CIRCUIT: A BENCHMARK FOR CIRCUIT INTERPRETATION AND REASONING CAPABILITIES OF LLMS

Lejla Skelic, Yan Xu, Matthew Cox, Wenjie Liu, Tao Yu, Ruonan Han



TERAHERTZ INTEGRATED ELECTRONICS GROUP

Ruonan Han

Associate Professor, Director of the Center for Integrated Circuits & Systems







Motivation

Large Language Models (LLMs) hold promise in automating various aspects of analog circuit design.

LLMs have already been integrated into frameworks that automate aspects of the design process.

Existing benchmarks fail to assess LLMs' reasoning capabilities specific to analog circuits; hence, the effectiveness of LLMs in real-world circuit design may be limited.

Therefore, we curated the CIRCUIT dataset, comprising 510 structured question-answer pairs that test LLMs on various circuit topologies and configurations.

Results show that **even advanced LLMs**, like OpenAI's GPT-40, **struggle with circuit reasoning tasks**, highlighting the need for further advancements.



Project Overview



Automatic Evaluation

Human Evaluation

Figure 1: Overview of the CIRCUIT dataset and experiment setup.



School of Engineering



CIRCUIT Dataset | Curation

The dataset was made by adapting problems from various sources. **Problems** were **adapted into templates that can feature multiple numerical setups**.

We created 5 numerical setups for each template.

Some templates have associated diagrams, and some diagrams have corresponding netlists with syntax adapted to our dataset.



Template:

Calculate the current I in Amperes in the given circuit.

Numerical Setup: Assume V = 5 V and R = 100 Ohms.

Netlist: V N1 0 R N1 0 ; I

Figure 2: Example datapoint from the CIRCUIT dataset.

CIRCUIT Dataset | Statistics

The dataset consists of 510 questions derived from 102

templates, with 5 numerical setups each. 93 templates include diagrams, 79 of which include netlists.

Templates are divided into 4 categories and 5 levels (Figure 3).





Figure 3: Templates distribution across categories and levels.



Evaluation | Metrics and Methods

Metrics

Global $A_{global} = \frac{\# \text{ correctly answered questions}}{\# \text{ total questions}}$

Template
$$A_{\text{template},k/n} = \frac{\sum_{i=1}^{m} A_{t_i,k}}{m}$$
, where $A_{t_i,k} = \begin{cases} 1 & \text{if at least } k \text{ out of } n \text{ setups} \\ & \text{are answered correctly} \\ 0 & \text{otherwise} \end{cases}$

Methods

Automatic: final numerical answer **Human**: final numerical answer + errors in the solution steps + qualitative analysis



Experimental Setup

Models

GPT 4-turbo, GPT 40, Gemini 1.5 Pro

Experiments

Prompts: 0-shot, 0-shot with netlists, 1-shot, and 1-shot with netlists

12 experiments: 3 models x 4 prompts

Evaluation

Automatic and human

Final numerical answer within 0.001



		Accuracies (%)					
Model	Prompt	Entire dataset (102 templates)					
		Glob.	Template				
			5/5	4/5	3/5		
	0-s	38.4	18.6	30.4	40.2		
GPT	0-s w/ netlists	38.2	19.6	32.4	35.3		
4-turbo	1-s	39.2	15.7	32.4	40.2		
	1-s w/ netlists	38.2	22.6	31.4	34.3		
	0-s	46.7	27.5	35.3	48.0		
CPT 4o	0-s w/ netlists	48.0	27.5	37.3	47.1		
GF140	1-s	39.6	23.5	33.3	38.2		
	1-s w/ netlists	43.1	24.5	34.3	43.1		
Gemini 1.5-pro	0-s	36.3	18.6	29.4	33.3		
	0-s w/ netlists	34.7	13.7	25.5	33.3		
	1-s	32.0	10.8	21.6	30.4		
	1-s w/ netlists	32.2	13.8	23.5	33.3		

Table 1.1: Accuracies for the Entire CIRCUIT Dataset.



	Prompt	Accuracies (%)							
Model		Ques	stions Wi	thout Ne	tlists	Questions With Netlists			
		Glob	Template			Glob	Template		
		0100.	5/5	4/5	3/5	0100.	5/5	4/5	3/5
	0-s	617	39.1	56.5	60.9	31.7	12.7	22.8	34.2
GPT 4-turbo	0-s w/ netlists	01.7				31.4	13.9	25.3	27.9
	1-s	62.6	39.1	56.5	69.6	32.4	8.9	25.3	31.7
	1-s w/ netlists	63.5	43.5	60.9	65.2	30.9	16.5	22.8	25.3
GPT 40	0-s	67.0	47.8	65.2	69.6	40.8	21.5	26.6	41.8
	0-s w/ netlists					42.5	21.5	29.1	40.5
	1-s	67.8	56.5	65.2	65.2	31.4	13.9	24.1	30.4
	1-s w/ netlists	63.5	34.8	52.2	69.6	37.2	21.5	29.1	35.4
Gemini 1.5-pro	0-s	55.7	26.1	56.5	56.5	30.6	16.5	21.5	26.6
	0-s w/ netlists					28.6	10.1	16.5	26.6
	1-s	56.5	26.1	43.5	65.2	24.8	6.3	15.2	20.3
	1-s w/ netlists	53.0	21.7	43.5	56.5	26.1	11.4	17.7	26.6

Table 1.2: Accuracies on CIRCUIT Dataset Subsets

(Questions without netlists - 23 templates; Questions with Netlists - 79 templates).





Automatic

Best performing prompt varies by the model.

GPT 40 is the best-performing model on the dataset (Table 1).

- **Template accuracy** decreases as the value of *k* increases, **highlighting the sensitivity and the inconsistency of the models across numerical setups**.
- **Template accuracies reveal the narrowness of the range of topologies understood by the models**. Global accuracy gives a false impression of model performance.

Questions without netlists yield higher average scores.

Human - GPT 40 responses

Results consistent with automatic evaluation, highlighting its reliability.

Mathematical and formatting errors are rare; reasoning errors predominant.

Dataset Subset	Prompt	GPT 40 Response Error Rate (%) by Error Type						
		Math	Formatting	Reasoning	Topology	Direction		
Questions With Netlists	0-s w/ netlists	7.1	1.3	58.5	36.2	4.1		
	1-s w/ netlists	8.4	0.5	61.8	39.2	3.5		
Questions W/O Netlists	1-s	1.7	0.0	34.8	16.5	4.4		

Table 2: Human Error Analysis of GPT-40 Responses.





Results | Qualitative - GPT 40

Strengths

Generally used appropriate tools, formulas, and identified circuit elements correctly.

Main Challenges

Most mistakes stem from topology reasoning —misinterpreting component connections, directions, and interactions. Occasional math errors: approximations affecting answer precision requirements and careless mistakes in equation manipulation.

Role of Netlists

Netlists improved topology understanding but weren't consistently referenced in responses.

Other Issues

Occasional conceptual misunderstandings, instruction lapses, and circuit configuration hallucinations.





Conclusions

LLM Capabilities

Models show some capacity for reasoning in analog circuits, but their understanding of topologies is limited.

New Metrics

The pass@k/n metric provides deeper insight into model performance across varied setups, beyond a single accuracy score.

Dataset Design

CIRCUIT's structured, transparently homogeneous design helps reveal nuanced LLM limitations, such as poor generalization across numerical setups.

Automatic Evaluation

Combining global accuracy with template pass@k/n enabled by CIRCUIT's structured design allows quick, scalable, and nuanced assessment of LLM performance.

Future Directions

Expanding the dataset to more complex topologies and design tasks, refining prompting methods, etc.



Q & A



Automatic Evaluation

Human Evaluation





Dataset Subset	Prompt	GPT 40 Response Accuracies (%)								
		Automatic				Human				
		Glob	Template			Glob	Template			
		0100.	5/5	4/5	3/5	0100.	5/5	4/5	3/5	
Questions With Netlists	0-s w/ netlists	42.5	21.5	29.1	40.5	36.5	17.7	27.9	35.4	
	1-s w/ netlists	37.2	21.5	29.1	35.4	31.9	19.0	27.9	31.7	
Questions W/O Netlists	1-s	67.8	56.5	65.2	65.2	63.5	52.2	65.2	65.2	

Table 3: The accuracy of GPT-40 responses evaluated automatically versus by humans



Model	Prompt	GPT 40 Accuracy (%) per category					
		Analog	Basic	Power	Radio Frequency		
Questions With Netlists	0-s w/ netlists	30.6	49.4	30.0	20.0		
	1-s w/ netlists	28.2	45.0	26.7	10.0		
Questions W/O Netlists	1-s	33.3	80.0	100.0	60.0		

Table 4: Global Accuracy per Category Derived from Human Analysis Results.





Results | Qualitative | Misunderstand. Topologies



GPT's RESPONSE:

Let's analyze and solve for the power delivered to R3 in the given circuit step by step.

```
Step 1: Identify the Loop Equations
```

We will use Kirchhoff's Voltage Law (KVL) for the two loops with given voltage sources and resistors.

Loop 1 (left loop): -V1 + I1R1 + I3R3 = 0

Loop 2 (right loop): V2 + I2R2 - V3 - I3R3 = 0

Step 2: Write the Equations in Terms of Currents Expressing the above in terms:

The left loop does not include R3 but does include another resistor and another voltage source. Additionally, the signs in the KVL equations are inconsistent.



Results | Qualitative | Success with Netlists



Step by step solution:

Netlist:

I 0 N1 U R1 N1 0 10 ; I R2 N1 N2 20 V N2 0 30

Problem: Find the current I in A in the circuit in the figure 133-01. Assume that U = 0.6A. **Full Solution:** Using KVL, we get that $-30V + 20\Omega(I - U) + 10\Omega(I) = 0$. Therefore, 30 + 20U = 30I, i.e. $I = 1 + \frac{2}{2}U$ Answer: 1.4

GPT's RESPONSE:

1. From the netlist: - A current source $U=0.6\,A$ is connected between ground (0) and node N1.- $R1 = 10 \,\Omega$ is connected between N1 and ground (0). - $R2 = 20 \,\Omega$ is connected between N1 and N2. - A voltage source V = 30 V is connected between ground (0) and N2. 2. Current U = 0.6 A flows from ground to N1. 3. Apply Kirchhoff's Current Law (KCL) at node N1: Add 30 to both sides: $I_U = I_{R1} + I_{R2}$ where - I_U is the current from the current source, which is $0.6\,A.$ - I_{R1} is the current through R1. - I_{R2} is the current through R2. Divide by 3: 4. Calculate the current through R1: $I_{R1} = \frac{V_{N1}}{R1} = \frac{V_{N1}}{10}$ 7. Calculate I_{R1} , the current through R1: 5. Calculate the current through R2:

$$I_{R2} = \frac{V_{N1} - V_{N2}}{R2} = \frac{V_{N1} - 30}{20}$$

Using KCL:

T Al Hard

 $0.6 = \frac{V_{N1}}{10} + \frac{V_{N1} - 30}{20}$

6. Simplify the equation to solve for V_{N1} : Multiply through by 20 to clear the denominators:

 $12 = 2V_{N1} + V_{N1} - 30$

 $12 = 3V_{N1} - 30$

 $42 = 3V_{N1}$

 $V_{N1} = 14 V$

$$V_{R1} = \frac{V_{N1}}{10} = \frac{14}{10} = 1.4 A$$

8. The current I labeled in the diagram is the current through

R1, which is I_{R1} .

Final Numerical Answer: 1.400000 EEE CONCOLORENGING

zman - concept of Computing

18

Results | Qualitative | Misunderstand. Directions + Failure with Netlists



GPT's RESPONSE:

•••

1. From the netlist and the circuit diagram (Figure 134-01): - U = 0.4V - R1 = 4Ω - I1 = 1A (current flowing from N2 to ground through the current source)

The current is flowing from ground to N2.

Results | Qualitative | Minor Reasoning Error



Netlist:

V1 N1 0 ; i1 R1 N1 N2 ; i2 R2 N2 0 ; i3 R3 N2 0 ; i4

Problem: Determine the current i_3 in Amperes in the circuit in the figure 7-01. Assume $V_1 = 10V$, $R_1 = 5\Omega$, $R_2 = 10\Omega$, and $R_3 = 50\Omega$.

Full Solution: Using KCL, we find that $-i_1 = i_2 = i_3 + i_4$. We also have two independent KVL equations: $-v_1 + v_2 + v_3 = 0$ and $-v_3 + v_4 = 0$. Solving with constitutive relations, we get that $i_3 = \frac{V_1 R_3}{R_1 R_2 + R_2 R_3 + R_1 R_3}$. Answer: 0.625

GPT's RESPONSE:

····

•••

The voltage drop across the parallel combination of R2 and R3 is the same as the voltage across R1.

R1 and R2 || R3 form a voltage divider.





Results | Qualitative | Math Errors

GPT's RESPONSES:

(1)
...
$$R_{out} = \frac{1}{250 \times 10^{-3} \text{ s}} = \frac{1}{0.25} \text{ k} = 4 \text{ k}$$
Small computation mistake.
Therefore, the small-signal output resistance *R* is:

Therefore, the small-signal output resistance R_{out} is: Final Numerical Answer: 4.000000

(2)

•••

 $= \frac{6806.25}{6187.5} \approx 1.099$ Approximation resulting in the FNA being outside the 0.001 error range. Final Numerical Answer: 1.099000

(3)

... 4. Convert Back to Noise Figure in dB:

 $NF_{\text{total}} = 10 \log_{10}(1.904893) = 10 \times 0.280173$

Small error in the log approximation.

 $NF_{\rm total} \approx 2.801732~{\rm dB}$ Final Numericul Aliswel . 2.001/32



MIT AI Hardware Program





Results | Qualitative | Not Following Instructions

GPT's RESPONSE:

Since the solution requested is to find $v_b(t)$ and an explicit time was not provided, the final answer must be left in the general solution form, which is:

 $v_{b}(t) = 5 \exp(-2t)$

Therefore, for a general time t:

Final Numerical Answer: $v_b(t) = 5 \exp(-2t)$ The value of *t* was given in the setup.

Failed to follow the formatting instructions in the prompt – the correct answer format is Final Numerical Answer: Unknown in the case that insufficient information was provided.







Netlist: V N1 0 10 Rp N1 Vp 100 OPA N3 Vp Vn R1 Vn 0 R2 N3 Vn R3 N3 0

Problem: Find the current in Amperes through resistor R_1 in the circuit in the figure 140-01. Assume that the op-amp is ideal. Assume that $R_1 = 2\Omega$, $R_2 = 2\Omega$, and $R_3 = 3\Omega$. Full Solution: Since we have an idea op-amp, $V_p = V_n$ and $i_p = i_n = 0$. Therefore, $I_s = 0$, and thus $V_n = 10V$. From here we get that $I_1 = \frac{10V}{R_1}$ Answer: 5

GPT's RESPONSE:

... 3. Calculate Vp:

•••

- Since the non-inverting input (Vp) is connected to the junction of the 100 Ohm resistor and the 10 V source, we need to find this voltage.

- Voltage drop across the 100 resistor is not needed since Vp can be assumed to be at virtual ground if Vn = Vp.

Vn=Vp at virtual ground is a common configuration for op-amps in negative feedback, but it is not true for this topology.

