Research Statement: Steven J Miller

sjm1@williams.edu

Research and Applied Interests: Distribution of zeros and n-level statistics for families of L-functions, especially families of elliptic curves with rank over $\mathbb{Q}(T)$, Random Matrix Theory, Random Graphs, Elliptic Curves, Additive, Analytic, Combinatorial and Computational Number Theory, Probability Theory and Statistics, Benford's Law, Cryptography, Sabermetrics, Linear Programming and Operations Research.

I earned my PhD in 2002 and have written over 100 papers in the areas above. After a high level overview of my work, I include the project description from my current NSF grant; this does an excellent job summarizing my recent work and problems I am currently studying. After that is a brief summary of my thesis and a representative sample of earlier results.

I have worked on many of these projects with students in high school (20), college (over 270) and graduate school or postdocs (over 10). In many instances I ran guided vertically research groups, where the graduate students had their own projects and also helped mentor the undergraduates and high school students. I have been able to maintain my active research program while at a liberal arts college which prides itself on its commitment to teaching.

My research has been continuously supported by individual NSF grants since 2006.
Papers and talks are available at http://www.williams.edu/Mathematics/sjmiller/.

Contents

1 Research Interests: Summary

While my primary training is in analysis and number theory (and I do love working on problems in these fields), I have collaborated with numerous people outside of pure mathematics, writing papers in Accounting, Computer Science, Economics, Engineering, Geology, Marketing, Physics, Sabermetrics and Statistics. I enjoy these projects for many reasons. Frequently these are interesting problems which lead to interesting theoretical issues. Further, I am often able to use these papers in classes I teach, and I find these projects are a great way to interest and excite students.

My main area of research is number theory. Since Riemann's investigations 150 years ago, zeros of L-functions have been known to be intimately connected to solutions to many problems in number theory. In the last few decades finer properties of the zeros have helped understand problems such as the observed preponderance of primes congruent to 3 mod 4 over 1 mod 4 as well as the growth of the class number. Random Matrix Theory has become a powerful tool to model the behavior of these zeros, suggesting both the answers as well as new questions to ask. My primary interest is in the distribution of zeros near the central point for families of L-functions, especially families of elliptic curves with rank over $\mathbb{Q}(T)$. Related to this, I am also investigating constructions for moderate to high rank one-parameter families of elliptic curves, lower order corrections to n-level densities of elliptic curves (where the arithmetic lives), and the influence of forced zeros at the central point on the distribution of the first zero above the central point. Additionally, I am studying the low zeros of Dirichlet characters with square-free modulus (which have applications to how primes are distributed in arithmetic progressions) and Rankin-Selberg convolutions of GL_n and GL_m families of L-functions (which highlights how the behavior of complex families can be understood in terms of the behavior of the building blocks). Finally, much of my recent work involves determining lower order terms in the behavior of zeros of L-functions. These terms are sadly inaccessible through standard random matrix theory, which misses arithmetic; however, through recent bold conjectures such as the L-functions Ratios Conjecture we have an excellent predictive method. I have studied these lower order terms in a variety of interesting families; the most exciting recent application (joint with several colleagues) a model which explains the observed repulsion of zeros near the central point in families of quadratic twists of elliptic curve L-functions, which we are currently extending to more general families. Similar to the Berry-Essen theorem from probability (where the universality of the Central Limit Theorem is due to the first and second moments, and the higher moments affect only the rate of convergence), I have shown similar phenomena exist in number theory (where the higher moments of the Fourier coefficients in families of L-functions affect the rate of convergence to the Random Matrix Theory predictions; in other words, the arithmetic of the families live only in the converge rates).

I am investigating numerous problems in Random Matrix Theory and Random Graphs, especially ensembles with few degrees of freedom (order N independent matrix elements, instead of order N^2). These provide fascinating windows to see new behavior and have numerous applications (k-regular graphs are used to construct cheap and efficient networks). Along these lines, I am also studying several problems on the boundary of Probability Theory, Number Theory and Analysis, such as proving that the distribution of the first digits of $|L(s, f)|$ near the critical line and iterates of the $3x+1$ map follows Benford's Law of digit bias (the first digit is a 1 about 30% of the time). These problems have led to results ranging from the distribution of digits of order statistics to a generalization of the central limit theorem for random variables modulo 1. With some colleagues and students I am extending these results and working on applications (I have been in contact with the Criminal Investigative Division of the IRS, helped organize the first conference on Benford's law, and am currently editing the first book on the theory and applications of the law). Using recent strong concentration results I proved a conjecture on the size of the sumset to the difference set in additive number theory / probability theory, with fascinating behavior at the critical threshold which I am continuing to explore.

I have also worked on and am pursuing several applied projects in Probability, Statistics, Linear Algebra and Cryptography, such as closed-form Bayesian inferences for the multinomial logit model, a binary integer linear programming problem for movie distributors, bounding incomplete multiple exponential sums arising in Computer Science, extreme cases of the Cramer-Rao inequality, modeling baseball games, and determining the security of certain signature schemes in cryptography, as well as studying the propagation of viruses in networks. I am also interested in computational aspects of these problems, writing algorithms to investigate many of these topics, from zeros of elliptic curve L-functions and moments of Dirichlet L-functions over function fields to random matrix theory and graph theory to Bayesian inference and linear programming.

2 Current NSF Grant

2.1 Background

The proposed research covers three major, but related, areas: distribution of zeros of L-functions (and corresponding problems for ensembles of random matrix theory), problems in additive number theory (especially those related to Zeckendorf decompositions, the structure of sum and difference sets, problems in point configurations and Ramsey theory), and Benford's law (in particular dependent random variables and fragmentation problems). We first provide some background for the problems, then summarize the PI's work under previous grants, and end with a description of the proposed research. The PI has extensive experience in supervising undergraduate, graduate, post-doc and junior faculty research; in addition to many of these projects being central in current investigations, several provide a very accessible introduction to higher mathematics, which works well with the PI being at an RUI institution and having numerous REU students. The PI will continue mentoring students and junior colleagues, giving public lectures, and writing introductory material (such as textbooks, conference resources, and survey articles).

In attempting to describe the energy levels of heavy nuclei [BFFMPW,Wig1,Wig2], researchers were confronted with daunting calculations for a many bodied system with extremely complicated interaction forces. Unable to explicitly calculate the energy levels, physicists developed Random Matrix Theory (RMT) to predict general properties of the systems. Surprisingly, the same model is an excellent predictor of many (but not all!) properties of zeros and values of L-functions [Con1, KS1, KS2]. The PI plans to continue his studies of several statistics, especially the behavior of zeros near the central point and the 1-level density in various families, which is defined by

$$\lim_{N\to\infty} \frac{1}{|\mathcal{F}_N|} \sum_{f\in\mathcal{F}_N} \sum_j \phi\left(\widetilde{\gamma}_{j;f}\right),$$

with ϕ an even Schwartz function, $\mathcal{F} = \cup\mathcal{F}_N$ a family of L-functions ordered by conductor, and (assuming GRH) the zeros of $L(s,f)$ are $1/2+i\gamma_{j;f}$ (the $\widetilde{\gamma}_{j;f}$ are the normalized imaginary parts of the non-trivial zeros). The Katz and Sarnak conjectures [KS1,KS2] state that as the conductors tend to infinity, the behavior of the zeros near the central point agree with the scaling limit of normalized eigenvalues of a classical compact group. In particular, while RMT does an excellent job predicting the main term, it misses the arithmetic, which surfaces as lower order corrections, which the PI proposes to isolate.

The development of RMT was motivated by Statistical Mechanics. For example, consider a room with N air molecules (which either move left or right at a constant speed v) and a macroscopic quantity such as pressure. To calculate the pressure one must know the position and velocity of each molecule, an unrealistic goal. Instead one approximates the pressure for a given configuration of molecules by finding the average pressure over all configurations, as most configurations have a pressure close to the system average. These central limit theorem type results led to modeling the energy levels of heavy nuclei by the eigenvalues of ensembles of matrices. At first physicists studied real symmetric and complex Hermitian matrices with entries independently drawn from Gaussians [Meh], though nowadays the ensembles are usually taken to be the classical compact groups with Haar measure [Con1,KS1,KS2]. The PI will continue his studies of the distribution of eigenvalues of structured ensembles; these families have significantly fewer degrees of freedom than the full families of real symmetric and complex Hermitian matrices, and provide an excellent opportunity to see new behavior. If the independent matrix elements are chosen from a nice distribution, the limiting spectral distribution is reduced to integrating the trace of powers of the matrix over the space, which converts the problem to a combinatorial one of determining which matching contribute in the limit (the answer depends on the structure of the ensemble).

The second major theme concerns problems in additive number theory, mostly related to Zeckendorf decompositions and sizes of sum and difference sets. Every integer can be written uniquely as a sum of non-adjacent Fibonacci summands (if we label $F_1 = 1$, $F_2 = 2$). Lekkerkerker computed the mean number of summands needed for integers in $[F_n, F_{n+1})$; other researchers showed the fluctuations here and in other recurrence relations are Gaussian. Numerous other questions can be asked, especially for signed decompositions and for the distribution of gaps between summands (both in the bulk as well as the longest gap). Another topic concerns sum and difference sets. Let A be a finite set of integers, and set $A+A$ and $A-A$ to be the set of all sums (respectively differences) of elements in A. As addition is commutative but subtraction is not, it was believed that 'most' of the time $|A+A| < |A-A|$. It thus came as a surprise when Martin and O'Bryant [MO] proved that if each element from $\{0, 1, \dots, N\}$ is in A with positive probability p then as $N \to \infty$ a positive percentage of A are sum dominated (i.e., $|A+A| > |A-A|$).

The final topic is Benford's law of digit bias. In many systems the probability of observing a first digit of d is not 1/9 but $\log_{10}(1 + 1/d)$; if this holds we say the system obeys Benford's law. Benford's law for the original sequence is equivalent to the logarithms modulo 1 being equidistributed, which suggests Fourier analytic techniques will be useful. The main problems studied are related to fragmentation questions.

2.2 Results from Prior Support

The PI is currently supported by NSF Grant DMS1265673 (Analysis, Number Theory and Combinatorics: \$135,610): *Low-Lying Zeros of L-functions & Problems in Additive Number Theory*, August 15, 2013 to August 15, 2016, as well as two extensions (for travel / stipends for students and to support his math riddles page); from 2006 to 2013 he was supported by NSF Grants DMS0600848 and DMS0970067. Additionally, (1) the PI received support from the DHS Center at Rutgers for work on a cryptography book for non-math majors, (2) VCTAL to develop mathematical modules on cryptography and streaming information for high school students (part of a large project to bring computational thinking to the classrooms, received small stipend), (3) NSF Grant DMS1347804 (the Williams College SMALL REU), where he was the PI and program director for two years (the PI did not receive any funds from this grant, but administered it and it supported some of his students and colleagues), (4) a supplement to the SMALL REU grant to travel with students and colleagues to conferences such as SACNAS and the Math Alliance's Field of Dreams (where he is an Alliance Mentor) to encourage students to pursue careers in mathematics, to inform them about research opportunities both as undergraduate and graduate students, and to give talks about careers at liberal arts institutions, (5) participant on the most recent SACNAS grant (travel / lodging reimbursed to talk to students, to give a talk and to help with writing the grant), and (6) AIM/ICERM REUF grant NSF Grant DMS1239280 to support to mentor four junior faculty in improving their research and working with students (small stipend and travel expenses). All papers are on the arXiv and the PI's homepage.

2.2.1 Intellectual Merits

On DMS1265673 the PI had 32 papers appear (in journals such as Algebra & Number Theory; Communications in Number Theory and Physics; Experimental Mathematics; Fibonacci Quarterly (4x); Journal of Combinatorial Theory, Series A; Journal of Mathematical Analysis and Applications; Journal of Number Theory (3x); Int Math Res Notices; Physical Review E; Random Matrices: Theory and Applications (2x)), 6 more are to appear, 10 are under review, and 10+ are in preparation and serve as starting points for the new proposed research. The published or accepted work is joint with over 50 undergraduates, 5 graduate students and 2 post-docs, as well as 7 junior faculty the PI is mentoring, and is listed in the main body of the bibliography with the other references (and for the convenience of the reader is also listed separately by category at the end of the bibliography). To leave room to describe the proposed research, we just give a brief summary of a representative sample of the work below.

1. *L-Functions (4 papers):* [AMil] (*Low-lying zeroes of Maass form L-functions*): We determine the 1-level density for Maass forms for support up to $(-2, 2)$ under smooth weighting. Previous work could not break $(-1, 1)$. Briefly, we first write down the explicit formula to convert the relevant sums over zeroes to sums over Hecke eigenvalues. We then average and apply the Kuznetsov trace formula and

reduce the difficulty to bounding an integral of shape

$$\int_{-\infty}^{\infty} J_{2ir}(X)\frac{rh_T(r)}{\cosh(\pi r)}dr,$$

where these J are Bessel functions, and h_T is a certain weight function. We break into cases: X "small" and X "large". For X small, we move the line of integration from $\mathbb{R}$ down to $\mathbb{R}-iR$ and take $R \to +\infty$, converting the integral to a sum over residues. The difficulty then lies in bounding a sum of residues of shape

$$T\sum_{k\geq 0}(-1)^k J_{2k+1}(X)\frac{P\left(\frac{2k+1}{2T}\right)}{\sin\left(\frac{2k+1}{2T}\pi\right)},$$

where P is closely related to h. To do this (after a few tricks), we apply an integral formula for these Bessel functions, switch summation and integration, apply Poisson summation, apply Fourier inversion, and then apply Poisson summation again. The result is a sum of Fourier coefficients, to which we apply the stationary phase method one by one. This yields the bound for X small. To handle X large, we use a precise asymptotic for the $J_{2ir}(X)$ term from Dunster [Du] (as found in [ST]). In fact, for X large it is enough to simply use the oscillation of $J_{2ir}(X)$ to get cancelation. It is worth noting that the same considerations would also be enough for the case of X small were the asymptotic expansion convergent.

[AMPT] (*The n-Level Density of Dirichlet L-Functions over $\mathbb{F}_q[T]$*): Hughes and Rudnick [HR] computed 1-level density statistics for low-lying zeros of the family of primitive Dirichlet L-functions of fixed prime conductor $Q \to \infty$, and verified the unitary symmetry predicted by random matrix theory. We compute 1- and 2-level statistics of the analogous family of Dirichlet L-functions over $\mathbb{F}_q(T)$. In our situation the test function is periodic and our results are only restricted by a decay condition on its Fourier coefficients, and include error terms.

[FMil] (*Surpassing the Ratios Conjecture in the 1-level density of Dirichlet L-functions*): We study the 1-level density of low-lying zeros of Dirichlet L-functions in the family of all characters modulo q, with $Q/2 < q \leq Q$. For test functions whose Fourier transform is supported in $(-3/2, 3/2)$, we calculate this quantity *beyond* the square-root cancellation expansion arising from the L-function Ratios Conjecture of Conrey, Farmer and Zirnbauer [CFZ] (they agree in the region the conjecture is believed), and isolate a new lower-order term which is not predicted by this powerful conjecture. This is the first family where the 1-level density is determined well enough to see a term which is not predicted by the Ratios Conjecture, and proves that the exponent of the error term $Q^{-\frac{1}{2}+\epsilon}$ in the Ratios Conjecture is best possible. The analysis requires several results on smooth sums of primes in arithmetic progression, combining divisor switching techniques and precise estimates on the mean value of smoothed sums of the reciprocal of Euler's totient function. We also show how various conjectures on the distribution of primes in arithmetic progression increase the support, from modest conjectures increasing up to $(-4, 4)$ to stronger ones (such as Montgomery's conjecture) giving arbitrary support.

[BMMPT-B] (*Gaps between zeros of GL(2) L-functions*): We extend previous work on large gaps between zeros of $\zeta(s)$ and Dirichlet L-functions to cuspidal newforms on GL(2). Our arguments are general and hold for elements of the Selberg class, and thus the applicability in the end is reduced to proving a given L-function has the conjectured behavior. We use Wirtinger's inequality to relate gaps to second moments of L-functions and their derivatives. We analyze in great generality the shifted second moments which arise and derive the needed main and lower order terms.

2. *Random Matrix Theory (3 papers):* [GKMN] (*The expected eigenvalue distribution of large, weighted d-regular graphs*): McKay [McK] proved the limiting spectral measures of the ensembles of d-regular graphs with N vertices converge to Kesten's measure as $N \to \infty$. We generalize and consider d-regular graphs with random weights, drawn from some distribution $\mathcal{W}$, on the edges. We establish the existence of a unique 'eigendistribution' (a weight distribution $\mathcal{W}$ such that the associated limiting spectral distribution is a rescaling of $\mathcal{W}$). Initial investigations suggested that the eigendistribution was

the semi-circle distribution, which by Wigner's Law is the limiting spectral measure for real symmetric matrices. We prove this is not the case, though the deviation between the eigendistribution and the semi-circular density is small (the first seven moments agree, and the difference in each higher moment is $O(1/d^2)$). The main ingredient in the proof is a combinatorial result about closed acyclic walks in large trees.

[MSTW] (*Limiting Spectral Measures for Random Matrix Ensembles with a Polynomial Link Function*): We studied ensembles of matrices where there is a link function L such that if $L(i,j) = L(i',j')$ then $a_{i,j} = a_{i',j'}$. Using the Method of Moments and an analysis of the resulting Diophantine equations, we show that the spectral measures associated with linear bivariate polynomials converge in probability and almost surely to universal non-semicircular distributions. We prove that these limiting distributions approach the semicircle in the limit of large values of the polynomial coefficients. We then prove that the spectral measures associated with the sum or difference of any two real-valued polynomials with different degrees converge in probability and almost surely to a universal semicircular distribution (we need this difference in order to deal with some resulting integrals). Many of the formulas were first conjectured by extensive numerical analysis.

[BLMST] (*Distribution of eigenvalues of weighted, structured matrix ensembles)*: Given a structured ensemble such that (i) each random variable occurs $o(N)$ times in each row and (ii) the limiting rescaled spectral measure $\widetilde{\mu}$ exists, we introduce a parameter to continuously interpolate between these two behaviors. We fix a $p \in [1/2, 1]$ and study the ensemble of signed structured matrices by multiplying the $(i,j)^{\text{th}}$ and $(j,i)^{\text{th}}$ entries of a matrix by a randomly chosen $\epsilon_{ij} \in \{1,-1\}$, with $\text{Prob}(\epsilon_{ij} = 1) = p$ (i.e., the Hadamard product). For $p = 1/2$ we prove that the limiting signed rescaled spectral measure is the semi-circle. For all other p, we prove the limiting measure has bounded (resp., unbounded) support if $\widetilde{\mu}$ has bounded (resp., unbounded) support, and converges to $\widetilde{\mu}$ as $p \to 1$. Notably, these results hold for Toeplitz and circulant matrix ensembles. The proofs are by Markov's Method of Moments. The analysis of the $2k^{\text{th}}$ moment for such distributions involves the pairings of $2k$ vertices on a circle. The contribution of each pairing in the signed case is weighted by a factor depending on p and the number of vertices involved in at least one crossing. These numbers are of interest in their own right, appearing in problems in combinatorics and knot theory. The number of configurations with no vertices involved in a crossing is well-studied, and are the Catalan numbers. We discover and prove similar formulas for configurations with $4, 6, 8$ and 10 vertices in at least one crossing. We derive a closed-form expression for the expected value and determine the asymptotics for the variance for the number of vertices in at least one crossing. As the variance converges to 4, these results allow us to deduce properties of the limiting measure.

3. *Zeckendorf decompositions (7 papers):* The main results here are in [BILMT] (*Gaps between summands in generalized Zeckendorf decompositions*). Earlier work established that for positive linear recurrence relations (these relations have non-negative coefficients and the first coefficient is positive), every positive integer has a unique legal decomposition in terms of the sequence $\{G_n\}$, and as $n \to \infty$ the number of summands used in decompositions of $m \in [G_n, G_{n+1})$ approaches a Gaussian. We prove that the distribution of gaps between summands converges to essentially a geometric random variable on average, and almost surely for the individual gap measures attached to each m; we also show the distribution of the longest gap converges to similar behavior as seen in the longest run of heads in a biased coin. The proofs on the gap measure follow from counting arguments applied to dependent events and Levy's criteria, while the longest gap is more technical and uses Rouche's theorem to prove needed results about general characteristic polynomials.

   In [DDKMMV1] (*Generalizing Zeckendorf's Theorem to f-decompositions*), [DDKMMV2] (*A Generalization of Fibonacci Far-Difference Representations and Gaussian Behavior* and [CFHMN] (*Generalizing Zeckendorf's Theorem: The Kentucky Sequence*) we explore different generalizations of previous work. The first introduces a new notion of a legal decomposition, where now if a_n is a summand then $a_{n-1}, a_{n-2}, \dots, a_{n-f(n)}$ are unavailable, the second allows positive and negative summands to be used, and the third is a recurrence where the first coefficient is zero. We obtain similar results, although interestingly a related sequence to the Kentucky one, currently being investigated with the same authors, does not have unique decompositions.

In [BDEMMTW1] (*Gaussian Behavior of the Number of Summands in Zeckendorf Decompositions in Small Intervals*) we show that it is not necessary to look at the natural intervals $[F_n, F_{n+1})$ and that similar behavior is obtained in almost all sub-intervals, while in [MW2] (*Gaussian Behavior in Generalized Zeckendorf Decompositions*) we provide a non-technical general overview of results in the area.

The paper [BDEMMTW1] (*Benford Behavior of Zeckendorf Decompositions*) will be described in greater detail in the Benford section.

4. *More Sums Than Differences Sets (4 papers):* [ILMZ2] (*Finding and Counting MSTD sets*) and [MilV] (*Most Subsets are Balanced in Finite Groups*) are two conference proceedings report on work with students on a variety of MSTD sets. The first includes (among other results) a powerful generalization of the base expansion method which allows us to completely resolve many questions, including proving a positive percentage of sets having a given linear combination greater than another linear combination, and a proof that a positive percentage of sets are k-generational sum-dominant (meaning A, $A+A$, ..., $kA = A + \cdots + A$ are each sum-dominant). The second looks at similar questions in finite groups. We show that if we take subsets of larger and larger finite groups uniformly at random, then not only does the probability of a set being sum-dominated tend to zero but the probability that $|A+A| = |A-A|$ tends to one, and hence a typical set is balanced in this case. The cause of this marked difference in behavior from subsets of the integers is that subsets of $\{0, \dots, n\}$ have a fringe, whereas finite groups do not. We give a detailed analysis of dihedral groups, where the results are in striking contrast to what occurs for subsets of integers. Specifically, even though almost all subsets of dihedral groups are balanced as the size grows, more sets are sum-dominated than difference-dominated.

   In [DKMMW] (*Sums and Differences of Correlated Random Sets*): Martin and O'Bryant [MO] showed a positive proportion of subsets of $\{0, \dots, n\}$ are sum-dominant. We generalize and study sums and differences of pairs of *correlated* sets (A, B) ($a \in \{0, \dots, n\}$ is in A with probability p, and a goes in B with probability ρ_1 if $a \in A$ and probability ρ_2 if $a \notin A$). If $|A+B| > |(A-B) \cup (B-A)|$, we call (A, B) a *sum-dominant* (p, ρ_1, ρ_2)*-pair*. We prove for any fixed $\vec{\rho} = (p, \rho_1, \rho_2)$ in $(0,1)^3$, (A, B) is a sum-dominant (p, ρ_1, ρ_2)-pair with positive probability, which approaches a limit $P(\vec{\rho})$. We investigate p decaying with n, generalizing results of Hegarty-Miller [HeMi] on phase transitions, and find the smallest sizes of MSTD pairs. The main ingredient in the proof is bounding probabilities of fringe structures, especially through Bayes' theorem.

   [DKMMWW] (*Sets Characterized by Missing Sums and Differences in Dilating Polytopes*): We generalize investigations of integers in a growing interval to the lattice points in a dilating polytope. Specifically, let P be a polytope in $\mathbb{R}^D$ with vertices in $\mathbb{Z}^D$, and let $\rho_n^{s,d}$ now denote the proportion of subsets of $L(nP)$ that are missing exactly s sums in $L(nP) + L(nP)$ and exactly $2d$ differences in $L(nP) - L(nP)$. The geometry of P has a significant effect on the limiting behavior of $\rho_n^{s,d}$. We introduce a geometric characteristic of polytopes called local point symmetry, and show that $\rho_n^{s,d}$ is bounded below by a positive constant as $n \to \infty$ if and only if P is locally point symmetric. We prove that the proportion of subsets in $L(nP)$ that are missing exactly s sums and at least $2d$ differences remains positive in the limit, independent of the geometry of P (this is the correct higher-dimensional analogue of being a More Sums Than Differences set, as the difference set frequently has far more candidate points than the sumset), and if P is additionally point symmetric, then the proportion of sum-dominant subsets of $L(nP)$ also remains positive in the limit.

5. *Benford's Law (3 papers and 1 book):* The main contribution is the edited book [Mil8] (*Theory and Applications of Benford's Law*), where the PI also wrote two chapters and collaborated on a third. This is the first book to have the theory and applications extensively developed together.

   The leading digits of the Fibonacci numbers obey Benford's law; in [BDEMMTW1] (*Benford Behavior of Zeckendorf Decompositions*) we prove that additionally the number of summands in Zeckendorf decompositions with leading digit d follows Benford's law. The main idea is to prove the result when we randomly choose summands in a legal decomposition to be taken with a fixed probability p, and then note that a special choice of p is equivalent to uniformly choosing integers.

[IMS] (*Equipartitions and a distribution for numbers: A statistical model for Benford's law*) considers fragmentation problems, and establishes new upper and lower bounds for the distribution of fragments with a given leading digit.

6. *General Number Theory (5 papers):* The two papers [ACM] (*Newman's conjecture in various settings*) and [CMMRSY] (*Newman's Conjecture in Function Fields* explore Newman's conjecture, which quantifies the maxim that if RH is true it is just barely true, in various settings. [AMPT] (*Sets of Special Primes in Function Fields*) and [ABMRS] (*Generalized Ramanujan Primes*) look at some special sets of primes (the first involves the totient function and constructing a non-trivial set of primes $\mathcal{P}$ where if $p \in \mathcal{P}$ so is r for any $r|p-1$, while the second involves determining interval sizes so that a given number of primes always lie in such intervals). [CHMW] (*Continued Fraction Digit Averages and Maclaurin's Inequalities*) generalizes the arithmetic mean - geometric mean results on digits of continued fractions to all symmetric means, and determines a phase transition.

7. *Sabermetrics (3 papers):* [CGLMP] (*Pythagoras at the Bat*) summarizes research on the Pythagorean Won-Loss formula to estimate a team's winning percentage, providing new justifications using just calculus and elementary algebra applied to baseball models; [HJM] (*The James Function*) extracts from the standard Pythagorean formula a set of conditions desirable for any predictor and determines the class of functions satisfying those conditions; [LMil] (*Relieving and Readjusting Pythagoras*) extends standard models on ballgames to include ballpark effects and rare events (among others), as well as introducing a Weibull-basis to obtain tractable closed-form predictive statistics.

8. *General / Surveys (9 papers):* The first five papers are a mix of survey articles and preliminary announcements on work in progress, much of which will be expanded on in the *New Problems* section. [MMRW] (*Lower-Order Biases in Elliptic Curve Fourier Coefficients in Families*) is a continuation of work begun in the PI's thesis on observed biases in the second moments of the Fourier coefficients of elliptic curve L-functions, and consequences on low-lying zeros. [OFMT-B] (*From Quantum Systems to L-Functions: Pair Correlation Statistics and Beyond*) reviews of the development of random matrix theory and its connections to number theory. [AAILMZ] (*Maass waveforms and low-lying zeros*) contains weaker results than [AMil] but a significantly easier exposition, highlighting the issues of applying the Kuznetsov trace formula. [FrMil] (*Determining Optimal Test Functions for Bounding the Average Rank in Families of L-Functions*) returns to a problem from [ILS]; they determine the optimal test function to bound the average rank in a family by using the 1-level density for support in $(-2, 2)$; we remove the support restriction by introducing a series of delay differential equations. Finally [MMRT-BW] (*Some Results in the Theory of Low-lying Zeros*) gives an elementary and accessible account of low-lying zeros, concentrating on families of Dirichlet characters and highlighting how the arithmetic of the family enters the analysis.

   [Mil4, Mil5, Mil6, Mil7] (*The Pi Mu Epsilon 100th Anniversary Problems: Parts I-IV*): As 2014 marked the 100$^{\text{th}}$ anniversary of Pi Mu Epsilon, the PI (who is the Problem Editor of Pi Mu Epsilon) decided to celebrate with 100 problems related to important mathematics milestones of the past century, meant to provide a brief tour through some of the most exciting and influential moments in recent mathematics. Several leading experts in the relevant fields authored pieces (the PI was the primary author on 47 entries, and edited all).

2.2.2 Broader Impacts

For almost twenty years, the PI has worked extensively with undergraduates; in the last three years he supervised 43 summer undergraduate research students and 7 senior theses (3 more co-advised). He has mentored and published with 6 graduate students, 2 postdocs and 6 junior faculty. Some of the proposed problems represent continuing work of the PI and colleagues at other institutions, while others are meant to introduce students and junior faculty to the fields and research. As such, many of these problems are meant to serve as ***springboards*** to get people interested and involved, and the final projects, though inspired by the questions below, may differ in the end. The PI has used this model very successfully with his previous grants, finding students respond well to the freedom to have input in problem selection (while at the same time appreciating and benefiting from some direction, which ensures much of the work will be of interest to senior researchers in the field). The resulting projects have appeared in good to top journals, have often been related to the major research interests of the PI, and have helped his students decide whether or not to pursue graduate studies in mathematics (and, if so, in which sub-division).

He has incorporated research projects into his classes during the academic year, resulting in several papers. His students have presented at both undergraduate and research conferences, including the Maine - Québec Number Theory Conference, the Combinatorial and Additive Number Theory conference, the Young Mathematicians Conference at Ohio State, and the Automorphic Forms Workshop (University of Michigan). In addition to working with all on writing papers and giving talks, the PI has also worked with them in refereeing papers for journals and writing reviews for MathSciNet, in speaking at summer programs for talented math students, in mentoring later colleagues of the PI, and doing fun math units in local schools.

The PI is also active in mathematics education. He has written an introductory textbook in cryptography, just completed one on an introduction to Benford's law, and is finishing books on Probability and Operations Research. Related to this, the PI is writing modules for high school math courses (so far on the science of encryption and information protection, and streaming information), and working with math teachers at all levels through a variety of outreach efforts (from his math riddles webpage to continuing education classes in the Teachers as Scholars program to volunteering in schools).

2.3 Proposed Problems

The PI proposes to study a large number of problems in several related areas of number theory. He will work extensively with undergraduates (both thesis students at Williams and summer students), graduate students, postdocs and junior faculty (at Williams and at other institutions). For the past six years he has had students and junior colleagues working with him on multiple projects simultaneously; this model has worked very well for the PI and his colleagues, and the current plan is to build on this strength. Thus, in the interest of space, the project descriptions are a little short in order to highlight the breadth of problems being studied; however, many of these problems are natural outgrowths of previous work of the PI and more information can be found in the section on previous work.

2.3.1 Low-lying zeros of L-functions and Random Matrix Theory

The 1-level density is an excellent statistic to investigate properties of zeros of L-functions near the central point, which is often where interesting arithmetic occurs. All families of L-functions investigated to date have scaling limits agreeing with the scaling limit of a classical compact group (unitary, orthogonal or symplectic), or a trivial modification (such as elliptic curve L-functions, where there is an extra copy of the identity matrix which models the family zeros at the central point). The classical compact groups all have different densities, though the three orthogonal types cannot be distinguished for test functions supported in $(-1, 1)$.

There are many reasons to desire as large support as possible. In addition to being able to distinguish the orthogonal flavors, often breaking $(-1, 1)$ (or more) is equivalent to understanding finer arithmetic questions. Examples include Dirichlet L-functions, where larger support is related to the distribution of the primes in congruence classes, to cuspidal newforms (where Iwaniec, Luo and Sarnak [ILS] show the equivalence of larger support to exponential sums over primes.

1. *Petersson Formulas for arbitrary level and low-lying zeros of cuspidal newforms.* Many investigations of cuspidal newforms restrict the level to square-free (or sometimes even prime) to avoid the technical obstructions arising from the factorization of the level and the resulting inclusion-exclusion needed to obtain a tractable basis. A notable recent exception is the work of Blomer and Milićević [BM], where they obtain a basis for cusp forms for general N. The PI proposes to extend this work and generalize

arguments in [ILS] and [Rou] to develop a tractable Petersson formula for such N. Preliminary investigations are promising. In his Summer '15 REU, his students have begun the work and have corrected several minor mistakes in [BM], and have developed the resulting Petersson formula in the simpler case where the argument of the Fourier coefficient is relatively prime to the level: if $(n, N) = 1$ then we have shown

$$\Delta^{\star}_{k,N}(n) \ = \ \frac{k-1}{12} \sum_{LM=N} \mu(L) M \prod_{p^2|M} \left(\frac{p^2}{p^2-1}\right)^{-1} \sum_{(m,M)=1} \frac{1}{m} \Delta_{k,M}(m^2, n),$$

where

$$\Delta_{k,N}(m,n) \ = \ \frac{12}{(k-1)\nu(N)} \sum_{LM=N} \frac{M}{\phi(M)} \sum_{f \in \mathrm{H}^{\star}_k(M)} \frac{1}{\mathrm{Z}(1,f)}$$
$$\times \sum_{d|L} \overline{\left(\sum_{\ell|(d,m)} \xi_d(\ell) \ell^{1/2} \lambda_f(\tfrac{m}{\ell})\right)} \left(\sum_{\ell|(d,n)} \xi_d(\ell) \ell^{1/2} \lambda_f(\tfrac{n}{\ell})\right)$$

with the linear combinations chosen to give us orthonormal bases. In addition to being of interest in its own right, this formula (and the generalization to arbitrary n) will be of use in computing statistics such as the 1-level density. In particular, the PI will use it to investigate lower order terms in the 1-level density. While in many families the main terms agree with Random Matrix Theory, the lower order terms depend on arithmetic of the family, and thus we hope to see how the factorization of N effects the lower order terms (the first lower order term plays an important role in determining the behavior of zeros at the central point for finite conductors). In developing the Petersson formula it was essential that $(n, N) = 1$; we hope to remove this restriction,, which would allow us to split the families by sign of the functional equation.

2. *Optimal test functions for bounding order of vanishing in families at the central point.* In [ILS] the authors determine the optimal test functions to obtain the best bounds on excess rank at the central point using the 1-level density when the support of the Fourier transform of the test function lies in $(-2, 2)$ for the orthogonal families. In [FrMil] the PI and his recent thesis student, Jesse Freeman, determine the optimal test functions for the 1-level density for support in $(-3, 3)$. We proved that if ϕ is an even Schwartz test function such that $\mathrm{supp}(\widehat{\phi}) \subset [-2s, 2s]$, with $s = \sigma/2$, then for $2 < \sigma < 3$ the optimal test function is given by $\widehat{\phi} = g_{\mathcal{G},\sigma} * \mathfrak{g}_{\mathcal{G},\sigma}$. Here $*$ represents convolution, $\mathfrak{g}_{\mathcal{G},\sigma}(x) = \overline{g_{\mathcal{G},\sigma}}(-x)$, and $g_{\mathcal{G},\sigma}$ is given by

$$g_{\mathrm{SO(even)},\sigma}(x) \ = \ \lambda_{\mathrm{SO(even)},\sigma} \begin{cases} c_{1,\mathcal{G},\sigma} \cos\left(\frac{|x|}{\sqrt{2}}\right) & |x| \le \sigma/2 - 1 \\ \cos\left(\frac{|x|}{2} - \frac{(\pi+1)}{4}\right) & \sigma/2 - 1 \le |x| \le 2 - \sigma/2 \\ \frac{c_{1,\mathcal{G},\sigma}}{\sqrt{2}} \sin\left(\frac{x-1}{\sqrt{2}}\right) + c_{3,\mathcal{G},\sigma} & 2 - \sigma/2 < |x| < \sigma/2 \\ 0 & |x| \ge \sigma/2, \end{cases}$$

and

$$g_{\mathrm{O},\sigma}(x) \ = \ \begin{cases} \frac{1}{1+\sigma/2} & |x| < \sigma/2 \\ 0 & |x| \ge \sigma/2 \end{cases}$$

for $\mathcal{G} \ = \ \mathrm{O}$, and

$$g_{\mathcal{G},\sigma}(x) \ = \ \lambda_{\mathcal{G},\sigma} \begin{cases} c_{1,\mathcal{G},\sigma} \cos\left(\frac{|x|}{\sqrt{2}}\right) & |x| \le \sigma/2 - 1 \\ \cos\left(\frac{|x|}{2} + \frac{(\pi-1)}{4}\right) & \sigma/2 - 1 \le |x| \le 2 - \sigma/2 \\ \frac{-c_{1,\mathcal{G},\sigma}}{\sqrt{2}} \sin\left(\frac{x-1}{\sqrt{2}}\right) + c_{3,\mathcal{G},\sigma} & 2 - \sigma/2 < |x| < \sigma/2 \\ 0 & |x| \ge \sigma/2 \end{cases}$$

for $\mathcal{G} = \mathrm{SO(odd)}$ or Sp. Here, the $c_{i,\mathcal{G}}$ and $\lambda_\mathcal{G}$ are easily explicitly computed. In the proposed work we will extend these results two ways: (1) arbitrary support, and (2) generalize to n-level densities. The work on (1) uses the Fredholm alternative, Paley-Weiner theory, and a system of delay differential equations. Specifically, we deduce an optimality criterion which holds for all $s \in \mathbb{R}^+$, where $\mathrm{supp}(\hat{\phi}) \subset [-2s, 2s]$. Our kernels give us a system of location-specific integral equations, which we will use to extend the results to larger support.

3. *Biases in second moments of Fourier coefficients of families of L-functions.* In his thesis [Mil1] the PI noticed that for one parameter families of elliptic curves over $\mathbb{Q}[T]$ that the second moments of the $a_t(p)$ exhibited a bias. By this we mean the following: Michel proved that if $j(T)$ is non-constant then $\sum_{t \bmod p} a_t(p)^2 = p^2 + O(p^{3/2})$, with the lower order terms having co-homological interpretations. In every family he investigated the first term which did not average to zero had a negative average. More evidence was recently obtained by him and his students in [MMRW], where they proved this bias exists in numerous families. The PI proposes to continue this investigation to other families of L-functions, such as Dirichlet L-functions, cuspidal forms, and symmetric power lifts. The main tools are averaging formulas (orthogonality of characters, Petersson formula, and formulas for Legendre sums). The bias of these terms are very important in the n-level densities, and play a major role in identifying lower order terms (which in the next project are related to the behavior of zeros near the central point).

4. *Models for zeros at the central point for finite conductors.* In [DHKMS1, DHKMS2] the PI and his collaborators introduced the Excised Orthogonal Ensemble to model the behavior of zeros of quadratic twists of elliptic curve L-functions at the central point. This new approach succeeded beautifully in explaining the observed repulsion which vanishes as the conductors go to infinity but is strongly seen in the range experimentally reachable. The excised ensemble has two parameters, one determining an effective matrix size, and the other restricting the orthogonal matrices to those whose characteristic polynomial at $\theta = 1$ is a certain size in absolute value; this is responsible for the 'excision', and incorporates the discreteness of the elliptic curve L-functions at the central point.

   We plan on extending this model to other families, including Dirichlet L-functions (and provide an interpretation for the discrepancy seen there from the limiting behavior as the matrix sizes tend to infinnity) and cuspidal newforms on GL(2) and symmetric power lifts. The PI has begun investigations of these cases with his REU students and Nathan Ryan at Bucknell. One way of obtaining the effective matrix size is to match the first lower order terms inn number theory and random matrix theory; interestingly when this happens we obtain a *negative* size for quadratic Dirichlet characters (this is not terminal, as the RMT formulas can be analytically continued and negative values make sense). We plan on gathering numerical evidence for this and related families and testing the new theoretical predictions. Many of the calculations are first done through the L-Function Ratios Conjecture [CFZ1, CFZ2].

5. *Large gaps between zeros of L-functions.* In [BMMPT-B] the PI and his colleagues generalized existing techniques to study large gaps between zeros of GL(2) L-functions. Combining mean-value estimates of Montgomery and Vaughan with a method of Ramachandra, we prove a formula for the mixed second moments of derivatives of $L(1/2 + it, f)$. As an application, we use an argument of Hall to show that there are infinitely many gaps between consecutive zeros of $L(1/2 + it, f)$ that are at least $\sqrt{3} = 1.316\dots$ times the average spacing. The linking factor between the moments and gaps is Wirtinger's inequality: Let $y : [a, b] \to \mathbb{C}$ be a continuously differentiable function, and suppose that $y(a) = y(b) = 0$. Then
$$\int_a^b |y(x)|^2 dx \ \le\ \left(\frac{b-a}{\pi}\right)^2 \int_a^b |y'(x)|^2\, dx.$$
   Building on this success and other recent work in the field, we propose to go further by using mollifiers and fourth moments. We will expand on the ideas in [HY] and [HH] to attack shifted sums. Our first task will be the case where we average over the family, as then we can exploit additional cancellation through the Petersson formula; if we are successful there we will consider the more general case of

a fixed form. Preliminary investigations give four terms to study, arising from the Kronecker delta versus the Bessel-Kloosterman term in the Petersson formula, and from whether or not the resulting quadratic form is zero or non-zero. To date we have had success in analyzing the Bessel-Kloosterman piece attached to the non-zero associated value, and are optimistic that similar techniques will handle the other cases.

6. *Alternative formulations of the Katz-Sarnak determinantal expansions and applications.* The PI and C. Hughes [HuMi] computed the n-level density for cuspidal newforms. The main difficulty was a multidimensional Bessel-Kloosterman sum. We converted this to a 1-dimensional sum, at the cost of replacing the test function with an n-fold convolution. While this facilitated the analysis, it led to difficulties in comparing with random matrix theory. We derived a new and more tractable expression for the n-level densities. This new formula facilitated comparisons for test functions supported in $(-\frac{1}{n-1}, \frac{1}{n-1})$ by simplifying the combinatorial arguments. This work suggests that additional terms emerge whenever the support breaks $(-\frac{1}{n-k}, \frac{1}{n-k})$. Building on this success, the PI plans on further developing the combinatorics to increase the support where this alternative formula holds. As an application, the n-level density of cuspidal newforms, currently known only up to $(-\frac{1}{n-1}, \frac{1}{n-1})$, will be expanded (currently our new methods give $(-\frac{1}{n-2}, \frac{1}{n-2})$. The plan is to generalize Hypothesis S from [ILS], which the authors use to break $(-2, 2)$ for the 1-level density, to an n-dimensional analogue. These are very delicate calculations; as [ILS] remark, an exponential sum similar to that in Hypothesis S does not have the conjectured cancelation.

2.3.2 Classical Random Matrix Theory

1. *Structured matrices.* The PI has extensive experience [HaMi,MMS,JMP,KKMSX] in studying the limiting spectral measure of families of structured matrices. Let A be an $N \times N$ matrix with some structure (perhaps it is Toeplitz and constant along diagonals), with independent entries drawn from a fixed density p with mean 0, variance 1 and finite higher moments. The empirical spectral measure

$$\mu_{A,N}(x)dx \;=\; \frac{1}{N}\sum_{i=1}^{N}\delta\left(x-\frac{\lambda_i(A)}{\sqrt{N}}\right)dx$$

often converges (weakly, almost surely) to a limiting spectral measure. The determination of the limit if often deduced to delicate combinatorics, whose solution depends on finding the proper combinatorial vantage. The PI plans on further studies of interesting ensembles. One of the more interesting ensembles are checkerboard matrices, where every k entries in row i vanish, starting at entry $i \bmod k$. So far we have shown that, properly normalized, the limiting spectral measure of most of the eigenvalues converges to the semi-circle; however, there is a bounded number of eigenvalues that escape – the plan is to determine the proper normalization to focus on those eigenvalues and see (in what sense) they have a limiting behavior.

2.3.3 Generalized Zeckendorf Decompositions

1. *Non-Positive Linear Recurrences.* Much of the previous literature in generalized Zeckendorf decompositions (see for example [BILMT]) assumes that the linear recurrence has non-negative coefficients and positive leading term. Working with four junior faculty (this work was begun at a REUF program at AIM and continue at ICERM), the PI is investigating what happens when the leading coefficient vanishes. In some cases unique decompositions still exist, the greedy algorithm still leads to valid decompositions, and the fluctuation of the number of summands is Gaussian and the gaps between summands is geometric decay; the situation is strikingly different in other cases. We have associated a sequence to the Fibonacci Spiral (the sequence of squares which tile the plane, declaring a decomposition legal if we never use two terms who share an edge). Every integer has a decomposition, but the decompositions are not unique (by solving an associated recurrence we have shown the number

grows exponentially), and the greedy algorithm only terminates successfully in a decomposition approximately 93% of the time. We plan on continuing this work to see if the fluctuation in the number of summands and gaps are more universal and still exhibit the same behavior. One of the main ingredients in these problems is a new method which bypasses many of the technical obstructions in previous work (see for example [MW1]). There the difficulty was to show certain coefficients were positive, which required involved analysis of the polynomials associated to the initial recurrence. This analysis could only be done in the special case listed earlier. We are developing a new technique which uses a partial fraction decomposition of a two variable generating function, and then show the needed coefficients must be positive or the number of numbers in our interval would be too small. We have already applied this argument to some recurrences where the leading term is zero, but these have unique decomposition. We plan on extending this method to the Fibonacci Quilt and other cases, developing a useful tool to handle many similar problems.

2. *Geometric Notions of Legal Decompositions.* In addition to the Fibonacci Spiral, the PI will explore other notions leading to legal decompositions, and see how the resulting geometry affects the behavior.

2.3.4 Sum and Difference Sets

1. *Phase transitions.* The PI and P. Hegarty [HeMi] proved that the existence of a positive percentage of sets $A \subset \{0, \dots, n-1\}$ being sum-dominant (i.e., $|A+A| > |A-A|$) as $n \to \infty$ is due to choosing each $k \in \{0, \dots, n-1\}$ to be in A with a positive probability independent of n. If all k are still chosen equally but with some probability $p(n)$ decaying to zero, then with probability one a set is difference dominated. Finer behavior of the relative sizes depends on the decay rate, and there is a phase transition when $p(n) = n^{-1/2}$. The PI plans on exploring the presence of phase transitions in other phenomena, especially when there are h summands, each weighted by ± 1. The critical threshold is now $p(n) = n^{-(h-1)/h}$; we currently have results up to and including this critical exponent (the main idea involves generalizing the combinatorics from [HeMi] to bound the number of repeated elements), and are investigating the case of slow decay (i.e., exponents less than $(h-1)/h$).

2. *Explicit constructions of generalized MSTD sets.* In [ILMZ1] the PI and his students proved many properties about generalized MSTD sets, including that for any k a k-generational set exists, but no set works for all k (a set A is *k-generational* if $A, 2A, \dots, kA$ are all MSTD). We have already improved previous constructions from $|A| = \Omega(k!^2)$ to $\Omega(k)$ by replacing the base-expansion technique with a more delicate fringe argument. We plan on exploring how far we can push these fringe arguments. Other recent successes include proving a positive percentage of sets are bi-MSTD (a set $A \subseteq [0, n]$ is *bi-MSTD* if A and A^c are both MSTD, with A^c the set of elements in $[0, n]$ not in A); we plan on trying to push this further (for what m can a positive percentage of sets be written as a disjoint union of m MSTD sets?) Related to the above, we will explore higher-dimensional versions of these problems. In previous work [DKMMWW] the author and his colleagues found the right interpretation of being MSTD in higher dimensions was to have fewer missing sums than missing differences (as the difference set was just naturally so much larger); the behavior depended greatly on the geometry of the polytope. There are many additional questions we will pursue in higher dimensions, including the need to find computationally efficient algorithms so that the behavior of the quantities on the dimension can be numerically isolated.

2.3.5 Ramsey Theory for Sets Avoiding 3-term Geometric Progressions

Joint with Nathan McNew, Towson

1. *Finite Fields and Non-commutative sets.* Last summer the PI and Nathan McNew generalized earlier work on sets avoiding three term arithmetic progressions from subsets of $\mathbb{Z}$ to subsets of number

fields, to see how the structure of the number field affects the behavior. We propose to look at two other natural extensions. The first is finite fields, the second are non-commutative sets (such as the quaternions or families of matrices). We have made excellent progress on the finite field case; the main difficulty involved a combinatorial inclusion / exclusion of polynomials of different lengths. The non-commutative case is considerably harder. The proposed research will look at various free groups and see how the structure (number of elements and their orders) affect the answer; a nice application will be to understand $\mathrm{PSL}(2, \mathbb{Z})$. We have spent the past few months learning the quaternion background, and are now in a position to deal with the issues arising from the non-commutativity. The main problem is figuring out how elements of a given norm can arise as multiples or multiples of squares of numbers of smaller norm. The analysis often reduces to counting lattice points in 4-dimensional space.

2.3.6 Point Configuration

Joint with Eyvi Palsson, Virginia Tech

1. *Erdős Distance Problem.* The Erdős distinct distance problem asks what is the least number of distinct distances among N points in the plane. This problem, which Erdős later described as his most striking contribution to geometry, turned out to be quite hard. Erdős conjectured that asymptotically the least amount of distinct distances was on the order of $N/\sqrt{\log(N)}$. This was essentially solved by Guth and Kat [GK] in 2010 when they obtained the lower bound $N/\log(N)$ although the problem is still open in higher dimensions. Throughout the years Erdős explored and made many conjectures regarding the structure of sets that asymptotically obtain the lower bound. Many of these conjectures turned out to be even harder than the original question and are still unsolved. One of the key ingredients from [GK] was the polynomial ham sandwich theorem. We will apply this new technique to some open problems. We have already shown that optimal sets in higher dimensions must have many points on a single hyperplanes and also on a single hypersphere. We conjecture in 3-D that optimal sets must lie in a special lattice that is based on the hexagonal triangular lattice in the plane, which we propose to explore further.

2. *Point Configurations.* The point configuration problems in geometric measure theory and combinatorics can be thought of as continuous multipoint analogues of the well known Erdős distinct distance problem. Recently Palsson and his collaborators proved that say if a subset of the Euclidean space has large enough Hausdorff dimension then it is guaranteed that you have many distinct patterns of a particular type. Few sharpness examples have been found, apart from the trivial ones, so it is not entirely clear what the right conjectures should be. We will extend the few sharpness examples that exist from distances to higher order configurations, such as triangles, and take them into higher dimensions. Some delicate number theory arose when they had to count the number of triangles in an integer lattice adapted to a paraboloid, such as looking at how many numbers up to kn^2 are the sum of k calculator numbers from the table up to n.

2.3.7 Benford's Law

Many systems exhibit a digit bias. For example, the first digit base 10 of the Fibonacci numbers, or of 2^n, equals 1 not 10% or 11% of the time, as one would expect if all digits were equally likely, but about 30% of the time. This phenomenon, known as Benford's Law, has many applications, ranging from detecting tax fraud for the IRS to analyzing round-off errors in computer science. The central question is determining which data sets follow Benford's law. There has been a lot of work on independent random variables; below we propose a project involving dependent random variables.

1. *Dependent random variables and fragmentation.* Inspired by natural processes such as particle decay, the PI will explore various models for the decomposition of conserved quantities (see [Lem]). The PI has already proved that often the distribution of lengths converges to Benford behavior as the number of divisions grow. The main difficulty is that the resulting random variables are dependent, which

requires a careful analysis of the dependencies and tools from Fourier analysis to obtain quantified convergence rates. The solution proceeds by quantifying levels of dependence and how often two pieces share a given number of cuts, and then using earlier work of his quantifying the convergence. The PI will extend these results to more complicated fragmentation models, especially those in higher dimensions. There is an extensive literature in physics on fragmentation. Though much of it is not rigorous by mathematical standards, there are many good ideas and models worth exploring. Much of this work will be joint with Frederick Strauch (physics department, Williams College) and math/physics majors at Williams College. One project, joint with Nathan McNew, is to look at the distribution of digits of prime factors of an integer.

2. *Matrix groups.* With C. Manack the PI proposes to investigate the distribution of digits from random matrix ensembles. In many cases we have already determined these laws; in the compact case the behavior is reduced to that of coordinates of points on spheres. The arguments are a mix of classic results about Haar measure, Lie Theory, and standard analysis.

3. *Fraud detection.* The PI has written several papers with Mark Nigrini (professor of accounting) on the theory and application of Benford's law, especially with an eye to detecting data fraud [MN1,MN2,NM1,NM2]. He has submitted to the IRS a new method to detect fraud using Benford's law, and hopes to work with them on refining such tests.

2.4 Intellectual Merit and Broader Impact

The PI is a professor at Williams College with a 2-2 teaching load, which is what his load was during his previous two grants. He thus has experience balancing teaching and research, in particular in getting undergraduates, graduate students and junior faculty involved. The PI plans on investigating many of these problems with students and junior colleagues, a model which has worked well for him throughout his career. Thus, while the PI is proposing to study a large number of problems, his goal is to create an environment (both during the academic school year as well as the summer) where students will be colleagues, and graduate students and junior faculty will gain experience in designing and supervising research programs.

2.4.1 Intellectual Merit of the Proposed Work

The questions the PI proposes to study range from some of the most fundamental in the subject to some more standard questions (which are not only very appropriate for undergraduates, but serve as excellent introductions to the subject and often lead to deeper questions). We describe the intellectual merit of a few of them below.

The problems fall naturally into several groups; in the interest of space we describe just two of them. The first concerns the distribution of zeros of L-functions and the eigenvalues of matrix ensembles. These are some of the most fundamental and important objects in mathematics, encoding the answers to numerous problems. The models being explored have applications in physics as well.

The primary goal is to obtain a better understanding of the arithmetic of families of L-functions. The 1-level density provides an excellent window to see such behavior. The proposed work will explore a variety of families, developing techniques to isolate the arithmetic contributions. In the course of these investigations numerous combinatorial challenges will surface. It is quite likely that the techniques and results used to surmount these difficulties will be of independent interest in allied fields.

The second major theme is the distribution of the number of summands and gaps between them. This is similar in spirit to the studies of zeros of L-functions and eigenvalues of matrix ensembles. The two main questions are how many objects are there, and how are they distributed. As recurrence relations model a variety of phenomena, the techniques and results might be of independent interest in related fields (the distribution of the largest gap can be cast as an extreme value problem, which is studied by many researchers). The PI is developing a new technique which will bypass the technical obstructions that have plagued the subject, forcing people to derive difficult bounds on roots of polynomials associated to the recurrence relation;

the new method avoids these difficulties and replaces it with a simple counting problem. We have successfully applied this in some systems, and plan on further developing this new method. Additionally, the proposed additive number theory problems involve understanding phase transitions, another popular research topic cutting across many fields.

2.4.2 Broader Impacts of the Proposed Work

The PI is in his eighth year (and is tenured) at Williams College, which prides itself on its excellence in teaching and efforts to attract people to mathematics. While the national average of math majors is about 1%, at Williams it is 10%. There are a large number of students with strong backgrounds who are looking to see the connections and applications of higher mathematics. The PI will work with this talented pool of students, both through his classes (where he constantly introduces research projects), senior thesis students, and summer REU students (the summer REU program at Williams is one of the largest in the nation, attracting over 200+ applications each year).

The PI has extensive experience in involving undergraduates, graduate students and junior faculty in research, having supervised almost 300 undergraduates in research and over 20 graduate students. Much of the PI's work has focused on mentoring older students (upperclassmen or graduate students) in how to mentor younger students; the PI plans to build on these successes in his future research groups. Many of his students have continued to graduate school in mathematics or allied disciplines, and have found their experiences extremely helpful in understanding what academic life is like. A typical comment from the anonymous end-of-summer reviews was: "*Besides research, I learned a lot in writing good papers, giving presentations, choosing grad school, and more, which all help me a lot now as well as the career in math in the future.*"

The PI maintains numerous websites of material for beginning students and professors interested in starting undergraduate research programs. These files are available on his homepage, and range from notes on L-functions to reports on panels on undergraduate research.

The PI is strongly committed to providing opportunities to highly motivated students of all backgrounds. In his past three REUs, the PI has had 14 women and several under-represented minorities (the PI is an Alliance Mentor of the Math Alliance, was part of a recent grant to support SACNAS, and has been extensively involved in outreach activities such as the Field of Dreams and SACNAS meetings). His students are constantly encouraged and invited to attend conferences and mingle with professional mathematicians; in the past three years all undergraduates whom he has worked with under his current grant were invited to present at research conferences.

Recognizing that the colleagues of tomorrow are the college students and high school students of today, the PI does many activities to encourage people to explore and pursue mathematics. These range from giving lectures at programs such as the Ross Program at Ohio State, PROMYS at Boston University and Hampshire College's summer program to being a research mentor at PROMYS (advising high school students and their college councilors) to writing interesting and expository articles (ranging from articles for the Monthly to a survey of the path from nuclear physics to number theory) to visiting and giving talks in junior high and high schools. He has also served as a visiting committee for high school math departments, and is on the school committee of the regional high school, where he has actively worked on collaborative programs. Examples include one set of students in his Operations Research class working with the principal to optimize daily schedules to increase the accessibility of course offerings, and another coordinating scheduling for Math Blast, when all tenth graders from several local schools come to Williams and we try to optimize assignments to general math/stat lectures.

In order to excite young students, the PI has been running a math riddles page for over a decade (located at http://mathriddles.williams.edu/). The site is often the number one hit when googling 'math riddles'. It gets over 4000 hits per month, and is being used in junior high and high schools around the world. The PI corresponds extensively with teachers and students about the mathematics behind these riddles. The PI has created a student / teacher corner for the website, to facilitate the use of these riddles in classrooms. The PI plans on continuing this expansion.

The results from this grant will be disseminated through a variety of channels. In addition to traditional journal publications, the PI is working on books on probability, linear programming and a list of 100 problems related to each of the past 100 years (in honor of Pi Mu Epsilon's centennial), all with the goal of making these important subjects accessible to a wide audience. The PI will also present the results at colleges and conferences, as well as organize meetings on these subjects. All code developed will be made freely available, and all results posted on the arXiv.

Finally, the Benford law research (in particular, the work with Mark Nigrini on order statistics [NM2]) has already found applications in data integrity. The PI plans on continuing these and related studies (to detect dependence), and will continue his discussions with workers at the Internal Revenue Service about the issues and applications with fraud detection.

The references below are from the NSF Research Description

References

[AAILMZ] L. Alpoge, N. Amersi, G. Iyer, O. Lazarev, S. J. Miller and L. Zhang, *Maass waveforms and low-lying zeros*, to appear in "Analytic Number Theory: In honor of Helmut Maier's 60th birthday," Springer-Verlag.

[AMil] L. Alpoge and S. J. Miller, *The low-lying zeros of level 1 Maass forms*, Int Math Res Notices (2014), 24 pages, doi:10.1093/imrn/rnu012.

[ABMRS] N. Amersi, O. Beckwith, S. J. Miller, R. Ronan and J. Sondow, *Generalized Ramanujan Primes*, Combinatorial and Additive Number Theory, CANT 2011 and 2012 (Melvyn B. Nathanson, editor), Springer Proceedings in Mathematics & Statistics (2014), 1–13.

[ACM] J. Andrade, A. Chang and S. J. Miller, *Newman's conjecture in various settings*, Journal of Number Theory **144** (2014), 70–91.

[AMPT] J. Andrade, S. J. Miller, K. Pratt and M.-T. Trinh, *Sets of Special Primes in Function Fields*, INTEGERS **14** (2014), #A18.

[AMPT] J. Andrade, S. J. Miller, K. Pratt and M.-T. Trinh, *The n-Level Density of Dirichlet L-Functions over $\mathbb{F}_q[T]$*, Communications in Number Theory and Physics **8** (2014), no. 3, 1–29.

[BFMT-B] O. Barrett, F. W. K. Firk, S. J. Miller and C. Turnage-Butterbaugh, *From Quantum Systems to L-Functions: Pair Correlation Statistics and Beyond*, to appear in Open Problems in Mathematics (editors John Nash Jr. and Michael Th. Rassias), Springer-Verlag.

[BMMPT-B] O. Barrett, B. McDonald, S. J. Miller, C. Turnage-Burrerbaugh and K. Winsor, *Gaps between zeros of GL(2) L-functions*, Journal of Mathematical Analysis and Applications **429** (2015), 204–232.

[BLMST] O. Beckwith, V. Luo, S. J. Miller, K. Shen and N. Triantafillou, *Distribution of eigenvalues of weighted, structured matrix ensembles*, Integers: Electronic Journal Of Combinatorial Number Theory **15** (2015), paper A21, 28 pages.

[BDEMMTW1] A. Best, P. Dynes, X. Edelsbrunner, B. McDonald, S. J. Miller, C. Turnage-Butterbaugh and M. Weinstein, *Benford Behavior of Zeckendorf Decompositions*, Fibonacci Quarterly **52** (2014), no. 5, 35–46.

[BDEMMTW1] A. Best, P. Dynes, X. Edelsbrunner, B. McDonald, S. J. Miller, C. Turnage-Butterbaugh and M. Weinstein, *Gaussian Behavior of the Number of Summands in Zeckendorf Decompositions in Small Intervals*, Fibonacci Quarterly **52** (2014), no. 5, 47–53.

[BM] V. Blomer and D. Milićević, *The Second Moment of Twisted Modular L-Functions*, Geom. Funct. Anal. Vol. **25**, 2000, 453–516.

[BILMT] A. Bower, R. Insoft, S. Li, S. J. Miller and P. Tosteson, *Gaps between summands in generalized Zeckendorf decompositions* (and an appendix on *Extensions to Initial Segments* with Iddo Ben-Ari), Journal of Combinatorial Theory, Series A **135** (2015), 130–160.

[BFFMPW] BFFMPW T. Brody, J. Flores, J. French, P. Mello, A. Pandey, and S. Wong, *Random-matrix physics: spectrum and strength fluctuations*, Rev. Mod. Phys., vol. 53 (1981), no. 3, 385–479.

[CFHMN] M. Catral, P. Ford, P. Harris, S. J. Miller and D. Nelson, *Generalizing Zeckendorf's Theorem: The Kentucky Sequence*, Fibonacci Quarterly **52** (2014), no. 5, 68–90.

[CHMW] F. Cellarosi, D. Hensley, S. J. Miller and J. Wellens, *Continued Fraction Digit Averages and Maclaurin's Inequalities*, Experimental Mathematics **24** (2015), no. 1, 23–44.

[CMMRSY] A. Chang, D. Mehrle, S. J. Miller, T. Reiter, J. Stahl and D. Yott, *Newman's Conjecture in Function Fields*, Journal of Number Theory **157** (2015), 154–169.

[Con] J. B. Conrey, *L-functions and Random Matrices*. Pages 331–352 in Mathematics unlimited — 2001 and beyond, Springer-Verlag, Berlin, 2001.

[CFZ1] J. B. Conrey, D. W. Farmer and M. R. Zirnbauer, *Autocorrelation of ratios of L-functions*, preprint. http://arxiv.org/abs/0711.0718

[CFZ2] J. B. Conrey, D. W. Farmer and M. R. Zirnbauer, *Howe pairs, supersymmetry, and ratios of random characteristic polynomials for the classical compact groups*, preprint. http://arxiv.org/abs/math-ph/0511024

[CGLMP] T. Corcoran, J. Gossels, V. Luo, S. J. Miller and J. Porfilio, *Pythagoras at the Bat*, book chapter in Social Networks and the Economics of Sports (edited by Panos M. Pardalos and Victor Zamaraev), Springer-Verlag, 2014, pages 89–114.

[CLM] V. Cuff, A. Lewis and S. J. Miller, *The Weibull distribution and Benford's law*, to appear in Involve.

[DDKMMV1] P. Demontigny, T. Do, A. Kulkarni, S. J. Miller, D. Moon and U. Varma, *Generalizing Zeckendorf's Theorem to f-decompositions*, Journal of Number Theory **141** (2014), 136–158.

[DDKMMV2] P. Demontigny, T. Do, A. Kulkarni, S. J. Miller, D. Moon and U. Varma, *A Generalization of Fibonacci Far-Difference Representations and Gaussian Behavior* (joint with Philippe Demontigny, Thao Do, Archit Kulkarni and Umang Varma), Fibonacci Quarterly **52** (2014), no. 3, 247–273.

[DKMMW] T. Do, A. Kulkarni, S. J. Miller, D. Moon and J. Wellens, *Sums and Differences of Correlated Random Sets*, Journal of Number Theory **147** (2015), 44–68.

[DKMMWW] T. Do, A. Kulkarni, S. J. Miller, D. Moon, J. Wellens and J. Wilcox, *Sets Characterized by Missing Sums and Differences in Dilating Polytopes* (with Thao Do, Archit Kulkarni, David Moon and Jake Wellens), Journal of Number Theory **157** (2015), 123–153.

[DHKMS1] E. Dueñez, D. K. Huynh, J. P. Keating, S. J. Miller and N. Snaith, *The lowest eigenvalue of Jacobi Random Matrix Ensembles and Painlevé VI*, Journal of Physics A: Mathematical and Theoretical **43** (2010) 405204 (27pp).

[DHKMS2] E. Dueñez, D. K. Huynh, J. P. Keating, S. J. Miller and N. Snaith, *Models for zeros at the central point in families of elliptic curves* (with Eduardo Dueñez, Duc Khiem Huynh, Jon Keating and Nina Snaith), J. Phys. A: Math. Theor. **45** (2012) 115207 (32pp).

[DM1] E. Dueñez and S. J. Miller, *The low lying zeros of a $GL(4)$ and a $GL(6)$ family of L-functions*, Compositio Mathematica **142** (2006), no. 6, 1403–1425.

[DM2] E. Dueñez and S. J. Miller, *The effect of convolving families of L-functions on the underlying group symmetries*, Proceedings of the London Mathematical Society, 2009; doi: 10.1112/plms/pdp018.

[Du] T.M. Dunster, *Bessel Functions of Purely Imaginary Order, with an Application to Second-Order Linear Differential Equations Having a Large Parameter*, SIAM Journal of Mathematical Analysis, **21** (1990), no. 4.

[FiMi] F. W. K. Firk and S. J. Miller, *Nuclei, Primes and the Random Matrix Connection*, invited submission to Symmetry **1** (2009), 64–105; doi:10.3390/sym1010064.

[FiMil] D. Fiorilli and S. J. Miller, *Surpassing the Ratios Conjecture in the 1-level density of Dirichlet L-functions*, Algebra & Number Theory **Vol. 9** (2015), No. 1, 13–52.

[FrMil] J. Freeman and S. J. Miller, *Determining Optimal Test Functions for Bounding the Average Rank in Families of L-Functions*, to appear in proceedings for Ram Murty's 60th birthday, Contemporary Math. series of AMS (jointly with CRM).

[Gao] P. Gao, *N-level density of the low-lying zeros of quadratic Dirichlet L-functions*, Ph. D thesis, University of Michigan, 2005.

[GJMMNPP] J. Goes, S. Jackson, S. J. Miller, D. Montague, K. Ninsuwan, R. Peckner and T. Pham, *A unitary test of the L-functions Ratios Conjecture*, Journal of Number Theory **130** (2010), no. 10, 2238–2258.

[GoMi] J. Goes and S. J. Miller, *Towards an 'average' version of the Birch and Swinnerton-Dyer Conjecture*, Journal of Number Theory **130** (2010), no. 10, 2341–2358.

[GKMN] L. Goldmakher, C. Khoury, S. J. Miller and K. Ninsuwan, *The expected eigenvalue distribution of large, weighted d-regular graphs*, Random Matrices: Theory and Applications **3** (2014), no. 4, 1450015 (22 pages).

[GV] D. A. Goldston and R. C. Vaughan, *On the Montgomery-Hooley asymptotic formula*, Sieve methods, exponential sums and their applications in number theory (ed. G. R. H. Greaves, G. Harman and M. N. Huxley, Cambridge University Press, 1996), 117–142.

[GK] L. Guth and N. H. Katz, *On the Erdős distinct distances problem in the plane*, Annals of Mathematics **181** (2015), 155–190.

[HJM] C. N. B. Hammond, Warren P. Johnson and S. J. Miller, *The James Function*, Mathematics Magazine **88** (2015) 54–71.

[HaMi] C. Hammond and S. J. Miller, *Distribution of eigenvalues for the ensemble of real symmetric Toeplitz matrices*, Journal of Theoretical Probability **18** (2005), no. 3, 537–566.

[HeMi] P. Hegarty and S. J. Miller, *When almost all sets are difference dominated*, Random Structures and Algorithms **35** (2009), no. 1, 118–136.

[HuMi] C. Hughes and S. J. Miller, *Low lying zeros of L–functions with orthogonal symmetry*, Duke Mathematical Journal **136** (2007), no. 1, 115–172.

[HH] T. Hulse and J. Hoffstein, *Multiple Dirichlet Series and Shifted Convolutions, joint with Jeffrey Hoffstein*, preprint (2015).

[HY] C. P. Hughes and M. P. Young, *The twisted fourth moment of the Riemann zeta function*, J. Reine Angew. Math. **641** (2010), 203–236.

[HMM] D. K. Huynh, S. J. Miller and R. Morrison, *An elliptic curve family test of the Ratios Conjecture*, Journal of Number Theory **131** (2011), 1117–1147.

[HR] C. Hughes and Z. Rudnick, *Linear Statistics of Low-Lying Zeros of L-functions*, Quart. J. Math. Oxford **54** (2003), 309–333.

[IMS] J. Iafrate, S. J. Miller and F. Strauch, *Equipartitions and a distribution for numbers: A statistical model for Benford's law*, Physical Review E **91** (2015), no. 6, 062138 (6 pages).

[ILS] H. Iwaniec, W. Luo and P. Sarnak, *Low lying zeros of families of L-functions*, Inst. Hautes Études Sci. Publ. Math. **91** (2000), 55–131.

[ILMZ1] G. Iyer, O. Lazarev, S. J. Miller and L. Zhang, *Generalized More Sums Than Differences Sets*, Journal of Number Theory **132** (2012), no. 5, 1054–1073.

[ILMZ2] G. Iyer, O. Lazarev, S. J. Miller and L. Zhang *Finding and Counting MSTD sets*, Combinatorial and Additive Number Theory, CANT 2011 and 2012 (Melvyn B. Nathanson, editor), Springer Proceedings in Mathematics & Statistics (2014), 79–98.

[JMP] S. Jackson, S. J. Miller and V. Pham, *Distribution of eigenvalues for highly palindromic real symmetric Toeplitz matrices*, Journal of Theoretical Probability **25** (2012), 464–495.

[JKKKM] D. Jang, J. U. Kang, A. Kruckman, J. Kudo and S. J. Miller, *Chains of distributions, hierarchical Bayesian models and Benford's Law*, Journal of Algebra, Number Theory: Advances and Applications, volume 1, number 1 (March 2009), 37–60.

[KS1] N. Katz and P. Sarnak, Random Matrices, Frobenius Eigenvalues and Monodromy, AMS Colloquium Publications, AMS, Providence, RI 1999.

[KS2] N. Katz and P. Sarnak, *Zeros of zeta functions and symmetries*, Bull. AMS, vol. 36 (1999), 1–26.

[KeSn] J. P. Keating and N. C. Snaith, *Random matrices and L-functions*, Random matrix theory, J. Phys. A **36** (2003), no. 12, 2859–2881.

[KKMSX] M. Kologlu, G. S. Kopp, S. J. Miller, F. Strauch and W. Xiong, *The Limiting Spectral Measure for Ensembles of Symmetric Block Circulant Matrices*, to appear in the Journal of Theoretical Probability.

[KKMW] M. Kologlu, G. S. Kopp, S. J. Miller and Y. Wang, *On the number of summands in Zeckendorf decompositions*, Fibonacci Quarterly **49** (2011), no. 2, 116–130.

[Lem] D. S. Lemons, *On the number of things and the distribution of first digits*, Am. J. Phys. **54** (1986), no. 9, 816–817.

[LMil] V. Luo and S. J. Miller, *Relieving and Readjusting Pythagoras*, By The Numbers – The Newsletter of the SABR Statistical Analysis Committee **25** (2015), no. 1, 5–14.

[MMRW] B. Mackall, S. J. Miller, C. Rapti and K. Winsor, *Lower-Order Biases in Elliptic Curve Fourier Coefficients in Families*, to appear in the Conference Proceedings of the Workshop on Frobenius distributions of curves at CIRM in February 2014.

[MMRT-BW] B. Mackall, S. J. Miller, C. Rapti, C. Turnage-Butterbaugh and K. Winsor, *Some Results in the Theory of Low-lying Zeros* (and an appendix with Megumi Asada, Eva Fourakis, Kevin Yang), in Families of automorphic forms and the trace formula (editors), Simons Symposia series, Springer-Verlag.

[MO] G. Martin and K. O'Bryant, *Many sets have more sums than differences*, Additive combinatorics, 287–305, CRM Proc. Lecture Notes **43**, Amer. Math. Soc., Providence, RI, 2007.

[MMS] A. Massey, S. J. Miller and J. Sinsheimer, *Distribution of eigenvalues of real symmetric palindromic Toeplitz matrices and circulant matrices*, Journal of Theoretical Probability **20** (2007), no. 3, 637–662.

[McK] B. McKay, *The expected eigenvalue distribution of a large regular graph*, Linear Algebra Appl. **40** (1981), 203–216.

[Meh] M. Mehta, Random Matrices, 2nd edition, Academic Press, Boston, 1991.

[Mil1] S. J. Miller, 1*- and* 2*-level densities for families of elliptic curves: evidence for the underlying group symmetries*, Compositio Mathematica **140** (2004), 952–992.

[Mil2] S. J. Miller, *Investigations of zeros near the central point of elliptic curve L-functions* (with an appendix by E. Dueñez), Experimental Mathematics **15** (2006), no. 3, 257–279.

[Mil3] S. J. Miller, *Lower order terms in the 1-level density for families of holomorphic cuspidal newforms*, Acta Arithmetica **137** (2009), 51–98.

[Mil4] S. J. Miller, *The Pi Mu Epsilon 100th Anniversary Problems: Part I* (with James M. Andrews, Avery T. Carr and many of my students), The Pi Mu Epsilon Journal **13** (2013), no. 9, 513–534.

[Mil5] S. J. Miller, *The Pi Mu Epsilon 100th Anniversary Problems: Part II* (with James M. Andrews, Avery T. Carr and many of my students), The Pi Mu Epsilon Journal **13** (2014), no. 10, 577–608.

[Mil6] S. J. Miller, *The Pi Mu Epsilon 100th Anniversary Problems: Part III* (with many of my students), The Pi Mu Epsilon Journal **14** (2014), no. 1, 65–99.

[Mil7] S. J. Miller, *The Pi Mu Epsilon 100th Anniversary Problems: Part IV* (with many of my students), The Pi Mu Epsilon Journal **15** (2015), no. 2, 100–134.

[Mil8] *Theory and Applications of Benford's Law*, Princeton University Press, 2015, 438 pages.

[MN1] S. J. Miller and M. Nigrini, *The Modulo* 1 *Central Limit Theorem and Benford's Law for Products*, International Journal of Algebra **2** (2008), no. 3, 119–130.

[MN2] S. J. Miller and M. Nigrini, *Order statistics and Benford's law* (with Mark Nigrini), International Journal of Mathematics and Mathematical Sciences, Volume 2008 (2008), Article ID 382948, 19 pages. doi:10.1155/2008/382948

[MOS] S. J. Miller, B. Orosz and D. Schienerman, *Explicit constructions of infinite families of MSTD sets*, Journal of Number Theory **130** (2010) 1221–1233.

[MilPe] S. J. Miller and R. Peckner, *Low-lying zeros of number field L-functions*, Journal of Number Theory **132** (2012), 2866–2891.

[MPR] S. J. Miller, S. Pegado and L. Robinson , *Explicit Constructions of Large Families of Generalized More Sums Than Differences Sets*, Integers **12** (2012), #A30.

[MilSc] S. J. Miller and D. Scheinerman, *Explicit constructions of infinite families of MSTD sets*, Additive Number Theory: Festschrift In Honor of the Sixtieth Birthday of Melvyn B. Nathanson (David Chudnovsky and Gregory Chudnovsky, editors), Springer-Verlag, 2010. Also an appendix (with Peter Hegarty): *When almost all sets are not MSTD sets*.

[MSTW] S. J. Miller, K. Swanson, K. Tor and K. Winsor*Limiting Spectral Measures for Random Matrix Ensembles with a Polynomial Link Function*, Random Matrices: Theory and Applications **4** (2015), no. 2, 1550004 (28 pages).

[MilV] S. J. Miller and K. Vissuet, *Most Subsets are Balanced in Finite Groups*, Combinatorial and Additive Number Theory, CANT 2011 and 2012 (Melvyn B. Nathanson, editor), Springer Proceedings in Mathematics & Statistics (2014), 147–157.

[MW1] S. J. Miller and Y. Wang, *From Fibonacci numbers to Central Limit Type Theorems*, Journal of Combinatorial Theory, Series A **119** (2012), no. 7, 1398–1413.

[MW2] S. J. Miller and Y. Wang, *Gaussian Behavior in Generalized Zeckendorf Decompositions*, Combinatorial and Additive Number Theory, CANT 2011 and 2012 (Melvyn B. Nathanson, editor), Springer Proceedings in Mathematics & Statistics (2014), 159–173.

[Rou] D. Rouymi, *Formules de trace et non-annulation de fonction L autotrophes au niveau p^v*, Acta Arithmetica Vol. **147**, 2011, no. 1, 1–32.

[Rub] M. Rubinstein, *Low-lying zeros of L–functions and random matrix theory*, Duke Math. J. **109**, (2001), 147–181.

[ST] P. Sarnak and J. Tsimerman, *On Linnik and Selberg's Conjecture about Sums of Kloosterman Sums*, preprint.

[Wig1] E. Wigner, *On the statistical distribution of the widths and spacings of nuclear resonance levels*, Proc. Cambridge Philo. Soc., vol. 47 (1951), 790–798.

[Wig2] E. Wigner, *Statistical Properties of real symmetric matrices*. Pages 174–184 in Canadian Mathematical Congress Proceedings, University of Toronto Press, Toronto, 1957.

3 Summary of Thesis Results

Following Brumer-Heath-Brown [BH-B], Iwaniec-Luo-Sarnak [ILS], Katz-Sarnak [KaSa1, KaSa2], Rubinstein [Rub] and Silverman [Si], I used the 1- and 2-level densities to study the distribution of low-lying zeros for one-parameter rational families of elliptic curves of rank r over $\mathbb{Q}(T)$; these densities are defined by summing test functions at scaled zeros of the L-functions. Katz and Sarnak [KaSa1, KaSa2] predict that, to each family of L-functions $\mathcal{F}$, there is an associated symmetry group $G(\mathcal{F})$ (a classical compact group) which governs the distribution of the low-lying zeros. In other words, the behavior of zeros in a family of L-functions near the central point is well modeled by the behavior of eigenvalues near 1 of a classical compact group. For families of elliptic curves of rank 0 over $\mathbb{Q}(T)$, we expect $G(\mathcal{F})$ to be $SO(\text{even})$ if all curves have even functional equation, O if half are even, half odd, and $SO(\text{odd})$ if all are odd. If the family has rank r over $\mathbb{Q}(T)$, the densities are trivially modified to take into account the r expected zeros at the central point, and we will still refer to these as $O, SO(\text{even}), SO(\text{odd})$. The 1-level densities for the Unitary and Symplectic are distinguishable from the three Orthogonal groups for test functions of arbitrary small support; unfortunately, the three orthogonal groups all agree for functions supported in $(-1, 1)$. The 2-level densities for the orthogonal groups, however, are distinguishable for functions supported in arbitrarily small neighborhoods of the origin.

Modulo standard conjectures (which can be verified for many specific cases), for small support I showed the densities agree with Katz and Sarnak's predictions [Mil1]. The difficulty is that the logarithm of the analytic conductors must be extremely well controlled or oscillatory behavior drowns out the main term. In particular, it is not enough to confine the logarithms of the conductors to lie in $[\log N^d, \log 2N^d]$ as $N \to \infty$. The conductors are controlled by careful sieving and deriving explicit formulas via Tate's algorithm. Further, the densities confirm that the curves' L-functions behave in a manner consistent with having r zeros at the central point, as predicted by the Birch and Swinnerton-Dyer conjecture. By studying the 2-level densities of some constant sign families, we find the first examples of families of elliptic curves where we can distinguish $SO(\text{even})$ from $SO(\text{odd})$ symmetry.

Similar to the GUE universality Rudnick and Sarnak [RS] found in studying n-level correlations of L-functions, our universality follows from the sums of $a_t^2(p)$ in our families. For non-constant $j(T)$, this follows from a Sato-Tate law proved by Michel [Mic]; however, for many of our families we are able to show this by a direct calculation. The effect of the rank over $\mathbb{Q}(T)$ surfaces through sums of $a_t(p)$.

Finally, while the n-level densities for these families are universal, potential lower order correction terms have been observed in several families. These family dependent corrections are of size $1/\log N$; unfortunately, trivial estimation of the errors lead to terms of size $\log\log N/\log N$. I have recently completed a detailed analysis for many families [Mil5], where these corrections can be isolated. These corrections show how the arithmetic of the family enters as lower order corrections; in particular, families with and without complex multiplication behave differently. I am currently exploring the relation between these lower order terms and finer properties of the behavior of low zeros for small conductors.

4 Representative Subset of Previous Research

4.1 Previous Research.

Below is a representative sample of some of my older work.

(i) Elliptic curves and additive number theory

(a) While most of my work in elliptic curves is related to the behavior of zeros near the central point, with graduate students Scott Arms and Alvaro Lozano-Robledo [ALM] I also proved results on the geometric side. Specifically, generalizing an idea from my thesis, we gave a new, novel construction of one-parameter families of elliptic curves of moderate rank over $\mathbb{Q}(T)$. The key idea is to use Rosen and Silverman's [RSi] proof of Nagao's conjecture for rational surfaces to interpret averages of the Fourier coefficients as the rank. These averages are sums of Legendre sums; while these are impossible to evaluate in general, by careful construction these can be determined for special families. One interesting consequence is that families of moderate rank are constructed *without* having to enumerate rational solutions and showing that they are linearly independent (in other words, no height matrices are needed). The families constructed here are of great use in other investigations, as their Legendre sums are more tractable than the general family, and thus it is possible to isolate lower order correction terms (which depend on the arithmetic).

(b, c) Let A be a finite set of integers, and set $A+A$ and $A-A$ to be the set of all sums (respectively differences) of elements in A. As addition is commutative but subtraction is not, it was believed that 'most' of the time $|A+A| < |A-A|$. It thus came as a surprise when Martin and O'Bryant [MO] proved that if each element from $\{0,1,\dots,N\}$ is in A with positive probability p then as $N\to\infty$ a positive percentage of A are sum dominated (i.e., $|A+A| > |A-A|$). I and Peter Hegarty [HeMi] investigated the relationship between the sizes of the sum and difference sets attached to a subset of $\{0,1,...,N\}$ chosen randomly according to a binomial model with parameter $p(N)$, with $N^{-1} = o(p(N))$. We showed that the random subset is almost surely difference dominated, as $N\to\infty$, for any choice of $p(N)$ tending to zero, thus confirming a conjecture of Martin and O'Bryant. The proofs involve applying recent strong concentration results to our situation. It is worth noting that interesting transitional behavior is observed when $p(N)$ crosses the threshold of $N^{1/2}$. More precisely, the main result is

Theorem Let $p:\mathbb{N}\to(0,1)$ be any function such that $N^{-1} = o(p(N))$ and $p(N) = o(1)$. For each $N\in\mathbb{N}$ let A be a random subset of I_N chosen according to a binomial distribution with parameter $p(N)$. Then, as $N\to\infty$, the probability that A is difference dominated tends to one. More precisely, let $\mathscr{S},\mathscr{D}$ denote respectively the random variables $|A+A|$ and $|A-A|$. Then the following three situations arise:
(i) $p(N) = o(N^{-1/2})$: Then $\mathscr{S} \sim \frac{(N\cdot p(N))^2}{2}$ and $\mathscr{D}\sim 2\mathscr{S} \sim (N\cdot p(N))^2$.
(ii) $p(N) = c\cdot N^{-1/2}$ for some $c\in(0,\infty)$: Define the function $g:(0,\infty)\to(0,2)$ by $g(x) := 2\left(\frac{e^{-x}-(1-x)}{x}\right)$. Then $\mathscr{S}\sim g\left(\frac{c^2}{2}\right)N$ and $\mathscr{D}\sim g(c^2)N$.
(iii) $N^{-1/2} = o(p(N))$: Let $\mathscr{S}^c := (2N+1)-\mathscr{S}$, $\mathscr{D}^c := (2N+1)-\mathscr{D}$. Then $\mathscr{S}^c \sim 2\cdot\mathscr{D}^c \sim \frac{4}{p(N)^2}$.

In a related project, with my undergraduate student Dan Scheinerman I developed a new construction for families of more sum than difference sets; at the time of their construction, they were the densest known. The construction was generalized to more general linear combinations of A with Brooke Orvosz and David Newman, as well as two students in a number theory class (Sean Pegado and Luc Robinson) at Williams.

(d) A beautiful theorem of Zeckendorf states that every positive integer can be written uniquely as a sum of non-consecutive Fibonacci numbers. Once this has been shown, it is natural to ask how many Fibonacci numbers are needed. Lekkerkerker proved that the average number of such summands needed for integers in $[F_n, F_{n+1})$ is $n/(\phi^2+1)$, where ϕ is the golden mean. With summer students Murat Kologlu, Gene Kopp and Yinghui Wang [KKMW, MW], we showed that the fluctuation about the mean is Gaussian for the Fibonacci numbers. Unlike previous number-theoretic approaches, we attack the problem through combinatorics and generating functions. Our techniques can be generalized to apply to a large class of linear recurrence relations, leading to 10+ papers with students and junior faculty in the ensuing years.

(ii) L-functions and the Ratios Conjecture

Much of my previous research on L-functions splits naturally into two cases, determining the main term and determining lower order corrections for the 1-level density, defined by:

$$\lim_{N\to\infty} \frac{1}{|\mathcal{F}_N|} \sum_{f\in\mathcal{F}_N} \sum_j \phi\left(\widetilde{\gamma}_{j;f}\right).$$

Here ϕ is an even Schwartz function, $\mathcal{F} = \cup \mathcal{F}_N$ is a family of L-functions ordered by conductor, and (assuming GRH) we may write the zeros of $L(s,f)$ as $1/2 + i\gamma_{j;f}$ (the $\widetilde{\gamma}_{j;f}$ are the normalized imaginary parts of the non-trivial zeros). The Katz and Sarnak conjectures [KaSa1], [KaSa2] state that as the conductors tend to infinity, the behavior of the zeros near the central point agree with the scaling limit of normalized eigenvalues of a classical compact group.

(a) For families where the signs of the functional equations are all even and there is no corresponding family with odd functional equations, a "folklore" conjecture (for example, see page 2877 of [KeSn1]) states that the symmetry is symplectic, presumably based on the observation that SO(even) and SO(odd) symmetries in the examples known to date arise from splitting orthogonal families according to the sign of the functional equations. *A priori* the symmetry type of a family with all functional equations even is either symplectic or SO(even). Let ϕ be a fixed Maass form and let H_k denote the space of cusp forms of full level. All L-functions in the families (formed by Rankin-Selberg convolution) $\{\phi \times H_k\}$ and $\{\phi \times \mathrm{sym}^2 H_k\}$ (by $\mathrm{sym}^2 H_k$ we mean the set $\mathrm{sym}^2 f$ for $f \in H_k$) have even functional equations, and neither family seems to naturally arise from splitting sign within a full orthogonal family. By calculating the 1-level density I and Dueñez proved in [DM1] that the symmetry of the first is symplectic (as predicted); however, the second family has orthogonal symmetry (we cannot distinguish between SO(even), O and SO(odd) due to the small-support restriction on the allowable test functions; however, analyzing the 2-level density allowed us to discard O and SO(odd)). Thus our calculations are only consistent with the symmetry being SO(even). In particular, this work proved that the theory of low-lying zeros is more than just a theory of the distribution of signs of functional equations.

(b) Building on this work, I and Dueñez in [DM2] considered the Rankin-Selberg convolution of more general families. We define an NT-good family of L-functions to be a family where there is good control over the conductors, the cardinality of the family, and a good enough averaging formula to evaluate the needed prime sums to compute the 1-level density; known examples include Dirichlet L-functions, cuspidal newforms, as well as twists and symmetric powers of these. The main result relates how the zeros of compound families formed by Rankin-Selberg convolution are distributed in terms of how the constituent families are distributed. Explicitly,

Thereom. Let $\mathcal{F}$ and $\mathcal{G}$ be NT-good families of unitary automorphic cuspidal representations of $\mathrm{GL}_n(\mathbb{A}_\mathbb{Q})$ and $\mathrm{GL}_m(\mathbb{A}_\mathbb{Q})$ with trivial central character, with symmetry constants $c_\mathcal{F}$ and $c_\mathcal{G}$. Assume $\mathcal{F} \times \mathcal{G}$ is an NT-good family. Then the family $\mathcal{F} \times \mathcal{G}$ (which is the limit of $\mathcal{F}_N \times \mathcal{G}_M$, where N and M tend to infinity together) has symmetry constant $c_{\mathcal{F}\times\mathcal{G}} = c_\mathcal{F} \cdot c_\mathcal{G}$.

The proof follows from understanding the moments of the Satake parameters, with the main terms determined from the first and second moments. It is the universality of these moments that is responsible for the universality of the results. Numerous specific examples are given, in particular families of elliptic curves with rank. The difficulty there is in controlling the conductors.

(c) The last result involving the main term of the zeros near the central point is for the family of cuspidal newforms of prime level N split by sign of the functional equation. Iwaniec, Luo and Sarnak [ILS] showed that the main term of the 1-level density in these families agrees with random matrix theory when the Fourier transform of the test function is supported in $(-2, 2)$. This impressive calculation depends on a difficult analysis of the non-diagonal term in the Petersson formula, namely the Bessel-Kloosterman sums. With Chris Hughes [HuMi] generalized these results to the n-level density (which improves estimates of high order of vanishing), avoiding analyzing n-dimensional analogues of the Bessel-Kloosterman piece by cleverly changing variables. The cost of this switch was difficult combinatorics needed to show agreement with random matrix theory; this is because random matrix theory was expecting an n-dimensional integral whereas we changed test functions to a one-dimensional integral which can be evaluated by the techniques in [ILS]. (It is not the case that the combinatorics can always be analyzed; in Gao's thesis [Gao] the number theory computations are thought to agree with random matrix theory throughout his range, but this can only be shown for a restricted window.) This led to the discovery of a new formula for the n-level density, which after some combinatorics was shown to agree with the determinantal expansions of Katz-Sarnak [KaSa1, KaSa2]. This formula is significantly easier for comparison purposes in restricted ranges. We proved

Theorem Let $n \geq 2$, $\text{supp}(\widehat{\phi}) \subset (-\frac{1}{n-1}, \frac{1}{n-1})$, and define

$$R_n(\phi) \ = \ (-1)^{n-1}2^{n-1}\left[\int_{-\infty}^{\infty} \phi(x)^n \frac{\sin 2\pi x}{2\pi x}\,\mathrm{d}x - \frac{1}{2}\phi(0)^n\right], \quad \sigma_\phi^2 \ = \ 2\int_{-1}^{1} |y|\widehat{\phi}(y)^2\,\mathrm{d}y.$$

Assume GRH for $L(s, f)$ and for all Dirichlet L-functions. As $N \to \infty$ through the primes, the centered n^{th} moment for $H_k^{\pm}(N)$ (the family of weight k cuspidal newforms of level N and functional equation either even or odd) agrees with RMT and equals $(2m-1)!!\sigma_\phi^{2m} \pm R_{2m}(\phi)$ if $n = 2m$ is even and $\pm R_{2m+1}(\phi)$ if $n = 2m+1$ is odd. One application of these results (where it is essential that we are able to evaluate the relevant sums without using the Petersson weights) is to bound high vanishing in the family. Specifically, we prove the following (which for large r provides better bounds than the previous records):

Theorem: Consider the families of weight k cuspidal newforms split by sign, $H_k^{\pm}(N)$. Assume GRH for all Dirichlet L-functions and all $L(s, f)$. For each n there are constants r_n and c_n such that as $N \to \infty$ through the primes, for $r \geq r_n$ the probability of at least r zeros at the central point is at most $c_n r^{-n}$; equivalently, the probability of fewer than r zeros at the central point is at least $1 - c_n r^{-n}$.

(d) In addition to studying the main term, I've also investigated the arithmetic dependence of lower order terms in various families. These investigations are important as the main term is independent of the arithmetic. In [Mil3] I studied families of elliptic curves (using many of the families constructed in [ALM] and discussed earlier). By isolating the first correction term to the 1-level density, we csn see differences depending on whether or not the families had complex multiplication or what the torsion group was.

(e, f) Building on analogies with Random Matrix Theory and previous number theory conjectures, recently Conrey, Farmer and Zirnbauer [CFZ1], [CFZ2] created the L-functions Ratios Conjecture (or recipe) to predict sums of ratios of L-functions in a family. These quantities allow us to compute almost any desired statistic, from n-level correlations and densities to mollifiers and moments, to name a few. The predictions are believed to be correct to square-root cancelation in the family's cardinality. To put this in context, a typical family of one-parameter elliptic curves over $\mathbb{Q}(T)$ with T specialized to be in $[N, 2N]$ has an error term of size $O(\log\log N/\log N)$, much larger than $N^{-1/2+\epsilon}$! There is little basis for such a small conjectured error term, other than the philosophy of square-root cancelation. The recipe is as follows: let

$$R_{\mathcal{F}_N}(\alpha, \gamma) \ := \ \sum_{f \in \mathcal{F}_N} \frac{L(1/2+\alpha, f)}{L(1/2+\gamma, f)}.$$

1. Use the approximate functional equation to expand the numerator into two sums plus a remainder. The first sum is over m up to x and the second over n up to y, where xy is of the same size as the analytic conductor (typically one takes $x \sim y$). We ignore the remainder term.

2. Expand the denominator by using the generalized Mobius function.

3. Execute the sum over $\mathcal{F}_N$, keeping only main (diagonal) terms; however, before executing these sums replace any product over epsilon factors (arising from the signs of the functional equations) with the average value of the sign of the functional equation in the family.

4. Extend the m and n sums to infinity (i.e., complete the products).

5. Differentiate with respect to the parameters, and note that the size of the error term does not significantly change upon differentiating.

6. A contour integral involving $\frac{\partial}{\partial\alpha} R_{\mathcal{F}_N}(\alpha,\gamma)\Big|_{\alpha=\gamma=s}$ yields the 1-level density.

What is remarkable is that in many of the steps above we throw away error terms that are the same size as main terms, yet at the end we obtain (conjecturally) perfect agreement! I've studied families of quadratic characters, cuspidal newforms (not split by sign) and number field L-funtions; for suitably restricted test functions, we've [GJMMNPP, Mil1, Mil4, MilMo, MilPe] proved the Ratios' prediction (complete with square-root cancelation) is correct. For example, for quadratic characters the Ratios Conjecture predicts one of the lower order terms is

$$\frac{2}{\log X}\int_{-\infty}^{\infty} \phi(\tau)\frac{\zeta'}{\zeta}\left(1+\frac{4\pi i\tau}{\log X}\right) d\tau;$$

which implies that the lower order term depends on the zeros of the Riemann zeta function. This was numerically observed by Rubinstein, and the improvement in the fit by incorporating these terms was powerfully demonstrated by Stopple's computations. Additional applications of knowing these lower order terms is to finding $N_{\text{effective}}$, the optimal size matrices to model behavior of zeros at a finite conductor (and not in the limit).

(iii) Random Graphs and Random Matrix Ensembles

Let A be a real $N \times N$ symmetric matrix with eigenvalues $\lambda_i(A)$. We can form a measure by placing a mass of size $1/N$ at each (normalized) eigenvalue. If our ensemble is large enough, we can average over the family and a generic matrix will have behavior close to the system average. While these ensembles typically do not directly correspond to families of L-functions, they are nevertheless useful in building intuition as to how small sub-families can behave, as well as being interesting in their own right.

(a) Building on joint work with Chris Hammond [HaMi], I studied the distribution of eigenvalues of real symmetric palindromic Toeplitz matrices with undergraduates John Sinsheimer and Adam Massey, both of whom have continued to graduate school. Many authors had noticed that the density of normalized eigenvalues of real symmetric Toeplitz matrices was close to, but not equal to, the standard normal. In [HaMi] we interpreted the discrepancy in terms of Diophantine obstructions to systems of equations, and conjectured that forcing the first row to be a palindrome would remove these obstructions. We proved this in [MMS]. The difficulty in this project, as is frequently the case in studying random matrix ensembles, was developing the combinatorics to obtain closed form expressions for the moments. Similar to many other problems in the field, it is straightforward to show the averages of the moments of the measures converge; however, it is quite difficult to determine the precise value of the average moments. One reason this ensemble is so difficult to study is that it has of the order N degrees of freedom, far less than the full family of all real symmetric matrices (which is of order $N^2/2$). Through Cauchy's interlacing formula, there are explicit formulas for the eigenvalues of these matrices, and as one application we obtain a central limit theorem for weighted sums of random variables.

(b) Joint with undergraduates Tim Novikoff and Anthony Sabelli [MNS] (who are now graduate students in applied math at Cornell), I studied the distribution of the second largest eigenvalue in families of d-regular

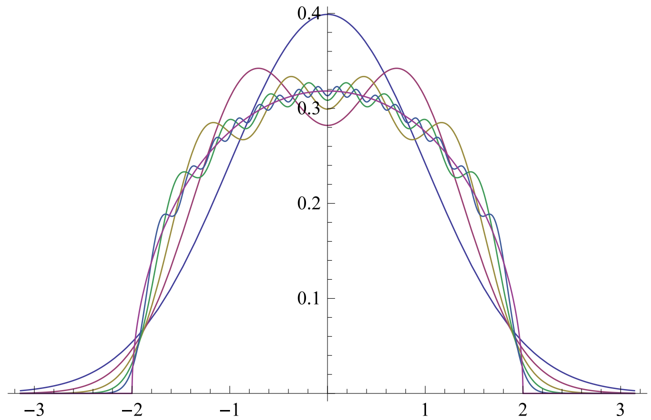

Figure 1: Plots for $f_1, f_2, f_4, f_8, f_{16}$ and the semi-circle density.

graphs. Recently Friedman [Fr] proved Alon's conjecture [Al] for many families of d-regular graphs, namely that given any $\epsilon > 0$ "most" graphs have their largest non-trivial eigenvalue at most $2\sqrt{d-1}+\epsilon$ in absolute value; if the absolute value of the largest non-trivial eigenvalue is at most $2\sqrt{d-1}$ then the graph is said to be Ramanujan. These graphs have important applications in communication network theory, allowing the construction of superconcentrators and nonblocking networks, coding theory and cryptography. As many of these applications depend on the size of the largest non-trivial positive and negative eigenvalues, it is natural to investigate their distributions. We showed these are well-modeled by the $\beta = 1$ Tracy-Widom distribution for several families. If the observed growth rates of the mean and standard deviation as a function of the number of vertices holds in the limit, then in the limit approximately 52% of d-regular graphs from bipartite families should be Ramanujan, and about 27% from non-bipartite families (assuming the largest positive and negative eigenvalues are independent). The key difficulty in interpreting the numerical investigations is that, appropriately normalized, the three Tracy-Widom distributions and the standard normal are very close to each other; the best test was looking at the percentage of eigenvalues to the right of the mean.

(c) Given an ensemble of $N \times N$ random matrices, the first question to ask is whether or not the empirical spectral measures of typical matrices converge to a limiting spectral measure as $N \to \infty$. While this has been proved in many thin patterned ensembles sitting inside all real symmetric matrices, frequently there is no nice closed form expression for the limiting measure. Further, current theorems provide few pictures of transitions between ensembles. Building on earlier work with my students [HaMi, MMS, JMP], I continued to explore highly patterned matrices this past summer with two REU students, Murat Kologlu and Gene Kopp [KKM]. We considered the ensemble of symmetric period m–circulant matrices with entries i.i.d.r.v. These matrices have toroidal diagonals periodic of period m. We view m as a "dial" we can "turn" from the highly structured symmetric circulant matrices, whose limiting eigenvalue density is a Gaussian, to the ensemble of all real symmetric matrices, whose limiting eigenvalue density is a semi-circle. The limiting eigenvalue densities f_m show a visually stunning convergence to the semi-circle as $m \to \infty$, which we prove. In contrast to most studies of patterned matrix ensembles, we find explicit closed form expressions for the densities. We prove that f_m is the product of a Gaussian and a certain even polynomial of degree $2m-2$. The proof is by derivation of the moments from the eigenvalue trace formula. The new feature, which allows us to obtain closed form expressions, is converting the central combinatorial problem in the moment calculation into an equivalent counting problem in algebraic topology. The explicit formula is then obtained using topology (especially Euler characteristic), generating functions, and complex analysis. I explored the effect of changing the structure on the answer with one of my thesis students, Wentao Xiong, who was also being mentored by Murat and Gene. See Figure 1.

(d) We introduced a new family of $N \times N$ random real symmetric matrix ensembles, the k-checkerboard

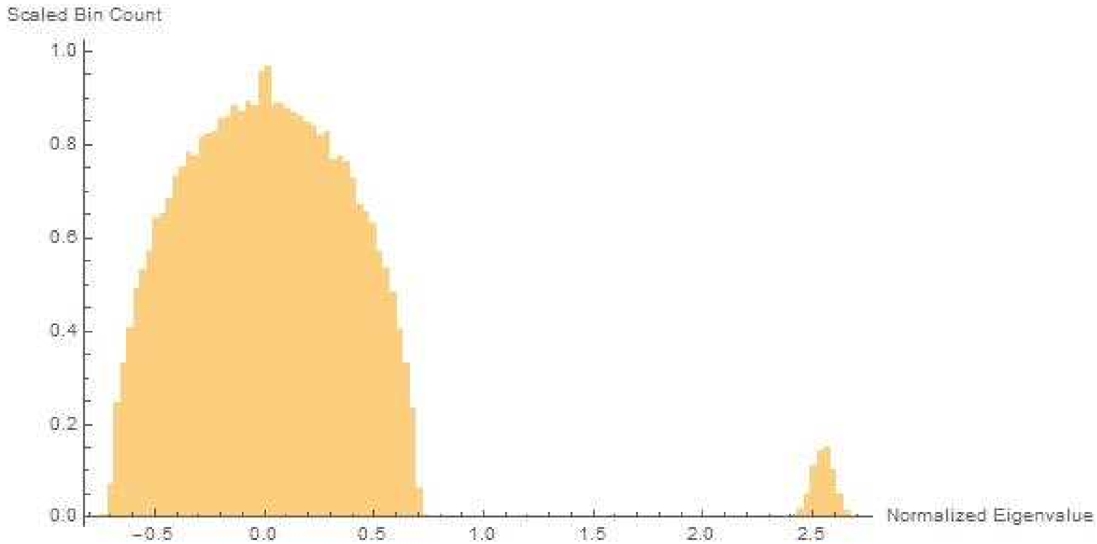

Figure 2: A histogram, normalized appropriately to achieve unit mass, of the scaled eigenvalue distribution for 100×100 2-checkerboard real matrices with $w = 1$ after 500 trials.

matrices. Fix $D = \mathbb{R}, \mathbb{C}$ or $\mathbb{H}$, $k \in \mathbb{N}$, $w \in \mathbb{R}$. Then the $N \times N$ (k, w)-checkerboard ensemble over D is the ensemble of matrices $M = (m_{ij})$ given by

$$m_{ij} \;=\; \begin{cases} a_{ij} & \text{if } i \not\equiv j \bmod k \\ w & \text{if } i \equiv j \bmod k \end{cases}$$

where $a_{ij} = \overline{a_{ji}}$ and

$$a_{ij} \;=\; \begin{cases} r_{ij} & \text{if } D = \mathbb{R} \\ \frac{r_{ij}+b_{ij}\hat{i}}{\sqrt{2}} & \text{if } D = \mathbb{C} \\ \frac{r_{ij}+b_{ij}\hat{i}+c_{ij}\hat{j}+d_{ij}\hat{k}}{2} & \text{if } D = \mathbb{H} \end{cases}$$

with r_{ij}, b_{ij}, c_{ij}, and d_{ij} i.i.d. random variables with mean 0, variance 1, and finite higher moments, and the probability measure on the ensemble given by the natural product probability measure. We refer to the $(k, 1)$-checkerboard ensemble over D simply as the k-checkerboard ensemble over D. The limiting spectral measure has two components which can be determined explicitly; see Figure 2. All but k eigenvalues are in the bulk, and their behavior, appropriately normalized, converges to the semi-circle as $N \to \infty$; the remaining k are tightly constrained near N/k and their distribution converges to the $k \times k$ hollow GOE ensemble (this is the density arising by modifying the GOE ensemble by forcing all entries on the main diagonal to be zero). Similar results hold for complex and quaternionic analogues. We are able to isolate each regime separately through appropriate choices of weight functions for the eigenvalues and then an analysis of the resulting combinatorics.

(e) Continuing the work I did with Chris Hughes, one of my thesis students (Jake Levinson) and I extended that analysis to derive alternatives to the determinantal formuls of Katz and Sarnak for the n-level densities for larger support than done in [HuMi]. The motivation for this is Peng Gao's thesis [Gao]. Gao was able to determine the number theory for the n-level density for families of Dirichlet L-functions, unfortunately, due to the difficulty of the combinatorics, he was not able to show that his answer always agrees with the random matrix theory prediction (though we do believe they always agree). We derived more tractable formulas for these expansions to facilitate these comparisons, and showed agreement for $n \leq 7$, reducing the general case to a Fourier-combinatorial identity.

(f) I completed investigations begun with Leo Goldmakher years ago at AIM and extended with undergraduate Eve Ninsuwan at Williams on the density of eigenvalues of weighted d-regular graphs. McKay

[McK] computed the density of normalized eigenvalues (it is different from the semi-circle seen for the GOE ensemble, but does converge to the semi-circle as $d \to \infty$). Goldmakher and I noticed that if the adjacency matrix elements are weighted by multiplying by independent identically distributed random variables, then if the resulting distribution is to be the semi-circle then the first 9 moments of the weighting distribution must agree with the semi-circleŠs distribution. This suggests that the semi-circle is a fixed point of this weighting procedure. The combinatorics become quite involved, and interestingly for the eight and higher moments there is a difference of order $1/d^2$ with the momenets of the eigendensity and the semi-circle. We found a formula for the eigendensity in terms of new combinatorial objects.

(iv) Benford's law

Many mathematical and natural phenomena satisfy Benford's law (or a close approximation), where the probability of a leading digit of d is $\log_{10}(1 + \frac{1}{d})$. While the proofs are typically related to equidistribution theory, and thus fall in the province of standard number theoretic investigations, the numerous applications (especially for fraud detection and data integrity tests) ensure that the subject is of interest to many.

(a) Funded in part by NSF Grant DMS0753043, I helped organize the *Conference on the Theory and Applications of Benford's Law* in 2007, and edited an introductory book on the subject with the participants and other world experts. It was the first conference on the subject, and participants came from mathematics, statistics, biology, engineering, computer science, accounting and industry, to name a few. The discussions there have led to many projects.

(b, c) One of the most important questions is why Benford's law is so prevalent, and how rapidly it sets in (such estimates are essential in prosecuting fraud). In [JKKKM, MN1] we answer these questions for data that is the product of a growing number of independent random variables. The key ingredients are Fourier analysis and the Mellin transform, which allow us to prove a version of the Central Limit Theorem for sums of independent random variables modulo 1 with (at times) exponentially decaying error term. For example, if we consider a product of n independent uniform random variables on $[0, k]$, the difference between the cumulative distribution function here and that of Benford's law is bounded by

$$\frac{k}{s}\frac{(\log k)^{n-1}}{\Gamma(n)} + \left(\frac{1}{2.9^n} + \frac{\zeta(n)-1}{2.7^n}\right) 2\log_{10} s,$$

where $\zeta(n)$ is the Riemann zeta function and $\log_{10} s$ is the probability of a Benford random variable having mantissa of s or less. These results have been converted to new tests for fraud by Nigrini and Miller [MN2], [NM2], which have been shared at an invited address at the Boston headquarters of the IRS.

(v) Applied Projects

(a: Incomplete Exponential Sums) Exponential sums have a rich history, and estimates of their size have numerous applications, ranging from uniform distribution to solutions to Diophantine equations to L-functions to the Circle Method, to name a few. Consider the following incomplete exponential sum:

$$S(f, n, q) \ = \ \sum_{x_1=\pm 1} \cdots \sum_{x_n = \pm 1} x_1 \cdots x_n e^{2\pi i f(x_1,\dots,x_n)/q},$$

with f a non-homogenous quadratic. Proving non-trivial exponentially decreasing upper bounds for $S(f, n, m)$ will provide insight into the computational complexity of a class of boolean circuits needed to compute the parity of n binary inputs. A theorem that shows that the norm of $S(f, n, m)$ is c^n, where $c < 1$ will show that the size (number of binary gates) of these circuits computing parity has to grow exponentially fast. These lower bound results are of great interest to the theoretical computer science community. Using Ramsey-theoretic techniques, Alon and Beigel [AB] proved that for each fixed n, d and m there exists a positive constant $b_{d,m,n}$ such that $|S(f, n, m)| \ < \ b_{d,m,n}$ and $\lim_{n\to\infty} b_{d,m,n} = 0$; note the resulting sequences converge very slowly to 0. In terms of computational complexity, this only tells us that the minimum circuit size required to compute parity of n bits tends to infinity with n. It is of far more interest, from the computational point of view, to show exponentially fast growth in minimum circuit size. This is generally interpreted as showing that parity circuits of the required kind cannot feasibly be built. We have currently obtained sharp bounds on average, and have solved the problem for n small; this project is joint with Eduardo Dueñez and Amitabha Roy and Howard Straubing [DMRS].

(b: Empirical Bayes Inference in the Multinomial Logit Model) Whether it's the 20,000+ hits based on a www.google.com search or the 1000+ hits on www.jstor.org, the multinomial logit (MNL) model plays a very prominent role in many literatures as a basis for probabilistic inferences. One of the recent advances regarding the MNL model is the ability to incorporate heterogeneity into the response coefficients; unfortunately, this leads to increased numerical computation. Once one combines the MNL kernel, a Bernoulli random variable with logit link function, with a heterogeneity distribution, closed-form inference is unavailable due to the non-conjugacy of the product Bernoulli likelihood and the heterogeneity distribution (prior).

Eric Bradlow (professor of marketing and statistics, University of Pennsylvania). Kevin Dayaratna (graduate student in marketing at the University of Maryland) and I [MBD] derived a closed-form solution to the heterogeneous MNL problem; unfortunately, the closed-form expansion requires too many terms to be computationally feasible at present. We reduce the number of computations by several orders of magnitude by rewriting the expansion in terms of the number of solutions to systems of Diophantine equations. This allows us to have one very long initial calculation, with all subsequent calculations involving only 10 or 20 terms (instead of 10^8 and higher), and is now applicable for some problems.

(c: Binary Integer Linear Programming) With Joshua Eliashberg (professor, Wharton), Sanjeev Swami (professor, India Institute of Technology Kanpur) Chuck Weinberg (professor, University of British Columbia) and Berend Wierenga (professor, Erasmus University Rotterdam), I solved a binary integer linear programming problem to allow movie theaters to optimally schedule movies each day, taking into account a variety of managerial constraints in reel time. We are currently expanding our model to include additional constraints, and it is being implemented at movie theaters in Amsterdam. This work [EHHHMSWW] won the best paper award for 2009 for the International J. of Research in Marketing.

(d: Dynamical Systems) Leo Kontorovich (postdoc, Weizmann Institute), Amitabha Roy and I are studied various models for the propagation of viruses in different systems. We are primarily concerned with how infections are transmitted in various networks. For certain configurations we have derived a differential equation whose fixed points answer the problem. Numerical and theoretical investigations suggested what the limiting behavior should be. We developed techniques to analyze the resulting equations and proved our conjecture on the limiting behavior in most cases.

(e: Sabermetrics) Sabermetrics is the application of mathematical tools to analyze baseball. The field really took off with Bill James' (who later helped build the Red Sox championship teams of '04 and '08) work in the 70's and 80's, and has since become an major force in the baseball industry [Le]. The subject poses numerous questions of both theoretical and practical interest, and it is often highly non-trivial to derive a mathematically tractable model for a baseball event which captures the essential features. It has been noted that in many professional sports leagues a good predictor of a team's end of season won-loss percentage is Bill James' Pythagorean Formula ${\rm RS_{obs}}^\gamma/({\rm RS_{obs}}^\gamma + {\rm RA_{obs}}^\gamma)$, where ${\rm RS_{obs}}$ (resp. ${\rm RA_{obs}}$) is the observed average number of runs scored (allowed) per game and γ is a constant for the league; for baseball the best agreement is when γ is about 1.82. This formula is often used in the middle of a season to determine if a team is performing above or below expectations, and estimate their future standings (the principles involved, however, have enormous application, as the question could just as easily be asked about which mutual funds or stocks are over- or underperforming, and thus determine when to buy or sell). In [Mil4] I showed how this formula is a consequence of a reasonable model of a baseball game. I have presented this result at numerous conferences, discussed my model with Bill James at the Boston Red Sox, improved the model with a thesis student, and studied many similar problems.

(vi) Zeros of Elliptic Curve L-functions

My main ongoing research project involves the observed repulsion of zeros near the central point by zeros at the central point. By the Birch and Swinnerton-Dyer Conjecture, if an elliptic curve has geometric rank r, its L-function should vanish to order r at the central point, and these curves offer an exciting laboratory to test the conjectures of Random Matrix Theory. I have introduced two 'natural' models for the random matrix analogues into the literature, what I call the independent model (where the forced zeros do not interact with the remaining zeros) and the interaction model (where they do) [Mil3]. In my thesis I proved that as the conductors tend to infinity the distribution of zeros agrees with the independent model (which is the same as the interaction model with no forced zeros). The interaction model is related to the classical Bessel kernels of RMT, and gives a very different prediction as to the behavior of the first few zeros when there are forced zeros. I and my students (at Princeton, AIM and Ohio State) wrote code to construct large numbers of elliptic curves and study the effect on the location of the first zero above the central point. Extensive calculations and theoretical modeling were done with Eduardo Dueñez, Jon Keating and Nina Snaith (professors, University of Bristol) and their student Duc Khiem Huynh (in fact, work related to this become Duc Khiem's thesis). Unlike the excess rank investigations, however, as we increase the conductor we see a marked change in data; specifically, the repulsion decreases. The detailed numerical investigations I have run [Mil3] provided several clues which helped us determine the correct non-limiting behavior. In particular, we observe that the repulsion increases with rank, decreases with the size of the conductor, and all the zeros are shifted by the same amount. This suggests the right model for *finite* conductors is the interaction model, with parameters a function of the average rank for conductors of a given size. As excess rank has long been observed for finite conductors, this explains how we can have repulsion in rank 0 families over $\mathbb{Q}(T)$ (where there are no zeros at the central point to repel other zeros!).

Amazingly, all the zeros appear to be repelled equally. There is a random matrix ensemble (Jacobi ensembles) with very similar properties, namely there is a variable parameter (which may be thought of as corresponding to the number of zeros at the central point), and as that parameter is increased the remaining zeros are all equally repelled. We call this the interaction model, and expect it to model the low-lying zeros for small conductors. We need two pieces of information for the analysis. The first is related to discretizing the random matrix ensemble. The second is the matrix size. This corresponds to the problem of finding $N_{\text{effective}}$ for modeling zeros of $\zeta(s)$ at finite heights, done by Keating-Snaith and others. We are constructing a rigorous theory of these low-lying zeros, especially the first zero above the central point. There are numerous technical difficulties, ranging from the fact that values of elliptic curve L-functions are discretized at the central point (with the discretization depending on arithmetic) to the difficulty in determining the lower-order terms in the one-level density to determine $N_{\text{effective}}$. We are using the Ratios Conjecture to predict these terms, and then using that as a guide for the number theory computations. Much progress has been made (including writing code to solve a Painleve VI equation that arises), and our prediction does an outstanding job of describing the data for the family of quadratic twists (*see Figure 3*). In particular, incorporating discretization and the lower order arithmetic terms captures the repulsion. I am extending our theory to general families of L-functions.

Using these observations, we found the correct model for finite conductors. This is similar to Keating and Snaith's observation [KeSn1, KeSn2] that zeros of L-functions at height T should not be modeled by the infinite scaling limits of matrices, but by $N \times N$ matrices with $N \approx \log T$. To date we have written one paper on the algorithms needed to solve related Painleve VI differential equations [DHKMS], and one on the number theory.

The references below are from my thesis and previous research

References

[Al] N. Alon, *Eigenvalues and expanders*, Combinatorica **6** (1986), no. 2, 83–96.

[AB] N. Alon and R. Beigel, *Lower bounds for approximations by low degree polynomials over* $\mathbb{Z}_m$. In *Sixteenth Annual IEEE Conference on Computational Complexity*, IEEE Computer

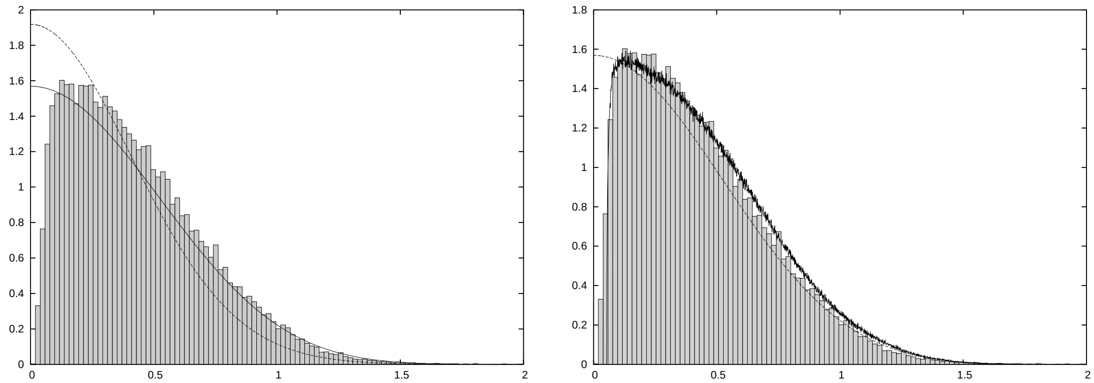

Figure 3: (left) Distribution of lowest zeros for $L_{E_{11}}(s, \chi_d)$ with $0 < d < 400,000$ (bar chart), distribution of lowest eigenvalue of SO(2N) with N_{eff} (solid), standard N_0 (dashed). (right) Distribution of lowest zeros for $L_{E_{11}}(s, \chi_d)$ with $0 < d < 400,000$ (bar chart), distribution of lowest eigenvalue of SO(2N) effective N of N_{eff} = 2 (solid) with discretisation, distribution of lowest eigenvalue of SO(2N) effective N of $N_{\mathrm{eff}} = 2.32$ (dashed) without discretisation.

Society Press (2001), 184-187.

[ALM] S. Arms, A. Lozano-Robledo and S. J. Miller, *Constructing One-Parameter Families of Elliptic Curves over $\mathbb{Q}(T)$ with Moderate Rank*, Journal of Number Theory **123** (2007), no. 2, 388–402.

[Ben] F. Benford, *The law of anomalous numbers*, Proceedings of the American Philosophical Society **78** (1938), 551-572.

[BBH] A. Berger, Leonid A. Bunimovich and T. Hill, *One-dimensional dynamical systems and Benford's Law*, Trans. Amer. Math. Soc. **357** (2005), no. 1, 197–219.

[BH] A. Berger and T. Hill, *Newton's method obeys Benford's law*, The Amer. Math. Monthly **114** (2007), no. 7, 588-601.

[Be] M. V. Berry, *Semiclassical formula for the number variance of the Riemann zeros*, Nonlinearity **1** (1988), 399–407.

[BeKe] M. V. Berry and J. P. Keating, *The Riemann zeros and eigenvalue asymptotics*, Siam Review **41** (1999), no. 2, 236–266.

[BBLM] E. Bogomolny, O. Bohigas, P. Leboeuf and A. G. Monastra, *On the spacing distribution of the Riemann zeros: corrections to the asymptotic result*, Journal of Physics A: Mathematical and General **39** (2006), no. 34, 10743–10754.

[BoKe] E. B. Bogomolny and J. P. Keating, *Gutzwiller's trace formula and spectral statistics: beyond the diagonal approximation*, Phys. Rev. Lett. **77** (1996), no. 8, 1472–1475.

[BrDu] J. Brown and R. Duncan, *Modulo one uniform distribution of the sequence of logarithms of certain recursive sequences*, Fibonacci Quarterly **8** (1970) 482–486.

[BH-B] A. Brumer and R. Heath-Brown, *The average rank of elliptic curves V*, preprint.

[CFKRS] B. Conrey, D. Farmer, P. Keating, M. Rubinstein and N. Snaith, *Integral moments of L-functions*, Proc. London Math. Soc. (3) **91** (2005), no. 1, 33–104.

[CFZ1] J. B. Conrey, D. W. Farmer and M. R. Zirnbauer, *Autocorrelation of ratios of L-functions*, preprint. http://arxiv.org/abs/0711.0718

[CFZ2] J. B. Conrey, D. W. Farmer and M. R. Zirnbauer, *Howe pairs, supersymmetry, and ratios of random characteristic polynomials for the classical compact groups*, preprint. http://arxiv.org/abs/math-ph/0511024

[CS] J. B. Conrey and N. C. Snaith, *Applications of the L-functions Ratios Conjecture*, to appear in the Proc. London Math. Soc. http://arxiv.org/abs/math/0509480

[Dia] P. Diaconis, *The distribution of leading digits and uniform distribution mod 1*, Ann. Probab. **5** (1979), 72–81.

[DHKMS] E. Dueñez, D. K. Huynh, J. Keating, S. J. Miller and N. Snaith), *The lowest eigenvalue of Jacobi Random Matrix Ensembles and Painlevé VI*, Journal of Physics A: Mathematical and Theoretical **43** (2010) 405204 (27pp).

[DM1] E. Dueñez and S. J. Miller, *The low lying zeros of a $GL(4)$ and a $GL(6)$ family of L-functions*, Compositio Mathematica **142** (2006), no. 6, 1403–1425.

[DM2] E. Dueñez and S. J. Miller, *The effect of convolving families of L-functions on the underlying group symmetries*, Proceedings of the London Mathematical Society, 2009; doi: 10.1112/plms/pdp018.

[DMRS] E. Dueñez, S. J. Miller, A. Roy and H. Straubing, *Incomplete quadratic exponential sums in several variables*, Journal of Number Theory **116** (2006), no. 1, 168–199.

[EHHHMSWW] J. Eliashberg, Q. Hegie, J. Ho, D. Huisman, S. J. Miller, S. Swami, C. B. Weinberg and B. Wierenga, *Silver Scheduler: a demand-driven modeling approach for the construction of micro-schedules of movies in a multiplex*, Intern. J. of Research in Marketing (2009), doi:10.1016/j.ijresmar.2008.09.004 (lead article, and won the IJRM Best Paper Award for 2009).

[FI] E. Fouvry and H. Iwaniec, *Low-lying zeros of dihedral L-functions*, Duke Math. J. **116** (2003), no. 2, 189-217.

[Fr] J. Friedman, *A proof of Alon's second eigenvalue conjecture*, Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing, 720–724 (electronic), ACM, New York, 2003.

[Gao] P. Gao, *N-level density of the low-lying zeros of quadratic Dirichlet L-functions*, Ph. D thesis, University of Michigan, 2005.

[GJMMNPP] (with J. Goes, S. Jackson, S. J. Miller, D. Montague, K. Ninsuwan, R. Peckner and T. Pham, *A unitary test of the L-functions Ratios Conjecture*, Journal of Number Theory **130** (2010), no. 10, 2238–2258.

[GV] D. A. Goldston and R. C. Vaughan, *On the Montgomery-Hooley asymptotic formula*, Sieve methods, exponential sums and their applications in number theory (ed. G. R. H. Greaves, G. Harman and M. N. Huxley, Cambridge University Press, 1996), 117-142.

[Gü] A. Güloğlu, *Low-Lying Zeros of Symmetric Power L-Functions*, Internat. Math. Res. Notices 2005, no. 9, 517-550.

[Ha] R. W. Hamming, *On the distribution of numbers*, Bell Syst. Tech. J. **49** (1970), 1609-1625.

[HaMi] C. Hammond and S. J. Miller, *Distribution of eigenvalues for the ensemble of real symmetric Toeplitz matrices*, Journal of Theoretical Probability **18** (2005), no. 3, 537–566.

[HeMi] P. Hegarty and S. J. Miller, *When almost all sets are difference dominated*, Random Structures and Algorithms **35** (2009), no. 1, 118–136.

[Hi1] T. Hill, *The first-digit phenomenon*, American Scientist **86** (1996), 358–363.

[Hi2] T. Hill, *A statistical derivation of the significant-digit law*, Statistical Science **10** (1996), 354–363.

[HuMi] C. Hughes and S. J. Miller, *Low-lying zeros of L-functions with orthogonal symmtry*, Duke Mathematical Journal, **136** (2007), no. 1, 115–172.

[HR] C. Hughes and Z. Rudnick, *Linear Statistics of Low-Lying Zeros of L-functions*, Quart. J. Math. Oxford **54** (2003), 309–333.

[Hu] W. Hurlimann, *Benford's Law from 1881 to 2006: a bibliography*, http://arxiv.org/abs/math/0607168.

[ILS] H. Iwaniec, W. Luo and P. Sarnak, *Low lying zeros of families of L-functions*, Inst. Hautes Études Sci. Publ. Math. **91**, 2000, 55–131.

[JMP] S. Jackson, S. J. Miller and V. Pham, *Distribution of eigenvalues for highly palindromic real symmetric Toeplitz matrices*, to appear in the Journal of Theoretical Probability.

[JKKKM] D. Jang, J. U. Kang, A. Kruckman, J. Kudo and S. J. Miller, *Chains of distributions, hierarchical Bayesian models and Benford's Law*, Journal of Algebra, Number Theory: Advances and Applications, volume 1, number 1 (March 2009), 37–60.

[KaSa1] N. Katz and P. Sarnak, *Random Matrices, Frobenius Eigenvalues and Monodromy*, AMS Colloquium Publications **45**, AMS, Providence, 1999.

[KaSa2] N. Katz and P. Sarnak, *Zeros of zeta functions and symmetries*, Bull. AMS **36**, 1999, $1-26$.

[Ke] J. P. Keating, *Statistics of quantum eigenvalues and the Riemann zeros*, in Supersymmetry and Trace Formulae: Chaos and Disorder, eds. I. V. Lerner, J. P. Keating & D. E Khmelnitskii (Plenum Press), 1–15.

[KeSn1] J. P. Keating and N. C. Snaith, *Random matrix theory and $\zeta(1/2+it)$*, Comm. Math. Phys. **214** (2000), no. 1, 57–89.

[KeSn2] J. P. Keating and N. C. Snaith, *Random matrix theory and L-functions at $s = 1/2$*, Comm. Math. Phys. **214** (2000), no. 1, 91–110.

[KeSn3] J. P. Keating and N. C. Snaith, *Random matrices and L-functions*, Random matrix theory, J. Phys. A **36** (2003), no. 12, 2859–2881.

[KiVu] J. H. Kim and V. H. Vu, *Concentration of multivariate polynomials and its applications*, Combinatorica **20** (2000), 417–434.

[KKM] M. Kologlu, G. Kopp and S. J. Miller, *The Limiting Eigenvalue Density for the Ensemble of Symmetric Period m–Circulant Matrices*, preprint.

[KKMW] M. Kologlu, G. Kopp, S. J. Miller and Y. Wang, *On the number of summands in Zeckendorf decompositions*, preprint.

[KoMi] A. Kontorovich and S. J. Miller, *Benford's Law, Values of L-functions and the $3x+1$ Problem*, Acta Arith. **120** (2005), 269–297.

[Knu] D. Knuth, *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*, Addison-Wesley, third edition, 1997.

[Le] M. Lewis, *Moneyball: The Art of Winning an Unfair Game*, W. W. Norton & Company, 2004.

[MO] G. Martin and K. O'Bryant, *Many sets have more sums than differences*. To appear in : Proceedings of CRM-Clay Conference on Additive Combinatorics, Montréal 2006.

[MMS] A. Massey, S. J. Miller and J. Sinsheimer, *Distribution of eigenvalues of real symmetric palindromic Toeplitz matrices and circulant matrices*, Journal of Theoretical Probability **20** (2007), no. 3, 637–662.

[McK] B. McKay, *The expected eigenvalue distribution of a large regular graph*, Linear Algebra Appl. **40** (1981), 203–216.

[MS] F. Mezzadri and N. C. Snaith (eds.), *Recent perspectives in random matrix theory and number theory*, LMS Lecture Note Series, 322, Cambridge University Press, Cambridge, 2005.

[Mic] P. Michel, *Rang moyen de familles de courbes elliptiques et lois de Sato-Tate*, Monat. Math. **120**, (1995), 127-136.

[Mil1] S. J. Miller, 1- *and* 2*-level densities for families of elliptic curves: Evidence for the underlying group symmetries*, Compositio Mathematica **140** (2004), no. 4, 952–992.

[Mil2] S. J. Miller, *Variation in the number of points on elliptic curves and applications to excess rank*, C. R. Math. Rep. Acad. Sci. Canada **27** (2005), no. 4, 111–120.

[Mil3] S. J. Miller, *Investigations of zeros near the central point of elliptic curve L-functions*, Experimental Mathematics **15** (2006), no. 3, 257–279.

[Mil4] S. J. Miller, *A derivation of the Pythagorean Won-Loss Formula in baseball*, Chance Magazine **20** (2007), no. 1, 40–48 (an abridged version appeared in The Newsletter of the SABR Statistical Analysis Committee **16** (February 2006), no. 1, 17–22).

[Mil5] S. J. Miller, *Lower order terms in the* 1*-level density for families of holomorphic cuspidal newforms*, Acta Arithmetica **137** (2009), 51–98.

[Mil6] S. J. Miller, *A symplectic test of the L-functions Ratios Conjecture*, Int Math Res Notices (2008) Vol. 2008, article ID rnm146, 36 pages, doi:10.1093/imrn/rnm146.

[MBD] S. J. Miller, E. T. Bradlow and K. Dayaratna, *Closed-form Bayesian inference for the logit model via polynomial expansions*, Quantitative Marketing and Economics **4** (2006), no. 2, 173–206.

[MilMo] S. J. Miller and D. Montague, *An Orthogonal Test of the L-functions Ratios Conjecture, II*, to appear in Acta Arithmetica.

[MNS] S. J. Miller, T. Novikoff and A. Sabelli, *The distribution of the second largest eigenvalue in families of random regular graphs*, Experimental Mathematics **17** (2008), no. 2, 231–244.

[MN1] S. J. Miller and M. Nigrini, *The Modulo* 1 *Central Limit Theorem and Benford's Law for Products*, International Journal of Algebra **2** (2008), no. 3, 119–130.

[MN2] S. J. Miller and M. Nigrini, *Order statistics and Benford's law*, International Journal of Mathematics and Mathematical Sciences, Volume 2008 (2008), Article ID 382948, 19 pages. doi:10.1155/2008/382948

[MilPe] S. J. Miller and R. Peckner, *Low-lying zeros of number field L-functions*, preprint.

[MW] S. J. Miller and Y. Wang, *From Fibonacci numbers to Central Limit Type Theorems*, preprint.

[Mon] H. Montgomery, *The pair correlation of zeros of the zeta function*, Analytic Number Theory, Proc. Sympos. Pure Math. **24**, Amer. Math. Soc., Providence, 1973, $181-193$.

[Nig1] M. Nigrini, *Digital Analysis and the Reduction of Auditor Litigation Risk*. Pages 69–81 in *Proceedings of the 1996 Deloitte & Touche / University of Kansas Symposium on Auditing Problems*, ed. M. Ettredge, University of Kansas, Lawrence, KS, 1996.

[Nig2] M. Nigrini, *The Use of Benford's Law as an Aid in Analytical Procedures*, Auditing: A Journal of Practice & Theory, **16** (1997), no. 2, 52–67.

[NM1] M. Nigrini and S. J. Miller, *Benford's Law applied to hydrology data – results and relevance to other geophysical data*, Mathematical Geology **39** (2007), no. 5, 469–490.

[NM2] M. Nigrini and S. J. Miller, *Data diagnostics using second order tests of Benford's Law*, Auditing: A Journal of Practice and Theory **28** (2009), no. 2, 305–324. doi: 10.2308/aud.2009.28.2.305

[OS1] A. E. Özlük and C. Snyder, *Small zeros of quadratic L-functions*, Bull. Austral. Math. Soc. **47** (1993), no. 2, 307–319.

[OS2] A. E. Özlük and C. Snyder, *On the distribution of the nontrivial zeros of quadratic L-functions close to the real axis*, Acta Arith. **91** (1999), no. 3, 209–228.

[Rai] R. A. Raimi, *The first digit problem*, Amer. Math. Monthly **83** (1976), no. 7, 521–538.

[RR] G. Ricotta and E. Royer, *Statistics for low-lying zeros of symmetric power L-functions in the level aspect*, preprint. http://arxiv.org/abs/math/0703760

[RSi] M. Rosen and J. Silverman, *On the rank of an elliptic surface*, Invent. Math. **133** (1998), 43-67.

[Ro] E. Royer, *Petits zéros de fonctions L de formes modulaires*, Acta Arith. **99** (2001), no. 2, 147-172.

[Rub] M. Rubinstein, *Low-lying zeros of L–functions and random matrix theory*, Duke Math. J. **109**, (2001), 147–181.

[RS] Z. Rudnick and P. Sarnak, *Zeros of principal L-functions and random matrix theory*, Duke Journal of Math. **81**, 1996, $269-322$.

[Si] J. Silverman, *The average rank of an algebraic family of elliptic curves*, J. reine angew. Math. **504**, (1998), 227-236.

[TW] C. Tracy and H. Widom, *Distribution functions for largest eigenvalues and their applications*, ICM Vol. I (2002), 587–596.

[Vu1] V. H. Vu, *New bounds on nearly perfect matchings of hypergraphs: Higher codegrees do help*, Random Structures and Algorithms **17** (2000), 29–63.

[Vu2] V. H. Vu, *Concentration of non-Lipschitz functions and Applications*, Random Structures and Algorithms **20** (2002), no.3, 262-316.

[Yo1] M. Young, *Lower-order terms of the 1-level density of families of elliptic curves*, Internat. Math. Res. Notices 2005, no. 10, 587–633.

[Yo2] M. Young, *Low-lying zeros of families of elliptic curves*, J. Amer. Math. Soc. **19** (2006), no. 1, 205–250.

[Za] I. Zakharevich, *A generalization of Wigner's law*, Comm. Math. Phys. **268** (2006), no. 2, 403–414.