

VIPUL SINGH – Applying to MS program, Computer Science at Carnegie Mellon University

My fascination with computer science stems from an innate desire to be good and quick at reasoning and solving puzzles. In fact, the first non-trivial algorithms I designed were to solve 9X9 Sudokus. I have been constantly evolving and modifying my own techniques there and today stand as one of the top 25 solvers in India. Probabilities, number theoretical properties, statistics and mathematical modelling amuse me and form the basis for my interest in **machine learning, natural language processing** and theoretical computer science, **quantum computation** in particular.

At the school level, my proficiency at maths and science helped me excel in many Olympiads. I featured in the National Top 35 in Astronomy and Chemistry in 2010, and ranked 41st in the world at the **International Physics Olympiad** in Croatia, bagging a silver medal. Based on my outstanding performance in the test for the coveted **KVPY** scholarship, I was chosen as a part of the 30 member Indian contingent to the **Asian Science Camp** 2009, Japan where I got to interact with 7 Nobel laureates.

I entered the Bachelor of Technology program in Computer Science at IIT Bombay after ranking **5th in India in IIT-JEE 2010** and have been **Department Rank 1** throughout. I shall be completing a minor degree in Mathematics and have taken up extra courses on Probability Theory, Computational Algebra, Quantum Mechanics, Convex Optimization etc. to build the foundations for topics that interest me. I've derived a lot of confidence from my exploits so far and hope to keep up my consistency with persistent efforts.

I am working on a project titled '**Applications of Quantum Computing Principles to Natural Language Processing (NLP)**' as my undergraduate thesis under the guidance of **Prof. Pushpak Bhattacharyya**. In many tasks in NLP, we've run into efficiency roadblocks. E.g.: the Word Sense Disambiguation (WSD) Problem was conceived long ago but the best known algorithms till date don't provide accuracy better than 65-70%. In this thesis, we seek to study whether quantum computing can be used to give efficient algorithms for such problems which are part of NLP. We have extensively studied various search tasks and algorithms in the NLP domain such as iterative scaling, swarm intelligence, Hidden and Maximum Entropy Markov Models and their Viterbi algorithms. The use of quantum computing in any of these entails the need for us to express the initial configuration as a quantum state and then evolve it with time according to some operator that builds a high coefficient for the state that we ultimately wish to measure. We now intend to delve deeper into the **application of Grover search to the classical Viterbi algorithm**.

In the summer of 2012, I took up a research internship with **Prof. Thomas Henzinger** and Dr. Ashutosh Gupta at **IST Austria** on **verification of the delayed Continuous Time Markov Chain (CTMC) model** applied to gene regulatory circuits. Delayed CTMCs are an extension of CTMCs that allow for the specification of a lower bound on the time interval between an event's initiation and its completion. I proved that there exists an optimum discretization for the model when applied to a 1-particle system that minimizes the error in approximating a continuous process by a discrete model. One major lesson from this internship for me was on how to work in a large research group, communicate my ideas to more experienced people and be open to their suggestions in order to avoid being stuck on some issue for long.

Before this, I had spent a few weeks in December 2011 working on **quantum entanglement** with **Prof. Prasanta Panigrahi** at IISER Kolkata. I came across various aspects of quantum computing such as the No-cloning and No-deletion theorems. My task was to quantify entanglement and we came up with ideas like **(i) representation of the state as a point in an n-dimensional cube and looking at distance from centre (ii) using K-map reduction**. Further, we tried to explain the failure of W-state in teleportation, using this quantification. As part of a 3-member team, I had earlier developed a *debugger-cum-simulator for building quantum networks* along with a macro-based language layer to input functions using scientific symbols and implemented algorithms such as Grover's, Deutsch-Jozsa and Fast Fourier Transform. These projects along with my undergraduate thesis have given me lot of insight into quantum computing and its fundamentals.

Another research area that I have been exploring this year is '**Clustering in Time-Series Forecasting**' with **Prof. Bernard Menezes**. Clusters in time series pertaining to retail sales of various items - furniture, apparel, housing, beverages, etc. have been identified. I have been understanding, at a very fundamental level, the basis of this clustering. This has proved to be a challenging task, requiring a detailed study of the mathematical characteristics of various cluster-specific forecasting models. I also opted for the Foundations of Machine Learning course offered by Prof. Sunita Sarawagi this semester and achieved the maximum grade. In the project on *identifying device labels from accelerometer readings*, I obtained a better understanding of various classification and clustering methods via application on real-world data. These projects, along with the huge role of Machine Learning in NLP constitute my exposure to this vast field.

In the summer of 2013, I worked with Team Timeline Identity at **Facebook HQ, Menlo Park** and built a basic version of a new in-house application collection for the user timeline. I got a chance to try my hands on tools such as mercurial, javascript and HPHP (HipHop for PHP), a series of PHP execution engines and improvements created by Facebook. In addition, I fixed several bugs related to timeline visibility, as well as a new experimental timeline maps application. This was a huge learning curve and gave me confidence that I could work successfully in a company.

Though I was hugely delighted by the experience at Facebook and have been offered a job at Samsung, San Jose, I wish to continue with further studies to realize my goal of contributing to research. At IIT Bombay, I have been a **Teaching Assistant** for a variety of courses spanning physics, mathematics and computer science. I have thoroughly enjoyed these stints and this has led me to develop a strong liking for teaching. I plan to continue into a PhD after the Masters degree and then indulge in top-class research both as a student and later, hopefully as a professor.

At CMU, due to the plethora of available opportunities, I believe that I will be able to gain higher exposure and with it, a deeper understanding of various problems in my fields of interest. I would specifically like to focus on machine learning and language technologies. Having done a course on Linear Programming (LP), I was intrigued by **Prof. Noah Smith's** formulation of NLP tasks like parsing as Integer LP problems. During my thesis and while preparing for a seminar on 'Maximum Entropy Markov Models' that I had delivered this April as part of the course on Artificial Intelligence, I came across **Prof. Ronald Rosenfeld's** work on statistical language modelling and maximum entropy language models, as well as **Prof John Lafferty's** on probabilistic models and conditional random fields, and found them to be quite interesting. I would be delighted to work with any of them. I also wish to undertake the course on 'Quantum Computation and Quantum Information Theory' by **Prof. Robert Griffiths** and **Prof. Daniel Stahlke** from the Physics Department and work with them as part of some inter-disciplinary project.

Being a premier and esteemed institute, CMU attracts some of the best brains from all over the world. Having worked and interacted with some students during my internship at Facebook, I feel CMU is a wonderful place to be at, one that gives an individual a lot of scope for expansion and a vast multitude of research options. Being part of such an elite and diverse group would foster intellectual growth and enable me to work to my potential. It will be a privilege for me to spend a fruitful and rewarding time at the Pittsburgh campus.