

## Dr. Tirthajyoti Sarkar

37331 DUCKLING TER,  
FREMONT, CA 84536

Phone: (408) 823-3208,  
E-mail: tirthajyoti@gmail.com

### SUMMARY

Semiconductor technologist with 15+ years of experience in R&D and product innovation. Currently focused on applying data analytics/machine learning to semiconductor/electronics domain.

**LinkedIn profile:** <https://www.linkedin.com/in/tirthajyoti-sarkar-2127aa7/>

**Data Science, Machine Learning blogs:** <https://medium.com/@tirthajyoti>

**GitHub Homepage:** <https://tirthajyoti.github.io>

### PROFESSIONAL POSITIONS

**Position:** Senior Principal Engineer  
**Organization:** ON Semiconductor, Sunnyvale, CA  
**December 2009 - Present**

**Responsibilities:** (a) Leading AI/ machine-learning based projects: (a) design optimization framework development, (b) deep learning for chip optimization, (c) AI-based power IC, (b) Power Semiconductor technology and new product development (NPD) for applications in Automotive, Cloud infrastructure, gaming and AI, mobile, and industrial systems.

-----  
**Position:** Consulting Scientist  
**Organization:** AutonomiQ, Palo Alto, CA  
**April 2018 - Present**

**Responsibilities:** (a) Advising the software development/data science team about automatic test data synthesis and machine learning based generative model building, (b) Idea generation for the Natural Language Processing (NLP) module of the automated software testing platform.

-----  
**Position:** Author (Packt Publishing)  
**Responsibilities:** Authoring multiple Data Science books  
**September 2018 – Present**

-----  
**Position:** Postdoctoral Research Associate and Research Assistant  
**Institution:** University of Illinois at Chicago  
**August 2003 – November 2009**  
**Responsibilities:** Worked on Federal agency-sponsored R&D projects (NSF, DOE, ONR)

### FORMAL EDUCATION (Chronological Order)



**Master of Science (MS) in Analytics**, focus on data analysis, Machine Learning, and Big Data. *Georgia Tech, College of Engineering.*  
**August 2018 - now**



**Doctor of Philosophy (Ph.D.) in Electrical Engineering**, *University of Illinois at Chicago (UIC)*, Electrical and Computer Engineering.  
**March 2009**



**Bachelor of Technology (B.Tech.) in Instrumentation Engineering**, *Indian Institute of Technology (IIT), Kharagpur, India.*  
**June 2003**

### CONTINUING PROFESSIONAL EDUCATION (AI, Machine learning, Data science)



**Stanford Continuing Studies**, “Artificial Intelligence: An Introduction To Neural Networks And Deep Learning”, Stanford Univ.  
**July – August 2018**



**MIT Professional Program**, “Data Science: Data to Insights”, Massachusetts Institute of Technology, Continuing Education.  
**May 2017 – July 2017**

## CORE COMPETENCIES

- ✓ **Proven innovator:** Inventor/co-inventor on 4 issued U.S. patents, multiple U.S. patents are in application stage
- ✓ **Proven communicator of scientific study:** Author of 30+ international journal and conference papers in Tier-I category (IEEE or equivalent); Author of 2 book chapters/monographs. Keynote lectures in top IEEE conferences.
- ✓ 15+ years' experience with numerical computing and finite-element simulations
- **Software technology stack experience:**
  - *GitHub public profile:* <https://github.com/tirthajyoti>
  - *Machine Learning:* NumPy, SciPy, Scikit-Learn, Caret, kernlab
  - *Data wrangling, NLP, others:* Pandas, BeautifulSoup, SymPy, PySpark, SpaCy, NLTK, dplyr
  - *Deep Learning:* TensorFlow, Keras, PyTorch
  - *Statistical modeling:* Statsmodels, R-stats, JMP (SAS)
  - *Scientific computation:* MATLAB/Simulink, Octave
  - *Statistical Visualization:* Seaborn, ggplot2, Tableau, Plotly
  - *Web technologies (basic experience):* HTML5, CSS, JavaScript
  - *Cloud technologies (basic experience):* AWS, Google Colaboratory for machine learning
- ✓ **Analytics in manufacturing, quality, and product development:** 10+ years of experience in analysis of semiconductor data - multivariate statistics, advanced graphing, control charts, hypothesis testing.

## BOOK, OPEN-SOURCE PACKAGES & OTHER PERSONAL MACHINE LEARNING PROJECTS

**BOOK: *Hands on Mathematics for Data Scientists*:** Working on a book covering essential mathematics topics for machine learning and data science – set algebra, functions, calculus, statistics, optimization techniques, and linear algebra. Expected to be published in December 2019.



**Data Wrangling with Python**  
Dr. Tirthajyoti Sarkar, Shubhadeep Roychowdhury  
February 2019 New Release!

Simplify your ETL processes with these hands-on data hygiene tips, tricks, and best practices.

Quick links: [Table of contents](#) [What will you learn?](#)

Packt Subscription <b>FREE</b> \$9.99/m after trial	eBook <b>\$22.40</b> RRP \$31.99 <span style="color: red;">Save 29%</span>	Print + eBook <b>\$39.99</b> RRP \$39.99
---	--	--

**BOOK: *Data Wrangling with Python*:** Principal author of a book/courseware for [Packt publishing](#) on data wrangling techniques using Python. This course teaches the fundamentals of data scraping, cleaning, imputation, statistical plotting, and formatting for a machine learning pipeline from diverse sources like internet, legacy database vaults, or Excel financial tables. Published in January 2019.

**Python Package: Pydbgen:** This is a lightweight Python library for generating random database tables. Useful for beginners in data science when they want

to create SQL database tables with synthetic data for practicing machine learning and data extraction algorithms. It can generate Pandas DataFrame, MySQL and SQLite tables, and Excel files with random but contextual data such as name, address, city, zip code, phone number, birthday, license plate, organization, job title, etc. [Here is the detailed documentation](#)

**Python Package: UCI-ML API:** This is a simple and intuitive API written in Python to interface with the famous UC Irvine Machine Learning repository. It can help a user easily search and download relevant datasets or selectively choose a dataset based on its size or machine learning task category (regression, classification, clustering etc.). [Here is the detailed documentation.](#)

**Python Package: Design-of-Experiment:** Design of Experiment (DOE) is a critical activity for any scientist, engineer, or statistician planning to conduct scientific research. This set of codes is a collection of functions which wrap around the core packages (pyDOE and DiversiPy) and generate various types of DOE matrices (random, Latin hypercube, face centered design, factorial matrix) from an arbitrary range of input variables. [Read the detailed documentation here.](#)

**Random regression/classification problem set generation using symbolic input:** For beginners in data science and machine learning, a common problem is to get hands on good, clean data set for practicing various algorithms. Regression and classification are two most common supervised machine learning tasks. In this project, we demonstrate

a controllable way to generate these problems based on a well-defined function (involving linear, nonlinear, rational, or even transcendental terms) using symbolic expression from user. [Read my article on Medium about this project](#)

## **CURRENT MACHINE LEARNING BASED PROJECTS (PROPRIETARY)**

### **Semiconductor device design automation pipeline with machine learning**

**Goal:** This project aims to automate/aid the complex technology development and device design tasks in the field of high-power semiconductors, using machine learning, statistical modeling, and advanced optimization.

**Tools/Techniques used:** Scikit-learn, SciPy, JMP, R-part, nonlinear regressions, regularization, and various cross-validation strategies, decision tree, random forest, derivative-free optimization.

### **Deep learning-based semiconductor design feature extraction:**

**Goal:** To use deep learning framework to mimic 'high-level' design experience of human experts by classification of designs into categories such as 'sub-optimal' or 'aggressive'.

**Tools/Techniques used:** TensorFlow, Keras, PIL, OpenCV, Scikit-image

### **Low power neural network for Power IC controller**

**Goal:** Implement and embed deep learning function inside a controller IC for optimizing power conversion efficiency with a limited compute power and memory budget.

**Tools/Techniques used:** Keras/TensorFlow, SciPy.

### **Neural network-in-loop-PSICE modeling**

**Goal:** Build and deploy machine learning models and advanced optimization modules for finding best semiconductor die design for a given electrical target using physically-scalable SPICE modeling data.

**Tools/Techniques used:** Scikit-learn, Keras/TensorFlow, SciPy.

### **Python-based data analytics framework for semiconductor manufacturing:**

**Goal:** Build Python-based notebooks for analyzing silicon wafer data involving failure analysis, statistical plotting, normality analysis, causality discovery by multi-variate regression, outlier detection with Gaussian mixture models.

**Tools/Techniques used:** Scikit-learn, Statsmodels, Pandas, Seaborn.

## **OTHER PROFESSIONAL EXPERIENCE/SERVICE**

- Speaker and contributor, ValleyML – a non-profit working with ACM, IEEE Silicon Valley AI/CS/SSCS Chapters and other AI related organizations to cover the state-of-the art advances in AI technology
- Editorial Associate, *Towards Data Science*, Online publication
- Senior Member, IEEE; Electron Device Society (EDS) and Power Electronics Society (PELS)
- Chair, Semiconductor Committee, Power Supply Manufacturers' Association (PSMA)
- Industry Expert member, *Wide and Narrow bandgap technologies for Sustainable Energy Systems*, IEEE
- Topic Chair, *Sustainable Energy*, IEEE ECCE Conference, 2016, 2017, 2018
- Technical Track Chair - International Transportation Electrification Conference, 2017.
- Full member, Sigma Xi Scientific Research Society
- Visiting Lecturer, Indian Institute of Technology (IIT) Bombay, India, 2011-2012.