

Lipyeow Lim

Curriculum Vitae

Home Address

2339 Valorie St.
Fremont, CA 94539, USA.
(808) 203-8018

Office Address

1680 East West Rd, POST 317
Honolulu, HI 96822, USA.
(808) 956-3495

email: lipyeow@hawaii.edu

<http://www2.hawaii.edu/~lipyeow>

Employment

- Jan 2019–Present *Senior Staff Software Engineer.*
FireEye, Inc.
- Aug 2014–Present *Associate Professor.*
Information and Computer Sciences Department
University of Hawai‘i at Mānoa, HI, U.S.A.
- Mar 2012– Dec 2018 *Short Term Consultant*
Social protection & health projects in Mongolia, China, Malaysia, and Vietnam
The World Bank
- Aug 2009–Jul 2014 *Assistant Professor.*
Information and Computer Sciences Department
University of Hawai‘i at Mānoa, HI, U.S.A.
- Oct 2004–Jul 2009 *Research Staff Member.*
IBM T. J. Watson Research Center, NY, U.S.A.

Education

- Aug 2004 *Ph.D., Computer Science*
Department of Computer Science, Duke University, NC, U.S.A.
- Dissertation: Online methods for database optimization.
Advisor: Jeffrey Scott Vitter
Committee: Min Wang, Ronald Parr, and Jun Yang
GPA: 3.9
- Sep 2000 *M.Sc., Information Systems and Computer Science*
School of Computing, National University of Singapore, Singapore.
- Dissertation: A theoretical look at pixel ordering.
Advisor: Philip M. Long
Committee: Ghim-Hwee Ong and Wee-Sun Lee
- May 1999 *B.Sc. (accelerated 3-year honors), Information Systems and Computer Science*
School of Computing, National University of Singapore, Singapore.
- Thesis Project: Implementation of the mobile IPv4 configuration option for PPP IPCP (RFC 2290).
Advisor: Yong-Chiang (Y. C.) Tay
Committee: Kam-Hong Shum and Kwok-Yan Lam
GPA: 3.5

Research Experience

- Aug 2014–Present *Associate Professor.*
University of Hawai'i at Mānoa, HI, U.S.A.
Research Area: Data Management Systems, Big Data
- Aug 2009–Jul 2014 *Assistant Professor.*
University of Hawai'i at Mānoa, HI, U.S.A.
Research Area: Data Management Systems
- Oct 2004–Jul 2009 *Research Staff Member.*
IBM T. J. Watson Research Center, NY, U.S.A.
Research Area: Optimizing Data Management Systems
- Sep 2000–Sep 2004 *Research Assistant (Advisor: Prof. Jeffrey Scott Vitter).*
Department of Computer Science, Duke University, NC, U.S.A.
Research Area: On-line Methods for Database Optimization
- Jun–Sep 2003 *Research Intern (Mentor: Min Wang)*
IBM T. J. Watson Research Center, NY, U.S.A.
Research Topic: Statistics Collection for IBM DB2XML Optimization.
- Jun–Aug 2001 *Research Intern (Mentor: Min Wang)*
IBM T. J. Watson Research Center, NY, U.S.A.
Research Topic: On-line Statistics for Native XML Databases. (Published in VLDB 2002).
- May–Aug 2000 *Research Intern (Mentor: Min Wang)*
IBM T. J. Watson Research Center, NY, U.S.A.
Research Topic: Keeping Web Indexes Up-to-date Using The Landmark-Diff Method. (Published in WWW 2003).
- Jul 1998–Aug 1999 *Research Assistant (Advisor: Prof. Philip M. Long).*
School of Computing, National University of Singapore, Singapore.
Research Area: Pixel Ordering in Image Compression.
- Jul 1997–Jun 1998 *Member of the Mobile-IP Research Group (Advisor: Prof. Y. C. Tay).*
School of Computing, National University of Singapore, Singapore.
Research Area: Inter-operability of the PPP and MobileIP networking protocols.

Teaching Experience

- Aug 2009–Present *Assistant & Associate Professor.*
 University of Hawai‘i at Mānoa, HI, U.S.A.
 ICS 321: Data Storage & Retrieval
 ICS 421: Database Systems
 ICS 491: Big Data Analytics
 ICS 621: Algorithms
 ICS 624: Advanced Data Management
 CID 705/ ICS 691E: Data Science Foundations
- Jan–Apr 2001 *Teaching Assistant* (to Prof. Curry Guinn)
 Department of Computer Science, Duke University, NC, U.S.A.
 CPS 170: Artificial Intelligence (Undergraduate level).
- Sep–Dec 2000 *Teaching Assistant* (to Prof. Lars Arge)
 Department of Computer Science, Duke University, NC, U.S.A.
 CPS 230: Analysis of Algorithms (Graduate level).
- Jan–Jun 1999 *Teaching Assistant* (to Prof. Kian-Lee Tan)
 School of Computing, National University of Singapore, Singapore.
 CS2102: Introduction to Database Systems
- Jan–Jun 1998 *Teaching Assistant* (to Prof. Weng-Fai Wong)
 School of Computing, National University of Singapore, Singapore.
 IC1273: C Programming & Unix Tools

Students Mentored

Undergraduate Students Advised

1. Glenn Galvizo (Spring & Fall 2017)
2. Derek Chan (Spring 2017)
3. Robert Patch (Fall 2014)
4. Kevon Adonis (Spring 2014)
5. Jasmine Ishigami (Spring 2014)
6. Nicholas Davis (Fall 2013)
7. Nathan Kanekuni (Fall 2013)
8. Kevon Adonis (Fall 2013)
9. Kyle Mulleady (Spring 2013)
10. Alexander Cam Liu (Fall 2012)
11. Joshua Kulhavy Sutherland (Fall 2012)
12. Christopher Foo (Fall 2011, Spring 2012)
13. Zachary Heilbron (Fall 2010)
14. Joseph Gum (Fall 2010)

Graduate Students Advised

1. Jonas Krause (Ph.D. 2020). Thesis: CNN-based Plant Species Categorization Using Natural Images.

2. Kelsea Hosoda (CIS Ph.D. 2019). Thesis: Hawaiian Morphemes: Identification, Usage, and Application in Information Retrieval.
3. Gavin Sujita (M.S. 2018). Project: Classifying Recidivism using LSI, ASUS, and Machine Learning
4. Ling-chih Yao (M.S. 2019). Project: Parallelizing String Similarity Join Algorithms.
5. Tianli Mo (Ph.D. 2019). Thesis: Energy-efficient query processing on mobile devices.
6. Mariam Doliashvili (M.S. 2017). Project: Predicting Emoji Usage for Emoji Recommender System.
7. Michael Sommer (M.S. 2017). Project: Classifying Job Post Text Documents with the Malaysian Standard Classification of Occupations.
8. Carlos Andrade (M.S. 2016). Thesis: Probabilistic Models for One-Day Ahead Solar Irradiance Forecasting in Renewable Energy Applications.
9. Todd Taomae (M.S. 2015). Project: Short-Term Solar Irradiance Forecasting and Weather Analysis using Gridded Data.
10. Alexander Cam Liu (M.S. 2015). Project: An Energy Efficient Approach to Indoor Localization
11. Christopher Foo (M.S. 2014). Project: Data Analytics for Solar Energy Management.
12. Randall Parabolicoli (M.S. 2014). Project: Data Analytics for Gamification.
13. Christopher Mullins (M.S. 2013). Thesis: Query-Aware Compression of Join Results.
14. Andrea Connell (M.S. 2012). Project: Towards Continuous Weather Forecasting Using Stream Computing.
15. Benjamin Karsin (M.S. 2011, co-advised with Henri Casanova). Thesis: Parallel XPath Query Evaluation on Multi-core Processors.

Graduate Students (Thesis Committee Member)

1. Benjamin Karsin (Ph.D. ICS)
2. Andre Pattantyus (Ph.D Meteorology)
3. Mike Lum (Ph.D Astronomy),
4. Matt Wolff (Ph.D ICS)
5. Michael Gowanlock (Ph.D ICS)
6. Anthony Christe (M.S. ICS)
7. Serge Negrashov (M.S. ICS)
8. David Schanzenbach (M.S. ICS)
9. Lyneth Peou (M.S. ICS).

Honors and Awards

- | | |
|-----------|---|
| 2008 | <i>Best Paper Award,</i>
ER 2008 Conference. |
| 2007 | <i>Outstanding Technical Achievement Award — in appreciation for Watson Contribution to DB2/XML</i>
International Business Machines. |
| 2003–2004 | <i>IBM PhD Fellowship.</i>
International Business Machines. |
| 2002/2003 | <i>Outstanding Service to the Department Award.</i>
Department of Computer Science, Duke University, NC, U.S.A. |
| 1999/2000 | <i>Graduate Fellowship.</i>
Department of Computer Science, Duke University, NC, U.S.A. |

- 1995/1996 *On the Dean's List.*
School of Computing, National University of Singapore, Singapore.
- 1996/1997 *On the Dean's List.*
School of Computing, National University of Singapore, Singapore.

Publications

1. *An Experimental Survey of Evaluation Strategies for Constellation Queries*
Glenn Galvizo and Lipyeow Lim Scientific and Statistical Database Management Conference (SSDBM 2019), 2019.
2. *A guided multi-scale categorization of plant species in natural images*
Jonas Krause, Kyungim Baek, and Lipyeow Lim CVPR Workshop on Computer Vision Problems In Plant Phenotyping (CVPPP 2019), 2019.
3. *What's That Plant? WTPlant is a Deep Learning System for Identifying Plants in Natural Images*
Jonas Krause, Lipyeow Lim, Kyungim Baek, and Gavin Sujita BMVC Workshop on Computer Vision Problems In Plant Phenotyping (CVPPP 2018), 2018.
4. *WTPlant (What's That Plant?): a Deep Learning System for Identifying Plants in Natural Images (Demonstration Paper)*
Jonas Krause, Gavin Sujita, Lipyeow Lim, Kyungim Baek ACM International Conference on Multimedia Retrieval (ICMR), 2018.
5. *Parallelizing String Similarity Join Algorithms*
Ling-Chih Yao, Lipyeow Lim Australasian Database Conference, 2018.
6. *30min-Ahead Gridded Solar Irradiance Forecasting using Satellite Data*
Todd Taomae, Lipyeow Lim, Duane Stevens, Dora Nakafuji Australasian Database Conference, 2018.
7. *A Differentially Private Matching Scheme for Pairing Similar Users of Proximity Based Social Networking applications*
Michael Sommer, Lipyeow Lim and Depeng Li 51st Hawaii International Conference on System Sciences (HICSS'18), January 2018.
8. *Cloud-based query evaluation for energy-efficient mobile sensing*
Tianli Mo, Lipyeow Lim, Sougata Sen, Archan Misra, Rajesh Krishna Balan, Youngki Lee Pervasive Mobile Computing Journal, March, 2017.
9. *Developing Data-Driven Hawaiian Language Vocabulary Lists using Preserved Documents (Abstract)*
Kelsea Hosoda, Lipyeow Lim. International Conference on Language Documentation and Conservation (ICLDC) 2017
10. *Low-latency XPath Query Evaluation on Multi-Core Processors*
Benjamin Karsin, Lipyeow Lim and Henri Casanova. HICSS 2017
11. *Probabilistic Models for One-Day Ahead Solar Irradiance Forecasting in Renewable Energy Applications*
Carlos Andrade, Lipyeow Lim, Duane Stevens, et al. ICMLA 2015

12. *Data Analytics for Solar Energy Management (Abstract)*
Lipyeow Lim, Duane Stevens, Sen Chiao, et al. ICEM 2015
13. *Cloud-based Query Evaluation for Energy-Efficient Mobile Sensing*
Tianli Mo, Sougata Sen, Lipyeow Lim, Archan Misra, Rajesh Krishna Balan and Youngki Lee MDM 2014
14. *Cost-Optimal Execution of Boolean Query Trees with Shared Streams*
Dounia Zaidouni, Yves Robert, Frederic Vivien, Henri Casanova, Lipyeow Lim International Parallel and Distributed Processing Symposium 2014
15. *Elastic Data Partitioning for Cloud-based SQL Processing Systems*
Lipyeow Lim Scalable Cloud Data Management Workshop 2013
16. *Energy Efficient Collaborative Query Processing Framework for Mobile Sensing Services.*
Jin Yang, Tianli Mo, Lipyeow Lim, Kai-Uwe Sattler and Archan Misra. IEEE 14th International Conference on Mobile Data Management (MDM) 2013.
17. *Semantic queries by example.*
Lipyeow Lim, Haixun Wang and Min Wang. EDBT 2013.
18. *Query-aware compression of join results.*
Christopher M. Mullins, Lipyeow Lim and Christian A. Lang. EDBT 2013.
19. *Adaptive data acquisition strategies for energy-efficient, smartphone-based, continuous processing of sensor streams.*
Lipyeow Lim, Archan Misra and Tianli Mo. Distributed and Parallel Databases 2012.
20. *The case for cloud-enabled mobile sensing services.*
Sougata Sen, Archan Misra, Rajesh Balan and Lipyeow Lim. Proceedings of the first edition of the MCC workshop on Mobile cloud computing 2012.
21. *Optimizing Sensor Data Acquisition for Energy-Efficient Smartphone-based Continuous Event Processing.*
Archan Misra and Lipyeow Lim. The 12th International Conference on Mobile Data Management 2011.
22. *Optimizing Access across Multiple Hierarchies in Data Warehouses.*
Lipyeow Lim and Bishwaranjan Bhattacharjee. HICSS 2011.
23. *Optimizing content freshness of relations extracted from the web using keyword search.*
Mohan Yang, Haixun Wang, Lipyeow Lim and Min Wang. SIGMOD 2010.
24. *Statistics-based parallelization of XPath queries in shared memory systems.*
Rajesh Bordawekar, Lipyeow Lim, Anastasios Kementsietsidis and Bryant Wei-Lun Kok. EDBT 2010.
25. *LinkedCT: A Linked Data Space for Clinical Trials.*
Oktie Hassanzadeh, Anastasios Kementsietsidis, Lipyeow Lim, Renée J. Miller and Min Wang. CoRR abs/0908.0567 2009.
26. *Efficient Index Compression in DB2 LUW.*
Bishwaranjan Bhattacharjee, Lipyeow Lim, Timothy Malkemus, George A. Mihaila, Kenneth A. Ross, Sherman Lau, Cathy McCarthur, Zoltan Toth and Reza Sherkat. PVLDB 2(2) 2009.
27. *Linkage Query Writer.*

- Oktie Hassanzadeh, Reynold Xin, Renée J. Miller, Anastasios Kementsietsidis, Lipyeow Lim and Min Wang. PVLDB 2(2) 2009.
28. *Challenges on Modeling Hybrid XML-Relational Databases.*
Mirella M. Moro, Lipyeow Lim and Yuan-Chi Chang. Open and Novel Issues in XML Database Applications: Future Directions and Advanced Technologies 2009.
 29. *A framework for semantic link discovery over relational data.*
Oktie Hassanzadeh, Anastasios Kementsietsidis, Lipyeow Lim, Renée J. Miller and Min Wang. CIKM 2009.
 30. *Semantic queries in databases: problems and challenges.*
Lipyeow Lim, Haixun Wang and Min Wang. CIKM 2009.
 31. *Parallelization of XPath queries using multi-core processors: challenges and experiences.*
Rajesh Bordawekar, Lipyeow Lim and Oded Shmueli. EDBT 2009.
 32. *A declarative framework for semantic link discovery over relational data (Poster).*
Oktie Hassanzadeh, Lipyeow Lim, Anastasios Kementsietsidis and Min Wang. WWW 2009.
 33. *Profile-based Retrieval of Records in Medical databases..*
Anastasios Kementsietsidis, Lipyeow Lim and Min Wang. AMIA 2009.
 34. *Modeling and Querying E-Commerce Data in Hybrid Relational-XML DBMSs.*
Lipyeow Lim, Haixun Wang and Min Wang. ER 2008.
 35. *Optimizing Hierarchical Access in OLAP Environment (Poster).*
Lipyeow Lim and Bishwaranjan Bhattacharjee. ICDE 2008.
 36. *Supporting ontology-based keyword search over medical databases..*
Anastasios Kementsietsidis, Lipyeow Lim and Min Wang. AMIA 2008.
 37. *Efficient Update of Indexes for Dynamically Changing Web Documents.*
Lipyeow Lim, Min Wang, Sriram Padmanabhan, Jeffrey Scott Vitter and Ramesh C. Agarwal. World Wide Web Journal 10(1) 2007.
 38. *Persisting and querying biometric event streams with hybrid relational-XML DBMS.*
Daby M. Sow, Lipyeow Lim, Min Wang and Kyu Hyun Kim. DEBS 2007.
 39. *Semantic Data Management: Towards Querying Data with their Meaning.*
Lipyeow Lim, Haixun Wang and Min Wang. ICDE 2007.
 40. *Supporting ranking and clustering as generalized order-by and group-by.*
Chengkai Li, Min Wang, Lipyeow Lim, Haixun Wang and Kevin Chen-Chuan Chang. SIGMOD 2007.
 41. *Schema advisor for hybrid relational-XML DBMS.*
Mirella Moura Moro, Lipyeow Lim and Yuan-Chi Chang. SIGMOD 2007.
 42. *Unifying Data and Domain Knowledge Using Virtual Views.*
Lipyeow Lim, Haixun Wang and Min Wang. VLDB 2007.
 43. *Preserving XML queries during schema evolution (Poster).*
Mirella Moura Moro, Susan Malaika and Lipyeow Lim. WWW 2007.
 44. *Cost-based optimization in DB2 XML.*
Andrey Balmin, Tom Eliaz, John Hornibrook, Lipyeow Lim, Guy M. Lohman, David E.

- Simmen, Min Wang and Chun Zhang. IBM Systems Journal 45(2) 2006.
45. *Real Time Business Performance Monitoring and Analysis Using Metric Network*. Pu Huang, Hui Lei and Lipyeow Lim. ICEBE 2006.
 46. *Managing E-Commerce Catalogs in a DBMS with Native XML Support*. Lipyeow Lim and Min Wang. ICEBE 2005.
 47. *CXHist : An On-line Classification-Based Histogram for XML String Selectivity Estimation*. Lipyeow Lim, Min Wang and Jeffrey Scott Vitter. VLDB 2005.
 48. *On-line Methods for Database Optimization*. Lipyeow Lim. PhD Thesis, Department of Computer Science, Duke University, North Carolina, USA. 2004.
 49. *SASH: A Self-Adaptive Histogram Set for Dynamically Changing Workloads*. Lipyeow Lim, Min Wang and Jeffrey Scott Vitter. VLDB 2003.
 50. *Dynamic maintenance of web indexes using landmarks*. Lipyeow Lim, Min Wang, Sriram Padmanabhan, Jeffrey Scott Vitter and Ramesh C. Agarwal. WWW 2003.
 51. *XPathLearner: An On-line Self-Tuning Markov Histogram for XML Path Selectivity Estimation*. Lipyeow Lim, Min Wang, Sriram Padmanabhan, Jeffrey Scott Vitter and Ronald Parr. VLDB 2002.
 52. *Wavelet-Based Cost Estimation for Spatial Queries*. Min Wang, Jeffrey Scott Vitter, Lipyeow Lim and Sriram Padmanabhan. SSTD 2001.
 53. *Characterizing Web Document Change*. Lipyeow Lim, Min Wang, Sriram Padmanabhan, Jeffrey Scott Vitter and Ramesh C. Agarwal. WAIM 2001.
 54. *A Theoretical Look at Pixel Ordering*. Lipyeow Lim. Masters Thesis, School of Computing, National University of Singapore, Singapore. 1999.

Patents

1. *Query-aware compression of join results*. Lipyeow Lim, Christian Lang. Patent 8,423,522 issued 4/16/2013.
2. *Optimizing queries to hierarchically structured data*. Lipyeow Lim, Rajesh Bordawekar, Anastasios Kementsietsidis, Bryant Wei Lun Kok. Patent 8,285,711 issued on 10/9/2012
3. *Compression of sorted value indexes using common prefixes*. Lipyeow Lim, Bishwaranjan Bhattacharjee, Sherman Lau, Timothy Malkemus, Cathy McArthur, George Mihaila, Sherkat Reza, Zoltan Toth. Patent 8,255,398 issued 08/28/2012.
4. *Statistics collection using path-identifiers for relational databases*. Lipyeow Lim, George Mihaila, Min Wang. Patent 8,229,924 issued 07/24/2012.
5. *Ontology-based searching in database systems*. Lipyeow Lim, Anastasios Kementsietsidis, Min

- Wang. Patent 8,135,730 issued 03/13/2012.
6. *Method and apparatus for encoding list of variable length structures to support bi-directional scans* Lipyeow Lim, Bhattacharjee, B. Malkemus, T. Mihaila, G. Patent 8,126,929 issued 02/28/2012.
 7. *Identifying and annotating shared hierarchical markup document trees.* Lipyeow Lim, Min Wang, Haixun Wang. Patent 8,108,765 issued 01/31/2012.
 8. *Classification-based method and apparatus for string selectivity estimation.* Lipyeow Lim and Min Wang. Patent 7,987,180 issued 07/26/2011.
 9. *Method, apparatus and system for business performance monitoring and analysis using metric network.* Hui Lei, Pu Huang and Lipyeow Lim. Patent 7,895,152 issued 02/22/2011.
 10. *Compressability estimation of non-unique indexes in a database management system.* Lipyeow Lim, Bhattacharjee, B. Mihaila, G. Patent 7,895,171 issued 02/22/2011.
 11. *Method and system for indexing and serializing data.* Haixun Wang, Lipyeow Lim, Min Wang, Xiaohui Gu. Patent 7,752,192 issued 07/06/2010.
 12. *Method for Supporting Ontology-Related Semantic Queries in DBMSs with XML Support.* Lipyeow Lim, Min Wang and Haixun Wang. Patent 7,730,098 issued 06/01/2010
 13. *Querying data and an associated ontology in a database management system.* Lipyeow Lim, Min Wang, Haixun Wang. Patent 7,693,812 issued 4/6/2010.
 14. *Generating database schemas for relational and markup language data from a conceptual model.* Mirella Moro, Lipyeow Lim, Yuan-chi Chang, Sharon Adler. Patent 7,630,993 issued 12/08/2009.
 15. *Statistics collection using path-identifiers for relational databases.* Lipyeow Lim, George Mihaila, Min Wang. Patent 7,613,682 issued 11/03/2009.
 16. *Method for searching deep web services* Zhen Zhang, Lipyeow Lim, Min Wang and Yuan-Chi Chang. Patent 7,533,085 issued 05/12/2009.
 17. *Method and apparatus for organizing data sources.* Zhen Zhang, Lipyeow Lim, Min Wang and Yuan-Chi Chang. Patent 7,529,740 issued 05/05/2009.
 18. *Method, apparatus and system for business performance monitoring and analysis using metric network.* Hui Lei, Pu Huang and Lipyeow Lim. Patent 7,509,308 issued 03/24/2009.
 19. *Statistics collection using path-value pairs for relational databases.* Min Wang, Lipyeow Lim and George Mihaila. Patent 7,472,108 issued 12/30/2008.
 20. *Classification-based method and apparatus for string selectivity estimation.* Lipyeow Lim and Min Wang. Patent 7,395,270 issued 7/1/2008.

21. *Dynamic Maintenance of Web Indices Using Landmarks*. Lipyeow Lim, Min Wang, Sriram Padmanabhan, Ramesh Agarwal. Patent 7,299,404 issued 11/20/2007.
22. *Semantic Link Discovery*. Oktie Hassanzadeh, Anastasios Kementsietsidis, Lipyeow Lim, Min Wang. Application 20110106836. 05/05/2011.
23. *Semantic-Aware Record Matching*. Oktie Hassanzadeh, Anastasios Kementsietsidis, Lipyeow Lim, Min Wang. Application 20110106821. 05/05/2011.
24. *Semantic query by example*. Lipyeow Lim, Haixun Wang, Min Wang. Application 20110078187. 03/31/2011.
25. *Method and apparatus for selecting an optimal delete-safe compression method on list of delta encoded integers* Lipyeow Lim, Bhattacharjee, B. Malkemus, T. Mihaila, G. Application 20090248723. 10/01/2009.
26. *Method and system for combining ranking and clustering in a database management system*. Chengkai Li, Min Wang, Haixun Wang, Lipyeow Lim. Application 20080270374. 10/30/2008.
27. *Method and Apparatus for Providing Direct Access to Unique Hierarchical Data Items*. Lipyeow Lim, George Mihaila, Yuan-chi Chang. Application 20080183657. 07/31/2008.
28. *Processing queries on hierarchical markup data using shared hierarchical markup trees*. Lipyeow Lim, Min Wang, Haixun Wang. Application 20080091649. 04/17/2008.

Grants

- 2014 Hawaiian Electric Company Award PXP-14-752 at \$69,924 for Short Term Solar Irradiance Forecasting and Analysis.
- 2013 University Corporation for Atmospheric Research (UCAR) Subaward Z13-13582. Funded at \$5,833 for Year 1. Co-Investigator. *Nowcasting/Forecasting High-Resolution Irradiance for Photovoltaic Power Generation over Hawai'i*.
- 2010 IBM Scalable Data Analytics for a Smarter Planet Innovation Award. Funded at \$20,000. Principle Investigator. *Scalable Data Analytics Infrastructure for Smart Wind Energy Management*.

Invited Talks

1. *Query-Aware Compression of Join Results*. Computer Science Seminar, School of Computing, National University of Singapore, Singapore, June 2013.
2. *Semantic Query By Example*. The 16th International Conference on Extending Database Technology (EDBT), Genoa, Italy, March 2013.
3. *Optimizing Sensor Data Acquisition for Energy-Efficient Smartphone-based Continuous Event Processing*. The 12th IEEE International Conference on Mobile Data Management, Lulea, Sweden, June 2011.

4. *Optimizing Sensor Data Acquisition for Energy-Efficient Smartphone-based Continuous Event Processing*. Computer Science Seminar, School of Computing, National University of Singapore, Singapore, June 2011.
5. *Optimizing Sensor Data Acquisition for Energy-Efficient Smartphone-based Continuous Event Processing*. Research Seminar, School of Information Systems, Singapore Management University, Singapore, June 2011.
6. *Optimizing Access Across Multiple Hierarchies in Data Warehouses*. Hawaii International Conference on System Sciences, Kauaii, Hawaii, USA, January 2011.
7. *Optimizing Content Freshness of Relations Extracted from the Web using Keyword Search*. ACM SIGMOD International Conference on Management of Data, Indianapolis, Indiana, USA, June 2010.
8. *Parallelization of XPath Queries using Multi-Core Processors: Challenges and Experiences*. International Conference on Extending Database Technology (EDBT), March 2009.
9. *Supporting ontology-based keyword search over medical databases*. American Medical Informatics Association (AMIA) Annual Symposium, Washington, DC, USA, November 2008.
10. *Schema Advisor for Hybrid Relational-XML DBMS*. ACM SIGMOD International Conference on Management of Data, Beijing, China, June 2007.
11. *Unifying Data and Domain Knowledge Using Virtual Views*. International Conference on Very Large Data Bases (VLDB), Vienna, Austria, September 2007.
12. *SASH: A Self-Adaptive Histogram Set for Dynamically Changing Workloads*. International Conference on Very Large Data Bases, Berlin, Germany, September 2003.
13. *Dynamic maintenance of web indexes using landmarks*. The 12th International World Wide Web Conference, Budapest, Hungary, May 2003.
14. *Wavelet-Based Cost Estimation for Spatial Queries*. The 7th International Symposium on Spatial and Temporal Databases (SSTD), Redondo Beach, CA, USA, July 2001.

Service

University Service

- | | |
|-----------|---|
| 2013-2017 | Curriculum Committee Chairperson,
Information and Computer Sciences Department,
University of Hawai'i at Mānoa, HI, U.S.A.. |
| 2009-2013 | Graduate Committee Member,
Information and Computer Sciences Department,
University of Hawai'i at Mānoa, HI, U.S.A.. |
| 2009-2013 | Curriculum Committee Member,
Information and Computer Sciences Department,
University of Hawai'i at Mānoa, HI, U.S.A.. |

- 2010-2012 Undergraduate Research Council Member,
University of Hawai‘i at Mānoa, HI, U.S.A..
- 2002-2003 Graduate Student Liaison to the Faculty.
Department of Computer Science, Duke University, NC, U.S.A..

Research Community Service

- 2017 *Local Arrangements Co-Chair* for IEEE CLUSTER 2017
Program Committee Member for CIKM 2017
Program Committee Member for ICDE 2017
Program Committee Member for IEEE SWC 2017
- 2016 *Program Committee Member* for DASFAA 2016
- 2015 *Local Arrangements Chair* for VLDB 2015
Program Committee Member for SBBB 2015
- 2014 *Reviewer* for JIDM
Program Committee Member for WWW 2014 Demo Track
Program Committee Member for EDBT 2014 Industrial Track
- 2013 *Program Committee Member* for VLDB 2013
Reviewer for Hawaii International Conference on System Sciences (HICSS) 2013
Program Committee Member for IEEE International Conference on Big Data (IEEE Big-Data) 2013
Program Committee Member for ACM International Workshop on Data Warehousing and OLAP (DOLAP) 2013
Program Committee Member for International Workshop for Analytics Services in the Cloud (ASC) 2013
Reviewer for Transactions on Knowledge and Data Engineering Journal
- 2012 *Local Arrangements Chair* for CIKM 2012
Program Committee Chair for International Conference on Web-Age Information Management (WAIM) 2012
Program Committee Member for CIKM 2012
Program Committee Member for International Conference on Database and Expert Systems Application (DEXA) 2012
Program Committee Member for ACM International Workshop on Data Warehousing and OLAP (DOLAP) 2012
Program Committee Member for VLDB 2012
Program Committee Member for International Workshop on Analytics Services on the Cloud (ASC) 2012
Reviewer for Data and Knowledge Engineering (DKE) Journal
- 2011 *Local Arrangements Chair* for the IEEE International Conference on Machine Learning And Applications (ICMLA) 2011
Program Committee Member for the International Symposium on Database Programming

- Languages (DBPL) 2011
Program Committee Member for DEXA 2011
Program Committee Member for WAIM 2011
Reviewer for HICSS 2011
Reviewer for DKE Journal
Reviewer for the IEEE Transactions on Knowledge and Data Engineering (TKDE) Journal
Program Committee Member for ICDE 2011
- 2010 *Program Committee Member* for WAIM 2010
Program Committee Member for VLDB 2010
Program Committee Member for the SIGMOD PhD Workshop on Innovative Database Research 2010
Program Committee Member for DEXA 2010
- 2009 *External Reviewer* for VLDB 2009
Program Committee Member for DEXA 2009
Program Committee Member for ICDE 2009
Program Committee Member for International Conference on Extending Database Technology (EDBT) 2009
- 2008 *Program Committee Member* for ACM International Workshop on Web Information and Data Management (WIDM) 2008
Program Committee Member for VLDB 2008
External Reviewer for the Asia-Pacific Web Conference (APWEB) 2008
- 2007 *Reviewer* for the World Wide Web Journal
External Reviewer for WAIM 2007
External Reviewer for DEXA 2007
External Reviewer for VLDB 2007
- 2006 *External Reviewer* for VLDB 2006
- 2005 *External Reviewer* for VLDB 2005
External Reviewer for DEXA 2005
- 2004 *Reviewer* for the World Wide Web Conference Poster Track.
- 2003 *Reviewer* for IEEE Intelligent Systems magazine.

Projects

- Jan 2017–Present *Plant Identification in Natural Images using Deep Learning*
 Use deep learning techniques to analyze natural images in order to identify plant(s) in the image. The vision is to create a mobile application that hikers, students, tourists can use to learn about plants in the state of Hawai'i by simply taking a picture of the plant using their smartphone camera.
- Nov 2010–Present *Energy Efficient Complex Event Processing on Smartphones*
 Investigate novel query processing algorithms for complex event processing

in mobile devices like smartphones. Advance mobile applications like tele-health and context sensing performs continuous queries on streams of sensor data which maybe on the phone or tethered to the phone. The optimization objectives are to conserve power on the phone and sensors and to maximize the operational lifetime of the entire system.

Jan 2012–Dec 2017

Data Analytics for Renewable Energy Management

Investigate the use of data mining techniques to analyze various sensor and weather model output data to extract interesting patterns for renewable energy management.

Jan 2009–Dec 2012

Query-Aware Compression of Join Query Results.

Investigate the use of redundancy information from query execution plans to compress query results. Query results are often transmitted to client devices over a communication network and hence any reduction in data size has a significant impact on network usage, latency, and energy consumption.

Jan 2011–Jun 2012

Stream Computing for Continuous Wind Forecasting.

Studied the feasibility of applying stream computing techniques to the problem of continuous wind forecasting using numerical weather prediction models (WRF). Continuous forecasting aims to continuously assimilate the most current observation data and provide up-to-date forecast to be used in the management of renewable energy.

Aug 2008–Dec 2012

Parallel Evaluation of XPath Queries on Multi-Core Architectures.

Designed parallel algorithms for evaluating XPath queries on multi-core processors in order to minimize the latency of queries on large XML databases. While inter-query parallelism has increased the throughput of XML queries, this project focus on intra-query parallelism to reduce the latency of individual queries.

Jan 2009–Jun 2010

Refreshing Structured Content Extracted from the Web Using Keyword Search.

Studied the problem of maintaining structured relations that are extracted from web sources using only a keyword search interface. Design update algorithms and heuristics to minimize the amount of probing of web sources.

Jan 2009–Jul 2009

Performance optimization of IBM's internal Tyto RDF Store.

Collected usecases and performance problems in the usage of Tyto in IBM's Websphere products. Designed special indexes and techniques for performance enhancement of RDF processing.

Apr 2008–Jul 2009

Database Adapters for IBM InfoSphere Streams.

Designed, prototyped and implemented adapters for integrating InfoSphere Streams, IBM's stream computing middleware, with IBM database servers.

Jul 2008–Jul 2009

Artemis First-of-a-kind Project with UOIT

Led the design and implementation of the Data Integration Manager, a key component in the Artemis project. The Artemis project aims to monitor premature babies using streaming technology in order to detect critical conditions such as sepsis before they happen.

Jan 2006–Jul 2009

Semantic Search in Database and Information Retrieval Systems.

Investigated the problem of searching data using semantics from an ontol-

ogy in databases. Existing semantic search applications then lack scalability and the ability to integrate with existing information systems. The thrust of this research is to support ontology and semantic search in a scalable and integrated fashion in databases and information systems.

- Oct 2007–Mar 2008 *Index Compression Line Item in IBM DB2 LUW Cobra Release.*
Designed, prototyped and benchmarked index compression schemes for IBM's DB2 database product. The resultant compression schemes achieved a compression rate of 50% which exceeded the targeted rate of 40%.
- Mar 2005–Dec 2007 *Data management in a relational-XML model.*
Studied the problem of managing data using hybrid relational-XML databases. The entity-relationship model and the normal forms were used to design schemas for relational data. Hybrid relational-XML databases allows XML columns which cannot be modelled adequately by the normal forms.
- Mar 2005–Dec 2006 *XML Schema Evolution.*
Studied the problem of XML schema evolution with a particular focus on how schema evolution will impact applications and how applications need to evolve with the schema. One component of this project is the development of a XML schema difference tool.
- Oct 2004– Jun 2006 *IBM DB2XML Product Development.*
Developed the XML runstats component in DB2 v9.1 (Viper). In contrast to prototyping, this is actual product development work encompassing design, documentation, coding, testing, defect resolution etc.
- Sep 2003–Dec 2004 *Classification-based Histograms.*
Designed a histogram technique that used Bayesian classifiers to remember the values associated with a bucket approximately. Histograms are used to store statistics for database query optimization.
- Jun–Sep 2003 *IBM DB2XML Product Prototyping.*
Internship project at IBM T. J. Watson Research Center. Prototyped the runstats routine in DB2XML, IBM's new native XML DBMS. The runstats routine collects XML data statistics used in query optimization.
- Jul 2002–Oct 2004 *Self-Adaptive Set of Histograms.*
Studied the problem of building a set of histograms to store statistics for relational database query optimization without scanning the underlying data. Statistical graphical models are used to learn which attributes to build histograms on and how to build these histograms. Collaborators: Jeffrey S. Vitter (Purdue) and Min Wang (IBM).
- Jun 2001–Dec 2006 *XML Path Selectivity.*
Studied the problem of collecting statistics for query optimization in native XML databases without scanning the underlying data. Started as an internship project at IBM T. J. Watson Research Center, this research produced XPathLearner (VLDB 2002), an on-line method for estimating the selectivity of XML path expressions. Collaborators: Jeffrey S. Vitter (Purdue), Min Wang (IBM) and Ronald Parr (Duke).
- Mar–May 2002 *Finding Motifs from Gene Expression and Sequence Data.*

- Designed an efficient algorithm to find candidate motifs in upstream regions by combining information from both gene expression data and gene sequencing data. Course project for graduate level course *Computational Functional Genomics*. Collaborators: Allister Bernard and Ryan Deering.
- May 2000–May 2001 *Updating Web Indexes.*
Studied the problem of keeping web indexes up-to-date. Started as an internship project at IBM T. J. Watson Research Center, this research produced the novel landmark-diff method (published in WWW 2003) for updating web indexes. Collaborators: Jeffrey S. Vitter (Purdue), Min Wang (IBM), Sriram Padmanabhan (IBM), Ramesh Agarwal (IBM).
- Sep–Nov 2000 *Pixel Ordering with Markov Decision Processes (MDPs).*
Studied the use of Markov Decision Processes to find an optimal path to traverse all the pixels of an image for optimal compression using a predictive encoder. Course project for graduate-level course *Markov Decision Processes*.
- Mar–May 2000 *Using Prediction to Manage Power-Aware Memory.*
Conducted research on using different prediction methods to predict memory usage patterns for powering up/down of memory chips. Course project for graduate-level course *Operating Systems*. Collaborators: Martin Gilbert.
- Nov 1999–Mar 2000 *Selectivity Estimation in Spatial Database.*
Conducted research on using wavelets to compress sparse 2D histograms that are used in query optimization in spatial databases. Collaborators: Jeffrey S. Vitter (Duke) and Min Wang (IBM).
- Jul 1998–Aug 1999 *Image Compression Research.*
Conducted research on image compression techniques using predictive-differential coding, discrete cosine transform (DCT), wavelet transforms, hybrid transforms and pixel ordering. Advisor: Philip M. Long.
- Jul 1997–Jun 1998 *Mobile-IP Research.*
Implemented RFC 2290 *MIPv4 Option for PPP IPCP* on Linux's PPP source. The scope of the project covers issues related to the inter-working of the MIPv4 protocol and the PPP protocol. This is the first implementation of this RFC in the world.
- Jul–Dec 1997 *Cryptanalysis Research.*
Implemented and benchmarked the linear cryptanalysis method (Mitsuru Matsui) and the differential cryptanalysis method on a reduced rounds DES. Course project for the graduate-level course *Cryptographic systems*.
- Jul–Dec 1997 *Parallel Matrix Multiplication.*
Implemented and benchmarked parallel versions of a block-matrix multiplication algorithm and the Strassen's matrix multiplication algorithm using the Message Passing Interface (MPI) standard. Course project for the graduate-level class *Parallel and Distributed Algorithms*.

References

1. *Jeffrey S. Vitter*, Kansas University, (785) 864-4904, jsv@ku.edu.
2. *Min Wang*, Senior Vice President, Head of Visa Research minw83@gmail.com.
3. *Jun Yang*, Duke University, (919) 660-6587, junyang@cs.duke.edu.
4. *Sriram Padmanabhan*, IBM Silicon Valley Labs, (408) 463-2712, srp@us.ibm.com.

Personal

- Citizen of Singapore
- U.S. Permanent Resident
- Language Ability: English, Mandarin, German, Cantonese and the Teochew dialect.