# LLM Benchmark Report

Generated by script

July 2, 2025

## Summary Statistics

| Model | Total | Correct | Partial | Incorrect | Errors | Avg. Score |
|---|---|---|---|---|---|---|
| o3-mini | 14 | 11 | 1 | 2 | 0 | 0.82 |

## Problem Details

### Problem 1 (Paper: 2506.22909v1)

#### Problem Statement

Background: Consider an ensemble of $N$ FitzHugh–Nagumo oscillators whose fast variables are denoted $x_i(t)$ $(i = 1, \ldots, N)$. Assume that the interaction between the oscillators is diffusive and global: every oscillator is coupled with equal strength to every other oscillator. The diffusive coupling must (i) be proportional to the difference between the state of an interacting neighbour and the state of the oscillator itself and (ii) contain an overall coupling strength $\sigma > 0$. Furthermore, because each oscillator receives input from $N$ units (including itself), the interaction term must be normalised by the ensemble size $N$.

Task: Using these requirements, derive the explicit analytical expression for the global–diffusive coupling term $f_i(x_1, x_2, \ldots, x_N)$ that enters the activator equation of oscillator $i$.

#### Ground Truth Solution

$$\boxed{f_i(x_1, x_2, \ldots, x_N) = \frac{\sigma}{N}\sum_{j=1}^{N}(x_j - x_i)}$$

#### Model Outputs

**Model: o3-mini (Score: 1.0)**

**Model Solution:**

$$f_i(x_1, x_2, \ldots, x_N) = \sigma\left(\frac{1}{N}\sum_{j=1}^{N} x_j - x_i\right)$$

**Judge's Evaluation:** The ground truth expression is:

$$f_i(x_1, x_2, \ldots, x_N) = \frac{\sigma}{N}\sum_{j=1}^{N}(x_j - x_i)$$

The model-generated expression is:

$$f_i(x_1, x_2, \ldots, x_N) = \sigma\left(\frac{1}{N}\sum_{j=1}^{N} x_j - x_i\right)$$

Let's analyze the two expressions:

- The ground truth sums the differences $x_j - x_i$ over $j$, then multiplies by $\frac{\sigma}{N}$. - The model expression rewrites this as $\sigma$ times the difference between the average $\frac{1}{N}\sum_j x_j$ and $x_i$.

Mathematically, these two are equivalent because:

$$\frac{1}{N}\sum_{j=1}^{N}(x_j - x_i) = \frac{1}{N}\sum_{j=1}^{N}x_j - \frac{1}{N}\sum_{j=1}^{N}x_i = \frac{1}{N}\sum_{j=1}^{N}x_j - x_i$$

(since $x_i$ does not depend on $j$, the sum over $j$ of $x_i$ is $Nx_i$, so dividing by $N$ gives $x_i$).

Therefore, the model-generated expression is mathematically equivalent to the ground truth expression.

**Score: 1**

## Problem 2 (Paper: 2506.22909v1)

### Problem Statement

Background: Consider an ensemble of $N$ FitzHugh–Nagumo oscillators arranged on a ring with periodic boundary conditions. For each node $i\,(=1,\dots,N)$ the diffusive, symmetric, non-local coupling acting on the fast variable $x_i(t)$ is

$$f_i^{(\text{nonloc})}(x_1,\dots,x_N)=\frac{\sigma}{2R}\sum_{j=i-R}^{i+R}\bigl(x_j-x_i\bigr),\qquad 1\le R<\frac{N}{2},\tag{1}$$

where the indices are understood modulo $N$, $R$ is the coupling radius (number of neighbours on either side) and $\sigma>0$ is the coupling strength. Define the global (ensemble) mean

$$\langle x\rangle=\frac{1}{N}\sum_{j=1}^{N}x_j.$$

For the maximal possible radius on a ring of even size, $R=N/2$, every oscillator is linked to all others (“global” coupling).

Task: Starting from Eq. (1), set $R=N/2$ and simplify the coupling term to obtain its explicit global form solely in terms of the states $x_j$ and $x_i$.

### Ground Truth Solution

$$\boxed{f_i^{(\text{global})}(x_1,\dots,x_N)=\frac{\sigma}{N}\sum_{j=1}^{N}(x_j-x_i)}$$

### Model Outputs

### Model: o3-mini (Score: 1.0)

**Model Solution:**

$$f_i^{(\text{nonloc})}=\sigma\left(\langle x\rangle-x_i\right)$$

**Judge’s Evaluation:** The ground truth expression is:

$$f_i^{(\text{global})}(x_1,\dots,x_N)=\frac{\sigma}{N}\sum_{j=1}^{N}(x_j-x_i)$$

The model-generated expression is:

$$f_i^{(\text{nonloc})}=\sigma\left(\langle x\rangle-x_i\right)$$

where $\langle x\rangle=\frac{1}{N}\sum_{j=1}^{N}x_j$ is the average of the $x_j$.
If we expand the ground truth sum:

$$\frac{\sigma}{N}\sum_{j=1}^{N}(x_j-x_i)=\frac{\sigma}{N}\left(\sum_{j=1}^{N}x_j-\sum_{j=1}^{N}x_i\right)=\frac{\sigma}{N}\left(\sum_{j=1}^{N}x_j-Nx_i\right)=\sigma\left(\frac{1}{N}\sum_{j=1}^{N}x_j-x_i\right)=\sigma(\langle x\rangle-x_i)$$

Thus, the model-generated expression is mathematically equivalent to the ground truth expression, differing only in notation and naming of the function.

Therefore, the model-generated expression is completely correct.

Score: 1

## Problem 3 (Paper: 2506.23181v1)

### Problem Statement

Background: Consider an ensemble of $N$ phase oscillators whose dynamics (with vanishing inertia) is governed by

$$\dot{\theta}_i(t) = \omega_i + \frac{K_2}{N^2} \sum_{j=1}^{N} \sum_{l=1}^{N} \sin\bigl(2\theta_j - \theta_l - \theta_i\bigr), \qquad i = 1, \ldots, N.$$

Each natural frequency $\omega_i$ is drawn from the Lorentzian distribution

$$g(\omega) = \frac{1}{\pi\,[\,\omega^2 + 1\,]}.$$

Introduce the family of complex order parameters

$$z_p(t) = r_p(t) e^{i\psi_p(t)} = \frac{1}{N} \sum_{j=1}^{N} e^{ip\theta_j(t)}, \qquad p = 1, 2, \ldots$$

and note that for the present model the mean-field form of the single–oscillator equation is

$$\dot{\theta} = \omega + K_2 r_2 r_1 \sin\bigl(\psi_2 - \psi_1 - \theta\bigr).$$

In the thermodynamic limit $N \to \infty$, let $\rho(\theta, \omega, t)$ be the probability density of oscillators with phase $\theta$ and frequency $\omega$; it obeys the continuity equation

$$\partial_t \rho + \partial_\theta\bigl[\,\rho\dot{\theta}\,\bigr] = 0.$$

The Ott–Antonsen ansatz expands $\rho$ as

$$\rho(\theta, \omega, t) = \frac{g(\omega)}{2\pi}\Bigl[1 + \sum_{n=1}^{\infty}\bigl(\alpha^n e^{in\theta} + \alpha^{*\,n} e^{-in\theta}\bigr)\Bigr],$$

where $|\alpha(\omega, t)| \leq 1$. Substituting this series into the continuity equation and collecting the $e^{i\theta}$ harmonics gives the Ott–Antonsen evolution on the invariant manifold:

$$\dot{\alpha} = -i\omega\,\alpha + \frac{K_2 r_2 r_1}{2}\bigl(e^{-i(\psi_2 - \psi_1)} - \alpha^2 e^{i(\psi_2 - \psi_1)}\bigr).$$

For the Lorentzian $g(\omega)$, the contour-integration identities

$$r_1 e^{-i\psi_1} = \alpha(-i,\, t), \qquad r_2 e^{-i\psi_2} = \alpha^2(-i,\, t)$$

imply $r_2 = r_1^2$ and $\psi_2 = 2\psi_1$. Evaluating the Ott–Antonsen flow at $\omega = -i$ therefore yields a closed autonomous equation for the real amplitude $r_1(t)$.

Task: Derive the differential equation that governs the temporal evolution of the first-order coherence $r_1(t)$ in the thermodynamic limit, and express $\dot{r}_1$ solely as a function of $r_1$ and the coupling constant $K_2$.

### Ground Truth Solution

$$\boxed{\dot{r}_1 = -\,r_1 + \frac{K_2}{2}\bigl(r_1^3 - r_1^5\bigr)}$$

### Model Outputs

### Model: o3-mini (Score: 1.0)

### Model Solution:

$$\dot{r}_1 = -r_1 + \frac{K_2}{2}\, r_1^3\,(1 - r_1^2)$$

**Judge's Evaluation:** The ground truth expression is:

$$\dot{r}_1 = -r_1 + \frac{K_2}{2}(r_1^3 - r_1^5)$$

The model-generated expression is:

$$\dot{r}_1 = -r_1 + \frac{K_2}{2}r_1^3(1 - r_1^2)$$

Expanding the model-generated expression inside the parentheses:

$$r_1^3(1 - r_1^2) = r_1^3 - r_1^5$$

This matches exactly the term inside the parentheses in the ground truth expression. The rest of the terms are identical.

Therefore, the model-generated expression is mathematically equivalent to the ground truth expression.

Score: 1

## Problem 4 (Paper: 2506.23181v1)

### Problem Statement

Background: Consider an infinite population (continuum limit N $\rightarrow$ ) of phase oscillators whose dynamics is governed solely by a three–body Kuramoto interaction with inertia

m +=$\omega + \frac{K_2}{N^2} \sum_{j,l=1}^{N} \sin\big(2\theta_j - \theta_l - \theta\big)$,

where m ¿ 0 is the moment of inertia and the intrinsic frequency  is drawn from a probability density g() (take g() to be any even, normalised function). Introduce the complex order parameters

z$_p = r_p e^{i\psi_p} = \frac{1}{N} \sum_{j=1}^{N} e^{ip\theta_j}, \qquad p = 1, 2.$

In the thermodynamic limit the mean–field form of the equation of motion for a single oscillator reads

m +=$\omega + K_2 r_1 r_2 \sin(\psi_2 - \psi_1 - \theta) \equiv \omega - q \sin\theta$,

with the shorthand q:=K$_2 r_1 r_2. Inthestationaryregimeoscillatorssplitintotwoclasses.(i) Lockedoscillatorssatisfy||q;$
$0 andthefixedpointis = arcsin(/q).(ii) Driftingoscillatorssatisfy|| > qandrunaroundthecircle.$

For any p1 the order parameter can be decomposed into contributions of the locked and drifting sub–ensembles,

r$_p = r_p^\ell + r_p^d, \qquad r_p^\ell = \int_{|\omega| \le q} \cos\big(p\,\theta(\omega)\big)\, g(\omega)\, d\omega.$

Task: Show that the locked contribution r$^\ell_p canbewrittenintheclosedform$

r$^\ell_p = q \int_{-\pi/2}^{\pi/2} \cos\theta\, \cos(p\theta)\, g\big(q\sin\theta\big)\, d\theta.$

(Here p is an integer, usually p=1 or 2 in applications.)

### Ground Truth Solution

$$\boxed{r_p^\ell = q \int_{-\pi/2}^{\pi/2} \cos\theta\, \cos(p\theta)\, g\big(q\sin\theta\big)\, d\theta}$$

### Model Outputs

### Model: o3-mini (Score: 1.0)

### Model Solution:

$$r_p^\ell = q \int_{-\pi/2}^{\pi/2} \cos\theta\, \cos(p\theta)\, g\big(q\sin\theta\big)\, d\theta.$$

**Judge's Evaluation:** The model-generated expression matches the ground truth expression exactly in all mathematical aspects: the integral limits, the integrand, the variables, and the coefficients are all the same. The only difference is the presence of a trailing period in the model expression, which is a punctuation mark and does not affect the mathematical correctness or equivalence of the expression.

Therefore, the model expression is completely correct.

Score: 1

## Problem 5 (Paper: 2506.23496v1)

### Problem Statement

Background:

Consider a finite, reversible chemical reaction network (CRN) with species–index set $\mathcal{S}$ and reaction–index set $\mathcal{R}$. For every reaction $r \in \mathcal{R}$ let

$$r: \qquad r^- \longrightarrow r^+, \qquad \Delta r := r^+ - r^- \in \mathbb{Z}^{\mathcal{S}}$$

denote, respectively, the reactant complex, the product complex, and the stoichiometric change-vector. Let $q \in \mathbb{R}^{\mathcal{S}}_{>0}$ be the current concentration vector and $q^{\mathrm{eq}} \in \mathbb{R}^{\mathcal{S}}_{>0}$ the unique detailed–balance equilibrium in the same stoichiometric compatibility class.

Mass–action kinetics:

$$J_r(q) = k_r\, q^{\,r^-}, \qquad J_r^*(q) = k_r^*\, q^{\,r^+},$$

with the detailed–balance condition

$$k_r \left(q^{\mathrm{eq}}\right)^{r^-} = k_r^* \left(q^{\mathrm{eq}}\right)^{r^+} =: \Phi_r \qquad \forall r \in \mathcal{R}.$$

Stoichiometric matrix $\mathbb{S} \in \mathbb{Z}^{\mathcal{S}\times\mathcal{R}}$ has columns $\Delta r$. Define the (signed) reaction flux difference and the fixed throughput (species) current

$$\delta J_r := J_r - J_r^*, \qquad v_{\mathrm{ext}} := \mathbb{S}\,\delta J \in \mathbb{R}^{\mathcal{S}}.$$

Maintenance cost (entropy–production rate, EPR)

$$\dot{\Sigma}(q) := \frac{1}{2}\sum_{r\in\mathcal{R}} \left(J_r^* - J_r\right)\, \ln\!\left(\frac{J_r^*}{J_r}\right).$$

Throughout, ln acts component-wise and products of vectors are standard inner products: $a^T b$.

Task:

Derive an explicit expression for the maintenance cost in terms of the throughput current and the logarithmic concentration deviation, and show that it can be written solely as a scalar product of species-space vectors; that is, prove that

$$\boxed{\dot{\Sigma}(q) = -\ln\!\left(\frac{q}{q^{\mathrm{eq}}}\right)^{T} v_{\mathrm{ext}}}.$$

### Ground Truth Solution

$$\boxed{\dot{\Sigma}(q) \;=\; -\;\ln\!\left(\frac{q}{q^{\mathrm{eq}}}\right)^{T} v_{\mathrm{ext}}}$$

### Model Outputs

### Model: o3-mini (Score: 1.0)

**Model Solution:**

$$\dot{\Sigma}(q) = -\ln\!\left(\frac{q}{q^{\mathrm{eq}}}\right)^{T} v_{\mathrm{ext}}\,.$$

**Judge's Evaluation:** The model-generated expression is:

$$\dot{\Sigma}(q) = -\ln\left(\frac{q}{q^{\mathrm{eq}}}\right)^{T} v_{\mathrm{ext}}$$

The ground truth expression is:

$$\dot{\Sigma}(q) = -\ln\left(\frac{q}{q^{\mathrm{eq}}}\right)^{T} v_{\mathrm{ext}}$$

Both expressions are mathematically identical. The only difference is in formatting and spacing, which does not affect correctness. The variables and structure match exactly, and the expression is fully equivalent.

Score: 1

## Problem 6 (Paper: 2506.23496v1)

### Problem Statement

Background: Consider a reversible, detailed-balanced chemical reaction network (CRN) $\mathcal{G} = (\mathcal{S}, \mathcal{R})$ with stoichiometric vectors $\Delta r \in \mathbb{R}^{\mathcal{S}}$ and detailed-balance fluxes $\Phi_r > 0$ for every one–way reaction $r \in \mathcal{R}$. For any vector of (small) concentration deviations $\dfrac{q}{q^{\text{eq}}} - \mathbf{e}$ in the linear-response regime, the induced species–current is

$$\dot{q} = -\tfrac{1}{2} \sum_{r \in \mathcal{G}} \Phi_r \, \Delta r \, \Delta r^T \Big( \tfrac{q}{q^{\text{eq}}} - \mathbf{e} \Big).$$

Define the CRN conductance matrix

$$\boxed{C(\mathcal{G}) := \tfrac{1}{2} \sum_{r \in \mathcal{G}} \Phi_r \, \Delta r \, \Delta r^T} \qquad (C(\mathcal{G}) \succeq 0, \; C(\mathcal{G})\mathbf{e} = \mathbf{0}),$$

and its Moore–Penrose pseudoinverse, called the resistance matrix,

$$R(\mathcal{G}) := C(\mathcal{G})^{+}.$$

Let $\mathcal{G}' \subset \mathcal{G}$ be a strict partial network (obtained by removing one or more reactions but retaining all species), with conductance matrix $C(\mathcal{G}')$ and resistance matrix $R(\mathcal{G}')$ defined analogously. Because fewer dyadic terms are summed, $C(\mathcal{G}') \preceq C(\mathcal{G})$.

Throughout, $A \succeq 0$ means that $A$ is positive semidefinite, and $X \preceq Y$ means $Y - X \succeq 0$ (Löwner order).

Task: Show that for every vector $v$ in the image of $C(\mathcal{G}')$ the inequality

$$v^T \big( R(\mathcal{G}') - R(\mathcal{G}) \big) v \; \geq 0$$

holds. Equivalently, on $\operatorname{Im}\big(C(\mathcal{G}')\big)$ the resistance matrices satisfy $R(\mathcal{G}) \preceq R(\mathcal{G}')$. Give the final inequality exactly as a single boxed expression.

### Ground Truth Solution

$$\boxed{v^T \big( R(\mathcal{G}') - R(\mathcal{G}) \big) v \; \geq 0 \quad \text{for all } v \in \operatorname{Im}\big(C(\mathcal{G}')\big)}$$

### Model Outputs

### Model: o3-mini (Score: 0.5)

### Model Solution:

$$v^T \Big( R(\mathcal{G}') - R(\mathcal{G}) \Big) v \geq 0 \quad \text{for all } v \in \operatorname{Im}\big(C(\mathcal{G}')\big) \quad \Longleftrightarrow \quad R(\mathcal{G}) \preceq R(\mathcal{G}').$$

**Judge's Evaluation:** The model-generated expression includes the entire ground truth expression exactly as given, with only minor differences in formatting (e.g., spacing and parentheses style), which are irrelevant to mathematical equivalence. However, the model expression adds an additional equivalence statement "$\Longleftrightarrow R(\mathcal{G}) \preceq R(\mathcal{G}')$" which is not present in the ground truth.

Since the rubric specifies that the final expression must match the ground truth exactly (except for variable names), the addition of this equivalence is an extra statement, not part of the original expression. This means the model expression is not exactly the same as the ground truth.

However, the added equivalence is a correct and standard interpretation of the inequality involving positive semidefinite ordering, so it is not an error but an extension.

Given the rubric, this is more than a minor error or omission; it is an addition. But since the original expression is fully contained and correct, and the addition is mathematically valid, this is a borderline case.

The rubric states partial credit only if the answer nearly matches the ground truth solution, with one minor error or omission. Here, the model expression is the ground truth plus an additional equivalence, which is not an error but an addition.

Therefore, the model expression is not incorrect, but it is not exactly the same as the ground truth.

Given the strict rubric, I would assign a score of 0.5 for being extremely close and correct but not exactly matching the ground truth expression.

Score: 0.5

## Problem 7 (Paper: 2506.23991v1)

### Problem Statement

Background:

Let the phase space be the Cartesian product $M = \Sigma \times C$, where $\Sigma \simeq \mathbb{R}^n$ is coordinatised by $\sigma = (\sigma_1, \ldots, \sigma_n)$ and $C \simeq \mathbb{R}^m$ by $c = (c_1, \ldots, c_m)$. On $M$ there is a Poisson bracket whose Poisson tensor (anchor) in the splitting $T^*M = T^*C \oplus T^*\Sigma$ has the block form

$$\pi^\# = \begin{pmatrix} \{c,c\} & \{c,\sigma\} \\ \{\sigma,c\} & \{\sigma,\sigma\} \end{pmatrix}, \qquad \big(\{c,c\}\big)^* = -\{c,c\},\ \big(\{c,\sigma\}\big)^* = -\{\sigma,c\}.$$

Consider the constraint submanifold

$$\Sigma = \{(\sigma,c) \in M \ :\ c = 0\}.$$

Assume the (possibly singular) matrix $\{c,c\}$ satisfies the kernel–nesting condition

$$\ker\{c,c\} \subset \ker\{\sigma,c\}. \tag{A}$$

For a smooth function $f : \Sigma \to \mathbb{R}$ write $f_\sigma := \partial f/\partial\sigma \in \mathbb{R}^n$ and let $\Delta f/\Delta c \in \mathbb{R}^m$ be any vector that solves

$$\{c,\sigma\}\, f_\sigma + \{c,c\}\, \frac{\Delta f}{\Delta c} = 0. \tag{B}$$

(The solvability of (B) is guaranteed by (A); the solution is determined up to addition of an element of $\ker\{c,c\}$.)

Task:

Using only the data given above, derive the explicit Poisson bracket $\{f,g\}_{\mathrm{PD}}$ on the constraint manifold $\Sigma$ for arbitrary smooth $f, g : \Sigma \to \mathbb{R}$. Express the result solely in terms of $\{c,c\}, \{\sigma,c\}, \{\sigma,\sigma\}$, $f_\sigma, g_\sigma$, and the vectors $\Delta f/\Delta c$, $\Delta g/\Delta c$ defined through (B).

### Ground Truth Solution

$$\boxed{\{f,g\}_{\mathrm{PD}} = f_\sigma^{\mathsf{T}}\, \{\sigma,\sigma\}\, g_\sigma - \Big(\frac{\Delta f}{\Delta c}\Big)^{\mathsf{T}} \{c,c\} \Big(\frac{\Delta g}{\Delta c}\Big)}$$

### Model Outputs

### Model: o3-mini (Score: 0.0)

### Model Solution:

$$\{f,g\}_{\mathrm{PD}} = f_\sigma\, \{\sigma,\sigma\}\, g_\sigma \ +\ f_\sigma\, \{\sigma,c\}\, \frac{\Delta g}{\Delta c} \ +\ \frac{\Delta f}{\Delta c}\, \{c,\sigma\}\, g_\sigma \ +\ \frac{\Delta f}{\Delta c}\, \{c,c\}\, \frac{\Delta g}{\Delta c}\,.$$

**Judge's Evaluation:** The ground truth expression is:

$$\{f,g\}_{\mathrm{PD}} = f_\sigma^{\mathsf{T}}\, \{\sigma,\sigma\}\, g_\sigma - \left(\frac{\Delta f}{\Delta c}\right)^{\mathsf{T}} \{c,c\} \left(\frac{\Delta g}{\Delta c}\right)$$

The model-generated expression is:

$$\{f,g\}_{\mathrm{PD}} = f_\sigma\, \{\sigma,\sigma\}\, g_\sigma + f_\sigma\, \{\sigma,c\}\, \frac{\Delta g}{\Delta c} + \frac{\Delta f}{\Delta c}\, \{c,\sigma\}\, g_\sigma + \frac{\Delta f}{\Delta c}\, \{c,c\}\, \frac{\Delta g}{\Delta c}$$

**Comparison:**

1. The ground truth has two terms: one involving only $\sigma$ derivatives and one involving only $c$ derivatives, with a minus sign between them. 2. The model expression has four terms: the $\sigma-\sigma$ term, two cross terms involving $\{\sigma,c\}$ and $\{c,\sigma\}$, and the $c-c$ term. 3. The model uses plus signs for all terms, while the ground truth has a minus sign before the $c-c$ term. 4. The ground truth expression uses transposes on the vectors $f_\sigma$ and $\frac{\Delta f}{\Delta c}$, which is important for matrix multiplication; the model expression omits these transposes. 5. The model includes cross Poisson brackets $\{\sigma,c\}$ and $\{c,\sigma\}$ terms, which are

absent in the ground truth. 6. The ground truth expression is a simplified or reduced form, possibly assuming the cross Poisson brackets vanish or are zero, or that the variables are independent.

**Conclusion:**

- The model expression is not equivalent to the ground truth; it contains extra terms and has incorrect signs. - The omission of transposes and the presence of cross terms are significant differences. - The sign difference on the last term is also important. - These are more than one minor error.

Therefore, the model expression is **incorrect** relative to the ground truth.

Score: 0

## Problem 8 (Paper: 2506.23991v1)

### Problem Statement

Background:

Consider the ideal two–fluid Maxwell model on the three–torus $\mathbb{T}^3$ with dynamical fields $\big(n_i, \boldsymbol{u}_i, n_e, \boldsymbol{u}_e, \boldsymbol{E}, \boldsymbol{B}\big)$, where

• $n_\sigma$ is the number–density, $m_\sigma$ the mass and $q_\sigma$ the charge ($q_i = -Z_i q_e$) of species $\sigma \in \{i, e\}$; • $\boldsymbol{u}_\sigma$ is the species velocity; • $\boldsymbol{E}$ and $\boldsymbol{B}$ are the electric and magnetic fields, with $\nabla\cdot\boldsymbol{B} = 0$.

Introduce the dimensionless parameter $\epsilon > 0$ and the constants $\epsilon_0, \mu_0$. For arbitrary functionals $F, G$ of the six fields define the (Hamiltonian) Poisson bracket

$$
\begin{aligned}
\{F, G\} = \sum_{\sigma\in\{i,e\}} \int_{\mathbb{T}^3} & -\frac{1}{m_\sigma}(F_{n_\sigma}\nabla\cdot G_{\boldsymbol{u}_\sigma} + F_{\boldsymbol{u}_\sigma}\cdot\nabla G_{n_\sigma}) \\
& + \frac{\nabla\times\boldsymbol{u}_\sigma}{m_\sigma n_\sigma}\cdot\big(F_{\boldsymbol{u}_\sigma}\times G_{\boldsymbol{u}_\sigma}\big) + \frac{q_\sigma}{m_\sigma\epsilon^2\epsilon_0}\big(F_{\boldsymbol{u}_\sigma}\cdot G_{\boldsymbol{E}} - F_{\boldsymbol{E}}\cdot G_{\boldsymbol{u}_\sigma}\big) \\
& + \frac{q_\sigma\boldsymbol{B}}{\epsilon m_\sigma^2 n_\sigma}\cdot\big(F_{\boldsymbol{u}_\sigma}\times G_{\boldsymbol{u}_\sigma}\big)\, d^3x \\
& + \int_{\mathbb{T}^3}\frac{1}{\epsilon\epsilon_0}\big(F_{\boldsymbol{E}}\cdot\nabla\times G_{\boldsymbol{B}} - G_{\boldsymbol{E}}\cdot\nabla\times F_{\boldsymbol{B}}\big)\, d^3x,
\end{aligned} \tag{1}
$$

where $F_\chi$ denotes the functional derivative $\delta F/\delta\chi$.

Impose Gauss'-law as a holonomic constraint and eliminate the electron density:

$$
\boxed{n_e = Z_i n_i + \frac{\epsilon^2\epsilon_0}{q_e}\nabla\cdot\boldsymbol{E}}. \tag{2}
$$

Thus the constrained phase space is the submanifold

$$
N = \big\{(n_i, \boldsymbol{u}_i, \boldsymbol{u}_e, \boldsymbol{E}, \boldsymbol{B}) \mid n_e \text{ given by } (\textbf{??})\big\},
$$

with independent variables $(n_i, \boldsymbol{u}_i, \boldsymbol{u}_e, \boldsymbol{E}, \boldsymbol{B})$. For functionals $F, G$ of these five fields let their electron–density derivatives be understood by the chain rule via (2): $F_{n_e} \equiv (\partial F/\partial n_e)$ etc.

Task:

Using bracket (1) and the constraint (2), compute the Poisson bracket $\{F, G\}_{\text{red}}$ obtained by restricting (1) to the submanifold $N$ (i.e. treat $n_e$ everywhere as the function (2) and discard $\delta/\delta n_e$ variations). Express $\{F, G\}_{\text{red}}$ solely in terms of the independent variables $(n_i, \boldsymbol{u}_i, \boldsymbol{u}_e, \boldsymbol{E}, \boldsymbol{B})$, their functional derivatives, and the constants introduced above.

### Ground Truth Solution

$$
\boxed{
\begin{aligned}
\{F, G\}_{\text{red}} = & \int_{\mathbb{T}^3}\frac{1}{\epsilon\epsilon_0}\big(F_{\boldsymbol{E}}\cdot\nabla\times G_{\boldsymbol{B}} - G_{\boldsymbol{E}}\cdot\nabla\times F_{\boldsymbol{B}}\big)\, d^3x \\
& + \sum_{\sigma\in\{i,e\}}\int_{\mathbb{T}^3}\Big[-\frac{1}{m_\sigma}\big(F_{n_\sigma}\nabla\cdot G_{\boldsymbol{u}_\sigma} + F_{\boldsymbol{u}_\sigma}\cdot\nabla G_{n_\sigma}\big) \\
& \qquad + \frac{\nabla\times\boldsymbol{u}_\sigma}{m_\sigma n_\sigma}\cdot\big(F_{\boldsymbol{u}_\sigma}\times G_{\boldsymbol{u}_\sigma}\big) \\
& \qquad + \frac{q_\sigma}{m_\sigma\epsilon^2\epsilon_0}\big(F_{\boldsymbol{u}_\sigma}\cdot G_{\boldsymbol{E}} - F_{\boldsymbol{E}}\cdot G_{\boldsymbol{u}_\sigma}\big) \\
& \qquad + \frac{q_\sigma\boldsymbol{B}}{\epsilon m_\sigma^2 n_\sigma}\cdot\big(F_{\boldsymbol{u}_\sigma}\times G_{\boldsymbol{u}_\sigma}\big)\Big]\, d^3x,
\end{aligned}
}
$$

with $q_i = -Z_i q_e$, $n_e = Z_i n_i + (\epsilon^2\epsilon_0/q_e)\nabla\cdot\boldsymbol{E}$, and where all functional derivatives of $F, G$ are taken with respect to the independent fields $(n_i, \boldsymbol{u}_i, \boldsymbol{u}_e, \boldsymbol{E}, \boldsymbol{B})$

**Model Outputs**

**Model: o3-mini (Score: 1.0)**

**Model Solution:**

$$
\begin{aligned}
\{F, G\}_{\text{red}} = & -\int_{\mathbb{T}^3} \frac{1}{m_i}\Big[F_{n_i}\, \nabla\cdot G_{\boldsymbol{u}_i} + F_{\boldsymbol{u}_i}\cdot \nabla G_{n_i}\Big]\, d^3x \\
& + \int_{\mathbb{T}^3} \frac{1}{m_i\, n_i}\Big(\nabla\times \boldsymbol{u}_i\Big)\cdot\Big(F_{\boldsymbol{u}_i}\times G_{\boldsymbol{u}_i}\Big)\, d^3x \\
& + \int_{\mathbb{T}^3} \frac{q_i}{m_i \epsilon^2 \epsilon_0}\Big[F_{\boldsymbol{u}_i}\cdot G_{\boldsymbol{E}} - F_{\boldsymbol{E}}\cdot G_{\boldsymbol{u}_i}\Big]\, d^3x \\
& + \int_{\mathbb{T}^3} \frac{q_i\, \boldsymbol{B}}{\epsilon m_i^2\, n_i}\cdot\Big(F_{\boldsymbol{u}_i}\times G_{\boldsymbol{u}_i}\Big)\, d^3x \\
& + \int_{\mathbb{T}^3} \frac{1}{m_e\Big(Z_i n_i + \frac{\epsilon^2\epsilon_0}{q_e}\nabla\cdot \boldsymbol{E}\Big)}\Big(\nabla\times \boldsymbol{u}_e\Big)\cdot\Big(F_{\boldsymbol{u}_e}\times G_{\boldsymbol{u}_e}\Big)\, d^3x \\
& + \int_{\mathbb{T}^3} \frac{q_e}{m_e \epsilon^2 \epsilon_0}\Big[F_{\boldsymbol{u}_e}\cdot G_{\boldsymbol{E}} - F_{\boldsymbol{E}}\cdot G_{\boldsymbol{u}_e}\Big]\, d^3x \\
& + \int_{\mathbb{T}^3} \frac{q_e\, \boldsymbol{B}}{\epsilon m_e^2\Big(Z_i n_i + \frac{\epsilon^2\epsilon_0}{q_e}\nabla\cdot \boldsymbol{E}\Big)}\cdot\Big(F_{\boldsymbol{u}_e}\times G_{\boldsymbol{u}_e}\Big)\, d^3x \\
& + \int_{\mathbb{T}^3} \frac{1}{\epsilon\epsilon_0}\Big[F_{\boldsymbol{E}}\cdot(\nabla\times G_{\boldsymbol{B}}) - G_{\boldsymbol{E}}\cdot(\nabla\times F_{\boldsymbol{B}})\Big]\, d^3x\ .
\end{aligned}
$$

**Judge's Evaluation:** The ground truth expression is a compact form that sums over species $\sigma \in \{i, e\}$, with terms involving $F_{n_\sigma}$, $F_{\boldsymbol{u}_\sigma}$, $G_{n_\sigma}$, $G_{\boldsymbol{u}_\sigma}$, and the electromagnetic fields $\boldsymbol{E}, \boldsymbol{B}$. It includes the electron density expressed as

$$n_e = Z_i n_i + \frac{\epsilon^2 \epsilon_0}{q_e}\nabla\cdot \boldsymbol{E}.$$

The model-generated expression expands the sum explicitly into two separate integrals for ions and electrons, substituting $n_e$ accordingly. It also rearranges the terms but keeps the same structure and coefficients.

Key points of comparison:

1. **Ion terms:** - The model has the ion terms with the correct signs and coefficients:

$$-\frac{1}{m_i}(F_{n_i}\nabla\cdot G_{\boldsymbol{u}_i} + F_{\boldsymbol{u}_i}\cdot\nabla G_{n_i})$$

matches the ground truth. - The vorticity cross product term:

$$\frac{\nabla\times \boldsymbol{u}_i}{m_i n_i}\cdot(F_{\boldsymbol{u}_i}\times G_{\boldsymbol{u}_i})$$

matches exactly. - The electromagnetic coupling terms with $q_i$ and $\boldsymbol{B}$ have the correct coefficients and signs.

2. **Electron terms:** - The model substitutes $n_e = Z_i n_i + \frac{\epsilon^2\epsilon_0}{q_e}\nabla\cdot \boldsymbol{E}$ correctly in denominators. - The structure of the electron terms matches the ground truth, including the signs and coefficients. - The model correctly uses $m_e$ and $q_e$ for electron terms.

3. **Electromagnetic field terms:** - The curl terms involving $F_{\boldsymbol{E}}$ and $G_{\boldsymbol{B}}$ are present with the correct coefficients and signs.

4. **Overall structure:** - The model expression is a fully expanded version of the ground truth sum. - No missing terms or incorrect coefficients are observed. - The signs and factors of $\epsilon, \epsilon_0, m_\sigma, q_\sigma$ are consistent. - The model uses consistent notation and the same variables.

**Conclusion:** The model-generated expression is mathematically equivalent to the ground truth expression, just expanded and with the electron density substitution explicitly written out. There are no errors in signs, coefficients, or terms.

Score: 1

## Problem 9 (Paper: 2506.24079v1)

### Problem Statement

Background:

Let $A'$ and $A$ be finite–dimensional quantum systems and let $R \simeq A'$ be an isomorphic reference system. For any pure state $\psi_{RA'} \in (RA')$ we write

$$\mathcal{N}(\psi_{RA'}) \equiv ({}_R \otimes \mathcal{N}_{A' \to A})(\psi_{RA'}) =: \rho_{RA}.$$

The uniformly–mixing ("trace–to–identity") channel

$$\mathcal{R}^1_{A' \to A} : \; \rho_{A'} \longmapsto (\rho_{A'})\, 1_A$$

acts trivially on $R$, so that

$$({}_R \otimes \mathcal{R}^1)(\psi_{RA'}) = \psi_R \otimes 1_A.$$

For positive operators $\rho, \sigma$ with $(\rho) \subseteq (\sigma)$, the quantum relative entropy is

$$D(\rho \| \sigma) = \big[\rho\big(\log \rho - \log \sigma\big)\big] \, .$$

The (von Neumann) entropy of a bipartite state $\rho_{RA}$ is

$$S(RA)_\rho = -[\rho_{RA} \log \rho_{RA}],$$

and its conditional entropy is

$$S(A|R)_\rho = S(RA)_\rho - S(R)_\rho.$$

For a quantum channel $\mathcal{N}_{A' \to A}$, its channel relative entropy with respect to $\mathcal{R}^1$ is

$$D[\mathcal{N} \| \mathcal{R}^1] = \sup_{\psi_{RA'}} D\big(\mathcal{N}(\psi_{RA'}) \big\| \psi_R \otimes 1_A\big),$$

and the channel entropy is defined by

$$S[\mathcal{N}] := -D[\mathcal{N} \| \mathcal{R}^1].$$

Task: Using only the definitions above, express $S[\mathcal{N}]$ as an optimization over the conditional entropy of the channel outputs and state the resulting formula.

### Ground Truth Solution

$$\boxed{S[\mathcal{N}] = \inf_{\psi_{RA'} \in (RA')} S(A|R)_{\mathcal{N}(\psi_{RA'})}}$$

### Model Outputs

### Model: o3-mini (Score: 1.0)

### Model Solution:

$$S[\mathcal{N}] = \inf_{\psi_{RA'}} S(A|R)_{\mathcal{N}(\psi_{RA'})}.$$

**Judge's Evaluation:** The model-generated expression matches the ground truth expression exactly in mathematical content and structure. The only difference is the omission of the set notation "$\in (RA')$"$inthesubscriptoftheinfimum, butthemodelstillspecifiesthedomainoftheinfimumasover \psi_{RA'}$, which is sufficient and standard notation. The period at the end is a minor formatting difference and does not affect correctness.

Since the model's expression is mathematically equivalent and fully captures the meaning of the ground truth expression, it should receive full credit.

Score: 1

## Problem 10 (Paper: 2506.24079v1)

### Problem Statement

Background: Let $A$ be a quantum system with finite-dimensional Hilbert space $\mathcal{H}_A$ and bounded Hamiltonian $\widehat{H}_A$. 1. For any quantum state $\rho_A$, its von Neumann entropy is $S(\rho_A) = -\operatorname{tr}[\rho_A \log \rho_A]$. 2. The thermal state on $A$ at inverse temperature $\beta$ is $\gamma_A^\beta = \dfrac{e^{-\beta \widehat{H}_A}}{Z_A^\beta}, \qquad Z_A^\beta = \operatorname{tr}[e^{-\beta \widehat{H}_A}]$; the map $\beta \mapsto \langle \widehat{H}_A \rangle_{\gamma^\beta} := \operatorname{tr}[\widehat{H}_A \gamma_A^\beta]$ is strictly decreasing, so for every admissible energy $E$ there is a unique $\beta(E)$ with $\langle \widehat{H}_A \rangle_{\gamma^{\beta(E)}} = E$.

3. A quantum channel is a completely positive, trace-preserving (CPTP) linear map $\mathcal{N}_{A' \to A} : \mathcal{B}(\mathcal{H}_{A'}) \to \mathcal{B}(\mathcal{H}_A)$. Its mean output energy is $\langle \widehat{H}_A \rangle_{\mathcal{N}} := \sup_{\rho_{A'} \in (A')} \operatorname{tr}[\widehat{H}_A \, \mathcal{N}(\rho_{A'})]$.

4. For any reference system $R \simeq A'$, the entropy of the channel $\mathcal{N}_{A' \to A}$ is $S[\mathcal{N}] := \inf_{\psi_{RA'} \in (RA')} S(A|R)_{\mathcal{N}(\psi)} = \inf_{\psi_{RA'}} \big[ S(RA)_{\mathcal{N}(\psi)} - S(R)_\psi \big]$, where $S(A|R)_\rho = S(AR)_\rho - S(R)_\rho$ is the conditional entropy.

5. A replacer channel $\mathcal{R}^\omega_{A' \to A}$ outputs a fixed state $\omega_A$ for every input: $\mathcal{R}^\omega_{A' \to A}(\rho_{A'}) = \omega_A$. Its entropy equals the state entropy: $S[\mathcal{R}^\omega] = S(\omega_A)$.

6. The absolutely thermalizing channel at inverse temperature $\beta$ is the replacer channel $\mathcal{T}^\beta_{A' \to A} = \mathcal{R}^{\gamma^\beta}_{A' \to A}$, always outputting $\gamma_A^\beta$.

Task: For a fixed average energy $E$, determine the maximal possible channel entropy

$$\mathcal{M}(E) := \max_{\mathcal{N}_{A' \to A} \,:\, \langle \widehat{H}_A \rangle_{\mathcal{N}} = E} S[\mathcal{N}],$$

and express it in terms of the thermal state $\gamma_A^{\beta(E)}$ introduced above.

### Ground Truth Solution

$$\boxed{\mathcal{M}(E) = S\big(\gamma_A^{\beta(E)}\big)}$$

### Model Outputs

### Model: o3-mini (Score: 1.0)

**Model Solution:**

$$\mathcal{M}(E) = S\big(\gamma_A^{\beta(E)}\big) = -\operatorname{tr}\Big[\gamma_A^{\beta(E)} \log \gamma_A^{\beta(E)}\Big]$$

**Judge's Evaluation:** The model-generated expression includes the ground truth expression $\mathcal{M}(E) = S\big(\gamma_A^{\beta(E)}\big)$ exactly as given, with the same structure and notation (only minor formatting differences such as spacing and the use of vs. plain M). Additionally, the model provides an explicit formula for the entropy $S$, namely $-\operatorname{tr}[\gamma_A^{\beta(E)} \log \gamma_A^{\beta(E)}]$, which is a correct and standard expression for the von Neumann entropy of a density matrix.

Since the ground truth expression is just the definition of $\mathcal{M}(E)$ as the entropy of $\gamma_A^{\beta(E)}$, and the model's expression matches this exactly and supplements it with a correct explicit formula, the model's expression is mathematically equivalent and fully correct.

Score: 1

## Problem 11 (Paper: 2506.24097v1)

### Problem Statement

Background: Consider an infinite spin-1/2 chain that is invariant under translation by $s$ sites. Denote by $\mathcal{S}$ the one–site right–shift super-operator, $\mathcal{U}(\,\cdot\,) = U^{\dagger}(\,\cdot\,)U$ the Heisenberg-picture single-period propagator produced by a fixed, spatially homogeneous quantum circuit, and $\langle A, B\rangle \equiv 2^{-N}\,\mathrm{tr}(A^{\dagger}B)$ the local Hilbert–Schmidt inner product (for any finite $N$; translation invariance allows $N \to \infty$ at the end).

1. Local basis. Let $\surd = \{\sigma^x, \sigma^y, \sigma^z\}$ and $\surd_0 = \{\} \cup \surd$. For a positive integer $r$ define the set of Pauli strings $\mathcal{P}^{(r)} = \Big\{p_1 \otimes \cdots \otimes p_r \;\Big|\; p_1 \in \surd,\ p_{j\neq 1} \in \surd_0\Big\}$. Every $b \in \mathcal{P}^{(r)}$ is called a local density of (maximal) support $r$.

2. Extensive operators. For a quasi-momentum $k \in [0, 2\pi)$, density $b \in \mathcal{P}^{(r)}$ and an "in-cell" index $m \in \{0, \ldots, s-1\}$ define the extensive operator

$$B_k^{(m,b)} = \sum_{j=0}^{N/s-1} e^{-ikj}\,\mathcal{S}^{\,sj+m}(b) \quad (N \to \infty).$$

Different triples $(k, m, b)$ constitute an orthonormal basis with respect to $\langle\!\langle \cdot, \cdot \rangle\!\rangle = \dfrac{s}{N}\langle \cdot, \cdot \rangle$.

3. Truncation. Assume that during one period the circuit enlarges the support of any operator by at most $\delta r$ sites on each side, and that $\delta r \le s$. Fix an integer $r \ge 1$ and let $\mathcal{U}_k^{(r)}$ be the propagator restricted to the subspace spanned by $\{B_k^{(m,b)} \,|\, b \in \mathcal{P}^{(r)},\ m = 0, \ldots, s-1\}$ (i.e. we apply $\mathcal{U}$ and then project back onto that subspace).

Task: Show that, under the above assumptions ($\delta r \le s$), the matrix elements of the truncated propagator in the extensive–operator basis read

$$\big[\mathcal{U}_k^{(r)}\big]_{(m,b),(m',b')} = \sum_{j=-1}^{1} e^{ikj}\,\big\langle \mathcal{S}^{\,sj+m}(b),\, \mathcal{U}\big(\mathcal{S}^{\,m'}(b')\big)\big\rangle,$$

with $b, b' \in \mathcal{P}^{(r)}$ and $m, m' \in \{0, \ldots, s-1\}$.

### Ground Truth Solution

$$\boxed{\big[\mathcal{U}_k^{(r)}\big]_{(m,b),(m',b')} = \sum_{j=-1}^{1} e^{ikj}\,\Big\langle \mathcal{S}^{\,sj+m}(b),\, \mathcal{U}\big(\mathcal{S}^{\,m'}(b')\big)\Big\rangle}$$

### Model Outputs

### Model: o3-mini (Score: 1.0)

### Model Solution:

$$\big[\mathcal{U}_k^{(r)}\big]_{(m,b),(m',b')} = \sum_{j=-1}^{1} e^{ikj}\,\Big\langle \mathcal{S}^{\,sj+m}(b),\, \mathcal{U}\big(\mathcal{S}^{\,m'}(b')\big)\Big\rangle.$$

**Judge's Evaluation:** The model-generated expression matches the ground truth exactly in all aspects: the summation limits, the exponential factor, the inner product, and the arguments of the operators. The only difference is minor formatting (spacing and the use of  vs. ;), which does not affect mathematical equivalence. Variable names and indices are consistent.

Score: 1

## Problem 12 (Paper: 2506.24097v1)

### Problem Statement

Background: Consider a one–dimensional quantum spin-½ chain at infinite temperature with exactly one conserved U(1) charge, the local $z$-magnetisation density $\sigma^z(x,t)$. On hydrodynamic space–time scales the connected equilibrium correlation function of this density obeys ordinary diffusion with diffusion constant $D > 0$, so that for $t > 0$

$$\langle \sigma^z(x,t)\, \sigma^z(0,0) \rangle = \frac{1}{\sqrt{4\pi D\, t}} \exp\!\Big(-\frac{x^2}{4Dt}\Big). \tag{1}$$

The associated local spin current density $j(x,t)$ is defined through the (continuous) continuity equation

$$\partial_t \sigma^z(x,t) + \partial_x j(x,t) = 0. \tag{2}$$

Task: Using Eqs. (1) and (2), derive an explicit closed-form expression (as a function of $x$, $t$ and $D$) for the connected equilibrium current–current correlation function $\langle j(x,t)\, j(0,0)\rangle$.

### Ground Truth Solution

$$\boxed{\langle j(x,t)\, j(0,0)\rangle = \frac{x^2 - 2Dt}{8\sqrt{\pi D}\, t^{5/2}} \exp\!\Big(-\frac{x^2}{4Dt}\Big)}$$

### Model Outputs

### Model: o3-mini (Score: 1.0)

**Model Solution:**

$$\langle j(x,t)\, j(0,0)\rangle = \frac{1}{\sqrt{4\pi D\, t}} \exp\left(-\frac{x^2}{4Dt}\right) \left(\frac{x^2}{4t^2} - \frac{D}{2t}\right).$$

**Judge's Evaluation:** Let's analyze both expressions carefully.

**Ground Truth:**

$$\langle j(x,t)j(0,0)\rangle = \frac{x^2 - 2Dt}{8\sqrt{\pi D}\, t^{5/2}} \exp\left(-\frac{x^2}{4Dt}\right)$$

**Model-Generated:**

$$\langle j(x,t)j(0,0)\rangle = \frac{1}{\sqrt{4\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right) \left(\frac{x^2}{4t^2} - \frac{D}{2t}\right)$$

—

**Step 1: Simplify the model expression to see if it matches the ground truth**

Rewrite the model expression denominator inside the square root:

$$\sqrt{4\pi Dt} = 2\sqrt{\pi Dt}$$

So the prefactor is:

$$\frac{1}{2\sqrt{\pi Dt}}$$

Multiply this by the bracketed term:

$$\left(\frac{x^2}{4t^2} - \frac{D}{2t}\right)$$

So the full prefactor is:

$$\frac{1}{2\sqrt{\pi Dt}} \left(\frac{x^2}{4t^2} - \frac{D}{2t}\right) = \frac{1}{2\sqrt{\pi Dt}} \left(\frac{x^2}{4t^2} - \frac{D}{2t}\right)$$

Multiply numerator and denominator inside the bracket by $t^{3/2}$ to get a common power of $t$ in the denominator:

$$= \frac{1}{2\sqrt{\pi D t}} \left( \frac{x^2}{4t^2} - \frac{D}{2t} \right) = \frac{1}{2\sqrt{\pi D} t^{1/2}} \left( \frac{x^2}{4t^2} - \frac{D}{2t} \right)$$

Rewrite the terms inside the bracket with a common denominator $4t^2$:

$$\frac{x^2}{4t^2} - \frac{D}{2t} = \frac{x^2}{4t^2} - \frac{2Dt}{4t^2} = \frac{x^2 - 2Dt}{4t^2}$$

So the prefactor becomes:

$$\frac{1}{2\sqrt{\pi D} t^{1/2}} \cdot \frac{x^2 - 2Dt}{4t^2} = \frac{x^2 - 2Dt}{8\sqrt{\pi D} t^{5/2}}$$

This matches exactly the prefactor in the ground truth expression.

The exponential terms are identical.

—

**Conclusion:**

The model-generated expression is mathematically equivalent to the ground truth expression, just written in a different form.

—

**Score: 1**

The model expression is completely correct.

—

**Final answer:**

Score: 1

## Problem 13 (Paper: 2506.24115v1)

### Problem Statement

Background: Consider the quantum double D(G) of a finite group G. Anyons are labelled by pairs (C, ) where

- CG is a conjugacy class, • is an irreducible representation of the centraliser Z(g) of an arbitrary element gC.

For two anyons a=(C,) and b=(C,) the modular matrices are

T = (g)/dim (gC), (1)

$S_{ab} = \frac{1}{|G|} \sum_{g \in C} \sum_{h \in C' gh,hg} (h) (g)^*.(2)$

Here (h) is the character of evaluated at h, —G—=—S—=6, and $_{gh,hg} = 1 when g and h commute, 0 otherwise.$

For G=S fix the ordering of the 8 anyon types as

A=( 1, ),B=( 1, sgn ),C=( 1, 2),

D=( s,rs,sr, $_Z), E = (s, rs, sr, sgn_Z),$

F=( r,r$^2$, ),G=( r,r$^2$, ),H=( r,r$^2$, ),

where , sgn, 2 are the three irreps of S, $_k(r) = e^{2ik/3}(k = 0, 1, 2) are the irreps of Z and_Z, sgn_Z the irreps of Z.$

Task: Using equations (1)–(2) and the above ordering, compute the complete S and T matrices of the D(S) quantum-double phase.

### Ground Truth Solution

$$S = \frac{1}{6}\begin{pmatrix} 1 & 1 & 2 & 3 & 3 & 2 & 2 & 2 \\ 1 & 1 & 2 & -3 & -3 & 2 & 2 & 2 \\ 2 & 2 & 4 & 0 & 0 & -2 & -2 & -2 \\ 3 & -3 & 0 & 3 & -3 & 0 & 0 & 0 \\ 3 & -3 & 0 & -3 & 3 & 0 & 0 & 0 \\ 2 & 2 & -2 & 0 & 0 & 4 & -2 & -2 \\ 2 & 2 & -2 & 0 & 0 & -2 & -2 & 4 \\ 2 & 2 & -2 & 0 & 0 & -2 & 4 & -2 \end{pmatrix}, \qquad T = \mathrm{diag}\big(1,\, 1,\, 1,\, 1,\, -1,\, 1,\, e^{\frac{2\pi i}{3}},\, e^{-\frac{2\pi i}{3}}\big)$$

### Model Outputs

### Model: o3-mini (Score: 0.0)

**Model Solution:**

$$T \;=\; \mathrm{diag}\Big(1,\; 1,\; 1,\; 1,\; -1,\; 1,\; e^{2\pi i/3},\; e^{4\pi i/3}\Big),$$

$$S \;=\; \frac{1}{6}\begin{pmatrix} 1 & 1 & 2 & 3 & 3 & 2 & 2 & 2 \\ 1 & 1 & 2 & 3 & 3 & 2 & 2 & 2 \\ 2 & 2 & 4 & 0 & 0 & -2 & -2 & -2 \\ 3 & 3 & 0 & 3 & -3 & 0 & 0 & 0 \\ 3 & 3 & 0 & -3 & 3 & 0 & 0 & 0 \\ 2 & 2 & -2 & 0 & 0 & 2 & -1 & -1 \\ 2 & 2 & -2 & 0 & 0 & -1 & 2 & -1 \\ 2 & 2 & -2 & 0 & 0 & -1 & -1 & 2 \end{pmatrix}.$$

**Judge's Evaluation:** Let's compare the two expressions carefully.

—

**Comparison of T:**

- Ground truth: $T = \mathrm{diag}\big(1, 1, 1, 1, -1, 1, e^{\frac{2\pi i}{3}}, e^{-\frac{2\pi i}{3}}\big)$

- Model: $T = \mathrm{diag}\big(1, 1, 1, 1, -1, 1, e^{2\pi i/3}, e^{4\pi i/3}\big)$

Note that $e^{-\frac{2\pi i}{3}} = e^{\frac{4\pi i}{3}}$ because $e^{2\pi i} = 1$. So the last two diagonal entries are equivalent.

So the model's $T$ is mathematically equivalent to the ground truth $T$.

—

**Comparison of S:**

Ground truth matrix $S$ (only showing differences):

- Row 2 (indexing from 1): Ground truth: $1, 1, 2, -3, -3, 2, 2, 2$ Model: $1, 1, 2, 3, 3, 2, 2, 2$ The signs of the 4th and 5th elements differ ($\pm 3$).

- Row 4: Ground truth: $3, -3, 0, 3, -3, 0, 0, 0$ Model: $3, 3, 0, 3, -3, 0, 0, 0$ The second element differs in sign.

- Row 5: Ground truth: $3, -3, 0, -3, 3, 0, 0, 0$ Model: $3, 3, 0, -3, 3, 0, 0, 0$ The second element differs in sign.

- Rows 6, 7, 8: Ground truth:

$$\begin{pmatrix} 2 & 2 & -2 & 0 & 0 & 4 & -2 & -2 \\ 2 & 2 & -2 & 0 & 0 & -2 & -2 & 4 \\ 2 & 2 & -2 & 0 & 0 & -2 & 4 & -2 \end{pmatrix}$$

Model:

$$\begin{pmatrix} 2 & 2 & -2 & 0 & 0 & 2 & -1 & -1 \\ 2 & 2 & -2 & 0 & 0 & -1 & 2 & -1 \\ 2 & 2 & -2 & 0 & 0 & -1 & -1 & 2 \end{pmatrix}$$

Here, the entries differ significantly: the ground truth has 4, -2, -2, etc., while the model has 2, -1, -1, etc. These are not minor sign errors but different values.

—

**Summary:**

- The $T$ matrix is correct (equivalent). - The $S$ matrix has multiple sign errors in rows 2, 4, and 5. - The last three rows of $S$ differ in multiple entries, not just minor errors. - These errors are more than one minor error; they are multiple and significant.

—

**Conclusion:**

The model's expression for $T$ is correct, but the expression for $S$ contains multiple errors, both in signs and in numerical values. This is more than one minor error, so the expression is not partially correct.

—

**Score: 0**

## Problem 14 (Paper: 2506.24115v1)

### Problem Statement

Background: Consider the non-Abelian anyon type $D$ of the $D(S_3)$ quantum–double topological phase. Its internal Hilbert space is three–dimensional and will be written in the ordered basis

$$\big\{|s\rangle,\ |rs\rangle,\ |sr\rangle\big\},$$

where the kets indicate the flux label carried by the anyon inside the plaquette ($s, rs, sr \in S_3$). Under the electromagnetic-exchange $\mathbb{Z}_2$ symmetry $\mathcal{G}_{\rm em}$ the basis vectors transform as

$$\begin{aligned}
\mathcal{G}_{\rm em}: \quad |s\rangle &\longmapsto \frac{1}{\sqrt{3}}\Big(|s\rangle + |rs\rangle + |sr\rangle\Big),\\
|rs\rangle &\longmapsto \frac{1}{\sqrt{3}}\Big(|s\rangle + e^{2\pi i/3}|rs\rangle + e^{-2\pi i/3}|sr\rangle\Big),\\
|sr\rangle &\longmapsto \frac{1}{\sqrt{3}}\Big(|s\rangle + e^{-2\pi i/3}|rs\rangle + e^{2\pi i/3}|sr\rangle\Big).
\end{aligned}$$

Task: Express this action as a $3 \times 3$ matrix $\rho^D(\mathcal{G}_{\rm em})$ in the ordered basis $\{|s\rangle, |rs\rangle, |sr\rangle\}$; give the explicit matrix entries.

### Ground Truth Solution

$$\boxed{\rho^D\big(\mathcal{G}_{\rm em}\big) = \frac{1}{\sqrt{3}}\begin{pmatrix}1 & 1 & 1\\ 1 & e^{2\pi i/3} & e^{-2\pi i/3}\\ 1 & e^{-2\pi i/3} & e^{2\pi i/3}\end{pmatrix}}$$

### Model Outputs

### Model: o3-mini (Score: 1.0)

### Model Solution:

$$\rho^D(\mathcal{G}_{\rm em}) = \frac{1}{\sqrt{3}}\begin{pmatrix}1 & 1 & 1\\ 1 & e^{2\pi i/3} & e^{-2\pi i/3}\\ 1 & e^{-2\pi i/3} & e^{2\pi i/3}\end{pmatrix}$$

**Judge's Evaluation:** The model-generated expression matches the ground truth exactly in structure, coefficients, and exponents. The only differences are minor formatting details such as spacing and the use of

$$1mm$$

vs

$$2pt$$

for line spacing, which do not affect the mathematical content or equivalence. The variables and terms are identical, and the matrix entries correspond perfectly.

Score: 1