluetooth-enabled Cellular phones

Jesse Louis-Rosenberg
Saugerties HS
Saugerties, NY

Under the direction of
Prof. Larry Rudolph
MIT CS & AI Labs

With computing spreading more to handhelds and cellular phones, it becomes of interest to create pervasive computing environments. This paper explores the implementation and design of public display/cellular phone interfaces. The connection protocol and implementation were outlined. The security strengths and weaknesses of such a system were analyzed. It also looks at ways of improving the user interface, focusing on automatic connections based on proximity. A specification was designed for interacting with an advertising display. This paper also looks at the obstacles encountered in implementing a client application for the cellular phone.

Hypersurfaces in Spheres and de Sitter Space

Amos Lubin
Hanover HS
Norwich, VT

Under the direction of
Mr. Huadong Pang
MIT

For complete hypersurfaces of S^{n+1}, we provide an inequality relating the length of the Ricci curvature tensor and the scalar curvature and show that equality holds if and only if the hypersurfaces are totally geodesic. We provide an analogous inequality for complete timelike hypersurfaces of M_1^{n+1}, but we show that equality can hold only if $n = 2$.

Evidence for the Mechanism of the DNA-Induced Melanogenic Responce

Sarah Marks
Pace Academy
Decatur, GA

Under the direction of
Hee-Young Park, Ph.D.
Boston University School of Medicine

The melanogenic response is a first-line defense for cells against dangerous UV radiation that may cause both DNA and cellular damage. DNA photo damage appears to induce a melanogenic response by upregulation of tyrosinase, the rate-limiting enzyme in melanin biosynthesis. However, the mechanism of upregulation is unclear. To determine by which pathway DNA-damage-induced upregulation of ty-

rosinase proceeds, levels of tyrosinase, PKC-β, and MITF were measured in using Western-blot analysis in response to T-oligo, a DNA-damaging mimic that is homologous to the 3' overhang of telomeres. Tyrosinase and MITF were both found to be upregulated in a time-dependent manner and PKC-β was found to be suppressed in a time-dependent manner. To elucidate the mechanism of PKC-β regulation, it was hypothesized that the p53 protein tumor response system might regulate its expression; a CAT assay was performed to test whether PKC-β mRNA is upregulated. Our results suggest that p53 may actually suppress expression of PKC-β, explaining its suppression by T-oligo. This study suggests a possible mechanism for understanding DNA-induced melanogenesis which may have practical applications in treating pigmentation disorders and skin cancer.

In vitro Z-DNA binding conditions of ZE3L, a protein found in *vaccinia* virus

Kacey Meaker
Humble HS
Humble, TX

Under the direction of
Dr. Ky Lowenhaupt
MIT

Z_{E3L} is a protein present in *vaccinia* virus and other pox viruses. This protein uses Z-DNA binding to counteract the host's interferon effects. The protein readily binds *in vivo* but not *in vitro*. Cobalt hexaamine can allow Z_{E3L} to bind to Z-DNA *in vitro*, but cobalt hexaamine is not present in the cell. Other substances, including: spermine, spermidine, Denhardt's solution, and polyethylene glycol do not affect the way in which Z_{E3L} binds to Z-DNA. More testing is needed to discover other chemicals present in cells that will allow Z_{E3L} to bind *in vitro*. This experimentation is important to gaining knowledge about this protein so that eventually an inhibitor therapy can be developed to treat *vaccinia* virus and its close relative, smallpox.

A Computer Interface for Group Decision Making

Patrick Mihelich
Park Tudor HS
Carmel, IN

Under the direction of
Mr. Walter Bender
MIT Media Lab

Traditional group decision making has well-documented problems relating to inefficiency of communication and excessive emphasis on information which is commonly known to the whole group. Computer interfaces for use in face-to-face group meetings are implemented and tested to determine which most effectively address these problems. This paper introduces a novel layout for the shared display of an interface and compares it to a more traditional chat interface. Preliminary data suggest that the greater support of the new layout for organization of information is beneficial to the group decision making process.

Improving the Efficiency of the Heat Exchange in a Nuclear Reactor

Youssef Moussaoui
Casablanca American School
Casablanca, Morocco

Under the direction of
Prof. Mike Driscoll
MIT

As nuclear energy becomes a more important source of power, increasing its thermodynamic efficiency becomes necessary. Previously, however, it has proved challenging to derive a useful formula for efficiency in nuclear reactors. In this paper, an analytical expression is derived for $\frac{W}{Q}$. It is then tested using a computer simulation program for nuclear reactors. It was found that the data from the computer simulation corresponds almost perfectly with the derivation, proving its accuracy.

Covering Systems of Integers

Jennifer Nan
Norman North HS
Norman, OK

Under the direction of
Mr. Michael Ching
MIT

We define a composition operation on systems of congruence classes that cover the integers and find algorithms for determining whether such a system is "prime" with respect to this operation. We begin to investigate how these notions could help to find a covering system composed of only distinct, odd moduli.

Microlensing of Thousands of Light Sources

Abdel-Azim Osman
Cairo American College
El Haram, Cairo, Egypt

Under the direction of
Dr. Rosanne Di Stefano
Harvard-Smithsonian Ctr

Microlensing of thousands of background sources of light alters the observed light curve. The light curve of sources from a microlensing event and the light curve from the same sources without a microlensing event are compared, showing that the microlensing event light curve is smoother and has more variation than the light curve without a microlensing event. Different densities of light sources reveal that the baseline luminosity is proportional to the density of the background light sources. Also, baseline luminosity with different Einstein ring radii are compared to the baseline luminosity of the light sources, showing that the baseline luminosity is proportional to the square of the Einstein ring radius. The greater the speed of the lens, the more frequently maximums and minimums will occur. An alteration of the speed, density, and Einstein ring radius will have the result of each alteration's independent effect on the curve. The proportionalities are the same whether there is a microlensing event or not.

Development of Robotic Foosball as a Versatile Platform for Robotics Research and Contests

Oded Padon
Handesa'im Sch, Tel Aviv U
Ganey Tikva, Israel

Under the direction of
Mr. Oren Zuckerman
MIT Media Lab

This paper considers the development of a *programmable robotic foosball table* as a versatile platform for robotics research. A *programmable robotic foosball table* is a foosball table with a computer-controlled mechanical system to control the rods, and a means of sensing the position of the ball and the rods. We discuss the design of the system that we have created, and its various implementations. In particular, we suggest its use in an international robotic-programming contest.

Gamma-Ray Burst Detection Through Orbit Subtraction

Sara Peek
MSMS
Ocean Springs, MS

Under the direction of
Dr. George Ricker
Center for Space Research

Despite the quality of instrumentation on the gamma-ray burst detection satellite HETE and on the ground, not every gamma-ray burst is detected, and many random events are shown as false positives. Particles in space and in the earth's atmosphere, as well as known sources of gamma and X-rays, produce substantial background noise. This triggers many false readings and makes true bursts difficult to detect. In this paper a method is proposed to subtract background noise from HETE data to aid in burst detection and reduce the number of false positives triggered. Several variations are presented on the subtraction of previous similar orbits from a test orbit. As with all current triggering algorithms, orbit subtraction is more sensitive for shorter gamma-ray events. However, the method also shows promise in the detection of the more elusive "very long" GRBs. This attribute is much more valuable, as there is currently no sensitive algorithm for detecting bursts on timescales of 400–500 seconds and longer.

Chandra Observations of the Orion Trapezium

Vishal Prabhu
D.G. Ruparel College
Mumbai, India

Under the direction of
Dr. Norbert Schulz
MIT Center for Space Research

This paper uses the *Chandra* observations of the Orion Trapezium, taken over a span of two years, to study the X-ray characteristics of the the cluster. We detect six proplyds (young stars with protoplanetary disks) in the field, obtain their spectra, and compare them with those of the other late-type stars. We estimate the temperatures and column densities for the detected flares and proplyds and confirm that the X-ray emissions are related to the central stars rather than the protoplanetary disks.

Facilitating Parametric Analysis of Articular Surface Formation in Chicken Embryos through μMRI Compatible Programmable Mechanical Actuation of Limb Motion

David A. Puckett
Helena HS
Clancy, MT

Under the direction of
Brian Snyder, M.D., Ph.D.
BIDMC

Recent research efforts into the process of articular formation in embryonic chicken hind limbs have been hampered by a difficulty in performing a parametric analysis of the factors effecting this process. Specifically, the previous apparatuses used in this research required custom-made mechanical parts and circuitry developed specifically for the parameter that one wished to test. Also, the apparatuses used could actuate relatively few limbs at one time, leading to the requirement of multiple repeated trials in order to gain statistically relevant data and a corresponding reduction in time-usage efficiency. Finally, in order to more accurately model the process of joint formation, images capturing the developing joint at various points in its development had to be taken via μMRI. However, due to ferrous metals present in the device used to actuate the limb, the requirement for imaging meant that the limb had to be killed and preserved before any imaging could take place. Thus, the purpose of this study was to design and develop a device that facilitated the μMRI imaging of chicken embryo hind limbs and allowed for increased ease of parametric analysis of the data gained through this imaging by remedying the drawbacks of the previous model through the removal of ferrous materials, the replacement of the hardwired controller design with a programmable controller design, and increasing the test capacity of the actuator.

On the Effectiveness of Dark Energy and General Relativity in Modeling the Universe

Drew Reese
Oldham County HS
Crestwood, KY

Under the direction of
Prof. Edmund Bertschinger
MIT

The cause of the accelerating expansion of the universe has drawn much speculation in recent years. Dark energy is the most popular explanation for this phenomenon. Using data from Type Ia supernovae, models of a universe with and without dark energy are contrasted to determine which explanation is more accurate. While universes with dark energy model the data more closely than ones without, the possibility for viable alternatives through modifications to general relativity remains strong.

Optimization of Bayesian Inference Routines for the Analysis of X-ray Flares from Sagittarius A*

Alec Resnick
FW Springstead HS
Brooksville, FL

Under the direction of
Dr. Frederick K. Baganoff
MIT Center for Space Research

Difficulties in offboard statistical analysis of raw spectra and event data from the *Chandra* X-ray Observatory (CXO) are discussed. Reliability gaps in this data are rectified at the analytic-algorithm level through the incorporation of good time intervals (GTIs) and compensation for dead time (DT) into a Bayesian-blocks analysis script. A one-sided and two-sided Kolmogorov-Smirnov test are implemented in S-lang. The established IDL script for the Kolmogorov-Smirnov test is revealed as flawed and its invalidation is documented. A cumulative Poisson probability function S-lang subroutine is defined and a discussion of the inherent caveats of the function's application to time-tagged data ensues.

The implementation of GTI-correction in the Bayesian Blocks scripting is evaluated in context of a test case utilizing recent *Chandra* data from the nonstellar, Sgr A* radio source believed to be coincident with a supermassive black hole at the galactic center. The modified Bayesian-blocks analysis is applied to the *Chandra* data. Preliminary conclusions regarding the algorithm's effectiveness and the need for the generalization of a GTI implementation script is discussed.

The degree distributions and generation of general random graphs

Ho Seung Paul Ryu
Pacific Grove HS
Pacific Grove, CA

Under the direction of
Dr. Muriel Médard
MIT

This project investigates fiber-optic backbones to find the degree distribution of weighted random graphs. The power law has become the new standard in massive random network research, and this project confirms that weighted fiber-optic backbones appear

to follow the power-law distribution to a large degree. When weight is removed, the degrees of the networks seem to follow a distribution similar to the Poisson distribution. Discussion on the degree distribution and generation of random networks is given, with further possibilities for research.

Correcting Automatic Speech Recognition Errors Using Statistical Analysis of Word Co-occurrences

Arup Sarma
Shawnee Mission East
Overland Park, KS

Under the direction of
Dr. David Palmer
Virage Adv Tech Group

Automatic Speech Recognition (ASR) systems, which seek to automatically transcribe spoken audio, frequently make transcription errors that are difficult to detect. In this paper we describe a novel algorithm that detects and corrects ASR output errors using statistical analysis of word co-occurrence frequencies from training data. The probabilities for candidate ASR errors and their replacement words are determined from these frequencies. If a candidate error is phonetically similar enough to a given keyword and is less likely to occur in a given context, then it is a likely ASR error that can be replaced with the more likely keyword. The results of our experiments indicate that it is possible to use our method to correct a large number of the ASR errors in a document with few or no false positives.

An Investigation of Irreducible Polynomials in $\mathbb{Z}_p$ Using Galois Theory

Brianna Satinoff
Palm Harbor Univ HS
Palm Harbor, FL

Under the direction of
Mr. Michael Hill
MIT

Traditional notions of factorization are altered in finite fields such as $\mathbb{Z}_p$ for primes p. In this paper we examine the implications of Kronecker's Theorem on finite fields. This theorem states that, for any irreducible polynomial over a field F, there exists a field extension of F in which the polynomial factors completely.

We first use Kronecker's Theorem to determine the number of irreducible polynomials of degree n in $\mathbb{Z}_p$. After presenting a purely numerical approach, we develop a formula based on different field extensions of $\mathbb{Z}_p$. We later discuss the set of polynomials $x^{p^n} - x$ as a partially ordered set.

In addition, we group the irreducible polynomials in $\mathbb{Z}_p$ into polynomials for which $p(x) = p(x+1)$, and others. After establishing properties of polynomials for which $p(x) = p(x+1)$, we suggest a way to determine how many irreducible polynomials of degree n in $\mathbb{Z}_p$ have this property.

Evaluation of High-Efficiency Production of Hydrogen Using the Supercritical CO_2 Power Cycle and High-Temperature Steam Electrolysis

Mudassar Shahid
Salamat ICAS
Lahore, Pakistan

Under the direction of
Prof. Michael J. Driscoll
MIT

Hydrogen has the potential to replace fossil fuels as the main source of energy in the transportation sector, thereby reducing emissions and ending our dependence on petroleum. Nuclear energy is a promising candidate for addressing the expected increase in the worldwide demand for hydrogen. The purpose of this paper is to identify the most attractive, efficient, and economical way of using nuclear energy for hydrogen production. The major technology scenario which is assessed in our work is the high-temperature steam-electrolysis (HTES) approach coupled to advanced reactor concepts. Two different paths are possible using this technology. In this work we assess which path is feasible from both economic and efficiency points of view.

Determining the Accuracy of Jet-Detection Algorithms for Relativistic Heavy-Ion Collisions

John Shen
Montgomery Blair HS
Gaithersburg, MD

Under the direction of
Dr. Maarten Ballintijn
MIT

We tested jet-finding algorithms using tracker data from simulated events from the compact muon solenoid (CMS). Here, we adapt jet finding algorithms from calorimeter-based algorithms, the UA1 cone, UA2 window, CMS cone, and CMS window algorithms, and apply them to simulated heavy-ion collision data. We determine the accuracy and spatial resolution of the CMS window jet finder by examin-

ing the distances from detected jets to real jets. At high energies, we independently verify the accuracy by monitoring the angular separation between dijets. The energy resolution of these algorithms is found to depend on the aperture or clustering size of the algorithm. Over all jets, about 30% of all jets are correctly identified, but the CMS window algorithm correctly identifies 70% of jets with energies above 50 GeV.

The Observability of Gravitational Radiation from Binary X-ray Pulsars

Aditi Shrikumar
Int'l School, Bangalore
Bangalore, Karnataka, India

Under the direction of
Mr. Craig B. Watkins
MIT

In this paper, we assess the detectability of gravitational radiation produced by the four known binary X-ray pulsar systems. Observations indicate that some countertorque prevents these pulsars from spinning faster due to accretion. We attribute this countertorque to the emission of gravitational radiation and thereby derive an expression for gravitational wave strain in terms of the rate of mass accretion, distance from earth, and spin frequency. We find it unlikely that present technology can detect gravitational radiation from these sources.

Effect of Threat on Efficiency and Detection

Ansam Sinjab
Hariri HS II
Beirut, Lebanon

Under the direction of
Dr. Jeremy Wolfe
BWH; Harvard Med Sch

People look at things and classify them as attractive, scary, or disturbing. One can infer that things that affect our emotions are the things that capture our visual attention. Such perceptual factors play an important role in triggering our attention and thus directing our reactions and physical or mental responses to the things we see. The aim of this research is to examine a relation between threat, the resulting fear, and visual attention.

Analysis of the Negative Magnetoresistance Effect in $Co/SrTiO_3/Ni_{80}Fe_{20}$ Magnetic Tunnel Junctions

Robert Sinnott
Deerfield Academy
Barrington, IL

Under the direction of
Dr. Jagadeesh Moodera
Francis Bitter Magnet Laboratory

In May of 1999, a negative junction magnetoresistance was found to occur in cobalt–strontium titanate–lanthanum strontium manganese oxide (LSMO) junctions. Here, we measure the magnetic-field dependence of tunneling between cobalt and Permalloy across the dielectric insulating barrier of strontium titanate and compare our measurements to previous results from cobalt–strontium titanate–LSMO junctions. By using a high-vacuum evaporative-deposition technique, we constructed junctions that demonstrated the negative magnetoresistance. The study of this effect and its dependence on temperature and the voltage drop across the junction suggests possible magnon interactions as the cause of this negative magnetoresistance effect.

Formation of Protective Groups For Alcohols

Rola Sleiman
Hariri HS II
Lebanon, Beirut

Under the direction of
Dr. Mohammed Movassaghi
MIT

Metal vinylidene is a reactive compound used in the formation of carbon-carbon bonds. These bonds are very important for organic synthesis, in which very complex compounds are to be formed. To form this compound, 5-hexyn-1-ol is the starting material. It has two functional groups, the alcohol (OH), and the alkyne. The alcohol reacts and transforms under the same conditions through which the reaction would take place, which is unwanted. Thus the alcohol in 5-hexyn-1-ol should be blocked or protected by changing into another form, that is O-TBS-5-hexyn-1-ol. This study is about how to protect the alcohol in 5-hexyn-1-ol, or in other words, how to form O-TBS-5-hexyn-1-ol, silyl ether.

Is shininess a basic feature of visual search?

Matthew Sng Under the direction of
Raffles JC Dr. Jeremy M. Wolfe
Singapore Brigham and Women's Hospital

The human eye is capable of distinguishing shiny objects from matte objects easily. It is still unclear how exactly the visual system is able to differentiate between a reflection and a mere surface feature on a subject. In addition, it is able to identify the material of the surface with considerable accuracy. This ability to detect shininess is not dependent on the context in which the shiny object is placed. Given this ability, can humans search efficiently for a shiny object amongst matte objects? Do shiny objects pop out of a visual scene? This paper investigates searches for glossy objects that are displayed in 2D. The results indicate that search is inefficient. However, we believe that this arose due to flaws in experimental design and that the results for a 2D stimulus are not representative of searches in the real, 3-dimensional world.

Determination of Clonality in Human Meninges

Marta Swierczynska Under the direction of
University of Warsaw Mia MacCollin, M.D.
Szczecin, Poland Harvard Medical School

Most models of tumorigenesis assume that tumors are monoclonal in origin. This conclusion is based largely on studies using X-chromosome–linked markers in females. One important factor, often ignored in such studies, is a distribution of X-inactivated cells in tissues. Recent research has shown that in normal females, cells with the same X chromosome inactivated are grouped together to form patches. As the patches turned out to be relatively large, they have the potential to confound the assessment of tumor clonality. The presence of large patches reduces the probability that tumors will be located at patch borders, and thus X-inactivation studies are heavily biased toward showing that tumors are clonal. Thus, it is important to know the patch size in different human tissues.

In this project we tried to estimate the patch size in human meninges by using the PCR-based assay called HUMARA. We present the optimization of this assay for use in human meninges. Establishing the patch size in this tissue will enable us to understand better the brain tumor development, thus allowing for more effective treatment.

Using Wireless Sensor Networks to Monitor the Physical World

Shane Treadway Under the direction of
Kubasaki Prof. Ruaidhri M. O'Connor
Okinawa, Japan MIT

Wireless sensor networks (WSN) are an emerging technology to help monitor large physical systems. WSN's could be used to measure the sway of a large building or keep track of the condition of a bridge exposed to the weather. Small wireless sensors could be embedded in the system, monitoring the system from the inside. For instance, monitoring the integrity of a bridge under traffic loading would be an example of this method.

This technology combines networked radio communications, microprocessors, various sensors, and smart materials. New advances in optical communications allow transmitters to be smaller and may eventually render radio communications obsolete.

In this project wireless remote sensors were tested on a cantilever-beam system. The acceleration of the cantilever tip was measured over time and the average damping coefficient of the motion was calculated.

Next, the wireless sensors were used to measure the damping effect of piezoceramic on a cantilever system. The piezoceramics were connected in a simple resistor circuit and used to strain out of phase with the cantilever in an attempt to further damp the oscillatory motion.

The Role of Cdc42 and its Downstream Effectors in Mouse Submandibular Gland Development

Peter Wang Under the direction of
IMSA Maria A. Kukuruzinska, Ph.D.
Naperville, IL Boston University

Traditionally, integrins and cadherins have been seen as instrumental in the cell's ability to migrate and to establish polarity and stable cell-cell contacts during development. However, they have recently been shown to regulate a variety of intracellular signalling cascades through a family of small Rho GTPases,

Cdc42, RhoA, and Rac. These adhesion-based receptors have been implicated in various cellular processes including proliferation, differentiation, and actin cytoskeletal organization. In our studies $\alpha 3\beta 1$ integrin-deficient and wildtype mouse submandibular gland (SMG) were used as model systems for studying the downstream effectors of Cdc42. In particular, we focused on WASP, which is central in actin cytoskeletal organization in many cell types. However, our preliminary data suggest that cytodifferentation in SMG may be mediated by a mechanism other than WASP.

Statistical Mechanics of DNA Bending

Shuyu Wang Under the direction of
Okemos HS Dr. Udayan Mohanty
East Lansing, MI Boston College

We propose a statistical mechanical model for the electrostatic contribution to the persistence length of B-DNA in a multivalent ionic solution. In this model, the phosphate charges along the double helix are projected onto the symmetry axis of the DNA. The charges interact via a screened Debye-Hückel potential. The Gibbs free energy for assembling the charges from an infinite distance is determined by a technique due to Manning. Each phosphate charge is renormalized by the linear charge density parameter to account for the phenomenon of counterion condensation. By comparing the difference in the Gibbs free energies of a weakly bent and straight DNA, the electrostatic contribution to the persistence length is determined for several DNA conformations. Finally, we calculate the end-to-end distribution of a weakly bent DNA by exploiting a formalism due to Zandi et al. The end-to-end distribution allows us to address two fundamental and pressing issues regarding DNA structure. We determine for the first time not only the contribution of the electrostatic part to the total persistence length, but also the fraction of phosphate charge that has to be neutralized in order that the end-to-end distribution of a polyelectrolyte DNA may resemble that of a neutral DNA. These predictions pave the way toward the design of new experiments that can probe the stiffness of neutral and polyelectrolyte DNA *in vivo* and *in vitro*.

On the k-ordered Hamiltonicity of n-dimensional Hypercubes

Jeremy Warshauer Under the direction of
San Marcos HS Deniss Čebikins
San Marcos, TX MIT

A graph is *k-ordered* if, for any sequence of k vertices, there is a cycle containing these vertices in the given order. If the cycle is also a hamiltonian cycle, then G is said to be *k-ordered hamiltonian*. Let $\sigma(G)$ be the maximum integer k such that G is k-ordered hamiltonian. The n-dimensional hypercube, denoted Q_n, has vertex set consisting of all binary strings of length n, and two vertices are connected by an edge if they differ in exactly one coordinate. It is shown that $\lfloor \frac{2n-1}{3} \rfloor \leq \sigma(Q_n)$, for $3 \leq n$. A computational method was developed to efficiently generate all sequences of k vertices of Q_n and determine whether there is a hamiltonian cycle traversing each of these sequences. This data then allowed an exact value of $\sigma(Q_n)$ to be determined, for $n = 2, 3$, and 4.

Neural Basis of Inhibition: A Study of Antisaccades Using fMRI and MEG

Bobby Xu Under the direction of
IMSA Dr. Dara Manoach
Naperville, IL MGH

Inhibition is a cognitive ability that allows humans to respond flexibly rather than reflexively to events. The purpose of this investigation is to elucidate the neural basis of inhibition in healthy humans. Combining magnetoencephalography and functional MRI, two unique brain-imaging instruments, allows both key brain regions and the timing of neuronal processes involved with inhibition to be resolved. The study found strong evidence that the dorsolateral prefrontal cortex is an important component in programming inhibitory signals.

Molecular characterization of the *acerebellar* mutant in *Brachydanio rerio* through whole-mount *in situ* hybridization

Jessica Yang
Iolani School
Mililani, HI

Under the direction of
Dr. Craig Albertson
Forsyth Dental Institute

The *acerebellar* mutant in *Brachydanio rerio* lacks the midbrain-hindbrain boundary and displays defects in the craniofacial skeleton, the most notable of which is asymmetry. Since most craniofacial structures arise from neural crest cells, the observed phenotypic defects are hypothesized to result from abnormalities in the cranial neural crest. To test this hypothesis, whole-mount *in situ* hybridization was performed utilizing various probes that marked stages of cranial neural crest induction, migration, and differentiation. Results suggest that directional asymmetry is statistically significant and that abnormalities in migration are the source of that asymmetry. The molecular characterization of the *acerebellar* mutant may facilitate future advancements in detailing the pathways of craniofacial development, and hence a better understanding of craniofacial disorders.

The Effect of Gas Pressure on Carbon Nanotube Synthesis by Thermal Chemical Vapor Deposition

Yen T. Yeh
Westview HS
Avondale, AZ

Under the direction of
Prof. Zhifeng Ren
Boston College

The effect of gas pressure on the structure of carbon nanotubes (CNTs) has been systematically investigated in the chemical vapor deposition process. The yield of CNTs (defined as the weight ratio of CNTs vs. catalyst) decreases significantly with the gas pressure, going from 547.5% at 5 torr to 157.9% at 200 torr. At low reacting gas pressure the CNTs have completely hollow cores, whereas at high pressure the CNTs have a bamboo structure. The results show that the structure and yield of carbon nanotubes are strongly affected by the growth gas pressure.

Quantitative Trait Loci Modulating Corpus-Callosum Size in the Mouse Brain

Ning Zhou
Wayzata HS
Plymouth, MN

Under the direction of
Dr. Glenn D. Rosen
BIDMC/Harvard Medical School

The corpus callosum is the brain's primary pathway for communication between the left and right hemispheres, and abnormal function and anatomy of the corpus callosum have been associated with a variety of disorders in humans and animals. In this study, we mapped quantitative trait loci (QTLs) modulating corpus-callosum size in mice. By using mouse-brain sections and stereological techniques, we estimated the midsagittal corpus-callosum size in 191 BXD recombinant inbred mice and their parent strains (C57BL/6J and DBA/2J). Analysis showed midsagittal corpus-callosum size to be a heritable trait ($h^2 = .42$). We detected a suggestive QTL modulating corpus-callosum size near the marker S10Gnf071.990 on chromosome 10 at approximately 73 Mb, in close proximity to a previously defined QTL modulating striatal volume. We also found correlations between corpus-callosum size and other behavioral and anatomical phenotypes. Overall, our results implicate a genetic role in modulation of corpus-callosum size.