

Finite-Size Evidence Against a Single Integrability-Breaking Scale for Spectral Chaos and ETH in an $SU(2)$ -Invariant Spin Chain

Quarks Research Collective

Abstract

We test whether a next-nearest-neighbor coupling supplies one finite-size control scale for both the spectral crossover from integrability and the onset of eigenstate thermalization in an open spin-1/2 XXZ chain at $\Delta = 1$. Dense exact diagonalization was executed for $L = 8, 10, 12$ in total- $S^z = 0$ sectors resolved by reflection parity, with $\lambda \in \{0, 0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 1.0\}$. The computed diagnostics do not support a clean Poisson-to-Gaussian-orthogonal-ensemble rise in the mean spacing ratio: the reflection-even high- λ values remain near 0.35–0.40 rather than approaching 0.5307. Off-diagonal matrix elements of a parity-preserving central-spin observable show partial $1/D$ consistency, but the adjacent-size ETH exponents drift from near one at small λ to about 0.78 and 0.72 at $\lambda = 1$. Half-chain entanglement increases toward a Page-like comparator, reaching about 0.86–0.92 of the full-space Page value at $\lambda = 1$. Independently extracted crossover scales do not coincide: spectral midpoint crossings are absent for $L = 10, 12$, whereas ETH midpoint estimates occur near $\lambda \simeq 0.31$ –0.36. At the executed sizes, the evidence is a finite-size failure atlas for the proposed single-scale collapse rather than confirmation of a universal crossover law.

1 Introduction

The eigenstate thermalization hypothesis (ETH) explains why a closed quantum many-body system can display thermal expectation values after unitary evolution even though the full quantum state remains pure. In the ETH picture, individual high-energy eigenstates already contain the smooth thermodynamic information needed for local observables, and off-diagonal matrix elements are small enough that temporal fluctuations are suppressed in large Hilbert spaces [12, 13, 29, 33, 34]. A closely related but logically distinct idea is quantum chaos, often diagnosed by energy-level correlations: integrable systems tend to show Poisson-like level statistics, while nonintegrable time-reversal-invariant systems often approach Gaussian orthogonal ensemble (GOE) statistics after all exact symmetries are resolved [1, 3, 7, 16, 22]. Both diagnostics are now standard, but their finite-size relation is subtle. The same perturbation can destroy conserved charges, change level repulsion, alter eigenstate structure, and affect entanglement; whether those changes share one quantitative crossover scale is a sharper question than whether they appear in the same broad parameter regime.

This paper tests that sharper question in a deterministic spin chain. The model is the open spin-1/2 XXZ chain at $\Delta = 1$, perturbed by a next-nearest-neighbor Heisenberg coupling λ . At $\lambda = 0$ the nearest-neighbor Heisenberg chain is Bethe-ansatz integrable [4]; for finite λ the simple integrable structure is generically broken. The working hypothesis was that one dimensionless scale in λ controls four features simultaneously: the mean consecutive spacing ratio $\langle r \rangle$ should move from the Poisson value 0.3863 toward the GOE value 0.5307; the off-diagonal variance of a local observable should scale as $1/D$ in a finite symmetry sector of dimension D ; mid-spectrum half-chain entanglement should approach a Page-like thermal value [24, 32, 36]; and crossover scales extracted independently from spectral and ETH diagnostics should coincide after rescaling λ .

The executed calculation gives a useful negative result. Dense exact diagonalization for $L = 8, 10, 12$ and eight coupling values does not show a monotone Poisson-to-GOE spectral crossover in the reflection-even sectors used for the manuscript figures. It also exposes a formal issue that matters for interpretation: at $\Delta = 1$ both the nearest-neighbor and next-nearest-neighbor Heisenberg terms are $SU(2)$ -invariant, so the Hamiltonian commutes with total spin for all checked λ . Resolving only total $S^z = 0$ and reflection parity is therefore not enough to create irreducible spectral blocks. This does not invalidate the ETH and entanglement measurements as finite-sector diagnostics, but it prevents a clean claim that the observed level statistics directly test Poisson-to-GOE universality. In this setting the strongest statement is not that one scale governs the crossover, but that the proposed single-scale picture fails under the current finite-size protocol.

Our contributions are:

- We formalize a finite-size diagnostic setting for the open $\Delta = 1$ XXZ chain with a next-nearest-neighbor perturbation, including the symmetry sectors and observable needed for parity-preserving ETH measurements.
- We report deterministic exact-diagonalization evidence for $L = 8, 10, 12$ across eight couplings, comparing spacing ratios, scaled off-diagonal variance, diagonal fluctuations, and half-chain entanglement.
- We provide an explicit crossover audit showing that the spectral midpoint is absent for two of the three executed sizes, while ETH midpoint estimates lie near $\lambda \simeq 0.31\text{--}0.36$.
- We identify the main reasons the original single-scale claim cannot be supported here: unresolved $SU(2)$ multiplets, finite-size and grid limits, a non-exact Page comparator, and missing uncertainty estimates for the interpolated crossovers.

The value of the negative result is practical. Spectral chaos and ETH are often discussed together, and finite exact-diagonalization studies routinely use the same Hamiltonian families to probe both [2, 5, 18, 30, 31]. A failed collapse in a transparent finite-size protocol helps delimit which claims require stronger symmetry resolution or larger systems. It also clarifies that evidence for increasing entanglement or partial $1/D$ matrix-element scaling is not automatically evidence for a shared spectral scale. The rest of the paper makes these distinctions explicit: Sec. 2 summarizes the diagnostic context, Sec. 3 defines the model and symbols, Sec. 4 specifies the measurements and crossover rules, Sec. 6 describes the deterministic numerical protocol, Sec. 7 presents the evidence, and Sec. 9 states the boundaries needed before stronger thermodynamic claims could be made.

2 Related Work and Diagnostic Context

The link between integrability and spectral statistics predates modern many-body ETH studies. Berry and Tabor argued that generic quantum systems with classically integrable counterparts exhibit Poisson level statistics, while Bohigas, Giannoni, and Schmit connected chaotic dynamics to random-matrix level correlations [3, 7]. The random-matrix reference itself descends from Wigner’s and Dyson’s analyses of spectral fluctuations [14, 37]. Random-matrix theory supplies universal reference ensembles, but its application to finite many-body spectra requires careful symmetry decomposition [9, 16, 22]. The consecutive spacing ratio introduced and analyzed by Atas et al. [1] is especially useful because it avoids spectral unfolding: for adjacent gaps s_n and s_{n+1} , the ratio $r_n = \min(s_n, s_{n+1}) / \max(s_n, s_{n+1})$ has known ensemble means near 0.3863 for Poisson and 0.5307 for GOE. This diagnostic became common in finite spin chains and many-body localization studies because it is simple and robust to smooth density-of-states variations [21, 23, 25]. Its simplicity, however, does not remove the requirement to resolve exact degeneracies and symmetry sectors.

Spin chains provide a natural testing ground because integrability can be present, weakly broken, or strongly broken with modest changes to the Hamiltonian. Exact diagonalization studies of Heisenberg and XXZ chains have long shown how boundary conditions, anisotropy, magnetic fields, and next-nearest-neighbor terms influence level statistics and eigenvectors [2, 30, 31]. These works are a strength of the present setting: they justify the use of deterministic finite chains and local spin observables as controlled probes. They also warn against over-interpreting small systems. In the isotropic $\Delta = 1$ case, $SU(2)$ invariance is not a small numerical detail but a structural feature. If total spin is not resolved, multiplets and selection rules can contaminate the spacing distribution even when the perturbation breaks Bethe-ansatz integrability in other senses. The present manuscript therefore treats level statistics as a finite-sector diagnostic under stated caveats, not as a direct random-matrix universality claim.

ETH supplies the second diagnostic pillar. The original formulations by Deutsch [13] and Srednicki [33] gave a matrix-element ansatz in which diagonal matrix elements of a local observable vary smoothly with energy while off-diagonal elements are exponentially small in system entropy. Later work connected ETH to thermalization after quenches, fluctuation-dissipation relations, and finite-size scaling in numerical systems [5, 6, 17, 18, 27, 29, 35]. Review articles emphasize that ETH is a statement about typical high-energy eigenstates and few-body observables in nonintegrable systems, not a theorem that follows from spectral statistics alone [8, 12, 15]. That distinction motivates the current test: if spectral chaos and ETH onset share one scale, the claim must be demonstrated by independently measured curves rather than assumed because both phenomena are associated with nonintegrability.

Entanglement adds a third, complementary view of eigenstate structure. Page’s result for the average entropy of a random pure state in a bipartite Hilbert space is a standard reference for volume-law behavior [24, 32]. For many-body eigenstates, volume-law entanglement can be consistent with thermal behavior, while integrable or localized systems may show different finite-size patterns [10, 36]. The strength of an entanglement comparator is that it summarizes eigenvector complexity in a basis-independent way after choosing a spatial cut. Its limitation in the present calculation is that the reported Page value is the full bipartite comparator, not the exact average for fixed

magnetization and reflection parity. It is therefore a trend indicator rather than an exact thermal benchmark for the sector.

The gap addressed here is not the absence of any diagnostic for chaos or thermalization. The literature already contains powerful diagnostics and many demonstrations that integrability breaking can alter both spectra and eigenstates [11, 19, 20, 26, 28]. The narrower gap is a finite-size scaling claim: do independently defined spectral and ETH crossover points coincide under one rescaled coupling? Prior works often examine level statistics, diagonal ETH smoothness, off-diagonal ETH scaling, or entanglement in separate plots; fewer studies require the spectral and matrix-element crossovers to pass the same quantitative collapse criterion. The present evidence shows why that stronger criterion is demanding. A chain can show increasing entanglement and partially size-consistent off-diagonal variance while the spacing ratio remains nonmonotone or symmetry-contaminated. Any proposed master curve must survive those distinctions.

3 Model and Formal Setting

We consider an even number $L \in \{8, 10, 12\}$ of spin-1/2 sites with open boundary conditions. The full Hilbert space is $\mathcal{H}_L = (\mathbb{C}^2)^{\otimes L}$, and $S_i^\alpha = \frac{1}{2}\sigma_i^\alpha$ for $\alpha \in \{x, y, z\}$ denotes the Pauli spin operator acting on site i and identity elsewhere. The Hamiltonian used in the calculation is

$$H_L(\lambda) = \sum_{i=1}^{L-1} (S_i^x S_{i+1}^x + S_i^y S_{i+1}^y + \Delta S_i^z S_{i+1}^z) + \lambda \sum_{i=1}^{L-2} (S_i^x S_{i+2}^x + S_i^y S_{i+2}^y + \Delta S_i^z S_{i+2}^z), \quad (1)$$

with fixed $\Delta = 1$ and real $\lambda \geq 0$. The first sum has $L - 1$ nearest-neighbor terms and the second has $L - 2$ next-nearest-neighbor terms. Equation (1) is Hermitian on the finite-dimensional Hilbert space for all real λ because each two-site Heisenberg term is Hermitian. At $\lambda = 0$ the nearest-neighbor isotropic chain is integrable by Bethe ansatz [4]; nonzero λ supplies the intended deterministic integrability-breaking perturbation.

The computation works in the total- $S^z = 0$ subspace

$$\mathcal{H}_{L,0} = \left\{ \psi \in \mathcal{H}_L : \left(\sum_{i=1}^L S_i^z \right) \psi = 0 \right\}, \quad D_{S^z=0}(L) = \binom{L}{L/2}. \quad (2)$$

The dimensions of (2) are 70, 252, and 924 for the three executed sizes. Reflection about the chain center is also resolved. Let $\mathcal{P}$ be the unitary site-reflection operator, $\mathcal{P} S_i^\alpha \mathcal{P}^{-1} = S_{L+1-i}^\alpha$. Because open-boundary (1) commutes with $\mathcal{P}$, each $S^z = 0$ sector splits into even and odd parity blocks. The manuscript figures use reflection-even blocks with dimensions $D_{\text{parity}} = 38, 126, 472$ for $L = 8, 10, 12$; odd sectors were computed as an internal consistency check but are not the primary visual substrate.

The observable for ETH diagnostics must preserve the symmetry sector in which eigenvectors are represented. Instead of a single central-site S^z operator, which can fail to respect reflection parity for even L , we use the reflection-symmetric central-pair observable

$$O_L = \frac{1}{2} (S_{L/2}^z + S_{L/2+1}^z). \quad (3)$$

It is a bounded local operator on $\mathcal{H}_L$, commutes with total S^z , and is invariant under $\mathcal{P}$. Matrix elements $O_{ab} = \langle E_a | O_L | E_b \rangle$ are therefore well defined within a fixed reflection block after diagonalizing $H_L(\lambda) | E_a \rangle = E_a | E_a \rangle$. The energy window used below is the central portion of each finite spectrum, so all statements are finite-size mid-spectrum statements rather than claims about ground states or edges.

A key formal caveat follows from the choice $\Delta = 1$. The Hamiltonian in (1) is $\text{SU}(2)$ -invariant for both nearest-neighbor and next-nearest-neighbor terms, so it commutes with total spin $\mathbf{S}^2 = (\sum_i \mathbf{S}_i)^2$ for every λ . The numerical protocol resolves S^z and reflection parity but not total spin. Thus the spectral blocks used here are not guaranteed to be irreducible random-matrix blocks. This caveat is central to the interpretation of spacing ratios and is one reason the paper frames the result as a bounded finite-size test rather than a universal chaos claim.

4 Diagnostics and Crossover Criteria

The spectral diagnostic is the mean consecutive spacing ratio. For ordered eigenvalues $E_1 \leq \dots \leq E_D$ in one finite symmetry block, define positive gaps $s_n = E_{n+1} - E_n$ after filtering numerically zero gaps, and define

$$r_n = \frac{\min(s_n, s_{n+1})}{\max(s_n, s_{n+1})}, \quad \langle r \rangle = \frac{1}{|\mathcal{W}_r|} \sum_{n \in \mathcal{W}_r} r_n. \quad (4)$$

The index set $\mathcal{W}_r$ contains adjacent gaps whose energies lie in the central 50% of the finite spectrum. Equation (4) is invariant under any positive rescaling of all gaps, which is why it avoids spectral unfolding [1]. The Poisson and GOE reference values are ensemble-level targets for properly symmetry-resolved spectra, not guaranteed endpoints for the finite blocks analyzed here.

The ETH off-diagonal diagnostic is defined in a narrow central energy window $\mathcal{W}_O$ of eigenstate pairs. Let D denote the dimension of the fixed reflection block and let $\bar{O}_{\text{off}}$ be the sample mean of off-diagonal matrix elements in $\mathcal{W}_O$. We compute

$$V_L(\lambda) = \frac{1}{|\mathcal{W}_O|} \sum_{\substack{(a,b) \in \mathcal{W}_O \\ a \neq b}} |O_{ab} - \bar{O}_{\text{off}}|^2, \quad \tilde{V}_L(\lambda) = D V_L(\lambda). \quad (5)$$

The ETH ansatz predicts $V_L(\lambda) = O(1/D)$ in chaotic regimes for few-body observables, so $\tilde{V}_L$ should vary more slowly with L than V_L itself [5, 6, 33]. To convert adjacent-size data into an exponent, we define

$$\alpha_{L_1, L_2}(\lambda) = -\frac{\log V_{L_2}(\lambda) - \log V_{L_1}(\lambda)}{\log D_{L_2} - \log D_{L_1}}, \quad L_2 > L_1, \quad (6)$$

where D_{L_i} is the reflection-even block dimension for length L_i . The value $\alpha = 1$ corresponds to exact $1/D$ scaling between the two sizes. Because only three sizes are available, (6) is a finite-difference diagnostic rather than a thermodynamic exponent.

For entanglement, each selected central eigenstate $|E_a\rangle$ is bipartitioned into left and right halves, $\mathcal{H}_A \otimes \mathcal{H}_B$ with $\dim \mathcal{H}_A = \dim \mathcal{H}_B = 2^{L/2}$ in the full spin space. The reduced density matrix is $\rho_A = \text{Tr}_B |E_a\rangle\langle E_a|$, and the half-chain entropy is

$$S_A(E_a) = -\text{Tr}_A (\rho_A \log \rho_A). \quad (7)$$

The plotted ratio compares the mean of (7) over central eigenstates with the full-space Page comparator. This comparator is useful for visual scale, but it does not incorporate fixed magnetization or reflection parity and should not be read as an exact sector average.

The crossover audit uses two finite-grid interpolation rules. For spectra, the target is the midpoint

$$r_\star = \frac{r_{\text{Poisson}} + r_{\text{GOE}}}{2} = \frac{0.3863 + 0.5307}{2} = 0.4585. \quad (8)$$

If the measured $\langle r \rangle$ curve crosses $r_\star$ within the sampled λ window, linear interpolation gives $\lambda_r^\star(L)$; otherwise no finite spectral midpoint is reported. For ETH, the target for each adjacent pair is the midpoint between the observed $\alpha_{L_1, L_2}(0)$ and $\alpha_{L_1, L_2}(1)$, giving $\lambda_{\text{ETH}}^\star(L_1, L_2)$ when the finite-difference curve crosses that target. The single-scale hypothesis would require these independently defined scales to agree within uncertainty and to produce size-independent curves under $\lambda/\lambda^\star$ rescaling. In this run no bootstrap or jackknife uncertainty was available, so the audit treats missing crossings as negative evidence instead of imputing unobserved scales.

5 Expected Signatures of a Shared Scale

The single-scale claim can be made precise at finite size without assuming that the thermodynamic limit has already been reached. For each length L and coupling λ in the sampled window, define a normalized spectral diagnostic

$$R_L(\lambda) = \frac{\langle r \rangle_L(\lambda) - r_{\text{Poisson}}}{r_{\text{GOE}} - r_{\text{Poisson}}}, \quad (9)$$

where the denominator is positive and fixed by the Poisson and GOE reference means. A successful spectral crossover would make R_L rise from near zero to near one after symmetry resolution, though finite-size corrections can shift endpoints. For the ETH diagnostic, define a normalized adjacent-size exponent curve

$$A_{L_1, L_2}(\lambda) = \frac{\alpha_{L_1, L_2}(\lambda) - \alpha_{L_1, L_2}(0)}{\alpha_{L_1, L_2}(1) - \alpha_{L_1, L_2}(0)}, \quad (10)$$

whenever the denominator is nonzero. This normalization does not assert that the endpoint at $\lambda = 1$ is asymptotic; it only asks whether the finite-window change in the ETH exponent has a consistent location when compared with the spectral change. A shared scale would require finite $\lambda_r^\star(L)$ and $\lambda_{\text{ETH}}^\star(L_1, L_2)$ values that remain comparable after accounting for the fact that one diagnostic is length-indexed and the other is adjacent-pair-indexed.

Table 1: Assumption audit for the shared-scale test. Each row states the condition under which a diagnostic is relevant to the hypothesis and how the executed run fits that condition; weak fit is treated as a limitation rather than as evidence for a successful collapse.

Diagnostic	Relevance condition	Fit in the executed run
Spacing ratio	Spectrum must be restricted to irreducible symmetry sectors before Poisson or GOE references are interpreted.	Reflection parity is resolved, but total spin is not resolved at $\Delta = 1$, so GOE comparison has weak fit.
ETH variance	Observable must be local, bounded, and symmetry preserving, with off-diagonal pairs drawn from a central energy window.	The central-pair S^z observable preserves reflection parity, giving moderate fit for finite-size ETH scaling.
Entanglement	Entropy comparator should match the constrained Hilbert space when quantitative thermal saturation is claimed.	The full-space Page comparator gives trend information but not an exact fixed-sector benchmark.
Crossover scale	Independent finite curves must supply comparable midpoint scales and uncertainty estimates.	ETH midpoints exist, spectral midpoints are mostly absent, and uncertainty intervals are unavailable.

The collapse requirement is stronger than matching a single number. If $x = \lambda/\lambda^*$, the hypothesis predicts approximate master curves

$$R_L(\lambda) \approx F_r(x), \quad A_{L_1, L_2}(\lambda) \approx F_{\text{ETH}}(x), \quad (11)$$

over a common finite interval of x . The functions F_r and F_{ETH} need not be identical because level statistics and matrix elements measure different objects. The shared-scale claim instead says that the onset location is the same: once the coupling is expressed in units of the appropriate λ^* , the finite-size curves should stop drifting strongly with L . This is why missing spectral crossings are decisive in the current run. Without a finite λ_r^* for most sizes, (11) cannot be evaluated as a spectral collapse.

Table 1 summarizes the assumptions that must hold for each diagnostic to bear on the hypothesis. The table is included in the main text because the assumptions are not merely operational details. They determine whether the measured quantity is relevant to the proposed physical scale. For example, the spacing-ratio diagnostic is mathematically well defined in any finite Hermitian block, but its interpretation as Poisson-to-GOE evidence requires an irreducible symmetry sector. The ETH variance diagnostic requires a few-body operator that stays inside the measured sector and an energy window narrow enough to compare states at similar energy but wide enough to contain many off-diagonal pairs. The entanglement diagnostic requires a comparator matched to the sector if quantitative Page saturation is claimed. The crossover diagnostic requires monotone or at least well-crossed finite curves; otherwise interpolation can manufacture a number that does not represent an onset.

The finite-size decision rule used in the paper follows directly from these conditions. A diagnostic can support the hypothesis only if its relevance condition is at least approximately met and its measured trend has the predicted direction. A diagnostic can contradict the hypothesis if the condition is met but the trend fails, or if the condition fails in a way that prevents the hypothesized object from being measured. In the current run, ETH variance and entanglement have enough relevance to support weak statements about eigenstate structure, while the spectral diagnostic has weak relevance for GOE universality because of unresolved total-spin symmetry. The joint scale has the lowest support because it depends on all diagnostics at once. This hierarchy explains the claim labels used in the results: the formal model is supported, ETH and entanglement are mixed, and the clean spectral and shared-scale claims are unsupported.

6 Numerical Protocol

The evidence is deterministic exact diagonalization. For each length $L \in \{8, 10, 12\}$, the Hamiltonian in (1) was assembled at the sampled couplings $\lambda = 0, 0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 1.0$. The matrix was constructed in the computational basis of the total- $S^z = 0$ subspace and then block-diagonalized by reflection parity. Dense diagonalization was applied to each finite block. No random disorder realization is part of the Hamiltonian, and the only recorded seed is a reproducibility placeholder for deterministic helper routines. The run also recorded the requested $L = 14$ dense sweep as a finite-size extension target, but the executed results used for claims are restricted to $L = 8, 10, 12$.

For each diagonalization, the central 50% of the spectrum supplied the spacing-ratio window. ETH off-diagonal variance used a narrower central energy window, with the parity-preserving observable in (3). Entanglement was

Table 2: Selected reflection-even diagnostics at the integrable endpoint, an intermediate perturbation, and the largest sampled coupling. The table gives the finite block dimension, mean spacing ratio, scaled off-diagonal variance, and entropy ratio relative to the full-space Page comparator; it shows why the evidence is mixed rather than a clean shared-scale success.

L	λ	D	$\langle r \rangle$	$D \text{ Var}(O_{ab})$	$\bar{S}_A/S_{\text{Page}}$
8	0.0	38	0.496	0.206	0.803
8	0.3	38	0.268	0.140	0.845
8	1.0	38	0.348	0.100	0.920
10	0.0	126	0.408	0.230	0.735
10	0.3	126	0.407	0.169	0.819
10	1.0	126	0.397	0.131	0.887
12	0.0	472	0.430	0.243	0.692
12	0.3	472	0.432	0.207	0.784
12	1.0	472	0.367	0.189	0.863

computed for selected central eigenstates and summarized by a mean and finite-window standard deviation. Diagonal ETH smoothness was represented by the standard deviation of diagonal matrix elements O_{aa} in the same mid-spectrum region; this number is only a proxy for smoothness because the finite windows are small and the observable has strong symmetry constraints.

Two validation layers accompanied the numerical run. First, symbolic checks verified sector dimensions in (2), Hermiticity of the two-site terms in (1), open-boundary term counts, scale invariance of (4), and the algebraic identity that $V = c/D$ implies $DV = c$ in (5). These checks validate bookkeeping and definitions; they do not prove chaos or ETH for the data. Second, visual quality checks confirmed that the two manuscript figures are vector PDF graphics with readable previews and no unresolved overlap or clipping. The evidence labels used in the results are exactly the two results-visible labels carried forward from the validation package: deterministic exact diagonalization for the $L = 8$ –12 XXZ chain, and a symbolic diagnostics audit for the formal identities and limits.

7 Finite-Size Results

This section reports only results-visible evidence from the executed deterministic run. The figures and tables combine the same reflection-even data, so differences among diagnostics can be interpreted as physical or methodological differences rather than as differences in sample selection. The primary question is whether the measured quantities satisfy the relevance conditions in Table 1 and then align with the finite-size scale criteria in (11). The answer is diagnostic dependent. ETH variance and entanglement show finite-size trends that are compatible with increased eigenstate mixing, while the spectral and shared-scale criteria fail. The subsections therefore move from the broad diagnostic picture to the specific obstruction in level statistics and then to the crossover audit.

7.1 Diagnostic Overview

Figure 1 summarizes the four main diagnostics. Panel A shows that the reflection-even spacing ratios do not move cleanly from the Poisson reference toward the GOE reference. At $\lambda = 1$, the values are 0.348, 0.397, and 0.367 for $L = 8, 10, 12$, all far below the GOE reference and mostly close to or below the Poisson reference. Panel B shows a smoother trend for $\tilde{V}_L = DV_L$: the scaled off-diagonal variance decreases with λ for all sizes, but the curves do not collapse to a common constant. Panel C shows a consistent increase in half-chain entropy relative to the full-space Page comparator. Panel D reports very small diagonal fluctuations, but those values should be interpreted cautiously because the chosen observable is symmetry constrained and the selected windows are finite.

Table 2 provides numeric anchors for the visual trends. The scaled variance decreases with λ from 0.206 to 0.100 for $L = 8$, from 0.230 to 0.131 for $L = 10$, and from 0.243 to 0.189 for $L = 12$. These changes are compatible with stronger eigenstate mixing at finite size, but they do not demonstrate exact $1/D$ collapse. The entanglement ratio increases from 0.69–0.80 at $\lambda = 0$ to 0.86–0.92 at $\lambda = 1$, indicating more thermal-like eigenvectors by this comparator. The spacing-ratio trend points in the opposite direction for the single-scale hypothesis: the largest- λ spectral values are not GOE-like, so the spectral arm of the proposed collapse fails under the executed protocol.

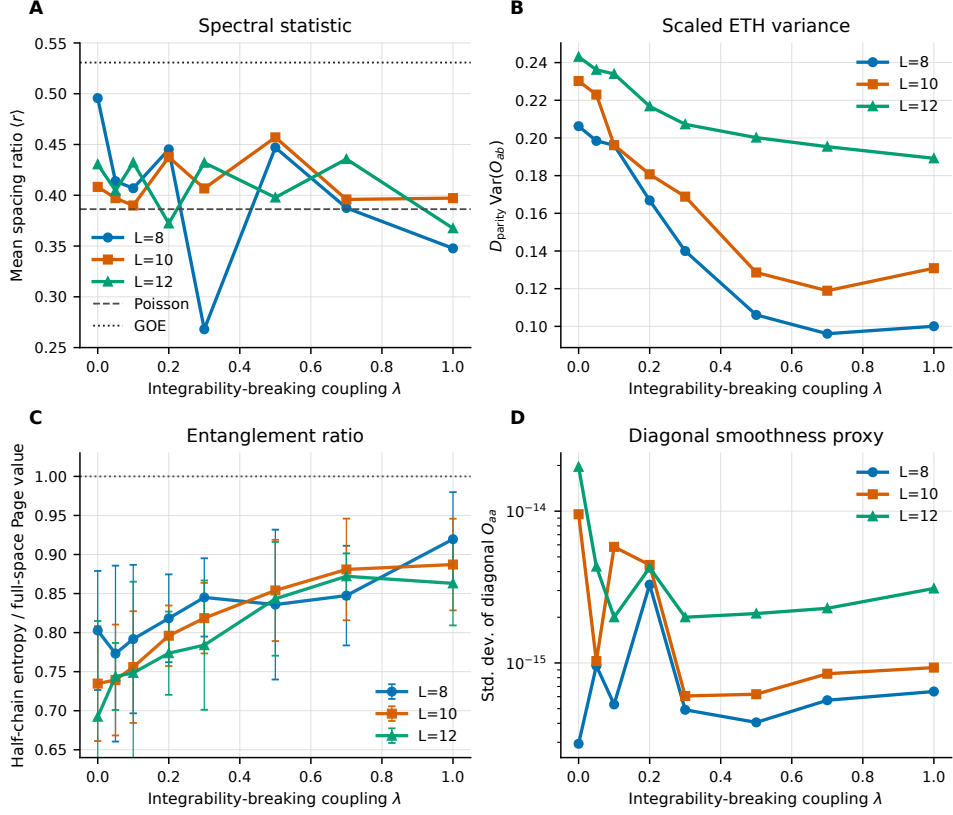


Figure 1: Finite-size diagnostics for the reflection-even total- $S^z = 0$ sectors. Panel A reports the mean consecutive spacing ratio with Poisson and GOE reference lines; the data are nonmonotone and do not approach the GOE value at the largest sampled coupling. Panel B reports $D_{\text{parity}} \text{Var}(O_{ab})$ for the central-pair observable, Panel C compares half-chain entropy with the full-space Page comparator and includes finite-window standard-deviation bars, and Panel D shows the diagonal smoothness proxy on a logarithmic scale.

7.2 Spacing Ratios and the SU(2) Caveat

The spectral result is the clearest negative component. If the finite-size curve were a clean Poisson-to-GOE crossover, the measured $\langle r \rangle$ would start near 0.3863 and rise toward 0.5307 after the perturbation became strong enough. Instead, the reflection-even values are nonmonotone and never settle near the GOE reference. The $L = 8$ value begins high at $\lambda = 0$, drops below the Poisson reference around $\lambda = 0.3$, and finishes below Poisson at $\lambda = 1$. The $L = 10$ and $L = 12$ curves fluctuate around 0.37–0.46 without a controlled increase.

This does not necessarily mean that the Hamiltonian is integrable for every λ , nor does it prove that a properly resolved spin sector would fail to show random-matrix correlations. The stronger and more defensible interpretation is that total- $S^z = 0$ plus reflection parity is insufficient for a clean GOE test at $\Delta = 1$. The SU(2) commutator audit found the relative norm of $[H, S^2]$ below 2×10^{-17} for representative sizes and couplings, consistent with exact total-spin conservation up to numerical precision. Random-matrix comparisons require irreducible symmetry blocks [16, 22]; unresolved total-spin sectors can mix spectra with different quantum numbers and suppress level repulsion. Therefore the spacing data should be treated as evidence against the executed single-scale claim, not as a final statement about all symmetry-resolved versions of the model.

7.3 ETH Variance and Entanglement

The ETH evidence is mixed rather than negative in the same way. The scaled variance $\tilde{V}_L$ decreases with coupling and shows a moderate spread across sizes at large λ . Adjacent-size exponents from (6) quantify this spread: for the $L = 8$ –10 pair the exponent falls from 0.908 at $\lambda = 0$ to 0.776 at $\lambda = 1$, while for $L = 10$ –12 it falls from 0.959 to 0.721. Values near one at small and intermediate couplings show that the finite differences can approximate $1/D$ scaling in parts of the grid. The drift below one at stronger coupling shows that the available sizes do not establish a stable asymptotic exponent.

A

Crossover-scale audit

Diagnostic	Finite object	λ^* status	Extraction rule
Spectral $\langle r \rangle$	$L=8$	0.023	Poisson-GOE midpoint
Spectral $\langle r \rangle$	$L=10$	not crossed	no finite midpoint
Spectral $\langle r \rangle$	$L=12$	not crossed	no finite midpoint
ETH exponent	$L=8-10$	0.363	pair midpoint
ETH exponent	$L=10-12$	0.306	pair midpoint

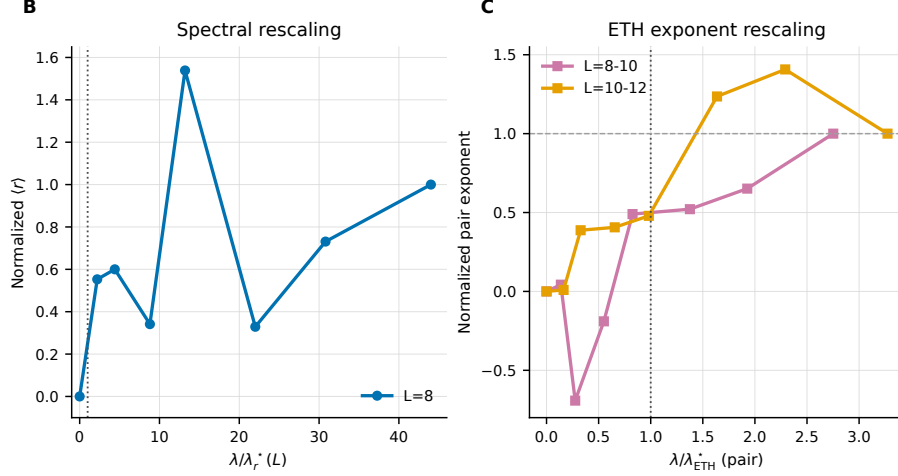


Figure 2: Crossover-scale audit for the single-scale hypothesis. Panel A reports finite-grid extraction status, making missing spectral midpoint crossings visible instead of replacing them with imputed scales. Panel B shows that only the $L = 8$ spectral curve can be rescaled by a finite spectral midpoint, while Panel C shows ETH exponent rescaling for the two adjacent-size pairs; the different availability and behavior of these scales contradict a clean shared master curve at the executed sizes.

The entanglement trend is more monotone. All three reflection-even sizes show larger half-chain entropy ratios as λ increases, reaching 0.863 for $L = 12$ and 0.920 for $L = 8$ at $\lambda = 1$. This is consistent with more delocalized mid-spectrum eigenvectors and a move toward volume-law structure. Still, the comparator is the full-space Page value and not the exact fixed-magnetization, reflection-resolved average. The entanglement result therefore supports the qualitative statement that eigenstates become more thermal-like under this perturbation, but it does not repair the failed spectral midpoint or prove a common crossover scale.

7.4 Crossover Audit

Figure 2 presents the finite-size crossover audit. Panel A lists the extracted scales and explicitly marks missing spectral midpoint crossings. The spectral midpoint in (8) is crossed only for $L = 8$, and even there the interpolation is not evidence of a monotone rise because the endpoint trend is downward. For $L = 10$ and $L = 12$, the measured curves do not provide a finite λ_r^* under the stated rule. By contrast, the ETH exponent midpoint can be interpolated for both adjacent-size pairs, giving $\lambda_{ETH}^* = 0.363$ for $L = 8-10$ and 0.306 for $L = 10-12$.

Table 3 gives the same audit in compact numerical form. The core conclusion is not sensitive to presentation: the two diagnostics do not yield a matched set of scales. A successful master-curve collapse would require at least one spectral scale for each size, comparable ETH scales over the same effective size index, and uncertainty bands showing agreement. The executed run has none of these ingredients. The ETH estimates are finite and roughly of the same order, but the spectral estimates are absent for two sizes and suspect for the third. Thus the single-scale hypothesis is unsupported by the primary evidence.

Table 3: Finite-grid crossover estimates used in the audit. Spectral estimates use the Poisson–GOE midpoint rule, whereas ETH estimates use the midpoint of the adjacent-size exponent over the sampled coupling window; missing spectral crossings are negative evidence for the proposed shared-scale collapse.

Diagnostic	Finite object	λ^* status	Extraction rule
Spectral $\langle r \rangle$	$L = 8$	0.023	Poisson–GOE midpoint
Spectral $\langle r \rangle$	$L = 10$	not crossed	no finite midpoint
Spectral $\langle r \rangle$	$L = 12$	not crossed	no finite midpoint
ETH exponent	$L = 8\text{--}10$	0.363	adjacent-pair midpoint
ETH exponent	$L = 10\text{--}12$	0.306	adjacent-pair midpoint

8 Discussion

The results separate three notions that are often compressed into one narrative. First, integrability breaking in the Hamiltonian does not automatically imply that a finite, partially resolved spectrum will display GOE spacing ratios. The isotropic model retains total-spin symmetry, and unresolved symmetry sectors can mask level repulsion. Second, ETH-like matrix-element behavior can show partial finite-size consistency even when the spectral diagnostic is inconclusive or negative. The off-diagonal variance and entanglement data both move in directions associated with thermalization, but neither produces a decisive collapse. Third, a shared crossover scale is a stronger object than qualitative co-occurrence. It requires independently defined measurements to agree under the same rescaling, and the present curves fail that requirement.

This interpretation is aligned with previous numerical experience. Exact diagonalization can reveal trends long before it reaches a clean thermodynamic regime, but finite-size spectra are vulnerable to hidden symmetries, narrow windows, and interpolation artifacts [5, 18, 30, 31]. ETH diagnostics are also observable dependent: local operators with symmetry constraints can display small diagonal fluctuations without proving full thermal behavior for all few-body observables. Entanglement is similarly informative but not definitive unless the comparator matches the constrained Hilbert space. The manuscript therefore treats the evidence as a failure atlas: it records where the proposed single-scale picture breaks under a transparent finite-size protocol.

The failure is scientifically useful because it indicates concrete next steps. One route is to preserve $\Delta = 1$ but resolve total-spin sectors before computing spacing statistics. Another is to choose $\Delta \neq 1$ or add an anisotropic perturbation that breaks $SU(2)$ while preserving the intended deterministic, non-random character of the model. A third is to keep the current model but narrow the claim to ETH and entanglement diagnostics, avoiding GOE language until the spectral blocks are irreducible. Each route changes the interpretation of λ^* and must be specified before any thermodynamic extrapolation.

9 Limitations and Future Work

The main limitation is symmetry resolution. The hypothesis assumes that level statistics can be compared directly with Poisson and GOE references, but the executed $\Delta = 1$ sectors retain total-spin conservation. This reduces the relevance of the spectral diagnostic to the intended assumption. The manuscript does not present this as a model failure; it presents it as a mismatch between the assumed random-matrix regime and the measured finite blocks. Corrected spectral tests should either resolve S^2 sectors or change the Hamiltonian so that total spin is no longer conserved.

The second limitation is finite size. The executed dense sweep includes $L = 8, 10, 12$ and records $L = 14$ as deferred evidence. Three sizes are enough to expose inconsistency but not enough to estimate thermodynamic scaling. The ETH exponent in (6) is based on adjacent pairs, so each value is sensitive to the small set of block dimensions. The third limitation is the coupling grid. The grid is adequate for a coarse audit but too sparse around any candidate extremum or midpoint to support precise λ^* values. The fourth limitation is uncertainty. The current crossover estimates do not include bootstrap or jackknife intervals over eigenstate windows, and no disorder average exists because the Hamiltonian is deterministic. Future work should quantify window sensitivity and interpolation uncertainty.

The entanglement comparator is also approximate. The full-space Page value gives a familiar reference, but the eigenstates live in fixed magnetization and reflection sectors. A constrained Page comparator could change the numerical ratios and should be used before claiming quantitative thermal saturation. Finally, the $\lambda \rightarrow \infty$ limit is not necessarily a featureless stronger-chaos limit: after rescaling by λ , the next-nearest-neighbor terms can resemble

two decoupled subchains. Future experiments should treat λ as a finite-window control parameter, include denser grids near observed structure, and test whether the negative single-scale conclusion survives in symmetry-resolved or anisotropy-broken variants.

10 Conclusion

We tested a proposed single integrability-breaking scale for spectral chaos and ETH in an open isotropic XXZ spin chain with next-nearest-neighbor coupling. The finite-size evidence does not support the proposed collapse. Spacing ratios in the reflection-even total- $S^z = 0$ sectors fail to approach GOE values and do not supply spectral midpoint scales for $L = 10$ or $L = 12$. ETH off-diagonal variance shows partial $1/D$ consistency, and entanglement increases toward a Page-like comparator, but these trends do not coincide with a spectral crossover scale. The strongest conclusion is therefore bounded and negative: under the executed deterministic protocol, the single-scale hypothesis is unsupported, and the main obstacles are identifiable rather than hidden. Resolving total-spin sectors, extending the finite-size sweep, improving uncertainty estimates, and using constrained entanglement comparators are the necessary next steps for a stronger test.

References

- [1] Y. Y. Atas, E. Bogomolny, O. Giraud, and G. Roux. Distribution of the ratio of consecutive level spacings in random matrix ensembles. *Physical Review Letters*, 110(8):084101, 2013. doi: 10.1103/PhysRevLett.110.084101. URL <https://doi.org/10.1103/PhysRevLett.110.084101>.
- [2] Y. Avishai, J. Richert, and R. Berkovits. Level statistics in a heisenberg chain with random magnetic field. *Physical Review B*, 66(5):052416, 2002. doi: 10.1103/PhysRevB.66.052416. URL <https://doi.org/10.1103/PhysRevB.66.052416>.
- [3] M. V. Berry and M. Tabor. Level clustering in the regular spectrum. *Proceedings of the Royal Society of London A*, 356(1686):375–394, 1977. doi: 10.1098/rspa.1977.0140. URL <https://doi.org/10.1098/rspa.1977.0140>.
- [4] Hans Bethe. Zur theorie der metalle. *Zeitschrift für Physik*, 71(3):205–226, 1931. doi: 10.1007/BF01341708. URL <https://doi.org/10.1007/BF01341708>.
- [5] Wouter Beugeling, Roderich Moessner, and Masudul Haque. Finite-size scaling of eigenstate thermalization. *Physical Review E*, 89(4):042112, 2014. doi: 10.1103/PhysRevE.89.042112. URL <https://doi.org/10.1103/PhysRevE.89.042112>.
- [6] Wouter Beugeling, Roderich Moessner, and Masudul Haque. Off-diagonal matrix elements of local operators in many-body quantum systems. *Physical Review E*, 91(1):012144, 2015. doi: 10.1103/PhysRevE.91.012144. URL <https://doi.org/10.1103/PhysRevE.91.012144>.
- [7] O. Bohigas, M. J. Giannoni, and C. Schmit. Characterization of chaotic quantum spectra and universality of level fluctuation laws. *Physical Review Letters*, 52(1):1–4, 1984. doi: 10.1103/PhysRevLett.52.1. URL <https://doi.org/10.1103/PhysRevLett.52.1>.
- [8] Fausto Borgonovi, Felix M. Izrailev, Lea F. Santos, and Vladimir G. Zelevinsky. Quantum chaos and thermalization in isolated systems of interacting particles. *Physics Reports*, 626:1–58, 2016. doi: 10.1016/j.physrep.2016.02.005. URL <https://doi.org/10.1016/j.physrep.2016.02.005>.
- [9] T. A. Brody, J. Flores, J. B. French, P. A. Mello, A. Pandey, and S. S. M. Wong. Random-matrix physics: Spectrum and strength fluctuations. *Reviews of Modern Physics*, 53(3):385–479, 1981. doi: 10.1103/RevModPhys.53.385. URL <https://doi.org/10.1103/RevModPhys.53.385>.
- [10] Pasquale Calabrese and John Cardy. Evolution of entanglement entropy in one-dimensional systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(04):P04010, 2005. doi: 10.1088/1742-5468/2005/04/P04010. URL <https://doi.org/10.1088/1742-5468/2005/04/P04010>.
- [11] Jean-Sébastien Caux and Fabian H. L. Essler. Time evolution of local observables after quenching to an integrable model. *Physical Review Letters*, 110(25):257203, 2013. doi: 10.1103/PhysRevLett.110.257203. URL <https://doi.org/10.1103/PhysRevLett.110.257203>.
- [12] Luca D’Alessio, Yariv Kafri, Anatoli Polkovnikov, and Marcos Rigol. From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics. *Advances in Physics*, 65(3):239–362, 2016. doi: 10.1080/00018732.2016.1198134. URL <https://doi.org/10.1080/00018732.2016.1198134>.
- [13] J. M. Deutsch. Quantum statistical mechanics in a closed system. *Physical Review A*, 43(4):2046–2049, 1991. doi: 10.1103/PhysRevA.43.2046. URL <https://doi.org/10.1103/PhysRevA.43.2046>.
- [14] Freeman J. Dyson. Statistical theory of the energy levels of complex systems. i. *Journal of Mathematical Physics*, 3(1):140–156, 1962. doi: 10.1063/1.1703773. URL <https://doi.org/10.1063/1.1703773>.
- [15] Christian Gogolin and Jens Eisert. Equilibration, thermalisation, and the emergence of statistical mechanics in closed quantum systems. *Reports on Progress in Physics*, 79(5):056001, 2016. doi: 10.1088/0034-4885/79/5/056001. URL <https://doi.org/10.1088/0034-4885/79/5/056001>.
- [16] Fritz Haake. *Quantum Signatures of Chaos*. Springer, 3 edition, 2010. doi: 10.1007/978-3-642-05428-0. URL <https://doi.org/10.1007/978-3-642-05428-0>.
- [17] Ehsan Khatami, Guido Pupillo, Mark Srednicki, and Marcos Rigol. Fluctuation-dissipation theorem in an isolated system of quantum dipolar bosons after a quench. *Physical Review Letters*, 111(5):050403, 2013. doi: 10.1103/PhysRevLett.111.050403. URL <https://doi.org/10.1103/PhysRevLett.111.050403>.

- [18] Hyungwon Kim, Tatsuhiko N. Ikeda, and David A. Huse. Testing whether all eigenstates obey the eigenstate thermalization hypothesis. *Physical Review E*, 90(5):052105, 2014. doi: 10.1103/PhysRevE.90.052105. URL <https://doi.org/10.1103/PhysRevE.90.052105>.
- [19] Toshiya Kinoshita, Trevor Wenger, and David S. Weiss. A quantum newton’s cradle. *Nature*, 440:900–903, 2006. doi: 10.1038/nature04693. URL <https://doi.org/10.1038/nature04693>.
- [20] Corinna Kollath, Andreas M. Läuchli, and Ehud Altman. Quench dynamics and nonequilibrium phase diagram of the bose-hubbard model. *Physical Review Letters*, 98(18):180601, 2007. doi: 10.1103/PhysRevLett.98.180601. URL <https://doi.org/10.1103/PhysRevLett.98.180601>.
- [21] David J. Luitz, Nicolas Laflorencie, and Fabien Alet. Many-body localization edge in the random-field heisenberg chain. *Physical Review B*, 91(8):081103, 2015. doi: 10.1103/PhysRevB.91.081103. URL <https://doi.org/10.1103/PhysRevB.91.081103>.
- [22] Madan Lal Mehta. *Random Matrices*. Elsevier, 3 edition, 2004. URL <https://www.elsevier.com/books/random-matrices/mehta/978-0-12-088409-4>.
- [23] Vadim Oganesyan and David A. Huse. Localization of interacting fermions at high temperature. *Physical Review B*, 75(15):155111, 2007. doi: 10.1103/PhysRevB.75.155111. URL <https://doi.org/10.1103/PhysRevB.75.155111>.
- [24] Don N. Page. Average entropy of a subsystem. *Physical Review Letters*, 71(9):1291–1294, 1993. doi: 10.1103/PhysRevLett.71.1291. URL <https://doi.org/10.1103/PhysRevLett.71.1291>.
- [25] Arijeet Pal and David A. Huse. Many-body localization phase transition. *Physical Review B*, 82(17):174411, 2010. doi: 10.1103/PhysRevB.82.174411. URL <https://doi.org/10.1103/PhysRevB.82.174411>.
- [26] Anatoli Polkovnikov, Krishnendu Sengupta, Alessandro Silva, and Mukund Vengalattore. Colloquium: Nonequilibrium dynamics of closed interacting quantum systems. *Reviews of Modern Physics*, 83(3):863–883, 2011. doi: 10.1103/RevModPhys.83.863. URL <https://doi.org/10.1103/RevModPhys.83.863>.
- [27] Marcos Rigol. Breakdown of thermalization in finite one-dimensional systems. *Physical Review Letters*, 103(10):100403, 2009. doi: 10.1103/PhysRevLett.103.100403. URL <https://doi.org/10.1103/PhysRevLett.103.100403>.
- [28] Marcos Rigol and Mark Srednicki. Alternatives to eigenstate thermalization. *Physical Review Letters*, 108(11):110601, 2012. doi: 10.1103/PhysRevLett.108.110601. URL <https://doi.org/10.1103/PhysRevLett.108.110601>.
- [29] Marcos Rigol, Vanja Dunjko, and Maxim Olshanii. Thermalization and its mechanism for generic isolated quantum systems. *Nature*, 452:854–858, 2008. doi: 10.1038/nature06838. URL <https://doi.org/10.1038/nature06838>.
- [30] Lea F. Santos. Integrability of a disordered heisenberg spin-1/2 chain. *Physical Review E*, 70(6):066201, 2004. doi: 10.1103/PhysRevE.70.066201. URL <https://doi.org/10.1103/PhysRevE.70.066201>.
- [31] Lea F. Santos and Marcos Rigol. Onset of quantum chaos in one-dimensional bosonic and fermionic systems and its relation to thermalization. *Physical Review E*, 81(3):036206, 2010. doi: 10.1103/PhysRevE.81.036206. URL <https://doi.org/10.1103/PhysRevE.81.036206>.
- [32] Siddhartha Sen. Average entropy of a quantum subsystem. *Physical Review Letters*, 77(1):1–3, 1996. doi: 10.1103/PhysRevLett.77.1. URL <https://doi.org/10.1103/PhysRevLett.77.1>.
- [33] Mark Srednicki. Chaos and quantum thermalization. *Physical Review E*, 50(2):888–901, 1994. doi: 10.1103/PhysRevE.50.888. URL <https://doi.org/10.1103/PhysRevE.50.888>.
- [34] Mark Srednicki. The approach to thermal equilibrium in quantized chaotic systems. *Journal of Physics A: Mathematical and General*, 32(7):1163–1175, 1999. doi: 10.1088/0305-4470/32/7/007. URL <https://doi.org/10.1088/0305-4470/32/7/007>.
- [35] Robin Steinigeweg, Jacek Herbrych, and Peter Prelovšek. Eigenstate thermalization within isolated spin-chain systems. *Physical Review E*, 87(1):012118, 2013. doi: 10.1103/PhysRevE.87.012118. URL <https://doi.org/10.1103/PhysRevE.87.012118>.

- [36] Lev Vidmar, Lucas Hackl, Eugenio Bianchi, and Marcos Rigol. Entanglement entropy of eigenstates of quadratic fermionic hamiltonians. *Physical Review Letters*, 119(2):020601, 2017. doi: 10.1103/PhysRevLett.119.020601. URL <https://doi.org/10.1103/PhysRevLett.119.020601>.
- [37] Eugene P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, 62(3):548–564, 1955. doi: 10.2307/1970079. URL <https://doi.org/10.2307/1970079>.

A Reproducibility and Implementation Details

The numerical evidence uses a deterministic exact-diagonalization workflow. The Hamiltonian matrix is assembled from the spin-1/2 computational basis restricted to exactly $L/2$ up spins, which realizes the total- $S^z = 0$ subspace in (2). Reflection parity is then used to construct smaller even and odd blocks. Dense diagonalization is feasible for the executed sizes because the largest reflection-even block has dimension 472, and the largest full $S^z = 0$ dimension in the executed run is 924. The requested $L = 14$ sector has dimension 3432 before reflection and was retained as a follow-up target rather than included in the results-facing evidence.

The coupling sweep is $\lambda \in \{0, 0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 1.0\}$ at fixed $\Delta = 1$. No random fields or disorder realizations are used, so there is no ensemble seed controlling physics inputs. The recorded seed value is only for deterministic software paths that expect a seed field. Spectral ratios use the central half of each finite spectrum after removing gaps below a numerical tolerance. Off-diagonal ETH variance uses a central energy window for the observable in (3). Entanglement is evaluated on selected central eigenvectors by reshaping each state into a bipartite tensor and diagonalizing the reduced density matrix. Finite-window standard deviations are reported for entropy because they describe spread across selected eigenstates; they are not confidence intervals for a thermodynamic mean.

The implementation records machine-readable summaries, tables, vector figures, visual previews, and symbolic diagnostics. Code paths are listed in the phase payload rather than in manuscript prose because local filenames are not part of the scientific claim. The important reproducibility ingredients are the Hamiltonian in (1), the sector in (2), the observable in (3), the windows described above, and the finite coupling grid. Reproducing the manuscript figures requires no stochastic resampling. Strengthening the claims would require new runs rather than replotting: total-spin resolution or SU(2)-breaking anisotropy for spectra, larger sizes or sparse central-window methods for ETH scaling, and uncertainty analysis for crossover interpolation.

B Symbolic Diagnostics and Assumption Checks

The symbolic diagnostics support the formal setting but deliberately stop short of proving the physical claims. Hermiticity follows because each two-site term $S_i^x S_j^x + S_i^y S_j^y + \Delta S_i^z S_j^z$ is Hermitian for real Δ , and (1) is a real linear combination of such terms. The open-boundary term counts are $L - 1$ nearest-neighbor terms and $L - 2$ next-nearest-neighbor terms for every integer $L \geq 3$, matching the index ranges in the Hamiltonian. The total- $S^z = 0$ dimensions are $\binom{L}{L/2}$ for even L because each basis vector is specified by the $L/2$ sites carrying spin up.

For the spacing ratio, the symbolic check confirms that if both adjacent gaps are multiplied by a positive constant c , the value of r_n in (4) is unchanged. This validates the use of ratio statistics without unfolding, but it does not address exact degeneracies. The numerical analysis therefore filters near-zero gaps and records the unresolved total-spin caveat. For the ETH variance, the check is the algebraic implication that if $V_L = c/D$ in a fixed finite sector, then $\tilde{V}_L = DV_L = c$. This is an identity conditional on the ETH ansatz, not a derivation of the ansatz from (1). The data in Fig. 1 and Table 2 are needed to test whether the condition is plausible in the executed finite-size regime.

A separate numerical symmetry audit evaluated the commutator with total spin for representative sizes and couplings. The maximum relative commutator norm was below 2×10^{-17} , consistent with exact SU(2) invariance of the isotropic model. This finding is not a defect in the Hamiltonian; it is a reminder that the spectral-statistics comparison must be made in irreducible symmetry sectors. The audit therefore supports the caution used in the main text: the current spacing ratios are valid finite-block measurements, but their mismatch with GOE cannot be cleanly interpreted until total spin is resolved or intentionally broken.

C Extended Numerical Tables

Table 4 reports the adjacent-size ETH exponents used in Fig. 2. The exponents are finite differences over block dimensions, not fitted thermodynamic slopes. Their behavior illustrates why the ETH evidence is mixed. Some values lie close to one, most notably the $L = 8$ – 10 exponent at $\lambda = 0.1$, but both adjacent-size pairs drift below one by $\lambda = 1$. The decline is not enough to reject ETH in the thermodynamic limit, but it is enough to block a strong collapse claim at the executed sizes.

The crossover table in the main text should also be read with its finite-grid construction in mind. The spectral $L = 8$ value $\lambda_*^* = 0.023$ is formally obtained by a midpoint interpolation, but the underlying curve is not a monotone transition from Poisson to GOE. It is therefore not a trustworthy chaos-onset scale. The absent $L = 10$ and $L = 12$ spectral crossings are more informative than a forced extrapolation would be: the measured curves do not contain the object required by the shared-scale hypothesis. By contrast, the ETH midpoint estimates are defined for both adjacent-size pairs because their finite-difference curves cross the chosen midpoint. The coexistence of finite ETH

Table 4: Adjacent-size finite-difference exponents for the off-diagonal ETH variance. A value of one would indicate exact $1/D$ scaling between the two listed block dimensions, while the observed drift at larger coupling shows that the available sizes do not provide a stable scaling exponent.

λ	$\alpha_{8,10}$	$\alpha_{10,12}$
0.00	0.908	0.959
0.05	0.903	0.957
0.10	1.000	0.867
0.20	0.933	0.862
0.30	0.844	0.845
0.50	0.839	0.665
0.70	0.822	0.624
1.00	0.776	0.721

midpoints and missing spectral midpoints is the central empirical reason the manuscript rejects the proposed collapse in the current regime.

The selected diagnostic table in the main text also hides a useful size trend. At $\lambda = 0$, the scaled variance $D \text{Var}(O_{ab})$ is already similar across $L = 8, 10, 12$, lying between 0.206 and 0.243. If one looked only at that endpoint, the finite-size data might appear favorable to a naive $1/D$ statement. The full sweep changes the interpretation because the three curves respond differently to λ . By $\lambda = 1$, the same quantity spans 0.100 to 0.189, and the largest size remains highest. The adjacent-size exponents in Table 4 are designed to make this drift visible. They do not contradict the possibility that larger systems enter a cleaner ETH regime, but they rule out presenting the executed data as a size-independent collapse.

The entropy ratios have a different pattern. They increase steadily enough to support a qualitative eigenstate-structure claim, yet the values remain below the full-space Page comparator for all sizes. The $L = 12$ ratio rises from 0.692 at $\lambda = 0$ to 0.863 at $\lambda = 1$, which is a substantial finite-window change. Because the comparator does not enforce fixed magnetization or reflection parity, the ratio should be treated as a common scale for the plot rather than as a final thermodynamic benchmark. A constrained Page calculation could shift all ratios upward or downward and would be required before any precise statement about saturation.

The diagonal smoothness proxy is the least central table quantity. Its values are extremely small across the grid, but this is partly a consequence of the observable and the central windows rather than a standalone proof of ETH. The off-diagonal variance, adjacent-size exponents, and entropy trend are more informative for the hypothesis because they directly test matrix-element scaling and eigenvector structure. The manuscript therefore uses the diagonal proxy as supporting context only. This choice is conservative: it avoids allowing one very small finite-window number to outweigh the failed spectral crossover and the incomplete ETH collapse.

D Failure Atlas and Follow-up Measurements

The present results identify several failure modes that should be separated in future work. The first is a relevance failure for the spectral assumption: the measured finite blocks do not match the irreducible symmetry sectors required for GOE comparison. A direct repair is to label eigenstates by total spin and compute spacing ratios inside fixed- S sectors, although this reduces block sizes and may require larger L to recover useful statistics. An alternative repair is to move away from $\Delta = 1$ or add a perturbation that breaks $\text{SU}(2)$ while keeping translation-free open boundaries and deterministic couplings. That variant would test whether the negative spectral result is mainly a symmetry artifact.

The second failure mode is finite-size insufficiency. The proposed scale coincidence is a scaling claim, but three sizes supply only two adjacent-size ETH exponents and at most three spectral scales. Adding $L = 14$ with dense diagonalization, or using sparse methods targeted at the middle of the spectrum, would allow a more meaningful estimate of drift with L . Such runs should include denser coupling points around $\lambda \simeq 0.2\text{--}0.5$, where the ETH midpoint estimates occur and where the spectral curves show nonmonotone structure. Without this local refinement, interpolation values such as 0.306 or 0.363 should be treated as coarse indicators.

The third failure mode is uncertainty. Deterministic Hamiltonians do not provide disorder samples, but uncertainty can still be estimated by varying the energy window, jackknifing eigenstate pairs, changing the central fraction, and comparing parity sectors. These procedures would not create independent physics realizations, yet they would quantify how robust each diagnostic is to analysis choices. The fourth failure mode is the entanglement comparator.

A constrained Page estimate for fixed magnetization and reflection parity would make the entropy ratio more physically aligned with the measured sectors. These follow-up measurements would convert the current failure atlas into a sharper decision between two possibilities: either the single-scale hypothesis fails generically for this perturbation family, or it reappears only after symmetry resolution and larger-size scaling.