Bin Shao

Contact Information

Microsoft Research (Beijing, China)
Bldg 2, No. 5 Dan Ling Street, Beijing, PRC 100080

Office: +86 (10) 59174329
E-mail: binshao@microsoft.com

Summary

Bin Shao is a senior principal research manager at Microsoft Research AI for Science. He joined Microsoft after receiving his Ph.D. degree from Fudan University in July 2010. His research interests include ab initio molecular dynamics, computational quantum chemistry, computational biology, machine learning, and distributed computing. The results of his research have appeared in many top conferences and leading journals. Microsoft Graph Engine, which was built by Bin's team, is serving many Microsoft products and services, such as Microsoft Satori knowledge graph, Bing search, MSN, Xbox, and Cognitive Services. Bin is the initiator and the main developer of LightAIMD, which is an open-source lightweight AIMD (*ab initio* molecular dynamics) simulation program.

Education

School of Computer Science, Fudan University, Shanghai, China

Ph.D. Sep. 2005 – Jul. 2010

School of Computer Science and Technology, Shandong University, Jinan, China

B.E. Sep. 2001 – Jul. 2005

Selected Publications

He Zhang, Siyuan Liu, Jiacheng You, Chang Liu, Shuxin Zheng, Ziheng Lu, Tong Wang, Nanning Zheng, *Bin Shao*. Overcoming the barrier of orbital-free density functional theory for molecular systems using deep learning, Nature Computational Science, March 2024.

Yusong Wang, Tong Wang, Shaoning Li, Xinheng He, Mingyu Li, Zun Wang, Nanning Zheng, *Bin Shao*, Tie-Yan Liu, Enhancing geometric representations for molecules with equivariant vector-scalar interactive message passing, Nature Communications, January 2024.

Yongge Li, Fusong Ju, Zhiyuan Chen, Yiming Qu, Huanhuan Xia, Liang He, Lijun Wu, Jianwei Zhu, *Bin Shao*, Pan Deng. CREaTor: zero-shot cis-regulatory pattern modeling with attention mechanisms, Genome Biology (2023) 24:266.

Tong Wang, Xinheng He, Mingyu Li, *Bin Shao*, Tie-Yan Liu. AIMD-Chig: Exploring the conformational space of a 166-atom protein Chignolin with ab initio molecular dynamics, Scientific Data 10, 549 (2023).

Zun Wang, Hongfei Wu, Lixin Sun, Xinheng He, Zhirong Liu, *Bin Shao*, Tong Wang, Tie-Yan Liu. Improving machine learning force fields for molecular dynamics simulations with fine-grained force metrics, The Journal of Chemical Physics, Volume 159, Issue 3, Cover Story, July 2023.

Shiqi Gong, Xinheng He, Qi Meng, Zhiming Ma, *Bin Shao*, Tong Wang, Tie-Yan Liu. Stochastic Lag Time Parameterization for Markov State Models of Protein Dynamics, The Journal of Physical Chemistry B 2022 126 (46), Cover Story, 2022.

Weitao Du, He Zhang, Yuanqi Du, Qi Meng, Wei Chen, Nanning Zheng, *Bin Shao*, Tie-Yan Liu. Equivariant graph neural networks with complete local frames, ICML 2022.

Siyuan Liu, Yusong Wang, Yifan Deng, Liang He, *Bin Shao*, Jian Yin, Nanning Zheng, Tie-Yan Liu, Tong Wang. Improved drug-target interaction prediction with intermolecular graph transformer, Briefings in Bioinformatics, 2022.

Jun Lan, Xinheng He, Yifei Ren, Ziyi Wang, Huan Zhou, Shilong Fan, Chenyou Zhu, Dongsheng Liu, *Bin Shao*, Tie-Yan Liu, Qisheng Wang, Linqi Zhang, Jiwan Ge, Tong Wang, Xinquan Wang. Structural insights into the SARS-CoV-2 Omicron RBD-ACE2 interaction, Cell Research, April 2022.

He Zhang, Fusong Ju, Jianwei Zhu, Liang He, *Bin Shao*, Nanning Zheng, Tie-Yan Liu. Co-evolution Transformer for Protein Contact Prediction. NeurIPS 2021.

Yao Li, Tong Wang, Juanrong Zhang, *Bin Shao*, Haipeng Gong, Yusong Wang, Xinheng He, Siyuan Liu, Tie-Yan Liu. Exploring the Regulatory Function of the N-terminal Domain of SARS-CoV-2 Spike Protein Through Molecular Dynamics Simulation. Advanced Theory and Simulations, Cover Story, 2021.

Siyuan Liu, Tong Wang, Qijiang Xu, *Bin Shao*, Jian Yin, Tie-Yan Liu. Complementing Sequence-derived Features with Structural Information Extracted from Fragment Libraries for Protein Structure Prediction. BMC Bioinformatics 22, 351 (2021).

Wenze Ding, Qijiang Xu, Siyuan Liu, Tong Wang, *Bin Shao*, Haipeng Gong, Tie-Yan Liu. SAMF: a Self-adaptive Protein Modeling Framework. Bioinformatics 2021.

Fusong Ju, Jianwei Zhu, *Bin Shao*, Lupeng Kong, Tie-Yan Liu, Wei-Mou Zheng, Dongbo Bu. CopulaNet: Learning residue co-evolution directly from multiple sequence alignment for protein structure prediction. Nature Communications 12, 2535, May 2021.

Wentao Xu, Shun Zheng, Liang He, *Bin Shao*, Jian Yin, Tie-Yan Liu. SEEK: Segmented Embedding of Knowledge Graphs. In ACL 2020: 2020 Annual Conference of the Association for Computational Linguistics.

Jingping Liu, Yanghua Xiao, Ao Wang, Liang He, *Bin Shao*. CapableOf reasoning: A step towards commonsense oracle. In SIGIR 2020: 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval.

Fanjin Zhang, Xiao Liu, Jie Tang, Yuxiao Dong, Peiran Yao, Jie Zhang, Xiaotao Gu, Yan Wang, *Bin Shao*, Rui Li and Kuansan Wang. OAG: Toward Linking Large-scale Heterogeneous Entity Graphs. In KDD 2019: Proceedings of the 25th ACM SIGKDD conference on Knowledge Discovery and Data Mining, Anchorage, Alaska, USA, August 4-8, 2019, pages 2585-2595.

Bin Shao, Haixun Wang, and Yatao Li. Trinity: A Distributed Graph Engine on a Memory Cloud. In SIGMOD 2013: Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data, New York, New York, USA, June 22-27, 2013, pages 505-516.

Liang He, *Bin Shao*, Yatao Li, Huanhuan Xia, Yanghua Xiao, Enhong Chen, Liang Jeff Chen. Stylus: A Strongly-Typed Store for Serving Massive RDF Data. In PVLDB 2018: Proceedings of the 44th International Conference on Very Large Data Bases, Volume 11, pages 203-216.

Hongbin Ma, *Bin Shao*, Yanghua Xiao, Jeff Liang Chen, and Haixun Wang.G-SQL: Fast Query Processing via Graph Exploration. In PVLDB 2016: Proceedings of the 42th International Conference on Very Large Data Bases, New Delhi, India, September 5-9, 2016, pages 900-911.

Bin Shao, Haixun Wang and Yanghua Xiao. Managing and Mining Large Graphs: Systems and Implementations (Tutorial). In SIGMOD 2012: Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, Scottsdale, Arizona, USA, May 20-24, 2012, pages 589-592.

Zichao Qi, Yanghua Xiao, *Bin Shao*, Haixun Wang. Toward a Distance Oracle for Billion-Node Graphs. In PVLDB 2014: Proceedings of the 40th International Conference on Very Large Data Bases, Hangzhou, China, September 1-5, 2014, pages 61-72.

Bin Shao, Yatao Li. Parallel Processing of Graphs. In: Fletcher G., Hidders J., Larriba-Pey J. (eds) Graph Data Management. Data-Centric Systems and Applications, Springer, 2018, pages 143-162.

Da Yan, Hongzhi Chen, James Cheng, Zhenkun Cai, *Bin Shao*. Scalable De Novo Genome Assembly Using Pregel. In ICDE 2018: Proceedings of the 34th IEEE International Conference on Data Engineering. April 2018.

Liang He, *Bin Shao*, Yanghua Xiao, Yatao Li, Tie-Yan Liu, Enhong Chen, Huanhuan Xia. Neurally-Guided Semantic Navigation in Knowledge Graph. IEEE Transactions on Big Data. February 2018.

Bin Shao, Yatao Li, Haixun Wang, and Huanhuan Xia. Trinity Graph Engine and its Applications. Bulletin of the Technical Committee on Data Engineering, Volume 40, Number 3, September, 2017, pages 18-29.

Yanghua Xiao, *Bin Shao*. Billion-Node Graph Challenges. Bulletin of the Technical Committee on Data Engineering, Volume 40, Number 3, September, 2017, pages 89-99.

Lu Wang, Yanghua Xiao, *Bin Shao*, Haixun Wang. How to Partition a Billion-Node Graph. In ICDE 2014: Proceedings of the 30th IEEE International Conference on Data Engineering, Chicago, IL, USA, March 31-April 4, 2014, pages 568-579.

Kai Zeng, Jiacheng Yang, Haixun Wang, *Bin Shao*, and Zhongyuan Wang. A Distributed Graph Engine for Web Scale RDF Data. In PVLDB 2013: Proceedings of the 39th international conference on Very Large Data Bases, Riva del Garda, Trento, August 26-30, pages 265-276.

Zhao Sun, Hongzhi Wang, Haixun Wang, *Bin Shao*, and Jianzhong Li. Efficient Subgraph Matching on Billion Node Graphs. In PVLDB 2012: Proceedings of the 38th international conference on Very Large Data Bases. 5(9):788-799, 2012.

Huanhuan Xia, Tun Lu, *Bin Shao*, Guo Li, Xianghua Ding, Ning Gu. A partial Replication Approach for Anywhere Anytime Mobile Commenting. In CSCW 2014: Proceedings of the 17th ACM conference on CSCW and social computing, Baltimore, MD, USA, February 15-19, pages 530-541.

Huanhuan Xia, Tun Lu, *Bin Shao*, Xianghua Ding, Ning Gu. Hermes: On Collaboration across Heterogeneous Collaborative Editing Services in the Cloud. In CSCWD 2014: 18th IEEE International Conference on CSCW in Design, Hsinchu, Taiwan, May 21-23, 2014, pages 655-660.

Liang He, *Bin Shao*, Yatao Li, Enhong Chen. Distributed Real-Time Knowledge Graph Serving. In BigComp 2015: Second International Conference on Big Data and Smart Computing, Jeju island, Korea, February 9-12, 2015. Invited Paper.

Bin Shao, Du Li, Tun Lu and Ning Gu. An Operational Transformation Based Synchronization Protocol for Web 2.0 Applications. In CSCW 2011: Proceedings of the 2011 ACM Conference on Computer Supported Cooperative Work, Hangzhou, China, Mar. 19-23, 2011.

Bin Shao, Du Li, and Ning Gu. An Algorithm for Selective Undo of Any Operation in Collaborative Applications. In Group 2010: Proceedings of Group 2010 Conference, Sanibel Island, FL., USA, Nov. 7-10, 2010, pages 131-140.

Bin Shao, Tun Lu, and Ning Gu. Key Techniques of Consistency Maintenance in Real-time Collaboration. Computer Engineering, vol.36, no.22, pages 1-6, Nov. 2010. Invited Paper.

Bin Shao, Du Li, and Ning Gu. A Sequence Transformation Algorithm for Supporting Cooperative Work on Mobile Devices. In CSCW 2010: Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work, Savannah, GA, USA, Feb. 6-10, 2010, pages 159-168.

Bin Shao, Du Li, and Ning Gu. A Fast Operational Transformation Algorithm for Mobile and Asynchronous Collaboration. IEEE Transactions on Parallel and Distributed Systems, vol.21, no.12, pages 1707-1720, Dec. 2010.

Bin Shao, Du Li, and Ning Gu. ABTS: A Transformation-based Consistency Control Algorithm for Wide-area Collaborative Applications. In CollaborateCom 2009: The 5th International Conference on Collaborative Computing: Networking, Applications and Worksharing, Washington D.C., USA, Nov. 11-14, 2009. Invited Paper.

Ho Lee, *Bin Shao*, U Kang. Fast graph mining with HBase. Information Sciences. Volume 315, 2015, pages 56-66.

Yatao Li, *Bin Shao*. Hash storage based large graph generator. Communications of the CCF (Chinese). 8(11): 16 -20, 2012.

Bin Shao, Du Li, and Ning Gu. An Optimized String Transformation Algorithm for Real-time Group Editors. In ICPADS 2009: The 15th International Conference on Parallel and Distributed Systems, Shenzhen, China, Dec. 8-11, 2009, pages 376-383.

Liping Gao, *Bin Shao*, Lin Zhu, Tun Lu, and Ning Gu. Maintaining Time and Space Consistencies in Hybrid CAD Environments: Framework and algorithms. Computers in Industry, Vol. 59, Issue 9, Dec. 2008, pages 894-904.

Sili Zhao, Du Li, Hansu Gu, *Bin Shao*, and Ning Gu. An Approach to Sharing Legacy TV/Arcade Games for Real-time Collaboration. In ICDCS 2009: The 29th International Conference on Distributed Computing Systems, Montreal, Canada, Jun. 22-26, 2009, pages 165-172.

Liping Gao, *Bin Shao*, Tun Lu, and Ning Gu. Maintaining Semantic Intention of Step-wise Operations in Replicated CAD Environments. In CSCWD 2008: The 12th International Conference on CSCW in Design, Xi'an, China, Apr. 16-18, 2008, pages 154-159.

Liping Gao, *Bin Shao*, and Ning Gu. Separating Data and View: Support View-wandering Between Different Trades During Engineering Design. In CSCWD 2007: The 11th International Conference on CSCW in Design, Melbourne, Australia, Apr. 26-28, 2007, pages 36-41.

Hansu Gu, Qiwei Zhang, and *Bin Shao*. Making AutoCAD Collaborative : Implementation and Application of CoAutoCAD. In ICPCA 2007: Proceedings of 2nd International Conference on Pervasive Computing and Applications, Birmingham, England, Jul. 26-27, 2007, pages 168-173.

Patent Applications

Tong Wang, *Bin Shao*, Tie-Yan Liu. Geometric transformer with interatomic positional encoding. 2023.

Chang Liu, Shuxin Zheng, *Bin Shao*, Tie-Yan Liu. Kinetic energy density functional machine learning model. 2023.

Tong Wang, *Bin Shao*, Tie-Yan Liu. Fragment-based quantum mechanical calculation of protein properties. 2023.

Tong Wang, *Bin Shao*, Tie-Yan Liu. Message passing graph neural network with vector-scalar message passing and run-time geometric computation. 2022.

Tong Wang, *Bin Shao*, Tie-Yan Liu. Binding analysis of protein molecule and ligand molecule. Sep 2021.

Tong Wang, *Bin Shao*, Tie-Yan Liu. Structural information from fragment libraries for protein structure prediction. January 2021.

Tong Wang, *Bin Shao*, Tie-Yan Liu. Protein structure folding framework with automatic constraint optimization, self-supervised iteration and deep decoy selection. January 2021.

Bin Shao, Huanhuan Xia, Tie-Yan Liu. Proactive data modeling. April 2018.

Yatao Li, Huanhuan Xia, *Bin Shao*, Tie-Yan Liu. Machine reasoning based on knowledge graph. June 2017.

Bin Shao, Haixun Wang, Wei Fang. US Patent Application 2015/0120775 A1, Answering relational database queries using graph exploration. Application Number: 14/062036. Publication date: Apr. 30, 2015.

Bin Shao, Du Li, Ning Gu. China Patent Application 201010176176.4, A Selective undo method based on operation effects relation order. Publication date: Oct. 6, 2010.