Circuit Complexity Limits of In-Context Learning

Eishkaran Singh
Department of Computer Science
Thapar Institute of Engineering and Technology
esingh3_be21@thapar.edu

Tanav Singh Bajaj
Department of Computer Science
University of British Columbia
tanav220@student.ubc.ca

Abstract

In-context learning enables transformers to solve tasks from demonstration examples without parameter updates. We establish the first theoretical framework connecting ICL to circuit complexity theory. We formalize ICL-solvability, prove impossibility theorems for NP-hard functions, derive exponential context length lower bounds, and demonstrate architectural limitations of attention mechanisms. Our results characterize which computational problems admit ICL solutions and provide tight resource bounds for learnable function class.

1 Introduction & Mathematical Framework

Large language models exhibit ICL capability by solving tasks through processing input-output examples in context without gradient-based learning. Despite empirical success across diverse domains, this phenomenon lacks theoretical foundations. We address fundamental questions regarding which functions transformers can learn in-context, what context lengths are necessary and sufficient, and which problems are fundamentally un-learnable through ICL. We connect ICL to circuit complexity theory by modeling transformers as circuit families and ICL as program synthesis from examples. The framework establishes connections between transformer computational resources and classical complexity measures, enabling precise analysis of ICL boundaries. We formalize ICL through complexity-theoretic definitions that capture the essential computational aspects of learning from examples.

ICL Problem $\rightarrow$ An ICL problem is defined as a tuple $\mathcal{P} = (D, f, k, n)$ where D represents the problem domain, $f : D \rightarrow D$ denotes the target function to be learned, k specifies the number of input-output examples provided in context, and n bounds the maximum allowable context length. Captures the resource constraints inherent in practical ICL scenarios while enabling theoretical analysis.

ICL-Solvability $\rightarrow$ A transformer T_θ with parameters θ is said to ICL-solve problem $\mathcal{P}$ with error ϵ if the following probabilistic condition holds: for a randomly sampled set of examples $S = \{(x_i, f(x_i))\}_{i=1}^{k}$ and a query input x, we have

$$\Pr_{S \sim D^k} \left[\Pr_{x \sim D} [T_\theta(S \| x) \neq f(x)] \leq \epsilon \right] \geq 1 - \delta$$

subject to the constraint $|S \| x| \leq n$, where $S \| x$ denotes the concatenation of examples and query within the context window. This captures the requirement that the transformer must generalize accurately to new inputs based on the provided examples, with high probability over both the choice of examples and test inputs.

Circuit Representation $\rightarrow$ We model transformers with L layers and hidden dimension d as circuit families $\{C_n\}$ where each circuit C_n has size $O(L \cdot d^2 \cdot n)$ and processes sequences of length at most n. The attention mechanism within each layer computes weighted combinations using softmax normalization over dot-product similarities between query and key vectors. This circuit representation enables analysis of ICL through established complexity theory while accounting for the specific architectural constraints imposed by transformer designs. The central insight underlying our theoretical framework is that ICL requires the transformer circuit to simultaneously extract computational procedures from input-output examples and execute these procedures on new inputs. This

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: What Can't Transformers Do?.

meta-computational process, where the transformer must parse examples to identify the underlying function and then apply this function to query inputs, fundamentally determines the computational complexity of ICL.

2 Impossibility Results

We establish the first impossibility theorems for in-context learning by connecting ICL to computational complexity theory.

Result 1 There exist functions $f \in \mathsf{NP}$ such that no polynomial-size transformer can ICL-solve f, regardless of context length. Consider the Boolean satisfiability function SAT : $\{0,1\}^m \to \{0,1\}$ where SAT$(\phi) = 1$ if and only if formula ϕ is satisfiable. Suppose a transformer T with polynomial circuit size can ICL-solve SAT. We construct a family of formulas $\{\phi_i\}$ where each ϕ_i has $2^{m/2}$ satisfying assignments distributed across exponentially many variable settings using cryptographic pseudorandom generators. Any polynomial-time algorithm distinguishing satisfiable from unsatisfiable instances would solve SAT in polynomial time, contradicting the Exponential Time Hypothesis. The number of examples required grows exponentially with m, exceeding any polynomial context bound. Extending this analysis to show architectural limitations of attention mechanisms.

Result 2 Functions computable by unbounded fan-in circuits cannot be ICL-solved by transformers with softmax attention. The softmax attention mechanism computes weighted averages with weights $\exp(q_i^T k_j)/\sum_\ell \exp(q_i^T k_\ell)$. For unbounded fan-in functions like MAJORITY on n inputs, attention must assign non-negligible weight to all n positions simultaneously. However, softmax normalization ensures individual weights decay as $O(1/n)$, insufficient to maintain constant precision across all inputs. We formalize this through a packing argument showing bounded-precision arithmetic cannot compute unbounded fan-in functions.

3 Context Length Lower Bounds

We derive tight bounds on context length requirements for different complexity classes through information-theoretic arguments.

Result 3 For functions computable by depth-d circuits, the minimum context length for ICL scales as $\Omega(2^d)$. Consider the nested XOR function f_d computable by depth-d circuits:

$$f_d(x_1, \ldots, x_{2^d}) = \bigoplus_{i=1}^{2^{d-1}} \left(\bigoplus_{j=1}^{2^{d-1}} x_{i \cdot 2^{d-1}+j} \right)$$

Through communication complexity analysis, the examples serve as a communication channel. For depth-d functions, information requirements scale exponentially with depth due to hierarchical structure. Each nesting level doubles information complexity, yielding the $\Omega(2^d)$ bound.

4 Related Work

The theoretical understanding of ICL has emerged through several approaches. [3] provided the first rigorous analysis by showing transformers learning linear functions can implement ridge regression and gradient descent with performance bounds depending on condition numbers. [7] established connections to iterative optimization, proving attention layers simulate gradient descent steps with each layer corresponding to one iteration. [1] framed ICL as approximate Bayesian inference, demonstrating transformers encode posterior distributions over function spaces. [9] interpreted ICL as meta-learning, arguing transformers acquire general-purpose learning algorithms during pre-training. [4] investigated expressivity, characterizing learnable function classes under architectural constraints. Our work departs fundamentally by focusing on complexity-theoretic limitations. We show ICL inherits the same computational limitations as classical algorithms, providing the first impossibility results in this domain.

5 Conclusion

We have established the first comprehensive theoretical framework for understanding in-context learning through circuit complexity analysis. Our impossibility theorems demonstrate that ICL faces fundamental computational barriers, with additional constraints from transformer architecture. The exponential context length lower bounds provide precise resource requirements for different complexity classes, while separation results establish an ICL difficulty hierarchy. The framework enables principled ICL analysis by connecting transformer computational resources to established complexity measures. This reveals that while ICL enables learning from examples without parameter updates, it cannot overcome fundamental computational limitations. The results provide theoretical insights and practical guidance for understanding when ICL will succeed or fail on specific problem classes.

References

[1] Ekin Akyürek, Dale Schuurmans, Jacob Andreas, Tengyu Ma, and Denny Zhou. What learning algorithm is in-context learning? Investigations with linear models. In *International Conference on Learning Representations*, 2023.

[2] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901, 2020.

[3] Shivam Garg, Dimitris Tsipras, Percy S Liang, and Gregory Valiant. What can transformers learn in-context? A case study of simple function classes. *Advances in Neural Information Processing Systems*, 35:19630–19642, 2022.

[4] Yingcong Li, Sébastien Bubeck, Ronen Eldan, Allie Del Giorno, Suriya Gunasekar, and Yin Tat Lee. Textbooks are all you need II: phi-1.5 technical report. *arXiv preprint arXiv:2309.05463*, 2023.

[5] Arvind Mahankali, Tatsunori B Hashimoto, and Tengyu Ma. One step of gradient descent is provably the optimal in-context learner with one layer of linear self-attention. *Advances in Neural Information Processing Systems*, 36, 2023.

[6] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

[7] Johannes Von Oswald, Eyvind Niklasson, Ettore Randazzo, João Sacramento, Alexander Mordvintsev, Andrey Zhmoginov, and Max Vladymyrov. Transformers learn in-context by gradient descent. In *International Conference on Machine Learning*, pages 35076–35104, 2023.

[8] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35:24824–24837, 2022.

[9] Sang Michael Xie, Aditi Raghunathan, Percy Liang, and Tengyu Ma. An explanation of in-context learning as implicit Bayesian inference. In *International Conference on Learning Representations*, 2022.