Let's pin down exactly what you want to show and then walk through the bridge step by step.

---

0. What needs to be proved?

They define the (infinite-temperature) OTOC as (Eq. (1))

$$O(t) \equiv \frac{\mathrm{Tr}\big(\rho\, W(t)\, V^\dagger W(t) V\big)}{\mathrm{Tr}\big(\rho\, W(t)^2\, V^\dagger V\big)}\,,$$

with

- $\rho = \mathbb{I}/N_H$ (infinite temperature),
- V unitary,
- W Hermitian and traceless ($\mathrm{Tr} W = 0$).

So for this case

$$O(t) = \frac{\mathrm{Tr}\big(W(t)\, V^\dagger W(t) V\big)}{\mathrm{Tr}\big(W(t)^2\big)}. \tag{1'}$$

The **global protocol** says: prepare a random state

$$|\psi_{u,k_0}\rangle = u|k_0\rangle$$

with u a global random unitary (CUE or any unitary 2-design). Then for each u measure:

- $X_u = \langle W(t)\rangle_{u,k_0} = \langle\psi_{u,k_0}|W(t)|\psi_{u,k_0}\rangle$,
- $Y_u = \langle V^\dagger W(t) V\rangle_{u,k_0} = \langle\psi_{u,k_0}|V^\dagger W(t) V|\psi_{u,k_0}\rangle$.

Then define the statistical correlation (Eq. (2)):

$$O(t) = \frac{1}{D(G)}\,\overline{X_u Y_u} \quad \text{with} \quad D(G) = \overline{X_u^2}, \tag{2}$$

where the overline is the ensemble average over random unitaries u .

Goal: Show that Eq. (2) (global protocol) reduces exactly to Eq. (1′).

So we need to show

$$\overline{X_u Y_u} = c\,\mathrm{Tr}\big(W(t) V^\dagger W(t) V\big), \qquad \overline{X_u^2} = c\,\mathrm{Tr}\big(W(t)^2\big), \tag{★}$$

with the same constant c , so the ratio gives Eq. (1′). The paper finds $c = [N_H(N_H+1)]^{-1}$.

---

1. Rewrite everything in density-matrix language

Fix a pure reference state

$$\rho_0 = |k_0\rangle\langle k_0|.$$

For a given random unitary u ,

$$|\psi_{u,k_0}\rangle = u|k_0\rangle, \quad \rho_u = |\psi_{u,k_0}\rangle\langle\psi_{u,k_0}| = u\rho_0 u^\dagger.$$

Then

$$X_u = \mathrm{Tr}(\rho_u W(t)) = \mathrm{Tr}\big(\rho_0\, u^\dagger W(t) u\big),$$
$$Y_u = \mathrm{Tr}(\rho_u V^\dagger W(t) V) = \mathrm{Tr}\big(\rho_0\, u^\dagger V^\dagger W(t) V u\big).$$

So their *product* for fixed u is

$$X_u Y_u = \mathrm{Tr}\big(\rho_0 u^\dagger W(t) u\big)\ \mathrm{Tr}\big(\rho_0 u^\dagger V^\dagger W(t) V u\big).$$

There's a standard trick: **product of traces = single trace on a doubled Hilbert space**.

Let $\mathcal{H}$ be the original Hilbert space and $\mathcal{H}^{\otimes 2} = \mathcal{H}_1 \otimes \mathcal{H}_2$ its double. Then for any operators A, B, C, D ,

$$\mathrm{Tr}(AB)\,\mathrm{Tr}(CD) = \mathrm{Tr}_{12}\big[(A\otimes C)(B\otimes D)\big].$$

Apply this with

$$A=\rho_0,\ B=u^\dagger W(t)u,\quad C=\rho_0,\ D=u^\dagger V^\dagger W(t)Vu,$$

to get

$$X_uY_u = \mathrm{Tr}_{12}\big[(\rho_0\otimes\rho_0)\,(u^\dagger\otimes u^\dagger)(W(t)\otimes V^\dagger W(t)V)(u\otimes u)\big]. \tag{3}$$

Similarly,

$$X_u^2 = \mathrm{Tr}_{12}\big[(\rho_0\otimes\rho_0)\,(u^\dagger\otimes u^\dagger)(W(t)\otimes W(t))(u\otimes u)\big]. \tag{4}$$

So the ensemble averages are

$$\overline{X_uY_u} = \mathrm{Tr}_{12}\big[(\rho_0\otimes\rho_0)\,\Phi\big(W(t)\otimes V^\dagger W(t)V\big)\big], \tag{5}$$

$$\overline{X_u^2} = \mathrm{Tr}_{12}\big[(\rho_0\otimes\rho_0)\,\Phi\big(W(t)\otimes W(t)\big)\big], \tag{6}$$

where I've defined the **twirling map**

$$\Phi(X) := \overline{(u^\dagger\otimes u^\dagger)X(u\otimes u)}.$$

Because the ensemble is a unitary 2-design, this average is exactly the Haar average over the unitary group, which is what Eq. (3) in the paper encodes.

So the whole problem reduces to: *under the Haar-twirl Φ , what happens to $W\otimes B$ with $B=W$ or $B=V^\dagger WV$?*

---

2. Structure of the twirl Φ\Phi on H⊗H\mathcal{H}\otimes\mathcal{H}

Key fact: the Haar twirl on two copies has only two invariant operators: the identity $\mathbb{I}\otimes\mathbb{I}$ and the **swap operator**

$$\mathrm{swap}(|i\rangle\otimes|j\rangle) = |j\rangle\otimes|i\rangle.$$

So for any X on $\mathcal{H}\otimes\mathcal{H}$,

$$\Phi(X) = \alpha(X)\,\mathbb{I}\otimes\mathbb{I} + \beta(X)\,\mathrm{swap}, \tag{7}$$

with α,β linear in X . One way to get α,β is to use the explicit 2-design identity for matrix elements (their Eq. (3)) — but structurally, all you need is that the twirl projects onto the span of $\{\mathbb{I},\mathrm{swap}\}$.

The coefficients can be conveniently expressed via

$$\mathrm{Tr}(X),\qquad \mathrm{Tr}(\mathrm{swap}\,X),$$

because these are the only two invariants compatible with that structure. Matching with the known Haar-integral (or explicitly using Eq. (3)), you get the formula they essentially encode diagrammatically in Fig. 2(a):

$$\Phi(X) = \frac{\mathrm{Tr}(X) - \frac{1}{N_H}\mathrm{Tr}(\mathrm{swap}\,X)}{N_H^2-1}\,\mathbb{I}\otimes\mathbb{I} + \frac{\mathrm{Tr}(\mathrm{swap}\,X) - \frac{1}{N_H}\mathrm{Tr}(X)}{N_H^2-1}\,\mathrm{swap}. \tag{8}$$

We won't re-derive (8) from scratch; it's exactly what you get if you rewrite their index identity (Eq. (3)) in operator form. The diagram in Fig. 2(a) is just a pictorial version of this formula.

---

3. Apply Φ\Phi to the numerator

Take

$$X = W(t)\otimes V^\dagger W(t)V.$$

We need $\mathrm{Tr}(X)$ and $\mathrm{Tr}(\mathrm{swap}\,X)$.

1. **Plain trace:**

   $$\mathrm{Tr}(X) = \mathrm{Tr}(W(t))\,\mathrm{Tr}(V^\dagger W(t)V).$$

   Since V is unitary, $\mathrm{Tr}(V^\dagger W(t)V) = \mathrm{Tr}(W(t))$ by cyclicity. And we assumed W (hence $W(t)$) is traceless, so

   $$\mathrm{Tr}(X) = [\mathrm{Tr}(W(t))]^2 = 0.$$

2. **Swap trace:**

   $$\mathrm{Tr}(\mathrm{swap}\,X) = \mathrm{Tr}\big(\mathrm{swap}\,(W(t)\otimes V^\dagger W(t)V)\big).$$

But for any operators A, B ,

$$\mathrm{Tr}(\mathrm{swap}\, A \otimes B) = \mathrm{Tr}(AB),$$

so

$$\mathrm{Tr}(\mathrm{swap}\, X) = \mathrm{Tr}\big(W(t)\, V^\dagger W(t) V\big).$$

Plug into (8) with $\mathrm{Tr}(X) = 0$:

$$\Phi\big(W(t) \otimes V^\dagger W(t) V\big) = \frac{-\frac{1}{N_H}\mathrm{Tr}(\mathrm{swap}\, X)}{N_H^2 - 1}\, \mathbb{I} \otimes \mathbb{I} + \frac{\mathrm{Tr}(\mathrm{swap}\, X)}{N_H^2 - 1}\, \mathrm{swap}.$$

Factor out $\mathrm{Tr}(\mathrm{swap}\, X) = \mathrm{Tr}(W(t) V^\dagger W(t) V)$:

$$\Phi\big(W(t) \otimes V^\dagger W(t) V\big) = \frac{\mathrm{Tr}(W(t) V^\dagger W(t) V)}{N_H^2 - 1} \left(-\frac{1}{N_H}\, \mathbb{I} \otimes \mathbb{I} + \mathrm{swap} \right). \tag{9}$$

Now plug this into (5):

$$\overline{X_u Y_u} = \mathrm{Tr}_{12}\Big[(\rho_0 \otimes \rho_0)\, \Phi\big(W(t) \otimes V^\dagger W(t) V\big) \Big].$$

Using (9):

$$\overline{X_u Y_u} = \frac{\mathrm{Tr}(W(t) V^\dagger W(t) V)}{N_H^2 - 1} \Big(-\frac{1}{N_H}\, \mathrm{Tr}_{12}(\rho_0 \otimes \rho_0) + \mathrm{Tr}_{12}[(\rho_0 \otimes \rho_0)\, \mathrm{swap}] \Big). \tag{10}$$

Now use simple identities for a **pure state** $\rho_0 = |k_0\rangle\langle k_0|$:

- $\mathrm{Tr}_{12}(\rho_0 \otimes \rho_0) = 1$,
- $\mathrm{Tr}_{12}[(\rho_0 \otimes \rho_0)\, \mathrm{swap}] = \mathrm{Tr}(\rho_0^2) = 1$.

(The second one is standard: $\mathrm{swap}(A \otimes B) = B \otimes A$; plug $A = B = \rho_0$ and trace.)

So the bracket in (10) becomes

$$-\frac{1}{N_H} \cdot 1 + 1 = 1 - \frac{1}{N_H} = \frac{N_H - 1}{N_H}.$$

Thus

$$\overline{X_u Y_u} = \frac{N_H - 1}{N_H (N_H^2 - 1)}\, \mathrm{Tr}(W(t) V^\dagger W(t) V) = \frac{1}{N_H (N_H + 1)}\, \mathrm{Tr}(W(t) V^\dagger W(t) V). \tag{11}$$

So we've got the constant

$$c = \frac{1}{N_H (N_H + 1)},$$

exactly as in Eq. (4) of the paper.

That proves the **numerator identity** in (★).

---

4. Apply Φ\Phi to the denominator

Now repeat the same for

$$X' = W(t) \otimes W(t).$$

We need:

- $\mathrm{Tr}(X') = [\mathrm{Tr}(W(t))]^2 = 0$ (again traceless W),
- $\mathrm{Tr}(\mathrm{swap} X') = \mathrm{Tr}(W(t)^2)$.

So by copying the previous logic, we get

$$\Phi(W(t) \otimes W(t)) = \frac{\mathrm{Tr}(W(t)^2)}{N_H^2 - 1} \left(-\frac{1}{N_H}\, \mathbb{I} \otimes \mathbb{I} + \mathrm{swap} \right).$$

Plug this into (6):

$$\overline{X_u^2} = \mathrm{Tr}_{12}\big[(\rho_0 \otimes \rho_0) \Phi(W(t) \otimes W(t)) \big] = \frac{\mathrm{Tr}(W(t)^2)}{N_H (N_H + 1)}. \tag{12}$$

So indeed the same constant c appears:

$$\overline{X_u^2} = c\,\mathrm{Tr}\big(W(t)^2\big), \quad c = \frac{1}{N_H(N_H+1)}.$$

This matches the statement in the text right after Eq. (4): the denominator $D(G)$ reduces to $c\,\mathrm{Tr}[W(t)^2]$.

---

5. Take the ratio: Eq. (2) → Eq. (1′)

Now just plug (11) and (12) into Eq. (2):

$$O(t) = \frac{\overline{X_u Y_u}}{D(G)} = \frac{c\,\mathrm{Tr}(W(t)V^\dagger W(t)V)}{c\,\mathrm{Tr}(W(t)^2)} = \frac{\mathrm{Tr}(W(t)V^\dagger W(t)V)}{\mathrm{Tr}(W(t)^2)}.$$

This is exactly Eq. (1′), i.e. Eq. (1) at infinite temperature with unitary V , so
the global random-measurement protocol reproduces the standard OTOC exactly.

In the paper this whole derivation is compressed into Eq. (4) and Fig. 2(b): the diagrams are just a compact way of representing (5)–(12).

---

6. Quick conceptual recap (no algebra)

If you want the “physics one-liner” in your head:

- Haar-randomizing the initial state is equivalent to averaging over pure states uniformly on the Hilbert space.
- The correlator $\overline{\langle W\rangle\langle V^\dagger W V\rangle}$ is essentially the Haar-twirl of $W \otimes V^\dagger W V$ projected with a fixed pure state on each copy.
- By Schur–Weyl, the Haar-twirl on two copies can only produce $\mathbb{I}$ and swap. The tracelessness of W kills the identity contribution, leaving only swap, which evaluates to $\mathrm{Tr}(WV^\dagger W V)$.
- Doing the same for $\overline{\langle W\rangle}^2$ gives $\mathrm{Tr}(W^2)$ with the same prefactor, hence the ratio is the usual infinite-T OTOC.

If you’d like, next step we can go *backwards* from the diagram in Fig. 2(b) and match each piece to the algebra above, so the diagrammatic proof feels completely transparent to you.

Great — these two questions are exactly the two conceptual bottlenecks everyone hits the first time they try to understand the derivation *cleanly*. I’ll answer both in a detailed, mathematically rigorous way, but at a physicist-friendly level.

Q1. Why does the twirling map Φ(X)\Phi(X) only produce a linear combination of I⊗I\mathbb{I}\otimes\mathbb{I} and SWAP?

This is the statement:

$$\Phi(X) := \int dU\,(U^\dagger \otimes U^\dagger)\,X\,(U\otimes U) = \alpha(X)\,\mathbb{I}\otimes\mathbb{I} + \beta(X)\,\mathrm{SWAP}. \tag{*}$$

You correctly noticed that Eq. (3) in the paper gives a 4-index identity:

$$\overline{u_{mm_0}u_{nn_0}u^*_{m'm_0}u^*_{n'n_0}} = \frac{\delta_{mm'}\delta_{nn'} + \delta_{mn'}\delta_{nm'}}{N_H(N_H+1)}.$$

This is the same identity hidden inside (*), but you want a conceptual explanation:
Why are the only invariant operators the identity and SWAP?

✔ This is an example of Schur–Weyl duality.

The twirling map Φ has the symmetry

$$\Phi(X) \equiv \int dU\,(U^\dagger \otimes U^\dagger)\,X\,(U\otimes U).$$

Twirling says: **project X onto the subspace of operators that commute with every U ⊗ U.**

This subspace is exactly the **commutant** of the representation $U \otimes U$.

A standard result in representation theory:

▶ The commutant of {U⊗U:U∈U(N)}\{U\otimes U: U\in U(N)\} is spanned by only two operators:

1. $\mathbb{I} \otimes \mathbb{I}$
2. SWAP

No other operator simultaneously commutes with all $U \otimes U$.

Why? A physical intuition:

- The action $U \otimes U$ treats the two copies of Hilbert space *identically*.
- So the only structures that survive after integrating over *all* U 's must be:
  - doing nothing to distinguish the two systems → **identity**
  - exchanging the roles of the two systems → **swap**

Everything else is washed away by group averaging.

A more formal way:

The representation $U \otimes U$ acting on $\mathcal{H} \otimes \mathcal{H}$ decomposes as:

$$\mathcal{H} \otimes \mathcal{H} = \mathcal{H}_{\mathrm{sym}} \oplus \mathcal{H}_{\mathrm{asym}}.$$

These are:

- symmetric subspace (dimension $N(N+1)/2$)
- antisymmetric subspace (dimension $N(N-1)/2$)

The **projector onto symmetric subspace** is

$$P_{\mathrm{sym}} = \frac{1}{2}(\mathbb{I} \otimes \mathbb{I} + \mathrm{SWAP}),$$

and onto antisymmetric:

$$P_{\mathrm{as}} = \frac{1}{2}(\mathbb{I} \otimes \mathbb{I} - \mathrm{SWAP}).$$

These two subspaces transform irreducibly under $U \otimes U$.
By **Schur's lemma**, the twirl must act as a scalar on each irrep:

$$\Phi(X) = a\, P_{\mathrm{sym}} + b\, P_{\mathrm{as}} = \alpha\, \mathbb{I} \otimes \mathbb{I} + \beta\, \mathrm{SWAP}.$$

Thus we get exactly eq. (7).

There is **no room** for other operators — any other operator changes under $U \otimes U$ and must average to zero.

So that explains why Eq. (3) collapses to the two-term structure in (*).

Q2. Why is

$$\mathrm{Tr}(\mathrm{SWAP}\, A \otimes B) = \mathrm{Tr}(AB)?$$

This is very simple once you write the trace in a basis.

Let $\{|i\rangle\}$ be any orthonormal basis.
Then the swap operator is:

$$\mathrm{SWAP}(|i\rangle \otimes |j\rangle) = |j\rangle \otimes |i\rangle.$$

Now compute:

$$\begin{aligned}
\mathrm{Tr}(\mathrm{SWAP}\, A \otimes B) &= \sum_{i,j} \langle i,j|\, \mathrm{SWAP}\, (A \otimes B)\, |i,j\rangle \\
&= \sum_{i,j} \langle j,i|\, A \otimes B\, |i,j\rangle \\
&= \sum_{i,j} \langle j|A|i\rangle \langle i|B|j\rangle \\
&= \sum_{i,j} A_{ji} B_{ij} \\
&= \sum_{i} (AB)_{ii} \\
&= \mathrm{Tr}(AB).
\end{aligned}$$

That's the entire reason.

This identity is used throughout quantum information in:

- purity measurements
- SWAP tests
- entanglement detection
- randomized benchmarking
- OTOC random measurement protocols (this paper)

SUMMARY

1. Why does Φ(X)=αI⊗I+βSWAP\Phi(X) = \alpha \mathbb{I}\otimes\mathbb{I} + \beta\text{SWAP}?

Because twirling projects onto operators that commute with all $U \otimes U$.
By Schur–Weyl duality, the only such operators are identity and swap.

2. Why does Tr(SWAP A⊗B)=Tr(AB)\mathrm{Tr}(\text{SWAP}\, A\otimes B)=\mathrm{Tr}(AB)?

Because writing SWAP in a basis and performing the trace reduces exactly to matrix multiplication.

If you want, next I can also do:

✔ a full diagrammatic reconstruction of Fig. 2(a) → Fig. 2(b)
✔ derive Eq. (8) explicitly from the 4-index identity (your Eq. (3))
✔ derive the constants α, β step-by-step
✔ show why the tracelessness of W kills the $\mathbb{I} \otimes \mathbb{I}$ term

Just tell me which direction you want to deepen.

Absolutely — let's do this *carefully* and in a way that you can audit line-by-line.

We start from the statement (your "Eq-(7)" / my earlier eq-(7)):

$$\Phi(X) = \int dU \, (U^\dagger \otimes U^\dagger) \, X \, (U \otimes U) = \alpha(X) \, I + \beta(X) \, S, \tag{7}$$

where $I := \mathbb{I} \otimes \mathbb{I}$ and $S := \mathrm{SWAP}$.

Your question: **How do we determine $\alpha(X)$ and $\beta(X)$ in a mistake-free way?**

The clean way is: use **two independent linear functionals** on operators on $\mathcal{H}^{\otimes 2}$. The natural ones are

$$\mathrm{Tr}(\cdot) \quad \text{and} \quad \mathrm{Tr}(S \, \cdot),$$

because I and S are linearly independent and these traces give a solvable 2×2 linear system.

Below is the derivation with *all steps* and *explicit checking*.

Step 1. Two invariants of the twirl

Define

$$\Phi(X) = \int dU \, (U^\dagger \otimes U^\dagger) \, X \, (U \otimes U),$$

with dU the normalized Haar measure (so $\int dU = 1$).

Claim A: Tr(Φ(X))=Tr(X)\mathrm{Tr}(\Phi(X))=\mathrm{Tr}(X).

Proof: Use cyclicity of trace and unitarity:

$$\begin{aligned}\mathrm{Tr}(\Phi(X)) &= \int dU\, \mathrm{Tr}\big((U^\dagger \otimes U^\dagger) X (U \otimes U)\big) \\ &= \int dU\, \mathrm{Tr}\big(X (U \otimes U)(U^\dagger \otimes U^\dagger)\big) \\ &= \int dU\, \mathrm{Tr}(X\, I) = \mathrm{Tr}(X).\end{aligned}$$

No assumptions on X . ✓

Claim B: Tr(S Φ(X))=Tr(S X)\mathrm{Tr}(S\,\Phi(X))=\mathrm{Tr}(S\,X).

Reason: S commutes with $U \otimes U$ for all U :

$$S(U \otimes U) = (U \otimes U)S.$$

(You can check on basis vectors: swap then apply same U to both copies equals apply U to both then swap.)

Now:

$$\begin{aligned}\mathrm{Tr}(S\,\Phi(X)) &= \int dU\, \mathrm{Tr}\Big(S(U^\dagger \otimes U^\dagger) X (U \otimes U)\Big) \\ &= \int dU\, \mathrm{Tr}\Big((U^\dagger \otimes U^\dagger) S X (U \otimes U)\Big) \quad (\text{move } S \text{ through}) \\ &= \int dU\, \mathrm{Tr}(SX) = \mathrm{Tr}(SX),\end{aligned}$$

using Claim A on the operator SX . ✓

So the two numbers

$$t_1 := \mathrm{Tr}(X), \qquad t_2 := \mathrm{Tr}(SX)$$

are *invariants* under the twirl (they are the same for X and for $\Phi(X)$).

Step 2. Write down the linear system for α,β\alpha,\beta

Assume

$$\Phi(X) = \alpha I + \beta S.$$

Compute the two traces on the RHS.

Compute Tr(Φ(X))\mathrm{Tr}(\Phi(X))

We need:

- $\mathrm{Tr}(I) = \mathrm{Tr}(\mathbb{I} \otimes \mathbb{I}) = N^2$,
- $\mathrm{Tr}(S) = N$ (swap has N fixed basis vectors $|i\rangle|i\rangle$).

So

$$\mathrm{Tr}(\Phi(X)) = \alpha\, \mathrm{Tr}(I) + \beta\, \mathrm{Tr}(S) = \alpha N^2 + \beta N. \tag{E1}$$

But by Claim A this equals $\mathrm{Tr}(X) = t_1$. Hence

$$\alpha N^2 + \beta N = t_1. \tag{L1}$$

Compute Tr(S Φ(X))\mathrm{Tr}(S\,\Phi(X))

We need:

- $SI = S$,
- $S^2 = I$,
  so

$$\mathrm{Tr}(S\,\Phi(X)) = \mathrm{Tr}\big(S(\alpha I + \beta S)\big) = \alpha\, \mathrm{Tr}(S) + \beta\, \mathrm{Tr}(S^2) = \alpha N + \beta N^2. \tag{E2}$$

By Claim B this equals $\mathrm{Tr}(SX) = t_2$. Hence

$$\alpha N + \beta N^2 = t_2. \tag{L2}$$

So we have a **2×2 linear system**:

$$\begin{pmatrix} N^2 & N \\ N & N^2 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}. \tag{LS}$$

Step 3. Solve the system (and double-check algebra)

The determinant is

$$\det = N^4 - N^2 = N^2(N^2-1) \neq 0 \quad (N > 1),$$

so unique solution.

Compute inverse explicitly:

$$\begin{pmatrix} N^2 & N \\ N & N^2 \end{pmatrix}^{-1} = \frac{1}{N^4 - N^2} \begin{pmatrix} N^2 & -N \\ -N & N^2 \end{pmatrix}.$$

Thus

$$\alpha = \frac{1}{N^4 - N^2}(N^2 t_1 - N t_2), \qquad \beta = \frac{1}{N^4 - N^2}(-N t_1 + N^2 t_2).$$

Factor out N^2 :

$$N^4 - N^2 = N^2(N^2 - 1).$$

So

$$\boxed{\alpha(X) = \frac{N\,t_1 - t_2}{N(N^2-1)}} \tag{A}$$

$$\boxed{\beta(X) = \frac{N\,t_2 - t_1}{N(N^2-1)}} \tag{B}$$

where $t_1 = \mathrm{Tr}(X)$ and $t_2 = \mathrm{Tr}(SX)$.

Quick internal consistency checks

Check 1: Plug $X = I$.
Then $\Phi(I) = I$. We must get $\alpha = 1, \beta = 0$.

Compute:

- $t_1 = \mathrm{Tr}(I) = N^2$,
- $t_2 = \mathrm{Tr}(SI) = \mathrm{Tr}(S) = N$.

Then

$$\alpha = \frac{N \cdot N^2 - N}{N(N^2-1)} = \frac{N^3 - N}{N(N^2-1)} = \frac{N(N^2-1)}{N(N^2-1)} = 1,$$

$$\beta = \frac{N \cdot N - N^2}{N(N^2-1)} = 0.$$

✓

Check 2: Plug $X = S$.
Then $\Phi(S) = S$. We must get $\alpha = 0, \beta = 1$.

Compute:

- $t_1 = \mathrm{Tr}(S) = N$,
- $t_2 = \mathrm{Tr}(S^2) = \mathrm{Tr}(I) = N^2$.

Then

$$\alpha = \frac{N \cdot N - N^2}{N(N^2-1)} = 0, \quad \beta = \frac{N \cdot N^2 - N}{N(N^2-1)} = 1.$$

✓

So (A)(B) are correct.

Step 4. Rewrite in the exact “Eq-(8)” form

Now substitute $t_1 = \mathrm{Tr}(X)$, $t_2 = \mathrm{Tr}(SX)$ into $\Phi(X) = \alpha I + \beta S$.

From (A):

$$\alpha = \frac{N\,\mathrm{Tr}(X) - \mathrm{Tr}(SX)}{N(N^2-1)} = \frac{\mathrm{Tr}(X) - \frac{1}{N}\mathrm{Tr}(SX)}{N^2-1}.$$

From (B):

$$\beta = \frac{N\,\mathrm{Tr}(SX) - \mathrm{Tr}(X)}{N(N^2-1)} = \frac{\mathrm{Tr}(SX) - \frac{1}{N}\mathrm{Tr}(X)}{N^2-1}.$$

Therefore

$$\boxed{\Phi(X) = \frac{\mathrm{Tr}(X) - \frac{1}{N}\mathrm{Tr}(SX)}{N^2-1}\, I + \frac{\mathrm{Tr}(SX) - \frac{1}{N}\mathrm{Tr}(X)}{N^2-1}\, S} \tag{8}$$

This is exactly the operator form I previously wrote as Eq. (8), and it is the precise “from Eq-(7) to Eq-(8)” bridge.

Step 5. Specialize to the paper’s XX (so you see why it collapses)

In the global protocol you use $X = W(t) \otimes V^\dagger W(t) V$. Then:

- $\mathrm{Tr}(X) = \mathrm{Tr}(W(t))\mathrm{Tr}(V^\dagger W(t) V) = \mathrm{Tr}(W(t))^2 = 0$ since W is traceless.
- $\mathrm{Tr}(SX) = \mathrm{Tr}(W(t) V^\dagger W(t) V)$ using $\mathrm{Tr}(SA \otimes B) = \mathrm{Tr}(AB)$.

So in that case Eq. (8) simplifies immediately to

$$\Phi(X) = \frac{-\frac{1}{N}\mathrm{Tr}(W(t)V^\dagger W(t)V)}{N^2-1}\, I + \frac{\mathrm{Tr}(W(t)V^\dagger W(t)V)}{N^2-1}\, S,$$

which is the simplification used later.

Where people usually get tripped up (and why this derivation avoids it)

- If you try to get α, β by “matching indices” directly from the paper’s Eq.(3), it’s easy to lose track of what’s being summed and what’s being held fixed.
- The trace-functional method above is robust: it only uses two facts
  1. $\Phi(X) \in \mathrm{span}\{I, S\}$
  2. Tr and $\mathrm{Tr}(S\,\cdot)$ are invariant under twirling.

No hidden steps; every identity is checkable.

If you want, next I can do the “index-level” derivation too: start from the 4- u identity in the paper (their Eq. 3), insert an explicit general $X_{ab,cd}$, do the contractions, and show it collapses to the same α, β . That route is longer but it makes the connection to their Eq.(3) maximally explicit.

Good questions. I’ll keep (1) short as requested, and be precise-but-compact for (2).

---

1. Relation to kk-designs — are you correct?

Yes, you are correct.
A twirling channel built from $U^{\otimes k}$ corresponds exactly to Haar moments of order k ; replacing Haar by a unitary k -design gives the *same twirl up to order* k. For the 2 -tensor case we used a unitary 2-design; higher tensor powers require higher designs.

---

2. BIG QUESTION — What subspace does higher-order twirling project onto?

Short answer (structural):

Yes.
The k -fold twirling channel

$$\Phi_k(X) \;=\; \int dU\,(U^\dagger)^{\otimes k}\,X\,U^{\otimes k}$$

projects X onto the **commutant of the representation** $U^{\otimes k}$.
This commutant is spanned by **permutation operators of the symmetric group** S_k.

Precise statement (this is the core result)

For $U \in U(N)$, acting on $\mathcal{H}^{\otimes k}$:

$$\boxed{\mathrm{Comm}\big(U^{\otimes k}\big) \;=\; \mathrm{span}\{\,P_\pi \mid \pi \in S_k\,\}}$$

where:

- S_k is the symmetric group on k elements,
- P_π permutes tensor factors according to π :

$$P_\pi\big(|i_1\rangle \otimes \cdots \otimes |i_k\rangle\big) = |i_{\pi^{-1}(1)}\rangle \otimes \cdots \otimes |i_{\pi^{-1}(k)}\rangle.$$

This is the **Schur–Weyl duality** statement in operator form.

Consequence for the twirling channel

Because Φ_k is a projection onto the commutant:

$$\boxed{\Phi_k(X) \;=\; \sum_{\pi \in S_k} c_\pi(X)\,P_\pi}$$

- The coefficients $c_\pi(X)$ are **linear functionals of** X.
- They are determined by overlaps such as
  $\mathrm{Tr}(P_\pi X)$, exactly like $\mathrm{Tr}(X)$ and $\mathrm{Tr}(\mathrm{SWAP}\,X)$ fixed the $k=2$ case.
- No operator outside the permutation algebra survives the average.

How this generalizes your k=2k=2 intuition

k	Surviving operators after twirl	Dimension
1	identity	1
2	$\{I,\ \mathrm{SWAP}\}$	2
3	all 6 permutations in S_3	6
general k	all $k!$ permutations	$k!$

So what you saw for $k=2$ is **not special** — it is the first nontrivial case of a fully general pattern.

Representation-theoretic meaning (why this must be true)

- $U^{\otimes k}$ decomposes $\mathcal{H}^{\otimes k}$ into irreducible representations labelled by **Young diagrams** with k boxes.
- The twirl acts as a **blockwise identity** on each irrep.
- Operators that survive must act *within multiplicity spaces only*.
- Those multiplicity spaces are exactly generated by permutations.

This is why:

- higher-order Haar averages are controlled by permutations,
- Weingarten calculus appears,

- randomized-measurement protocols at order k always reduce to permutation diagrams.

Physical intuition (important)

Twirling by $U^{\otimes k}$:

> "forgets everything about X except how it behaves under relabeling of the k copies."

Permutations are the **only memory** that survives complete unitary scrambling.

Bottom line

- ✔ Higher-order twirling **does project onto a specific subspace**.
- ✔ That subspace is the **group algebra of S_k**.
- ✔ The $k = 2$ $\{I, \mathrm{SWAP}\}$ structure is just the S_2 case.
- ✔ All higher-order randomized protocols (OTOCs, shadows, purity estimation, Rényi entropies) are controlled by this structure.

If you want, next we can:

- explicitly derive $\Phi_3(X)$,
- connect this to Weingarten functions diagrammatically,
- or explain how this structure controls **which correlators randomized measurements can or cannot access**.

You're right to pause here: **Appendix G is not "general ρ " in full generality** — it's a **finite-temperature (thermal) extension** built from a **high-temperature expansion** and a *symmetrized* thermal OTOC. That matters for correctness. The appendix title in the arXiv text is "Thermal OTOCs from global unitaries." ar5iv

Below I'll (i) restate *exactly* what they compute, (ii) derive the expansion cleanly (with internal consistency checks), and (iii) translate it into an experimental protocol.

---

1) What object Appendix G targets (and why it's "symmetrized")

They explicitly **do not** start from the "canonical" finite-T OTOC, but from a **symmetrized variant** (they cite Maldacena et al.). ar5iv

Let

$$\rho_\beta = \frac{e^{-\beta H}}{Z}, \qquad Z = \mathrm{Tr}(e^{-\beta H}),$$

and define the (Heisenberg) $W(t) = U^\dagger(t) W U(t)$.

The *symmetrized* thermal OTOC has the form (up to notation conventions)

$$F_\beta(t) := \mathrm{Tr}\Big(\rho_\beta^{1/2}\, W(t)\, \rho_\beta^{1/2}\, V^\dagger W(t) V\Big),$$

rather than $\mathrm{Tr}(\rho_\beta\, W(t)\, V^\dagger W(t) V)$. The appendix then expands this in β (high-T). ar5iv

They also assume H Hermitian, V Hermitian+unitary, and $W, V, \ldots$ traceless for their spin examples. ar5iv

---

2) High-temperature expansion to first order in β\beta

2.1 Normalize away annoying constants (consistency step)

A standard trick (and it's **legit**): shift the Hamiltonian by a constant,

$$H \mapsto H - \frac{\mathrm{Tr}(H)}{N}\,\mathbb{I},$$

which does **not** change ρ_β (it only rescales Z). So WLOG you can assume

$$\mathrm{Tr}(H) = 0. \tag{C0}$$

This ensures the first-order expansion of ρ_β is clean.

2.2 Expand ρβ\rho_\beta and ρβ1/2\rho_\beta^{1/2}

Using $e^{-\beta H} = \mathbb{I} - \beta H + O(\beta^2)$ and $Z = \mathrm{Tr}(e^{-\beta H}) = N + O(\beta^2)$ given (C0),

$$\rho_\beta = \frac{1}{N}(\mathbb{I} - \beta H) + O(\beta^2). \tag{Eϱ}$$

Similarly,

$$e^{-\beta H/2} = \mathbb{I} - \frac{\beta}{2}H + O(\beta^2),$$

and $\sqrt{Z} = \sqrt{N} + O(\beta^2)$, so

$$\rho_\beta^{1/2} = \frac{1}{\sqrt{N}}\left(\mathbb{I} - \frac{\beta}{2}H\right) + O(\beta^2). \tag{E√ϱ}$$

Consistency check: $(\rho_\beta^{1/2})^2 = \rho_\beta + O(\beta^2)$.
Indeed,

$$\left(\mathbb{I} - \frac{\beta}{2}H\right)^2 = \mathbb{I} - \beta H + O(\beta^2),$$

and prefactors match. ✓

2.3 Expand the symmetrized numerator

Let $A := W(t)$ and $B := V^\dagger W(t) V$. Then

$$F_\beta(t) = \mathrm{Tr}\left(\rho_\beta^{1/2} A \rho_\beta^{1/2} B\right).$$

Plug (E√ρ):

$$\rho_\beta^{1/2} A \rho_\beta^{1/2} = \frac{1}{N}\left(\mathbb{I} - \frac{\beta}{2}H\right) A \left(\mathbb{I} - \frac{\beta}{2}H\right) + O(\beta^2).$$

Multiply out to first order:

$$(\mathbb{I} - \tfrac{\beta}{2}H)A(\mathbb{I} - \tfrac{\beta}{2}H) = A - \frac{\beta}{2}(HA + AH) + O(\beta^2) = A - \frac{\beta}{2}\{H, A\} + O(\beta^2).$$

Therefore

$$F_\beta(t) = \frac{1}{N}\mathrm{Tr}(AB) - \frac{\beta}{2N}\mathrm{Tr}(\{H, A\}B) + O(\beta^2). \tag{Fexp1}$$

Now use cyclicity to rewrite the correction in a symmetric way:

$$\mathrm{Tr}(\{H, A\}B) = \mathrm{Tr}(HAB) + \mathrm{Tr}(AHB) = \mathrm{Tr}(HAB) + \mathrm{Tr}(HBA) = \mathrm{Tr}(H\{A, B\}).$$

So

$$\boxed{F_\beta(t) = \frac{1}{N}\mathrm{Tr}(AB) - \frac{\beta}{2N}\mathrm{Tr}(H\{A, B\}) + O(\beta^2).} \tag{Fexp}$$

This is the mathematically clean “first order in β ” statement they refer to. ar5iv

---

3) How to measure Fβ(t)F_\beta(t) using randomized measurements

Appendix G says: define **additional statistical correlations** (besides the usual ones) and you can access the thermal correction; operationally, you must also measure the **energy of random initial states**. ar5iv

The key point is: the high-T correction introduces **an extra insertion of H** in a trace. So you need **correlations that estimate traces with H included**.

3.1 The “infinite temperature” piece is unchanged

For each global random unitary u and a fixed reference $|k_0\rangle$, prepare

$$|\psi_u\rangle = u|k_0\rangle, \quad \rho_u = |\psi_u\rangle\langle\psi_u|.$$

Measure two expectation values (as in the main global protocol):

$$X_u := \langle A\rangle_u = \mathrm{Tr}(\rho_u A), \qquad Y_u := \langle B\rangle_u = \mathrm{Tr}(\rho_u B).$$

Then the averaged product $\overline{X_u Y_u}$ gives $\propto \mathrm{Tr}(AB)$ (the same 2-design identity you already worked through in the infinite-T proof).

So the first term $\frac{1}{N}\mathrm{Tr}(AB)$ is obtained exactly the same way as Eq. (2) in the main text. ar5iv

3.2 The thermal correction requires a new correlation involving HH

From (Fexp), the new object is

$$\mathrm{Tr}(H\{A,B\}) = \mathrm{Tr}(HAB) + \mathrm{Tr}(HBA).$$

Appendix G's operational message is: you can get this by *also measuring the energy* of the randomized initial state, i.e.

$$E_u := \langle H\rangle_u = \mathrm{Tr}(\rho_u H),$$

and then forming additional correlations. ar5iv

Important consistency note (critical):
To reconstruct a trace of a *three-operator product* (like $\mathrm{Tr}(HAB)$) from randomized pure-state data, you typically need **third moments** of Haar-random states, i.e. effectively a **unitary 3-design** (not just a 2-design). Appendix G explicitly uses CUE/Haar properties (their wording " _ -design properties of the CUE" is pointing at that). ar5iv
So: **the thermal extension is more demanding than the infinite-T proof.**

3.3 What you actually do in the lab (protocol view)

For each random unitary u , do **three** experiments (or three sets of shots):

1. **Measure** $X_u = \langle W(t)\rangle_u$
   Prepare $|\psi_u\rangle$, evolve with $U(t)$, measure W .
2. **Measure** $Y_u = \langle V^\dagger W(t) V\rangle_u$
   Prepare $|\psi_u\rangle$, apply V (or equivalently measure in rotated basis), evolve, measure W .
3. **Measure** $E_u = \langle H\rangle_u$ (energy of the randomized initial state)
   Prepare $|\psi_u\rangle$, measure energy (or estimate it from local terms if $H = \sum h_i$).
   This "energy measurement" requirement is explicitly stated. ar5iv

Then, in postprocessing:

- use $\overline{X_u Y_u}$ to estimate $\mathrm{Tr}(AB)$,
- use correlations involving E_u and (X_u, Y_u) to estimate $\mathrm{Tr}(H\{A,B\})$,
- combine them as in (Fexp) to get $F_\beta(t)$ to first order in β .

Appendix G's bottom-line formula (their Eq. (40)) says precisely that the thermal OTOC can be obtained by measuring (i) the infinite-T value (as before) and (ii) these "additional correlations" that include the energy. ar5iv

---

4) What I can and cannot guarantee from the source you're using

Because APS is blocked (403) and ar5iv **suppresses the explicit math symbols** in Appendix G, I can't quote their exact definition of the "additional correlations" (they are present as "(G)" / "(40)" placeholders in the HTML). ar5iv

What I *have* given you is the correct, checkable mathematical backbone:

- the symmetrized object they use,
- the **first-order** β expansion with the **anticommutator insertion** $H\{A,B\}$,
- the operational consequence: you must measure **energy of randomized initial states** and form extra correlations. ar5iv

If you want the **exact estimator** they propose (the precise combination of $\overline{E_u X_u Y_u}$, $\overline{E_u X_u Y_u}$, etc.), the best next move is: I open the **actual arXiv PDF** (not ar5iv HTML) and read Appendix G from the rendered equations. Then I can reproduce their Eq. (39)→(40) line-by-line with no ambiguity.

If you want that, tell me "use the PDF" and I'll do it.

You're asking for the **concrete estimator** for a *triple trace* like $\mathrm{Tr}(HAB)$, i.e. how to get it from randomized measurements. Here is the clean, internally-consistent answer.

What you can and cannot estimate directly

From **randomized measurement of expectation values** on random pure states $\rho_u = |\psi_u\rangle\langle\psi_u|$, the natural "raw data" are numbers like

$$h_u = \langle H\rangle_u = \mathrm{Tr}(\rho_u H), \quad a_u = \langle A\rangle_u = \mathrm{Tr}(\rho_u A), \quad b_u = \langle B\rangle_u = \mathrm{Tr}(\rho_u B),$$

and products like $h_u a_u b_u$, $h_u a_u$, etc.

These products correspond to **Haar moments**. In particular, to get *any* third-order trace information you need the **third moment**, i.e. effectively a **unitary 3-design** (Haar is fine; a 3-design also works).

Also: in Appendix G the correction you actually need is

$$\mathrm{Tr}\big(H\{A,B\}\big) = \mathrm{Tr}(HAB) + \mathrm{Tr}(HBA) = 2\,\Re\,\mathrm{Tr}(HAB)$$

for Hermitian H, A, B. So you **do not need** $\mathrm{Tr}(HAB)$ alone; you need the **symmetrized** combination (and it is real). That's why the protocol can work with simple real-valued correlations.

Key identity (the math engine): third moment of Haar-random pure states

Let $N = \dim \mathcal{H}$. If $|\psi\rangle$ is Haar-random and $\rho = |\psi\rangle\langle\psi|$, then

$$\mathbb{E}\big[\rho^{\otimes 3}\big] = \frac{P^{(3)}_{\mathrm{sym}}}{\dim(\mathrm{Sym}^3\mathcal{H})} \quad \text{with} \quad \dim(\mathrm{Sym}^3\mathcal{H}) = \frac{N(N+1)(N+2)}{6},$$

and

$$P^{(3)}_{\mathrm{sym}} = \frac{1}{6}\sum_{\pi\in S_3} P_\pi,$$

where P_π permutes the three tensor factors.

Now observe:

$$h\,a\,b = \mathrm{Tr}(\rho H)\,\mathrm{Tr}(\rho A)\,\mathrm{Tr}(\rho B) = \mathrm{Tr}\big(\rho^{\otimes 3}(H\otimes A\otimes B)\big).$$

Average over $|\psi\rangle$:

$$\mathbb{E}[hab] = \mathrm{Tr}\Big(\mathbb{E}[\rho^{\otimes 3}]\,(H\otimes A\otimes B)\Big) = \frac{\mathrm{Tr}\big(P^{(3)}_{\mathrm{sym}}(H\otimes A\otimes B)\big)}{\dim(\mathrm{Sym}^3\mathcal{H})}.$$

Compute $\mathrm{Tr}(P_\pi\, H\otimes A\otimes B)$ permutation-by-permutation. The result is the standard identity

$$\boxed{\mathbb{E}[hab] = \frac{\mathrm{Tr}(H)\mathrm{Tr}(A)\mathrm{Tr}(B) + \mathrm{Tr}(HA)\mathrm{Tr}(B) + \mathrm{Tr}(HB)\mathrm{Tr}(A) + \mathrm{Tr}(AB)\mathrm{Tr}(H) + \mathrm{Tr}(HAB) + \mathrm{Tr}(HBA)}{N(N+1)(N+2)}.} \tag{★}$$

Consistency check (critical)

Pick $H = A = B = \mathbb{I}$. Then LHS $\mathbb{E}[1] = 1$. RHS becomes

$$\frac{N^3 + 3N^2 + 2N}{N(N+1)(N+2)} = \frac{N(N+1)(N+2)}{N(N+1)(N+2)} = 1.$$

So (★) is correct.

How to isolate the term you need

Case used in Appendix G (and in your earlier derivation):

You can (and they do) assume tracelessness:

- $A = W(t)$ traceless if W is traceless.
- $B = V^\dagger W(t) V$ also traceless.
- You can always shift $H \to H - \frac{\mathrm{Tr}(H)}{N}\mathbb{I}$ so $\mathrm{Tr}(H) = 0$ without changing the Gibbs state (only changes Z).

So assume

$$\mathrm{Tr}(H) = \mathrm{Tr}(A) = \mathrm{Tr}(B) = 0. \tag{T0}$$

Then in (★) all "nuisance" terms vanish automatically, and you get

$$\boxed{\mathbb{E}[hab] = \frac{\mathrm{Tr}(HAB) + \mathrm{Tr}(HBA)}{N(N+1)(N+2)} = \frac{\mathrm{Tr}\big(H\{A,B\}\big)}{N(N+1)(N+2)}.} \tag{★★}$$

That's the clean estimator.

So:

$$\boxed{\mathrm{Tr}\big(H\{A,B\}\big) = N(N+1)(N+2)\ \overline{\langle H\rangle_u \langle A\rangle_u \langle B\rangle_u}} \qquad \text{(Estimator)}$$

where the overline is the empirical average over random unitaries u (a 3-design / Haar).

And since $\mathrm{Tr}(H\{A,B\}) = \mathrm{Tr}(HAB) + \mathrm{Tr}(HBA) = 2\Re\mathrm{Tr}(HAB)$, this is exactly what you need for the first-order thermal correction.

Concrete protocol steps to estimate it (operational, but math-aligned)

For each random unitary u drawn from a **global unitary 3-design** (Haar/CUE works):

1. **Prepare** $|\psi_u\rangle = u|k_0\rangle$.
2. **Estimate** $a_u = \langle A\rangle_u$ by measuring observable $A = W(t)$ on $|\psi_u\rangle$.
3. **Estimate** $b_u = \langle B\rangle_u$ by measuring observable $B = V^\dagger W(t) V$ on $|\psi_u\rangle$.
4. **Estimate** $h_u = \langle H\rangle_u$ by measuring the energy (or summing measured local terms if $H = \sum_j h_j$).
5. Compute the sample product $h_u a_u b_u$.
6. Average over u : $\overline{hab} = \frac{1}{N_U}\sum_u h_u a_u b_u$.
7. Multiply by $N(N+1)(N+2)$ to obtain $\mathrm{Tr}(H\{A,B\})$ via (Estimator).

That's it.

Where your earlier "Tr(HAB)\mathrm{Tr}(HAB)" request fits

- **You cannot get** $\mathrm{Tr}(HAB)$ **alone** from just the symmetric moment (hab) without additional structure, because the Haar third moment naturally produces $\mathrm{Tr}(HAB) + \mathrm{Tr}(HBA)$.
- But Appendix G **doesn't require** $\mathrm{Tr}(HAB)$ **alone**; it requires the symmetrized combination $\mathrm{Tr}(H\{A,B\})$. That's why the protocol is consistent.

If you want, next I can plug this estimator back into your first-order expansion

$$F_\beta(t) = \frac{1}{N}\mathrm{Tr}(AB) - \frac{\beta}{2N}\mathrm{Tr}(H\{A,B\}) + O(\beta^2),$$

and express $F_\beta(t)$ *entirely* in terms of measurable correlations $\overline{ab}$ and $\overline{hab}$, with the correct prefactors (including the same normalization constants as the main-text global protocol).

OK thanks. Now I understand. let me ask another few questions. Act as sophisticated physicist: no need to agree with me amicably. You can be critical, answer with reasoning. 1. The F_b is defined different to standard OTOC (only equal when b=0). For general b, when one obtained F_b using your protocol (or the protocol in the paper's appendix), do I need some transformation back to std OTOC? Or the F_b is enough to represent chaos itself? 2. The complexity of OTOC is not in W or V, which are often known, local operators; it is the U(t) that is hard, so its generator H, might be complicated and unknown. If this logic is correct, then estimation E=Tr[rho H] might be impossible when one don't know H (say the many body system's Hamiltonian is unknown)? Or we should always known H to form a justified OTOC problem, yet the only hard part is U(t) generated by H.

1) Do you need to "convert" FβF_\beta back to the standard thermal OTOC?

No universal conversion exists, and you usually **shouldn't** try.

What Appendix G uses is the **regularized/symmetrized** (often called "regulated") thermal OTOC

$$F_\beta(t) = \mathrm{Tr}\Big(\rho_\beta^{1/2} W(t)\, \rho_\beta^{1/2}\, V^\dagger W(t) V\Big),$$

not the more naive

$$F_\beta^{\mathrm{std}}(t) = \mathrm{Tr}\big(\rho_\beta\, W(t)\, V^\dagger W(t) V\big).$$

These coincide at $\beta = 0$ (infinite temperature), but for $\beta > 0$ they are **different objects**, and the difference is not a mere rescaling or a simple post-processing map that you can apply *without additional dynamical information*.

More importantly: **the regulated one is not a "weaker proxy."** It is often the *right* object if your goal is "quantum chaos" in the sense of:

- analyticity properties of thermal correlators,
- avoiding UV/ordering pathologies,
- compatibility with the standard derivations of the MSS chaos bound (Lyapunov bound) in thermal QFT/large- N settings.

So if your goal is "scrambling/chaos diagnostic," F_β is typically **enough** and, in many theoretical contexts, **preferable**.

When would you insist on the “standard” version? If you have a very specific operational definition tied to an experimental protocol or to a particular Keldysh contour ordering, or you want to compare to a model/paper that defines the OTOC unregularized. Even then, you don’t “transform” F_β into F_β^{std} — you **measure the one you care about** (or simulate it).

Bottom line: Treat F_β as its own, physically motivated chaos correlator; don’t assume it’s just an approximation to be converted back.

---

2) Doesn’t the protocol become impossible if HH is unknown?

Your logic is mostly right, but it hides an important distinction:

(a) If you truly don’t know HH and cannot measure it, then yes: the Appendix-G estimator is not well-defined.

Their first-order correction needs quantities like $\langle H \rangle_u$ and ultimately $\mathrm{Tr}(H\{A,B\})$. If “ H ” is only an abstract generator that you never characterize, you can’t claim you’re estimating *thermal* corrections of a Gibbs state $e^{-\beta H}$. In that sense, the protocol presupposes access to H.

(b) But in most “OTOC as chaos” settings, HH is not “unknown” in the relevant sense.

Even if $U(t)$ is hard to *simulate classically*, the experiment typically **implements** $U(t) = e^{-iHt}$. That means:

- you know the Hamiltonian model (at least its local term decomposition) well enough to engineer it,
- or you can measure energy by measuring those local terms,
- or you can calibrate/learn an effective Hamiltonian (to the precision required).

So: $U(t)$ being “hard” is a *computational hardness* statement, not “we have no knowledge of the generator.”

(c) If you’re in a regime where the system is a black box with unknown HH, you have options—but they change the problem.

- If you want a **scrambling diagnostic** without knowing H, you can still measure **infinite-temperature** OTOCs (or proxy quantities) via randomized protocols that do *not* require energy measurements.
- If you want **finite-temperature** scrambling relative to some Gibbs state, you’re implicitly committing to a Hamiltonian (or at least to a notion of “energy”), so you must either:
  1. **learn** H (Hamiltonian learning / spectroscopy), or
  2. **prepare the thermal state physically** and measure correlators on it—yet even then, you still need to know what “thermal” means (again: relative to which H ?).

So your critique is valid: Appendix G’s method assumes a degree of Hamiltonian access that the $\beta = 0$ protocol doesn’t.

Bottom line:

- For $\beta = 0$, you can often treat $U(t)$ as an operational black box.
- For $\beta > 0$ in Appendix G’s approach, you *cannot* —you need an “energy operator” you can evaluate, i.e. H (or a trusted effective H).

If you tell me your concrete scenario (analog simulator with known local couplings? noisy device with drift? genuinely unknown many-body system?), I can tell you which of (a)–(c) is the right model and what a realistic workaround looks like.

Sure. I’ll do two things:

1. give you a **clean, checked** β^2 expansion of $\rho^{1/2}$ and $F_\beta(t) = \mathrm{Tr}(\rho^{1/2} A \rho^{1/2} B)$;
2. tell you **what randomized data you must collect** and **which Haar-moment identities** you use to extract the new β^2 traces (and what design order you need).

I’ll keep it general enough to be correct, but I will use the standard harmless gauge choice $\mathrm{Tr}(H) = 0$ to avoid clutter (you can always shift $H \mapsto H - c\mathbb{I}$; it only changes Z , not ρ_β). Same for traceless A, B if you want the cleanest estimators, but I’ll flag where that matters.

---

1) Expand ρβ1/2\rho_\beta^{1/2} to second order

Let

$$\rho_\beta = \frac{e^{-\beta H}}{Z}, \qquad Z = \mathrm{Tr}(e^{-\beta H}), \qquad N = \dim \mathcal{H}.$$

Assume $\mathrm{Tr}(H) = 0$. Then

$$Z = N + \frac{\beta^2}{2}\mathrm{Tr}(H^2) + O(\beta^3),$$

so

$$Z^{-1/2} = N^{-1/2}\Big(1 - \frac{\beta^2}{4N}\mathrm{Tr}(H^2)\Big) + O(\beta^3).$$

Also

$$e^{-\beta H/2} = \mathbb{I} - \frac{\beta}{2}H + \frac{\beta^2}{8}H^2 + O(\beta^3).$$

Multiply them:

$$\boxed{\rho_\beta^{1/2} = \frac{1}{\sqrt{N}}\left[\mathbb{I} - \frac{\beta}{2}H + \frac{\beta^2}{8}H^2 - \frac{\beta^2}{4N}\mathrm{Tr}(H^2)\,\mathbb{I}\right] + O(\beta^3).} \tag{1}$$

Consistency check: squaring (1) gives $\rho_\beta = \frac{1}{N}(\mathbb{I} - \beta H + \frac{\beta^2}{2}H^2 - \frac{\beta^2}{2N}\mathrm{Tr}(H^2)\mathbb{I}) + O(\beta^3)$, which matches expanding $e^{-\beta H}/Z$. ✅

---

2) Expand Fβ=Tr(ρ1/2Aρ1/2B)F_\beta=\mathrm{Tr}(\rho^{1/2}A\rho^{1/2}B) to β2\beta^2

Let $A = W(t)$, $B = V^\dagger W(t) V$ (as in the appendix), but the algebra below is generic.

Write

$$\rho^{1/2} = \frac{1}{\sqrt{N}}(I + \beta K + \beta^2 L) + O(\beta^3)$$

with

$$K = -\frac{1}{2}H, \qquad L = \frac{1}{8}H^2 - \frac{1}{4N}\mathrm{Tr}(H^2)I.$$

Then

$$\rho^{1/2}A\rho^{1/2} = \frac{1}{N}\Big(A + \beta(KA + AK) + \beta^2(LA + AL + KAK)\Big) + O(\beta^3).$$

Compute each term:

- $KA + AK = -\frac{1}{2}(HA + AH) = -\frac{1}{2}\{H, A\}$.
- $KAK = \frac{1}{4}HAH$.
- $LA + AL = \frac{1}{8}\{H^2, A\} - \frac{1}{2N}\mathrm{Tr}(H^2)\,A$.

So

$$\rho^{1/2}A\rho^{1/2} = \frac{1}{N}\left[A - \frac{\beta}{2}\{H, A\} + \beta^2\left(\frac{1}{8}\{H^2, A\} + \frac{1}{4}HAH - \frac{\mathrm{Tr}(H^2)}{2N}A\right)\right] + O(\beta^3).$$

Now trace against B :

$$F_\beta = \frac{1}{N}\mathrm{Tr}(AB) - \frac{\beta}{2N}\mathrm{Tr}(\{H, A\}B) + \frac{\beta^2}{N}\left[\frac{1}{8}\mathrm{Tr}(\{H^2, A\}B) + \frac{1}{4}\mathrm{Tr}(HAHB) - \frac{\mathrm{Tr}(H^2)}{2N}\mathrm{Tr}(AB)\right] + O(\beta^3). \tag{2}$$

Using cyclicity,

$$\mathrm{Tr}(\{H, A\}B) = \mathrm{Tr}(H\{A, B\}), \qquad \mathrm{Tr}(\{H^2, A\}B) = \mathrm{Tr}(H^2\{A, B\}),$$

so a clean “physics” form is

$$\boxed{F_\beta = \frac{1}{N}\mathrm{Tr}(AB) - \frac{\beta}{2N}\mathrm{Tr}\big(H\{A, B\}\big) + \frac{\beta^2}{8N}\mathrm{Tr}\big(H^2\{A, B\}\big) + \frac{\beta^2}{4N}\mathrm{Tr}(HAHB) - \frac{\beta^2}{2N^2}\mathrm{Tr}(H^2)\mathrm{Tr}(AB) + O(\beta^3).} \tag{3}$$

That’s the full β^2 contribution. No handwaving: every term has a clear origin (two-sided dressing + normalization via $Z^{-1/2}$).

Sanity check: if $H = 0$, all corrections vanish and $F_\beta = \frac{1}{N}\mathrm{Tr}(AB)$. ✅

---

3) What randomized protocol learns the β2\beta^2 terms?

What changes relative to the β1\beta^1 appendix protocol?

- β^1 required **third moments** of Haar random states $\Rightarrow$ a **3-design** and correlators like $\overline{\langle H\rangle\langle A\rangle\langle B\rangle}$.
- At β^2 you now have terms involving
  - $\mathrm{Tr}(H^2\{A,B\})$ (still "third-order" in *operators* if you treat H^2 as one operator),
  - $\mathrm{Tr}(HAHB)$ (genuinely "4th-order" structure),
  - $\mathrm{Tr}(H^2)\mathrm{Tr}(AB)$ (product of two 2nd-order invariants).

So you need **up to 4th moments of Haar random states** $\Rightarrow$ **unitary 4-design** (or Haar/CUE).

---

4) Concrete estimators from Haar moment identities

Let $|\psi_u\rangle = u|k_0\rangle$, $\rho_u = |\psi_u\rangle\langle\psi_u|$. For each u , estimate:

- $a_u = \langle A\rangle_u = \mathrm{Tr}(\rho_u A)$
- $b_u = \langle B\rangle_u = \mathrm{Tr}(\rho_u B)$
- $h_u = \langle H\rangle_u = \mathrm{Tr}(\rho_u H)$
- $q_u = \langle H^2\rangle_u = \mathrm{Tr}(\rho_u H^2)$

(Operationally, q_u is obtained by measuring energy eigenvalues and squaring outcomes, or by measuring local terms of H^2 if feasible.)

You will form empirical averages (over random u) of the following products:

(i) 2-design piece: Tr(AB)\mathrm{Tr}(AB) and Tr(H2)\mathrm{Tr}(H^2)

With traceless A, B and traceless H ,

$$\overline{a_u b_u} = \frac{\mathrm{Tr}(AB)}{N(N+1)}, \qquad \overline{h_u^2} = \frac{\mathrm{Tr}(H^2)}{N(N+1)}.$$

These are standard 2nd-moment identities (same spirit as your $k=2$ twirl).

(ii) 3-design piece: Tr(H2{A,B})\mathrm{Tr}(H^2\{A,B\})

Apply the **third-moment identity** (same structure you used for β^1) but with H^2 treated as a single operator:

If $\mathrm{Tr}(A) = \mathrm{Tr}(B) = \mathrm{Tr}(H^2) = 0$ (note: $\mathrm{Tr}(H^2) \neq 0$ generally, so don't assume this), you must use the full identity; but in practice you can remove the trace part by defining the traceless operator

$$\widetilde{H^2} := H^2 - \frac{\mathrm{Tr}(H^2)}{N} I.$$

Then $\mathrm{Tr}(\widetilde{H^2}) = 0$, and you can estimate

$$\mathrm{Tr}(\widetilde{H^2}\{A,B\}) \propto \overline{\langle\widetilde{H^2}\rangle\, a_u\, b_u}.$$

Concretely, since $\langle\widetilde{H^2}\rangle_u = q_u - \frac{\mathrm{Tr}(H^2)}{N}$,
this requires knowing $\mathrm{Tr}(H^2)$, which you already got from $\overline{h_u^2}$.

This yields $\mathrm{Tr}(H^2\{A,B\})$ after adding back the identity part (which contributes $\propto \mathrm{Tr}(H^2)\mathrm{Tr}(\{A,B\})$, and for traceless A, B that extra piece is typically controlled).

(iii) 4-design piece: Tr(HAHB)\mathrm{Tr}(HAHB)

This is the genuinely new difficulty. You need a **4th-moment** identity.

A compact, correct way to state it (no missing terms) is:

$$\overline{\prod_{j=1}^{4}\mathrm{Tr}(\rho_u O_j)} = \frac{1}{N(N+1)(N+2)(N+3)} \sum_{\pi\in S_4} \prod_{c\in\mathrm{cycles}(\pi)} \mathrm{Tr}\Big(\prod_{j\in c} O_j\Big), \tag{4}$$

where O_1, O_2, O_3, O_4 are any operators and the product inside each cycle follows the cycle order.

Now choose

$$(O_1, O_2, O_3, O_4) = (H, A, H, B),$$

so the measurable LHS is

$$\overline{h_u\, a_u\, h_u\, b_u} = \overline{h_u^2\, a_u\, b_u}.$$

Expanding the RHS of (4), and **assuming** $\mathrm{Tr}(H) = \mathrm{Tr}(A) = \mathrm{Tr}(B) = 0$ so all permutations with 1-cycles drop, the surviving contributions are:

- the 2–2 cycle terms:

$$\mathrm{Tr}(H^2)\mathrm{Tr}(AB) \quad + \quad 2\,\mathrm{Tr}(HA)\mathrm{Tr}(HB),$$

- the 4-cycle terms:

$$\mathrm{Tr}(HAHB),\ \mathrm{Tr}(HABH),\ \mathrm{Tr}(H^2AB),\ \mathrm{Tr}(H^2BA),\ \mathrm{Tr}(HBHA),\ \mathrm{Tr}(HBAH).$$

So $\overline{h^2ab}$ gives you a **known linear combination** of six 4-cycle traces plus some "pairing" traces you can also estimate independently (e.g. $\mathrm{Tr}(H^2)$, $\mathrm{Tr}(AB)$, $\mathrm{Tr}(HA)$, $\mathrm{Tr}(HB)$ from 2-design correlations).

Critical point: one measured number $\overline{h^2ab}$ is not enough to isolate $\mathrm{Tr}(HAHB)$ alone, because several 4-cycle invariants appear.

How to isolate Tr(HAHB)\mathrm{Tr}(HAHB) in practice

You measure **multiple fourth-order correlators** corresponding to different orderings of (H, A, H, B) , e.g.

$$\overline{\langle H\rangle\langle A\rangle\langle H\rangle\langle B\rangle}, \quad \overline{\langle H\rangle\langle H\rangle\langle A\rangle\langle B\rangle}, \quad \overline{\langle H\rangle\langle B\rangle\langle H\rangle\langle A\rangle},$$

etc. Each choice corresponds to a different tuple (O_1, O_2, O_3, O_4) and hence a different linear combination of the same underlying trace invariants. Using 2–3 such correlators, you solve a small linear system to extract the specific invariant $\mathrm{Tr}(HAHB)$ (and typically also $\mathrm{Tr}(HBHA)$ and the H^2AB/H^2BA pair, which you need anyway for the $\{A, B\}$ term).

That's the honest answer: β^2 **forces you into 4-design + a linear inversion problem in the permutation-invariant basis**.

---

5) Summary: what you need for the β2\beta^2 upgrade

From (3), you must estimate five quantities:

1. $\mathrm{Tr}(AB)$ — 2-design via $\overline{ab}$
2. $\mathrm{Tr}(H\{A, B\})$ — 3-design via $\overline{hab}$ (what you already mastered)
3. $\mathrm{Tr}(H^2\{A, B\})$ — 3-design via $\overline{q\,a\,b}$ with $q = \langle H^2\rangle$ (plus trace-subtraction)
4. $\mathrm{Tr}(HAHB)$ — 4-design via several 4th-order correlators and linear inversion
5. $\mathrm{Tr}(H^2)$ — 2-design via $\overline{h^2}$

Then plug them into (3).

If you want, I can do the next step "for real": pick a minimal set of 4th-order correlators (say two or three), write out the explicit linear system for the 4-cycle invariants under the assumption $\mathrm{Tr}(H) = \mathrm{Tr}(A) = \mathrm{Tr}(B) = 0$, and show you exactly how $\mathrm{Tr}(HAHB)$ drops out. That's a bit algebra-heavy but completely doable and audit-friendly.

Short, direct, and critical answer:

Operationally, you do *not* measure $\mathrm{Tr}(AB)$ using the physical finite-temperature density matrix.
You measure it using **random pure states**, exactly as in the infinite-temperature protocol. The finite- T density matrix only enters **in the theory**, not in state preparation.

More explicitly:

- The randomized-measurement protocol **never prepares** ρ_β.
- It samples Haar-random pure states $|\psi_u\rangle$ (effectively $\mathrm{DM} = \mathbb{I}/N$ at the level of ensemble averages).
- Haar averages convert products of expectation values into **trace invariants** like $\mathrm{Tr}(AB)$, independent of the actual physical temperature.

The finite-temperature physics is then reconstructed **a posteriori** via the **analytic β -expansion**

$$F_\beta = \frac{1}{N}\mathrm{Tr}(AB) - \frac{\beta}{2N}\mathrm{Tr}(H\{A, B\}) + \frac{\beta^2}{\dots}(\dots),$$

where each coefficient is inferred from *separate randomized estimators*.

So:

- Yes, $\mathrm{Tr}(AB)$ is inferred exactly as if the system were at infinite temperature.
- This is **not a contradiction**: $\mathrm{Tr}(AB)$ is a **Hilbert–Schmidt invariant**, not a thermal expectation value.
- Finite temperature enters only through **how you recombine invariants**, not through the measurement ensemble.

If you tried to *physically* measure $\mathrm{Tr}(AB)$ using ρ_β , the protocol would indeed break — but **that is not what the scheme does**.

Bottom line:

Randomized measurement protocols probe operator algebra (Hilbert-space geometry), not the actual thermal state. Finite-T is reconstructed analytically, not prepared operationally.