2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC

Host Sponsor UNC Charlotte

Paper ID #52743

Work-In-Progress: Development of an Solid State Devices and Materials Novice Chatbot to Improve Critical Thinking

Vignesh Balasundaram Sathiya Devi, Columbia University in the City of New York
Dr. Christopher V.H.-H. Chen, Columbia University

Christopher V.H.-H. Chen, Ph.D., is a Lecturer in the Discipline of and Director of Masters Studies in Chemical Engineering at Columbia University. He completed his PhD in Chemical and Biological Engineering at Princeton University; MBA at Columbia Business School; and BS and MS in Chemical Engineering at Stanford University. His teaching and research interests include the application of case- and problem-based approaches to STEM teaching; improving critical thinking with educational technology (e.g., AI); integrating sustainability and other human considerations into the engineering training; and preparing graduate students as future leaders.

Work-in-Progress: Development of Solid-State Devices and Materials Novice Chatbot to Improve Critical Thinking

Abstract

Four years after the release of ChatGPT, debates on the use of generative AI in higher education continue. Educators are still seeking guidance and examples of how to integrate these tools into courses to intentionally prepare students for an AI-enabled future. In Electrical Engineering, some connections with AI are more obvious–where students are expected to use these tools to code and optimize systems, and more curricula are beginning to teach students how these systems work–but even these uses will require students to develop more complex skills to appropriately leverage these tools. In particular, critical thinking as an engineering skill (e.g., articulating and challenging assumptions in problem solving, designing experiments, and collaborating with other engineers)–skills that map to multiple ABET Student Outcomes–will become more important than ever, as AI easily generates persuasive but false arguments.

To prepare students for using AI in their work while helping them build critical thinking skills in a disciplinary context, we have developed a "novice" chatbot for Solid State Devices and Materials (SSDM), a required core course in the electrical engineering major that most students take in their junior year. The chatbot is introduced to students late in the course and is used generally as a resource for students to practice course concepts, and specifically in conjunction with targeted exercises to help students practice evaluating AI output. However, due to deployment challenges, no student data was collected during this course offering.

In this work-in-progress paper, we present a customized Retrieval-Augmented Generator (RAG) chatbot built on ChatGPT tuned to perform worse than the base Large Language Model (LLM). We share the methods used to test chatbot performance, with results showing a B+ average approaching our novice target, and proposed learning activities that focus on developing students' critical thinking skills (primarily through critiquing proposed AI-generated solutions). In particular, we targeted the following concepts, in which students frequently make mistakes: temperature dependence; drift; mobility; diffusion of carriers; and metal-semiconductor junctions and contacts. This is opposite to how many others have used chatbots to support student learning (e.g., automated TAs). The approach of implementing an intentionally errant AI chatbot is novel in the engineering education literature, with only one other known application in chemical engineering. As such, our project tests on whether the strategies employed to create a novice chatbot for chemical engineering still hold for a more conceptual, text-driven course. Our initial data shows that this concept seems to transfer, with the SSDM "novice" chatbot underperforming other commercial chatbots in a battery of questions designed to trigger misconceptions common to students. This paper presents a feasibility study of the degradation approach, demonstrating that a degraded (B+ or lower) performance level is achievable through RAG tuning and system prompting with classroom deployment and student outcome assessment are planned for Fall 2026 cycle. By sharing our experience, we hope to demonstrate that approaches to create novice chatbots are transferable, and that intentionally trained chatbots can help students learn to critically interact with AI in their future engineering practice.

Introduction

The rapid emergence of Artificial Intelligence (AI) has created ripples across both industry and academia. In industry, AI development is typically driven by access to large, proprietary datasets and the ability to integrate advanced models directly into existing engineering tools and workflows. Major electronic design automation (EDA) vendors have begun embedding AI capabilities into their platforms to streamline design flows, automate routine tasks, and support engineers in day-to-day decision-making. These industrial AI systems tend to be highly optimized for performance, tightly coupled with commercial design environments, and focused on improving productivity and design quality.

The empirical research on the viability and feasibility of current AI platforms as well as their alignment with and capacity to inform various pedagogical approaches to instruction is necessary for the long-term deployment of AI tools in education. In academic settings, however, AI development often occurs on a smaller scale, using localized or domain-specific datasets with inherently higher variation [1]. Rather than optimizing large-scale deployment, academic AI efforts typically aim to support small scale teaching and learning. As a result, AI tools in academia are more exploratory and aimed at enhancing the efficiency of administrative staff and students [2]. However, there are several difficulties and restrictions with using these new technologies, most of which have to do with ethics [3].

In the past, studies demonstrated the promise of a large language model (LLM)-based assistants in introductory engineering courses [4]. These systems are not designed to replace instructors, but to augment instruction and create richer, more interactive learning experiences for students. AI is transforming education, closing gaps, and fostering a more inclusive and productive learning environment by personalizing learning experiences, automating administrative tasks, and providing real-time feedback. Due to the significance of incorporating AI into education, considering all of AI's implications–intentional or not–is necessary for instructors [5].

The field of Electronics and Solid-State Devices and Materials (SSDM) form the base of modern technology, such as advancements in computing, and telecommunications, and is a foundational pillar of the electrical engineering curriculum. However, its conceptual foundation, built upon abstract quantum mechanical principles and complex physics, presents a significant pedagogical challenge, particularly for novice learners. A primary bottleneck for SSDM students is the inability to identify and articulate the underlying assumptions in device physics and to see how complex concepts are built upon foundational assumptions. Therefore, it is important for students to come up with questions to get an understanding of the underlying concepts. Prior studies indicate that students who engage in Dialogue-Based Learning (DBL) demonstrate stronger practical knowledge and potentially greater long-term knowledge retention compared with those in Lecture-Based Learning (LBL) [6]. However, variations in class size can make the implementation of DBL challenging in traditional classroom settings [7]. In this context, the proposed novice chatbot offers a practical solution to support a sort of dialogue-based instructional practice.

To address some of the challenges students have with identifying and articulating assumptions in device physics, we have developed a specialized AI chatbot, designed for SSDM students, with the primary objective of encouraging students to critique the output of chatbots and actively foster the critical thinking skills essential for a successful career in electrical engineering. This customized Retrieval-Augmented Generator (RAG) chatbot built on ChatGPT 4-o is trained on SSDM course materials, but then tuned to have degraded, “novice” responses to SSDM topics (targeting ~75% correct or a C level performance). This approach of creating an intentionally errant chatbot has not yet been tried in the electrical and computer engineering education literature, as far as the authors are aware. The only example of such a chatbot in chemical engineering [4]. By creating a novice AI, this chatbot is conceived not as a mere answer-giving tool, but as a partner that can engage students in collaborative problem-solving similar to that of working with peers in the classroom. As modern engineering is a profoundly collaborative endeavor, using AI as a synthetic team member could encourage students to justify their reasoning when presented with an alternative, or incorrect, solution by the “novice” chatbot.

Our approach of using design learning activities around an intentionally errant chatbot is grounded in productive failure theory [8], which posits that students who struggle to identify flaws in proposed solutions develop deeper conceptual understanding than those who receive correct answers directly. So, rather than shielding students from AI errors, exposing them to confident but incorrect reasoning activates the generative processing needed for durable learning, particularly for the kind of assumption-checking central to device physics. Keeping this framework in mind will help us develop and make use of our novice chatbot in SSDM.

This work-in-progress paper describes the development of the novice SSDM chatbot, its evaluation against other commercial chatbots in terms of accuracy, and the design of learning activities that integrate the chatbot into the course. Though we are still tuning the chatbot’s performance and evaluating students’ engagement with it, we hope that by sharing our experience we can encourage our colleagues to consider how to integrate generative AI tools in their courses to promote instead of discouraging critical thinking.

Methods

Chatbot Development and Testing. The chatbot is created using a base model of OpenAI’s ChatGPT 4-o and tuned similarly to a previously developed novice chatbot [4]. The chatbot is trained on course materials including lecture notes and textbook chapters, excluding problem sets, which are preprocessed into embeddings. The chatbot uses RAG to answer student questions using the embedded course content. The chatbot’s system prompt, which informs the chatbot of its purpose, was as follows:

> “You are a student in Solid State Devices and Materials. You commonly make mistakes of the following types: you confuse the direction of temperature dependence on resistance for various materials; you mistake the direction charge carriers move due to drift and carrier diffusion; and you make conceptual mistakes in metal-semiconductor junctions and contacts. Regardless of whether you are correct, you are always confident in your answer and will double down on your explanations. Your tone should be that of a learning college student.”

Students interacted with the chatbot through a simple web interface–an application programming interface (API) showing a chat window, which was accessible through Canvas, our learning management system (LMS). They could then type to send messages to the chatbot that accessed

ChatGPT via their API to generate responses. A sample exchange that shows the simple interface and how results are presented to students can be seen in Figure 1.

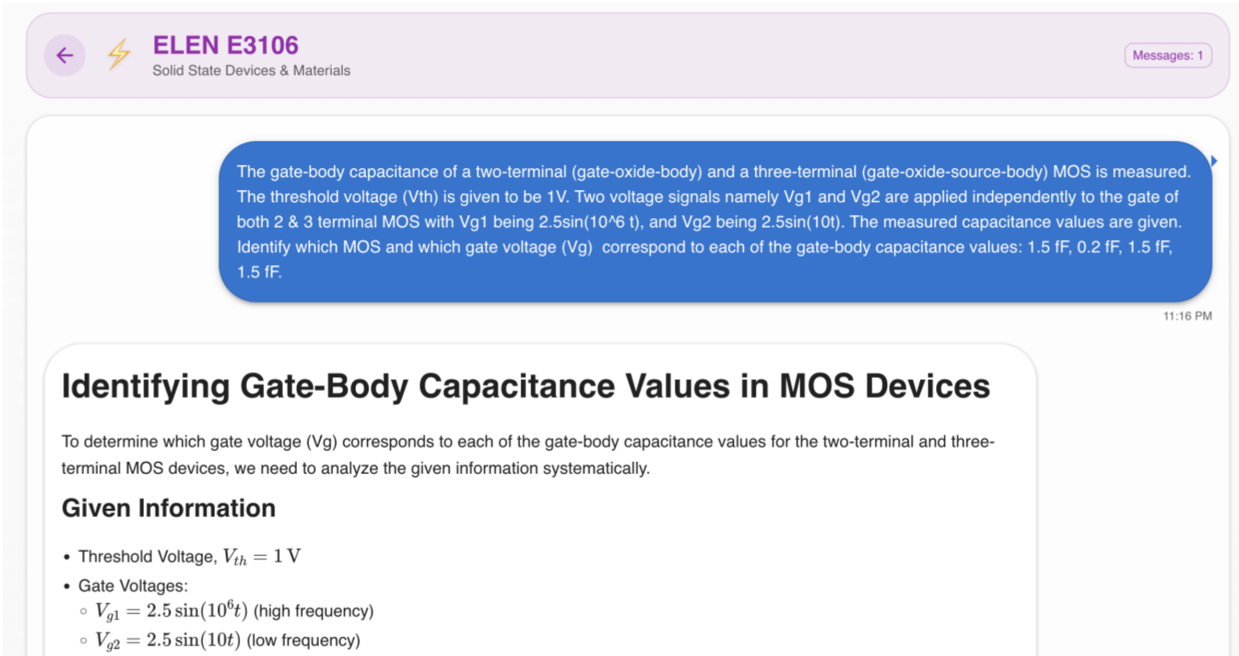

Figure 1. Screenshot of the SSDM "novice" chatbot with a sample question used for testing.

Testing and development of the chatbot was done by two graduate students involved with the undergraduate SSDM course. Based on their experience, and with engagement with the course instructor, the following concepts were targeted in which students frequently make mistakes: temperature dependence; drift; mobility; diffusion of carriers; and metal-semiconductor junctions and contacts (i.e., MOSFET-related questions). Three testing questions, designed to trigger common misconceptions in their respective topics, were developed for each of these five concept areas at different levels of depth: Concepts and Definitions, Qualitative Prediction, and Quantitative Application. This "misconception" inventory (shown in the Appendix) was then used to compare the performance of the SSDM novice chatbot versus commercial chatbots that students may already access. Namely, we compared against Anthropic Claude, ChatGPT (4-o and 5), and Google Gemini. Two graders then assessed the chatbot responses into a grade (A through F) with the following guidelines:

- **A** Correct with only insignificant errors (e.g., equation copy error or numerical error)
- **B** Mostly correct (e.g., minor conceptual or logical errors)
- **C** Not great (e.g., multiple errors but still generally in the right conceptual direction)
- **D** Mostly incorrect (e.g., multiple conceptual errors with some correct content)
- **F** Bad info or gibberish (e.g., scattershot response or nonsensical answer)

Once the two graders finished reviewing the responses, an average GPA was calculated for each chatbot (A $\rightarrow$ 4, F $\rightarrow$ 0) to holistically evaluate the chatbot's abilities in SSDM.

Course Description & Class Use of the Chatbot. SSDM is a required core course in the electrical engineering major that most students take in their junior year. Due to IT challenges in creating and serving our project (namely, access challenges due to new security rules put in place for which domains can access AI-related projects on our campus), the SSDM chatbot was deployed too late to collect data from students using the chatbot in the class. However, we had designed activities for students with the SSDM chatbot as study exercises for the final exam.

Results & Discussion

Evaluation of Commercial Chatbots for SSDM Concepts. We tested commercial chatbots with the 15-question "misconception" inventory on December 24 and 25, 2025. As these chatbots are always changing and being updated, our evaluation of these questions against these chatbots may not necessarily be representative of future performance. The two graders independently evaluated outputs from the chatbots on an A-F scale, and the scores were numerically averaged to evaluate the performance of these chatbots on our SSDM questions, summarized in Table 1. Each grader evaluated a total of 70 AI responses for this trial with 84% agreement, and an inter-rater reliability score (Cohen's κ) of 0.55. This indicates moderate agreement between the two graders, which is acceptable for our initial assessment. As this project continues, alignment between the graders will occur through further discussions of disagreed grades to ensure increased agreement and inter-rater reliability. The performance of these chatbots by problem type (Concepts and Definitions, Qualitative Prediction, Quantitative Application) and concept area are shown in Tables 2 and 3, respectively.

Table 1. Overall Performance of Commercial Chatbots on SSDM Concepts

Chatbot	ChatGPT 4-o	ChatGPT 5	Anthropic Claude	Google Gemini	SSDM Chatbot
Average Grade	3.14	3.39	3.93	3.79	3.35
Letter Grade	B	B+	A	A-	B+

The comparative study on how each model responds to the SSDM curriculum reveals that commercial AI models generally outperform the novice SSDM chatbot, as shown in Table 1. Anthropic Claude performed the best with an average letter grade of A, while our novice SSDM chatbot obtained a letter grade of B+, which is trending towards our target accuracy of 75%.

Table 2. Performance of Commercial Chatbots on SSDM Concepts by Problem Type

Problem Type	ChatGPT 4-o	ChatGPT 5	Anthropic Claude	Google Gemini	SSDM Chatbot
Concepts and Definitions	3.50	3.70	4.00	4.00	3.60
Qualitative Prediction	3.60	3.80	3.80	4.00	4.00
Quantitative Application	2.13	2.50	4.00	3.25	2.38

From Table 2, almost all the compared LLM models perform the best, except for the novice SSDM chatbot and ChatGPT 4-o, on both Concepts and Definition and Qualitative Prediction type of questions. We also observed that chatbots were generally more prone to mistakes in

Quantitative Application type questions, as was previously observed in the other novice chatbot paper [4], with Anthropic Claude as the exception, here. This behavior is preferred for our application, as we would still want our chatbot to help students understand the concepts and definitions and make qualitative predictions in the course, which the SSDM chatbot does (at least to a similar degree as ChatGPT, its base LLM). However, as our focus is in helping students develop the ability to critique proposed solutions through productive failure theory [8], degraded behavior in Quantitative Application type questions is primarily where we want the SSDM chatbot to make mistakes to facilitate our pedagogical intentions.

Table 3. Performance of Commercial Chatbots on SSDM Concepts by Concept Area

Concept Area	ChatGPT 4-o	ChatGPT 5	Anthropic Claude	Google Gemini	SSDM Chatbot
Temperature Dependence	3.67	3.33	4.00	4.00	2.75
Drift	3.33	4.00	4.00	4.00	4.00
Mobility	2.67	3.00	3.67	3.67	3.50
Diffusion of Carriers	4.00	4.00	4.00	4.00	4.00
MOSFET	2.33	2.83	4.00	3.33	2.50

When presenting the data by different SSDM concept areas, as shown in Table 3. All chatbots performed well in fundamental areas such as mobility and drift. MOS or MOSFET questions were found to be the area in which the chatbots performed the worst, with these complex questions often requiring intuitive reasoning and the ability to interconnect multiple concepts to answer. This particular question was able to cause all chatbots to produce an errant response because it combines CV characteristics of MOS, high and low frequency variations to CV characteristics and requires one to understand the behavior of two and three terminal MOS. We do note, though, that the initial response of Anthropic Claude was incorrect initially, but it was able to correct itself as conversation with the chatbot on the question continued.

As an example of how graders evaluated a specific question, consider Question 15 (Appendix). The correct answer requires recognizing that the 3-terminal MOS capacitance is frequency-independent in strong inversion (1.5 fF for both V_{g1} and V_{g2}), while the 2-terminal MOS capacitance is frequency-dependent (1.5 fF at low frequency V_{g2}, 0.2 fF at high frequency V_{g1}). Graders used this expected answer to assess whether chatbot responses correctly identified both the terminal configuration and the frequency dependence–any conflation of the two cases was marked down.

One notable takeaway identified during the chatbot testing was that the latest commercially available AI models do not limit their responses to the level of a teaching assistant for a sophomore-level SSDM course, even when prompted to do so. In contrast, the developed SSDM novice chatbot attempts to generate a convincing solution while limiting itself to the lecture notes that underline the RAG model, demonstrating the importance of using RAG to generate our novice chatbots. However, our novice chatbot does seem to hallucinate (e.g., create references–

such as chapter numbers or lecture slide decks–for information used in its responses) more than the commercial chatbots do. This is not necessarily a bad thing, as the novice chatbot is designed to encourage students to be more careful of AI output, but this observation is notable as part of our further development of the SSDM chatbot.

Based on our data, almost all chatbot models–both commercially available and customized–are suitable for dialogue-based learning activities related to SSDM courses. Differences in responses emerge when the chatbots are prompted with complex quantitative questions that require a certain degree of intuitive thinking.

Future Work

In this work-in-progress paper we have presented how we developed a novice SSDM chatbot and tested its performance against commercial chatbots. Our initial work shows that it is possible to transfer the concept of a "novice" chatbot from its original use in chemical engineering into electrical engineering. Due to technical challenges in creating and serving the chatbot, student access to the chatbot was offered too late to collect data. As such, we plan to use the chatbot in the future during its next run in Fall 2026 to better design activities for students to engage with the chatbot, and to collect data on how students are using the custom novice chatbot in the class.

AI Use Statement

ChatGPT was used to refine our statements and ensure the texts are grammatically correct. Perplexity AI was used as a tool in the literature review process.

References

[1] O. Ali, P. A. Murray, M. Momin, Y. K. Dwivedi, and F. T. Malik, "The effects of artificial intelligence applications in educational settings: Challenges and strategies," *Technological Forecasting and Social Change*, vol. 199, no. 123076, pp. 123076–123076, Feb. 2024, doi: https://doi.org/10.1016/j.techfore.2023.123076.

[2] M. Aleedy, E. Atwell, and S. Meshoul, "Using AI Chatbots in Education: Recent Advances Challenges and Use Case," 3rd International Conference on Sustainable and Innovative Solutions for Current Challenges in Engineering & Technology ICSISCET, November 2021.

[3] C. Kooli, "Chatbots in education and research: a critical examination of ethical implications and solutions," *Sustainability*, vol. 15, no. 7, p. 5614, Mar. 2023, doi: https://doi.org/10.3390/su15075614.

[4] C. V.H.-H. Chen and S. Ratanalert. "Development of an MEB Novice Chatbot to Improve Chemical Engineering Critical Thinking," 2025 ASEE Annual Conference & Exposition, June 2025, https://doi.org/10.18260/1-2--56281.

[5] F. Kamalov, D. S. Calonge, and I. Gurrib, "New Era of Artificial Intelligence in Education: Towards a Sustainable Multifaceted Revolution," *Sustainability*, vol. 15, no. 16, p. 12451, Jan. 2023, doi: https://doi.org/10.3390/su151612451.

[6] B. Zhao and D. D. Potter, “Comparison of Lecture-Based Learning vs Discussion-Based Learning in Undergraduate Medical Students,” *Journal of Surgical Education*, vol. 73, no. 2, pp. 250–257, Mar. 2016, doi: https://doi.org/10.1016/j.jsurg.2015.09.016.

[7] Y. Zheng and Y. Shi, “A Systematic Review of Dialogue-Based Teaching Intervention in Chinese Classrooms,” *ECNU Review of Education*, vol. 8, no. 2, pp. 581–606, May 2025, doi: https://doi.org/10.1177/20965311241310880.

[8] M. Kapur, “Examining productive failure, productive success, unproductive failure, and unproductive success in learning,” *Educational Psychologist*, vol. 51, no. 2, 2 Apr. 2016, pp. 289–299, https://doi.org/10.1080/00461520.2016.1155457.

Appendix: SSDM Misconception Inventory Questions for Chatbot Testing

#	Concept Area	Question Type	Question
1	Temperature Dependence	Concepts & Definitions	How does temperature affect the threshold voltage V_{th} of a MOSFET?
2	Temperature Dependence	Qualitative Prediction	Current of I_C flows through a simple common source BJT for given V_{BE}, the heat dissipated by the BJT in a unit time is given by $10 * K * I_c$, where K is a constant. The base voltage is fixed at 0.7, V_E is at ground. Is the above scenario in some sort of feedback? If yes, what feedback is it in?
3	Temperature Dependence	Quantitative Application	Silicon material is doped with donor atoms at temperature T_0, if the temperature is risen by 200K. Which quantity represents what N_D, n_i. The values are $5 * 10^{14} cm^{-3}$, $1.5 * 10^{10}$ and $7 * 10^{14}, 5 * 10^{25}$.
4	Drift	Concepts & Definitions	Drift is the primary mechanism by which charge carriers create current in an applied electric field within a semiconductor. If drift current density (J_{drift}) can be expressed in terms of carrier concentration, mobility, charge and electric field, such that $J_{drift} = \sigma * E$ (where E is the electric field, and sigma(σ) is a constant), what is the unit of sigma and what does it signify?
5	Drift	Qualitative Prediction	Two voltages V_1 *and* V_2 are applied across a silicon sample whose length is 1cm, the magnitude of V_1 is 1V while V_2 is 100V. If the total current density is due to diffusion and drift current. Predict which out of drift and diffusion will contribute majorly to the total current.
6	Drift	Quantitative Application	Find the value of current density for a compensated p-type material having the mobility of 500 $\frac{cm^2}{V*sec}$ and intensity of Electric field as 5 V. The carrier concentrations are $N_A = 10^{16} cm^{-3}$ and $N_D = 5 * 10^{15} cm^{-3}$.
7	Mobility	Concepts & Definitions	Electron mobility does not explicitly include a V_{sb} term. Does that mean mobility is independent of V_{sb}. How does mobility change with increased V_{sb}?
8	Mobility	Qualitative Prediction	Mobility ratio of electrons to holes is 3 when the total dopant concentration is $10^5 cm^{-3}$. What will be the ratio if the total dopant concentration is increased 3 times that of the original value. (Follow up) How does the ratio change if the total dopant concentration is increased by 10^{15} times from the original value?
9	Mobility	Quantitative Application	Two samples of Silicon (Si), one with high doping density, and one with low doping density, are subjected to an electric field (E) of $\frac{10^3 V}{cm}$, maintained at T=300K. Four mobility values are given: 470 $\frac{cm^2}{V*sec}$, 1400 $\frac{cm^2}{V*sec}$, 200 $\frac{cm^2}{V*sec}$, and 150 $\frac{cm^2}{V*sec}$. Two of these mobility values are for holes, and two are for electrons, at different doping densities. Calculate the electron drift velocity (v_n) and the hole drift velocity (v_p) in cm/s for these two samples, indicating for which doping densities and explain why the ordering of these drift velocities makes sense.
10	Diffusion of Carriers	Concepts & Definitions	A 4 cm long piece of n-type Germanium is uniformly doped with donors at $N_D = 10^{17} cm^{-3}$ at 300K. A linear voltage of $V(x) = 0.5(1 - \frac{x}{L})$ volts is applied across the length L, where x is the position and L=4 cm. What is the

			direction of electron flow, electric field E (magnitude and direction) inside the semiconductor? Express your answer in $\frac{V}{cm}$.

#	Concept Area	Question Type	Question
11	Diffusion of Carriers	Qualitative Prediction	Consider a uniformly doped n-type semiconductor bar under a low applied electric field $E < E_c$. The bar is illuminated by a small spot of light at x=0, creating a high concentration of excess minority carriers (holes) that then diffuse and recombine. In which direction will the hole diffusion current $J_{p,diffusion}$ flow from x=0, and why?
12	Diffusion of Carriers	Quantitative Application	An n-type silicon sample is uniformly illuminated with light which generates 10^{20} electron-hole pairs per cm^3 per second. The minority carrier lifetime in the sample is 1μs in steady state, the hole concentration in the sample is approximately 10^x, the integer number close to x is?
13	MOSFET	Concepts & Definitions	Why does the drain current equation predict that I_D reaches a peak and then falls like a parabola?
14	MOSFET	Qualitative Prediction	In modern technology nodes, it becomes difficult for the transistor to turn off even at low V_{GS} values. Do you think a negative V_{GS} value for an NMOS will completely push I_{DS} to 0? If not, why?
15	MOSFET	Quantitative Application	The gate-body capacitance of a 2-terminal (gate-oxide-body) and a 3-terminal (gate-oxide-source-body) MOSFET is measured. V_{th} =1V, $V_{g1} = 2.5\sin(10^6 t)$, and $V_{g2} = 2.5\sin(10t)$. The measured capacitance values are given. Identify which MOSFET and which V_g correspond to each capacitance value: 1.5 fF, 0.2 fF, 1.5 fF and 1.5 fF.