Georgia Institute of Technology
School of Electrical and Computer Engineering
Fall 2026

ECE 88xx: AI and Machine Learning for Semiconductor Manufacturing and Digital Twins

Instructor: Asif Khan
Email: akhan40@gatech.edu
Office: 153 Pettit
Office Hours: [Day/Time], or by appointment
Lecture: [Days] [Time], [Room]
Credit Hours: 3 (3-0-0-3)
Prerequisites: Graduate standing or permission of instructor. No prior machine learning experience required. Familiarity with semiconductor device physics or fabrication processes is expected.

1. Course Overview

This course introduces graduate students to the theory and practice of applying artificial intelligence and machine learning to semiconductor process technology, metrology, manufacturing, and digital twins. The semiconductor industry is undergoing a transformation driven by the increasing complexity of advanced process nodes and the explosion of data generated in modern fabrication facilities. AI/ML techniques and digital twin frameworks are now critical tools for process modeling, equipment monitoring, defect inspection, yield optimization, and accelerating technology development cycles.

Designed for students with a background in semiconductor devices and processes but no prior AI/ML experience, the course begins with a rigorous introduction to machine learning fundamentals using Python, then progressively applies these techniques to real-world semiconductor challenges. Topics include surrogate modeling for TCAD, physics-informed neural networks, virtual metrology, SEM defect classification, yield prediction, fault detection, digital twins for fab modules, and ML-assisted technology development.

The course emphasizes hands-on learning through four Python labs using open semiconductor datasets, two problem-set homeworks, and a substantial semester-long research project. Industry guest lectures provide direct exposure to how these methods are deployed in production fabs and R&D environments.

2. Course Outcomes

After successfully completing this course, students should be able to:

1. Design and evaluate ML models suitable for structured physical datasets in semiconductor environments.
2. Build ML-based surrogate models for semiconductor processes and apply optimization and control strategies.

3. Implement AI workflows for vitual metrology, defect classificiation, and wafer map defect pattern recognition.
4. Construct models for yield prediction, fault detection and classification, and predictive maintenance.
5. Apply AI methods to accelerate semiconductor R&D including surrogate modeling and materials informatics.
6. Describe how AI/ML methors are applied across the EDA pipeline and connect to manufacturing digital twins.
7. Critique research papers on AI/ML applications in semiconductor process technology and manufacturing.
8. Execute an original project applying AI/ML to a semiconductor process, metrology, or technology challenge.

3. Course Modules and Weekly Schedule

Module 1: Foundations (Weeks 1–4)

The first seven weeks gradually introduce AI concepts, taking students from zero Python experience through the complete ML toolkit needed for the application modules. We start with programming fundamentals and data handling, progress through supervised and unsupervised learning with semiconductor examples at every step, and conclude with neural network basics, model interpretability, and an introduction to physics-informed approaches. Every algorithm is taught in the context of structured physical data common in semiconductor environments.

Topics:

- Python programming and data handling for semiconductor environments (Jupyter, NumPy, Pandas, Matplotlib)
- Supervised and unsupervised learning: regression, classification, clustering, and dimensionality reduction
- Neural networks, deep learning fundamentals, and introduction to physics-informed approaches
- Model validation, interpretability (SHAP), and uncertainty quantification for semiconductor data

Datasets: Public UCI semiconductor datasets for introductory exercises

Outcome: Students will design and evaluate ML models suitable for structured physical datasets.

Module 2: Introduction to Semiconductor Processes and Process Control (Weeks 5–7)

Objective: Provide the semiconductor process context that grounds all subsequent application modules, and introduce AI-based approaches to process modeling and control.

This module bridges the gap between the ML foundations of Module 1 and the domain-specific applications that follow. Students review the key semiconductor fabrication processes (lithography, etch, deposition, CMP, implant) and the data they generate, then learn how ML can model these processes as an alternative to expensive physical simulation. The module introduces surrogate modeling, optimization techniques for recipe tuning, the fundamentals of

statistical and advanced process control, and how these components form the building blocks of digital twin systems for semiconductor manufacturing.

Topics:

- Semiconductor fabrication processes, process data, and statistical/advanced process control (SPC, APC, run-to-run)
- ML-based process modeling and TCAD surrogate models
- Optimization for recipe tuning (Bayesian optimization, design-of-experiments) and drift detection

Datasets: SECOM semiconductor manufacturing dataset (UCI)

Outcome: Students will build ML-based process models and apply optimization and control strategies to semiconductor fabrication problems.

Module 3: Metrology and Inspection (Weeks 8–10)

Objective: Apply AI to semiconductor measurement and inspection systems.

This module applies the ML toolkit to the measurement and inspection challenges central to semiconductor manufacturing and digital twin systems. Students learn to build virtual metrology models that predict wafer measurements from equipment sensor data, classify defect patterns on wafer maps and SEM images using transfer learning, and apply anomaly detection to high-volume fab data. The module also covers spectroscopic data inversion, CD and overlay prediction, and practical strategies for handling class imbalance and limited labeled data common in fab environments.

Topics:

- Virtual metrology and CD/overlay prediction from equipment sensor data
- Defect detection, classification, and wafer map analytics using image-based ML
- Anomaly detection and spectroscopic data inversion for high-volume fab data

Case Examples: SEM defect classification; inline CD prediction; wafer anomaly clustering

Datasets: WM-811K wafer map dataset (811,457 labeled wafer maps); SECOM sensor dataset for virtual metrology; SEM defect image collections

Outcome: Students will implement AI workflows for measurement data analysis and defect analytics.

Module 4: Yield and Smart Manufacturing (Weeks 11–12)

Objective: Develop AI-based optimization and control strategies for semiconductor fabrication yield and equipment health.

This module connects process and metrology data to manufacturing outcomes. Students learn how ML models can predict yield from inline measurements and process parameters, identify root causes of yield loss through feature importance and interpretability techniques, and detect equipment faults before they impact production. The module also covers predictive maintenance, reliability-aware process optimization, and the concept of digital twins for process chambers and fab modules.

Topics:

- Yield modeling, spatial yield analysis, and root-cause analysis with interpretable ML
- Fault detection and classification (FDC) from equipment sensor streams
- Predictive maintenance, tool health monitoring, and digital twins for process chambers

Case Examples: Yield loss root-cause analysis; tool drift prediction; CMP endpoint detection

Datasets: PHM 2016 CMP dataset (chemical-mechanical planarization sensor data); SECOM dataset for yield and fault detection

Outcome: Students will construct AI-driven models for yield prediction, fault detection, and equipment health monitoring in semiconductor manufacturing systems.

Module 5: Technology Development and Frontiers (Weeks 13–14)

Objective: Accelerate device and materials research using domain-informed AI.

This module surveys how AI is transforming semiconductor R&D — from materials discovery and device modeling to reliability prediction and autonomous experimentation. Students see how surrogate models can replace expensive TCAD simulations, how ML enables inverse design of device structures, and how data-driven compact models accelerate the path from device physics to circuit design. The module also introduces emerging topics including LLMs for fab knowledge management, AI for advanced packaging, and digital twin architectures that integrate sensing, control, and modeling. Includes an industry guest lecture on frontier applications.

Topics:

- Materials informatics, ML-driven discovery, and surrogate modeling for device physics
- Reliability and lifetime prediction; variability and defect statistics modeling
- Inverse design, data-driven compact modeling, and reinforcement learning for process tuning
- Digital twin architectures, LLMs for fab knowledge management, and AI for advanced packaging

Case Examples: ML-based retention prediction in emerging memories; surrogate electrostatics modeling for nanosheet FETs; statistical lifetime modeling under accelerated stress

Datasets: Materials Project database for materials informatics; published accelerated stress test datasets

Outcome: Students will develop AI models that accelerate semiconductor R&D and device innovation.

Module 6: AI in Electronic Design Automation (Week 15)

Objective: Survey how AI/ML is transforming electronic design automation, providing the design-side complement to the process and manufacturing focus of the course.

This module broadens the course perspective to the design side of the semiconductor pipeline. Industry guest lecturers introduce how ML is applied to key EDA challenges including placement and routing, timing optimization, layout verification, and design-technology co-optimization (DTCO). Students gain an understanding of how process-aware design decisions connect to the manufacturing and digital twin frameworks covered in earlier modules, closing the loop between design and fabrication.

Topics:

- ML for placement, routing, and timing/power optimization
- AI for lithography hotspot detection, layout verification, and design-technology co-optimization (DTCO)
- Generative AI for circuit and layout design

Datasets: ISPD and ICCAD benchmark circuits for placement and routing (demonstration by guest lecturers); open-source PDK examples

Outcome: Students will understand how AI/ML methods are applied across the EDA pipeline and articulate the connections between design automation and process-aware manufacturing.

Module 7: Project Presentations (Week 16)

Objective: Demonstrate integrated understanding by presenting original research applying AI/ML to semiconductor challenges.

Student teams present their semester research projects in a conference-style oral format (10 min presentation + 5 min Q&A). Presentations should clearly communicate the problem motivation, the AI/ML approach taken, key results, and honest discussion of limitations. The session concludes with a course synthesis reflecting on how the modules connect and where the field is heading. Final written reports are due at the end of Week 16.

Outcome: Students will present and defend an original project applying AI/ML to a semiconductor process, metrology, or technology development challenge.

Detailed Weekly Schedule

Week	Topic	Description	Deliverables
1	***MODULE 1: FOUNDATIONS*** **Course Introduction; Semiconductor Process Overview**	Overview of modern fab processes (lithography, etch, deposition, CMP, implant). Where AI and digital twins fit in the semiconductor value chain. Course logistics and project expectations.	*Survey: background self-assessment*
2	**Python for Process Data; ML Fundamentals I**	Python refresher (NumPy, Pandas, scikit-learn). Types of process data: tabular sensor logs, time-series equipment traces, wafer maps, SEM images. Supervised learning: regression and classification.	*Lab 1 assigned: Python + exploratory data analysis on fab sensor dataset*
3	**ML Fundamentals II: Model Selection and Validation**	Training/validation/test splits for process data. Cross-validation, overfitting, bias-variance trade-off. Feature engineering for semiconductor data (recipe parameters, chamber conditions, wafer history).	
4	**Deep Learning Foundations for Process Engineers**	Neural network fundamentals: MLPs, activation functions, backpropagation. Introduction to CNNs for image data. Practical PyTorch tutorial. Transfer learning basics.	*Lab 1 due; Lab 2 assigned: Build and train a regression model for process parameter prediction*
5	***MODULE 2: SEMICONDUCTOR PROCESSES AND PROCESS CONTROL***	Replacing expensive TCAD simulations with ML surrogates. Gaussian process regression for process modeling.	*HW 1 assigned: Process*

Week	Topic	Description	Deliverables
	ML for Process Modeling and TCAD Surrogate Models	Bayesian optimization for recipe tuning and design-of-experiments.	*modeling problem set*
6	**Physics-Informed Machine Learning for Semiconductor Processes**	Encoding physical laws (diffusion, reaction kinetics, heat transfer) into neural networks. PINNs for etch and deposition modeling. Hybrid physics-ML models. Domain adaptation.	*Lab 2 due*
7	**Industry Lecture: ML in Process Development**	Guest speaker from industry on real-world applications of ML in process R&D. Case study discussion.	*HW 1 due; Project proposal due*
8	***MODULE 3: METROLOGY AND INSPECTION*** **AI for Semiconductor Metrology**	Virtual metrology: predicting wafer measurements from equipment sensor data. CD-SEM and OCD model-based vs. ML-based measurement. Reducing metrology sampling while maintaining quality.	*Lab 3 assigned: Virtual metrology pipeline on sensor + metrology dataset*
9	**Defect Detection and Classification**	Wafer defect map analysis (WM-811K dataset). CNN architectures for SEM defect classification. Anomaly detection for novel defect types. Class imbalance strategies in fab data.	
10	**Advanced Inspection and Image Analysis**	SEM image segmentation and feature extraction. Overlay and pattern fidelity analysis with ML. Generative models for synthetic defect data augmentation. Active learning for labeling efficiency.	*Lab 3 due; HW 2 assigned: Metrology and inspection problem set*
11	***MODULE 4: YIELD AND SMART MANUFACTURING*** **Yield Prediction and Modeling**	Connecting process parameters to yield with ML. Spatial yield modeling across wafers and lots. Feature importance and interpretability for yield excursion root-cause analysis.	
12	**Fault Detection, Classification, and Smart Fab**	Real-time FDC from equipment sensor streams. Predictive maintenance for fab tools. Run-to-run control enhanced with ML. Digital twins of process chambers.	*HW 2 due; Lab 4 assigned: Fault detection on equipment sensor time-series data*
13	***MODULE 5: TECHNOLOGY DEVELOPMENT AND FRONTIERS*** **AI for Technology Development and DTCO**	ML-assisted exploration of new device architectures (GAA, CFET, backside power). Design-technology co-optimization. Generative models and RL for accelerating process-of-record development.	*Lab 4 due*
14	**Industry Lecture: Frontiers and Emerging Topics**	Guest speaker(s). Topics may include: LLMs for fab knowledge management,	

Week	Topic	Description	Deliverables
		autonomous experimentation, neuromorphic process control, AI in advanced packaging.	
15	***MODULE 6: AI IN EDA*** **AI in Electronic Design Automation**	Guest lectures on AI/ML in EDA. Topics include ML for placement and routing, timing and power optimization, lithography hotspot detection, layout verification, DTCO, and generative AI for circuit design. Discussion of how design-side and process-side AI connect.	
16	***MODULE 7: PROJECT PRESENTATIONS*** **Project Presentations and Course Wrap-Up**	Student teams present semester projects (oral presentation + Q&A). Course synthesis and future outlook. Final project reports due.	*Final project report due*

Note: Schedule is approximate and may be adjusted. Industry lecture dates will be confirmed during the semester.

4. Assignments and Labs

4.1 Python Labs (4 labs)

Hands-on Jupyter notebook labs guide students through building complete ML pipelines on semiconductor datasets. Each lab takes approximately two weeks and includes starter code, data exploration tasks, model building, and interpretation questions. All labs use Python with scikit-learn, PyTorch, and standard data science libraries.

Lab 1: Exploratory Data Analysis on Fab Sensor Data

Load and explore a semiconductor manufacturing sensor dataset. Practice data cleaning, visualization, feature statistics, and correlation analysis. Identify candidate features for process modeling. Dataset: UCI semiconductor manufacturing datasets.

Lab 2: Process Parameter Prediction with Regression Models

Build regression models (linear, random forest, neural network) to predict a target process outcome (e.g., film thickness, etch depth) from recipe and sensor parameters. Compare model performance, interpret feature importance. Dataset: SECOM semiconductor manufacturing dataset (UCI).

Lab 3: Virtual Metrology Pipeline

Construct an end-to-end virtual metrology system using equipment sensor traces to predict wafer-level metrology measurements. Address temporal alignment, feature extraction from time-series, and deployment considerations. Dataset: SECOM sensor dataset for virtual metrology.

Lab 4: Fault Detection on Equipment Sensor Streams

Apply anomaly detection and classification methods to equipment sensor time-series data. Detect tool faults and process excursions. Explore windowing strategies, autoencoders, and threshold-setting for production use. Dataset: PHM 2016 CMP dataset; WM-811K wafer map dataset.

4.2 Homework Assignments (2 assignments)

Problem sets that combine analytical and computational questions. HW 1 covers process modeling concepts (surrogate models, Bayesian optimization, physics-informed learning). HW 2 covers metrology, inspection, and yield modeling. Each includes short-answer conceptual questions and Python coding problems.

4.3 Paper Critiques

Throughout the semester, students will read and submit brief structured critiques (1 page each) of 4–5 assigned research papers. Critiques should summarize the paper's contribution, evaluate its methodology, identify limitations, and suggest extensions. This prepares students for the literature review component of their project.

5. Semester Project

The semester project is the centerpiece of the course (50% of grade: proposal + presentation + report). Teams of 2–3 students will identify a semiconductor process or metrology problem, acquire or generate appropriate data, apply ML techniques studied in the course, and present their findings in both oral and written form.

5.1 Project Scope

Projects should address a concrete problem in semiconductor process technology, metrology, or manufacturing. Example topics include:

- ML surrogate model for a specific etch or deposition process using TCAD-generated or experimental data
- Defect classification system for a particular defect type using publicly available SEM or wafer-map datasets
- Virtual metrology model for predicting CD or overlay from equipment sensor logs
- Physics-informed neural network for modeling a thermal or diffusion process step
- Yield prediction model correlating inline metrology and process parameters to die-level yield
- Anomaly detection system for chamber health monitoring using synthetic or real sensor data
- Bayesian optimization framework for process recipe tuning in a simulated environment
- Comparative study of ML architectures for a specific semiconductor data modality (time-series, images, tabular)

Students are encouraged to propose topics related to their own research. The instructor will help refine scope during office hours.

5.2 Project Milestones

Due	Milestone	Details
Week 7	Project Proposal (1–2 pages)	Problem statement, proposed ML approach, dataset plan, team members
Week 10	Progress Check-In	Brief in-class update: data acquired, initial results or obstacles
Week 16	Oral Presentation (10 min + 5 min Q&A)	Present motivation, methods, results, and conclusions to class
Week 16	Final Report (8–12 pages, conference format)	Introduction, related work, methodology, experiments, results, discussion, references

5.3 Evaluation Criteria

Projects are evaluated on: (1) problem motivation and relevance to semiconductor technology, (2) appropriate selection and application of ML methods, (3) rigor of experimental methodology (proper data splits, baselines, ablations), (4) quality of analysis and interpretation of results, (5) clarity of written report and oral presentation, and (6) novelty or insight contributed.

6. Grading

Component	Weight
Homework Assignments (2)	15%
Python Labs (4)	20%
Semester Project Proposal	5%

Semester Project Presentation	15%
Semester Project Final Report	30%
Paper Critiques / Participation	15%

Letter grades follow the Georgia Tech scale. The instructor reserves the right to curve final grades upward.

Late Policy: Assignments submitted up to 48 hours late will receive a 15% penalty. Submissions beyond 48 hours will not be accepted without prior arrangement. Project milestones cannot be submitted late.

7. Resources and References

There is no required textbook. The course draws on research papers, open-access tutorials, and instructor-prepared notes. The following references are recommended for supplementary reading:

7.1 Machine Learning Foundations

- A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 3rd ed. Sebastopol, CA, USA: O'Reilly Media, 2022.
- I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: https://www.deeplearningbook.org
- C. M. Bishop and H. Bishop, *Deep Learning: Foundations and Concepts*. Cham, Switzerland: Springer, 2024.

7.2 Semiconductor Processes and Manufacturing

- G. S. May and C. J. Spanos, *Fundamentals of Semiconductor Manufacturing and Process Control*. Hoboken, NJ, USA: Wiley-IEEE Press, 2006.

7.3 Key Research Papers (Curated Reading List)

A curated list of ~20 papers will be provided on Canvas at the start of the semester, organized by module. Papers will be drawn from venues including IEEE Transactions on Semiconductor Manufacturing, SPIE Advanced Lithography, IEDM, ICCAD, and relevant ML conferences (NeurIPS, ICML workshops on scientific ML).

7.4 Open Datasets

- **WM-811K Wafer Map Dataset:** 811,457 wafer maps with labeled defect patterns for classification tasks
- **SECOM Dataset (UCI):** Semiconductor manufacturing process sensor data with pass/fail labels for fault detection
- **PHM 2016 CMP Dataset:** Chemical-mechanical planarization process sensor data for predictive maintenance

7.5 Software and Computing

All labs and assignments will use Python 3.x with the following libraries: NumPy, Pandas, Matplotlib, scikit-learn, PyTorch, and Jupyter Notebooks. Students will have access to Georgia Tech's PACE computing cluster for GPU-accelerated training. Setup instructions will be provided in Week 1.

8. Course Policies

8.1 Academic Integrity

All students are expected to comply with the Georgia Tech Academic Honor Code: http://www.policylibrary.gatech.edu/student-affairs/academic-honor-code. Collaboration on labs and homework is encouraged for discussion, but all submitted code and writing must be your own (or your team's, for the project). Use of AI coding assistants (e.g., GitHub Copilot, ChatGPT) is permitted for labs and homework as a learning tool, but must be disclosed. AI tools may not be used for paper critiques.

8.2 Accommodations for Students with Disabilities

If you have a disability and require accommodations, please contact the Office of Disability Services at https://disabilityservices.gatech.edu/ to establish your eligibility. Then provide the instructor with your accommodation letter as early as possible so that appropriate arrangements can be made.

8.3 Absences and Attendance

Attendance is expected for all lectures and is essential for industry guest lectures. The Institute Absence Policy applies: http://www.catalog.gatech.edu/rules/4/. Students who must miss class should notify the instructor in advance and are responsible for obtaining notes and materials from classmates.

8.4 Student-Faculty Expectations Agreement

At Georgia Tech we believe that it is important to continually strive for an atmosphere of mutual respect, acknowledgment, and responsibility between faculty members and the student body. See http://www.catalog.gatech.edu/rules/22/ for an articulation of some combinated expectations that we have of each other and of ourselves.