

# FLORIN RUSU

*Mailing address:*  
SE2-210, 5200 N Lake Rd.  
Merced, CA 95343, USA

*Phone:* +1 (209) 228-4286  
*E-mail:* frusu@ucmerced.edu  
*Web:* <http://www.faculty.ucmerced.edu/frusu>

---

## Research Interests

- Big Data: design and build infrastructure for Big Data analytics
- Data science: approximate query processing and scalable machine learning
- Scientific data analysis: multi-dimensional array storage and in-situ processing
- Databases and large scale data management in general

## Education

- May 2009     *Ph.D. (Computer Science)*  
University of Florida, USA  
Dissertation: *Sketches for Aggregate Estimations over Data Streams*  
Advisor: Alin Dobra
- August 2008     *M.Sc. (Computer Science)*  
University of Florida, USA
- June 2004     *B.Eng. (Computer Science)*  
Technical University of Cluj-Napoca, Romania (5 year curriculum)  
Thesis: *DICOM-Compliant Integrated System for Medical Structured Reporting*  
Advisor: Sergiu Nedevschi

## Work Experience

- University of California at Merced, School of Engineering     September 2010–present  
Title: *Associate Professor*     July 2016–present  
Title: *Assistant Professor*     September 2010–June 2016
- Start the research program in large scale data management and Big Data infrastructure. Design and implement the GLADE parallel processing system and its extensions to array data processing (EXTASCID), parallel online aggregation (PF-OLA), and scalable machine learning. Supervise 6 Ph.D. students, 9 M.S. students, and several undergraduate students. Create two graduate courses and two undergraduate courses. Participate in various school and university-level committees. Receive a DOE Early Career Award in 2014 and a Hellman Faculty Fellowship in 2013.
- Lawrence Berkeley National Laboratory, Scientific Data Management (SDM) Group     August 2015–present  
Title: *Faculty Scientist*
- Work on the scientific data management and analysis applications required by the Palomar Transient Factory (PTF) project in astronomy. Collaborate closely with the SDM group members.
- Innovative Scheduling, Inc.     June 2009–June 2010  
Title: *Senior Software Engineer*
- Perform research, design, and implement scheduling algorithms for large scale transportation networks. Develop models for network optimization problems.

## Research Experience

University of Florida, CISE Department

January 2005–May 2009

Title: *Research Assistant*

- Perform research on different approximate query processing techniques with applications to databases and data streaming under the supervision of Alin Dobra and Christopher Jermaine.

IBM Almaden Research Center

May 2008–August 2008

Title: *Summer Intern*

- Design and implement algorithms for mining queries from a data warehouse under the supervision of Vijayshankar Raman, Lin Qiao, and Peter Haas.

Technical University of Cluj-Napoca, Computer Science Department

October 2002–June 2004

Title: *Junior Research Assistant*

- Work on the design and implementation of a medical structured reporting system for electrocardiography images under the supervision of Sergiu Nedeveschi. Joint work with University of Medicine and Pharmacy *Iuliu Hatieganu* from Cluj-Napoca.

## Publications

### Journal

1. Weijie Zhao, **Florin Rusu**, Kesheng Wu, and Peter Nugent. *Automatic Identification and Classification of Palomar Transient Factory Astrophysical Objects in GLADE*. International Journal of Computational Science and Engineering (IJCSE), Vol. 16, No. 4, pp. 337–349, July 2018. [Accepted in June 2016]
2. Anna Y. Q. Ho, S. R. Kulkarni, Peter E. Nugent, Weijie Zhao, **Florin Rusu**, S. Bradley Cenko, Vikram Ravi, Mansi M. Kasliwal, Daniel A. Perley, Scott M. Adams, Eric C. Bellm, Patrick Brady, Christoffer Fremling, Avishay Gal-Yam, David Alexander Kann, David Kaplan, Russ R. Laher, Frank Masci, Eran O. Ofek, Jesper Sollerman, and Alex Urban. *iPTF Archival Search for Fast Optical Transients*. The Astrophysical Journal Letters, Vol. 854, No. 1, February 2018.
3. M. M. Kasliwal, E. Nakar, L. P. Singer, D. L. Kaplan, D. O. Cook, A. Van Sistine, R. M. Lau, C. Fremling, O. Gottlieb, J. E. Jencson, S. M. Adams, U. Feindt, K. Hotokezaka, S. Ghosh, D. A. Perley, P.-C. Yu, T. Piran, J. R. Allison, G. C. Anupama, A. Balasubramanian, K. W. Bannister, J. Bally, J. Barnes, S. Barway, E. Bellm, V. Bhalerao, D. Bhattacharya, N. Blagorodnova, J. S. Bloom, P. R. Brady, C. Cannella, D. Chatterjee, S. B. Cenko, B. E. Cobb, C. Copperwheat, A. Corsi, K. De, D. Dobie, S. W. K. Emery, P. A. Evans, O. D. Fox, D. A. Frail, C. Frohmaier, A. Goobar, G. Hallinan, F. Harrison, G. Helou, T. Hinderer, A. Y. Q. Ho, A. Horesh, W.-H. Ip, R. Itoh, D. Kasen, H. Kim, N. P. M. Kuin, T. Kupfer, C. Lynch, K. Madsen, P. A. Mazzali, A. A. Miller, K. Mooley, T. Murphy, C.-C. Ngeow, D. Nichols, S. Nissanke, P. Nugent, E. O. Ofek, H. Qi, R. M. Quimby, S. Rosswog, **F. Rusu**, E. M. Sadler, P. Schmidt, J. Sollerman, I. Steele, A. R. Williamson, Y. Xu, L. Yan, Y. Yatsu, C. Zhang, W. Zhao. *Illuminating Gravitational Waves: A Concordant Picture of Photons from a Neutron Star Merger*. Science, October 2017.
4. Anna Y. Q. Ho, Peter Nugent, Weijie Zhao, **Florin Rusu** et al. *Multi-messenger Observations of a Binary Neutron Star Merger*. The Astrophysical Journal Letters, Vol. 848, No. L12, October 2017.
5. Dan Olteanu and **Florin Rusu**. *Special Issue on In-Database Analytics*. Distributed and Parallel Databases (DAPD), Vol. 35, No. 3–4, pp. 333–334, September 2017.
6. Chengjie Qin, Martin Torres, and **Florin Rusu**. *Scalable Asynchronous Gradient Descent Optimization for Out-of-Core Models*. Proceedings of the VLDB Endowment (PVLDB), Vol. 10, No. 10, pp. 986–997, June 2017. [Oral presentation at VLDB 2017 International Conference on Very Large Data Bases, Munich, Germany]

7. **Florin Rusu**, Zixuan Zhuang, Mingxi Wu, and Chris Jermaine. *Workload-Driven Antijoin Cardinality Estimation*. ACM Transactions on Database Systems (TODS), Vol. 40, No. 3, October 2015. [Poster at ACM SIGMOD 2016]
8. Yu Cheng and **Florin Rusu**. *SCANRAW: A Database Meta-Operator for Parallel In-Situ Processing and Loading*. ACM Transactions on Database Systems (TODS), Vol. 40, No. 3, October 2015.
9. **Florin Rusu**, Chengjie Qin, and Martin Torres. *Scalable Analytics Model Calibration with Online Aggregation*. IEEE Data Engineering Bulletin, Vol. 38, No. 3, pp. 30–44, September 2015. [Invited article]
10. Yu Cheng and **Florin Rusu**. *Formal Representation of the SS-DB Benchmark and Experimental Evaluation in EXTASCID*. Distributed and Parallel Databases (DAPD), May 2014.
11. Chengjie Qin and **Florin Rusu**. *PF-OLA: A High-Performance Framework for Parallel Online Aggregation*. Distributed and Parallel Databases (DAPD), August 2013. [Special issue on scalable data summarization on Big Data]
12. **Florin Rusu** and Alin Dobra. *GLADE: A Scalable Framework for Efficient Analytics*. Operating Systems Review (OSR), Vol. 46, No. 1, pp. 12–18, January 2012. [Selected as one of the best papers presented at LADIS 2011]
13. Alin Dobra, Chris Jermaine, **Florin Rusu**, and Fei Xu. *Turbo-Charging Estimate Convergence in DBO*. Proceedings of the VLDB Endowment (PVLDB), Vol. 2, No. 1, pp. 419–430, August 2009. [Oral presentation at VLDB 2009 International Conference on Very Large Data Bases, Lyon, France] [Acceptance rate 17%]
14. **Florin Rusu** and Alin Dobra. *Sketches for Size of Join Estimation*. ACM Transactions on Database Systems (TODS), Vol. 33, No. 3, September 2008.
15. **Florin Rusu** and Alin Dobra. *Pseudo-Random Number Generation for Sketch-Based Estimations*. ACM Transactions on Database Systems (TODS), Vol. 32, No. 2, June 2007. [Accepted at first submission]

### Conference & Workshop

1. Weijie Zhao, **Florin Rusu**, Bin Dong, Kesheng Wu, Anna Y. Q. Ho, and Peter Nugent. *Distributed Caching for Processing Raw Arrays*. Proceedings of the 2018 SSDBM International Conference on Scientific and Statistical Database Management (SSDBM 2018), Bolzano-Bozen, Italy, July 2018. [Acceptance rate 30%]
2. Chengjie Qin and **Florin Rusu**. *Dot-Product Join: Scalable In-Database Linear Algebra for Big Model Analytics*. Proceedings of the 2017 SSDBM International Conference on Scientific and Statistical Database Management (SSDBM 2017), Chicago, Illinois, June 2017. [Acceptance rate 26%]
3. Yu Cheng, Weijie Zhao, and **Florin Rusu**. *Bi-Level Online Aggregation on Raw Data*. Proceedings of the 2017 SSDBM International Conference on Scientific and Statistical Database Management (SSDBM 2017), Chicago, Illinois, June 2017. [Acceptance rate 26%]
4. Bin Dong, Kesheng Wu, Surendra Byna, Jialin Liu, Weijie Zhao, and **Florin Rusu**. *ArrayUDF: User-Defined Scientific Data Analysis on Arrays*. Proceedings of the 2017 ACM HPDC International Symposium on High-Performance Parallel and Distributed Computing (HPDC 2017), Washington, DC, June 2017. [Acceptance rate 19%]
5. Weijie Zhao, **Florin Rusu**, Bin Dong, Kesheng Wu, and Peter Nugent. *Incremental View Maintenance over Array Data*. Proceedings of the 2017 ACM SIGMOD International Conference on Management of Data (SIGMOD 2017), Chicago, Illinois, May 2017. [Acceptance rate 19%]

6. Borui Wang, Martin Torres, Dong Li, Jishen Zhao, and **Florin Rusu**. *Performance Implications of Processing-in-Memory Designs on Data-Intensive Applications*. Proceedings of the 2016 HUCAA Workshop on Heterogeneous and Unconventional Cluster Architectures and Applications (HUCAA 2016) held in conjunction with the International Conference on Parallel Processing (ICPP 2016), Philadelphia, Pennsylvania, August 2016.
7. Weijie Zhao, **Florin Rusu**, Bin Dong, and Kesheng Wu. *Similarity Join over Array Data*. Proceedings of the 2016 ACM SIGMOD International Conference on Management of Data (SIGMOD 2016), San Francisco, California, June 2016.
8. Weijie Zhao, Yu Cheng, and **Florin Rusu**. *Vertical Partitioning for Query Processing over Raw Data*. Proceedings of the 2015 ACM SSDBM International Conference on Scientific and Statistical Database Management (SSDBM 2015), San Diego, California, June 2015. [Acceptance rate 45%]
9. Chengjie Qin and **Florin Rusu**. *Speculative Approximations for Terascale Distributed Gradient Descent Optimization*. Proceedings of the 2015 DanaC Workshop on Data Analytics at Scale (DanaC 2015) held in conjunction with the ACM SIGMOD International Conference on Management of Data (SIGMOD 2015), Melbourne, Victoria, Australia, May 2015.
10. Yu Cheng and **Florin Rusu**. *Parallel In-Situ Data Processing with Speculative Loading*. Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data (SIGMOD 2014), Snowbird, Utah, June 2014. [Acceptance rate 25%]
11. **Florin Rusu**, Peter Nugent, and John K. Wu. *Implementing the Palomar Transient Factory Real-Time Detection Pipeline in GLADE: Results and Observations*. Proceedings of the 2014 International Workshop on Databases in Networked Information Systems (DNIS 2014), Aizu-Wakamatsu, Japan, March 2014, pp. 53–66.
12. Chengjie Qin and **Florin Rusu**. *Speeding-Up Distributed Low-Rank Matrix Factorization*. Proceedings of the 2013 International Conference on Cloud Computing and Big Data (CloudCom-Asia 2013), Fuzhou, China, December 2013, pp. 521–528.
13. Yu Cheng and **Florin Rusu**. *Astronomical Data Processing in EXTASCID*. Proceedings of the 2013 ACM SSDBM International Conference on Scientific and Statistical Database Management (SSDBM 2013), Baltimore, Maryland, July 2013, pp. 387–390. [Demo track]
14. Chengjie Qin and **Florin Rusu**. *Parallel Online Aggregation in Action*. Proceedings of the 2013 ACM SSDBM International Conference on Scientific and Statistical Database Management (SSDBM 2013), Baltimore, Maryland, July 2013, pp. 383–386. [Demo track]
15. Chengjie Qin and **Florin Rusu**. *Sampling Estimators for Parallel Online Aggregation*. Proceedings of the 2013 BNCOD British National Conference on Databases (BNCOD 2013), Oxford, Great Britain, July 2013, pp. 204–217. [Acceptance rate 38%]
16. Chengjie Qin and **Florin Rusu**. *Scalable I/O-Bound Parallel Incremental Gradient Descent for Big Data Analytics in GLADE*. Proceedings of the 2013 DanaC Workshop on Data Analytics in the Cloud (DanaC 2013) held in conjunction with the ACM SIGMOD International Conference on Management of Data (SIGMOD 2013), New York, New York, June 2013, pp. 16–20. [Acceptance rate 56%]
17. Yu Cheng, Chengjie Qin, and **Florin Rusu**. *GLADE: Big Data Analytics Made Easy*. Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data (SIGMOD 2012), Scottsdale, Arizona, May 2012, pp. 697–700. [Demo track] [Acceptance rate 42%]
18. **Florin Rusu** and Alin Dobra. *GLADE: A Scalable Framework for Efficient Analytics*. Proceedings of the 2011 LADIS Workshop on Large Scale Distributed Systems and Middleware (LADIS 2011) held in conjunction with the VLDB Conference on Very Large Databases (VLDB 2011), Seattle, Washington, September 2011.

19. **Florin Rusu** and Alin Dobra. *Sketching Sampled Data Streams*. Proceedings of the 2009 IEEE ICDE International Conference on Data Engineering (ICDE 2009), Shanghai, China, March 2009, pp. 381–392. [Acceptance rate 17%]
20. **Florin Rusu**, Fei Xu, Ravi Jampani, Luis Perez, Mingxi Wu, Chris Jermaine, and Alin Dobra. *The DBO Database System*. Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data (SIGMOD 2008), Vancouver, Canada, June 2008, pp. 1223–1226. [Demo track] [Acceptance rate 32%]
21. **Florin Rusu** and Alin Dobra. *Statistical Analysis of Sketch Estimators*. Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data (SIGMOD 2007), Beijing, China, June 2007, pp. 187–198. [Acceptance rate 14%]
22. **Florin Rusu** and Alin Dobra. *Fast Range-Summable Random Variables for Efficient Aggregate Estimation*. Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data (SIGMOD 2006), Chicago, Illinois, June 2006, pp. 193–204. [Acceptance rate 13%]
23. Sergiu Nedeveschi, **Florin Rusu**, Iulian Popa, Alexandru Smeu, Dan Olinic, and Calin Homorodean. *Structured Reporting of Echocardiography Images in a DICOM Environment*. Proceedings of Computers in Cardiology 2004, Chicago, Illinois, September 2004, pp. 685–688.
24. Sergiu Nedeveschi, Dan Olinic, Radu Popovici, **Florin Rusu**, Alexandru Smeu, and Calin Homorodean. *DICOM Compliant Environment for Structured Reporting in Echocardiography*. Proceedings of Computers in Cardiology 2003, Thessaloniki, Greece, September 2003, pp. 29–32.

#### Technical Report

1. Weijie Zhao, **Florin Rusu**, Bin Dong, Kesheng Wu, Anna Y. Q. Ho, and Peter Nugent. *Distributed Caching for Complex Querying of Raw Arrays*. CoRR abs/1803.06089, March 2018.
2. Yujing Ma, **Florin Rusu**, and Martin Torres. *Stochastic Gradient Descent on Highly-Parallel Architectures*. CoRR abs/1802.08800, February 2018.
3. Anna Y. Q. Ho, S. R. Kulkarni, Peter E. Nugent, Weijie Zhao, **Florin Rusu**, S. Bradley Cenko, Vikram Ravi, Mansi M. Kasliwal, Daniel A. Perley, Scott M. Adams, Eric C. Bellm, Patrick Brady, Christoffer Fremling, Avishay Gal-Yam, David Alexander Kann, David Kaplan, Russ R. Laher, Frank Masci, Eran O. Ofek, Jesper Sollerman, and Alex Urban. *iPTF Archival Search for Fast Optical Transients*. CoRR abs/1712.00949, December 2017.
4. M. M. Kasliwal, E. Nakar, L. P. Singer, D. L. Kaplan, D. O. Cook, A. Van Sistine, R. M. Lau, C. Fremling, O. Gottlieb, J. E. Jencson, S. M. Adams, U. Feindt, K. Hotokezaka, S. Ghosh, D. A. Perley, P.-C. Yu, T. Piran, J. R. Allison, G. C. Anupama, A. Balasubramanian, K. W. Bannister, J. Bally, J. Barnes, S. Barway, E. Bellm, V. Bhalerao, D. Bhattacharya, N. Blagorodnova, J. S. Bloom, P. R. Brady, C. Cannella, D. Chatterjee, S. B. Cenko, B. E. Cobb, C. Copperwheat, A. Corsi, K. De, D. Dobie, S. W. K. Emery, P. A. Evans, O. D. Fox, D. A. Frail, C. Frohmaier, A. Goobar, G. Hallinan, F. Harrison, G. Helou, T. Hinderer, A. Y. Q. Ho, A. Horesh, W.-H. Ip, R. Itoh, D. Kasen, H. Kim, N. P. M. Kuin, T. Kupfer, C. Lynch, K. Madsen, P. A. Mazzali, A. A. Miller, K. Mooley, T. Murphy, C.-C. Ngeow, D. Nichols, S. Nissanke, P. Nugent, E. O. Ofek, H. Qi, R. M. Quimby, S. Rosswog, **F. Rusu**, E. M. Sadler, P. Schmidt, J. Sollerman, I. Steele, A. R. Williamson, Y. Xu, L. Yan, Y. Yatsu, C. Zhang, W. Zhao. *Illuminating Gravitational Waves: A Concordant Picture of Photons from a Neutron Star Merger*. CoRR abs/1710.05436, October 2017.
5. Yu Cheng, Weijie Zhao, and **Florin Rusu**. *OLA-RAW: Scalable Exploration over Raw Data*. CoRR abs/1702.00358, February 2017.
6. Borui Wang, Martin Torres, Dong Li, Jishen Zhao, and **Florin Rusu**. *Performance Implications of Processing-in-Memory Designs on Data-Intensive Applications*. UC Merced EECS technical report, March 2016.

7. Chengjie Qin and **Florin Rusu**. *Dot-Product Join: An Array-Relation Join Operator for Big Model Analytics*. CoRR abs/1602.08845, February 2016.
8. Weijie Zhao, Yu Cheng, and **Florin Rusu**. *Workload-Driven Vertical Partitioning for Effective Query Processing over Raw Data*. CoRR abs/1503.08946, March 2015.
9. Chengjie Qin and **Florin Rusu**. *Speculative Approximations for Terascale Analytics*. CoRR abs/1501.00255, December 2014.
10. Yu Cheng and **Florin Rusu**. *Formal Representation of the SS-DB Benchmark and Experimental Evaluation in EXTASCID*. CoRR abs/1305.1609, May 2013.
11. **Florin Rusu** and Yu Cheng. *A Survey on Array Storage, Query Languages, and Systems*. CoRR abs/1302.0103, February 2013.
12. Chengjie Qin and **Florin Rusu**. *PF-OLA: A High-Performance Framework for Parallel On-Line Aggregation*. CoRR abs/1206.0051, June 2012.

## Poster

1. Yujing Ma, Martin Torres, and **Florin Rusu**. *Asynchronous Stochastic Gradient Descent on GPU: Is It Really Better than CPU?* Northern California (NorCal) Database Day, Redwood City, California, May 2018.
2. Ye Zhu and **Florin Rusu**. *High-Throughput Push-Based Storage Manager*. Northern California (NorCal) Database Day, Redwood City, California, May 2018.
3. Jun Hyung Shin, **Florin Rusu**, and Alex Suhan. *Novel Selectivity Estimation Strategy for Modern DBMS*. Northern California (NorCal) Database Day, Redwood City, California, May 2018.
4. Weijie Zhao, **Florin Rusu**, Bin Dong, Kesheng Wu, Anna Y. Q. Ho, and Peter Nugent. *Distributed Caching for Processing Raw Arrays*. Northern California (NorCal) Database Day, Redwood City, California, May 2018.
5. Zhiyi Huang, Weijie Zhao, **Florin Rusu**, Kai Zeng, and Rahul Potharaju. *Adaptive Online Aggregation with Randomness Detection*. Northern California (NorCal) Database Day, Redwood City, California, May 2018.
6. **Florin Rusu** and Yujing Ma. *Asynchronous Stochastic Gradient Descent on GPU: Is It Really Better than CPU?* Extremely Large Databases Conference (XLDB 2018), Palo Alto, California, May 2018. [Lightning talk and poster.]
7. Yujing Ma, Martin Torres, and **Florin Rusu**. *In-Depth Analysis of HOGWILD SGD on GPU*. Northern California (NorCal) Database Day, San Francisco, California, April 2017.
8. Zhiyi Huang, Weijie Zhao, **Florin Rusu**, and Kai Zeng. *Initialization-Free Online Aggregation with Adaptive Sampling*. Northern California (NorCal) Database Day, San Francisco, California, April 2017.
9. Weijie Zhao, **Florin Rusu**, Bin Dong, Kesheng Wu, and Peter Nugent. *Incremental View Maintenance over Array Data*. Northern California (NorCal) Database Day, San Francisco, California, April 2017.
10. Chengjie Qin, Martin Torres, and **Florin Rusu**. *Scalable HOGWILD! for Big Models*. Northern California (NorCal) Database Day, San Francisco, California, April 2017 & Bay Area Machine Learning Symposium (BayLearn), Sunnyvale, California, October 2016.
11. **Florin Rusu**. *Scalable In-Situ Exploration over Raw Data*. Conference on Innovative Data Systems Research (CIDR), Santa Cruz, California, January 2017.
12. Abdur Rafay, Abhineet Dubey, Zhiyi Huang, Xin Zhang, and **Florin Rusu**. *Recent Developments in GLADE*. NERSC Data Day, Berkeley, California, August 2016 & Northern California (NorCal) Database Day, Mountain View, California, April 2016.

13. Yu Cheng, Weijie Zhao, and **Florin Rusu**. *Online Aggregation on Raw Data*. Northern California (NorCal) Database Day, Mountain View, California, April 2016.
14. Chengjie Qin, Martin Torres, and **Florin Rusu**. *Dot-Product Join: An Array-Relation Join Operator for Big Model Analytics*. Northern California (NorCal) Database Day, Mountain View, California, April 2016.
15. Weijie Zhao, **Florin Rusu**, Bin Dong, and Kesheng Wu. *Similarity Join over Array Data*. Northern California (NorCal) Database Day, Mountain View, California, April 2016.
16. Chi Zhang and **Florin Rusu**. *When to Stop Random Walks in Social Networks*. Northern California (NorCal) Database Day, Mountain View, California, April 2016.
17. Chengjie Qin and **Florin Rusu**. *Database Parameter Server*. Bay Area Machine Learning Symposium (BayLearn), Menlo Park, California, October 2015.
18. Chengjie Qin and **Florin Rusu**. *Effective Model Calibration for Terascale Analytics*. Extremely Large Databases Conference (XLDB 2015), Palo Alto, California, May 2015. [Lightning talk and poster.]
19. Chengjie Qin and **Florin Rusu**. *Effective Model Calibration for Terascale Analytics*. Northern California (NorCal) Database Day, Santa Cruz, California, April 2015.
20. Weijie Zhao, Yu Cheng, and **Florin Rusu**. *Vertical Partitioning for Query Processing over Raw Data*. Northern California (NorCal) Database Day, Santa Cruz, California, April 2015.
21. Chengjie Qin and **Florin Rusu**. *Fast Parameter Tuning with Approximations at Scale*. Bay Area Machine Learning Symposium (BayLearn), Berkeley, California, October 2014.
22. Yu Cheng and **Florin Rusu**. *Parallel In-Situ Data Processing with Speculative Loading*. Northern California (NorCal) Database Day, IBM Almaden Research Center, California, April 2014.
23. Chengjie Qin and **Florin Rusu**. *Intra-Iteration Approximations for Parallel Gradient Descent Optimization*. Northern California (NorCal) Database Day, IBM Almaden Research Center, California, April 2014.
24. Zixuan Zhuang and **Florin Rusu**. *Workload-Driven Antijoin Cardinality Estimation*. Northern California (NorCal) Database Day, IBM Almaden Research Center, California, April 2014.
25. **Florin Rusu**, Yu Cheng, and Suzanne Sindi. *In-Situ Processing of Genomic Sequence Alignment Data*. Extremely Large Databases Conference (XLDB 2013), Palo Alto, California, September 2013. [Lightning talk and poster.]
26. Chengjie Qin and **Florin Rusu**. *Lightning-Fast, Dirt-Cheap Parallel Stochastic Gradient Descent for Big Data in GLADE*. Bay Area Machine Learning Symposium (BayLearn), Menlo Park, California, August 2013.
27. Yu Cheng and **Florin Rusu**. *Running Queries From Raw Files*. Northern California (NorCal) Database Day, Palo Alto, California, April 2013.
28. Chengjie Qin and **Florin Rusu**. *Scalable I/O-Bound Parallel Incremental Gradient Descent for Big Data Analytics in GLADE*. Northern California (NorCal) Database Day, Palo Alto, California, April 2013.
29. Yu Cheng and **Florin Rusu**. *EXTASCID: An Extensible System for the Analysis of Scientific Data*. Extremely Large Databases Conference (XLDB 2012), Palo Alto, California, September 2012. [Poster and system demo.]
30. Yu Cheng and **Florin Rusu**. *EXTASCI: An Extensible System for Scientific Data Analysis*. Northern California (NorCal) Database Day, Berkeley, California, April 2012.
31. Chengjie Qin and **Florin Rusu**. *PF-OLA: A High-Performance Framework for Parallel On-Line Aggregation*. Northern California (NorCal) Database Day, Berkeley, California, April 2012.

32. **Florin Rusu.** *GLADE: A Scalable Framework for Efficient Analytics.* Extremely Large Databases Conference (XLDB 2011), Menlo Park, California, October 2011.
33. **Florin Rusu.** *GLADE: A Highly-Scalable Architecture-Independent Framework for Efficient Analytics.* Northern California (NorCal) Database Day, Davis, California, April 2011.

## Talks

- *Distributed Caching for Processing Raw Arrays*
  - SSDBM 2018 Conference, Bolzano-Bozen, Italy, July 2018
- *Advanced Database Techniques for Processing Scientific Multi-Dimensional Data*
  - Invited talk at Temple University, Philadelphia, Pennsylvania, May 2018
  - Invited talk at UC Merced – LBNL Forum, Merced, California, February 2018
- *Stochastic Gradient Descent on Highly-Parallel Architectures: Multi-core CPU or GPU? Synchronous or Asynchronous?*
  - Invited talk at IBM Almaden Research, San Jose, California, December 2017
- *OLA-RAW: Scalable Exploration over Raw Data*
  - Invited talk at Amazon AWS, Palo Alto, California, October 2017
- *Scalable Asynchronous Gradient Descent Optimization for Out-of-Core Models*
  - VLDB 2017 Conference, Munich, Germany, August 2017
- *Dot-Product Join: Scalable In-Database Linear Algebra for Big Model Analytics*
  - SSDBM 2017 Conference, Chicago, Illinois, June 2017
- *Asynchronous Stochastic Gradient Descent on GPU: Is It Really Better than CPU?*
  - XLDB 2018 Conference, Palo Alto, California, May 2018
  - Invited talk at Google Research, Mountain View, California, June 2017
- *GLADE: A Scalable Big Data Analytics System*
  - University of Nevada Reno, Rochester Institute of Technology, Florida International University, Lehigh University, DePaul University, UC Davis, UNC Charlotte, Portland State University, February–April 2016
- *GLADE: A Big Data System for Scalable Interactive Exploration*
  - Invited talk at Cloudera, San Francisco, California, August 2015
- *Speculative Approximations for Terascale Distributed Gradient Descent Optimization*
  - DanaC 2015 Workshop at ACM SIGMOD Conference, Melbourne, Victoria, Australia, May 2015
  - XLDB 2015 Conference, Palo Alto, California, May 2015
- *Parallel In-Situ Data Processing Techniques*
  - Invited talk at HGST R&D (Western Digital), San Jose, California, December 2014
  - Invited talk at Lawrence Berkeley National Laboratory (LBNL), Berkeley, California, November 2014
- *Implementing the Palomar Transient Factory Real-Time Detection Pipeline in GLADE: Results and Observations*

- DNIS 2014 Workshop, Aizu-Wakamatsu, Japan, March 2014
- *Sampling Estimators for Parallel Online Aggregation*
  - BNCOD 2013 Conference, Oxford, Great Britain, July 2013
- *Scalable I/O-Bound Parallel Incremental Gradient Descent for Big Data Analytics in GLADE*
  - DanaC 2013 Workshop at ACM SIGMOD Conference, New York, New York, June 2013
- *Methods for Scalable Interactive Exploration of Massive Datasets*
  - University of California Santa Cruz, Santa Cruz, California, November 2012
  - California Polytechnic State University (Cal Poly), San Luis Obispo, California, October 2012
  - University of California Davis, Davis, California, September 2012
- *GLADE: A Scalable Framework for Efficient Analytics*
  - Invited talk at Lawrence Berkeley National Laboratory (LBNL), Berkeley, California, June 2012
  - LADIS 2011 Workshop at VLDB Conference, Seattle, Washington, September 2011
  - Invited talk at EMC Greenplum, San Mateo, California, May 2011
- *Scalable Approximate Query Processing*
  - Job talk given at New Mexico State University, University of California at Merced, and Utah State University, February–April 2010
- *Sketching Sampled Data Streams*
  - IEEE ICDE 2009 Conference, Shanghai, China, March 2009
- *Fast Randomized Algorithms for Data Streams*
  - Cornell University, DB Colloquium, Ithaca, New York, February 2009
- *Mining for Queries*
  - IBM Almaden Research Center, San Jose, California, August 2008
- *Statistical Analysis of Sketch Estimators*
  - ACM SIGMOD 2007 Conference, Beijing, China, June 2007
  - University of Florida, DB Seminar, Gainesville, Florida, January 2007
- *Fast Range-Summable Random Variables for Efficient Aggregate Estimation*
  - ACM SIGMOD 2006 Conference, Chicago, Illinois, June 2006
  - University of Florida, DB Seminar, Gainesville, Florida, February 2006

## Funding

- California Department of Education Mathematics and Science Partnership (CaMSP): *Team-E Science*, 2015-2018 (\$600,000)
- Department of Energy Early Career Award (DOE CAREER): *Scalable and Energy-Efficient Methods for Interactive Exploration of Scientific Data*, 2014-2019 (\$750,000)
- Hellman Faculty Fellowship: *Parallel On-Line Aggregation Techniques for Big Data Analytics*, 2013-2014 (\$20,000)
- Research gift from LogicBlox, 2013 (\$16,000)

## Teaching Experience

University of California at Merced, School of Engineering

Title: *Associate & Assistant Professor*

- CSE 100, Algorithm Design and Analysis (Fall 2017)
- CSE 111, Database Systems (Fall 2018, Fall 2016, Fall 2015, Fall 2014, Fall 2013, Fall 2012, Spring 2011)
- CSE 165/ENGR 140, Introduction to Object-Oriented Programming (Fall 2011)
- CSE 177/EECS 277, Database Systems Implementation (Spring 2017, Spring 2015, Spring 2012)
- CSE 195, Upper Division Undergraduate Research (Spring 2017, Summer 2016, Fall 2015)
- EECS 284, Large Scale Data Management (Spring 2013, Spring 2011)
- EECS 284, Big Data Systems & Analytics (Spring 2018, Spring 2016)
- EECS 290, EECS Seminar (Fall 2016, Spring 2012)

University of Florida, CISE Department

Title: *Teaching Assistant*

- Introduction to Computer Organization (Spring 2008)
- Database Management Systems (Fall 2005)

## Student Advising

University of California at Merced, School of Engineering

- *Ph.D.*:
  1. **Asoke Datta** (Fall 2018–present)
  2. **Yesdaulet Izenov** (Fall 2018–present)
  3. **Yujing Ma** (Fall 2016–present; intern: LBNL 2017)
  4. Martin Torres (Fall 2015–Summer 2018)
  5. Weijie Zhao (Fall 2014–Summer 2018; dissertation: *Advanced Database Techniques for Scientific Data Processing*; advanced to candidacy in Spring 2017; intern: LBNL 2017, 2016, 2015; first employment: Baidu Research USA)
  6. Chengjie Qin (Fall 2011–Fall 2016; dissertation: *GLADE-ML: A Database for Big Data Analytics*; advanced to candidacy in Fall 2013; intern: GraphSQL; first employment: GraphSQL)
  7. Yu Cheng (Fall 2011–Summer 2016; dissertation: *In-situ Data Processing Over Raw Files*; advanced to candidacy in Fall 2013; intern: Google; first employment: Turn)
- *M.Sc.*:
  1. **Twinkle Mistry** (Fall 2017–present)
  2. **Amar Saini** (Fall 2017–present)
  3. **Ye Zhu** (Fall 2016–present)
  4. **Jaspal Atwal** (Summer 2016–present)
  5. **Zhiyi Huang** (Summer 2016–present)
  6. Jun Hyung Shin (Spring 2017–Summer 2018; thesis: *Novel Selectivity Estimation Strategy for Modern DBMS*)
  7. Abhineet Dubey (Fall 2015–Summer 2018)

8. Xin Zhang (Fall 2015–Fall 2017; intern: LBNL 2016, Striim 2017; thesis: *Code Generation Techniques for Raw Data Processing*; first employment: Autonomic.ai)
  9. Zhiyi Huang (Fall 2015–Summer 2016; switch to Ph.D.)
  10. Abdur Rafay (Fall 2015–Summer 2016; thesis: *Multi-Query Optimization in GLADE*; first employment: FICO)
  11. Zixuan Zhuang (Fall 2013–Summer 2015; thesis: *An Experimental Study of Distributed Quantile Estimation*; intern: GraphSQL; first employment: GraphSQL)
- *B.S.:*
    1. Abraham Uribe (Summer 2017; UROC Undergraduate Research Scholar; work: NBA season prediction in Microsoft SQL Server for Ubuntu)
    2. Alexander Crosdale (Summer 2016–May 2017; UROC Undergraduate Research Scholar; work: implementation of TPC-H in MongoDB)
    3. Jonathan Rodriguez (October 2015–May 2017; work: parallel data partitioning in GLADE)
    4. Gobind Ball (Fall 2015; SoE Outstanding Undergraduate Student Award; work: code generation for JSON parsing; first employment: Zappos)
    5. Nhan Dao (Spring 2014; work: data loading in SciDB; first employment: eGain Corporation)
    6. Zachary Michaels (Spring 2013; work: run GLADE on NERSC supercomputers; first employment: Applied Medical)
    7. Roseller Velicaria (Summer 2012; work: in-memory hash join algorithms; first employment: JD Power and Associates)
    8. Jorge Ramirez Perez (May 2011–May 2012; work: efficient data loading in GLADE; first employment: Boeing)
  - *Visiting scholars:*
    1. Chi Zhang, Ph.D., Beijing Jiao Tong University, China (September 2015–September 2016)
    2. Felipe da Silva Oliveira, B.S., Universidade Federal de Pelotas, Brazil (June 2015–August 2015)

## Professional Service

### Steering Committee Member

- International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA): January 2017–present

### Editorial Board Member

- Associate editor for Distributed and Parallel Databases (DAPD): July 2017–present
- Guest editor for Distributed and Parallel Databases (DAPD) special issue on “Scientific and Statistical Database Management”: June 2017–present
- Guest editor for Distributed and Parallel Databases (DAPD) special issue on “In-Database Analytics”: August 2016–August 2017

### Paper Refereeing

- *Program Committee Member*
  - 2019: SIGMOD, PVLDB, ICDE

- 2018: PVLDB, SSDBM, SIGMOD DEEM Workshop, DASFAA, LOD, BigData Congress
- 2017: SIGMOD, ICDE, DASFAA, SSDBM, BIDMA, MOD, BigData Congress
- 2016: SIGMOD, S2 ICIOT, BDAA, DBKDA, MOD, BIDMA, FAB, SSDBM demo, BigData Congress
- 2015: SIGMOD, MOD, FAB, KDDA, SSDBM
- 2014: CIKM demo, BDSE, SSDBM demo, BigData Congress
- 2013: BigData Congress, BNCOD, ICDE
- 2012: VLDB demo
- 2011: SIGMOD demo
- *Journal Reviewer*
  - ACM Transactions on Database Systems (TODS)
  - Very Large Databases Journal (VLDBJ)
  - IEEE Transactions on Knowledge and Data Engineering (TKDE)
  - Springer Knowledge and Information Systems (KAIS)
  - Elsevier Information Sciences
  - MDPI Sensors, ISPRS International Journal of Geo-Information
  - IEICE Trans. Fundamentals/Commun./Electron./Inf. & Syst.
- *Conference Reviewer*: ICDE (2011, 2009, 2006), SIGMOD (2009 demo, 2006), VLDB (2007)

### Conference Organization

- *Demos Chair*: SSDBM (2017)
- *Organizing Committee*: XLDB (2016)
- *Proceedings Chair*: SIGMOD (2014)
- *Seminar*: UC Merced CITRIS Big Data Seminar (Fall 2013), UC Merced EECS Seminar (Fall 2016, Spring 2012)
- *Session Chair*: SIGMOD (2017, 2015, 2014), SSDBM (2018, 2017, 2015), BNCOD (2013)

### Panels

- *NSF CISE*: 2016, 2014, 2013
- *DOE ASCR*: 2018, 2017, 2014
- *Hellman Fellowship*: 2017, 2016

### Professional Society Membership

- Member of ACM and ACM SIGMOD special interest group
- Member of IEEE, IEEE Computer Society, and TCDE

## Academic Service

University of California at Merced, School of Engineering

- Representative of the Computer Science and Engineering Department to the School of Engineering Executive Committee (August 2018–present)
- Member/chair of the EECS TA Allocation Committee (August 2017–present, May 2015–September 2016)
- Chair of the Computer Science and Engineering Undergraduate Program (CSE Chair) (August 2017–June 2018)
- Member of the UC-System Committee on Academic Computing and Communications (UCACC) (December 2015–August 2017)
- Member of the Information, Privacy, Security Committee (IPSC) (January 2015–August 2017)
- Member of the Undergraduate Council (UGC) (September 2011–May 2014): member of the student admission sub-committee
- Co-chair of the Learning Management System (LMS) Evaluation Committee (June 2012–June 2013)
- Member of the Research Computing Infrastructure Committee (September 2011–June 2012)
- Guest speaker for the Fiat Lux Scholars program (November 2011)
- Chair/member of several faculty search and academic personnel committees

University of Florida, CISE Department

- Treasurer of the Romanian Student Organization (RSA) (September 2006–May 2008)
- Vice-President of the CISE ASCIE graduate student organization (May 2007–May 2008)
- Student member of the CISE Industrial Advisory Board (IAB) (September 2006–May 2007)

## Community Service

- **Merced County of Education Team-E Science Professional Development Seminar** (September 2016): Instructor for K3-8 teachers participating in the Team-E Science project. Introduce data acquisition and visualization for Arduino with ThingSpeak channels.
- **Merced County of Education Team-E Science Summer Institute 2** (June 2016): Instructor for K3-8 teachers participating in the Team-E Science project. Introduce programming in Scratch for Arduino with tools such as ScratchX and S4A.
- **Merced County of Education Team-E Science Summer Institute 1** (June 2015): Instructor for K3-8 teachers participating in the Team-E Science project. Introduce programming in Scratch.

## Honors and Awards

- Recipient of DOE Early Career Award (July 2014–June 2019)
- Recipient of Hellman Faculty Fellowship (July 2013–June 2014)
- Recipient of UF CISE Travel Grant to participate at IEEE ICDE 2009 Conference (March 2009).
- Recipient of UF Student Government Travel Grant to participate at ACM SIGMOD 2008 Conference (June 2008).
- Recipient of EDBT scholarship to attend the 8<sup>th</sup> EDBT Summer School on database technology for novel applications in Bozen-Bolzano, Italy (September 2007)
- Recipient of UF CISE Travel Grant to participate at ACM SIGMOD 2006 Conference (June 2006).

- Recipient of University of Florida outstanding achievement award for academic excellence (May 2006)
- Recipient of Technical University of Cluj-Napoca scholarship (1999–2001) and fellowship (2001–2004) for academic excellence

### **Personal Data**

- Male, Romanian & US citizen
- Languages spoken: *Romanian* (native), *English* (professional proficiency), *French* (intermediate), *Italian*, *Spanish* (elementary)