Shubham Chandak | CV

🌐 shubhamchandak94.github.io

Sr. Applied Scientist at AWS Annapurna ML (previously in S3). Interested in machine learning, data compression, DNA storage, bioinformatics and information theory.

Education

○ **Stanford University** — **Stanford, CA, USA**
PhD, Electrical Engineering, Advisor: Prof. Tsachy Weissman — *2016–2021*
Dissertation title: Compression of raw genomic data.

○ **Stanford University** — **Stanford, CA, USA**
MS, Electrical Engineering, GPA 4.2/4 — *2016–2018*

○ **Indian Institute of Technology Bombay** — **Mumbai, India**
B.Tech. in Electrical Engineering with Honours, CPI 9.99/10, Minor in Math — *2012–2016*

Academic Achievements

- **Centennial Teaching Assistant Award 2021** from Stanford University School of Engineering.
- **Siemens Inventor of the Year 2020** award in Open Innovation category for sensor time-series compression project.
- **Qualcomm Innovation Fellowship** Finalist in 2019.
- **Beckman Technology Development Award** in 2018 for project on DNA Storage.
- Recipient of **The Numerical Technologies Co-Founders Fellowship** in 2017-18 awarded to the top performers in Stanford Electrical Engineering PhD qualifying exam.
- **Institute Silver Medal** at IIT Bombay for best academic standing in B.Tech. Electrical Engineering in 2016.
- **Prof. K. C. Mukherjee Award** at IIT Bombay for best project among B.Tech. Electrical Engineering students in 2016.
- Awarded the **Aditya Birla Scholarship** for 2012-16 by India's premier business house viz. Aditya Birla Group.
- **Gold medalist (International Rank 9)** at the **44th International Chemistry Olympiad**, Washington DC, USA in 2012.
- **All India Rank 15** in **IIT-JEE 2012** (IIT-Joint Entrance Exam) among more than 470,000 candidates.
- Recipient of **KVPY** Fellowship by Govt. of India with **All India Rank 6** in 2010.
- **NTSE** scholarship (National Talent Search Exam conducted by NCERT, Govt. of India) in 2009.

Publications

Journal

- F. Müntefering, Y. Adhisantoso, S. Chandak, J. Ostermann, M. Hernaez and J. Voges; Genie: The First Open-Source ISO/IEC Encoder for Genomic Data, To appear in *Communications Biology* (2024).
- B. Lau, S. Chandak, S. Roy, K. Tatwawadi, M. Wootters, T. Weissman and H. Ji; Magnetic DNA random access memory with nanopore readouts and exponentially-scaled combinatorial addressing, *Sci Rep* 13, 8514 (2023).
- M.G. Cassar, C. Sebu, M. Pidcock, S. Chandak and B. Andrews; Optimal design of electrodes for functional electrical stimulation applications to single layer isotropic tissues, *COMPEL - The international journal for computation and mathematics in electrical and electronic engineering*, Vol. 42 No. 3, pp. 695-707 (2023).
- Q. Meng, S. Chandak, Y. Zhu and T. Weissman; Reference-free lossless compression of nanopore sequencing reads using an approximate assembly approach, *Sci Rep* 13, 2082 (2023).
- P. Tandon, S. Chandak, P. Pataranutaporn, Y. Liu, A. Mapuranga, P. Maes, T. Weissman and M. Sra; Txt2Vid: Ultra-Low Bitrate Compression of Talking-Head Videos via Text, in *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 107-118, Jan. 2023.
- S. K. Tabatabaei, B. Pham, C. Pan, J. Liu, S. Chandak, S. Shorkey, A. Hernandez, A. Aksimentiev, M. Chen, C. Schroeder and O. Milenkovic; Expanding the Molecular Alphabet of DNA-Based Data Storage Systems with Neural Network Nanopore Readout Processing, *Nano Letters*, 2022, 22 (5), 1905–1914.
- S. Chandak, K. Tatwawadi, S. Sridhar and T. Weissman; Impact of lossy compression of nanopore raw signal data on basecall and consensus accuracy, *Bioinformatics*, Volume 36, Issue 22-23, 1 December 2020, Pages 5313–5321.
- S. Chandak, K. Tatwawadi, I. Ochoa, M. Hernaez and T. Weissman; SPRING: A next-generation compressor for FASTQ data, *Bioinformatics*, Volume 35, Issue 15, 1 August 2019, Pages 2674–2676.
- S. Chandak, K. Tatwawadi and T. Weissman; Compression of genomic sequencing reads via hash-based reordering: algorithm and analysis, *Bioinformatics*, Volume 34, Issue 4, 15 February 2018, Pages 558–567.

○ N. Desai, C. Juvekar, S. Chandak and A. P. Chandrakasan, "An Actively Detuned Wireless Power Receiver With Public Key Cryptographic Authentication and Dynamic Power Allocation," in *IEEE Journal of Solid-State Circuits*, vol. 53, no. 1, pp. 236-246, Jan. 2018.

Conference

○ M.G. Cassar, C. Sebu, M. Pidcock, B. Andrews and S. Chandak; "Optimal Design of Concentric Electrodes for Functional Electrical Stimulation in multilayered isotropic tissues," *2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, Tenerife, Canary Islands, Spain, 2023, pp. 1-11.

○ R. Hulett, S. Chandak, and M. Wootters; "On Coding for an Abstracted Nanopore Channel for DNA Storage," *2021 IEEE International Symposium on Information Theory (ISIT), 2021*, pp. 2465-2470.

○ R. Prabhakar, S. Chandak, C. Chiu, R. Liang, H. Nguyen, K. Tatwawadi and T. Weissman; "Reducing latency and bandwidth for video streaming using keypoint extraction and digital puppetry," *2021 Data Compression Conference (DCC)*, Snowbird, UT, USA, 2021, pp. 360-360.

○ M. Goyal, K. Tatwawadi, S. Chandak and I. Ochoa; "DZip: improved general-purpose lossless compression based on novel neural network modeling," *2021 Data Compression Conference (DCC)*, Snowbird, UT, USA, 2021, pp. 153-162.

○ S. Chandak, J. Neu, K. Tatwawadi, J. Mardia, B. Lau, M. Kubit, R. Hulett, P. Griffin, M. Wootters, T. Weissman and H. Ji; "Overcoming high nanopore basecaller error rates for DNA storage via basecaller-decoder integration and convolutional codes," *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020, pp. 8822-8826.

○ S. Chandak, K. Tatwawadi, C. Wen, L. Wang, J.A. Ojea and T. Weissman; "LFZip: Lossy compression of multivariate floating-point time series data via improved prediction," *2020 Data Compression Conference (DCC)*, Snowbird, UT, USA, 2020, pp. 342-351.

○ M. Goyal, K. Tatwawadi, S. Chandak and I. Ochoa; "DZip: improved general-purpose lossless compression based on novel neural network modeling," *2020 Data Compression Conference (DCC)*, Snowbird, UT, USA, 2020, pp. 372-372.

○ S. Chandak, K. Tatwawadi, B. Lau, J. Mardia, M. Kubit, J. Neu, P. Griffin, M. Wootters, T. Weissman and H. Ji; "Improved read/write cost tradeoff in DNA-based data storage using LDPC codes," *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Monticello, IL, USA, 2019, pp. 147-156.

○ A. Bhown, S. Mukherjee, S. Yang, S. Chandak, I. Fischer-Hwang, K. Tatwawadi and T. Weissman; "Humans are still the best lossy image compressors," *2019 Data Compression Conference (DCC)*, Snowbird, UT, USA, 2019, pp. 575-575.

○ M. Goyal, K. Tatwawadi, S. Chandak and I. Ochoa; "DeepZip: Lossless Data Compression using Recurrent Neural Networks," *2019 Data Compression Conference (DCC)*, Snowbird, UT, USA, 2019, pp. 558-558.

○ N. V. Desai, C. Juvekar, S. Chandak and A. P. Chandrakasan, "21.8 An actively detuned wireless power receiver with public key cryptographic authentication and dynamic power allocation," *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, San Francisco, CA, 2017, pp. 366-367.

Patents

○ D. Pavlichin, S. Chandak, T. Weissman, C. Burgess. Multi-domain configurable data compressor/de-compressor. *U.S. Patent No. 11,966,597* issued 23 Apr 2024.

○ C. Wen, L. Wang, E. Solowjow, S. Chandak, P. Tandon, and T. Weissman. Embedded compression for product lifecycle data management. *International Patent Application No. PCT/US2022/075452.*

○ O. Milenkovic, C. Schroeder, S. K. Tabatabaei, A. Aksimentiev, A. Hernandez, C. Pan, J. Liu, S. Chandak, B. Pham, M. Chen, and S. Shorkey. DNA-Based Data Storage Systems. *U.S. Patent Application No. 18/171,764.*

○ P. Tandon, S. Chandak, P. Pataranutaporn, Y. Liu, A. M. Mapuranga, P. Maes, T. Weissman, and M. Sra. Systems and Methods for Performing Video Communication Using Text-Based Compression. *U.S. Patent Application No. 17/808,507.*

○ S. Chandak and Y.H. Cheung. Genomic information compression by configurable machine learning-based arithmetic coding. *Patent Application No. PCT/EP2021/067960.*

○ S. Chandak and Y.H. Cheung. Improved quality value compression framework in aligned sequencing data based on novel contexts. *U.S. Patent No. 11,916,576* issued 27 Feb 2024.

○ S. Chandak and Y.H. Cheung. System and method for effective compression, representation and decompression of diverse tabulated data. *U.S. Patent Application No. 17/767,070.*

○ S. Chandak, K. Tatwawadi, I. Ochoa, M. Hernaez, and T. Weissman. Systems and Methods for Compressing Genetic Sequencing Data and Uses Thereof. *U.S. Patent Application No. 16/545,751.*

○ C. Wen, L. Wang, J. Aparicio, S. Chandak, K. Tatwawadi, and T. Weissman. Embedded Deep Compression for Time Series Data. *U.S. Patent Application No. 17/439,836.*

○ A. P. Chandrakasan, N. Desai, C. Juvekar, and S. Chandak. Detuning for a resonant wireless power transfer system including cryptography. *U.S. Patent No. 10,651,687* issued 12 May 2020.

Magazine articles

- I. Hwang, S. Chandak, K. Tatwawadi and T. Weissman; Forget JPEG, How Would a Person Compress a Picture?, *IEEE Spectrum* (October 2021).

Preprints

- K. Tatwawadi and S. Chandak; Tutorial on algebraic deletion correction codes, *arXiv (2019)*.

Talks and Panels

- Hands-on Workshop on Learned Image Compression (joint with Pulkit Tandon). *CVIP 2023*, IIT Jammu (remote), November 5, 2023. *Invited.*
- Efficient storage of and in DNA: genomic data compression & DNA based storage. *Biochemical Engineering and Biotechnology Department Seminar*, IIT Delhi, April 28, 2022. *Invited.*
- Moderator for panel on DNA-based data storage at *Stanford Compression Workshop 2021*, February 25-26, 2021.
- Genomic data compression. *Roche*, November 13, 2020. *Invited.*
- Error correcting codes for DNA based data storage. *ISMB/ECCB 2019*, Basel, July 21-25, 2019.
- SPRING: A next-generation compressor for FASTQ data. *ISMB/ECCB 2019*, Basel, July 21-25, 2019. *Invited.*
- SPRING: A next-generation compressor for FASTQ data. *Stanford Compression Workshop 2019*, Stanford, February 15, 2019. *Invited.*
- SPRING: A practical compressor for short-read FASTQ data. *56th Annual Allerton Conference on Communication, Control, and Computing*, Urbana, IL, October 3-5, 2018. *Invited.*

Poster

- Impact of lossy compression of nanopore raw signal data on basecalling and consensus accuracy. *ISMB/ECCB 2021*, Virtual, July 25-30, 2021.
- Error correcting codes for DNA based data storage. *ISMB/ECCB 2019*, Basel, July 21-25, 2019.
- SPRING: A practical compressor for short-read FASTQ data. *ISMB 2018*, Chicago, IL, July 6-10, 2018.
- Compression of genomic sequencing reads with and without preserving the order. *2017 Biomedical Computation at Stanford Symposium*, Stanford University, 10 April, 2017.

Experience

- **AWS Annapurna ML**   **USA**
  Sr. Applied Scientist   *December 2023–present*

- **Amazon S3**   **USA**
  Sr. Applied Scientist   *December 2022–December 2023*

- **Amazon S3**   **USA**
  Applied Scientist   *June 2021–November 2022*

- **Philips Research North America**   **Cambridge, MA, USA**
  Data Compression Research Intern   *June–September 2019*
  Worked on genomic data compression, supervised by Patrick Cheung.

- **Massachusetts Institute of Technology**   **Cambridge, MA, USA**
  Summer Intern   *May–July 2015*
  Worked on hardware implementation of Elliptic Curve Cryptography for IoT applications, supervised by Prof. Anantha Chandrakasan.

- **Oxford Brookes University**   **Oxford, UK**
  Summer Intern   *May–July 2014*
  Worked on electrode design for Functional Electrical Stimulation using numerical analysis, supervised by Prof. Cristiana Sebu.

Teaching

- EE 274 - Data Compression, Theory and Applications, Autumn 2022-23 and Autumn 2023-24.
  - Co-designed and taught class on data compression at Stanford University. Gave lectures on information theory basics, entropy coding, non-iid sources and Lempel-Ziv compression.

- Teaching Assistantships at Stanford University:
  - ENGR 76 - Information Science and Engineering, Spring 2020-21.
    - Helped design the first iteration of the class aimed at introducing the principles and techniques underlying the design of modern information, communication, and decision-making systems. The course is now included in the list of courses satisfying the Engineering Fundamentals Requirement at Stanford.
  - EE 276 - Information Theory, Winter 2019-20, Winter 2020-21.
  - EE 376A - Information Theory, Winter 2018-19.
    - Helped redesign information theory course to be accessible to wider audience, helped organize outreach event at local school, mentored course projects, and gave guest lecture on entropy rate and universal compressors.
  - EE 178 - Probabilistic Systems Analysis, Autumn 2017-18.

Other Experience

- Co-organizer of Stanford Compression Workshop in 2019 and 2021.
- Collaboration project with Siemens on compression of sensor data from 2017-21.
- Mentor for STEM to SHTEM summer internship for high school students in 2020.
- Reviewed papers for Nature Communications, Bioinformatics, Bioinformatics Advances, Briefings in Bioinformatics, Journal of Bioinformatics and Computational Biology, Computational and Structural Biotechnology Journal, PLOS One, IEEE Transactions on Computers, IEEE Access, Frontiers in Bioengineering, Computers in Biology and Medicine, ISMB/ECCB, ISIT, DCC, ITW, ICML Neural Compression Workshop and CVIP.
- Technical Program Committee member in the 8th International Conference on Computer Vision & Image Processing (CVIP-2023) held on November 3-5, 2023 at Indian Institute of Technology Jammu (IIT Jammu).

Course Projects

- **CS 349T/EE 192T: Video and Audio Technology for Live Theater in the Age of COVID**   *Autumn 2020-21*
  The course focused on developing technologies to enable live theater performance over the internet.
  - Developed systems for emoji and audio-based audience feedback along with audiovisual effects in the video player.
  - Researched low-latency video transmission using WebRTC to enable live audience interaction.

- **Implementation and analysis of stabilizer codes in pyQuil**   *Spring 2018-19*
  Stabilizer codes form a large family of quantum error correcting codes that includes well-known codes such as Shor code, Steane code, CSS codes and toric codes.
  - As part of "CS 269Q: Quantum Computer Programming", we built a framework for encoding and decoding of general stabilizer codes on pyQuil and tested specific single qubit codes with standard quantum noise models.

- **Understanding the Amazon Rainforest from Space using CNNs**   *Spring 2016-17*
  - Participated in Kaggle contest "Understanding the Amazon from Space" by Planet Labs, as part of course project for "CS231N: Convolutional Neural Networks for Visual Recognition".
  - Tested various architectures for multi-class, multi-label prediction of weather and land-use features based on satellite images of the Amazon rainforest.

Coursework

- **Electrical Engineering:** Modern Coding Theory, Universal Schemes in Information Theory, Information Theory, Convex Optimization, Digital Signal Processing, Digital Communications, Control Systems, Microprocessors

- **Computer Science:** Advanced Cryptography, Cryptocurrencies & Blockchain Technologies, Quantum Computing, Convolutional Neural Networks for Visual Recognition, Probabilistic Graphical Models, Machine Learning, Automata & Complexity Theory, Web Applications, Audio/Video technology for live theater

- **Statistics & Probability:** Advanced Probability & Random Processes

- **Mathematics:** Probabilistic Methods, Measure Theory, General Topology, Graph Theory, Abstract Algebra, Fourier Analysis, Complex Analysis, Real Analysis, Partial Differential Equations

- **Genomics:** Genomics, The Human Genome Source Code

- **Humanities & Social Sciences:** Economics, Sociology, Environmental Studies

Technical and Personal Skills

- **Programming:**
  - Proficient in: Rust, C++, Python (incl. NumPy)
  - Experience with: JavaScript, Tensorflow, Keras, Julia, Matlab, VHDL, Arduino
- Introductory Piano Class at Stanford.
- **Languages:** Hindi, English