Neural Information Causality

A Random-Access Capacity Law for Representation Learning

Jeongho Bang[1,2] and Marcin Pawłowski[3,4]

[1] Institute for Convergence Research and Education in Advanced Technology (iCREAT), Yonsei University
[2] Department of Quantum Information, Yonsei University
[3] International Centre for Theory of Quantum Technologies (ICTQT), University of Gdańsk
[4] Institute of Theoretical Physics and Astrophysics, Faculty of Mathematics, Physics and Informatics, University of Gdańsk

M. Pawłowski et al., Nature 461, 1101 (2009)
J.Bang and M. Pawłowski, arXiv:2605.09316 (2026)

Core thesis

A query-separated representation is operationally a message.

- The encoder commits to a data-dependent interface before the query is known.
- The decoder later answers using only that interface, the query, and allowed side resources.
- Therefore random-access information must be bounded by independently certified capacity.

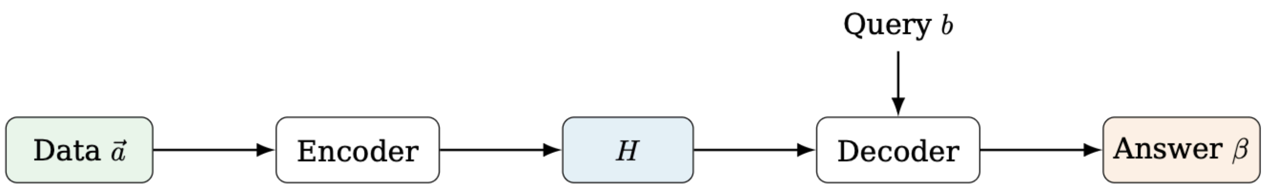
$$I_{\text{N-RAC}} \leq I(\vec{a} : H, B) \leq C_H$$

Why this question matters

Modern learning systems often rely on internal interfaces rather than direct data access.

- Memory-augmented networks and differentiable memory.
- Retrieval-augmented generation and query-time retrieval.
- Encoder–decoder pipelines with a hidden representation.
- Quantum- or correlation-assisted modules used inside learning architectures.

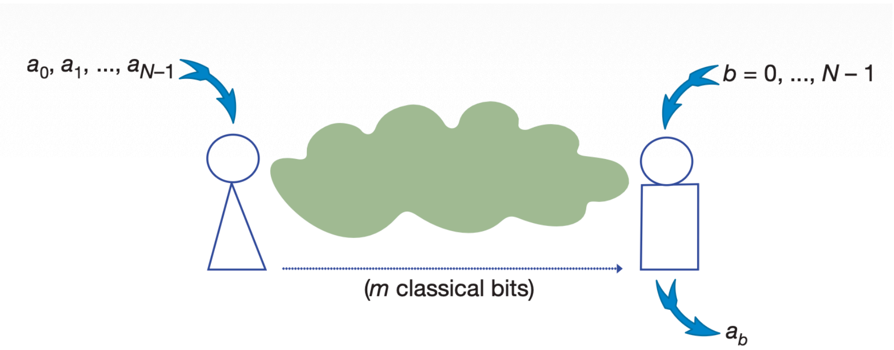
The diagnostic question: does the representation carry only its counted capacity?



*Query-separated structure

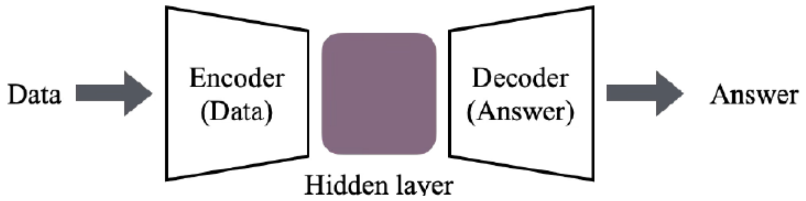
The shared causal skeleton

(a) Information Causality



M. Pawłowski et al., Nature 461, 1101 (2009)

(b) Deep Learning: Representation Learning



J.Bang and M. Pawłowski, arXiv:2605.09316 (2026)

- IC/RAC: Alice compresses data into an m-bit message.
- Representation learning: the hidden layer plays the same causal role.

M. Pawłowski et al., Nature 461, 1101 (2009)

Classical information causality: RAC form

- Alice receives N independent unbiased bits:

$$\vec{a} = (a_0, \ldots, a_{N-1}).$$

- Bob receives a uniformly random index b.
- Alice sends $x \in \{0,1\}^m$.
- Bob outputs β as a guess for a_b.

$$P_K = \Pr[\beta = a_K \mid b = K]$$

$$I_{\mathrm{RAC}} = \sum_{K=0}^{N-1} I(a_K : \beta \mid b = K)$$

Information causality: $I_{\mathrm{RAC}} \leq m$.

M. Pawłowski et al., Nature 461, 1101 (2009)

Accuracy is not the score, but it certifies score

The observable information score uses only classical variables $(\vec{a}, b, \beta)$:

$$I_{\mathrm{RAC}} = \sum_K I(a_K : \beta \mid b = K).$$

For unbiased target bits, binary estimation gives

$$I(a_K : \beta \mid b = K) \geq 1 - h(P_K).$$

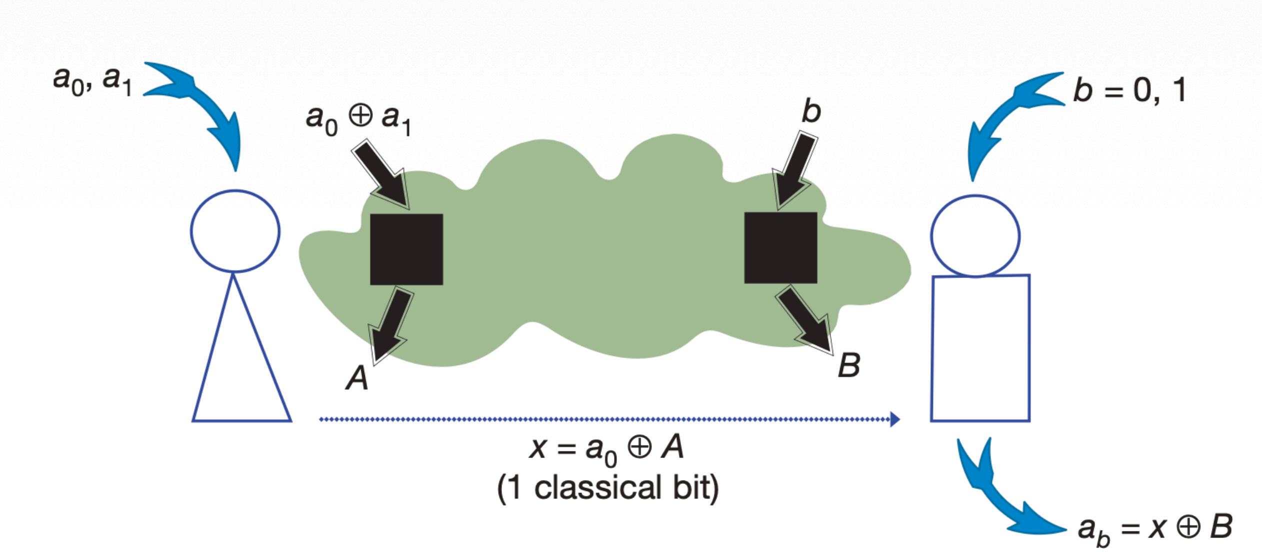
Therefore

$$I_{\mathrm{RAC}} \geq N - \sum_{K=0}^{N-1} h(P_K), \qquad I_{\mathrm{RAC}} \geq N[1 - h(P)] \text{ if } P_K \equiv P.$$

This converts random-access accuracy into an information-accounting test.

M. Pawłowski et al., Nature 461, 1101 (2009)

The seed case: $N = 2$, $m = 1$, and CHSH

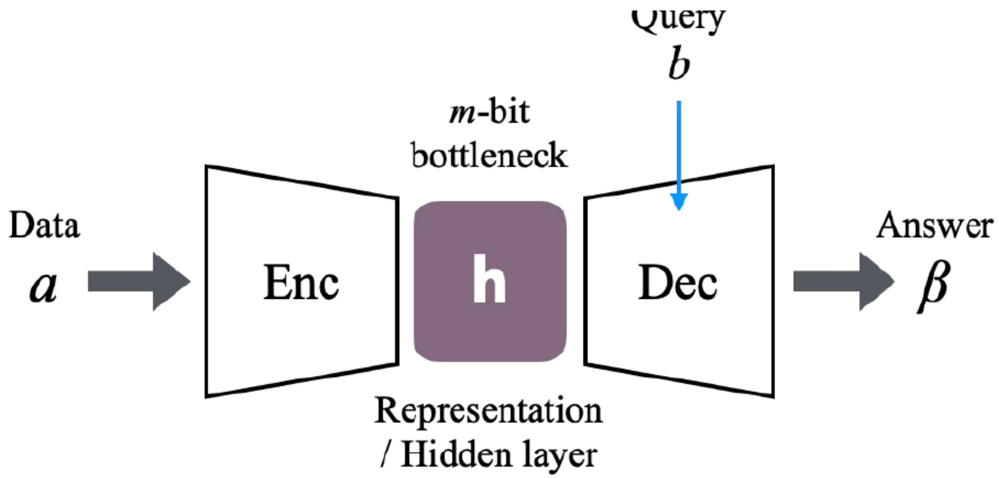


- Alice uses $s = a_0 \oplus a_1$.
- Bob uses $t = b$.
- The correlation cell outputs (A, B).
- Message and answer:

$$x = a_0 \oplus A, \quad \beta = x \oplus B.$$

Correct retrieval is exactly the CHSH winning event.

Neural-RAC: the query-separated primitive



- The encoder is query-blind.
- The decoder sees the query only after H is fixed.
- W is fixed before the fresh episode.
- H is the only data-dependent causal carrier.

$$H \leftarrow \mathrm{Enc}(\vec{a}; R_A, W), \qquad \beta \leftarrow \mathrm{Dec}(H, b; R_B, W).$$

Capacity must be certified independently

The effective bottleneck capacity is not inferred from success after the fact:

$$C_H(H) = \sup I(\vec{a} : H \mid B).$$

Useful certificates include:

- finite alphabet: $C_H \leq \log |\mathcal{H}|$;
- d coordinates quantized to q bits: $C_H \leq dq$;
- AWGN interface: $C_H \leq \frac{d}{2} \log(1 + P/\sigma^2)$;
- hard classical bottleneck: $H \in \{0,1\}^m$, so $C_H \leq m$.

Unconstrained ideal real values do not imply a finite capacity law.

Main theorem: embedding and capacity

Every query-separated Neural-RAC architecture induces a random-access communication experiment.

$$I_{\mathrm{N\text{-}RAC}} \leq I(\vec{a} : H, B)$$

If an independent interface certificate gives $I(\vec{a} : H, B) \leq C_H$, then

$$I_{\mathrm{N\text{-}RAC}} \leq C_H.$$

For a hard m-bit bottleneck:

$$I_{\mathrm{N\text{-}RAC}} \leq m.$$

Neural Information Causality

J.Bang and M. Pawłowski, arXiv:2605.09316 (2026)

Neural-IC as a capacity audit

The theorem becomes an evaluation protocol:

1. Fix all parameters W before the episode.
2. Sample a fresh database $\vec{a}$ independently of W and the resource.
3. Encode H before observing the query b.
4. Decode β using only (H, b) and allowed side resources.
5. Estimate $I_{\text{N-RAC}}$ and compare with the certified C_H.

If $I_{\text{N-RAC}} > C_H$, the model has an uncounted side channel or an invalid capacity certificate.

Capacity–accuracy tradeoff

Combining the accuracy lower bound with Neural-IC gives

Entropy defects $$N - \sum_{K=0}^{N-1} h(P_K) \leq C_H.$$ → a sanity check

For query-symmetric access:

$$C_H \geq N[1 - h(P)].$$

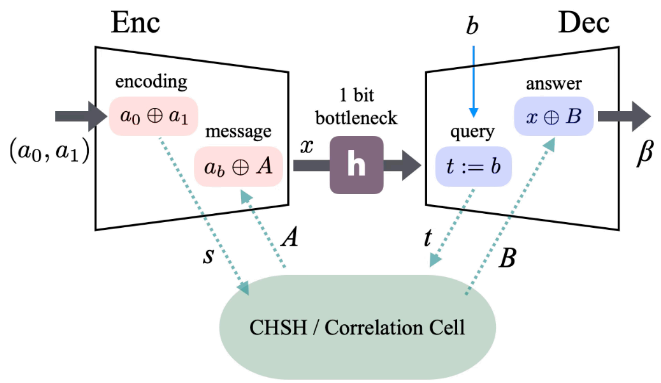
A tiny representation with uniformly high random-access accuracy implies one of four diagnoses:

- query leakage into the encoder;
- finite-precision leakage through real-valued interfaces;
- episode-specific weights or memory;
- uncounted nonclassical or post-quantum resources.

A CHSH correlation cell as a neural resource layer

The shared resource is modeled as independent no-signaling CHSH cells:

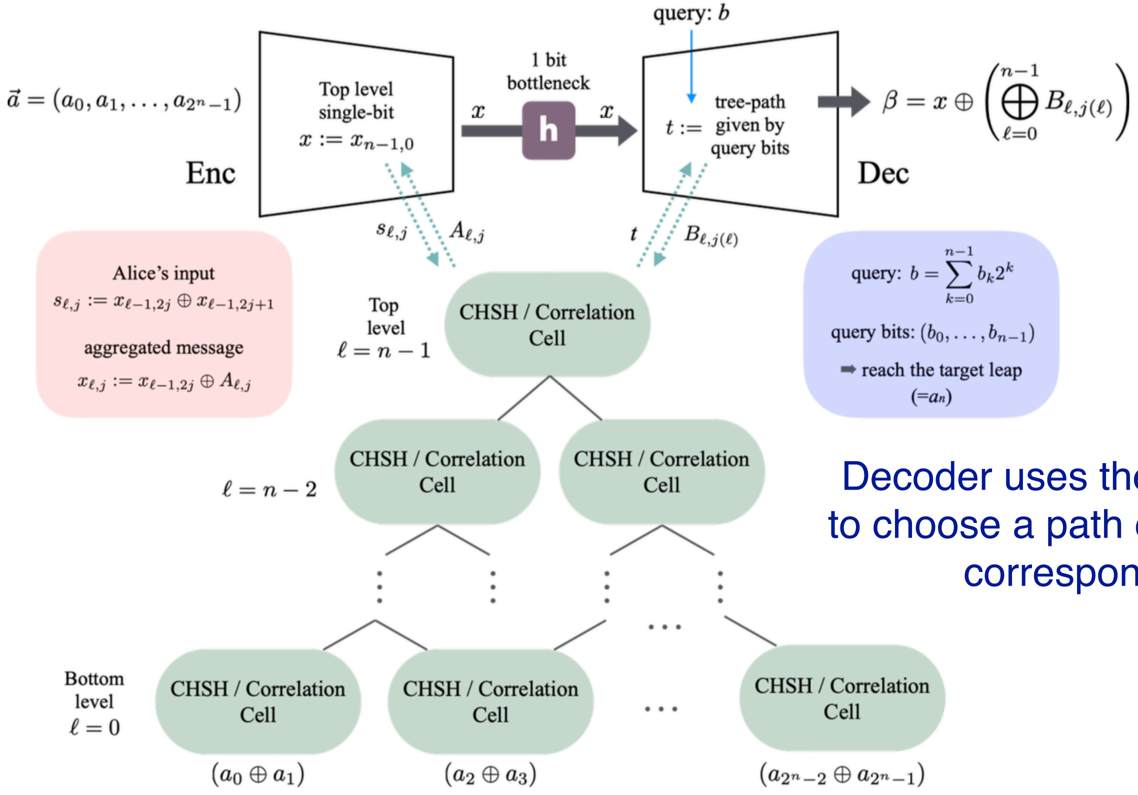
$$\Pr[A \oplus B = st \mid s, t] = \frac{1+E}{2}.$$



- E is the correlation bias.
- Classical local resources satisfy $E \leq 1/2$.
- Quantum theory permits $E \leq 1/\sqrt{2}$.
- Post-quantum no-signaling boxes could have stronger E.

What happens when such a cell is used repeatedly inside a query-separated architecture.

Nesting: a depth-n pyramid for $N = 2^n$



Higher levels recursively combine those messages

pairs of database bits are combined into intermediate messages

Decoder uses the binary expansion of the query to choose a path down the tree and combines the corresponding correlation outputs.

Alice aggregates to one bit x; Bob follows the query path through independent correlation cells.

Bias multiplication is the amplification mechanism

Each selected CHSH cell contributes an independent error bit. If local correctness bias is E, the parity of n errors has bias E^n.

*bias multiplication

$$P_K = \Pr[\beta = a_K \mid b = K] = \frac{1 + E^n}{2}$$

*Along a fixed query path, each CHSH cell contributes an independent error bit.

For $N = 2^n$:

$$I_{\text{N-RAC}}(n, E) = 2^n \left[1 - h\left(\frac{1 + E^n}{2}\right)\right].$$

Small single-cell advantages can become a depth-dependent random-access amplifier.

Tsirelson threshold from one-bit stability

The entropy deficit obeys

$$1 - h\left(\frac{1+\delta}{2}\right) \geq \frac{\delta^2}{2\ln 2}.$$

*quadratic lower bound

With $\delta = E^n$ and $N = 2^n$:

$$I_{\text{N-RAC}} \geq \frac{(2E^2)^n}{2\ln 2}.$$

- If $E > 1/\sqrt{2}$, then $2E^2 > 1$ and the score eventually exceeds one bit.
- Stability of a one-bit bottleneck for arbitrary depth therefore requires $E \leq 1/\sqrt{2}$.

Tsirelson's value appears as the stability frontier for representation-mediated random access.

Beyond isotropy: asymmetric branch biases

If the seed cell has different branch biases E_0 and E_1, a query path $b = (b_1, \ldots, b_n)$ gives

$$P_b = \frac{1 + \prod_{\ell=1}^{n} E_{b_\ell}}{2}.$$

The exact score for binary-symmetric branches is

$$\sum_{b \in \{0,1\}^n} \left[1 - h\left(\frac{1 + \prod_\ell E_{b_\ell}}{2}\right)\right].$$

A sufficient amplification criterion is

$$E_0^2 + E_1^2 > 1.$$

This keeps task-specific asymmetry visible instead of collapsing everything into one scalar too early.

Why quantum resources obey the accounting law

Quantum mutual information satisfies the information-calculus assumptions needed in the theorem:

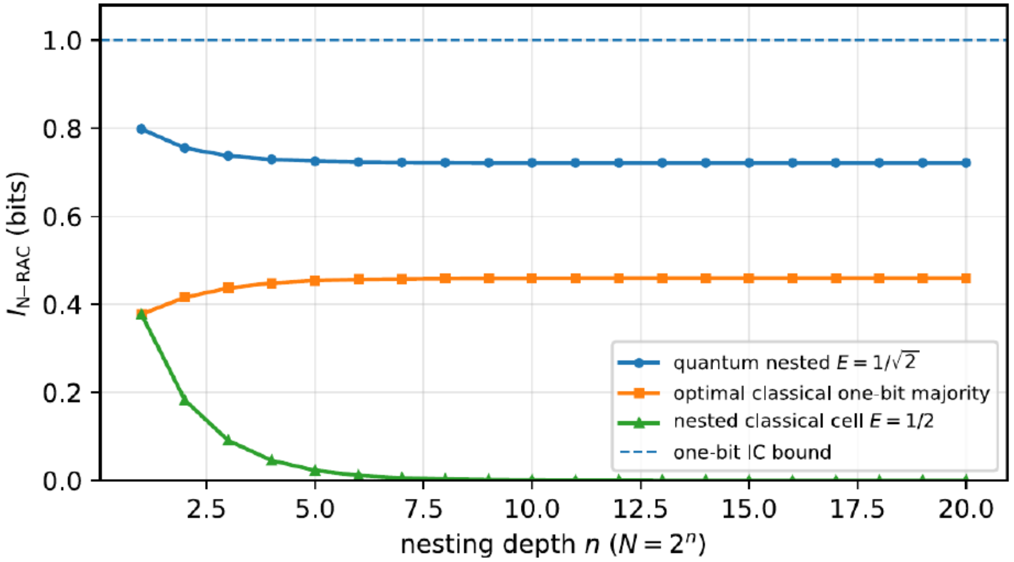
- consistency with Shannon mutual information on classical registers;
- product-state normalization;
- data processing under local CPTP maps;
- chain rule and nonnegative conditional mutual information;
- classical self-conditioning.

Therefore, quantum-assisted Neural-RAC protocols obey

$$I_{\mathrm{N\text{-}RAC}} \leq m$$

for a hard classical m-bit bottleneck, and $I_{\mathrm{N\text{-}RAC}} \leq C_H$ for certified interfaces.

Classical one-bit fair-access benchmark

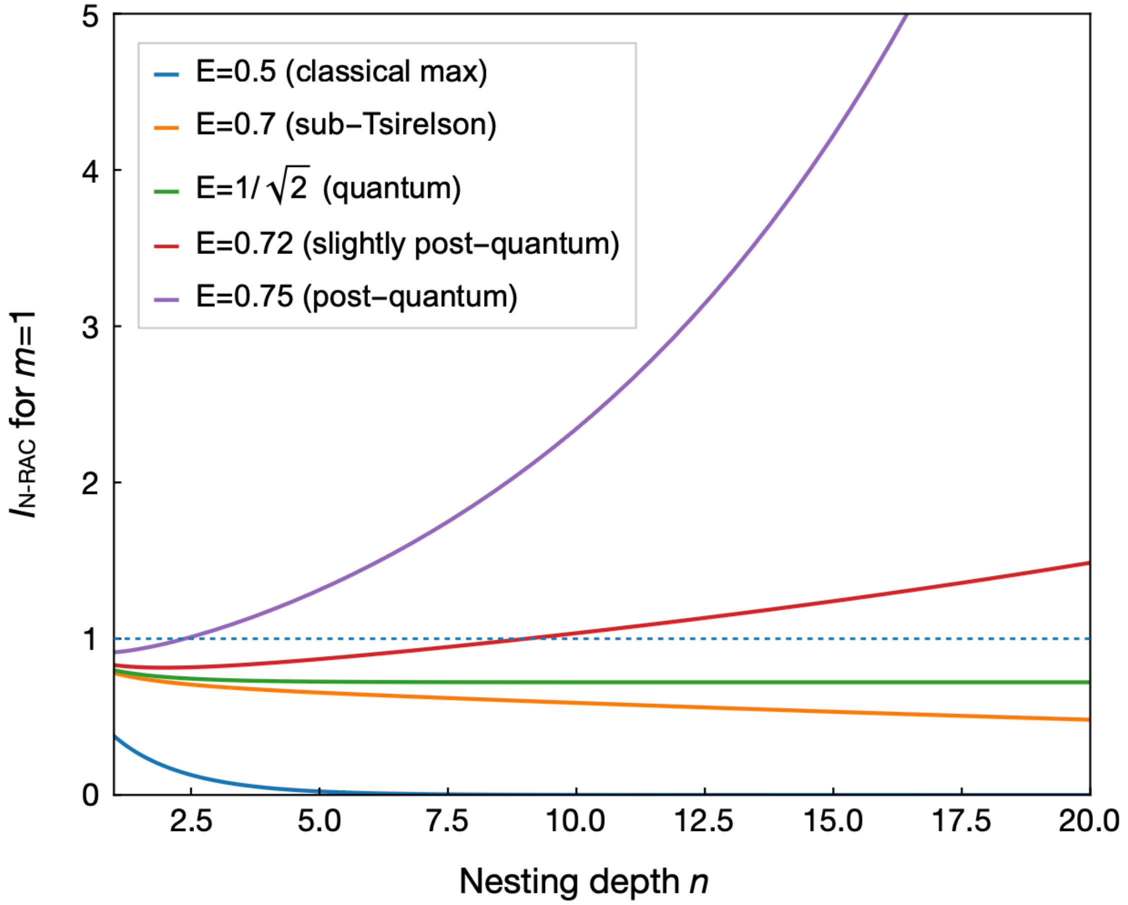


- One classical bit can saturate total information by storing one target bit.
- Fair access asks for average performance over all queries.

$$P^{\mathrm{cl}}_{\mathrm{avg}}(N,1) = \frac{1}{2} + 2^{-N}\binom{N-1}{\lfloor (N-1)/2 \rfloor}$$

Quantum correlations improve fair access without exceeding the one-bit information budget.

Closed-form stability probe: score versus depth

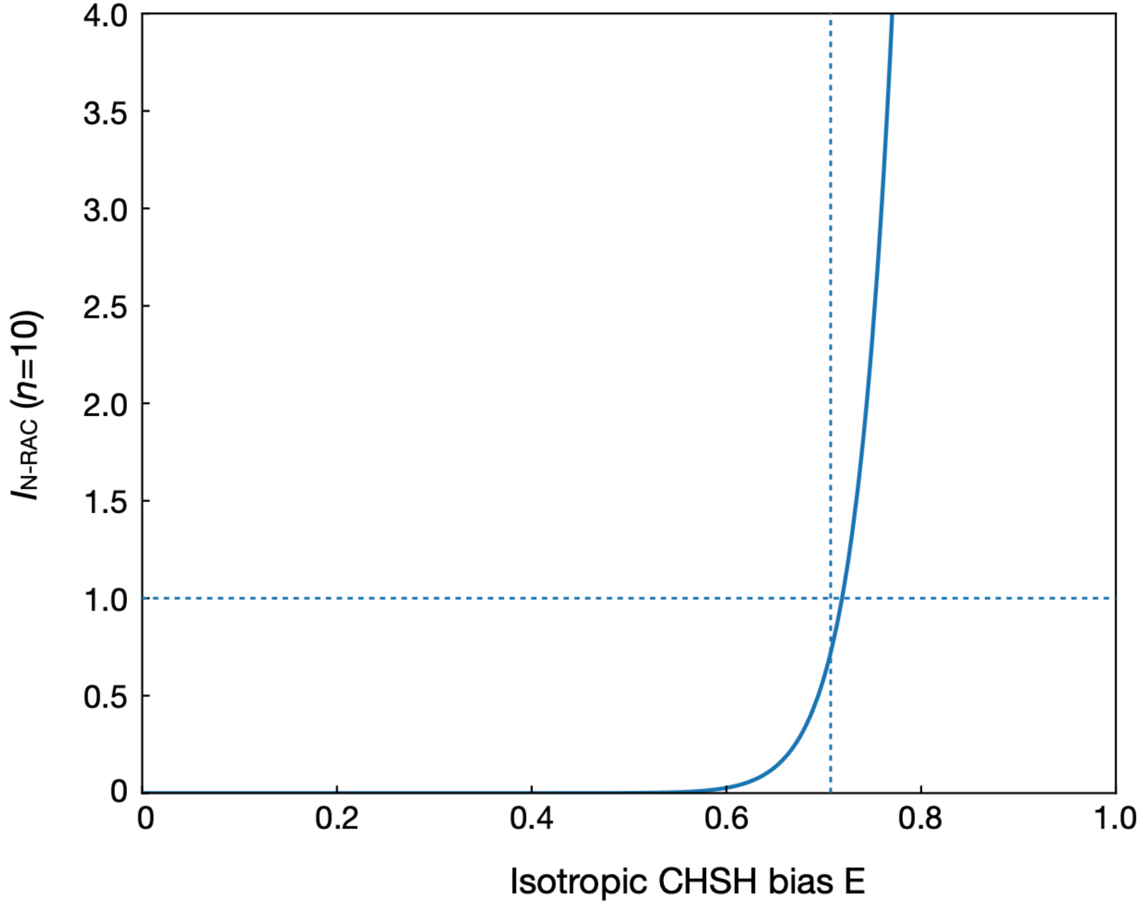


- Sub-Tsirelson curves remain stable.
- The quantum curve approaches a constant below one.
- Supercritical curves eventually cross the one-bit bound.

$$\lim_{n\to\infty} I_{\text{N-RAC}}\left(n, \frac{1}{\sqrt{2}}\right) = \frac{1}{2\ln 2}$$

$$\frac{1}{2\ln 2} \simeq 0.721.$$

Bias scan: shallow tests can hide supercriticality



At fixed depth $n = 10$, the finite-depth crossing is above the asymptotic threshold.

n	0.5	0.7	$1/\sqrt{2}$	0.72	0.75
1	.377	.780	.798	.832	.913
5	.023	.655	.725	.870	1.312
10	.001	.589	.721	1.036	2.344
20	.000	.482	.721	1.486	7.607

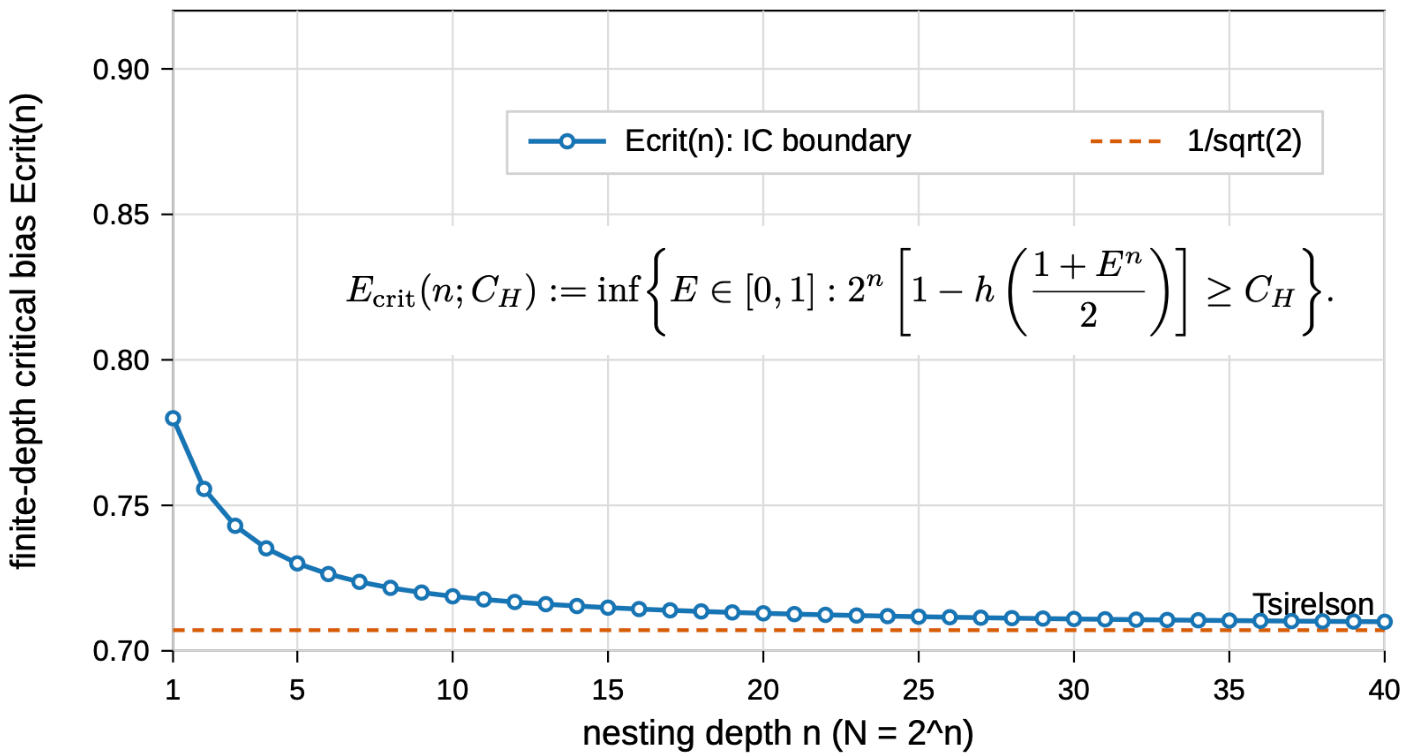
Values are $I_{\text{N-RAC}}(n, E)$; the one-bit limit is 1.

The vertical dashed line: Tsirelson value

The horizontal dashed line: Neural IC limit

The finite-depth tests have their own threshold.

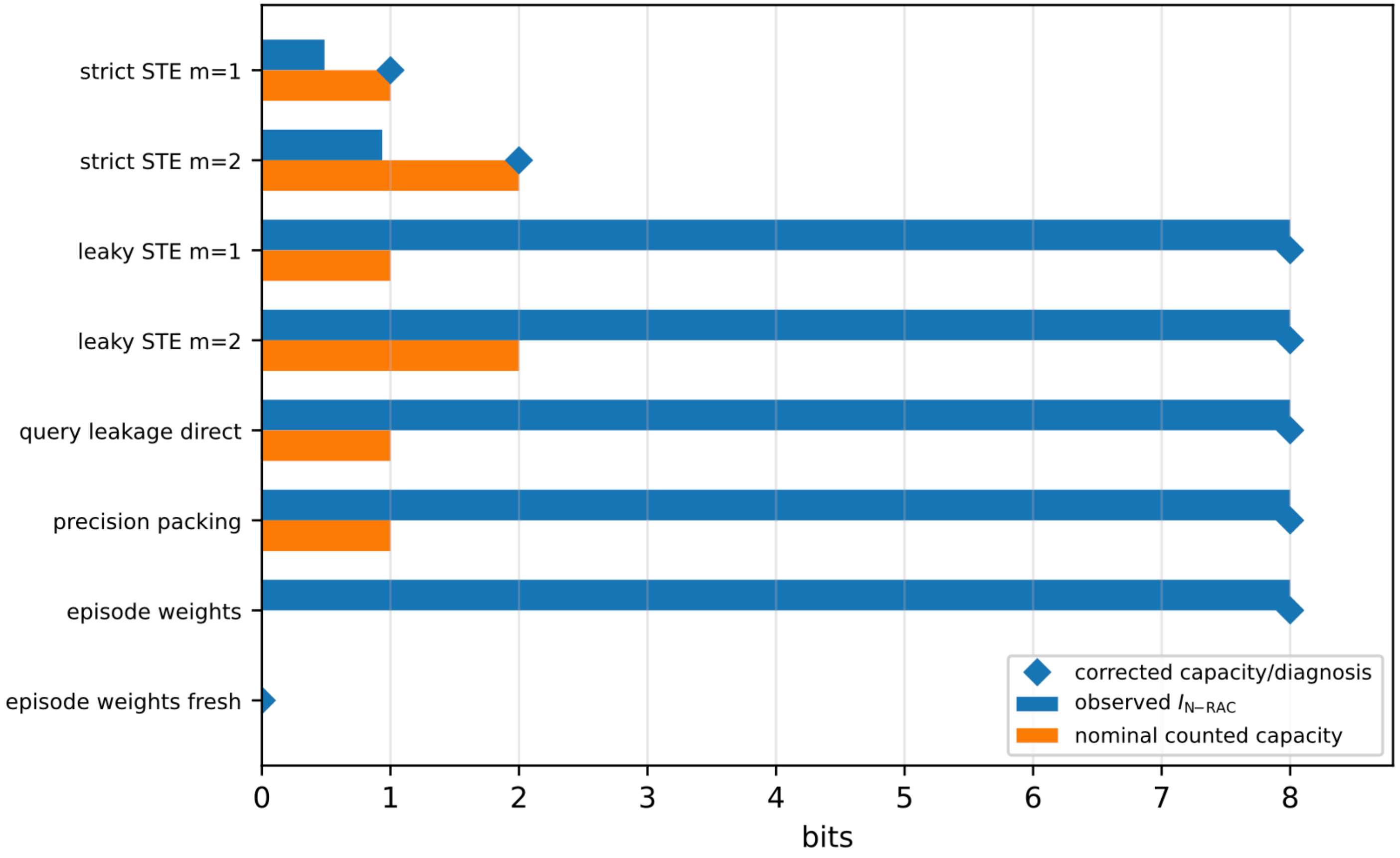
Finite-depth critical boundary



- $E_{\mathrm{crit}}(n)$ solves $I_{\text{N-RAC}}(n, E) = 1$.
- It approaches $1/\sqrt{2}$ from above.
- Slightly post-quantum correlations may need sufficient depth before violation is visible.

The operational boundary tightens and approaches the Tsirelson value from above

Leakage diagnostics: strict versus leaky bottlenecks



- Strict binary bottlenecks stay below counted capacity.
- Query leakage can directly transmit a_b.
- Precision packing hides many bits in nominally small real vectors.
- Episode-specific weights behave as uncounted memory.

Neural-IC therefore gives a structured way to debug unusual powerful representations.

Take-home messages

1. Neural-IC is an operational embedding of IC into query-separated representation learning.
2. The key law is not dimension-based; it is capacity-certified:

$$I_{\mathrm{N\text{-}RAC}} \leq I(\vec{a} : H, B) \leq C_H.$$

3. CHSH nesting turns correlation bias into an amplification mechanism.
4. Arbitrary-depth one-bit stability selects the Tsirelson threshold $E = 1/\sqrt{2}$.
5. Quantum correlations improve fair query-conditioned access without turning finite representations into oracles.

Thank you