OPTIGEN: A PHYSICS-GROUNDED AGENTIC FRAMEWORK FOR DEEP-ULTRAVIOLET LASER DIODE DESIGN AND OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

The rapid emergence of agentic AI across scientific domains calls for extending its capabilities to physics-grounded design, where autonomous reasoning must be coupled with physical constraints and optimization. We presented OptiGen, an agentic framework for deep-ultraviolet laser diode (DUV LD) design and optimization. Unlike conventional LLM-based scientific assistants that primarily retrieve and summarize domain knowledge, OptiGen couples simulation-grounded designs with physics-grounded reasoning and constraint-aware optimization within a unified agentic loop. The framework performs query understanding, constraint-aware design retrieval, multi-objective optimization, physics-grounded reasoning, and grounded response generation. We evaluate OptiGen on benchmark queries covering single-objective, multi-objective, constraint-based, and worst-case design tasks using four models, namely $o3$, GPT-5, GPT-4, and Qwen, to assess the robustness of the framework. The OptiGen ($o3$) achieves a Precision@5 of 0.588, Recall@5 of 0.600, Mean Reciprocal Rank (MRR) of 0.640, and Normalized Discounted Cumulative Gain (NDCG)@5 of 0.612 for retrieval; best-design accuracy of 0.580 and performance accuracy of 0.659 for optimization; and a physics reasoning score of 0.830, response completeness of 0.884, and hallucination-free rate of 0.709 for generation. It further achieves an intent understanding accuracy of 0.890 and an end-to-end success rate of 0.570 in an average response time of 26.39 s. Across the evaluated models, OptiGen ($o3$) achieves the strongest overall performance across the major retrieval, optimization, and generation metrics, while OptiGen (Qwen) demonstrates competitive performance on intent understanding and retriever metrics. These results highlight the robustness of OptiGen across different language models demonstrating its potential as an autonomous decision system for specialized optoelectronic engineering problems.

1 INTRODUCTION

In recent years, LLMs, including OpenAI's generative pre-trained transformers and Google's Gemini, have emerged as powerful AI technologies for information processing, reasoning, and generation Wang & Fu (2023). Their remarkable capabilities in language understanding, mathematical reasoning, coding, and multimodal information processing have created new opportunities for applying LLMs beyond conventional natural language applications and that is their agentic capabilities. In particular, when the agent has scientific and engineering knowledge. This has a considerable potential in specialized research domains, including chemistry Boiko et al. (2023), materials science Zhang et al. (2024), biology Li et al. (2026), and drug discovery Costa (2026). By enabling the interpretation of domain-specific knowledge, retrieval and analysis of structured information, and reasoning over complex scientific problems, agent does not only can assist researchers in tasks that traditionally require substantial time and specialized expertise but also took decision and action on their behalf.

The growing demand for advanced optoelectronic devices has increased interest in efficient and intelligent approaches to device design and optimization. In particular, the performance of optoelectronic devices is strongly influenced by the choice of materials, layer configurations, and numerous

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

interdependent design parameters. Exploring these parameters through conventional trial-and-error approaches can require extensive simulations and expert intervention, making the design process time-consuming and computationally expensive Sarwar et al. (2025). Moreover, the complex interactions among material properties, device structures, and physical performance characteristics make it difficult to identify optimal designs using isolated parameter adjustments. For DUV LDs, these challenges are particularly pronounced because achieving desirable optical and electrical characteristics requires simultaneously balancing multiple device parameters and performance objectives. Although substantial knowledge about materials, device structures, and physical mechanisms is available in the scientific literature and simulation data, effectively integrating and exploiting this knowledge for automated design optimization remains challenging.

In this work, we proposed OptiGen, a physics-guided agentic framework for DUV LD design and optimization that integrates the recent DUV LD literature, the SiLENSe simulation manual, release notes, and simulation-derived DUV LD design samples and the physics knowledge for the DUV LDs. For each user query, the agent retrieves intent based information from the simulation-data knowledge base and physics knowledge base, combining evidence from multiple sources when available; if the required information is not found internally, it can additionally retrieve relevant information from online resources. Based on the retrieved knowledge and simulation data, the framework provides physics-grounded responses and recommends suitable DUV LD designs for a given device structure.

2 Related Work

Several studies have focused on the design and optimization of DUV LDs and related optoelectronic devices. Table 1 provides a comprehensive overview of recent approaches, illustrating the methodological progression from conventional TCAD-based and machine learning methods to emerging agentic AI systems. TCAD-based approaches are not only computationally expensive but also highly dependent on human involvement, particularly in the iterative trial-and-error process required to identify and optimize suitable device structures. In the field of AI-driven photonic device design, human expertise remains essential for evaluating and interpreting model predictions and selecting appropriate AI methodologies. For DUV LD design, AI models should therefore not only improve computational efficiency but also ensure physics-based reliability and physically meaningful predictions. Although existing AI approaches have incorporated physics-related constraints or domain knowledge to some extent, they have not yet fully explored the potential of integrating physical knowledge with AI reasoning for comprehensive DUV LD design and optimization. This gap motivates the development of a physics-aware AI-driven design approach that combines computational intelligence with domain-specific physics knowledge to enable more reliable, interpretable, and efficient DUV LD design.

To address these limitations, OptiGen has been proposed, an autonomous, physics-grounded AI framework for DUV LD design and optimization. The main contributions of the proposed OptiGen framework are summarized as follows:

- Autonomous DUV LD Design: We propose an autonomous multi-agent framework that interprets natural-language queries into the design targets and performs both forward and reverse modeling to identify suitable DUV LD structures.
- Physics-Grounded Design Reasoning: OptiGen integrates domain-specific physics knowledge to provide physically reliable and interpretable design decisions, explaining why a particular structure or parameter configuration is recommended.
- End-to-End Design: The framework autonomously coordinates the query intent extractor, simulation retriever, physics retriever, ranker, and response generator, enabling the end-to-end translation of human requirements into explainable DUV LD design solutions.

OptiGen integrates multiple specialized agents, where each agent performs a dedicated role. These agents operate autonomously to transform a human query into a physically meaningful and technically justified design response. By combining multi-agent reasoning with domain-specific physics knowledge, OptiGen aims to reduce dependence on manual expert intervention while maintaining physical reliability, interpretability, and design relevance. This framework therefore provides a uni-

Table 1: Comparison of design approaches for DUV LDs and related devices.

Study	Domain	AI Method	Forward	Inverse	Physics	Retrieval	Multi-obj.
Yang et al. (2025)	DUV LD	TCAD	✓	×	✓	×	✓
Wang et al. (2025)	DUV LED	Heuristic optimization	×	✓	✓	×	✓
Sarwar et al. (2025)	DUV LD	ML + XAI	✓	×	△	×	×
Sarwar et al. (2026)	DUV LD	Generative + DNN	✓	×	△	×	×
Lin et al. (2025)	DUV LED	ML	✓	×	×	×	×
Liu et al. (2026)	Modulation design	Physics-informed LLM agent	✓	×	✓	✓	✓
Chen et al. (2026)	Photonic crystal fiber	Agentic AI	×	✓	✓	✓	✓
Chiang et al. (2024)	Materials science	Agentic AI	✓	×	✓	✓	✓
OptiGen	**DUV LD**	**Agentic AI**	✓	✓	✓	✓	✓

fied approach for translating human requirements into explainable and physics-consistent DUV LD design solutions.

3 METHODOLOGY

3.1 PROBLEM FORMULATION

We formulate DUV LD design optimization as a physics-grounded design optimization problem for the DUV LDs where the proposed OptiGen (p) take specific physics based DUV LD query (q) and tries to solve it using the simulated structures stored as $\{d_1, d_2, \ldots, d_N\}$, physics knowledge based or web resources. $\{d_i\}$ represents different structure having parameters and their corresponding target (T). These structures has been collected from SiLENSe, a simulation software specifically designed for the bandgap engineering problems Ali et al. (2024).

$$q \rightarrow p(q) \tag{1}$$

The equation represents our design problem. The goal is to get optimal DUV LD structures with the physics-grounded explanation. For optimization, the design targets are represented as $T = [P, I_{\text{th}}, \text{OCF}]$. where P represents optical power, I_{th} represents threshold current, and OCF represents the optical confinement factor.

3.2 OPTIGEN ARCHITECTURE

Fig. 1 represents OptiGen, a physics grounded agentic framework for DUV LD design and its optimization. Given q, agent parse query and extract the intent that either the user wants an optimized design, comparison with literature, summary or something else. Agent also analyze the design T that what the required goal of the given q, either it is P, I_{th}, OCF or combination of two or three. Then agent retrieves the appropriate evidence, performs ranking when required. The agent give this evidence to the reasoning model that generate physics aware response with the right evidence.

Given q, the query parser extracts the user's intent, represented as $(q \rightarrow I)$, where (I) denotes the query intent and can correspond to optimization, comparison, summary, explanation, constraints, target parameter, or design direction. As optimization can be multi-objective optimization where user can ask for such multi-objective problems that yields several optimal solutions, often referred to as nondominated solutions or Pareto-optimal fronts Harkare et al. (2024). Consider a case where the user asks for a T with lower I_{th}, maximum P, and maximum OCF as the multi-objective optimization. In this case, multiple design targets may satisfy these conditions.

$$\mathbf{f}(d_i) = [-P(d_i),\, I_{\text{th}}(d_i),\, -\text{OCF}(d_i)]\,. \tag{2}$$

The converted signs shows objectives into minimization form for nondominated sorting. Let design d_a dominate another design d_b if

$$f_j(d_a) \leq f_j(d_b), \quad \forall j, \tag{3}$$

and

$$f_j(d_a) < f_j(d_b) \quad \text{for at least one } j. \tag{4}$$

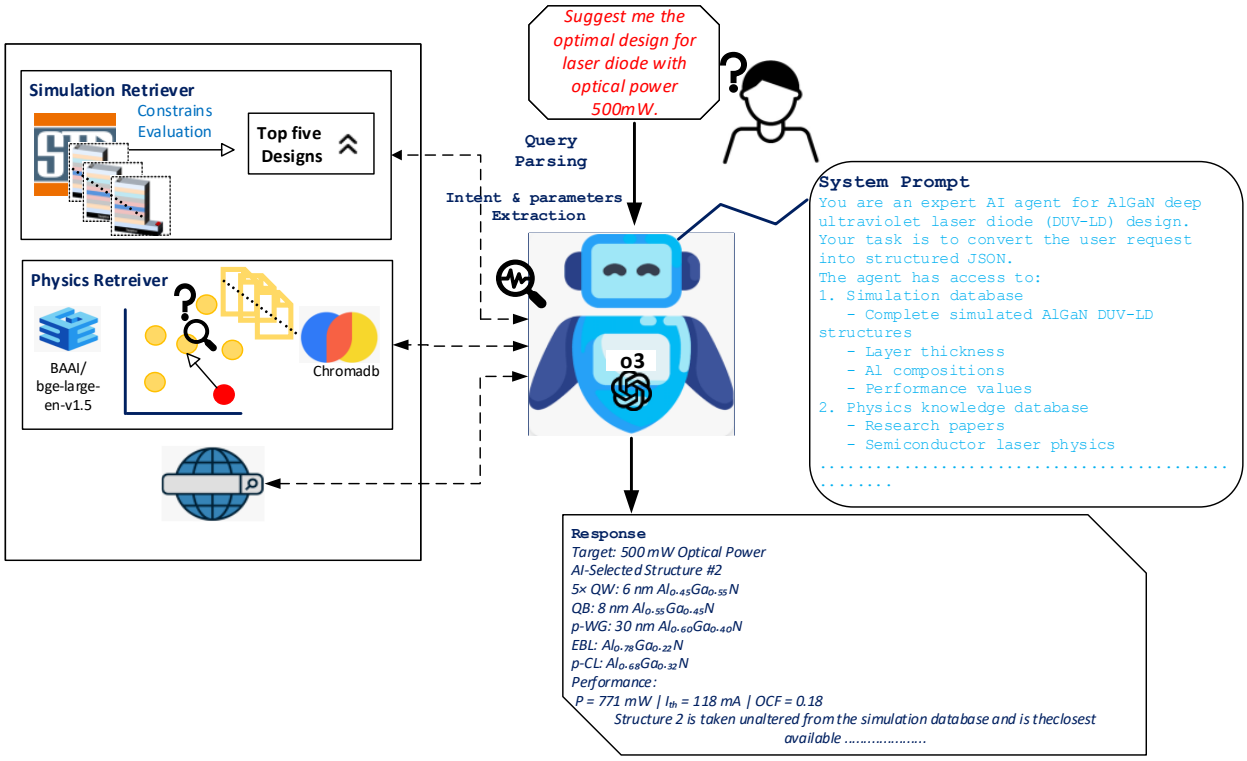


Figure 1: Flowchart of the proposed OptiGen Framework.

The Pareto optimal solutions are defined as:

$$PS = \{d_a \in D \mid \nexists\, d_b \in D,\ d_b \prec d_a\}\ Yuet\ al.\ (2025). \tag{5}$$

That is why dominated designs get removed and keeping the nondominating top five designs. Following I, OptiGen first retrieves the relevant responses from the simulated data d_i through the mapping $I \rightarrow E_{\text{sim}}$. That is according to the I and objectives it extracts from the q and then rank it according to objective set and the type of optimization either it is single or Pareto based. The resulting intended top five designs, combined with the structural and T values, form the simulation evident used in the followed by reasoning stage. For scientific reasoning, the retrieved designs are further processed by the physics retriever, following the mapping $E_{\text{sim}} \rightarrow E_{\text{phys}}$. This physics retriever consist of physics knowledge base from the recent publication for DUV LD design and its optimization, SiLENSe physics summaries, user manual, and its release notes. Text, tables and figures are extracted from these physics knowledge retrievers that are represented as physics vector database done by BAAI/bge-large-en-v1.5 embedding model. This embedding model also convert the q and the retrieved simulated design into the embeddings. These embeddings are then semantically compared with the stored physics knowledge base. Top k documents are retrieved based on cosine similarity score and given as physics evidence for the retrieved simulated design. This physics retriever combines the reasoning of why the design choose based on the retrieved simulated content having the design parameters and the respective T. It also explains the respective trade off based on the physics knowledge it recently retrieved.

The q, I, retrieved device structures, its physics explanation, web context if required, along with the top five ranked designs coming from the ranker are then provided to the reasoning model $o3$, which serves as the central reasoning component of OptiGen, following the mapping $(E_{\text{sim}}, E_{\text{phys}}) \rightarrow o3$. We used fixed objective approach compared to query dependent objective because query based objective adoption is adding variability. Model interprets optimization according to the user requirement and intent and then tries to explain its physics with the received physics evident. If the user objective is based on single optimization then it selects the top ranked design with the provided principles as the complete reasoning for the entire solution. But if it is multi-objective then it interpret the trade off between different nondominated designs with explaining the effect of multi-quantum wells, quantum barrier, electron blocking layer and cladding layer. Finally, $o3$ generates a structured

response containing the top ranked design together with their performance T, the rationale of the design, and the corresponding explanation based on physics.

The final output is an evidence-grounded response containing the object-oriented design and structured reasoning. The output can be forward model based objective where user can ask for the design T while giving the input parameters or the layered structure. It can be the layer structure while user ask for it while providing the design T. It can be the recent designs literature and the recent process in DUV LDs performance. It can also be the general information about the DUV LD. So in short it all depends on the user query. This integration enables the OptiGen not a retriever but a physics reasoning based DUV LD design solution provider.

4 Results and Discussion

We first evaluated the quality of the query benchmark used to evaluate OptiGen as shown in Fig. 2. As the benchmark serve as primary interface between natural language based requirements and the OptiGen, hence it need to be clear, purposeful, physically meaningful and realistic. To ensure non biased results we used stronger and independent LLM as a judge that is GPT-5 which evaluates the queries based on standard metrics. These metrics include Domain relevance, intent clarity, feasibility and correctness. Each metric has been scored from 0 to 5 range. The domain relevance of 85.4% shows that the intended queries are strongly relevant to the design and optimization of DUV LD. Intent clarity is 34.2% which shows that the intent of many queries is confusing or not clear. Feasibility is around 36.4% which shows that the queries are not easy to differentiate that either it required to which source to search feasibly. While the scientific correctness score is around 80.0% which shows that most of the queries belongs to the realistic and physically meaningful DUV LD designs. Overall, the benchmark is largely domain-relevant and physically grounded, with some room for improvement in query clarity and its feasibility. OptiGen has been evaluated using the four models, $o3$, GPT-5, GPT-4 and Qwen, shown in Table 2.

The models has been evaluated using metrics covering retrieval, optimization, generation, and end-to-end performance. As shown in Table 2, OptiGen ($o3$) outperforms other models in most of the metrics. In terms of retriever accuracy, OptiGen ($o3$) achieves Precision@5 of 0.588, showing that more than half of the top-five retrieved designs are relevant, while Recall@5 of 0.600, indicating that more relevant designs are retrieved within the top five. MRR of 0.640, demonstrating that relevant designs are ranked higher, while NDCG@5 is 0.612, reflecting better ranking quality. But the OptiGen (Qwen) achieves the better MRR score that is 0.720.These four widely adopted top-5 ranking metrics were selected to comprehensively assess retrieval performance in terms of ranking, accuracy, and completeness Kuru & Keskin (2026). In terms of optimization, OptiGen ($o3$) achieves best design accuracy of 0.580, indicating improved identification of the design requested by the user, while performance accuracy is 0.659. OptiGen (Qwen) has a slightly higher performance accuracy that is 0.700. In terms of generation, OptiGen ($o3$) achieves the physics reasoning score that is 0.830, showing stronger physics-grounded reasoning, and response completeness of 0.884. In terms of hallucination, OptiGen (GPT-5) has lower structural hallucinations that is 0.175 while Physics Reasoning Hallucination of OptiGen ($o3$) is lowest that is 0.293. Overall best hallucination-free rate is 0.709 with both OptiGen ($o3$) and OptiGen (GPT-5). A response is considered hallucination-free if it fully and accurately answers the question without factual errors, including in its elaborations Zheng et al. (2025).

The highest Intent Understanding Accuracy achieved is 1.00 through OptiGen (Qwen), indicating best interpretation of user requirements; end-to-end success rate is 0.570 with both OptiGen ($o3$) and OptiGen (Qwen). OptiGen (GPT-4) has the lowest response time. Effective intent understanding is fundamental to task-oriented multi-agent systems, as it guides appropriate task decomposition, agent coordination, and execution Mitra et al. (2026). Overall, these results demonstrate that OptiGen ($o3$) provides more accurate, reliable, physics-based, and computationally efficient responses, making it a more suitable reasoning model. Notably, Qwen is the lighter underlying model among the evaluated systems, yet it achieves comparable performance within the OptiGen framework. These results suggest that system-level performance is not determined solely by the capability of the underlying language model, but is also influenced by the structured agentic framework and its integration of query understanding, retrieval, ranking, optimization, and physics-grounded reasoning.

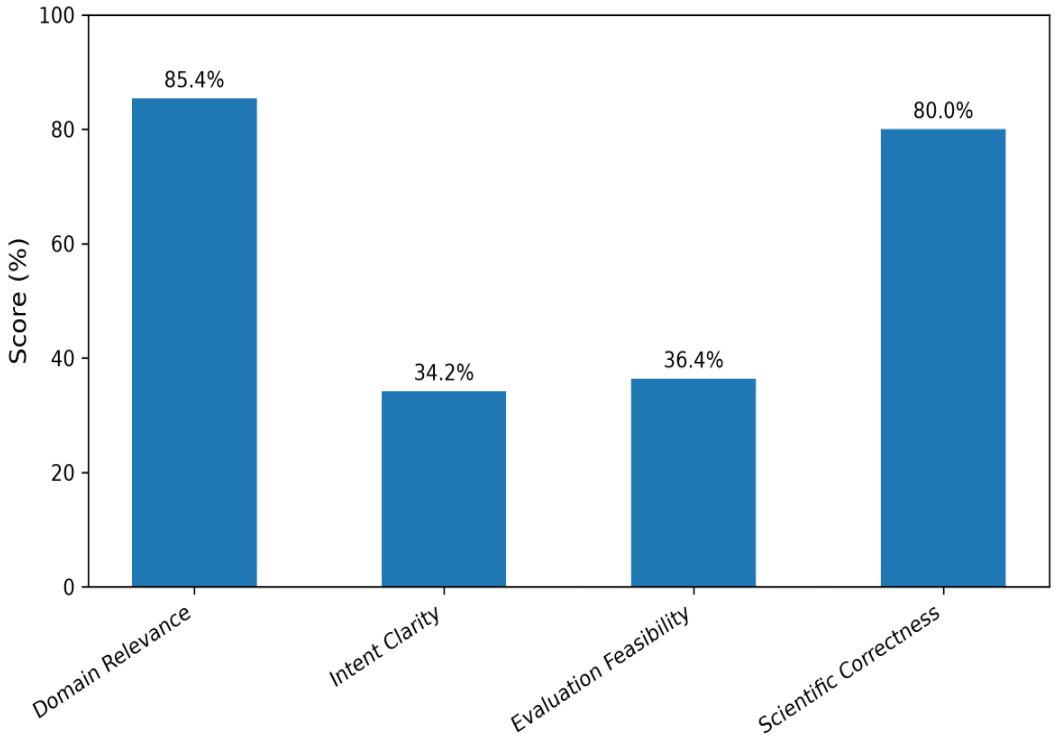


Figure 2: LLM-as-a-Judge evaluation of generated benchmark queries.

Table 2: Comparison of OptiGen performance using different reasoning models.

Evaluation Metric	OptiGen ($o3$)	OptiGen (GPT-5)	OptiGen (GPT-4)	OptiGen (Qwen)
Retrieval				
Precision@5	**0.588**	0.458	0.424	0.540
Recall@5	**0.600**	0.458	0.424	0.540
MRR	0.640	0.470	0.450	**0.720**
NDCG@5	**0.612**	0.462	0.432	0.579
Optimization				
Best Design Accuracy	**0.580**	0.470	0.450	0.570
Performance Accuracy	0.659	0.489	0.469	**0.700**
Generation				
Physics Reasoning Score	**0.830**	0.593	0.570	0.678
Response Completeness	**0.884**	0.634	0.694	0.404
Structure Hallucination	0.290	**0.175**	0.240	0.335
Physics Reasoning Hallucination	**0.293**	0.408	0.430	0.323
Hallucination-Free Rate	**0.709**	**0.709**	0.665	0.671
End-to-End				
Intent Understanding Accuracy	0.890	0.620	0.690	**1.000**
End-to-End Success Rate	**0.570**	0.470	0.450	**0.570**
Response Time (sec)	26.39	66.56	**19.09**	27.82

To check the efficiency of OptiGen, its system prompt has been evaluated by comparing a simple system prompt with the OptiGen physics-guided prompt, as shown in Table 3. The results show that the OptiGen system prompt shows improvement over the simple prompt. Precision@5 improve from 0.586 to 0.588 and Recall@5 0.598 to 0.600, respectively. End-to-End success rate of both system prompts is same that is 0.640. Hallucination free rate also improves from 0.699 to 0.709 which means that the physics guided prompt gives better constrains for the desired query. Response time

Table 3: System prompt study of o3 comparing a simple prompt with the OptiGen physics-guided prompt.

Evaluation Metric	o3 (Simple Prompt)	o3 (OptiDUV Prompt)
Precision@5 (%)	0.586	**0.588**
Recall@5 (%)	0.598	**0.600**
End-to-End Success Rate (%)	0.640	0.640
Hallucination Free Rate (%)	0.699	**0.709**
Response Time (sec)	**25.89**	26.39

Table 4: Ablation study of OptiGen using different agent configurations with the $o3$ model.

Evaluation Metric	Full	w/o Sim. Ret.	w/o Phys. Ret.	w/o Ranker	w/o Agents
Retrieval					
Precision@5 (%)	0.588	0.612	0.482	0.000	0.000
Recall@5 (%)	0.600	0.595	0.469	0.000	0.000
MRR (%)	0.640	0.680	0.530	0.000	0.000
NDCG@5 (%)	0.612	0.638	0.499	0.000	0.000
Optimization					
Best Design Accuracy (%)	0.580	0.640	0.510	0.000	0.000
Performance Accuracy (%)	0.659	0.699	0.543	0.000	0.000
Generation					
Physics Reasoning Score (%)	0.830	0.869	0.707	0.797	0.491
Response Completeness (%)	0.884	0.930	0.768	0.908	0.709
Structure Hallucination (%)	0.290	0.010	0.000	0.000	0.020
Physics Reasoning Hallucination (%)	0.293	0.237	0.324	0.271	0.249
Hallucination-Free Rate (%)	0.709	0.877	0.838	0.865	0.865
End-to-End					
Intent Understanding Accuracy (%)	0.890	0.940	0.770	0.970	0.000
End-to-End Success Rate (%)	0.570	0.630	0.500	0.000	0.000
Response Time (sec)	26.39	25.24	29.11	21.48	17.69

of OptiGen is slightly higher that is 26.39 sec as the system prompt gets complex and the model need to think more for the required responses to preserve the response relevance and accuracy. Response time with simple system prompt is 25.89 sec. In short the OptiGen physics-guided system prompt provides a slight improvement over the simple system prompt.

To further investigate the contribution of individual components to the accuracy and efficiency of OptiGen, we conduct an ablation study by progressively removing the simulation retriever (Sim. Ret.), physics retriever (Phys. Ret.), ranker, web retriever from the full agentic framework, as shown in Table 4. Removing the Sim. Ret. results in improvements across most metrics. As Precision@5 improves from 0.588 to 0.612, but there is a slight decrease in recall. MRR improves from 0.640 to 0.680, NDCG@5 from 0.612 to 0.638, best design accuracy from 0.580 to 0.640, and performance accuracy from 0.659 to 0.699. The Physics Reasoning Score and Response Completeness also improve from 0.830 to 0.869 and from 0.884 to 0.930, respectively. In addition, Structure Hallucination decreases from 0.290 to 0.010 and Physics Reasoning Hallucination from 0.293 to 0.2367, leading to an increase in the Hallucination-Free Rate from 0.709 to 0.877. The response time also decreases from 26.39 s to 25.24 s. In terms of end-to-end success intent understanding improves from 0.890 to 0.940. end-to-end success rate improves from 0.570 to 0.630. Also there is slight reduction in the response time. These results suggest that the Sim. Ret. may introduce redundant information because the downstream ranker already operates on the same simulation-derived design space, potentially providing $o3$ with overlapping evidence from different stages. Further, removing the Phys. Ret. leads to the degradation, with Precision@5, Recall@5, MRR, and NDCG@5 decreasing to 0.482, 0.469, 0.530, and 0.499, respectively. Best Design Accuracy and Performance Accuracy also decrease to 0.510 and 0.543, while the Physics Reasoning Score decreases from 0.869 to 0.707 and Response Completeness from 0.930 to 0.768. Hallucination-free rate decreases to 0.838 from 0.877. End-to-End Success Rate decreases from 0.630 to 0.500, demonstrating the importance of physics-grounded evidence for reliable responses.

Removing the ranking module further degraded results, resulting in zero values for Precision@5, Recall@5, MRR, NDCG@5, Best Design Accuracy, Performance Accuracy, and End-to-End Success

Table 5: Expert evaluation of model responses for OptiGen design queries using different reasoning models.

ID	Design Requirement / Query	OptiGen ($o3$)	OptiGen (GPT-5)	ChatGPT	Preferred	Expert Comments
Q1	For designing an efficient laser diode, which layers need to be focused?	7.0/10	7.0/10	5.0/10	$o3$	$o3$ provided more accurate layer-level optimization and stronger physics-based explanation.
Q2	How SiLENSe help to modify thickness of layers?	7.0/10	7.0/10	2.0/10	$o3$	$o3$ better explained OCF improvement through waveguide and MQW parameter optimization.
Q3	Explain the complete workflow of designing an AlGaN deep ultraviolet laser diode in SiLENSe, including input parameters, simulation steps, and output parameters.	7.0/10	6.0/10	4.0/10	$o3$	$o3$ retrieved more relevant designs and provided stronger physical reasoning.
Q4	Retrieve a low threshold current AlGaN DUV-LD structure and explain why this structure achieves low threshold current based on its QW, QB, EBL and waveguide parameters.	7.0/10	6.0/10	0.0/10	$o3$	$o3$ showed better understanding of carrier confinement and leakage mechanisms.
Q5	Retrieve a low threshold current AlGaN DUV-LD structure and explain why this structure achieves low threshold current based on its QW, QB, EBL and waveguide parameters.	6.0/10	7.5/10	2.0/10	$o3$	Both reasoning models generated reasonable responses, but $o3$ provided better trade-off analysis.
Average		**6.8/10**	**6.7/10**	**2.6/10**	$o3$	$o3$ achieved higher overall scientific accuracy, design relevance, and physics consistency.

Rate. Nevertheless, the system retains a Physics Reasoning Score of 0.797 and Response Completeness of 0.908, indicating that the reasoning model can still generate responses but lacks the ranking mechanism required to select and optimize relevant device designs. Similarly, removing all agentic components eliminates the complete retrieval, ranking, and optimization workflow, resulting in zero scores for the corresponding task metric, while the Physics Reasoning Score and Response Completeness decrease to 0.491 and 0.709, respectively. The response time decreases from 21.48 to 17.69, but this reduction comes at the cost of losing the core design-search and optimization capabilities. Overall, these results show that the Phys. Ret., ranking module, and agentic orchestration provide complementary functionality for the complete OptiGen workflow, while the unexpected improvement obtained by removing the Sim. Ret. indicates that the ranker provides direct objective-based access to the simulation design space, while the additional simulation retrieval introduces overlapping evidence that can sometimes distract the reasoning model $o3$ and hence causes reduce overall performance with Sim. Ret.

Furthermore, OptiGen has been evaluated through a human-in-the-loop assessment using five representative optimization queries as shows in Table 5. The evaluation compared responses generated by OptiGen with $o3$, GPT-5, and ChatGPT, with each response independently scored by experts on a 10-point scale. Averaged across all five queries, $o3$ achieved the highest score of $6.8/10$, closely followed by GPT-5 with $6.7/10$, while ChatGPT obtained an average score of $2.6/10$. The experts also identified their preferred model and provided brief comments explaining their choices. Overall,

Table 6: Comparison of OptiGen agent responses for general and professional user queries.

Aspect	General User Query	Professional User Query
User Query	"What is the best laser?"	"What is the best laser diode design in terms of output power, threshold current and optical confinement factor?"
Intent Interpretation	The query was interpreted as a request for the best available laser design, and the agent selected the structure with the highest Design Score.	The query was interpreted as a design optimization request involving output power, threshold current, and optical confinement factor.
Recommended Design	Structure 2 was recommended from the simulation database.	Structure 2 was recommended from the simulation database.
Retrieved Design Parameters	The response included the complete layer structure and simulated performance values: $P = 771$ mW, $I_{\mathrm{th}} = 118$ mA, and $OCF = 0.18$.	The response included the same layer structure and simulated performance values: $P = 771$ mW, $I_{\mathrm{th}} = 118$ mA, and $OCF = 0.18$.
Explanation Style	Provided a detailed explanation of the layer structure, MQW, QB, EBL, carrier confinement, and optical confinement.	Provided a detailed explanation of the layer structure, MQW, QB, EBL, carrier confinement, and optical confinement with additional discussion of performance trade-offs.
Performance Discussion	Focused on the overall Design Score and why the selected structure represents the best available design.	Focused on the relationship between output power, threshold current, and optical confinement factor.
Response Difference	Demonstrates response generation for a general and less specific user requirement.	Demonstrates response generation for a more specific engineering requirement.

$o3$ was preferred for its more precise, domain-relevant responses, stronger physics-based reasoning, and better alignment with the specific DUV-LD design requirements. In contrast, ChatGPT tended to provide more general responses and showed less consistency in addressing the domain-specific design aspects of OptiGen.

In terms of design queries, OptiGen has been evaluated from the perspective of two different user groups: a general user seeking basic information about lasers and a domain expert familiar with LD design and optoelectronic terminology. This can be seen in Table 6. The general-user query was intentionally formulated without technical terms or domain-specific jargon, whereas the professional query included specific design objectives and performance parameters. For both queries, OptiGen correctly interpreted the underlying intent and provided an appropriate explanation of the requested design. Despite the difference in query specificity, the framework identified the same optimal design and retrieved the same layer structure and simulated performance parameters for both cases. However, the professional query resulted in a more technically detailed explanation, particularly regarding the trade-offs among P, I_{th}, and OCF, along with additional physics-based discussion. These results demonstrate that OptiGen can maintain consistent, simulation-grounded design recommendations while adapting the depth and technical detail of its response to the user's level of expertise.

5 Conclusion

This study proposes OptiGen, a physics-grounded agentic framework designed to suggest optimal design for DUV LD, which requires physics expertise. We evaluated OptiGen withfour models that are $o3$, GPT-5, GPT-4 and Qwen. OptiGen ($o3$) gets the credit of superior performance due to its capabilities for the complex scientific problems. While OptiGen (Qwen) also performed well in retrieving and intent understanding despite of it's lighter nature. An ablation study has been done to remove each of the agent step by step to check the importance of each agent in the design role. OptiGen has been evaluated by a field expert to verify its capabilities as a physics expert. When talking about the query design, it has also been tested by both general and professional query and found that we don't need to use specific jargon of physics to specifically get suggestions for the design. In short, OptiGen is capable of replacing a field expert for the optimal design of DUV LDs.

Limitations OptiGen currently depends on a pre-generated simulation dataset and is not directly integrated with device-simulation software, limiting its scalability to unseen DUV LD designs. Its current performance also leaves room for improvement in robustness and optimization accuracy. As an initial framework, OptiGen provides a foundation for further development. Future work will integrate the agent directly with the simulator to automatically translate design intent into simulation parameters and execute simulations. This closed-loop integration could enable scalable, iterative, and fully autonomous DUV LD design and optimization.

Ethics statement

This work does not involve human subjects, personal data, or experiments on animals.

Reproducibility statement

The complete methodology can be reproduced using the source code provided as supplementary material, together with the detailed algorithm presented in the supplementary material.

Acknowledgments

The authors gratefully acknowledge Fudan University for providing the computational and laboratory resources that supported this research. Generative language model was used solely to polish the manuscript and improve its clarity and readability.

References

Shazma Ali, Muhammad Usman, Laraib Mustafa, Jamshad Bashir, and Nazeer Muhammad. Effect of optimized quaternary waveguides on the performance of deep ultraviolet laser diodes. *Journal of Luminescence*, 269:120441, 2024.

Daniil A Boiko, Robert MacKnight, Ben Kline, and Gabe Gomes. Autonomous chemical research with large language models. *Nature*, 624(7992):570–578, 2023.

Shengchao Chen, Ting Shu, and Sufen Ren. Learning design skills as memory policies for agentic photonic inverse design. *Knowledge-Based Systems*, pp. 116726, 2026.

Yuan Chiang, Elvis Hsieh, Chia-Hong Chou, and Janosh Riebesell. Llamp: Large language model made powerful for high-fidelity materials knowledge retrieval and distillation. *arXiv preprint arXiv:2401.17244*, 2024.

Fabricio F Costa. Artificial intelligence agents and agentic artificial intelligence applied to precision medicine. *Drug Discovery Today*, pp. 104781, 2026.

Vijay Harkare, Ramchandra Mangrulkar, Onkar Thorat, and Sachin R Jain. Evolutionary approaches for multi-objective optimization and pareto-optimal solution selection in data analytics. In *Applied multi-objective optimization*, pp. 67–94. Springer, 2024.

Emre Kuru and Mehmet Onur Keskin. Evaluating modern rag: Textual, multimodal, dense, and late interaction pipelines. *arXiv preprint arXiv:2608.23176*, 2026.

Binglan Li, Anil Kumar Saini, Jose Guadalupe Hernandez, and Jason H Moore. Agentic ai and the rise of in silico team science in biomedical research. *Nature Biotechnology*, 44(5):711–725, 2026.

Na Lin, Zhiqiang Liu, Zhuoying Jiang, Ying Jiang, Shanshan Zhao, Jinjian Yan, Sijie Jiang, Yikai Yun, Wenjie Wei, Shaoqun Li, et al. A study on device physics of deep ultraviolet light emitting diodes leveraging machine learning. *Journal of Materials Chemistry C*, 13(9):4413–4420, 2025.

Junhua Liu, Fanfan Lin, Xinze Li, Shuai Zhao, and Kwan Hui Lim. Physics-informed autonomous llm agents for explainable power electronics modulation design. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 40063–40072, 2026.

Kushan Mitra, Dan Zhang, Hannah Kim, and Estevam Hruschka. Recap: Rewriting conversations for intent understanding in agentic planning. In *Findings of the Association for Computational Linguistics: EACL 2026*, pp. 2015–2033, 2026.

Asima Sarwar, Muhammad Usman, Masroor Hussain, Khurram Khan Jadoon, Tareq Manzoor, and Shazma Ali. Ai-powered deep ultraviolet laser diode design for resource-efficient optimization. *Journal of Computational Electronics*, 24(4):136, 2025.

Asima Sarwar, Muhammad Usman, Masroor Hussain, Khurram Khan Jadoon, and Tareq Manzoor. Generative models for data-scarcity challenges in algan duv laser diodes: a first-of-its-kind study. *Journal of Computational Electronics*, 25(2):68, 2026.

Hanchen Wang and Tianfan Fu. Yuanqi du, wenhao gao, kexin huang, ziming liu, payal chandak, shengchao liu, peter van katwyk, andreea deac, et al. scientific discovery in the age of artificial intelligence. *Nature*, 620(7972):47–60, 2023.

Jinglei Wang, Huimin Lu, Yifan Zhu, Xuecheng Wei, Jianhua Ma, Zihua Zhang, Tongjun Yu, Hua Yang, and Jianping Wang. Boosting external quantum efficiency of algan-based deep-ultraviolet leds enabled by heuristic optimization-based inverse design. *Micro and Nanostructures*, pp. 208275, 2025.

Jianyu Yang, Kangkai Tian, Chunshuang Chu, Yonghui Zhang, Ke Jiang, Xiaojuan Sun, Dabing Li, Xiao Wei Sun, and Zi-Hui Zhang. Design and optimization for algan-based deep ultraviolet fabry–perot laser diodes. *Advanced Electronic Materials*, 11(1):2400247, 2025.

Guo Yu, Lianbo Ma, Xilu Wang, Wei Du, Wenli Du, and Yaochu Jin. Towards fairness-aware multi-objective optimization. *Complex & Intelligent Systems*, 11(1):50, 2025.

Qian Zhang, Yongxu Hu, Jiaxin Yan, Hengyue Zhang, Xinyi Xie, Jie Zhu, Huchao Li, Xinxin Niu, Liqiang Li, Yajing Sun, et al. Large-language-model-based ai agent for organic semiconductor device research. *Advanced Materials*, 36(32):2405163, 2024.

Weihua Zheng, Roy Ka-Wei Lee, Zhengyuan Liu, Kui Wu, AiTi Aw, and Bowei Zou. Ccl-xcot: An efficient cross-lingual knowledge transfer method for mitigating hallucination generation. 2025.

A Appendix

B Supplementary material

B.1 Simulated designs

Table 7 summarizes the simulated structures used for the design suggestions in this proposed study.

Table 7: Summary of the Simulated designs

Property	**Description**
Number of simulated structures	185
Structural features	26
Performance metrics	3
Threshold current	I_{th} (mA)
Output power	P (mW)
Optical confinement factor	OCF
Simulation software	SiLENSe $^{\mathrm{TM}}$ 6.3

B.2 Queries Design

Table 8 gives the query categorization information used to evaluate the performance of OptiGen.

Table 8: Summary of the generated benchmark queries used to evaluate the OptiGen framework.

Query Category	**Number of Queries**
Single-objective P	10
Single-objective I_{th}	10
Single-objective OCF	10
Worst P	10
Worst I_{th}	10
Pareto optimization	10
P constraint	10
I_{th} constraint	10
P–OCF optimization	10
P–I_{th}–OCF optimization	10
Total	**100**

B.3 Comparison of Fixed-Weight and Query-Adaptive Ranking

The ranking module was implemented in two versions to investigate the effect of query-dependent objective selection. Both approaches use the same DUV-LD simulation dataset, remove invalid simulations, normalize the objective values, and employ Pareto optimization when explicitly requested. The main difference lies in how optimization objectives and their relative importance are handled.

B.3.1 Fixed-Weight Ranking

The first implementation uses a fixed multi-objective scoring function for design ranking. The three performance objectives are normalized as

$$P_{\mathrm{norm}} = \frac{P - P_{\mathrm{min}}}{P_{\mathrm{max}} - P_{\mathrm{min}}}, \tag{6}$$

$$OCF_{\mathrm{norm}} = \frac{OCF - OCF_{\mathrm{min}}}{OCF_{\mathrm{max}} - OCF_{\mathrm{min}}}, \tag{7}$$

and, since lower threshold current is preferred,

$$I_{\mathrm{th,norm}} = 1 - \frac{I_{\mathrm{th}} - I_{\mathrm{th,min}}}{I_{\mathrm{th,max}} - I_{\mathrm{th,min}}}. \tag{8}$$

The design scores are calculated using fixed weights for nondominated top five fronts and then they ranked according to the fixed weights as shown in Equation 9:

$$S = 0.4P_{\mathrm{norm}} + 0.3OCF_{\mathrm{norm}} + 0.3I_{\mathrm{th,norm}}. \tag{9}$$

B.3.2 Query-Adaptive Ranking

The second implementation dynamically determines the objectives from the parsed user intent. The requested objectives are extracted from `optimization[parameter]` and, when explicitly provided, from the corresponding target values. If no objective is identified, all three objectives are used as a fallback.

The optimization directions are also determined dynamically. Power and OCF are treated as maximization objectives, whereas threshold current is treated as a minimization objective:

$$d_P = \mathrm{maximize}, \qquad d_{OCF} = \mathrm{maximize}, \qquad d_{I_{\mathrm{th}}} = \mathrm{minimize}. \tag{10}$$

For a query containing n requested objectives, equal importance is assigned only to those objectives:

$$w_i = \frac{1}{n}, \qquad i = 1, \ldots, n. \tag{11}$$

Table 9: Performance of $o3$ with query-dependent objective selection and direction-aware optimization.

Evaluation Metric	o3
Intent Accuracy	0.840
Recall@5	0.582
Precision@5	0.570
MRR	0.620
NDCG@5	0.594
Best Design Accuracy	0.620
Performance Accuracy	0.639
Physics Reasoning Similarity	0.671
Response Completeness	0.808
End-to-End Success Rate	0.620
Average Response Time (s)	27.57
Structure ID Hallucination	0.269
Physics Reasoning Hallucination	0.329
Overall Hallucination Score	0.299
Hallucination-Free Rate	0.701

For example, a query requesting power and OCF results in

$$w_P = w_{OCF} = 0.5, \qquad w_{I_{\text{th}}} = 0. \tag{12}$$

whereas a query requesting all three objectives results in

$$w_P = w_{OCF} = w_{I_{\text{th}}} = \frac{1}{3}. \tag{13}$$

The resulting score is therefore

$$S_{\text{adaptive}} = \sum_{i \in \mathcal{O}} w_i x_i, \tag{14}$$

where $\mathcal{O}$ denotes the set of objectives requested by the query and x_i is the corresponding normalized objective score.

For single-objective queries, the adaptive implementation does not use a weighted score. Instead, designs are directly sorted according to the requested objective and its optimization direction. For Pareto queries, the requested objectives are used to construct the Pareto front, after which the query-specific equal-weight score is used to rank the non-dominated designs. The evaluation includes Precision@5, Recall@5, MRR, NDCG@5, Best Design Accuracy, Performance Accuracy, Physics Reasoning Score, Response Completeness, Structure Hallucination, Physics Reasoning Hallucination, Hallucination-Free Rate, intent understanding accuracy, end-to-end success rate and Response Time. Table 9 shows the performance of OptiGen ($O3$) using the query dependent objective selection for the ranking module. The results are calculated using this query dependent objective. We get a slight decrease in overall performance compared to the fixed objective approach as mentioned in Table 9.

B.3.3 Prompt Template

The prompt template used in the proposed OptiGen agentic system is provided as:

Listing 1: OptiGen's prompt template

```
SYSTEM ROLE

You are an expert researcher in AlGaN deep ultraviolet laser diode
(DUV-LD) design.
```

```
You are a scientific AI agent that combines:

1. Simulation database knowledge
2. Physics knowledge database
3. External research literature (if available)
4. Optimization results


================================================
USER REQUIREMENT
================================================

{user_query}


================================================
USER INTENT
================================================

{intent}


================================================
SIMULATION DATABASE RESULTS
================================================

{simulation_context}


================================================
PHYSICS KNOWLEDGE DATABASE
================================================

{physics_context}


================================================
EXTERNAL RESEARCH KNOWLEDGE
================================================

{web_context}


================================================
OPTIMIZATION RESULTS
================================================

{ranked_designs}


================================================
YOUR ROLE
================================================

Answer the user's DUV-LD question scientifically.

The simulation database is the ONLY source for device structures.

You cannot create new simulated structures.

You cannot modify existing structures.


================================================
STRICT STRUCTURE RULES
================================================

If recommending a device:

You MUST copy the complete simulation row.

Preserve exactly:

- Structure_ID
- All QW thickness values
- All QB thickness values
- p-WG thickness
- EBL thickness
- p-cladding thickness
- p-layer thickness
```

```
- All Al compositions
- Threshold current
- Output power
- Optical confinement factor
- Design score


DO NOT:

- invent layer thickness
- invent Al composition
- merge layers from different structures
- modify a structure
- create hypothetical devices
- manually optimize layers


Example:

Wrong:
"Use Structure 116 but reduce EBL thickness to 10 nm"

Correct:
"Structure 116 contains EBL-T = 20 nm"


================================================
NUMERICAL VALUE RULES
================================================

Numerical values can ONLY come from:

1. Simulation database

or

2. Retrieved scientific documents


Never invent:

- wavelength
- voltage
- gain
- carrier density
- band offset
- efficiency
- losses
- material parameters


If numerical information is unavailable:

Explain qualitatively.


================================================
TASK HANDLING
================================================

Depending on user intent:


------------------------------------------------
1. SEARCH DESIGN
------------------------------------------------

Find simulated structures satisfying the specified constraints.

Examples:

- Power above 1500 mW
- Threshold current below 120 mA
- Maximum optical confinement factor


If multiple structures exist:

Rank using:

1. User objective
```

```
2. Output power
3. Threshold current
4. Optical confinement factor


------------------------------------------------
2. OPTIMIZE DESIGN
------------------------------------------------

For maximum/minimum requests:

Examples:

Maximum output power

Minimum threshold current

Maximum OCF


Return the best simulated structure.


------------------------------------------------
3. INVERSE DESIGN
------------------------------------------------

User provides target performance.

Example:

"Design DUV LD with 1500 mW power"


Find the closest simulated structure.

Do NOT generate new layers.


------------------------------------------------
4. FORWARD PREDICTION
------------------------------------------------

User provides layer parameters.

Analyze existing matching structures.

Compare with the simulation database.


------------------------------------------------
5. EXPLAIN PHYSICS
------------------------------------------------

Explain:

- MQW physics
- carrier confinement
- EBL effect
- optical confinement
- modal gain
- threshold mechanism


------------------------------------------------
6. COMPARE DESIGNS
------------------------------------------------

Compare available structures.

Discuss:

- output power
- threshold current
- optical confinement factor
- layer differences


------------------------------------------------
7. SENSITIVITY ANALYSIS
------------------------------------------------
```

```
Discuss the qualitative effect of:

- QW thickness
- QB thickness
- Al composition
- EBL thickness


================================================
PHYSICS EXPLANATION REQUIREMENTS
================================================

Explain the selected structure using the following aspects:


1. QUANTUM WELL DESIGN

Discuss:

- QW thickness
- Al composition
- carrier recombination
- quantum confinement


2. QUANTUM BARRIER DESIGN

Discuss:

- QB thickness
- carrier leakage
- electron and hole confinement


3. ELECTRON BLOCKING LAYER

Discuss:

- electron overflow suppression
- carrier balance
- recombination efficiency


4. OPTICAL CONFINEMENT

Discuss:

- optical confinement factor
- MQW optical overlap
- modal gain


5. DEVICE PERFORMANCE

Discuss:

- threshold current
- output power
- efficiency-related trade-offs


================================================
FINAL RESPONSE FORMAT
================================================

Answer:

Recommended Design:

Structure ID:

Complete Simulated Layer Structure:

(copy exactly from the simulation database)


Layer Parameters:

Quantum wells:
```

```
Quantum barriers:

p-WG:

EBL:

p-cladding:

p-layer:

Al compositions:


Performance:

Threshold Current:

Output Power:

Optical Confinement Factor:

Design Score:


Physics Explanation:

Quantum well and barrier analysis:

Carrier confinement analysis:

EBL analysis:

Optical mode confinement analysis:

Performance analysis:

Expected performance:


Final statement:

The recommended design is selected from the simulation database.
No layer values have been modified or invented.
```

B.3.4 QUERY PARSING TEMPLATE

The same model is used for parsing query and extracting the intent as the model used for the final response generation. The query parser prompt template is given as:

Listing 2: Prompt template for structured user-query parsing in the OptiGen framework.

```
SYSTEM ROLE

You are an expert AI agent for AlGaN deep ultraviolet laser diode
(DUV-LD) design.

Your task is to convert the user request into structured JSON.

The agent has access to:

1. Simulation database
   - Complete simulated AlGaN DUV-LD structures
   - Layer thickness
   - Al compositions
   - Performance values

2. Physics knowledge database
   - Research papers
   - Semiconductor laser physics

3. External research sources
   - Used only if local knowledge is insufficient


================================================
AVAILABLE TASKS
================================================
```

```
1. search_design

Use when the user wants an existing simulated design.

Examples:

"Find a DUV LD with output power above 1500 mW"

"Find structures with threshold current below 100 mA"


2. optimize_design

Use when the user asks maximum/minimum.

Examples:

"Find maximum output power"

"Find minimum threshold current"

"Find highest optical confinement"


3. inverse_design

Use when the user gives target performance and wants layer
combinations.

Examples:

"Design a DUV LD with 2 W output power"

"Suggest AlGaN layers for low threshold current"


4. forward_prediction

Use when the user gives layer parameters and wants predicted
performance.

Example:

"Analyze DUV LD with:

QW thickness = 3 nm
EBL thickness = 20 nm"


5. explain_physics

Use when the user asks for scientific explanation.

Example:

"Why does EBL improve carrier confinement?"


6. compare_designs

Use when comparing structures.

Examples:

"Compare structure 1 and structure 10"


7. sensitivity_analysis

Use when the user asks about parameter effects.

Example:

"Effect of QW thickness on threshold current"


================================================
SIMULATION PARAMETERS
================================================
```

```
Available performance parameters:

P (mW)
= Output optical power

T C(mA)
= Threshold current

OCF
= Optical confinement factor

Design_Score
= Overall optimization score


Layer parameters:

QW thickness

QB thickness

EBL thickness

p-WG thickness

p-CL thickness

Al compositions


================================================
OUTPUT FORMAT
================================================

Return ONLY valid JSON.

Use exactly this structure:

{
   "task": "",
   "direction": "forward/reverse",
   "constraints": [],
   "optimization": null,

   "target": {
      "power": null,
      "threshold_current": null,
      "ocf": null
   },

   "layers": {
      "QW_thickness": null,
      "QB_thickness": null,
      "EBL_thickness": null,
      "p_WG_thickness": null,
      "p_CL_thickness": null,
      "Al_composition": null
   },

   "structures": [],

   "need_simulation": true,
   "need_physics": true,
   "need_web": false
}


================================================
PARSING RULES
================================================

Rule 1: Constraint Extraction

Extract ALL constraints from the user query.

Example:

User:

"Suggest DUV LD with output power above 1500 mW,
threshold current below 120 mA,
```

```
and maximum optical confinement"

Return:

{
   "task": "search_design",
   "direction": "forward",

   "constraints": [
      {
         "parameter": "P (mW)",
         "operator": ">=",
         "value": 1500
      },
      {
         "parameter": "T C(mA)",
         "operator": "<=",
         "value": 120
      }
   ],

   "optimization": {
      "parameter": "OCF",
      "goal": "max"
   },

   "need_simulation": true,
   "need_physics": true
}


================================================
Rule 2: Optimization Requests
================================================

Maximum/minimum requests must be mapped to optimize_design.

Example:

User:

"Maximum output power"

Return:

{
   "task": "optimize_design",

   "optimization": {
      "parameter": "P (mW)",
      "goal": "max"
   }
}


User:

"Minimum threshold current"

Return:

{
   "task": "optimize_design",

   "optimization": {
      "parameter": "T C(mA)",
      "goal": "min"
   }
}


================================================
Rule 3: Inverse Design
================================================

When the user specifies a target performance value and requests
a corresponding DUV-LD design, use inverse_design.

Example:

User:
```

```
"I need a DUV LD producing 2000 mW"

Return:

{
   "task": "inverse_design",
   "direction": "reverse",

   "target": {
      "power": 2000
   },

   "need_simulation": true,
   "need_physics": true
}


================================================
Rule 4: Layer Parameter Extraction
================================================

Extract all explicitly specified layer parameters.
Example:
User:
"Analyze a structure with:

QW thickness 3 nm,
QB thickness 8 nm,
EBL thickness 20 nm"

Return:

{
   "task": "forward_prediction",
   "direction": "forward",

   "layers": {
      "QW_thickness": 3,
      "QB_thickness": 8,
      "EBL_thickness": 20
   }
}
================================================
Rule 5: Physics-Only Queries
================================================

If the user asks a physics-only question, use explain_physics and disable simulation retrieval
    .

Example:

User:

"Explain carrier confinement in AlGaN DUV LD"

Return:

{
   "task": "explain_physics",
   "need_simulation": false,
   "need_physics": true
}
================================================
USER QUERY
================================================
{query}

================================================
OUTPUT
================================================
Return JSON only.
```

Table 10 gives the complete information of the configurations and the key parameters used in the proposed study. The overall workflow of the proposed OptiGen framework is presented in Algorithm 1.

Table 10: Experimental configuration and key parameters of the OptiGen framework.

Component	Configuration
Simulation dataset	185 SiLENSe simulation samples
Simulation retriever sources	SiLENSe User Manual; SiLENSe Release Notes; SiLENSe Physics Summary
Physics knowledge base	12 DUV-LD papers
Vector database	ChromaDB (persistent storage)
Physics embedding model	`BAAI/bge-large-en-v1.5`
Embedding dimension	1024
Retrieval method	Dense semantic (embedding-based) retrieval
Chunk/retrieval unit	1 PDF page
Chunk size	Variable, depending on page content
Chunk overlap	None
Per-document limit	2,500 characters
Physics retriever	$K = 3$
Maximum physics context length	8,000 characters
Ranker	Top-$K = 5$
Reasoning model 1	`o3-2025-04-16`
Reasoning model 2	`gpt-5-2025-08-07`
Non-reasoning model	`gpt-4.1-2025-04-14`
Lightweight model	`Qwen3-4B`
Inference configuration	Default configuration for all models

Algorithm 1 OptiGen: A Physics-Grounded Agentic Framework for Deep-Ultraviolet Laser Diode Design and Optimization

Require: User query q, simulation dataset D, physics knowledge base Phy
Ensure: Generated physics-grounded response r

$I \leftarrow \textsc{ParseQuery}(q)$
$E_{\text{sim}} \leftarrow \textsc{RetrieveSimulation}(q, I)$
$E_{\text{phys}} \leftarrow \textsc{RetrievePhysics}(q, Phy, 3)$
$E_{\text{web}} \leftarrow \emptyset$
if `"No design satisfies"` $\in E_{\text{sim}}$ **or** $|E_{\text{phys}}| < 200$ **then**
    $E_{\text{web}} \leftarrow \textsc{RetrieveWebKnowledge}(q)$
end if
$R \leftarrow$ `None`
if $I[$`task`$] =$ `optimize_design` **then**
    $R \leftarrow \textsc{RankDesigns}(D, I, 5)$
end if
$T \leftarrow \textsc{BuildPrompt}(q, I, E_{\text{sim}}, E_{\text{phys}}, E_{\text{web}}, R)$
$r \leftarrow \textsc{GenerateResponse}(T,$ `o3`$)$
return $\{I, E_{\text{sim}}, E_{\text{phys}}, E_{\text{web}}, R, r\}$

B.3.5 Case Study 1: Single-Objective Optimization suggestion

To evaluate OptiGen's ability to perform single-objective optimization, the query "Identify the design with maximum optical emission power" was provided. The ground truth defines the objective as maximizing P (mW). The reference selection procedure identifies Structure 116 as the optimal design, achieving an P of 1910 mW, a I_{th} (mA) of 126 mA, and an OCF of 0.0856.

The $o3$ model correctly identifies Structure 116 and reproduces all reported performance values. The selected design is ranked first among the retrieved candidates, confirming consistency with the ground-truth result. The model also preserves the simulated layer parameters without modifying the design structure.

Physics Explanation. The retrieved physics knowledge links optical emission performance to optical confinement, quantum-well thickness, waveguide thickness, and electron leakage through the EBL. However, the generated response extends beyond these supplied rules by making additional claims about band offsets, polarization-induced electric fields, slope efficiency, wall-plug efficiency, and expected continuous-wave operation. These claims are not directly supported by the provided simulation results or retrieved physics knowledge.

B.3.6 CASE STUDY 2: WORST DESIGN SUGGESTION

For the query "Identify the lowest power AlGaN laser diode," OptiGen identifies a single-objective minimization task targeting P. The system retrieves Structure 140, reporting $P = 3.24$ mW, $I_{\mathrm{th}} = 408$ mA, and $OCF = 0.0795$, which match the ground truth.

Physics Explanation. The structure contains 3-nm AlGaN quantum wells and 8-nm barriers. The response suggests that these dimensions influence carrier confinement, inter-well coupling, and optical performance. It further associates the high Al composition in the wells and barriers with increased band offsets and polarization-related effects, while the 20-nm $\mathrm{Al}_{0.78}\mathrm{Ga}_{0.22}\mathrm{N}$ EBL is described as influencing electron blocking and hole injection.

The relatively low OCF (≈ 0.08) indicates limited overlap between the optical mode and the gain region, which may constrain modal gain and output power. In combination with the high threshold current of 408 mA, these characteristics are consistent with the low simulated output power of 3.24 mW. The retrieved physics knowledge specifically indicates that poor carrier confinement can increase leakage and reduce laser efficiency, while inappropriate MQW and barrier designs can degrade laser performance.

B.3.7 CASE STUDY 3: MULTI-OBJECTIVE OPTIMIZATION SUGGESTION

To evaluate OptiGen's ability to identify balanced laser diode designs, the query "Find balanced designs between output power and efficiency" was provided. The ground truth defines three objectives: maximizing output power, P (mW), minimizing threshold current, I_{th} (mA), and maximizing OCF. The reference selection procedure identifies Structure 0 as the rank-1 design, achieving an P of 846 mW, a I_{th} of 99.2 mA, and an OCF of 0.116. The reference Pareto front contains seven nondominated designs.

The $o3$ model correctly identifies Structure 0 and reproduces its simulated layer parameters and performance values. It also correctly reports the lowest threshold current among the five displayed candidate structures. However, the intent parser represents the optimization objective as maximizing a unified "Design_Score" rather than explicitly preserving the three objectives specified by the ground truth.

Physics Explanation. The generated response includes several unsupported physical explanations, such as claims regarding sub-band separation, polarization-induced interface charges, band offsets, tunnelling, and continuous-wave operation. These claims are not directly established by the retrieved physics rules or the supplied simulation results. Furthermore, the statement that Structure 0 has the best power-to-current efficiency among the five displayed structures is incorrect. Using P/I_{th}, Structure 116 achieves a higher value than Structure 0. The "Design Score" is also reported without a defined scoring formulation.

B.3.8 CASE STUDY 4: CONSTRAINT-BASED SUGGESTION

To evaluate OptiGen's ability to retrieve designs subject to a performance constraint, the query "Retrieve designs operating below 120 mA threshold current" was provided. The ground truth specifies the constraint $I_{\mathrm{th}} < 120$ mA and ranks feasible designs by minimizing I_{th}. The reference selection procedure identifies Structure 0 as the rank-1 design, with an P of 846 mW, a I_{th} of 99.2 mA, and an OCF of 0.116. Structures 1 and 2 are also identified as qualifying designs, with I_{th} of 106 mA and 118 mA, respectively.

The $o3$ model correctly identifies Structure 0 and reproduces its simulated performance values and reported layer parameters. It also retrieves Structures 1 and 2 as additional qualifying designs. The model correctly recognizes the I_{th} constraint; however, the intent parser represents the constraint as $I_{\mathrm{th}} \leq 120$ mA rather than the strict inequality specified in the ground truth, $I_{\mathrm{th}} < 120$ mA. Although this difference does not affect the three reported designs, it could affect the inclusion of a design with a I_{th} exactly equal to 120 mA.

Physics Explanation. The generated response includes several physical claims that are not directly supported by the retrieved physics rules or simulation results. These include a specific

conduction-band offset of approximately 0.21 eV, claims concerning hole injection and tunnelling, and predictions of slope efficiency, thermal roll-over, and high P operation above 800 mW. The statement that Structure 0 has the highest P among all qualifying designs is also not established by the provided ground truth, which reports only the three top-ranked designs.

B.3.9 CASE STUDY 5: PHYSICS-BASED EXPLANATION

To evaluate OptiGen's ability to provide a physics-grounded explanation, the query "Explain the relationship between optical confinement factor and threshold current" was provided. The response correctly identifies the relationship between the Γ and I_{th} through the $\Gamma g_{\mathrm{mat}} = \alpha_i + \alpha_m$. It explains that, for a fixed loss budget, a higher Γ increases the modal gain contribution and therefore reduces the threshold current required to reach lasing.

The $o3$ model further supports the explanation using simulated structures from the database. Structure 3 has an OCF of 0.0856 and an I_{th} of 133 mA, whereas Structure 0 has an OCF of 0.1160 and an I_{th} of 99.2 mA. The response appropriately notes that the relationship is not strictly monotonic across the complete dataset because other structural parameters can simultaneously affect the threshold current.

Some physics explanations are not fully aligned with the underlying principles due to limited domain knowledge, which may lead to unsupported or inaccurate physics interpretations.

B.4 EVALUATION METRIC

This section presents the mathematical formulations of the evaluation metrics used to assess OptiGen across different language models, including $o3$, GPT-5, GPT-4, and Qwen.

B.4.1 RETRIEVER METRICS

Precision@5: Precision@5 measures the proportion of retrieved top-5 designs that are relevant to the ground-truth design set:

$$\mathrm{Precision@5}_i = \frac{\left|\hat{d}_i^{(5)} \cap d_i^*\right|}{\left|\hat{d}_i^{(5)}\right|}. \tag{15}$$

Recall@5: Recall@5 measures the proportion of ground-truth relevant designs retrieved within the top-5 results.

$$\mathrm{Recall@5}_i = \frac{\left|\hat{d}_i^{(5)} \cap d_i^*\right|}{|d_i^*|}. \tag{16}$$

Mean Reciprocal Rank (MRR): MRR measures the reciprocal rank of the first relevant retrieved design, averaged across all queries:

$$\mathrm{MRR} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{RR}_i, \tag{17}$$

where

$$\mathrm{RR}_i = \begin{cases} \frac{1}{r_i}, & \text{if a relevant design is retrieved,} \\ 0, & \text{otherwise.} \end{cases} \tag{18}$$

Here, r_i denotes the rank of the first relevant design.

NDCG@5: NDCG@5 evaluates ranking quality by assigning greater importance to relevant designs appearing at higher ranks:

$$\mathrm{NDCG@5}_i = \frac{\mathrm{DCG@5}_i}{\mathrm{IDCG@5}_i}. \tag{19}$$

The discounted cumulative gain is defined as

$$\mathrm{DCG@5}_i = \sum_{j=1}^{5} \frac{\mathrm{rel}_{ij}}{\log_2(j+1)}, \tag{20}$$

where $\mathrm{rel}_{ij} = 1$ if the design at rank j is relevant and 0 otherwise.

B.4.2 Optimization Metrics

Best Design Accuracy: Best Design Accuracy measures the proportion of queries for which the ground-truth best design appears in the retrieved top-5 results:

$$\mathrm{BestDesignAcc} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}\left(d_i^* \in \hat{d}_i^{(5)}\right). \tag{21}$$

Here, d_i^* denotes the ground-truth best design, and $\mathbb{I}(\cdot)$ is the indicator function.

Performance Accuracy: Performance Accuracy measures the agreement between predicted and ground-truth device performance values using normalized absolute error:

$$A_{i,m} = \max\left(0,\, 1 - \frac{|\hat{d}_{i,m} - d_{i,m}^*|}{|d_{i,m}^*| + \epsilon}\right). \tag{22}$$

Here, $\hat{d}_{i,m}$ and $d_{i,m}^*$ denote the predicted and ground-truth values of metric m, respectively, with $\epsilon = 10^{-9}$.

The performance accuracy of a candidate design is computed as

$$A_{i,d} = \frac{1}{M} \sum_{m=1}^{M} A_{i,m}, \tag{23}$$

where M denotes the number of performance metrics.

B.4.3 Generation Metrics

Physics Reasoning Score: The physics reasoning score measures the cosine similarity between the generated response and the ground-truth physics-rule text in the embedding space:

$$\mathrm{PhysSim}_i = \frac{f(R_i) \cdot f(P_i)}{\|f(R_i)\|_2 \|f(P_i)\|_2}. \tag{24}$$

Here, R_i denotes the generated response, P_i represents the ground-truth physics rules, and $f(\cdot)$ is the sentence embedding function.

Response Completeness: Response completeness is estimated using a word-count-based proxy capped at one:

$$\mathrm{Completeness}_i = \min\left(\frac{W_i}{300}, 1\right). \tag{25}$$

Here, W_i denotes the number of words in the generated response.

Structure Hallucination: Structure hallucination measures the proportion of mentioned structure IDs that are not present in the retrieved design set:

$$\mathrm{StructHall}_i = \frac{|\{s \in M_i : s \notin S_i\}|}{|M_i|}. \tag{26}$$

Here, M_i is the set of structure IDs mentioned in the generated response, while S_i is the set of valid retrieved structure IDs. If no structure IDs are mentioned, the hallucination score is set to zero.

Physics Reasoning Hallucination: Physics reasoning hallucination is operationally defined as the complement of the physics similarity score:

$$\text{PhysHall}_i = 1 - \text{PhysSim}_i. \tag{27}$$

Hallucination-Free Rate: The hallucination-free rate is computed as the complement of the average structure-ID and physics discrepancy measures:

$$\text{HFR} = 1 - \frac{\text{StructureHallucination} + \text{PhysicsHallucination}}{2}. \tag{28}$$

B.4.4 End-to-End Metrics

Intent Understanding Accuracy: Intent Understanding Accuracy measures the proportion of predictions containing a valid intent object:

$$\text{IntentAcc} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\left(\hat{I}_i \text{ is valid}\right). \tag{29}$$

Here, $\hat{I}_i$ denotes the predicted intent representation.

End-to-End Success Rate: A query is considered successful when the ground-truth best design is retrieved within the top-5 results and the physics similarity score is at least 0.5:

$$\text{Success}_i = \mathbb{I}\left(d_i^* \in \hat{d}_i^{(5)} \wedge \text{PhysSim}_i \geq 0.5\right). \tag{30}$$

Average Response Time: Average response time measures the mean time required to process a query and generate a response:

$$\text{AverageTime} = \frac{1}{N}\sum_{i=1}^{N} t_i. \tag{31}$$

Here, t_i denotes the response time for query i, measured in seconds and N denotes the total number of queries.