Yuheng Bu

Assistant Professor
Department of Computer Science
University of California, Santa Barbara

Homepage: https://buyuheng.github.io/

Research Interests

My research interests lie in the intersection of *information theory*, *machine learning*, and *signal processing*. I have recently leveraged tools from information theory and signal processing to develop *theoretically justified* learning algorithms for diverse applications, including watermarking generative AI, fair machine learning, uncertainty quantification, and model compression. More broadly, the primary goal of my research is to lay information-theoretic foundations for trustworthy learning algorithms.

Academic Appointments

- Assistant Professor, Department of Computer Science, Jul. 2025 - present
  College of Engineering, **University of California, Santa Barbara**
- Assistant Professor, Department of Electrical and Computer Engineering, Aug. 2022 - Jun. 2025
  Herbert Wertheim College of Engineering, **University of Florida**
- Postdoctoral Research Associate, **Massachusetts Institute of Technology** Sep. 2019 - May 2022
  Advisor: Gregory W. Wornell
- Undergraduate Visiting Research (UGVR), **Stanford University** Jun. 2013 - Sep. 2013
  Only 18 students chosen from mainland China and Taiwan
  Advisor: Tsachy Weissman

Education

- Ph.D. in Electrical and Computer Engineering
  University of Illinois at Urbana-Champaign Jan. 2017 - Aug. 2019
  Advisor: Venugopal V. Veeravalli
  Thesis: "Information-theoretic Bounds in Learning Algorithms"
- Master in Electrical and Computer Engineering
  University of Illinois at Urbana-Champaign Aug. 2014 - Dec. 2016
  Advisor: Venugopal V. Veeravalli
  Thesis: "Estimation of KL Divergence: Optimal Minimax Rate"
- B.E. (**with honors**) in Electronic Engineering
  Tsinghua University, Beijing, China Aug. 2010 - Jul. 2014
  Double Major in **Economics**

Sponsored Research

- Collaborative Research: Elements: Empowering Semiconductor Device Research and Education through Integrated Machine Learning Models and Database
  Co-PI, in collaboration with Jing Guo (PI, UF), Erik Deumens (Co-PI, UF)
  08/15/2024-7/31/2027, **NSF**, total share: $383,998
- UF Research Opportunity Seed Fund (ROSF) 2023: Developing Graph Neural Networks to Quantify Uncertainty for Future of Remote Work in Cities.
  Co-PI, in collaboration with Shenhao Wang (PI, UF), Lingqian Hu (Co-PI, UF)
  09/01/2023-08/31/2025, **UF**, total share: $90,000

Full Publication List

Journal Papers

[1] F. Laakom, M. Gabbouj, J. Schmidhuber, **Y. Bu**. "Class-wise Generalization Error: an Information-Theoretic Analysis," *Transactions on Machine Learning Research (TMLR)*, Jul. 2025.

[2] G. Aminian*, **Y. Bu***, L. Toni, M. R. Rodrigues, G. W. Wornell (* equal contribution). "Information-theoretic Characterizations of Generalization Error for the Gibbs Algorithm," *IEEE Transactions on Information Theory*, vol. 70, no. 1, pp. 632-655, Jan. 2024.

[3] J. K. Lee*, **Y. Bu***, P. Sattigeri, R. Panda, G. W. Wornell, L. Karlinsky, R. S. Feris (* equal contribution). "A Maximal Correlation Framework for Fair Machine Learning," *Entropy*, vol. 24, no. 4, pp. 461, Mar. 2022.

[4] **Y. Bu**, W. Gao, S. Zou, V. V. Veeravalli. "Population Risk Improvement with Model Compression: An Information-Theoretic Approach," *Entropy*, vol. 23, no. 10, pp. 1255, Sep. 2021.

[5] **Y. Bu**, S. Zou, V. V. Veeravalli. "Tightening Mutual Information Based Bounds on Generalization Error," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 1, pp. 121-130, May 2020.

[6] C. Wilson, **Y. Bu**, V. V. Veeravalli. "Adaptive Sequential Machine Learning," *Sequential Analysis*, vol. 38, no. 4, pp. 545-568, Jan. 2020. **(16th Abraham Wald Prize)**

[7] **Y. Bu**, S. Zou, V. V. Veeravalli. "Linear-Complexity Exponentially-Consistent Tests for Universal Outlying Sequence Detection," *IEEE Transactions on Signal Processing*, vol. 67, no. 8, pp. 2115-2128, Apr. 2019.

[8] **Y. Bu***, S. Zou*, Y. Liang, V. V. Veeravalli (* equal contribution). "Estimation of KL Divergence: Optimal Minimax Rate," *IEEE Transactions on Information Theory*, vol. 64, no. 4, pp. 2648-2674, Apr. 2018.

Conference and Workshop Publications

[1] D. Yu, D. Zhuang, L. Jiang, R. Xu, X. Ye, **Y. Bu**, S. Wang, G. Wang. "UQGNN: Uncertainty Quantification of Graph Neural Networks for Multimodal Spatiotemporal Prediction," to appear in *Proc. International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL)*, Minneapolis, MN, Nov. 2025.

[2] H. He, Y. Liu, Z. Wang, Y. Mao, **Y. Bu**. "Distributional Information Embedding: A Framework for Multi-bit Watermarking," to appear in *Asia Pacific Workshop on Data Science and Information Theory (APWDSIT)*, Shenzhen, China, Oct. 2025.

[3] L. An, Y. Liu, Y. Liu, Y. Zhang, **Y. Bu**, S. Chang. "Defending LLM Watermarking Against Spoofing Attacks with Contrastive Representation Learning," in *Proc. Conference on Language Modeling (COLM)*, Montreal, Canada, Oct. 2025. (**acceptance rate: 32%**)

[4] F. Laakom, H. Chen, J. Schmidhuber, **Y. Bu**. "Fairness Overfitting in Machine Learning: An Information-Theoretic Perspective," in *Proc. International Conference on Machine Learning (**ICML**)*, Vancouver, Canada, Jul. 2025. (**acceptance rate: 29%**)

[5] Y. Liu, Y. Song, H. Ci, Y. Zhang, H. Wang, Z. Shou, **Y. Bu**. "Image Watermarks are Removable using Controllable Regeneration from Clean Noise," *International Conference on Learning Representations (**ICLR**)*, Singapore, Apr. 2025. (**acceptance rate: 32%**)

[6] J. J. Ryu, M. Shen, S. Ghosh, **Y. Bu**, P. Sattigeri, S. Das, G. W. Wornell. "Are Uncertainty Quantification Capabilities of Evidential Deep Learning a Mirage?" in *Proc. Conference on Neural Information Processing Systems (**NeurIPS**)*, Vancouver, Canada, Dec. 2024. (**acceptance rate: 26%**)

[7] Y. Zhu, **Y. Bu**. "Information-theoretic Analysis of the Gibbs Algorithm: An Individual Sample Approach," in *Proc. IEEE Information Theory Workshop (ITW)*, Shenzhen, China, Nov. 2024.

[8] D. Zhuang, **Y. Bu**, G. Wang, S. Wang, J. Zhao. "SAUC: Sparsity-Aware Uncertainty Calibration for Spatiotemporal Prediction with Graph Neural Networks," in *Proc. International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL)*, Atlanta, GA, Oct. 2024. Short

version presented at *NeurIPS Workshop on Temporal Graph Learning (TGL)*, New Orleans, LA, Dec. 2023. (**Spotlight talk**)

[9] B. Hu, **Y. Bu**, J. C. Príncipe. "Learning Orthonormal Features in Self-Supervised Learning using Functional Maximal Correlation," in *Proc. IEEE International Conference on Image Processing (ICIP)*, Abu Dhabi, United Arab Emirates, Oct. 2024. Short version presented at *NeurIPS Workshop on Self-Supervised Learning - Theory and Practice*, New Orleans, LA, Dec. 2023. (**Spotlight talk**)

[10] H. Chen, G. Aminian, **Y. Bu**. "An Algorithm for Computing the Capacity of Symmetrized KL Information for Discrete Channels," in *Proc. Allerton Conference on Communication, Control, and Computing (Allerton)*, Urbana, IL, Sep. 2024.

[11] C. Shi, **Y. Bu**, J. Fu. "Information-Theoretic Opacity-Enforcement in Markov Decision Processes," in *Proc. International Joint Conference on Artificial Intelligence (**IJCAI**)*, Jeju, Korea, Aug. 2024. (**acceptance rate: 15%**)

[12] Y. Liu, **Y. Bu**. "Adaptive Text Watermark for Large Language Models," in *Proc. International Conference on Machine Learning (**ICML**)*, Vienna, Austria, Jul. 2024. (**acceptance rate: 28%**)

[13] J. J. Ryu, X. Xu, H. S. M. Erol, **Y. Bu**, L. Zheng, G. W. Wornell. "Operator SVD with Neural Networks via Nested Low-Rank Approximation," in *Proc. International Conference on Machine Learning (**ICML**)*, Vienna, Austria, Jul. 2024. (**acceptance rate: 28%**) Short version presented at *NeurIPS Workshop on Machine Learning and the Physical Sciences*, New Orleans, LA, Dec. 2023.

[14] **Y. Bu**. "Towards Optimal Inverse Temperature in the Gibbs Algorithm," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Athens, Greece, Jul. 2024.

[15] A. Shah, M. Shen, J. J. Ryu, S. Das, P. Sattigeri, **Y. Bu**, G. W. Wornell. "Group Fairness with Uncertainty in Sensitive Attributes," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Athens, Greece, Jul. 2024.

[16] H. Chen, G. W. Wornell, **Y. Bu**. "Gibbs-Based Information Criteria and the Over-Parameterized Regime," in *Proc. International Conference on Artificial Intelligence and Statistics (**AISTATS**)*, Valencia, Spain, May 2024. (**acceptance rate: 28%**) Short version presented at *NeurIPS Workshop on Mathematics of Modern Machine Learning (M3L)*, New Orleans, LA, Dec. 2023.

[17] F. Laakom, **Y. Bu**, M. Gabbouj. "Information-Theoretic Generalization Bounds for the Subtask Problem," *ICML Workshop on PAC-Bayes Meets Interactive Learning*, Honolulu, HI, Jul. 2023.

[18] M. Shen, **Y. Bu**, G. W. Wornell. "On Balancing Bias and Variance in Unsupervised Multi-Source-Free Domain Adaptation," in *Proc. International Conference on Machine Learning (**ICML**)*, Honolulu, HI, Jul. 2023. (**acceptance rate: 28%**)

[19] **Y. Bu**, H. V. Tetali, G. Aminian, M. R. Rodrigues, G. W. Wornell. "On the Generalization Error of Meta Learning for the Gibbs Algorithm," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Taipei, Jun. 2023.

[20] A. Weiss, A. Lancho, **Y. Bu**, G. W. Wornell. "A Bilateral Bound on the Mean Squared Error for Estimation in Model Mismatch," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Taipei, Jun. 2023.

[21] M. Shen, S. S. Ghosh, P. Sattigeri, S. Das, **Y. Bu**, G. W. Wornell. "Reliable Gradient-free and Likelihood-free Prompt Tuning," in *Findings of the Association for Computational Linguistics: EACL 2023*, Dubrovnik, Croatia, May 2023.

[22] H. He, G. Aminian, **Y. Bu**, M. R. Rodrigues, V. Y. F. Tan. "How Does Pseudo-Labeling Affect the Generalization Error of the Semi-Supervised Gibbs Algorithm?" in *Proc. International Conference on Artificial Intelligence and Statistics (**AISTATS**)*, Valencia, Spain, Apr. 2023. (**acceptance rate: 29%**)

[23] M. Shen, **Y. Bu**, P. Sattigeri, S. S. Ghosh, S. Das, G. W. Wornell. "Post-hoc Uncertainty Learning using a Dirichlet Meta-Model," in *Proc. AAAI Conference on Artificial Intelligence (**AAAI**)*, Washington DC, Feb. 2023. (**acceptance rate: 19%**)

[24] A. Lancho, A. Weiss, G. C. Lee, J. Tang, **Y. Bu**, Y. Polyanskiy, G. W. Wornell. "Data-Driven Blind Synchronization and Interference Rejection for Digital Communication Signals," in *Proc. IEEE Global Communications Conference*, Rio de Janeiro, Brazil, Dec. 2022.

[25] G. C. Lee, A. Weiss, A. Lancho, J. Tang, **Y. Bu**, Y. Polyanskiy, G. W. Wornell. "Exploiting Temporal Structures of Cyclostationary Signals for Data-Driven Single-Channel Source Separation," in *Proc. IEEE International Workshop on Machine Learning for Signal Processing*, Xi'an, China, Aug. 2022. (**Best Student Paper Award**)

[26] A. Shah*, **Y. Bu***, J. K. Lee, P. Sattigeri, R. Panda, S. Das, G. W. Wornell (* equal contribution). "Selective Regression under Fairness Criteria," in *Proc. International Conference on Machine Learning (**ICML**)*, Baltimore, MD, Jul. 2022. (**acceptance rate: 22%**)

[27] G. Aminian*, **Y. Bu***, G. W. Wornell, M. R. Rodrigues (* equal contribution). "Tighter Expected Generalization Error Bounds via Convexity of Information Measures," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Espoo, Finland, Jun. 2022.

[28] J. K. Lee*, **Y. Bu***, P. Sattigeri, R. Panda, G. W. Wornell, L. Karlinsky, R. S. Feris (* equal contribution). "A Maximal Correlation Approach to Imposing Fairness in Machine Learning," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Singapore, May 2022.

[29] **Y. Bu***, G. Aminian*, L. Toni, M. R. Rodrigues, G. W. Wornell (* equal contribution). "Characterizing and Understanding the Generalization Error of Transfer Learning with Gibbs Algorithm," in *Proc. International Conference on Artificial Intelligence and Statistics (**AISTATS**)*, virtual, Mar. 2022. (**acceptance rate: 29%**)

[30] G. Aminian*, **Y. Bu***, L. Toni, M. R. Rodrigues, G. W. Wornell (* equal contribution). "An Exact Characterization of the Generalization Error for the Gibbs Algorithm," in *Proc. Conference on Neural Information Processing Systems (**NeurIPS**)*, virtual, Dec. 2021. (**acceptance rate: 26%**) Short version presented at *ICML Workshop on Information-Theoretic Methods for Rigorous, Responsible, and Reliable Machine Learning*, virtual, Jul. 2021.

[31] **Y. Bu***, J. K. Lee*, D. Rajan, P. Sattigeri, R. Panda, S. Das, G. W. Wornell (* equal contribution). "Fair Selective Classification via Sufficiency," in *Proc. International Conference on Machine Learning (**ICML**)*, virtual, Jul. 2021. (**Oral**, **Top 3%**)

[32] **Y. Bu**, T. Wang, G. W. Wornell. "SDP Methods for Sensitivity-Constrained Privacy Funnel and Information Bottleneck Problems," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, virtual, Jul. 2021.

[33] **Y. Bu**, W. Gao, S. Zou, V. V. Veeravalli. "Information-theoretic Understanding of Population Risk Improvement with Model Compression," in *Proc. AAAI Conference on Artificial Intelligence (**AAAI**)*, New York, NY, Feb. 2020.

[34] **Y. Bu**, K. Small. "Active Learning in Recommendation Systems with Multi-level User Preferences," *AAAI Workshop on Interactive and Conversational Recommendation Systems (WICRS)*, New York, NY, Feb. 2020.

[35] **Y. Bu**, J. Lu, V. V. Veeravalli. "Active and Adaptive Sequential Learning with Per Time-step Excess Risk Guarantees," in *Proc. IEEE Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2019.

[36] **Y. Bu**, S. Zou, V. V. Veeravalli. "Tightening Mutual Information Based Bounds on Generalization Error," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Paris, France, Jul. 2019.

[37] **Y. Bu**, J. Lu, V. V. Veeravalli. "Model Change Detection with Application to Machine Learning," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, May 2019.

[38] **Y. Bu**, S. Zou, V. V. Veeravalli. "Linear-Complexity Exponentially-Consistent Tests for Universal Outlying Sequence Detection," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Aachen, Germany, Jun. 2017.

[39] **Y. Bu***, S. Zou*, Y. Liang, V. V. Veeravalli (* equal contribution). "Estimation of KL Divergence Between Large-Alphabet Distributions," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Barcelona, Spain, Jul. 2016.

[40] **Y. Bu**, S. Zou, Y. Liang, V. V. Veeravalli. "Universal Outlying Sequence Detection for Continuous Observations," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016.

Patents

[1] M. Shen, **Y. Bu**, P. Sattigeri, S. Ghosh, S. Das, and G. W. Wornell. "Post-hoc uncertainty quantification for machine learning systems." U.S. Patent App. 18/538,726, Jun. 19, 2025.

[2] M. Shen, S. Ghosh, P. Sattigeri, S. Das, **Y. Bu**, and G. W. Wornell. "Reliable gradient-free and likelihood-free prompt tuning." U.S. Patent App. 18/510,612, May 15, 2025.

[3] A. Shah, M. Shen, S. Das, P. Sattigeri, **Y. Bu**, and G. W. Wornell. "Sensitive attribute driven predictive modeling." U.S. Patent App. 18/503,166, May 8, 2025.

[4] J. K. Lee, **Y. Bu**, D. Rajan, P. Sattigeri, S. Das, R. Panda, and G. W. Wornell. "Fair selective classification via a variational mutual information upper bound for imposing sufficiency." U.S. Patent App. 17/565,411, Jun. 29, 2023.

Tutorials

- "Characterizing the Generalization Error of Machine Learning Algorithms Via Information Measures," IEEE Information Theory Workshop (ITW) 2024

Invited Talks

- Adaptive Text Watermark for Large Language Models
  - Information Science and Learning Systems Seminar, Texas A&M Nov. 2024
  - Asilomar Conference on Signals, Systems, and Computers, Pacific Grove Oct. 2024
  - CS Colloquium, University of California, Santa Barbara (UCSB) Sep. 2024
  - Allerton Conference on Communication, Control, and Computing, UIUC Sep. 2024
  - Meta Fundamental AI Research (FAIR), Seattle Aug. 2024
  - Annual Conference on Information Sciences and Systems, Princeton University Mar. 2024
- Towards Consistent Uncertainty Quantification using Deep Models
  - 2024 Joint Statistical Meetings - Statistics & Data Science, Portland Aug. 2024
- Group Fairness with Uncertainty in Sensitive Attributes
  - Allerton Conference on Communication, Control, and Computing, UIUC Oct. 2023
- From Sensitivity-constrained Information Bottleneck to Fair Selective Prediction
  - Information Theory and Data Science Workshop (Virtual), NUS Jan. 2023
- Post-hoc Uncertainty Learning using a Dirichlet Meta-Model
  - Allerton Conference on Communication, Control, and Computing, UIUC Oct. 2022
- Can Information Theory Characterize Learning Algorithms?
  - ECE department, The Ohio State University (OSU) Mar. 2024

- Information Theory and Applications Workshop (ITA), UCSD Feb. 2023
- CS Colloquium, University of California, Santa Barbara (UCSB) Oct. 2022
- ECE department, University of Florida (UF) Mar. 2022
- ECE department, National University of Singapore (NUS) Jan. 2022
- ECE department, Hong Kong University of Science and Technology (HKUST) Jan. 2022
- EEE department, Hong Kong University (HKU) Dec. 2021

Honors & Awards

- 16th **Abraham Wald Prize**, *Sequential Analysis* 2023
- **MLSP** 2022 Best Student Paper Award 2022
- Outstanding Reviewer Award, **NeurIPS** 2021
- **Yi-Min Wang and Pi-Yu Chung** Research Award, UIUC 2019
- Nominee for Graduation Day at **IEEE ITA Workshop** 2019
- Student Travel Grant, **IEEE ISIT** 2016, 2017
- Student Travel Grant, **IEEE ICASSP** 2016
- **Outstanding graduate**, Tsinghua University 2014
- **National Scholarship** Granted by Ministry of Education of China (**top 2%**) 2012, 2013

Teaching Experience

- **University of Florida**:
  EEL 3850: Data Science for ECE (48 enrollment) Fall 2023, 2024
  EEL 6935: Information-Theoretic Methods in Machine Learning (23 enrollment) Spring 2023, 2024
- Teaching and developing contents for **MIT IDSS MicroMaster** online program:
  6.86x Machine Learning with Python–From Linear Models to Deep Learning Spring 2020
  14.310x Data Analysis for Social Scientists (live recitation and video recording) Fall 2019
- Teaching assistant at **UIUC**
  ECE 365: Data Science and Engineering Spring 2019
  ECE 398: Making Sense of Big Data Spring 2017, Fall 2018
  ECE 598: Computational Inference and Learning Fall 2016

Professional Activities

- **Panelist:** NSF CISE: CIF
- **Guest Editors:** Entropy Special Issue "Fairness in Machine Learning: Information Theoretic Perspectives," 2022-2023.
- **Organizer:** 6th and 7th Annual Workshop on Cognition & Control at University of Florida, 2023-2024
- **Reviewer:** Journal of Machine Learning Research (JMLR), Transactions on Machine Learning Research (TMLR), IEEE Transactions on Information Theory (TIT), IEEE Journal on Selected Areas in Information Theory (JSAIT), IEEE Transactions on Signal Processing (TSP), IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), IEEE Transactions on Neural Networks and Learning Systems (TNNLS), Transactions on Information Forensics & Security (TIFS), IEEE Transactions on Vehicular Technology (TVT), IEEE Journal on Selected Areas in Communications (JSAC)
  Conferences: NeurIPS, ICML, ICLR, AAAI, AISTATS, IJCAI, ISIT, ITW, ICASSP
- Area Chair: NeurIPS (2025)
- Technical Program Committee (TPC) Member: ISIT (2025, 2026), ITW (2025)