



Mehrnoosh Sadrzadeh

Professor of CS and Royal Academy of Eng. Research Chair

Structure-Informed Visual Language Understanding with Quantum Circuits

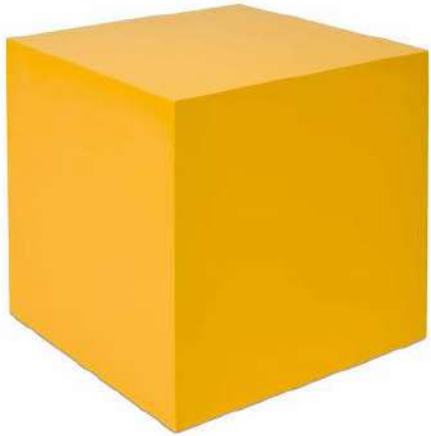


Vision Language Understanding

- Aligning image and text
- Given an image and many pieces of text, decide what is the best match.
- First large scale model: CLIP, OpenAI 2021
- There exists challenges
- Structural Reasoning or Generalisation!



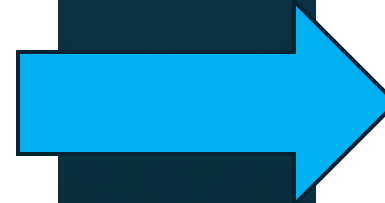
Challenge (1): Compositional **Attribute** Generalisation



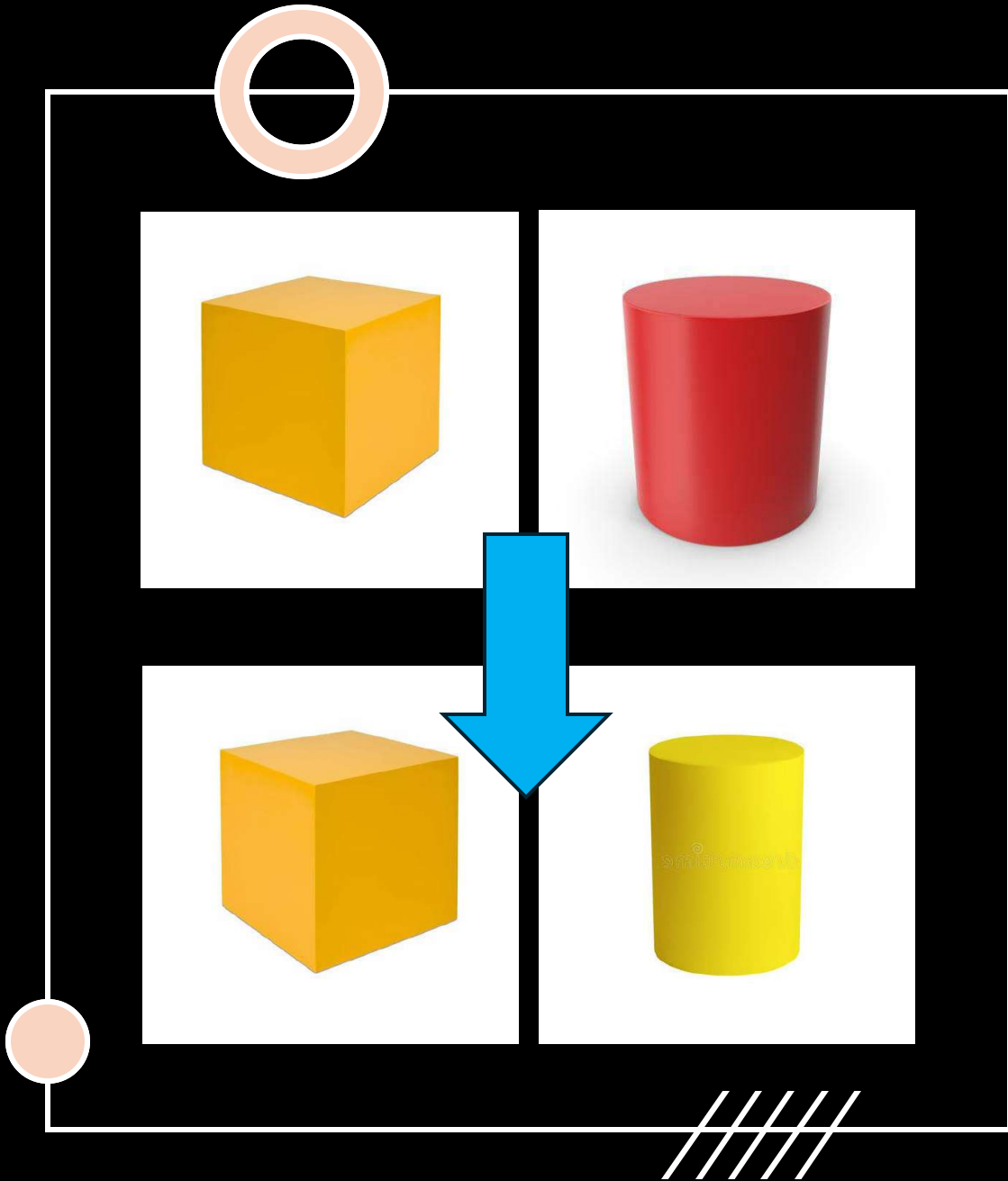
Yellow Cube



Red Cylinder

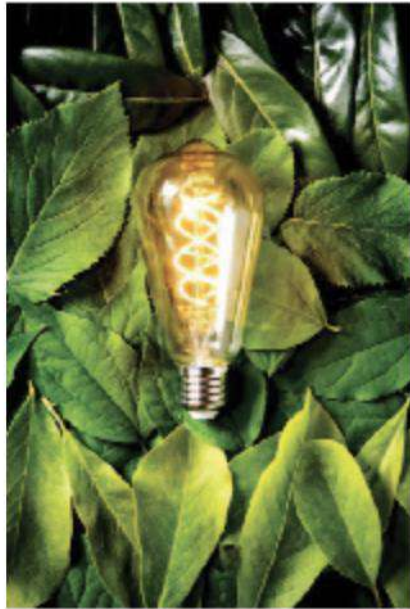


?



Challenge (2): Compositional
relation generalisation

Challenge (3): Vision Language Understanding



(a) some plants
surrounding a
lightbulb



(b) a lightbulb surrounding some plants



VLM challenges in practice:

Self driving cars:

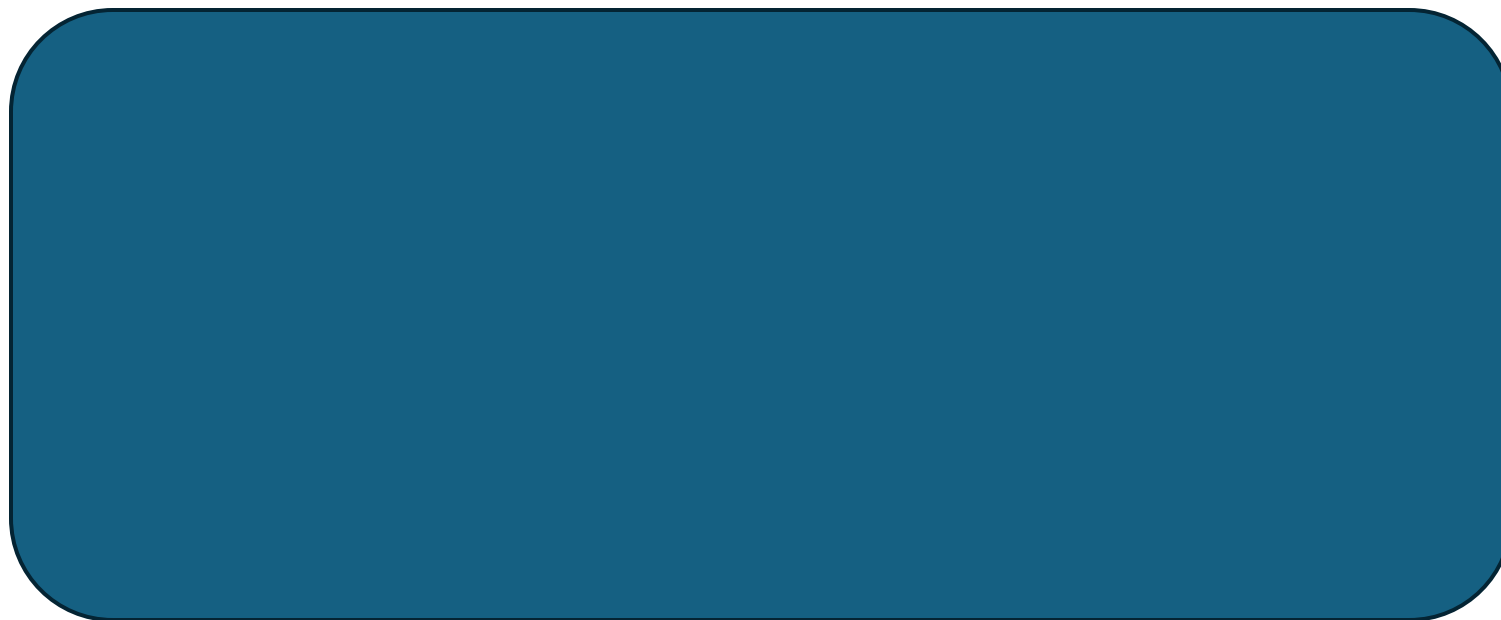
- Never seen a Yellow Porsche,
- See one when crossing the road,
- Stop or carry on walking?



Does CLIP Bind Concepts? Probing Compositionality in ...

by M Lewis · 2024 · Cited by 131 — **CLIP** can compose concepts in a single-**object** setting, but in situations where concept **binding** is needed, performance drops dramatically.

Model	Train	Val	Gen
CLIP	26.80	14.99	0.00





Does CLIP Bind Concepts? Probing Compositionality in ...

by M Lewis · 2024 · Cited by 131 — **CLIP** can compose concepts in a single-**object** setting, but in situations where concept **binding** is needed, performance drops dramatically.

Model	Train	Val	Gen
CLIP	26.80	14.99	0.00
CLIP-FT	49.59 _{0.44}	0.00 _{0.00}	0.00 _{0.00}



Does CLIP Bind Concepts? Probing Compositionality in ...

by M Lewis · 2024 · Cited by 131 — **CLIP** can compose concepts in a single-**object** setting, but in situations where concept **binding** is needed, performance drops dramatically.

Model	Train	Val	Gen
CLIP	26.80	14.99	0.00
CLIP-FT	49.59 _{0.44}	0.00 _{0.00}	0.00 _{0.00}
CSP	30.40 _{0.11}	0.12 _{0.01}	0.03 _{0.00}



Does CLIP Bind Concepts? Probing Compositionality in ...

by M Lewis · 2024 · Cited by 131 — **CLIP** can compose concepts in a single-**object** setting, but in situations where concept **binding** is needed, performance drops dramatically.

Model	Train	Val	Gen
CLIP	26.80	14.99	0.00
CLIP-FT	49.59 _{0.44}	0.00 _{0.00}	0.00 _{0.00}
CSP	30.40 _{0.11}	0.12 _{0.01}	0.03 _{0.00}
Add	25.41 _{0.13}	26.03 _{0.07}	25.47 _{0.18}
Mult	25.67 _{0.12}	25.95 _{0.09}	25.78 _{0.09}
Conv	24.83 _{0.06}	26.36 _{0.55}	24.95 _{0.11}

$$\sum_{j=0}^{n-1} \mathbf{a}_j \mathbf{b}_{i-j}$$



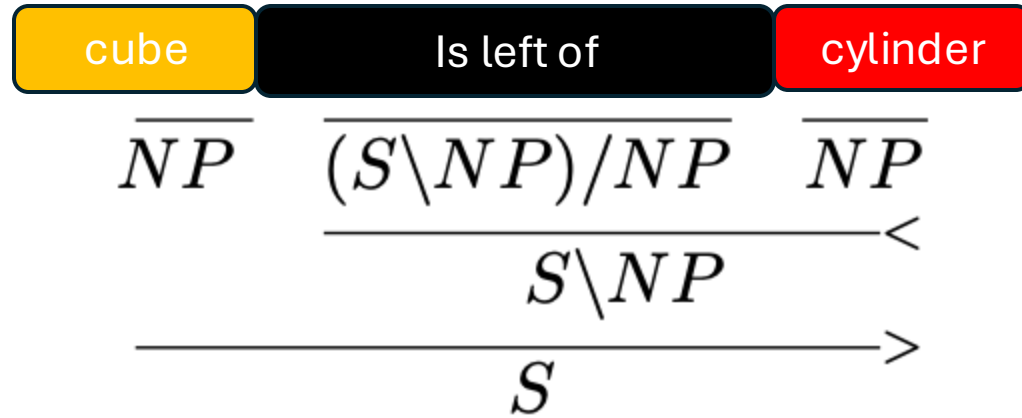
Does CLIP Bind Concepts? Probing Compositionality in ...

by M Lewis · 2024 · Cited by 131 — **CLIP** can compose concepts in a single-**object** setting, but in situations where concept **binding** is needed, performance drops dramatically.

Model	Train	Val	Gen
CLIP	26.80	14.99	0.00
CLIP-FT	49.59 _{0.44}	0.00 _{0.00}	0.00 _{0.00}
CSP	30.40 _{0.11}	0.12 _{0.01}	0.03 _{0.00}
Add	25.41 _{0.13}	26.03 _{0.07}	25.47 _{0.18}
Mult	25.67 _{0.12}	25.95 _{0.09}	25.78 _{0.09}
Conv	24.83 _{0.06}	26.36 _{0.55}	24.95 _{0.11}
TL	67.19 _{0.26}	0.00 _{0.00}	0.00 _{0.00}

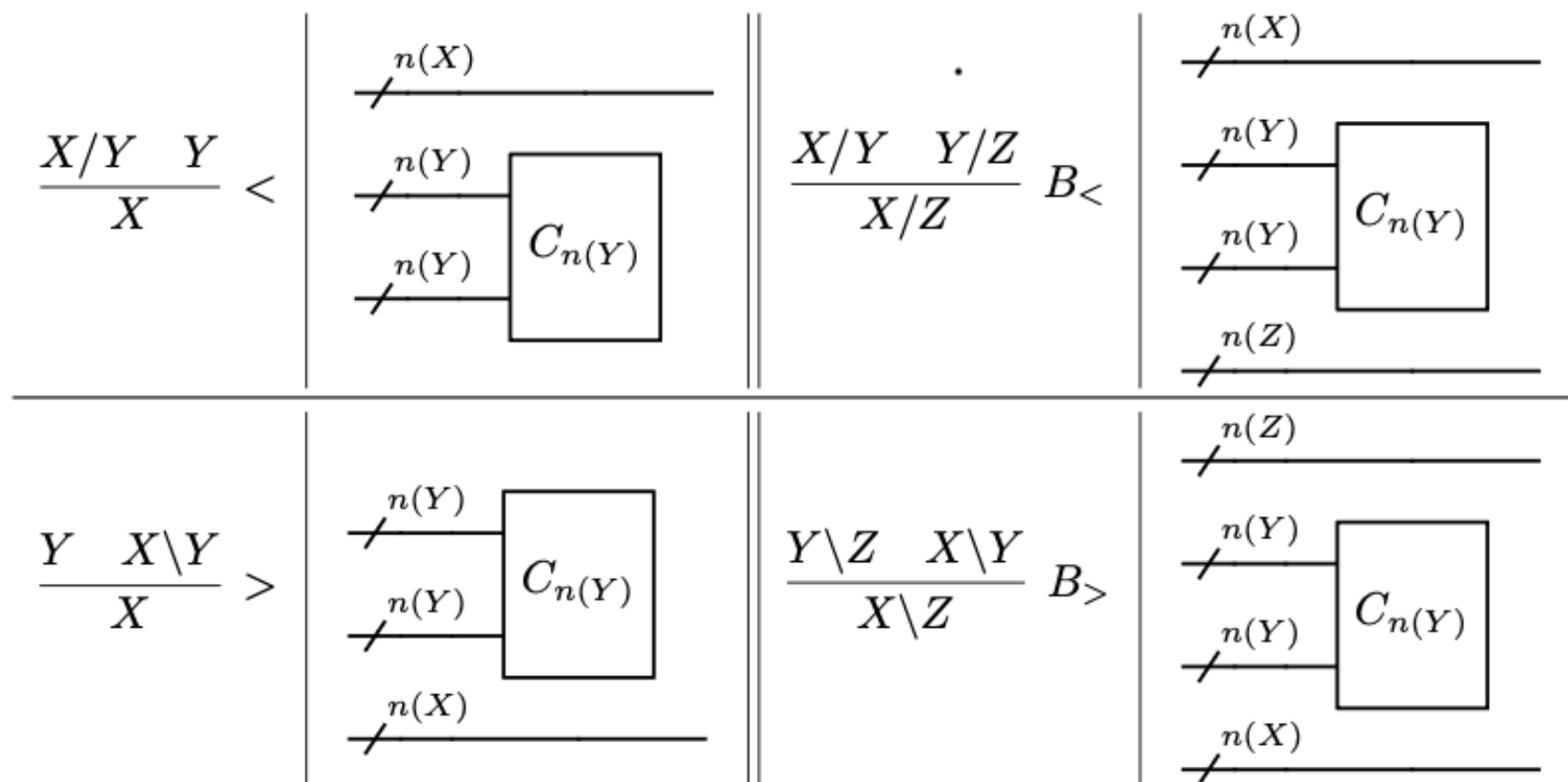
Tensor Logic

Tensor Logic

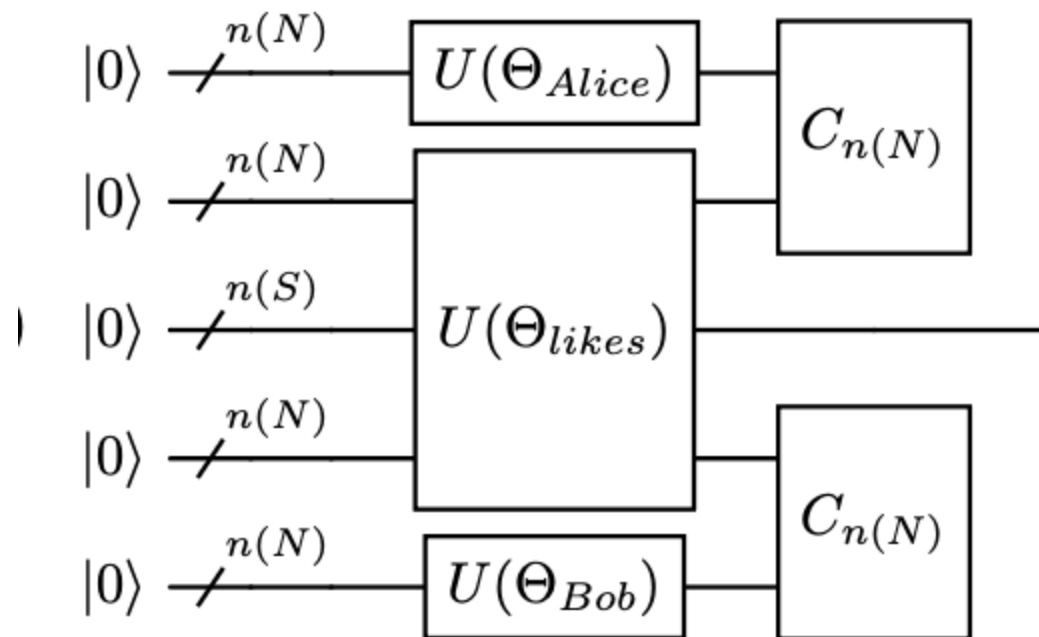
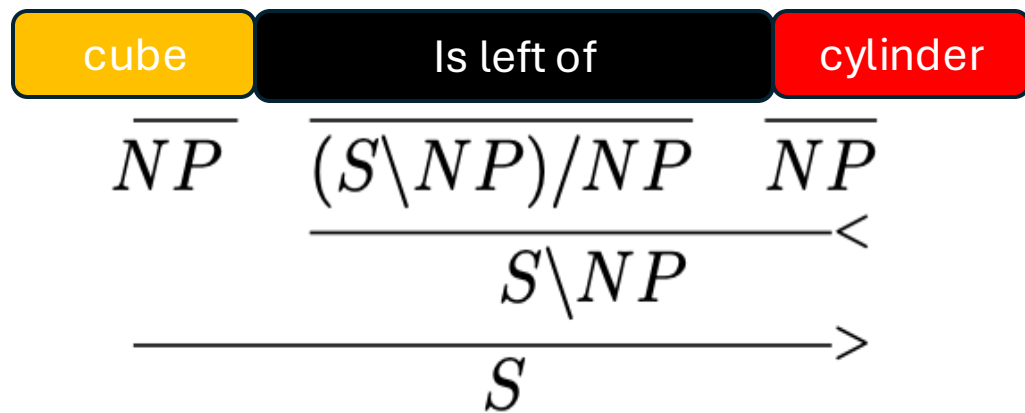


$$T_i^{\text{cube}} T_{ijk}^{\text{is left of}} T_k^{\text{cylinder}}$$

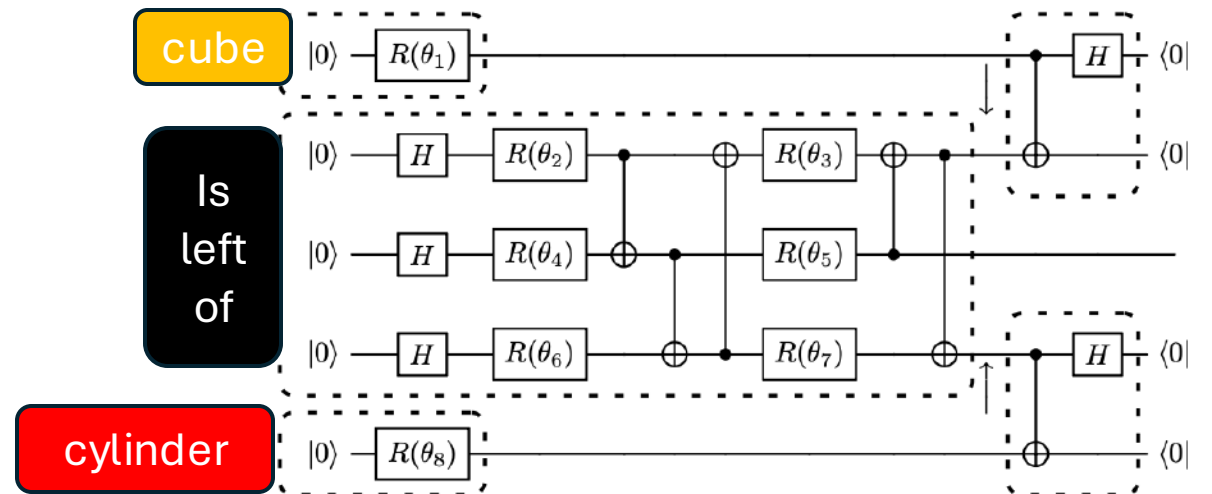
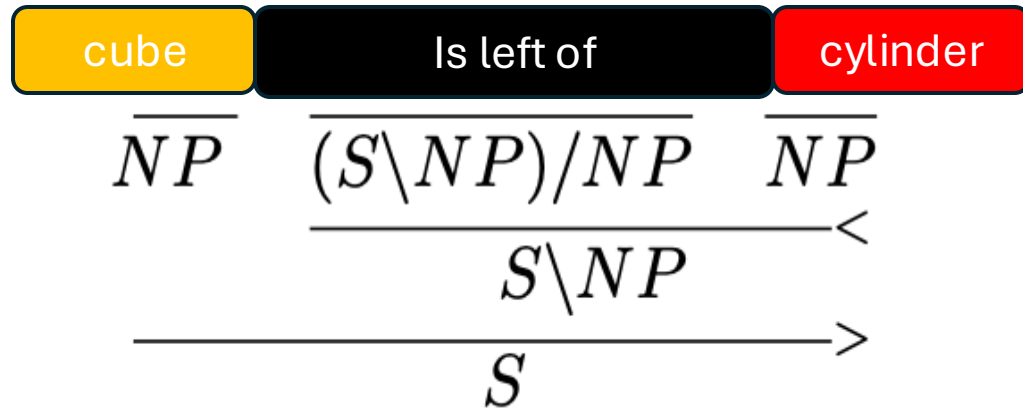
Quantum Tensor Logic



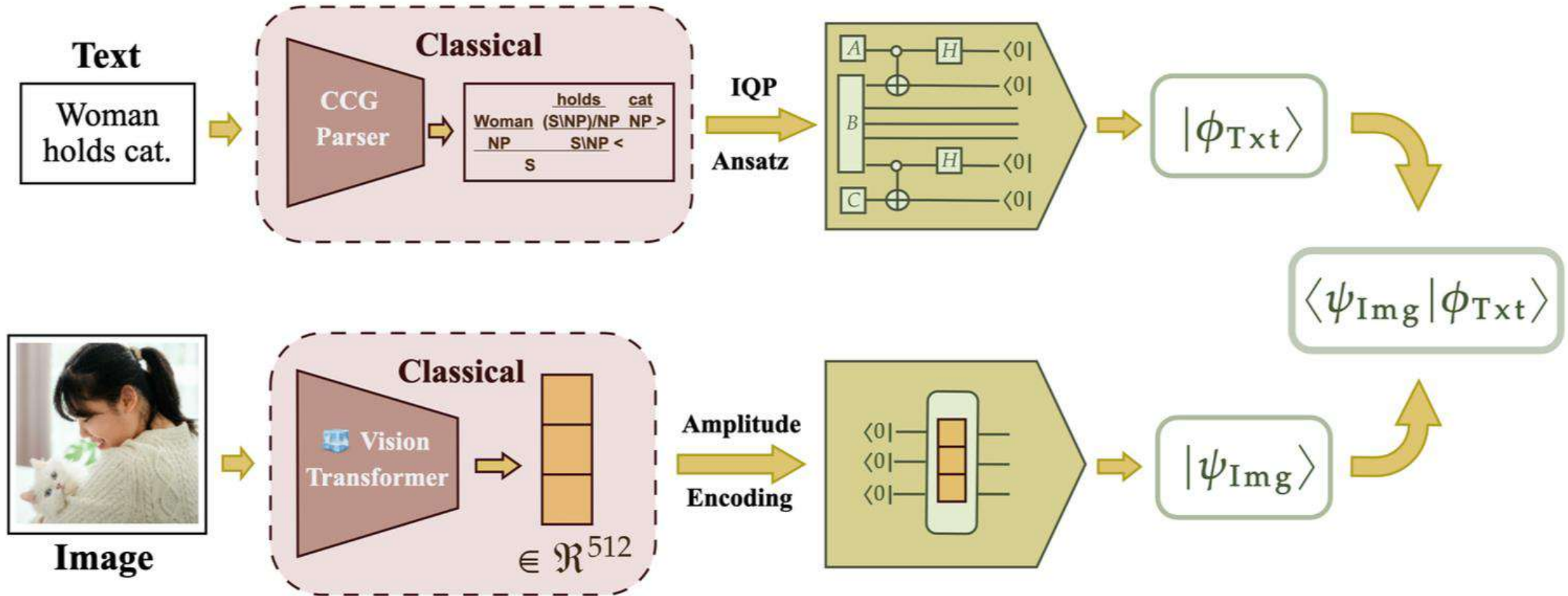
Example



Example



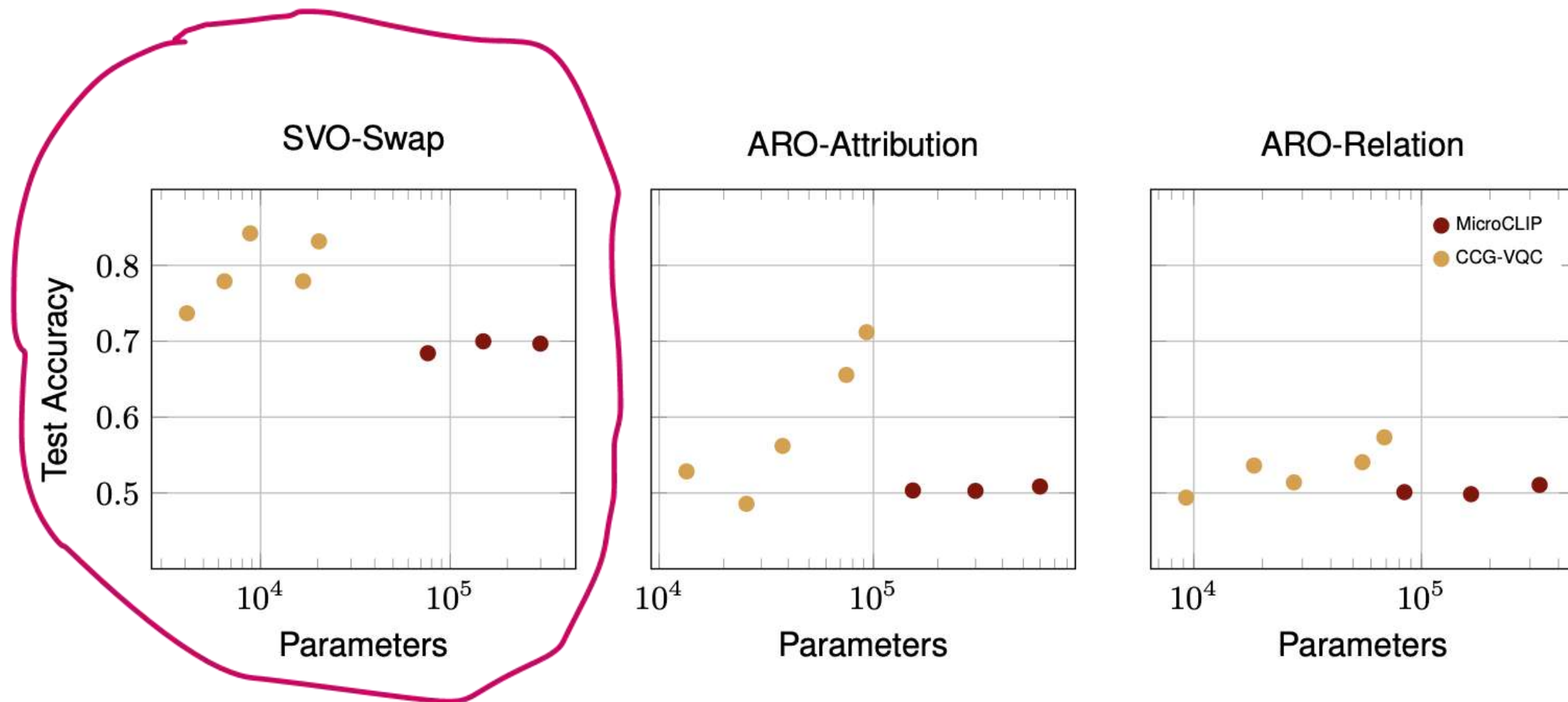
Architecture



Classical Simulation Results

	SVO-Swap	ARO	
		Attribution	Relation
QBoW	50.00	50.00	50.00
CCG-VQC _{SAP}	83.16	71.19	57.33
MicroCLIP-8	68.42	50.33	50.11
MicroCLIP-16	69.99	50.27	49.84
MicroCLIP-32	69.68	50.85	51.05
CLIP	57.89	61.00	51.53
OpenCLIP	63.16	59.13	50.71

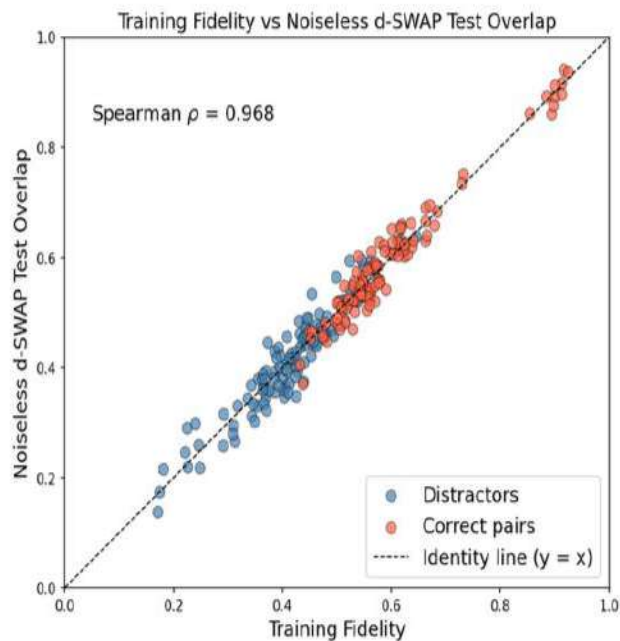
Parameter Efficiency



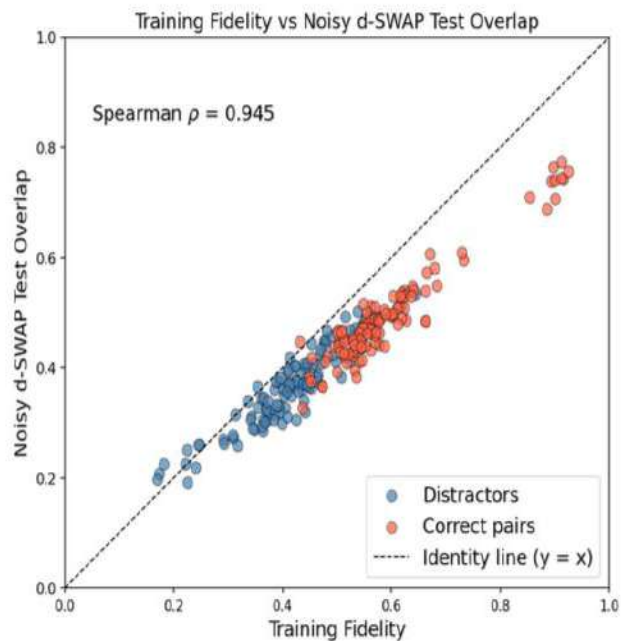
Results

Models	Method	Train	OOD-Valid	OOD-Test	Best Test	Parameters
MHE	Binary	99.00%	73.00%	85.88%	91.66	279
Collage Encoding						
	Angle Enc. with Entangl.	82.20%	61.05%	72.32%	75.00	426
	Ampl. Enc. with Entangl.	62.28%	40.00%	48.58%	50.00	426
Q-Multi-CLIP						
	Angle Enc.	76.60%	61.25%	55.33%	60.16	480
	Amplitude Enc.	59.02%	45.00%	50.00%	50.00	480
Baselines						
Q-Single-CLIP						
	Angle Enc.	55.26%	44.50%	53.09%	55.00	792
	Amplitude Enc.	50.64%	62.00%	42.93%	48.33	792
Classic-Rel-CLIP	CTP	75.94%	11.14%	50.00%	53.20	151,540,225

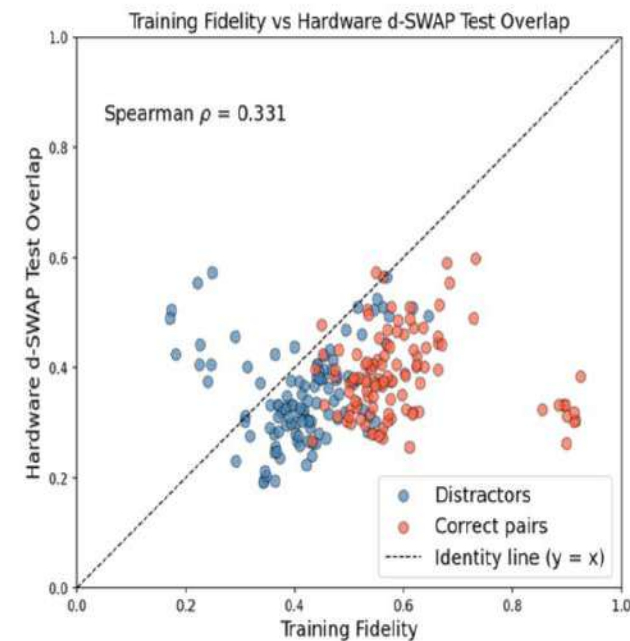
Hardware Runs



(a) IBM Aer (Noiseless)

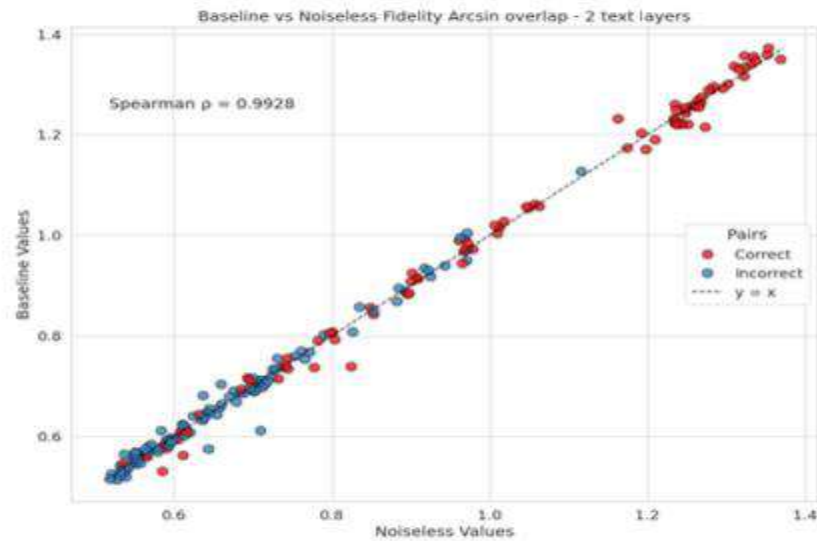


(b) IBM FakeMontrealV2 (Noisy)

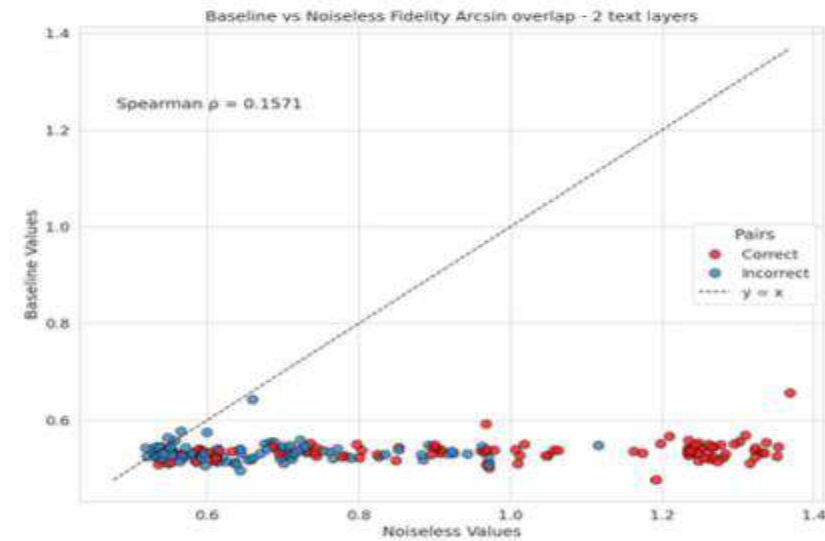


(c) IBM Marrakesh (Real Hardware)

Hardware Runs



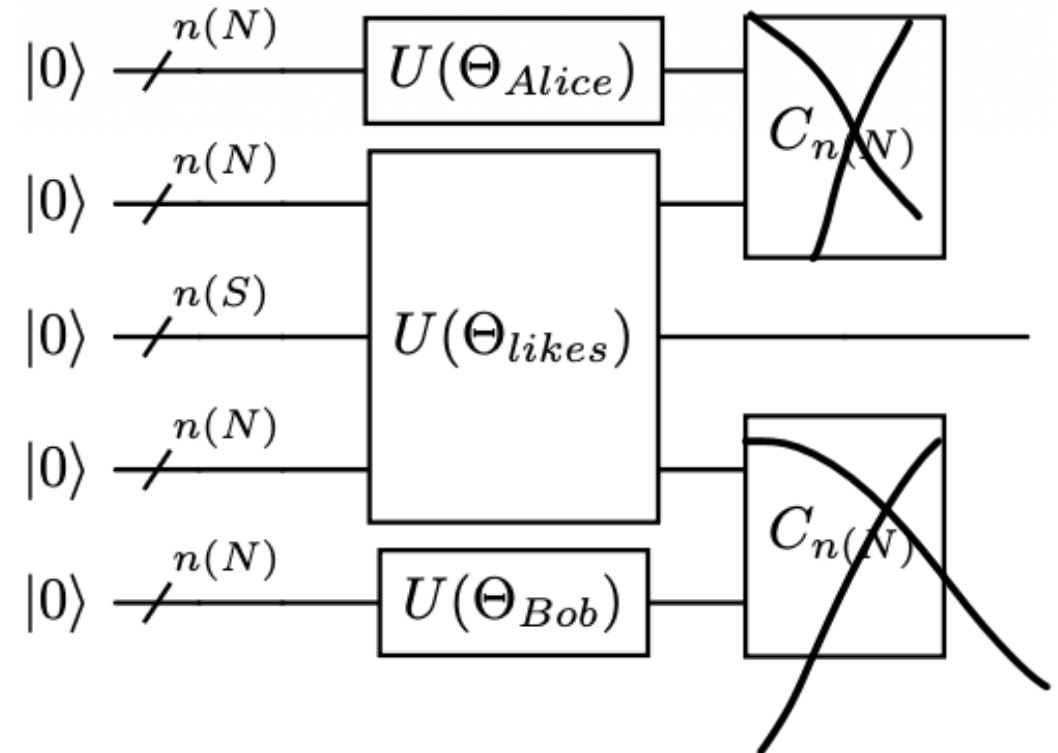
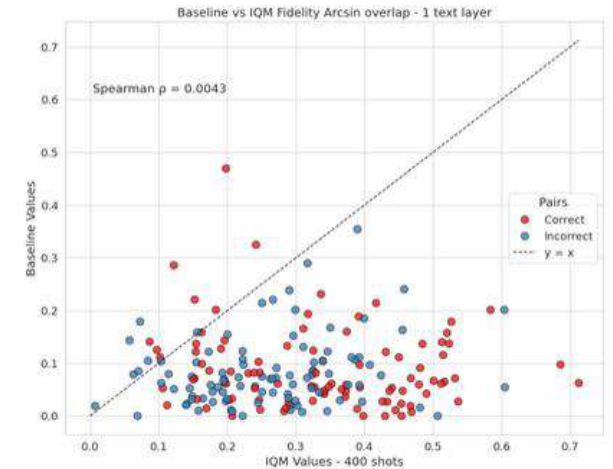
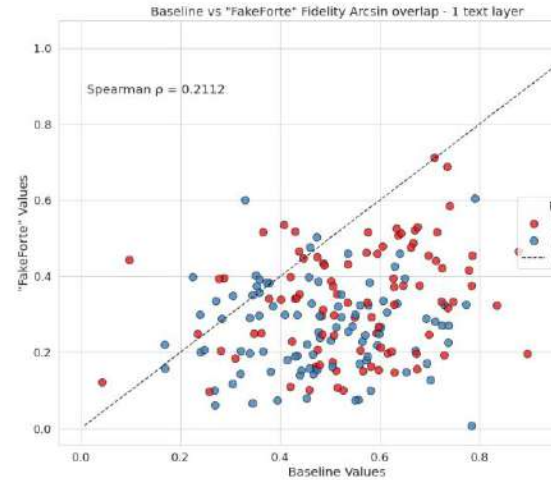
(a) Baseline vs Noiseless Values (2 text layers)



(b) Baseline vs IBM Kingston Values (2 text layers)

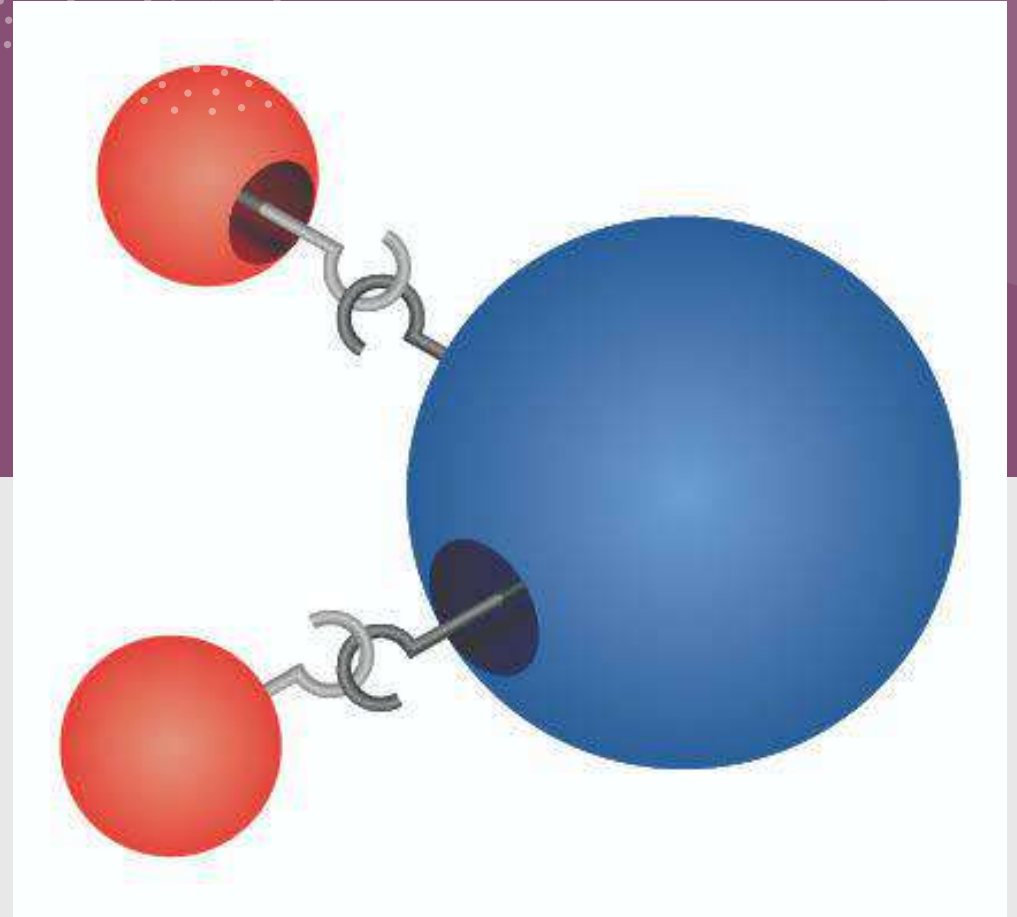
In Progress

- Why did the general purpose model not do well on HW?
- Tensor Contraction, Bell test, post-selection, lots of shots! £££
- Solution: replace with discard!
- Unknown theory.

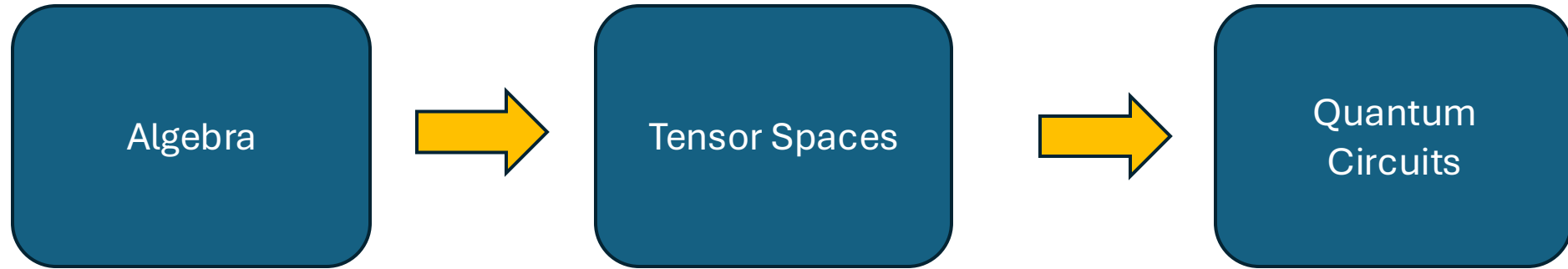


Future Directions

- Apply to quantum data:
 - Pregroups were based on valence theory
- Charles Sanders Peirce: atoms with hooks



Structure Informed Reasoning



$$(P, \cdot, 1, \leq, (-)^l, (-)^r)$$

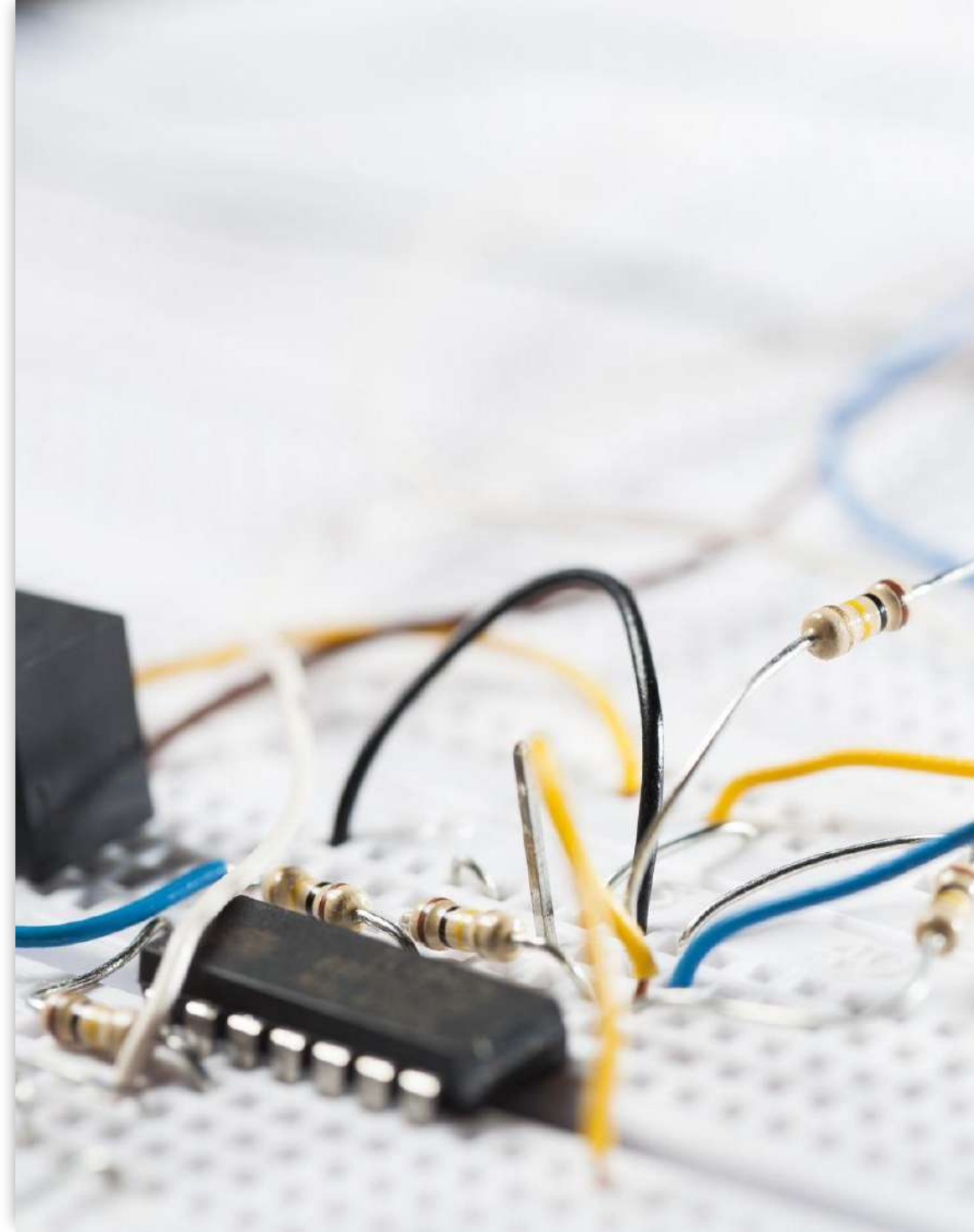
$$n \cdot (n^r \cdot s \cdot n^l) \cdot n \leq 1 \cdot s \cdot 1 \leq s$$

$$(Vect, \otimes, \mathbb{R}, V^*)$$

$$T_i^{\text{cube}} T_{ijk}^{\text{is left of}} T_k^{\text{cylinder}}$$

Quantum circuits are beneficial!

- 1- Can we do without the structure?
 - Yes, at the expenses of billions more parameters and lack of transparency
 - Transformers
- 2- Can we do without quantum circuits?
 - Yes, at the expense of millions more parameters and less speed
 - Tensor Networks
- 3- Can we do without tensor spaces?
 - Not sur!
 - Whenever there is a relation there is a tensor.



Thank you!

The IQM logo is displayed in white text on a dark, abstract background with green and blue light patterns.

IQM



IONQ



National Quantum
Computing Centre