

# SIHAN (SANDY) YUAN

San Francisco Bay, CA | US Permanent Resident

[sihany@stanford.edu](mailto:sihany@stanford.edu) | [Professional Website](#) | [Google Scholar](#) | [LinkedIn](#) | [GitHub](#)

## SUMMARY & HIGHLIGHTS

---

- Impactful **researcher**: 50+ peer-reviewed **publications** (16 first authored), with 3000+ citations.
- Strong **builder**: developed a slew of open-source **packages** including the fastest galaxy modeling **code**.
- Recognized **problem solver**: Gold Medalist in 2012 Asia Physics Olympiad (highest score).
- Proven **leader**: led a 70+ people working group in the \$100M DESI collaboration (**NYT feature**).
- Effective **communicator**: 30+ presentations/**lectures**. Won \$2000 in public speaking **contest**.

## EXPERIENCES

---

### Postdoctoral Research Fellow

*September 2021 - Present*

Stanford University, Stanford, CA

- **AI for Science**: building a human-AI collaboration framework to accelerate scientific research.
  - **Astro-MCP**: An MCP suite that allows LLM agents to directly access a variety of astronomy datasets and analyze them with popular software tools.
  - **ResearchBench**: A benchmark to measure AI's ability to conduct real-world scientific research, composed of verifiable end-to-end research tasks extracted from published papers. Submitted to NeurIPS.
  - **Astro-Agent**: an idea generation system that combines reasoning, debate, literature search while keeping human feedback in the loop. MCP enabled.
- **Bayesian Inference**: Combined my ML-accelerated **simulation-based inference pipeline** with **novel summary statistics** to achieve state-of-the-art **cosmology posterior constraints**.
- **Leadership**: Chaired a team of 70+ researchers across 10+ countries, directly contributing to the DESI collaboration's first **tentative discovery** of variable dark energy.

### Quantitative Research Intern

*June 2019 - August 2019*

Two Sigma, New York, NY

- Conducted quantitative analysis of stock performance relative to market cycles using linear regression.

### Graduate Research Fellow

*September 2016 - May 2021*

Harvard University, Cambridge, MA

- Developed the **fastest galaxy-halo connection code** (100+ citations, [GitHub](#)), widely used and adopted as the standard framework in international collaborations.
- Applied the above code to a dataset of millions of galaxies to constrain galaxy physics models.
- Constructed **novel clustering statistics** for feature detection in million-object datasets.

## EDUCATION

---

### Harvard University

*September 2016 - May 2021*

Ph.D. Astronomy and Astrophysics

GPA: 3.96/4.00

- *Relevant coursework: Stochastic Methods for Data Analysis, Inference and Optimization; Noise and Data Analysis in Astrophysics; Computing Foundations for Computational Science.*

### Princeton University

*September 2012 - May 2016*

A.B. Astrophysical Sciences

GPA: 3.95/4.00

- Honors: *summa cum laude, Phi Beta Kappa, Sigma Xi Book Award, Shapiro Prize For Academic Excellence*
- *Relevant coursework: Numerical Methods; Complex Analysis with Applications; Algorithms and Data Structures; Fundamentals of Statistics; Differential Equations.*

## TECHNICAL SKILLS

---

### Statistics

Bayesian inference, optimization, regression, clustering, sampling, causal inference.

### ML/AI

Generative models (transformers, diffusion & flows), agent, MCP, RAG, PyTorch.

### Programming

Python (advanced), bash, Git, parallel computing.

### Languages

Native Mandarin, bilingual in English.

## SERVICE & IMPACT

---

- Co-Investigator on James Webb Space Telescope Cycle 3 proposal 5907, *Stanford, CA* 2024
- Mentor to 2 Stanford graduate students and 1 undergrad, *Stanford, CA* 2024
- Working group chair in the Dark Energy Spectroscopic Instrument Collaboration, *Stanford, CA* 2022-2024
- KIPAC Diversity Equity & Inclusion Committee, *Stanford University* 2022-Present
- Journal Referee, *MNRAS*, *ApJ* 2020-Present
- Treasurer/Co-Founder, Open Labs At Harvard, *Harvard University* 2017-2018

## SELECTED PUBLICATIONS

---

8 out of 50+ (16 first authored papers)

1. *Robust cosmological inference from non-linear scales with  $k$ -th nearest neighbor statistics*  
**S. Yuan**, T. Abel, and R. H. Wechsler, 2024, MNRAS, 527 (2), 1993-2009 ([arXiv](#))
2. *Precise Cosmological Constraints from BOSS Galaxy Clustering with a Simulation-Based Emulator of the Wavelet Scattering Transform*  
G. Valogiannis, **S. Yuan**, C. Dvorkin, 2023, Phys. Rev. D, submitted ([arXiv](#))
3. *SUNBIRD: A simulation-based model for full-shape density-split clustering*  
C. Cuesta-Lazaro, E. Paillas, **S. Yuan**, et al., 2023, MNRAS, submitted ([arXiv](#))
4. *DESI 2024 VI: Cosmological Constraints from the Measurements of Baryon Acoustic Oscillations*  
DESI Collaboration including **S. Yuan**, 2024, JCAP, submitted ([arXiv](#))
5. *2D  $k$ -th nearest neighbor statistics: a highly informative probe of galaxy clustering*  
**S. Yuan**, A. Zamora, T. Abel, 2023, MNRAS, 522 (3), 3935-3947 ([arXiv](#))
6. *Stringent  $\sigma_8$  constraints from small-scale galaxy clustering using a hybrid MCMC+emulator framework*  
**S. Yuan**, L. H. Garrison, D. J. Eisenstein, and R. H. Wechsler, 2022, MNRAS, 515 (1), 871-896 ([arXiv](#))
7. *AbacusHOD: A highly efficient extended multi-tracer HOD framework and its application to BOSS and eBOSS data*  
**S. Yuan**, L. H. Garrison, B. Hadzhiyska, S. Bose, and D. J. Eisenstein, 2022, MNRAS, 510 (3): 3301-3320 ([arXiv](#))
8. *A Hybrid Deep Learning Approach to Cosmological Constraints From Galaxy Redshift Surveys*  
M. Ntampaka, D. J. Eisenstein, **S. Yuan**, and L. H. Garrison, 2020, ApJ, 889 (2): 151-166 ([arXiv](#))