

Sudip Roy

CONTACT INFORMATION

Email: sudip@cs.cornell.edu
Phone: (607) 229-2097
Address: 456 Gates Hall,
Cornell University,
Ithaca, NY 14853
Webpage: <http://www.cs.cornell.edu/~sudip/>

RESEARCH INTERESTS

- Transaction Processing
- Distributed Systems
- Scalable and Robust Diagnostics

EDUCATION

Cornell University [Aug, 2009- July, 2014 (expected)]

PhD Candidate, Computer Science (GPA 4/4)

- Advisor: Prof. Johannes Gehrke
- Thesis: “Lazy Transaction Execution Models”

Indian Institute of Technology, Kharagpur [2005-2009]

B.Tech. (Honors), Computer Science and Engineering (GPA 9.69/10)

- Advisor: Prof. P. P. Chakrabarti
- Thesis: “Qualitative Reasoning : Modeling and Reasoning about Incomplete Qualitative Temporal Information” (Best Undergraduate Thesis)
- Ranked 2nd in the graduating class

HONORS AND AWARDS

- Yahoo! Graduate Teaching Award, 2013
- Runner-up at Cornell Startup Weekend, 2012
- **SIGMOD 2011 Best Paper Award** for “Entangled Queries: Enabling Declarative Data-driven Coordination”
- Certificate of Achievement for outstanding contributions as a Teaching Assistant, Fall 2010
- Autodesk Fellowship, Cornell University, 2009-2010
- Bigyan Sinha Memorial Award, 2009
- O. P. Jindal Engineering and Management Scholarship, 2008-2009
- IIT Kharagpur Alumni (California Chapter) Award & Class of 1970 Alumni (US) Association Prize, 2007
- Rajendranath Das Memorial Award, 2006
- National Merit Scholarship (NTSE, India), 2003-2009

RESEARCH AND WORK EXPERIENCE

- Research Assistant, Prof. Johannes Gehrke, Cornell University, 2009-present
- Research Intern, Christian König, Data Management, Exploration and Mining Group, Microsoft Research, Summer 2013
- Research Intern, Jayant Madhavan, Structured Data Group, Google Research, Summer 2012
- Research Intern, Jayant Madhavan and Hector Gonzalez, Structured Data Group, Google Research, Summer 2012
- Research Intern, Alan Demers and Johannes Gehrke, Cornell University, Summer 2008
- Research Intern, Prof. Thambipillai Srikanthan, Nanyang Technological University, Singapore, Summer 2007
- Software Developer for DataResolve Technologies (part time), IIT Kharagpur, 2007-2008

ACADEMIC
PROJECTS

- **Homeostasis** 2013 - ongoing
Datastores today rely on distribution and replication to achieve improved performance and fault-tolerance. But the correctness of many applications depends on strong consistency properties — something that can impose substantial overheads, since it requires coordinating the behavior of multiple nodes. Homeostasis is a transaction processing system which minimizes the amount of coordination required among the nodes of a system without resulting in any external inconsistency. The key insight is to allow the state of the system to be internally inconsistent during execution, as long as this inconsistency is bounded and does not affect transaction correctness. In contrast to previous work, our approach uses program analysis to extract semantic information about permissible levels of inconsistency and is fully automated. [*Paper in submission*]
- **Quantum Databases** 2012 - 2013
In many applications, transactions compete for scarce physical or virtual resources represented by data items in the database. Quantum databases allows to defer the making of choices in transactions until an application or user forces the choices by observation. Such latent binding allows more transactions to succeed in environments with high contention.[CIDR '13]
- **Youtopia** 2010 - 2012
The rise of Web 2.0 has fueled the growth of applications to not only create, share and consume data, but also achieve complex tasks through online coordination. We proposed entangled queries and entangled transactions as two fundamental primitives for supporting data-driven coordination in databases. Our system, Youtopia, supports such primitives and provides scalable implementation of efficient algorithms for coordination.[TODS '12, VLDB '11, SIGMOD '11, SIGMOD Record '10]
- **Qualitative Reasoning** 2008 - 2009
In this project, we proposed a novel formal model to capture the semantics of complex processes and environments qualitatively. Using such model and inference rules bases on qualitative calculus and temporal logic, our tool enables efficient exploration of the qualitative behavior of the represented system. (Awarded Best Undergraduate Thesis)
- **Physical design for QCA circuits** 2007 - 2008
We designed efficient automatic partitioning and placement algorithm for synthesizable Quantum-dot Cellular Automata (QCA). We also synthesized various types of linear and additive CA rules using QCA cells.[GLSVLSI '08, HiPC '07]

INTERNSHIP
PROJECTS

- **Argus** Summer, 2013
Argus is a diagnostic tool which identifies anomalous behavior with respect to key performance indicators from telemetry data, and helps formulate data-driven hypotheses leading to root-cause diagnosis. Two key characteristics of the mining algorithms used in Argus are (i) robustness to the presence of outliers in the data and (ii) an ingrained notion of diversity to avoid presenting near-identical anomalies to the user. Argus is currently under active development and is used to find anomalies in telemetry data of Windows Azure. [*Working paper*]
- **Faceted Navigation for Google Fusion Tables** Summer, 2012
In this project, we designed and implemented offline and online algorithms for recommending interesting facets of structured data to a user automatically. Such recommendations assist in easy navigation and exploration of the data by non-expert users.
- **Spatial Query Processing for Google Fusion Tables** Summer, 2011
Visualization of large spatial datasets is often exploratory and imposes stringent requirements on data retrieval response times. We designed and implemented suitable spatial index structures which can leverage visualization constraints to improve query processing latencies. The implementation is currently deployed in Google Fusion Tables. [*Working paper*]

PUBLICATIONS

- **Writes that Fall in the Forest and Make no Sound: Semantics-Based Adaptive Data Consistency.**
Sudip Roy, Lucja Kot, Nate Foster, Johannes Gehrke, Hossein Hojjat, Christoph Koch
In *Computing Research Repository* (CoRR 2014)
- **Quantum Databases.**
Sudip Roy, Lucja Kot, Christoph Koch
In *Proceedings of 6th Biennial Conference on Innovative Data Systems Research* (CIDR 2013), Asilomar, CA
- **Entangled Queries: Enabling Declarative Data-Driven Coordination.**
Nitin Gupta, Lucja Kot, Sudip Roy, Gabriel Bender, Johannes Gehrke, Christoph Koch
In *ACM Transactions on Database Systems* (TODS 2012)
- **Entangled Transactions.**
Nitin Gupta, Milos Nikolic, Sudip Roy, Gabriel Bender, Johannes Gehrke, Christoph Koch
In *Proceedings of 37th International Conference on Very Large Data Bases* (VLDB 2011), Seattle, WA
- **Entangled Queries: Enabling Declarative Data-Driven Coordination.** (Best Paper)
Nitin Gupta, Lucja Kot, Sudip Roy, Gabriel Bender, Johannes Gehrke, Christoph Koch
In *Proceedings of the ACM SIGMOD International Conference on Management of Data* (SIGMOD 2011), Athens, Greece
- **Declarative Data-Driven Coordination in the Youtopia System.** (Demo)
Nitin Gupta, Lucja Kot, Sudip Roy, Gabriel Bender, Johannes Gehrke, Christoph Koch
In *Proceedings of the ACM SIGMOD International Conference on Management of Data* (SIGMOD 2011), Athens, Greece
- **Beyond Isolation: Research Opportunities in Declarative Data-Driven Coordination.**
Lucja Kot, Nitin Gupta, Sudip Roy, Johannes Gehrke, Christoph Koch
In *SIGMOD Record 2010*
- **A Layout-aware Physical Design Method for Constructing Feasible QCA Circuits.**
Mayur Bubna, Sudip Roy, Naresh Shenoy, Rajib Mall
In *Proceedings of the 18th Great Lake Symposium on VLSI* (GLSVLSI 2008), Orlando, Florida
- **Designing Cellular Automata Structures using Quantum-dot Cellular Automata.**
Mayur Bubna, Subhra Mazumdar, Sudip Roy, Rajib Mall
In *Proceedings of the 14th annual IEEE International Conference on High Performance Computing* (HiPC 2007), Goa, India

TEACHING
EXPERIENCE

- Courses TA'ed at Cornell University
- CS5320/CS4320: Big Data
 - CS5321/CS4321: Practicum in Database Systems
 - CS4320: Introduction to Database System
 - CS4321: Practicum in Database Systems

REFERENCES

Johannes Gehrke
Professor, Cornell University
Ithaca, NY 14853
johannes@cs.cornell.edu

Christian König
Researcher, Microsoft Research
Redmond, WA 98052
chrisko@microsoft.com

Christoph Koch
Professor, EPFL
Lausanne, Switzerland
christoph.koch@epfl.ch

Jayant Madhavan
Staff Software Engineer, Google Inc.
Mountain View, CA 94043
jayant@google.com