Peter Bailis

UC Berkeley EECS
Soda Hall 489-3
Berkeley, CA 94704

http://bailis.org/
pbailis@cs.berkeley.edu

INTERESTS

I am interested in data management, data-intensive computing, and distributed systems.

In my dissertation work, I have developed high-performance data management systems and algorithms that mitigate the fundamental trade-offs between coordination and programmability. My recent interests also include asynchrony in large-scale statistical learning and prediction.

EDUCATION

University of California, Berkeley *2011-2015 (expected)*
Ph.D. candidate, Computer Science.
Advisors: Joseph M. Hellerstein, Ali Ghodsi, Ion Stoica

Harvard College *2007-2011*
A.B. Computer Science.
Magna cum laude with Highest Honors in Computer Science

AWARDS & HONORS

- "Best of SIGMOD 2014" invitation to ACM TODS
- Communications of the ACM Research Highlight, 2014
- UC Berkeley EECS David J. Sakrison Memorial Prize, 2014-2015
- UC Berkeley Outstanding Graduate Student Instructor Award, 2014-2015
- "Best of VLDB 2012" invitation to the VLDB Journal
- National Science Foundation Graduate Research Fellowship, 2011-2016
- Berkeley Fellowship for Graduate Study, 2011-2014
- UC Berkeley EECS Chair's Excellence Award, 2011-2012
- Computing Research Association Outstanding Undergraduate Researcher, 2011
- Thomas Temple Hoopes Prize for Outstanding Senior Thesis, 2011
- Phi Beta Kappa, Harvard College Chapter, inducted 2011
- Best Student Paper, ANTS 2010
- Derek Bok Center Certificate for Distinction in Teaching, 2009

PROJECTS

Graduate (September 2011–Present)

Large-scale and distributed data management:

- INVARIANT CONFLUENCE is a necessary and sufficient condition for correct, coordination-free execution of distributed transactions that enables safe optimization of many DBMS workloads.
- My work on FERAL CONCURRENCY CONTROL examines the coordination requirements of common open-source applications as well as their incidence of data inconsistency due to races.
- RAMP TRANSACTIONS provide scalable multi-get, multi-put, secondary index update and retrieval, and materialized view maintenance without the use of blocking coordination.
- HIGHLY AVAILABLE TRANSACTIONS are the set of widely-deployed isolation guarantees that can be provided without sacrificing coordination-free execution or low latency.
- BOLT-ON CONSISTENCY upgrades the safety guarantees of underlying eventually consistent stores via a shim that provides causal ordering while maintaining availability of operation. EXPLICIT CAUSALITY helps limit causality tracking overheads via programmer annotations.
- PROBABILISTICALLY BOUNDED STALENESS quantifies the inconsistency of eventually consistent stores by providing an expectation on staleness in terms of versions and real time.
- I have performed two practitioner-oriented surveys on the incidence of NETWORK FAILURES in real system deployments as well as the use of EVENTUAL CONSISTENCY in the wild.

Scalable and coordination-avoiding machine learning systems:

- The BAP interface and implementation (PLASMA) provides support for asynchronous execution of convex programming tasks within existing distributed dataflow systems.
- VELOX is a nascent system to which I have contributed that provides scalable serving of data products—statistically-intensive machine learning tasks and predictions—with low latency.

Undergraduate (March 2009–May 2011)

With Professors Margo Seltzer, David Brooks, Radhika Nagpal, Matt Welsh, Dr. Vijay Reddi

- DIMETRODON uses operating system scheduler-level idle cycle injection for processor cooling.
- KARMA is an operating system for programming swarms of unmanned aerial vehicles.
- Within the RoboBees project, I studied the benefits of waggle dancing in honeybee foraging.

PUBLICATIONS

Journal Articles

- Peter Bailis, Shivaram Venkataraman, Michael J Franklin, Joseph M. Hellerstein, Ion Stoica.
  Quantifying Eventual Consistency with PBS.
  The VLDB Journal, April 2014, 23(2):279–302.
  "Best of VLDB 2012" Special Issue.

Conference Publications

- Peter Bailis, Alan Fekete, Michael J. Franklin, Ali Ghodsi, Joseph M. Hellerstein, Ion Stoica.
  Feral Concurrency Control: An Empirical Investigation of Modern Application Integrity.
  ACM SIGMOD 2015.
- Peter Bailis, Alan Fekete, Michael J. Franklin, Ali Ghodsi, Joseph M. Hellerstein, Ion Stoica.
  Coordination Avoidance in Database Systems.
  VLDB 2015.
- Dan Crankshaw, Peter Bailis, Joseph E. Gonzalez, Haoyuan Li, Zhao Zhang, Michael J. Franklin, Ali Ghodsi, Michael I. Jordan.
  The Missing Piece in Complex Analytics: Low Latency, Scalable Model Management and Serving with Velox.
  CIDR 2015.
- Peter Bailis, Alan Fekete, Ali Ghodsi, Joseph M. Hellerstein, Ion Stoica.
  Scalable Atomic Visibility with RAMP Transactions.
  ACM SIGMOD 2014.
  INVITED TO "BEST OF SIGMOD 2014" ACM TODS ISSUE.
- Peter Bailis, Aaron Davidson, Alan Fekete, Ali Ghodsi, Joseph M. Hellerstein, Ion Stoica.
  Highly Available Transactions: Virtues and Limitations.
  VLDB 2014.
- Peter Alvaro, Peter Bailis, Neil Conway, Joseph M. Hellerstein.
  Consistency Without Borders.
  ACM SoCC 2013 (Vision Track).
- Peter Bailis, Ali Ghodsi, Joseph M. Hellerstein, Ion Stoica.
  Bolt-on Causal Consistency.
  ACM SIGMOD 2013.
- Peter Bailis, Alan Fekete, Ali Ghodsi, Joseph M. Hellerstein, Ion Stoica.
  The Potential Dangers of Causal Consistency and an Explicit Solution.
  ACM SoCC 2012 (Vision Track).
- Peter Bailis, Shivaram Venkataraman, Michael J. Franklin, Joseph M. Hellerstein, Ion Stoica.
  Probabilistically Bounded Staleness for Practical Partial Quorums.
  VLDB 2012.
  INVITED TO "BEST OF VLDB 2012" VLDBJ ISSUE.
  SELECTED AS CACM RESEARCH HIGHLIGHT.

- Peter Bailis, Vijay Janapa Reddi, Sanjay Gandhi, David Brooks, Margo Seltzer.
  Dimetrodon: Processor-level Preventive Thermal Management via Idle Cycle Injection.
  DAC 2011.
- Karthik Dantu, Bryan Kate, Jason Waterman, Peter Bailis, Matt Welsh.
  Programming Micro-aerial Vehicle Swarms with Karma.
  ACM SenSys 2011.
- Peter Bailis, Radhika Nagpal, Justin Werfel.
  Positional Communication and Private Information in Honeybee Foraging Models.
  ANTS 2010.
  Awarded BEST STUDENT PAPER.

Workshop Publications

- Peter Bailis, Alan Fekete, Ali Ghodsi, Joseph M. Hellerstein, Ion Stoica.
  HAT, not CAP: Towards Highly Available Transactions.
  HotOS 2013.

Theses

- Peter Bailis.
  Dimetrodon: Processor-level Preventive Thermal Management via Idle Cycle Injection.
  Harvard College Undergraduate Honors Thesis, 2011.
  Awarded THOMAS TEMPLE HOOPES PRIZE.

Demonstrations

- Peter Bailis, Shivaram Venkataraman, Michael J. Franklin, Joseph M. Hellerstein, Ion Stoica.
  PBS at Work: Advancing Data Management with Consistency Metrics.
  ACM SIGMOD 2013.

Invited Articles

- Peter Bailis and Kyle Kingsbury.
  The Network is Reliable: An Informal Survey of Real-World Communications Failures.
  ACM Queue, 12(7), July 2014.
  Also appears in *Communications of the ACM* 57(9):48–55, September 2014.
- Peter Bailis, Shivaram Venkataraman, Michael J. Franklin, Joseph M. Hellerstein, Ion Stoica.
  Quantifying Eventual Consistency with PBS.
  Communications of the ACM (Research Highlight), 57(8):93–102, August 2014.
- Peter Bailis and Ali Ghodsi.
  Eventual Consistency Today: Limitations, Extensions, and Beyond.
  ACM Queue, 11(3), March 2013.
  Also appears in *Communications of the ACM* 56(3):55–63, May 2013.

Under review

- Peter Bailis, Joseph E. Gonzalez, Ali Ghodsi, Michael J. Franklin, Joseph M. Hellerstein, Michael I. Jordan, Ion Stoica.
  Asynchronous Complex Analytics in a Distributed Dataflow Architecture.
- Peter Bailis, Alan Fekete, Ali Ghodsi, Joseph M. Hellerstein, Ion Stoica.
  Scalable Atomic Visibility with RAMP Transactions (Invited journal article).

ACADEMIC TALKS

- **The Missing Piece in Complex Analytics: Low Latency, Scalable Model Management and Serving with Velox**
  CIDR 2015 (with Joey Gonzalez), January 2015, Asilomar, CA
- **The Case for Invariant-Based Concurrency Control**
  CIDR 2015 (short talk), January 2015, Asilomar, CA

- **Highly Available Transactions**
  VLDB 2014, September 2014, Hangzhou, China
  HotOS 2013, May 2013, Santa Ana Pueblo, NM
  CIDR 2013 (short talk), January 2013, Asilomar, CA
- **Velox: Serving the Berkeley Data Analytics Stack**
  AMPLab Summer Retreat (with Joey Gonzalez), May 2014, Santa Cruz, CA
  ASPIRE Lab All-Hands, April 2014, Berkeley, CA
- **Scalable Atomic Visibility with RAMP Transactions**
  ACM SIGMOD 2014, June 2014, Snowbird, UT.
- **Bridging the Gap: Opportunities in Coordination-Avoiding Database Systems**
  PaPEC Workshop (at EuroSys 2014), April 2014, Amsterdam, Netherlands.
- **Database Research at Berkeley**
  NorCal Database Day April 2014, Almaden, CA
- **Coordination and Consistency in Distributed Databases**
  BEARS Research Symposium (short talk), February 2014, Berkeley, CA
  AMPLab Winter Retreat, January 2014, Tahoe City, CA
  SCDM Workshop (at IEEE BigData) Invited Keynote, October 2013, Santa Clara, CA
  HPTS 2013, "Next Generation" panel, September 2013, Asilomar, CA
- **Bolt-on Causal Consistency**
  ACM SIGMOD 2013, June 2013, New York, NY
  AMPLab Winter Retreat, January 2013, Olympic Valley, CA
- **Probabilistically Bounded Staleness for Practical Partial Quorums**
  UC Irvine ISG Seminar, November, 2012, Irvine, CA
  VLDB 2012, August 2012, Istanbul, Turkey
  Harvard University SYRAH Seminar, March 2012, Cambridge, MA
- **The Potential Dangers of Causal Consistency and an Explicit Solution**
  ACM SoCC 2012, October 2012, San Jose, CA
- **Dimetrodon: Processor-level Preventive Thermal Management via Idle Cycle Injection**
  DAC 2011, June 2011, San Diego, CA
- **Positional Communication and Private Information in Honeybee Foraging Models**
  ANTS 2010, September 2010, Brussels, Belgium

INDUSTRY RESEARCH TALKS

- **Coordination and the Art of Scaling**
  SAP AG, August 2014, Dublin, CA
  CloudantCON 2014, June 2014, San Francisco, CA
- **Scalable Atomic Visibility with RAMP Transactions**
  Next-Generation Cassandra Conference (Invited Talk), June 2014, Austin, TX
- **Coordination and Consistency in Distributed Databases**
  SF Bay Area ACM Chapter Meeting, February 2014, Palo Alto, CA
  Apple, November 2013, Cupertino, CA
  RICON West 2013, October 2013, San Francisco, CA
- **Highly Available Transactions**
  Cisco, June 2013, San Jose, CA
  Cloudera, May 2013, San Francisco, CA

- **Bring Your Own ACID**
  RICON West 2012 (short talk) October 2012, San Francisco, CA

- **Probabilistically Bounded Staleness for Practical Partial Quorums**
  BashoChats 002, February 2012, San Francisco, CA
  HP Labs, May 2012, Palo Alto, CA
  CloudPhysics, June 2012, San Mateo, CA
  Facebook, June 2012, Menlo Park, CA
  LinkedIn, June 2012, Mountain View, CA
  Twitter, June 2012, San Francisco, CA
  Yelp, August 2012, San Francisco, CA

OTHER OUTREACH & TALKS

- **Papers We Love: Managing Update Conflicts in Bayou**
  PWL San Francisco December 2014, San Francisco, CA

- **ACM Queue Editorial Board**
  Guest, June and August 2014, San Francisco, CA
  Guest Expert, Eventual Consistency, February 2012, San Francisco, CA

- **Think Distributed Podcast**
  Live from RICON episode, December 2013
  Causality episode, August 2013
  Consensus episode, July 2013

- **Guest Lecture: Storage Systems**
  Berkeley I-School 253, October 2012, Berkeley, CA

LEADERSHIP & SERVICE

Berkeley Database Group *July 2013-present*
Organized weekly discussion seminar, including invited talks and paper readings, for Berkeley Database Research group (~8-15 Ph.D. students and faculty; $>$ 70 meetings), webmaster.

Berkeley Cloud Seminar *Fall 2014 semester-present*
Organized and curated weekly external research talk series on applied machine learning and large-scale cloud computing systems (30-50+ students and faculty; see cloud.berkeley.edu).

Berkeley EECS Graduate Admissions Committee *2014*
Database area student representative for Ph.D. and master's admissions. Reviewed applicant profiles, personally presented database area admission recommendation to committee.

Berkeley EECS Graduate Admissions Visit Day *2012-2014*
Operating Systems, Networking, and Databases visit day coordinator, 2014. Databases area peer advisor for prospective Ph.D. students, 2012-2014 (2 of 3 advisees attending, one deferred).

Berkeley Computer Science Graduate Students Association *2012-2013*
Elected Social Committee co-chair. Organized weekly Social Hours and semi-regular outings.

TinyToCS: Tiny Transactions on Computer Science *2012-present*
Co-founder and program committee member Volumes I, II and III. Peer-reviewed, 140 character paper summaries disseminated via social media (see @TinyToCS).

Program Committee: SoCC 2015, SIGMOD 2016 (Regular and Demos)
Editorial Board: Foundations and Trends in Databases (FnTDB, since 2015)
Organizing Committee: Workshop on Principles, Practice of Consistency (PaPoC, EuroSys 2015)
Organizing & Program Committee: Workshop on Principles, Practice of Eventual Consistency (PaPEC, EuroSys 2014)

External reviewer: NSDI '14, EuroSys '13, SIGCOMM '13, OSDI '12, ASPLOS '11, and others

TEACHING & MENTORING

Graduate Student Instructor, UCB Graduate Database Systems (CS 286) *Fall 2014*
Responsible for syllabus creation, occasional lectures, course research project advising, and assistance in revision of "Red Book" (*Readings in Database Systems*, 4th Edition; curating reading list and regular conference calls with MIT DB group). Student effectiveness rating: 4.9/5.0

Research mentor, undergraduate Aaron Davidson (graduated Dec. 2013) *09/2012-01/2014*
Aaron and I worked together to extend Probabilistically Bounded Staleness to account for partial failures and to implement Highly Available Transactions (on which he is second author). Aaron has deferred (unanimous) Ph.D. admissions and was Databricks's first engineering hire.

Research mentor, undergraduate Anirudh Todi (graduated Dec. 2012) *09-12/2012*
I worked with Anirudh to extend PBS. Anirudh now works on time series aggregation at Twitter.

Graduate Student Instructor, UCB Undergraduate Databases (CS 186) *Spring 2012*
Responsible for designing section material and projects, office hours, grading, and leading two 30-student weekly discussion sections. Student effectiveness rating: 3.9/5.0

Teaching Fellow, Harvard Computer Science I (CS 50) *Fall 2008*
Responsible for weekly section of 12 students, office hours, grading of problem sets and exams. Student effectiveness rating: 4.7/5.0. Received Certificate of Distinction in Teaching.

INDUSTRIAL EMPLOYMENT

Intern, Distributed Data Infrastructure, LinkedIn *Summer 2012*
I designed and implemented multi-datacenter, multi-master replication and conflict resolution in LinkedIn's Espresso database. I also developed staleness analysis for LinkedIn Voldemort.

Engineering Intern, Microsoft Server and Tools Division *Summer 2009*
I developed a UML XMI-based import/export tool for "Oslo" SQL Server Metadata repository.

REFERENCES

Joseph M. Hellerstein (co-advisor)
Chancellor's Professor of Computer Science, UC Berkeley
Chief Strategy Officer and Co-founder, Trifacta
hellerstein@cs.berkeley.edu

Ali Ghodsi (co-advisor)
Researcher, Computer Science, UC Berkeley
Vice President of Engineering and Co-founder, Databricks
alig@cs.berkeley.edu

Ion Stoica (co-advisor)
Professor, Computer Science, UC Berkeley
CEO and Co-founder, Databricks
istoica@cs.berkeley.edu

Michael J. Franklin
Thomas M. Siebel Professor of Computer Science
and Chair, Computer Science Division, EECS, UC Berkeley
franklin@cs.berkeley.edu

Eric Brewer
Professor, Computer Science
and Jim Gray Chair in Engineering, UC Berkeley
Vice President of Infrastructure, Google
brewer@cs.berkeley.edu

Samuel Madden
Professor, Computer Science, MIT CSAIL
madden@csail.mit.edu