# Learning the Shape of Information

Hirosi Ooguri

“Thermodynamic Approaches to Artificial Intelligence”
9 – 11 March 2026, DAIKIN AMICA KARUIZAWA

## Entanglement Entropies:

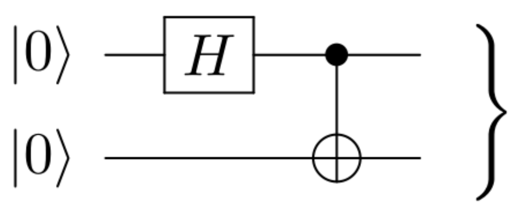
Entanglement entropy is a quantitative measure of quantum entanglement.

For $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_{\bar{A}}$, define $\rho_A = \mathrm{tr}_{\mathcal{H}_{\bar{A}}}[\,|\psi\rangle\langle\psi|\,]$.

Entanglement entropy: $S_\psi(A) = -\,\mathrm{tr}_{\mathcal{H}_A}\,[\,\rho_A \log(\rho_A)\,]$

In a certain precise sense,* the entanglement entropy $S_\psi(A)$ counts the **number of Einstein-Podolsky-Rosen (EPR) pairs** between $\mathcal{H}_A$ and $\mathcal{H}_{\bar{A}}$ in $|\psi\rangle$.

- Local Operation and Classical Communication




$$\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

# Entanglement Inequalities for General Quantum Systems:

$$S(A) = -\,\mathrm{tr}_{\mathcal{H}_A}\,[\,\rho_A \log(\rho_A)\,], \quad \rho_A = \mathrm{tr}_{\mathcal{H}_{\bar{A}}}[\,|\psi\rangle\langle\psi|\,]$$

Notation: $AB = A \cup B$

- Positivity: $S(A) \geq 0$
- Subadditivity: $S(A) + S(B) \geq S(AB)$
- Araki-Lieb: $S(A) + S(AB) \geq S(B)$
- Strong Subadditivity: $S(AB) + S(BC) \geq S(B) + S(ABC)$
- Weak Monotonicity: $S(AB) + S(BC) \geq S(A) + S(C)$

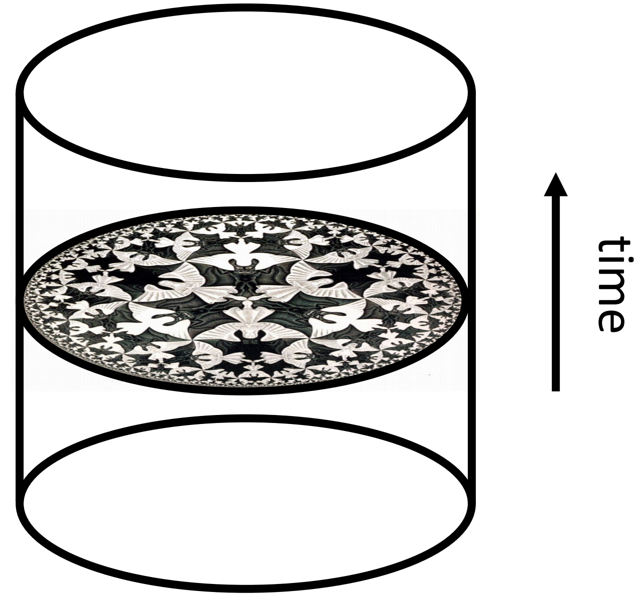
These are all the known inequalities, but we know there are more.
**Numerical evidence suggests that there are infinitely many inequalities.**

( For the classical Shannon entropy, it is known that the number of inequalities is infinite for more than 3 parties [Matus, 2007]. )

They quantify the asymptotic performance of information processing tasks.

# Holography of Quantum Gravity:

According to the **AdS/CFT correspondence**, quantum gravity in an asymptotically anti-de Sitter spacetime is equivalent/dual to a conformal field theory at the boundary.




Maldacena:hep-th/9711200

Quantum states described in terms of spacetime geometry in the dual gravitational theory have **special entanglement properties**.

These entanglement properties

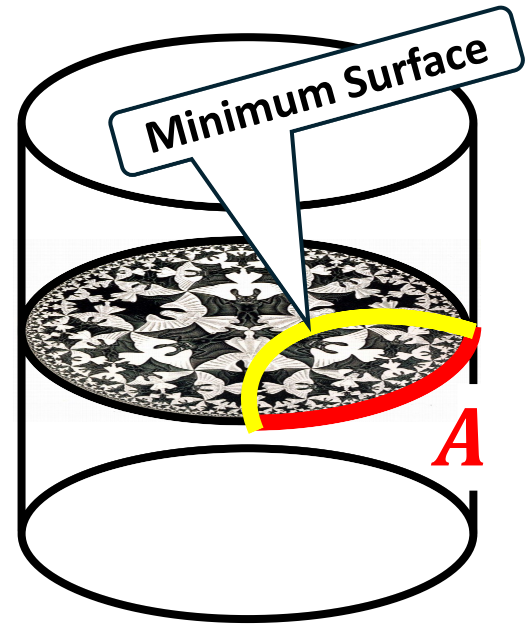
- are important for understanding the **microscopic mechanism** of the AdS/CFT correspondence,
- hold the key to solving **Hawking's black hole information paradox**,
- have applications in **quantum information theory**.

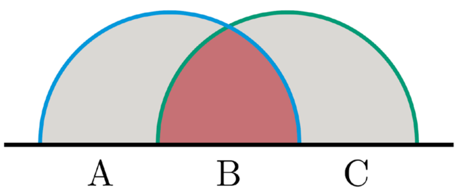
# Entanglement Entropies for Holographic Systems:

The Ryu-Takayanagi formula (2006)

$$S(A) = -\,\mathrm{tr}_{\mathcal{H}_A}\left[\,\rho_A \log(\rho_A)\,\right]$$

$$= \frac{1}{4G_N}\,\mathbf{Area}\left(\text{Minimum Surface Subtending } A\right)$$




Holographic proof of strong subadditivity

$$S(AB) + S(BC) \geq S(B) + S(ABC)$$

New holographic inequality: monogamy of mutual information

$$S(AB) + S(BC) + S(CA) \geq S(A) + S(B) + S(C) + S(ABC)$$

Hayden, Headrick, Maloney: 1107.2940

Laudauer:

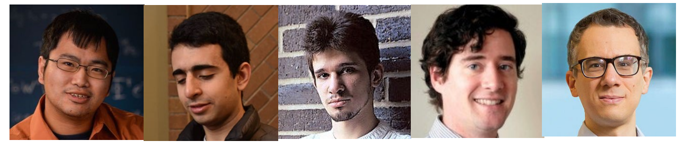
Information $\Longleftrightarrow$ Heat

Ryu-Takayanagi:

Information $\Longleftrightarrow$ Geometry

Eleven years ago, we launched a program to classify holographic entropy inequalities.

Bao, Nezami, Stoica, Sully, Walter + H.O.: 1505.07839



**We found:**

- **Finite combinatorial algorithm** to enumerate and classify all the holographic entanglement entropy inequalities
- There are only **finitely many** independent inequalities for a fixed number of parties
- **Complete classification** for 2, 3, and 4 parties
- **New families** of inequalities for 5 parties (later found to be complete)

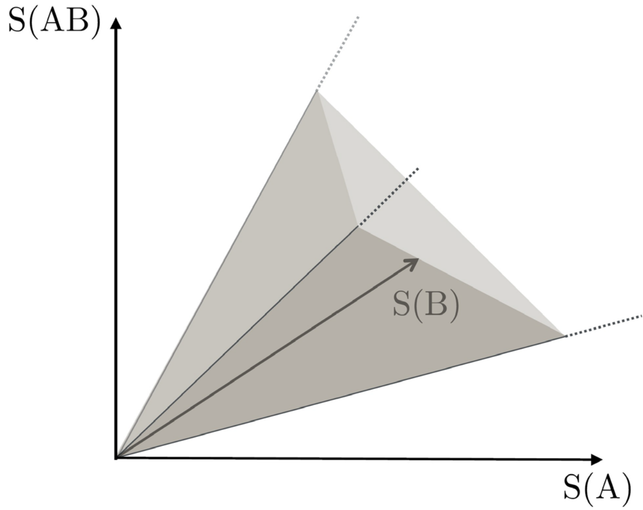
They contrast with **general quantum systems**, for which **infinitely many entropy inequalities** are expected, making the classification difficult.

# Holographic Entropy Cone

The entanglement entropies of $n$ parties $\{A_1, \ldots, A_n\}$ form a $(2^n - 1)$-dimensional **vector space.**

*E.g.*, $(S(A), S(B), S(AB))$ for $n = 2$.

- The entropy vector can be scaled by rescaling the bulk manifold $X$.
- To show the space is closed under addition, consider the disjoint union of two bulk manifolds $X \cup Y$.




The set of all holographic entropy vectors by varying parties $A_1, \ldots, A_n$ and the bulk manifold $X$ form a **convex cone**, called the holographic entropy cone. Every point in the entropy cone corresponds to some bulk geometry.

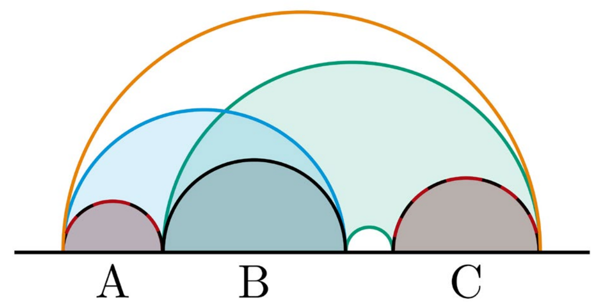
Facets $\Longleftrightarrow$ Inequalities

# Holographic Entropy Cone is Rational Polyhedral

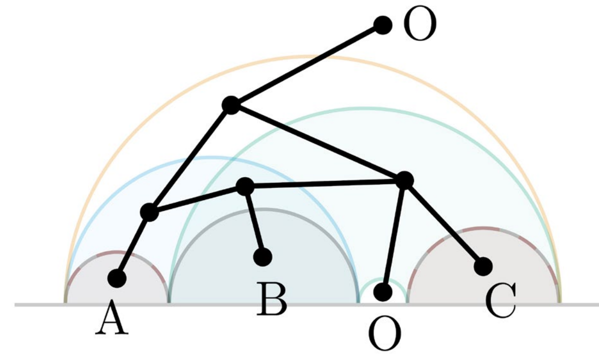
*i.e.*, there are only finite number of independent entropy inequalities.

To prove this, we reduce the Ryu-Takayanagi computation to a **combinatorial problem:**

- For each union $I$ $(= 1, \dots, 2^n - 1)$ of non-empty subsets of $\{A_1, \dots, A_n\}$, we consider the Ryu-Takayanagi surface $RT_I$. They cut the bulk into a **finite number of connected pieces**.



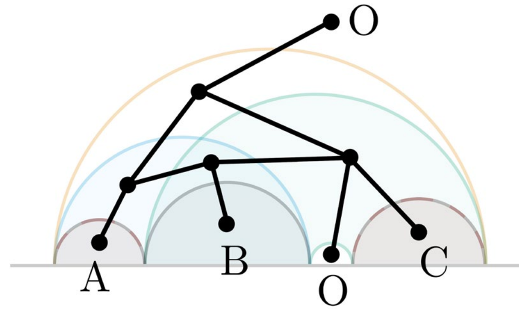

- **Graph model of the geometry:** Place a vertex in each connected piece and an edge connecting pieces separated by a segment of RT surfaces. Each edge is weighted by the area of the segment of the RT surface.



# Holographic Entropy Cone is Rational Polyhedral

*i.e.*, there are only finite number of independent entropy inequalities.

**Min-Cut:** Consider a cut separating the graph model into two, one containing all boundary vertices on $I$ and another containing $\bar{I}$. The discrete entropy is defined by minimizing the sum of edge weights.




**Discrete entropy = Ryu-Takayanagi entropy**

**Conversely,** for every graph model, one can construct a bulk geometry and boundary regions such that Ryu-Takayanagi entropies = discrete entropies.

**Universal graph model:** There is a set of graph transformations that preserves the combinatorial calculation of the holographic entropy. Using them, we can convert any holographic graph into a universal graph with $2^{2^n-1}$ vertices. Therefore, **determining the holographic entropy cone is a <u>finite combinatorial problem</u>.**

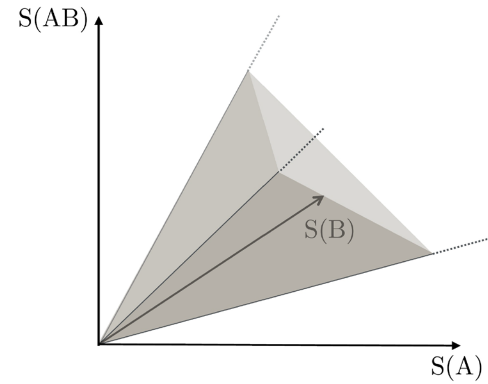
**It is an exact reduction, not an approximation.**

# Q: How do we know that we found all of the inequalities?

# A: Identify the **extreme rays**.

A rational convex polyhedron is a positive hull of a finite number of rays.

**Extreme rays = Rays that cannot be written as a proper convex combination of other elements of the cone.**




Once we have a candidate cone, we identify its extreme rays. For each ray, try to find a universal graph model such that the associated entropy vector lies on the ray. If we can do this for all the extreme rays, then the candidate cone is the true holographic entropy cone.

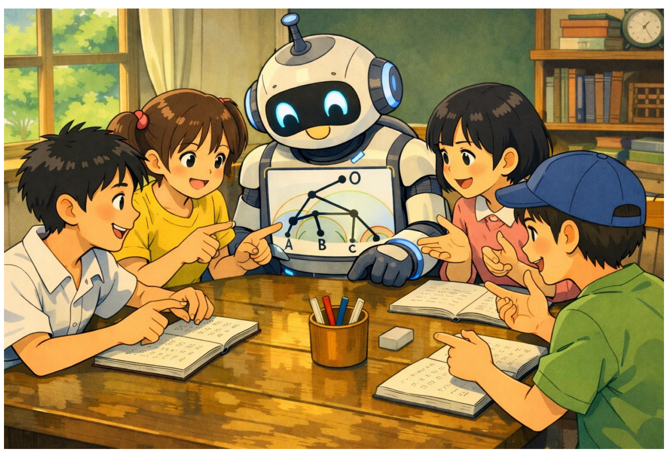
We carried out this program and **classified holographic entropy inequalities for $2, 3,$ and $4$ parties.**

**For $5$ parties, we found new inequalities.** It was subsequently shown that **they complete the classification for $5$ parties.** Hernández-Cuenca:1903.09148

We developed a greedy algorithm to implement “proofs by contraction.” However, classifying inequalities with $n \geq 6$ parties is found to be intractable, even with the aid of computers.

ChatGPT 5.2

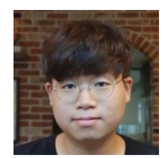
For 6 parties, the space of entropy vectors is $2^6 - 1 = 63$ dimensional, and the number of vertices in the universal graph model is $2^{2^6-1} \sim 10^{19}$.



ChatGPT 5.2

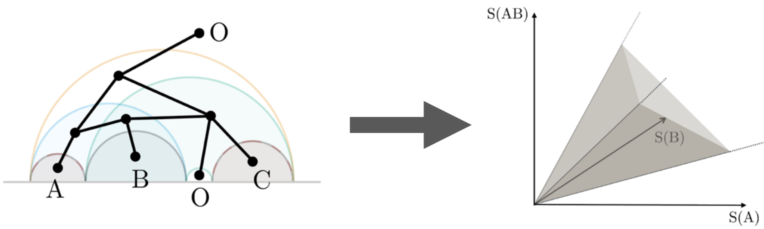
**Thus, we decided to ask AI for help.**

He, Lee + H.O.: 2601.19979

In addition, Claude accelerated our ability to test various hypotheses and iterate on experimental designs.

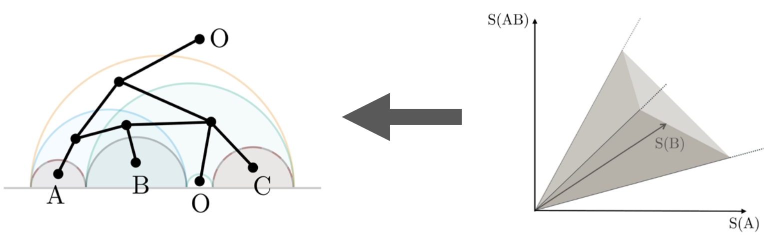
# Holographic Entropy Cone




Each weighted graph representing the spacetime geometry can be mapped to a vector inside of the holographic entropy cone by min-cuts.

## Can Reinforcement Learning (RL) solve the inverse problem?

Given a target entropy vector $\vec{S}_{\text{target}}$, our trained policy model attempts to find a weighted graph $G$ whose min-cut entropy vector $\vec{S}_G$ matches it.

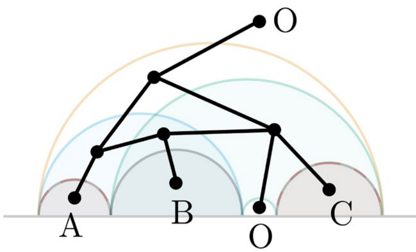


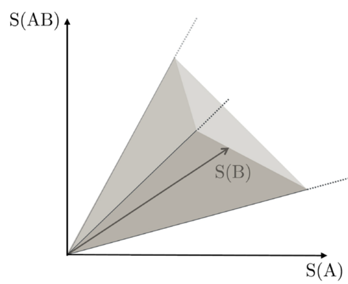

**Reward: cosine similarity**

$$R = \frac{\vec{S}_G \cdot \vec{S}_{\text{target}}}{\left\| \vec{S}_G \right\| \left\| \vec{S}_{\text{target}} \right\|}$$

# The policy model learns to take the right action for a given state.

**State:** The current weighted graph $G$
**Action:** Increase/decrease the weight on each edge
**Reward:** Cosine similarity $R$



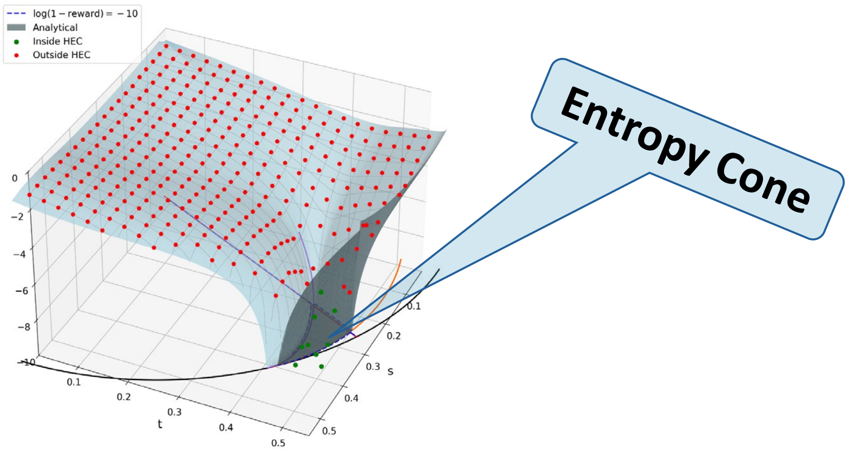



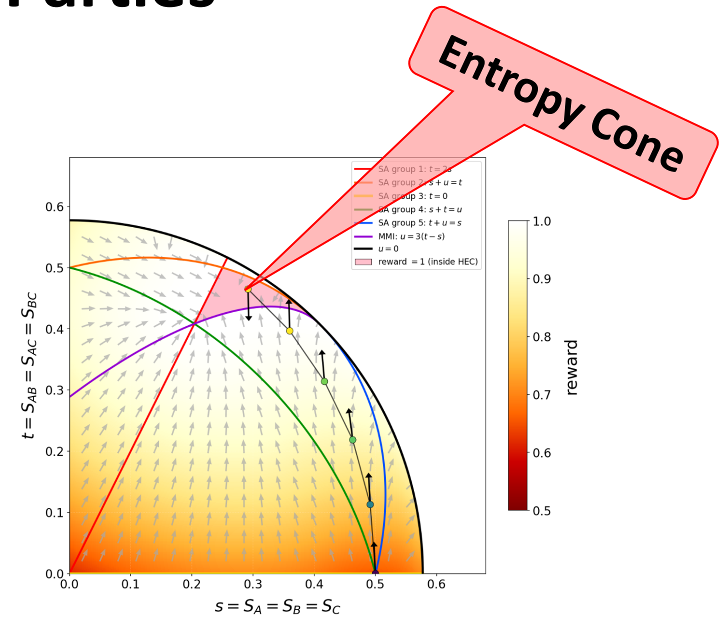

$$R = \frac{\vec{S}_G \cdot \vec{S}_{\text{target}}}{\left\|\vec{S}_G\right\| \left\|\vec{S}_{\text{target}}\right\|}$$

If $R = 1,\ \vec{S}_{\text{target}}$ is inside of the cone

We use the cosine similarity since the problem is scale-invariant.

# Proof of Concept at 3 Parties

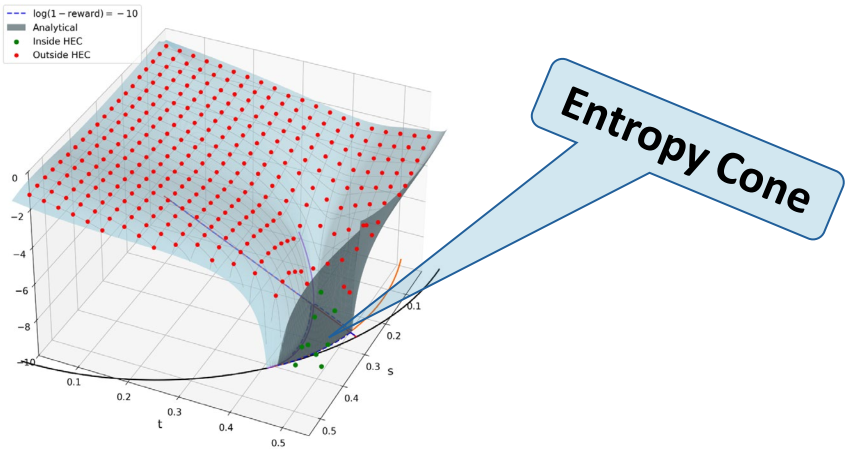


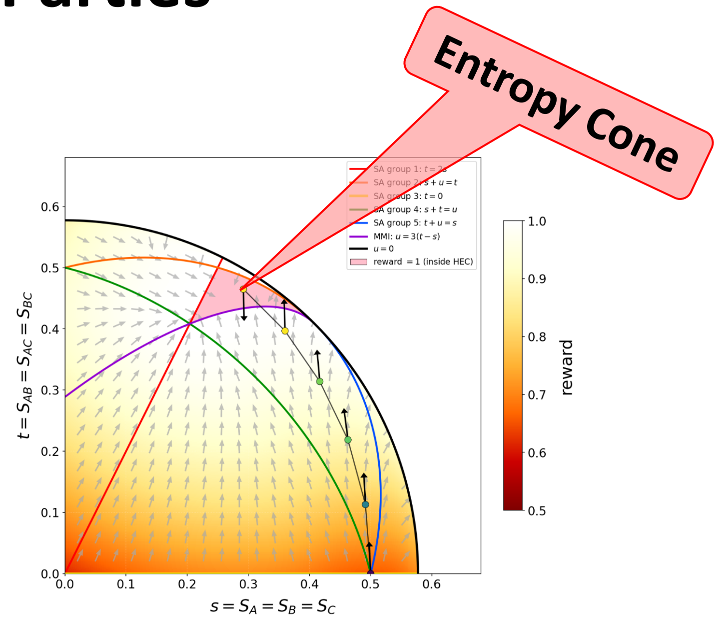

**Min-Cut entropy is not smooth.** RL is particularly useful for exploring high-dimensional, non-differentiable situations like this, where gradients are unavailable or unstable.

This enables us to find the optimal graph, even when $\vec{S}_{\mathrm{target}}$ lies outside the entropy cone, yielding a **navigator function, which flows toward the cone boundary and reveals unknown facets.**

# Proof of Concept at 3 Parties




Reward landscape on the 3-party symmetric slice, showing **agreement between analytical predictions (blue surface) and RL results (red points).**




Trajectory of the entropy vector along a gradient flow overlaid on the analytical reward landscape.

**The trained policy model re-discovered the monogamy of mutual information** (purple curve on the right panel).

# New Results at 6 Parties

Double-exponential growth of the number of inequalities:

- At $3$ and $4$ parties, the only new inequality is the monogamy of mutual information.
- At $5$ parties, the $5$ new families of inequalities found in our 2015 paper complete the set.
- **At 6 parties, $\mathbf{1800}$ new families of inequalities have been found so far.**

**Our trained policy model suggests that there are more inequalities at 6 parties.**

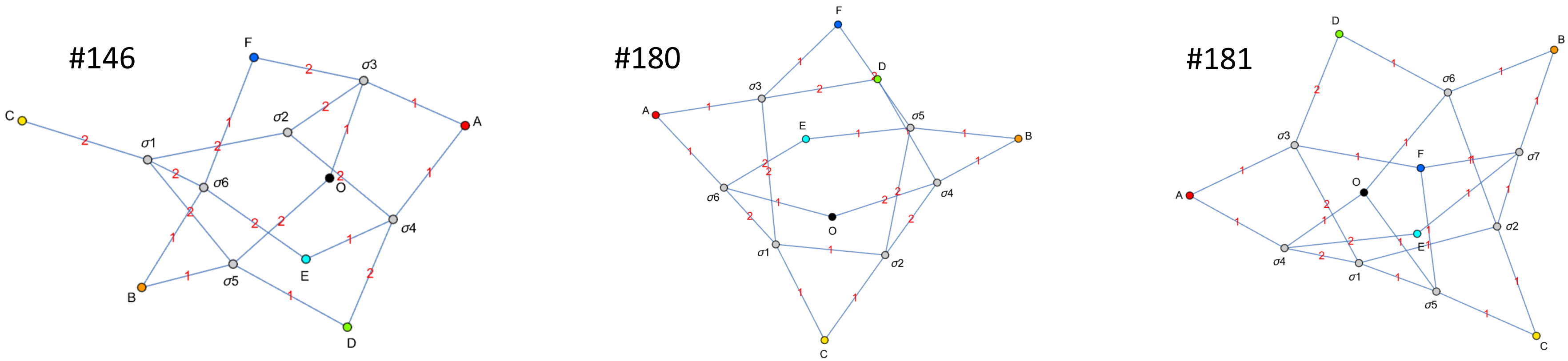
At 6 parties, there are $\mathbf{156}$ $\mathcal{S}_6$-orbits of candidate extreme rays that are genuinely new and **satisfy all known inequalities**.

Of these, **$\mathbf{150}$ have found graph realizations** and are therefore in the holographic entropy cone.

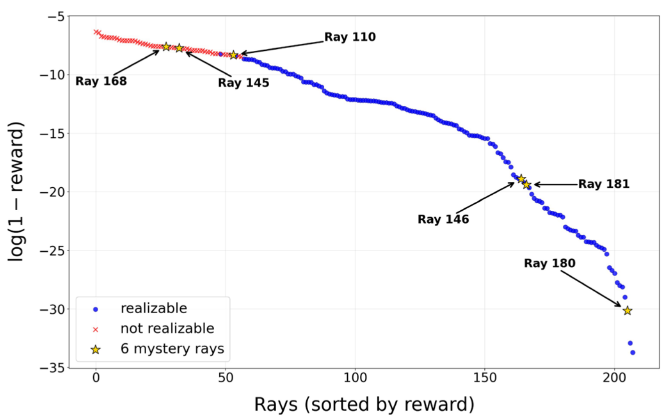
He, Hubeny, Rota: 2412.15364

However, **it was not known whether the remaining 6 orbits have graph realizations.**

# The trained policy model identified graph realizations for 3 of the 6 orbits.




## Once the policy model suggests a graph, we can verify it analytically.




The other 3 orbits (#110, #145, #168) have maximum rewards $R < 1$ and are surrounded by orbits that are known not to have graph realizations. Therefore, it is likely that they also do not have graph realizations.

These results imply the **presence of new inequalities that exclude #110, #145, and #168** from the holographic entropy cone.

# Claude

We did not use large language models in the search algorithm. However, we collaborated extensively with coding agents throughout the research process.

Claude Code was instrumental in **building our codebase and running comprehensive experiments**, such as the parameter sweeps that ensure our RL framework is exhaustive enough while remaining computationally reasonable.

**Claude designed a sophisticated gradient flow algorithm** that stabilizes the process of finding new facets from the reward landscape. **We verified graph realizations analytically.**

# Information $\Longleftrightarrow$ Geometry

- RL can act as a discovery engine for precise, verifiable structures in quantum gravity.
- We used a vanilla policy gradient algorithm. More sophisticated RL approaches could yield substantial improvements.
- We focused on static properties of quantum gravity. Some of our results have been generalized to dynamical time-dependent situations.

# Information $\Longleftrightarrow$ Geometry

- RL can act as a discovery engine for precise, verifiable structures in quantum gravity.
- We used a vanilla policy gradient algorithm. More sophisticated RL approaches could yield substantial improvements.
- We focused on static properties of quantum gravity. Some of our results have been generalized to dynamical time-dependent situations.