



CALIFORNIA STATE SCIENCE FAIR 2013 PROJECT SUMMARY

Name(s) Colin C. Aitken	Project Number S1401
Project Title Dots and Lines: A Combinatorial Interpretation of the Homotopy Groups of Finite Topologies	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The homotopy groups of a topological space are commonly studied as a topological invariant which gives information about the space's holes, homotopy type, and connectedness. McCord showed that the homotopy groups are precisely those of corresponding simplicial complexes, but gives no explicit way of constructing or interpreting elements of the homotopy groups of a finite space, making applications difficult. An attempt to construct an analogue of homotopy groups for graphs was examined by Atkin and Smith, but the groups they associate to a graph G do not in general correspond to the homotopy groups of a finite topology whose associated graph is G when such a topology exists.</p> <p>Methods/Materials My project presents a new definition of the homotopy groups of a graph G, in the spirit of Atkin by using only combinatorial methods, and shows that these homotopy groups of G are indeed the same as any topology whose associated graph is G. In particular, this implies that two topologies with the same associated graph have isomorphic homotopy groups. This new definition allows for a construction with no known analogue in topological homotopy theory, as a graph which reduces the dimension of each homotopy group. More explicitly, for a graph G, I construct a graph G^k whose nth homotopy group is the same as the $(n+k)$th homotopy group of G, for all nonnegative integers n. This also allows for algorithms converging to presentations for any homotopy group of any graph (and any simplicial complex after using McCord's correspondence.)</p> <p>Conclusions/Discussion The question of the existence of solutions to a certain class of communications routing-type problems is also considered, and using homotopy groups I provide a method of phrasing such problems in terms of groups, which generally simplifies the solution. In fact, in certain cases the existence of a solution can be determined within $O(v^3)$ time (where v is the number of vertices). These problems occur in a variety of situations including chip and network design.</p>	
Summary Statement This project provides a method of constructing all continuous functions (up to pointed homotopy equivalence) from n -spheres to finite topological spaces, generalizes the method to arbitrary graphs, and considers a variety of applications.	
Help Received Father listened while I practiced presentation, Dr. Kubelka at San Jose State University helped review an earlier version of the work (only pointed out errors in grammar/formatting - did not actually contribute to the research).	



**CALIFORNIA STATE SCIENCE FAIR
2013 PROJECT SUMMARY**

Name(s) Ulysse Carion	Project Number S1402
Project Title Creating a Chess-Playing Computer Program	
Objectives/Goals The goal of this project is to create a computer program that plays a relatively strong game of chess using programming techniques used by the top engines in use today.	
Abstract Methods/Materials Godot uses bitboards (64-bit numbers representing a chessboard) to implement board representation. When searching for moves, the Godot uses the most common methods of the day, including alpha-beta pruning, principal variation searching, history heuristics, iterative deepening, quiescent searching, static exchange evaluation, and null move pruning. The program evaluates positions by taking into account many factors that are typically an indication of a strong position. Godot can evaluate tens of thousands of positions per second. Godot also has an opening book based on a large database of thousands of very high-quality games. At the time of this writing, Godot's opening database has a little over 252,000 positions pre-programmed into it. The program is based on other chess engines, especially open-source ones such as Stockfish, Carballo, and Winglet.	
Results Despite being based on other programs, Godot has a distinctive "style" of play that has been repeatedly described as appearing "creative". Godot has achieved an Elo ranking online of above 2200 at 1-minute chess. It has also defeated multiple FIDE-titled players. Though certainly not nearly as strong as commercial chess engines, Godot certainly plays a very respectable game of chess. Godot can defeat 99.89% of humans at one-minute chess.	
Summary Statement Creating from scratch a chess program that can defeat over 99% of humans and can independently play online against others.	
Help Received Online computer chess community provided technical support.	



CALIFORNIA STATE SCIENCE FAIR 2013 PROJECT SUMMARY

Name(s) Julia Cho; Josh Kim	Project Number S1403
Project Title Creating a Clearer Voice Chat/Conference by Using Sound Localization	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals In a multi-speaker voice chat or online conference, it can be very difficult to differentiate between speakers; by using a sound localization engine, our project aims to create a clearer conference experience. The engine will increase the productivity and efficiency of voice chats, which many students, businessmen, and casual users utilize.</p> <p>Methods/Materials To create a clearer voice chat, we will allow users to assign a location to a person's voice on the audio spatial scene. We implement Interaural Time Difference (ITD) and Interaural Level Difference (ILD) between the left and right ears, which are essential cues that allow humans to localize sounds, based on the location of a person. Our GUI displays the participants of the chat, represented as icons. When an icon is moved, its x and y coordinates are sent to the sound localization engine. The engine applies ITD and ILD to the monaural voice of each participant, and combines the resulting voices into a stereo audio that is sent to the user through headphones. Materials used during this project include a Mac, a PC, and Java integrated development environments like Eclipse and NetBeans.</p> <p>Results To test ILD, we positioned the speakers from left to right in ascending speaker number (0-4 with 2 in the middle of the screen) and printed the amplitudes received by each ear from each speaker's voice. We received results that showed that the ILD engine works: speakers 0 and 1's left ear amplitudes were greater than the right ear's, speaker 2's amplitudes were equal, and speaker 3 and 4's left ear amplitudes were less than the right ear's. We tested the ITD engine in a similar fashion and received similar results: when laid out from left to right in ascending speaker order, speaker 0 had the greatest right ear delay and speaker 4 had the greatest left ear delay. Overall we could hear the effects of the sound localization engine very clearly. We received a mono audio and outputted a stereo audio, with ILD and ITD applied.</p> <p>Conclusions/Discussion The results of our testing, combined with our knowledge of sound localization show that our program achieves its engineering goal. ITD and ILD, combined with the visual cues presented in the GUI, help the user to experience an interactive audio spatial scene.</p>	
Summary Statement We are creating a program that will help users to differentiate between speakers in a voice chat/online conference by implementing ILD/ITD/visual cues from the GUI.	
Help Received Our dads helped the team in finding resources about sound localization and taught us needed Java programming skills. Josh's mom helped with posterboard.	



**CALIFORNIA STATE SCIENCE FAIR
2013 PROJECT SUMMARY**

Name(s) Richard K. Cho	Project Number S1404
Project Title A Computational Exploration of Quadratic Residues and Their Applications	
Abstract Objectives/Goals My objective was to learn how to more quickly compute the number of quadratic residues modulo semiprimes. This was because I proved last year that the number of quadratic residues can be used to factor semiprimes, which is potentially a groundbreaking result since the difficulty of semiprimes being hard to factor is at the backbone of the RSA encryption scheme, which is essential to secure data transmissions across the internet. To this extent a faster factoring algorithm was not created, but gains were made toward a faster algorithm. Evidence was found that the distribution of quadratic residues modulo primes is nonrandom, and patterns were found and proven in the distribution of quadratic residues modulo semiprimes.	
Methods/Materials My objective was to learn how to more quickly compute the number of quadratic residues modulo semiprimes. This was because I proved last year that the number of quadratic residues can be used to factor semiprimes, which is potentially a groundbreaking result since the difficulty of semiprimes being hard to factor is at the backbone of the RSA encryption scheme, which is essential to secure data transmissions across the internet. To this extent a faster factoring algorithm was not created, but gains were made toward a faster algorithm. Evidence was found that the distribution of quadratic residues modulo primes is nonrandom, and patterns were found and proven in the distribution of quadratic residues modulo semiprimes.	
Results During further exploration into the distribution of quadratic residues, many things were discovered. I used the Dieharder test suite in conjunction with my own C++ programs for evaluating the randomness of the distribution of quadratic residues modulo primes. My tests have shown evidence against randomness, although a concrete explanation as to why remains unknown. Symmetry and relation to factors have been found and proven of the distribution of quadratic residues modulo a semprime; previously a relation to just the number of quadratic residues was proven. Furthermore, I implemented a multithreaded version of my program to generate quadratic residues, achieving a 65% time reduction.	
Conclusions/Discussion It is intriguing that my data points to quadratic residues modulo primes to not be random. The question remains, why? An explanation needs to be found. This could reveal very important relations regarding the distribution of quadratic residues, which could in turn be used to factor semiprimes. Also, while I proved relations that could improve the speed of counting the number of quadratic residues modulo semiprimes, it	
Summary Statement The project is on quadratic residues modulo primes and semiprimes and tries to find a quicker method of factoring semiprimes, since this would have a major impact on encryption schemes that rely on semiprimes being hard to factor, like RSA.	
Help Received Dr. Ali Gurel helped me verify the proofs I wrote; Dr. William Wu helped me get started with multithreading	



**CALIFORNIA STATE SCIENCE FAIR
2013 PROJECT SUMMARY**

Name(s) Bianca N. De	Project Number S1405
Project Title Modeling the Effects of Vascularization and Tumor Location on the Rate of Tumor Growth	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The project aims to create an accurate iterative computer model of the growth and vascularization of a tumor that projects the radius, volume, number of cells, and volume of vascularized tissue of the tumor over time based on rate of vascularization and location.</p> <p>Methods/Materials Scilab, a mathematical programming language similar to Matlab was used to program the model. A model of oxygen diffusion throughout the tumor was created. The variations in oxygen concentration throughout the tumor then dictated cell division and blood vessel growth, which in turn changed the pattern of oxygen content. Many simpler models, such as a model of cell division rates and a model of oxygen diffusion from a capillary, were created and combined to form a comprehensive picture of tumor growth based on the processes within the tumor. Parameters were in place for rate of vascularization and tumor size at the start of vascularization, a value influenced by the location of a tumor. Values for these parameters were determined from publicly available data and the model was run with these various values in place.</p> <p>Results The model created fit the data used well. Changes in the #rate of vascularization# parameter had a more significant impact on the tumor growth dynamics than did changes in the #initial size# parameter.</p> <p>Conclusions/Discussion The model appears to be accurate despite the assumptions made in the creation of the model. These include the assumption that the tumor had a perfectly smooth surface, disregarding small projections, and the assumption that interstitial fluid pressure compressed the tumor into a sphere. Other assumptions include the assumption that no branching took place in the vasculature of the tumor. It can therefore be concluded that the factors disregarded by the model have minimal impact on the growth of the tumor. The model created has many practical applications, including treatment projections and use in epidemiology studies.</p>	
Summary Statement An iterative mathematical model of tumor growth and vascularization dictated by oxygen distribution in tissue, that predicted effects of varying tumor location and vascularization rate, was created, and then simulated in Scilab.	
Help Received	



**CALIFORNIA STATE SCIENCE FAIR
2013 PROJECT SUMMARY**

Name(s) James Gabbard; Kathleen Grant; Andrew Liu	Project Number S1407
Project Title New Methods of Generating Sets of Intransitive Dice	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals Although there are many well-known sets of intransitive dice, there are very few published methods for generating these sets, and very little published work that explains why these sets are intransitive. Our project responds to an article published by M.N. Deshpande in 2000, which outlines one such method. The objective of our project was to explore new methods of generating sets of intransitive dice, given a set of specific dimensions and allowing no repeated face values within sets.</p> <p>Methods/Materials Initially, we used an algorithm coded in the computing software Wolfram Mathematica 9 to find all possible intransitive sets of 3 dice with 3 faces each, using face values of 1-9. We then identified a pattern underlying each set, and attempted to expand this pattern to fit larger sets. After identifying each working pattern, we searched for a proof that demonstrates the reasons that the pattern generates intransitive sets, and proves that the pattern will generate intransitive sets of any size. We then experimented with systematically altering the newly generated sets.</p> <p>Results The algorithm discovered all 5 possible sets that fit our criteria. One of these sets had already been studied in Deshpande's article. Manipulating the other 4 led to the discovery of 3 new patterns that generate sets of intransitive dice. We were able to prove that these patterns yield intransitive sets by developing algebraic expressions for the probability of victory of each die over its neighbors. Sets altered by switching adjacent integers or removing a row of faces remained intransitive, while adding additional faces or dice proved unreliable.</p> <p>Conclusions/Discussion Based on the success of our experimentation, we conclude that there exist at least 3 other methods for generating sets of any size besides the method published by Deshpande. This supports our initial hypothesis. Our project expands not only on the work of M.N. Deshpande (2000), by creating similar methods of generating sets, but on the work of Schaefer and Schweig (2012) by providing additional proofs that intransitive sets of any size can be generated.</p>	
Summary Statement Our experiment responds to a question posed by Indian mathematician M. N. Deshpande, and encompasses 4 previously undiscovered methods (with accompanying proofs) of generating sets of intransitive dice.	
Help Received Dr. Brian Conrey helped write algorithm, taught us how to use computational software.	



**CALIFORNIA STATE SCIENCE FAIR
2013 PROJECT SUMMARY**

Name(s) Vamsee K. Gangaram; Anup Vatti	Project Number S1408
Project Title Nonstop High Speed Rail	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals Travel time is a crucial factor for the California high speed rail (HSR) project to be successful. It must be significantly faster than car, and comparable to air travel time. In order to increase ridership and revenues, the HSR needs to serve many stations en route. Travel time is increased due to deceleration and acceleration for each station, in addition to the stop time as shown below. Proposed stations for the HSR cause at least a 20% delay in travel time, and it will get worse as more stations are added. Our goal is to reduce travel time while serving all stations.</p> <p>Methods/Materials We propose to reduce travel times by keeping the main train always running and connecting a local tram at the end for boarding and departing purpose. 1. The high speed rail goes nonstop, slowing down a little at each station. 2. A tram attached to the train at the end disconnects from the train and brings passengers to the station. 3. Another tram picks up passengers from the station and attaches itself to the high speed rail at the end. 4. A moving pathway in each car to ease passenger movement from the tram to their assigned cars We will show the above concept(method) with 3D CAD drawings and animation.</p> <p>Results I am not able to add tables and graphs on this page. We estimated stopping and acceleration times using kinematics equations. Based on this and published stop times at each station, we computed time wasted for each station for the current method and our method. Using published proposed time table for California HSR, we computed the travel time reduction for each segment. Using that and published projections of HSR ridership and revenues, we deduced the demand (ridership and revenue) increase using time elasticity of -0.8 and value of time using \$42 per hour estimate for HSR. Results show 13.33% reduction in average travel time, 5.61 million increase in ridership,\$181.27 million dollars of revenue increase and \$477.92 million worth of value of time savings.</p> <p>Conclusions/Discussion We showed how travel times can be reduced by avoiding stops while serving all stations along the route. Our proposed method is easy to implement and 1. Reduces travel time by an average of 13.33% with potential ridership increase of 5.61 millions 2. Allows adding more stations (increasing ridership) without affecting travel time much. 3. Keeps scheduling simple and avoids overtake complexities.</p>	
Summary Statement We reduce train travel time by avoiding stops for the main train and instead use secondary trains for boarding and departures.	
Help Received Dad helped with getting the poster printed by a local printing company; We tried to contact California High Speed Rail Authority but no luck so far.	



**CALIFORNIA STATE SCIENCE FAIR
2013 PROJECT SUMMARY**

Name(s) Isis M. Grant	Project Number S1409
Project Title Cruise Control	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals My objective was to make a robot drive autonomously at a constant velocity, regardless of incline. To achieve this, I compared two different approaches to velocity control: a fixed-step form and an implementation of a Proportional-Integral-Derivative (PID) algorithm. I predicted that PID would reach the desired speed faster than the fixed-step approach, without sacrificing other crucial aspects of control.</p> <p>Methods/Materials First, I wrote Java programs to implement both approaches and measure selected quality metrics. I then tuned each of the methods by adjusting algorithmic parameters in order to strike a balance between responsiveness and accuracy. I tested the programs by placing the robot at the bottom of a steep, uneven slope and measuring its velocity and the quality metrics while driving up the slope under the control of each of the two algorithms.</p> <p>Conclusions/Discussion I concluded that PID was the superior algorithm, due to its shorter response time and greater overall accuracy. However, the fixed-step approach turned out to be more accurate than I originally hypothesized, exhibiting fewer oscillations and more consistent oscillation amplitude than the PID program.</p>	
Summary Statement I compared control algorithms for maintaining the constant velocity of a robot on a slope.	
Help Received FIRST Robotics Team 100 built robot.	



**CALIFORNIA STATE SCIENCE FAIR
2013 PROJECT SUMMARY**

Name(s) Andrew C. Haden	Project Number S1410
Project Title Developing High-Performance Hybrid iOS Apps	
Abstract Objectives/Goals My goal was to improve performance of hybrid iOS apps, which are combinations of both native app code (objective-c) and web code (html5/css3). Methods/Materials Using my laptop, iPhone, and Apple's XCode development environment, I designed, implemented, and tested solutions for caching, animation, and communication in a hybrid app. Caching performance was measured in percent of assets stored on disk and percent of assets stored in RAM, as well as page load time in milliseconds. Communication performance was measured with message transit time after varied numbers of sequential messages. Results I successfully improved caching, animation, and communication performance. Caching tests indicated that the heuristics I implemented were successful in storing assets on both RAM and on disk and load times increased significantly; communication tests likewise displayed high performances, but slowed as the number of sequential messages increased. Conclusions/Discussion Hybrid app technologies are becoming increasingly popular with app developers such as Facebook, LinkedIn and Evernote because they shorten development time and allow primarily web developers to participate in app development. In my project, I successfully created (essentially drop-in) solutions for the issue of performance and communication in hybrid apps.	
Summary Statement I developed drop-in solutions to performance and communication in hybrid iOS apps.	
Help Received None	



**CALIFORNIA STATE SCIENCE FAIR
2013 PROJECT SUMMARY**

Name(s) Benjamin Jacobs; Steven Lin	Project Number S1411
Project Title The Randomness of Humans and Computers	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The purpose of this project is to determine whether humans or computers can generate a more random sequence of numbers without the use of any materials. Randomness can be applied to almost anything, such as choosing a random number for a game.</p> <p>Methods/Materials This project tested randomness in sets of 100 numbers. There were 60 sets generated by humans and 60 sets generated by computers. The computers generated random by using the Math.random() method from Java, which is a commonly known linear congruential generator. The human side was tested by having each subject write down 100 numbers that they believed to be random. The numbers were then ran through a statistical test called the serial correlation analysis. This test checks for any repeating patterns in the numbers. The closer the score was to 0, the better the source generated random.</p> <p>Results The results showed that the computer had generated random sequences more efficiently than the humans.</p> <p>Conclusions/Discussion The statistical tests were not very accurate because a set of 100 numbers was not large enough to determine if there was a significant correlation. Further research should include testing with bigger sets, or with a certain group of humans. The result of using a diverse group of humans might be different than a concentrated group, such as a certain job like a mathematics professor who often deal with numbers.</p>	
Summary Statement This project will test how well a person can generate randomness without any resources in comparison to a commonly used computer generated algorithm.	
Help Received Science teacher helped make writing more clear.	



CALIFORNIA STATE SCIENCE FAIR 2013 PROJECT SUMMARY

Name(s) Rishabh Jain	Project Number S1412
Project Title The Early Detection of Wildfires through the Use of GOES Satellite Imagery, Edge Analysis, and Spectral-Based Hashing	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals Wildfires threaten thousands of people each occurrence including fire-fighters and citizens; furthermore, these wildfires are usually caused through natural means such as heat waves or lightning strikes, resulting in the wildfire spreading before it is detected and reported. Wildfires are extremely dangerous to report and as a result they are usually monitored and detected from space. Although this type of detection from space is practical, a large data set is presented, leading to the common usage of the MODIS, which does not provide intense coverage of one area. Therefore, in this project, the GOES is used as it provides the intense coverage needed to detect, classify, and analyze wildfires; however, the large data set must be searched and created effectively as to reduce computational power and data storage requirements necessary.</p> <p>Methods/Materials In this project, a novel wildfire detection algorithm from space was developed, using edge analysis of the wildfire hot spots as well as spectral-based hashing with suffix arrays as a means of classification and early detection of wildfires. By first retrieving a temperature mapping from the GOES and identifying hotspots from temperature irregularities, these outlines can then easily be stored in a persistent database. Then, the dataset is searched by identifying correlated sequences and through the usage of the novel suffix array algorithm. Deviation analysis is conducted on each wildfire through time, allowing for the classification and analysis of wildfires from space.</p> <p>Results The flagged hot spots were compared to known hot spots as provided by the NOAA, the system correctly identified the hot spots. Furthermore, the algorithm identified the core of the wildfire as a new correlation of pixels was created through the usage of deviation analysis over time, allowing for a variance system of importance per hotspot to be created. Moreover, the algorithm is measured for run-time and space requirements and its parts are found to be under quadratic run-time and to have linear space requirements</p> <p>Conclusions/Discussion The algorithm presented is shown to be a powerful method for the early detection of wildfires from space and is also shown to have monumental applicability to other fields such as cancer research, Alzheimer's research, and security systems.</p>	
Summary Statement Creating a novel algorithm that is as effective and more robust in determining wildfires from space than current implemented algorithms and has wide applications in diverse fields.	
Help Received Father allowed me to understand advanced mathematical conceptions; Chris Spenner aided in the process of editing and refining my project.	



CALIFORNIA STATE SCIENCE FAIR 2013 PROJECT SUMMARY

Name(s) Anna K. Jones	Project Number S1413
Project Title Statistics on a 611-Tree Data Set	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals I had seven different hypothesis, six of which were that a certain environmental factor, diameter of trunk for example, causes spiral grain in Bristlecone Pine trees. The seventh hypothesis is that spiral grain is due to genetics. Using statistics I tested each of the hypothesis, except for the seventh which is the null hypothesis, and if each of the hypothesis involving environmental factors were rejected the null hypothesis would be accepted.</p> <p>Methods/Materials Using a 611-tree data set that included the diameter of the trees, whether the tree stands or has fallen, the species of the tree, etc., I used the chi-square test of independence to conclude if there is a relationship between any of the environmental factors and spiral grain. A few of the factors in the data set were categorical, including the spiral grain, so the chi-square test was accurate. I also used a correlation test to see if this resulted in no relationship between environmental factors and spiral grain as well. The correlation test needs to have all factors as numerical so this test can't be accurate until the degree of spiral grain is recorded and tested. I changed the values of spiral grain from left twist, no twist, and right twist to -1, 0, and 1. This allowed me to do the test, but not accurately.</p> <p>Results The chi-square test for each of the numerical factors was rejected. The chi-square value for each test was greater than the critical value found from a distribution table using the degrees of freedom appropriate for that test and a .05 level of significance. The correlation tests for the environmental factors and spiral grain, although not completely accurate without recording and testing the degree of spiral grain which we do not have, also showed that there is no relationship between spiral grain and the environmental factors.</p> <p>Conclusions/Discussion The null hypothesis is accepted because the other hypothesis were rejected. This means that spiral grain is random and due to genetics. There will always be more environmental factors to test but at this moment in time we can accept that spiral grain is due to genetics.</p>	
Summary Statement Using statistics to test the relationship of environmental factors and spiral grain has concluded that spiral grain is not due to environmental factors, but genetics.	
Help Received My science teacher, Michael Wing, allowed me to use his data set of the Bristlecone Pines. My AP Statistics teacher advised that I use the chi-square test but I taught myself how to do the test.	



**CALIFORNIA STATE SCIENCE FAIR
2013 PROJECT SUMMARY**

Name(s) Amirali Kamgar	Project Number S1414
Project Title Gaussian Curvatures of Non-Euclidean Surfaces	
Abstract Objectives/Goals The objective is to compare and contrast the first five postulates of Euclid in elliptic geometry , hyperbolic geometry and Euclidean geometry and measure the Gaussian curvatures of non-Euclidean surfaces. Methods/Materials The objective is to compare and contrast the first five postulates of Euclid in elliptic geometry , hyperbolic geometry and Euclidean geometry and measure the Gaussian curvatures of non-Euclidean surfaces. Results The first five postulates of Euclid can be applied only in Euclidean geometry and only a few of them can be applied in elliptic and hyperbolic geometry. Also measured curvatures indicates the sphere has positive Gaussian curvature and hyperbolic surface has negative Gaussian curvature. Conclusions/Discussion The first five postulates of Euclid contradict each other in hyperbolic and elliptic geometry. For the sphere the measured curvature matched the theoretical curvature to within 1% error. For the hyperbolic paraboloid ,the result were less accurate ,the experimental curvatures were all in the same order of magnitude as theoretical and were all negative.	
Summary Statement My project is about Gaussian curvatures of Non-Euclidean surfaces and differences of the first five postulates of Euclid in the non-Euclidean geometry	
Help Received Used lab equipment at Ribet academy under the supervision of Mr. John shirajian (science teacher)	



CALIFORNIA STATE SCIENCE FAIR 2013 PROJECT SUMMARY

Name(s) Abraham P. Karplus	Project Number S1415
Project Title Arduino Data Logger	
Abstract Objectives/Goals The goal of this project is to create a software program, usable by non-programmers, for recording continuous streams of data from sensors. The data logger is physically an Arduino microcontroller connected to a computer. The Arduino reads from electronic sensors and sends the readings to the computer which displays them to the user and saves them in files. The data logger was originally designed for the new Applied Circuits for Bioengineers course at UCSC. However, it is general purpose, usable in other classes and labs and by hobbyists.	
Methods/Materials The hardware involved in the data logger is sensors, the Arduino, and a computer. Many sensors and devices, analog and digital, can be connected to the Arduino, e.g. a thermistor as a temperature sensor, or a string, pulley, and photogate for timing moving objects. Most models of Arduino microcontroller are supported, as are Mac OS X, Windows, and Linux on the personal computer. The software libraries I used, Tkinter, PySerial, and the Arduino IDE, handle some of the cross-platform compatibility. This project has mostly been writing the software for the computer and Arduino. On the computer, I used the Python programming language. My program provides a simple graphical interface for configuring the Arduino and displays a live feed and charts of the data received. I wrote the code for the Arduino in C++, using only libraries provided with the IDE.	
Results Currently, the data logger is functional and useful, with many enhancements planned. Sampling rates of a hundred Hertz for several minutes have been achieved, despite the tiny memory on the Arduino. The students in thre UCSC circuits course provided useful feedback on the data logger and identified several bugs. The code for the data logger is available at < bitbucket.org/abe_k/arduino_data_logger >.	
Conclusions/Discussion I learned a lot while creating the programs for the data logger. It was my first embedded programming project, for a microcontroller with very limited speed and memory. I learned about interrupt-driven programming (both timed and external) and low-level details of the AVR chips. On the Python end, I learned about multi-threaded programming along with Tkinter GUI toolkit. I also learned good software engineering practices, such as version control and program documentation. This project was both educational and enjoyable, and the resulting program is valuable to others.	
Summary Statement The purpose of this project is to create an inexpensive data logger to electronically record from sensors and save these readings.	
Help Received Father provided feedback on program; UCSC circuits course tested program; Writing tutor helped with poster	



**CALIFORNIA STATE SCIENCE FAIR
2013 PROJECT SUMMARY**

Name(s) Kevin K. Lee	Project Number S1416
Project Title Electromechanical Modeling of the Heart in Moving Domains Using the Phase-Field Method	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals To create a mathematical model of a beating heart that combines mechanical contraction with electrical excitation.</p> <p>Methods/Materials I developed a novel method for combining the propagation of electricity with mechanical contraction by using the phase-field method. I linked the Fenton-Karma model of the heart's electrodynamics to the physical motion of the heart while accounting for factors such as tissue elasticity and movement of the tissue. The contractile mechanics of the heart were represented by considering it as fluid with an elastic boundary using Darcy's Law, and its shape was evolved through a Cahn-Hilliard type equation. The equations were solved using a second-order accurate semi-implicit Crank-Nicholson scheme on an adaptive multigrid and MATLAB was used to analyze the results.</p> <p>Results I validated my algorithm by successfully observing its convergence toward a pre-constructed solution when the corresponding forcing function was incorporated. I also showed that the behavior of the solution is consistent with that presented in previous research which analyzed the heart in non-moving domains.</p> <p>Conclusions/Discussion I developed an efficient and widely applicable algorithm for more realistically simulating the beating of the heart. The algorithm is independent of the numerical scheme and can easily be applied to 3-D anatomical models of the heart. This gives drug developers a more complete tool in designing therapies for heart conditions, yields critical insight on the underlying mechanisms of fatal conditions like fibrillation, and enables dramatic improvements in their treatment and prevention.</p>	
Summary Statement I created a mathematical model of a beating heart that realistically represents the dynamics of contraction, which improves on existing models and allows for a more comprehensive analysis of the mechanisms of heart diseases.	
Help Received Professor Lowengrub helped me develop the model and allowed me to use the computing cluster in his lab. Esteban Meca helped me learn how to numerically solve the differential equations.	



**CALIFORNIA STATE SCIENCE FAIR
2013 PROJECT SUMMARY**

Name(s) George D. Morgan	Project Number S1417
Project Title Operating System Design and Development Continued	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The objective of my project is to apply modern programming paradigms to in-depth operating system design and development.</p> <p>Methods/Materials ~ A target architecture. The ARM architecture. ~ An target ARM platform. The AT91SAM7S321. ~ A build environment. Mac OS X Mountain Lion. ~ An integrated development environment. Xcode. ~ A package manager for Mac OS X. Homebrew and MacPorts. ~ An ARM toolchain. The GNU ARM Toolchain ported to Mac OS X. ~ An application performance analyzer. Instruments.</p> <p>Results Locality of reference and differential inheritance, along with modern programming paradigms such as object orientation and polymorphism, have the potential to enable a given operating system to more efficiently manage its resources and handle jobs. Inefficiencies such as duplication of code in memory, redundant instruction execution, and incoherent structure can be combatted, or altogether eliminated, resulting in an operating system that performs more efficiently than a traditional, non-paradigmatic kernel.</p> <p>Conclusions/Discussion In conclusion, the application of modern paradigmatic programming to the design and development of an operating system framework, on top of the concepts of object-orientation and reflection, which were introduced in last year's study, resulted in a dynamic kernel environment capable of operating at a higher level of abstraction than any other operating system has for over 40 years.</p>	
Summary Statement My project focuses on the application of modern paradigmatic programming to operating system design and development.	
Help Received None. All of the work done on this project was my own.	



**CALIFORNIA STATE SCIENCE FAIR
2013 PROJECT SUMMARY**

Name(s) Gilad E. Ronat	Project Number S1418
Project Title Graphical: Graphing Calculator App for the iPhone	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals Design and develop an intuitive, function-based graphing calculator app for iOS.</p> <p>Methods/Materials I planned the app's features, then sketched and drew wireframes for each screen in a sketchbook. The user interface of each screen was then designed in more detail. Next I created charts for object communication using the MVC paradigm on a whiteboard. I then set up the basic structure of the app in Xcode on a MacBook Pro running OS X Mountain Lion 10.8, and coded the simple functionality of each of the planned screens (basic equation editor, graph view, and table view) using the iOS SDK and Core Graphics. Each screen was revisited as I added more features, improved, and refined while testing along the way. The app was tested on an iPhone 4S running iOS 6.1 and finalized by fixing bugs.</p> <p>Results I succeeded in developing the graphing calculator app which accurately displays the graph of any function the user enters with fluid gesture support, custom UI, and the ability to analyze graphs.</p> <p>Conclusions/Discussion The process of iteration and testing is critical in software development to create a polished product. For instance, at first the graph view recalculated all the values of the function visible on screen for every change of the origin as the user panned around. While this ran smoothly on the iPhone Simulator backed by a computer processor, testing on the actual iPhone hardware revealed that this yielded terrible performance. To solve the issue, I had to implement a cache to keep the previously calculated values and use them while panning, triggering a refresh of the values once the user stops panning. I was able to overcome complex problems and challenges by breaking them down into smaller steps.</p>	
Summary Statement Using my knowledge of the iOS SDK along with Apple's docs, I created a graphing utility for the iPhone that will aid myself and other students once it arrives on the App Store.	
Help Received The project was completed entirely by myself.	



CALIFORNIA STATE SCIENCE FAIR 2013 PROJECT SUMMARY

Name(s) Gabriel S. Rossmann	Project Number S1419
Project Title Microfluidic Chip and Software Interface for Multi-Week Tissue Culturing Experiments	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The purpose of this engineering project is to design an end-to-end system to enable multi-week tissue culture experiments. Such experiments are not possible using traditional petri-dish culturing techniques nor commercially available microfluidic devices. The latter accommodate cells in suspension not whole tissue samples. Live imaging of cellular behaviors, drug screening and toxicological tests are some of the areas of research that could benefit from long duration tissue culturing experimentation.</p> <p>Methods/Materials The system I designed includes a novel large, open-top chamber, high fluidic resistance microfluidic device, an inexpensive two-way hardware controller, and an open source software stack supplemented by a custom software layer I developed. The equipment used includes my custom microfluidic chip fabricated by Stanford University's foundry, an Arduino, an open source, two-way single-board microcontroller, 8 solid state relays, 8 Festo solenoid valves to control the on-chip pumps, a Windows PC for AutoCad based microfluidic chip design, and a Linux laptop to write and test the controller software.</p> <p>Results Software tests confirmed that arbitrary experiments could be scripted and run remotely over the Internet. Hardware tests confirmed that software scripts could effectively control the microfluidic flows and that the scripts could actuate the valves at frequencies exceeding the 10Hz required for microfluidic peristaltic pumps. Flow tests on two versions of the chip confirmed that fluid could flow in and be evacuated from the culture chambers. Flow rates were compatible with cell survival. Cells were introduced by my customer in a fully functioning end-to-end system and were successfully cultured for 20 hours.</p> <p>Conclusions/Discussion Chip and software tests demonstrate that an automated microfluidic solution can fulfill the customer's requirement for long-term culturing. Based on new customer input and insights from the first set of experiments with cells, a second generation chip was designed and fabricated. It will be used to rigorously verify the biological efficacy of the microfluidic system employing cell tissues. A point-and-click interface will be added to enable non-programmers to automate experiments. A whole class of modern biological research involving cell tissues could be aided by the combination of these capabilities.</p>	
Summary Statement I designed a microfluidic chip and a software interface to enable multi-week tissue culturing experiments.	
Help Received Prof. Heller suggested microfluidic project and let me use equipment, Stanford U. foundry made the chips.	



**CALIFORNIA STATE SCIENCE FAIR
2013 PROJECT SUMMARY**

Name(s) Sonia S. Sachar	Project Number S1420
Project Title MASS: Memorizing Assistant for Social Software	
Objectives/Goals The best way to memorize information is by using a learning method called space repetition. This method is implemented via flashcard-based systems. The most popular algorithm that is used to implement space repetition learning is the Super Memo 2 (SM2) algorithm. I enhanced this algorithm with modifications such as using multiple repetitions in learning phase, using response time to deduce quality of retention, and using most effective repeat interval data from participants to set initial interval value for each word. The purpose of my project is to use my enhanced algorithm to build a Facebook application, SAT Memory Master, which helps users memorize vocabulary more effectively in a social setting.	
Abstract Methods/Materials 1)Researched learning techniques 2)Proposed a new algorithm (MASS) to aid in memorization process by modifying and adding to SM2 algorithm. 3)Built a Facebook application to help users memorize SAT vocabulary by using MASS algorithm. Designed the software such that different version of algorithm can be used by different users and performance data can be collected. 4)Conducted two experiments to compare new enhancements to the regular algorithm Experiment 1: Compared two different learning techniques found in the different algorithms Experiment 2: Compared changed initial intervals based on difficulty of words to set initial interval 5)Collected and compared data To build the Facebook application, the software tools used were PHP, HTML, MySQL, Facebook Plug-in, and Photoshop	
Results Experiment 1 showed that the MASS algorithm is 18% more efficient than the SM2 Algorithm. Experiment 2 showed that the MASS algorithm can be made 11-13% more efficient by customizing initial intervals for easier and hardier words.	
Conclusions/Discussion Since SAT Memory Master has been live on Facebook, it has attracted more than 1000 users in roughly five months. I have quantitatively shown that my MASS algorithm is 29% to 31% more efficient than SM2.	
Summary Statement This project introduces and implements a new algorithm as a Facebook application to help students memorize SAT vocabulary more efficiently in a social setting.	
Help Received Online Information such as user guides, articles, forums, etc.	



**CALIFORNIA STATE SCIENCE FAIR
2013 PROJECT SUMMARY**

Name(s) Zachary H. Seligman Karen	Project Number S1421
Project Title An Investigation of Shapes of Unvarying Height	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals This research seeks to examine the characteristics and traits of shapes of unvarying heights regardless the orientation. Through simple proofs using geometric/trigonometric principles, I derived the general formulas for the areas and perimeters of Reuleaux Polygons, constructed from an n-gon with an odd number of sides. Furthermore, to observe their general behavior as the number internal sides approaches infinity, the equations derived were evaluated with limits with cases of L'Hopital's Rule.</p> <p>Methods/Materials The construction is simple; using the longest diagonal from each of the vertices on any regular, odd-sided n-gon as a radius to create an arc.</p> <p>Results The perimeter of these shapes is based on the length of the arcs, the height and number of internal sides being the only variables. The area is particularly interesting as it encompasses both the area of the regular, odd-sided n-gon and the area of the little "smigums" allowing for the regular height. "Smigums" being defined as the difference between a sector and its corresponding internal triangle. Also, if the constructed shape is rotated around an axis of symmetry, it creates a three dimensional prism with the property of uniform height too.</p> <p>Conclusions/Discussion Although a circle is the first shape that comes to mind to many people when asked to think of a shape that has a uniform height regardless the orientation, this study showed that infinitely many shapes with a uniform height exist, all based on regular polygons with an odd number of sides.</p>	
Summary Statement This project sought to bring light to the behavior and characteristics of Reuleaux Polygons, bringing them into the third dimension, and potential application.	
Help Received Father gave guidance in the construction of the standing display board and the 3-dimensional models; Mother made sure I was fed so I could think properly; Annabel Adriatico gave tips on general logistics for the writing of the project; Nicolas Abruzzo helped flush out ideas for the direction to take in the	



**CALIFORNIA STATE SCIENCE FAIR
2013 PROJECT SUMMARY**

Name(s) Joshua G. Send	Project Number S1422
Project Title Two Dimensional Localization of Impacts Using Acoustic Time Delay Measurement	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals This project's main goal is to show that using time delays on the microsecond scale is a viable way of determining the source of a vibration caused by an impact on a two dimensional surface.</p> <p>Methods/Materials An Arduino Microprocessor connected to three analog-digital converters, which are connected to three accelerometers, is used to detect vibrations that propagate through a plastic board in the event of an impact. The Arduino transmits the set of acceleration data it collected from each sensor to the computer, which can then apply three data analysis algorithms. Because the general method of finding the intersection of three hyperbolas to determine the impact location depends on accurately determining the arrival time of the vibration, the three algorithms can be tuned interactively. The first method uses simple a threshold comparison, the second uses a least squares line fit through increasing maxima of rectified acceleration data, and the third uses wavelet decomposition to determine the starting time of the vibration.</p> <p>Results Three signal analysis methods were tested. Of the three, wavelet decomposition of the signal yielded the best results. Second best by a large margin was the basic threshold comparison that worked best closer to the center of the board. Lastly, the least squares line fit algorithm usually failed to get even 15 cm radius accuracy when the impact did not occur in the center. It was also found that applying a spline fit (which is required anyway for wavelet analysis) improved accuracy for the other two algorithms.</p> <p>Conclusions/Discussion The results show that it is completely possible to localize impacts on a hard surface, and the accuracy depends on analysis methods used, signal noise, and inherent limits of the hardware. The maximum theoretical accuracy, as restricted by sampling speed, is about a one centimeter radius, and the system developed can achieve a root mean square error of close to two centimeters.</p>	
Summary Statement This project uses multilateration based on the time differences of arrival times of a vibration at three accelerometers to localize the impact that caused the vibration.	
Help Received Dr. Uwe Send (father) helped understand the concepts of wavelets and decypher the wavelet library for Python.	



**CALIFORNIA STATE SCIENCE FAIR
2013 PROJECT SUMMARY**

Name(s) Dennis J. Shim	Project Number S1423
Project Title Beating Google: Development of the Novel Web Page Ranking Algorithm WordRank in a Search Engine	
Abstract Objectives/Goals To develop a new algorithm that will more effectively rank web pages so that a search engine can return more meaningful results, especially for specific queries. Methods/Materials A search engine with a crawler and a lookup function was created in Python. A novel page ranking algorithm named WordRank, which ranks pages based on user inputted keywords and a page's content, was designed and implemented. For a comparative study, the two well-known ranking algorithms HITS and PageRank were also implemented in the search engine. The results from each of these algorithms and different combinations of algorithms were compared using the number of relevant links and quality scores. Results A comparative analysis of PageRank, HITS, and WordRank was performed. It was observed that the search results from HITS were of lower quality than those from PageRank. PageRank was thus used in conjunction with WordRank. When a combination strategy was applied to the PageRank and WordRank scores, it was determined that the results from a PageRank to WordRank ratio of 50:50 were generally more relevant and had higher quality scores than 25:75 and 0:100 ratios. This 50:50 ratio was then compared to pure PageRank. It was found that the combination returned better results overall, but returned especially relevant results for specific queries. Conclusions/Discussion To create a viable search engine with a better page ranking algorithm than PageRank or HITS, the novel WordRank algorithm was successfully designed and implemented in Python. A 50:50 ratio of PageRank to WordRank returned the best search results, especially for specific queries. This combination has large potential use in a research search engine for users looking for specialized results.	
Summary Statement In this project, I developed the novel page ranking algorithm WordRank to return high quality results to specific queries for use in a search engine.	
Help Received Dr. James Li helped me with programming questions that I had.	



CALIFORNIA STATE SCIENCE FAIR 2013 PROJECT SUMMARY

Name(s) Yousuf M. Soliman	Project Number S1424
Project Title IlluminaMed: Developing New AI for Use in a Biomedical Image Analysis Toolkit	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The main aim of this research is to develop new, more effective methods of Artificial Intelligence and Fuzzy Logic Techniques and integrate them with biomedical imagery and spectroscopy analysis and interpretation. This will result in advancing the fields of both Artificial Intelligence and Radiological Scan Analyticity.</p> <p>Methods/Materials Throughout the course of this research I gained access to several MRI databases to train my AI to. I began my research by studying and developing new types of neural networks which would have both short and long term memory as well as larger output sets with faster training times. I developed several algorithms which would initially segment the MRI into the specified classes, in specific the white matter, grey matter, cerebral spine fluid, and edema. I then trained several neural networks to a training set of 750 MRI scans to find tumors effectively. I also trained illuminaMed to find and distinguish lesions in Multiple Sclerosis and ALS. Lastly, I trained illuminaMed to find patterns between brain activity and brain structure with the tendency of a patient to develop Alzheimer's.</p> <p>Results IlluminaMed was able to automatically segment MRI scans with a 98% accuracy while separating the pixels into the correct classes. IlluminaMed was then able to automatically detect tumors in MRI scans with a 93% accuracy as well as approximate its volume within a 5% error margin. Lastly IlluminaMed was able to find correlations between brain structure and activity with the tendency of a patient to develop Alzheimer's before symptoms begin to appear.</p> <p>Conclusions/Discussion In its current form, IlluminaMed is very practical and is in te process of being implemented into a Health Information System. Although IlluminaMed is a very powerful tool and is able to accomplish several tasks very efficiently, it still has a lot of room for potential. As I plan to continue research in this field I plan to further develop this tool to be able to analyze and segment many different types of biomedical scans. Another line of research would be to incorporate more efficient algorithms as to have this program be more lightweight and run on less demanding computational machines.</p>	
Summary Statement I developed a biomedical image analysis toolkit that can automatically segment and analyze MRI scans using AI for tumors, lesions in Multiple Sclerosis and ALS, as well as determine the tendency of a patient to develop Alzheimer's.	
Help Received Several professors for providing me access to MRI scan databases. Najwan Naserelldain for helping me learn to pprogram years ago.	



CALIFORNIA STATE SCIENCE FAIR 2013 PROJECT SUMMARY

Name(s) Prem M. Talwai	Project Number S1425
Project Title A Novel Mathematical Model of Targeted Cancer Therapy along p53 Proteasomal Degradation Pathways	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The objective of this project is to derive a mathematical model for describing multi-substrate enzyme kinetics, particularly focusing on the harmful degradation of the tumor suppressor protein p53, which is mediated by the oncogene MDM2. The mathematical model will then be used to determine whether the ubiquitin ligase MDM2 or the ubiquitin-conjugating enzyme E2D3 plays a larger role in the degradation process. This discovery will help identify new potential targets for cancer therapy.</p> <p>Methods/Materials A novel system of nonlinear partial differential equations incorporating initial concentration variables was derived in order model the enzyme mechanism. The nonlinear system was subsequently linearized using perturbation methods and solved using conventional linear algebraic techniques. The resulting solution to the model was validated against experimental results from gel-electrophoresis experiments conducted by Lai et al. Four dynamic modules were then constructed using the Mathematica program to capture the influence of initial concentration variables on the role played by E2D3 and MDM2 in the degradation process. The model was simulated for various values of the rate constants in order to confirm the observed patterns.</p> <p>Results Through growth rate analysis, it was concluded that although MDM2 and E2D3 initially compete for influence in the degradation process, MDM2 plays the larger role. Therefore the derived mathematical model demonstrates that targeting MDM2 with drugs such as Gleevec and Nutlin-3 will reduce unregulated p53 degradation and consequently redeem the integrity of the genome.</p> <p>Conclusions/Discussion My mathematical enzyme kinetic model allowed for the identification of MDM2 as a new potential target for cancer therapy. Furthermore, many of the relationships between enzymes, substrates, and products identified during the course of this project can be extrapolated to any biological process governed by a compulsory-order multi-substrate enzyme mechanism. The extension of the results obtained in this project may therefore lead to similar discoveries of key biomolecules driving various other vital biological processes. In future study, the computational results obtained in this project may be further verified by in vitro experimentation. In addition, a stochastic kinetics formalism may be adopted on the same problem in order to characterize random molecular motion and probabilistic enzyme binding.</p>	
Summary Statement I developed a novel mathematical enzyme kinetic model of MDM2-mediated degradation of p53 which enabled me to identify MDM2 as a new potential target for cancer therapy.	
Help Received	



CALIFORNIA STATE SCIENCE FAIR 2013 PROJECT SUMMARY

Name(s) Tanay Tandon	Project Number S1426
Project Title Clipped: Automated Text Summarization through Semantic Natural Language Processing and Clustered Machine Learning	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals Individuals and companies in today's world are inundated with a flood of data. The issue of information overload has made it exceedingly difficult to function efficiently as part of modern day society. Between web pages, social networks, news articles, and email; everyone parses through hundreds if not thousands of documents on a daily basis. People need a way to quickly determine what these documents mean, and extract the most important information from them. The purpose of this research is to develop a text summarization algorithm that makes use of novel grammatical models and linguistics to statistically extract the most relevant information from a text sample. The algorithm generates a bullet point summary of the text by discovering information concentration through subject-predicate relationships and keyword ranking based structures; and makes the process of reading the content more concise and efficient.</p> <p>Methods/Materials All development occurred in the server side language of PHP, and the entire algorithm was written independently. The program made use of the lexicon database known as the Brown Corpus to tag sentences through POS (Parts of Speech). The algorithm code was tested and deployed on a server with a user-testing beta to ensure scalability. The algorithm was developed in 3 major stages # the ICF (Initial Content Filtration), GPM (Grammatical Pattern Matching), and the CCR (Contextual Content Review). The algorithm was trained to identify 23 patterns of information concentration, involving subject-predicate positioning and keyphrase concentrations.</p> <p>Results The results of the study indicate that the algorithm holds high precision in summarization of content. The ROUGE metric generated an average F1 score of .5868, and an average precision score of .640. The ROUGE metric evaluates a summary through a similarity index run between a gold summary and the test summary. The summaries generated are concise and comprehensive, and the algorithm was released as a beta product on the iOS and Android Market # reaching the top of the app store within 2 months.</p> <p>Conclusions/Discussion The research indicates that the Clipped algorithm has the potential to change the way we consume content. The novel grammatical approach identifies information concentration more accurately, and the bullet-point summaries make the process of reading more efficient than ever before.</p>	
Summary Statement I developed a text summarization algorithm that analyzes information grammatically, and generates a bullet-point summary of the text sample; making the process of consuming information efficient.	
Help Received Parents helped make display board. Gary Griffiths of Trapit provided advice on application	



**CALIFORNIA STATE SCIENCE FAIR
2013 PROJECT SUMMARY**

Name(s) Manogna Vemulapati	Project Number S1427
Project Title Identification of CpG Islands in a DNA Sequence Using a Hidden Markov Model Trained in MapReduce	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The purpose of my project was to design and develop a software program to identify the locations of CpG islands in a DNA sequence. Hidden Markov Model (HMM) is the most popular statistical model used to locate CpG islands in DNA sequences. The accuracy of this model is limited by the size of the training data. The engineering goal of my project was to parallelize the training of the HMM so that (a) the training could be performed on larger data sets in reasonable time and (b) the resulting trained HMM could be used to more accurately identify the CpG islands.</p> <p>Methods/Materials I modeled the CpG island detection problem as a HMM by identifying the hidden states and emission symbols. There were 8 hidden states: A+,C+,G+,T+ corresponding to the bases within an island and A-,C-,G-,T- corresponding to bases outside an island. The observable emission symbols were A,C,G,T. A state such as C+ or C- emits the symbol C with a probability of 1 and similarly for other states. I used the MapReduce enabled version of Baum Welch algorithm from Apache Mahout project to perform unsupervised training on a sample contig from a human chromosome. I used the Amazon Elastic MapReduce platform to run the training. From the resulting trained HMM, I obtained the 8x8 state transition probability matrix. I used the Viterbi algorithm from HMM package in R to decode the hidden state sequence for given test contig from another human chromosome sequence. From the decoded hidden state sequence, I obtained the start and end positions of each CpG island in the test sequence.</p> <p>Results Unsupervised training on contig NT_028395 from human chromosome 22 (hg19) completed in about 2 hours on the MapReduce platform. From the trained HMM, I was able to identify about 80 percent of CpG islands for some test sequences from human chromosome 8. The same training using a non MapReduce version of Baum Welch algorithm did not complete even after few hours of running.</p> <p>Conclusions/Discussion I have observed that training the HMM on MapReduce cluster facilitates training on larger data sets in shorter time and also results in a more accurate identification of CpG islands in DNA sequences.</p>	
Summary Statement My project is about identifying CpG islands in a DNA sequence more accurately by training the HMM on large data sets.	
Help Received Father helped me with the setup and running jobs on Amazon MapReduce platform.	



CALIFORNIA STATE SCIENCE FAIR 2013 PROJECT SUMMARY

Name(s) Sohil Bammi; Brandon Betkolia; Eric Zhou	Project Number S1497
Project Title The Automated Creation of Randomized Relational Languages as a Web Application	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The goal of this program is to offer an engine for creating new, phonetically-correct languages via a web service. Furthermore, this program should effectively be used as an encryption device when paired with hashing algorithms. This should be further proven by experiments which will be carried out following the creation of the website.</p> <p>Methods/Materials The website MyVernacular was set up by purchasing a domain name. The engine was created in Perl and then implicated into the website. Then, a list of the 100 most popular passwords were taken and run through two industry standard hashing algorithms, MD5 and SHA1. Using an online decrypter we tested whether the hashes could be reversed and the password revealed. We repeated the same test three more times, but translated any English words to Italian, German, and a language generated by MyVernacular. Also, in order to test the memorability of the language we gathered 48 randomly selected students who were willing to participate in a memory test. They were randomly given a word chosen from three sets of words: English, German, and a randomly generated language from MyVernacular. All subject were told to spend one minute memorizing the word, and to not write down that word for later records. After three days, the subjects wrote down what they remember of the word. The number of successes (the word that the subjects remember matched exactly the word they were assigned) of each set of words was recorded.</p> <p>Results We were able to completely construct a website that allows a user to create a randomized relational language. In encryption test, all English passwords were cracked, and most of the German and Italian ones were effectively cracked as well. However, only two of the MyVernacular passwords were cracked. Furthermore, in a test of memorability, 91.7% of the test subjects given an English word remembered their designated word. 83.3% remembered the MyVernacular word. 66.7% remembered the Italian word. 50.0% remembered the German word.</p> <p>Conclusions/Discussion Overall, the engine was able to randomly create a language based on phonetics. MyVernacular was clearly used effectively in cryptography as an extra layer of abstraction between a password and a hash algorithm. Additionally, words served a practical function because of their memorability when compared to the original language (English) and the control languages (German and Italian).</p>	
Summary Statement This is a web service that creates a phonetically-correct and memorable language which can be implicated as an extra layer of abstraction in cryptography.	
Help Received Students participated in the memorization tests	



**CALIFORNIA STATE SCIENCE FAIR
2013 PROJECT SUMMARY**

Name(s) Jared A. Tramontano	Project Number S1498
Project Title A Generalized Formula for the A-th Element of a N-Nacci Recursive Sequence Using Complex Residues	
Abstract Objectives/Goals The objective of the project was to derive a closed form formula for the A-th element of a N-Nacci recursive sequence using the residue theorem from complex analysis. Methods/Materials In order to derive such a formula, I began with finding a closed form for the generating function for a N-Nacci recursive sequence. After this I showed that the generating function will be holomorphic for a suitable length of Z less than epsilon. I then proceeded to use Cauchy's Differentiation Formula to extract the A-th element of the sequence. Following this I used the Residue Theorem to relate the A-th element to the residues of the closed form generating function not about zero. Finally, I wrote two brief proofs to show that the generating function would only contain simple poles of multiplicity one, as well as a short proof of Binet's Formula. Results It was determined that the A-th element of any linear homogeneous recurrence relation with constant coefficients is equal to the negative sum of the residues of its closed form generating function not about zero. Conclusions/Discussion The results of this project show a deep connection between complex residues and linear homogeneous recurrence relations with constant coefficients. It provides a different method to solving recurrence relations than the standard linear algebraic method.	
Summary Statement I derived a closed form formula for any element within a linear homogeneous recurrence relation with constant coefficients using complex residues.	
Help Received Mathematical discussions with Dr. Gerhard Gierz, as well as Mr. Brian Drobot.	



**CALIFORNIA STATE SCIENCE FAIR
2013 PROJECT SUMMARY**

Name(s) Liang Zhou	Project Number S1499
Project Title Determining Combinatorial Sums Using Roots of Unity	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The objective is to determine an algorithmically efficient closed-sum formula for the combinatorial sum $nC_0 + nC_c + nC(2c) + nC(3c) + \dots$</p> <p>Methods/Materials Following a compilation of relevant background research and lemmas regarding roots of unity and Euler's formula, several numerical examples were analyzed in order to prove the main formula. Subsequently, the mechanics of the proof were discussed through an explanation of a clever application of the Binomial Theorem and its relation to the lemmas. The main formula was then further generalized to produce a corollary and various extensions. Finally, a Java program was developed to explore the efficiency of the main formula and its potential applications.</p> <p>Results The cyclic nature of roots of unity working in tandem with the combinatorial coefficients of the Binomial Theorem proved an especially powerful tool in "filtering" terms - with the appropriate substitutions, desired terms remained while unwanted ones were subsequently cancelled out. Using complexity analysis, simple computation of the combinatorial sum was determined to be $O(n^{2/c})$ while the main formula was $O(c \cdot \log(n))$.</p> <p>Conclusions/Discussion The project successfully determined an efficient formula for computing specific combinatorial series. The increased efficiency is especially important in large computations in applications such as encryption in computer science.</p>	
Summary Statement This project seeks to use the Binomial Theorem and roots of unity to prove an efficient formula for calculating combinatorial sums.	
Help Received n/a	