

Exploring Variational Entanglement Hamiltonians

Yanick S. Kind^{1,2} and Benedikt Fauseweh^{1,2}

¹Condensed Matter Theory, TU Dortmund University, Otto-Hahn-Straße 4, 44227 Dortmund, Germany

²Institute of Software Technology, German Aerospace Center (DLR), 51147 Cologne, Germany

Recent advances in analog and digital quantum-simulation platforms have enabled exploration of the spectrum of entanglement Hamiltonians via variational algorithms. In this work we analyze the convergence properties of the variationally obtained solutions and compare them to numerically exact calculations in quantum critical systems. We demonstrate that interpreting the cost functional as an integral permits the deployment of iterative quadrature schemes, thereby reducing the required number of measurements by more than an order of magnitude even in the presence of noise. We further show that a modified ansatz captures deviations from the Bisognano–Wichmann form in lattice models, improves convergence, improves trainability and provides a cost-function-level diagnostic for quantum phase transitions. Finally, we establish that a low cost value does not by itself guarantee convergence in trace distance. Nevertheless, it faithfully reproduces degeneracies and spectral gaps, which are essential for applications to topological phases.

Generating and controlling novel quantum states is a central objective in condensed-matter physics and in the development of emerging quantum devices. Complex many-body states are often highly entangled, rendering their analysis by classical numerical methods infeasible. Consequently, experiments that prepare such states frequently provide the only viable means of investigation—an idea at the heart of quantum simulation [1–4]. By tuning control parameters and performing measurements we can explore systems beyond classical computational limits, with notable examples in trapped-ion and ultracold-atom platforms [5–15].

Variational quantum algorithms (VQAs) [16] have emerged as a valuable tool in quantum simulation. These hybrid quantum–classical schemes leverage the complementary strengths of a classical computer to efficiently optimize scalar objective functions and of a quantum processor to represent, manipulate, and measure states in high-dimensional Hilbert spaces. Potential applications include the preparation of ground [17–20], excited [21–23], time-evolved [24–26], and Floquet states [27, 28].

Recently, Kokail *et al.* proposed a variational algorithm to learn the entanglement Hamiltonian of a given ground state [29, 30] and demonstrated it experimentally in trapped-ion systems [31, 32]. Ultracold atomic gases have also been used to variationally determine the entanglement Hamiltonian [33]. Combined with entanglement spectroscopy, the learned Hamiltonian enables determination of the low-lying entanglement spectrum without exponential resource costs. Since the seminal work by Li and Haldane [34], the entanglement spectrum has become an important diagnostic not only for topologically ordered phases but also for symmetry-protected topological phases [35, 36], tensor-network approaches [37], many-body-localized systems [38] and to detect quantum phase transitions [39–41]. Notably the measurement of the entanglement spectrum of a symmetry protected topological state was also realized on IBM quantum computers [42].

The variational algorithm employs an ansatz based on the Bisognano–Wichmann (BW) form of the entanglement Hamiltonian [43–45]. While this form is exact for ground states of relativistic quantum field theories (QFTs), it is generally only an approximation on a lattice. As a result, resolving spectral gaps and degeneracies with high fidelity can be demanding. Moreover, optimization of the cost function entails time evolution under the ansatz Hamiltonian and measurement of observables at multiple time points,

Benedikt Fauseweh: benedikt.fauseweh@tu-dortmund.de

leading to substantial measurement overhead.

In this paper we address these challenges by analyzing the convergence properties of the variational algorithm. We perform classical simulations for several one-dimensional spin models, contrast the simple right point integration rule [29, 31] with more sophisticated quadrature schemes, and demonstrate the superiority of the latter, even in the presence of noise. We further show that a BW violating ansatz can improve trainability and thereby lower the measurement costs of optimization and render the solution insensitive to observation time. Finally, we establish that a small cost value does not necessarily imply high fidelity as measured by the trace distance; nevertheless, it reliably indicates spectral gaps and degeneracies in the entanglement spectrum, as we demonstrate for the symmetry-protected Haldane phase in a $S = 1$ chain.

1 Theoretical foundation

1.1 Schmidt decomposition and entanglement Hamiltonian

Given a composite Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, composed of two subsystems A and B with their respective Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ of dimensions $d_A = \dim(\mathcal{H}_A)$ and $d_B = \dim(\mathcal{H}_B)$, spanned by the orthonormal bases $\{|\mu_A^i\rangle\}$ and $\{|\mu_B^j\rangle\}$, a general pure state $|\Psi\rangle \in \mathcal{H}$ can be written as

$$|\Psi\rangle = \sum_{i=1}^{d_A} \sum_{j=1}^{d_B} M_{ij} |\mu_A^i\rangle \otimes |\mu_B^j\rangle, \quad (1)$$

where the rank $\chi \leq \min(d_A, d_B)$ of the complex matrix M is called the Schmidt rank [46]. The entanglement matrix M can be brought into a diagonal form D via a singular value decomposition (SVD) [47]

$$M = UDV^\dagger. \quad (2)$$

The matrices U and V are of size $d_A \times \min(d_A, d_B)$ and $d_B \times \min(d_A, d_B)$, respectively, and obey $U^\dagger U = \mathbb{1}$ and $VV^\dagger = \mathbb{1}$. The non-negative entries (the singular values of M) of the diagonal matrix D with dimension $\min(d_A, d_B)$ are called Schmidt-coefficients [48] and can be expressed as

$e^{-\xi_\alpha/2}$. Using the SVD, Eq. (1) reads

$$|\Psi\rangle = \sum_{i=1}^{d_A} \sum_{j=1}^{d_B} \sum_{\alpha=1}^{\min(d_A, d_B)} e^{-\xi_\alpha/2} U_{i\alpha} V_{j\alpha}^* |\mu_A^i\rangle \otimes |\mu_B^j\rangle. \quad (3)$$

Defining a new orthonormal basis set $\{|\Phi_A^\alpha\rangle = \sum_{i=1}^{d_A} U_{i\alpha} |\mu_A^i\rangle\}$ and $\{|\Phi_B^\alpha\rangle = \sum_{j=1}^{d_B} V_{j\alpha}^* |\mu_B^j\rangle\}$ yields

$$|\Psi\rangle = \sum_{\alpha=1}^{\chi} e^{-\xi_\alpha/2} |\Phi_A^\alpha\rangle \otimes |\Phi_B^\alpha\rangle, \quad (4)$$

where $\{\xi_\alpha\}$ will be referred to as the Entanglement Spectrum (ES). Since the rank is preserved under a SVD, the number of non-zero singular values coincides with the Schmidt rank χ [49], and thus, the sum in Eq. (4) is restricted to χ . The lower and upper bound of summation will be dropped from now on as long as it is unambiguous. The reduced density matrix of the state $\hat{\rho} = |\Psi\rangle\langle\Psi|$ on subsystem A after tracing out the degrees of freedom related to subsystem B is given by

$$\hat{\rho}_A = \text{Tr}_B [\hat{\rho}] = \sum_{\alpha} e^{-\xi_\alpha} |\Phi_A^\alpha\rangle\langle\Phi_A^\alpha| =: e^{-\hat{H}_A}, \quad (5)$$

which defines the Entanglement Hamiltonian (EH) $\hat{H}_A$ [29, 50]. The EH and its non-negative eigenvalues $\{\xi_\alpha\}$, the ES, completely characterize all correlations in partition A [29] and reveal much more than the entanglement entropy [34] or the entanglement witness [41]. In general, it is hard to derive an analytical form of the EH especially for lattice theories. The BW theorem (Section 1.2) allows one to obtain the EH analytically for specific cases for QFTs.

1.2 Bisognano-Wichmann theorem

In a $d + 1$ -dimensional Poincaré invariant QFT with a local Hamiltonian density $\hat{\mathcal{H}}(\mathbf{x})$, the *exact* EH of the ground state for the special case of a bipartition of an infinite system A ($A = \{\mathbf{x} \in \mathbb{R}^d | x_1 > 0\}$) is

$$\hat{H}_A = \int_A d^d x \beta(x_1) \hat{\mathcal{H}}(\mathbf{x}) + c' \quad (6)$$

with $\beta(x_1) = \frac{2\pi}{c} x_1$ [29, 51, 52], whereby the speed of sound c of the underlying QFT is set to unity from now on. The constant c' ensures the normalization $\text{Tr}[\hat{\rho}_A] = 1$. This is the seminal BW-theorem. In Eq. (6) it becomes apparent that the

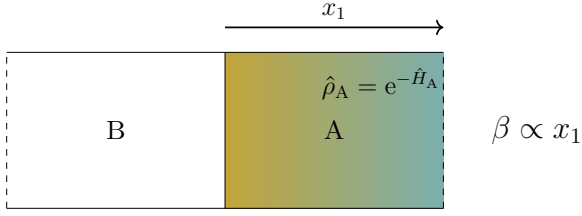


Figure 1: Interpretation of the reduced density matrix $\hat{\rho}_A$ as a thermal state with a locally varying temperature, the entanglement temperature. The inverse entanglement temperature takes the form of a linear ramp, and thus, the entanglement temperature decreases as $\propto 1/x_1$, indicated by the color gradient from orange to blue.

EH is a deformation of the system Hamiltonian [29]. Additionally, the reduced density matrix $\hat{\rho}_A$ can be interpreted as a thermal state with a locally varying entanglement temperature, which is very high near the entanglement cut (boundary between both partitions) and decreases with $1/x_1$ away from it [29], see also Fig. 1.

For lattice systems, it is straightforward to propose a discretized version of Eq. (6) s.t.

$$\hat{H}_A \approx \sum_{i \in A} g_i \hat{h}_i + c', \quad (7)$$

where the substitution $\beta(x_1) \rightarrow g_i$ and $\hat{\mathcal{H}}(x) \rightarrow \hat{h}_i$ with $\hat{h}_i$ as a quasi-local few-body operator for the i -th lattice site is utilized. A natural question is whether the BW theorem works for lattice systems since it is defined for relativistic QFTs at first. Although the presence of a lattice breaks the Lorentz invariance [51] (even when it is recovered as a low-energy symmetry[52]), analytical and numerical calculations [50–53] suggest that the discretized version of the BW theorem (7) is often a good first approximation for lattice systems.

1.3 Conformal extensions

For systems, which have conformal symmetry in addition to Lorentz invariance, the BW theorem (Eq. (6)) can be extended to different geometries [52]. In this work we focus one-dimensional systems. In case of a finite subsystem of length l in a ring of circumference L , the EH is given by

$$\hat{H}_A^{\text{CFT1}} = 2L \int_0^l dx \frac{\sin\left(\frac{\pi(l-x)}{L}\right) \sin\left(\frac{\pi x}{L}\right)}{\sin\left(\frac{\pi l}{L}\right)} \hat{\mathcal{H}}(x) + c'. \quad (8)$$

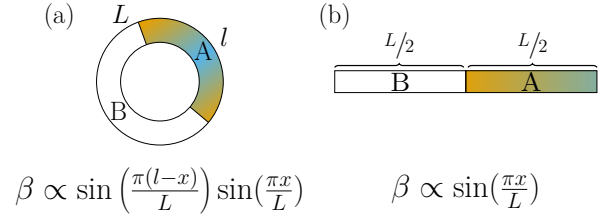


Figure 2: The BW theorem extended to one dimensional systems with conformal symmetry. Schematic representation of the entanglement temperature for (a) a subsystem of length l embedded in a system of length L with periodic boundary conditions and (b) a subsystem of length $L/2$ embedded at a boundary of an open system of length L . The colors indicate a high (orange) and low (blue) entanglement temperature.

Since the system obeys periodic boundary conditions (PBC), there are two entanglement cuts, where the inverse temperature increases approximately linearly for small distances from the entanglement cut in agreement with the BW theorem (Eq. (6)). For a finite partition of length $L/2$ at the edge of a finite open system of length L , the EH reads

$$\hat{H}_A^{\text{CFT2}} = 2L \int_0^{L/2} dx \sin\left(\frac{\pi x}{L}\right) \hat{\mathcal{H}}(x) + c', \quad (9)$$

again with a linear rise of the inverse entanglement temperature near the entanglement cut (see Fig. 2). For a finite subsystem of length l in an infinite composite system, the EH is given by

$$\hat{H}_A^{\text{CFT3}} = 2\pi \int_0^l dx x \left(\frac{l-x}{l}\right) \hat{\mathcal{H}}(x) + c'. \quad (10)$$

With these modified prefactors, we can therefore also describe conformal extensions, e.g., periodic boundary conditions, beyond the original BW theorem.

1.4 Quantum classical algorithm

The main goal of the algorithm, first presented in Ref. [29], is to learn the EH via a hybrid quantum-classical feedback loop (QCFL) utilizing the variational ansatz $\hat{H}_A^{\text{Var}}(\mathbf{g}) = \sum_i g_i \hat{h}_i$, which acts as a generator for the time evolution operator

$$\hat{U}_A(\mathbf{g}, t) = e^{-i\hat{H}_A^{\text{Var}}(\mathbf{g})t}, \quad (11)$$

acting on subsystem A for some time t s.t.

$$\hat{\rho}_A \rightarrow \hat{U}_A(\mathbf{g}, t) \hat{\rho}_A \hat{U}_A^\dagger(\mathbf{g}, t). \quad (12)$$

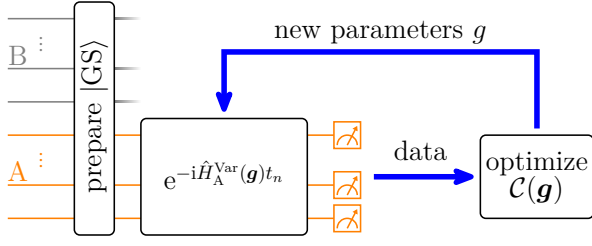


Figure 3: Quantum classical feedback loop (QCFL). The composite system is initialized with the ground state, $|\text{GS}\rangle$, of the system Hamiltonian. The subsystem A is then evolved under the variational ansatz $\hat{H}_A^{\text{Var}}(\mathbf{g})$ and some observables $\langle \hat{\mathcal{O}}_j^A \rangle_{t_n}$ are measured at time instances $\{t_n\}$. The cost function is then evaluated with the measurements and the new parameters suggested by the optimizer are used to repeat the procedure.

The parameters g_i are the variational parameters of the algorithm. The QCFL, see Fig. 3, works as follows:

1. Prepare an initial state $\hat{\rho}_A = \text{Tr}_B[|\text{GS}\rangle\langle\text{GS}|]$ with $|\text{GS}\rangle$ as the ground state of the composite system.
2. Evolve the subsystem A under the variational ansatz for some time $t_n > 0$, leaving the complementary subsystem untouched.
3. Evaluate the expectation values $\langle \hat{\mathcal{O}}_j^A \rangle_{t_n}$ after each time t_n .
4. Calculate a suitable cost function $\mathcal{C}(\mathbf{g})$ on a classical computer.
5. Repeat step 2 to 4 for different variational parameters and minimize $\mathcal{C}(\mathbf{g})$.

The expectation value after the subsystem A has been evolved under the variational ansatz reads

$$\langle \hat{\mathcal{O}}_j^A \rangle_{t_n} = \text{Tr}_A \left[\hat{\mathcal{O}}_j^A \hat{U}_A(\mathbf{g}, t_n) \hat{\rho}_A \hat{U}_A^\dagger(\mathbf{g}, t_n) \right], \quad (13)$$

where the operators $\hat{\mathcal{O}}_j^A$ are only defined on subsystem A and are restricted to be (quasi-)local. The optimal parameters $\mathbf{g}^{\text{opt}}$ are learned by minimizing the time variation of the observables s.t. $\langle \hat{\mathcal{O}}_j^A \rangle_{t_n} = \text{const.}$ A suitable cost function to be minimized is given as

$$\mathcal{C}(\mathbf{g}) = \sum_{j=1}^{N_O} \sum_{n=1}^{N_T} \left(\langle \hat{\mathcal{O}}_j^A \rangle_{t_n} - \langle \hat{\mathcal{O}}_j^A \rangle_0 \right)^2 \quad (14)$$

with N_O as the number of observables and N_T as the number of times the subsystem A is evolved

and each observable is measured. For sufficiently many observation times t_n and observables $\hat{\mathcal{O}}_j^A$, a cost function value of zero implies

$$[\hat{H}_A^{\text{Var}}(\mathbf{g}^{\text{opt}}), \hat{H}_A] = 0, \quad (15)$$

where $\hat{H}_A$ is the exact EH and $\mathbf{g}^{\text{opt}}$ are the optimal variational parameters. Equivalently, a cost function value of zero implies $[\hat{H}_A^{\text{Var}}(\mathbf{g}^{\text{opt}}), \hat{\rho}_A] = 0$, too, since the exact reduced density matrix $\hat{\rho}_A = \exp(-\hat{H}_A)$ is given by a power series in $\hat{H}_A$. This results in a thermalized subsystem A and the observables are constant in time. The precise choice of observables is not crucial, since an operator is expected to evolve into a complex operator under the dynamics as long as $[\hat{H}_A^{\text{Var}}(\mathbf{g}), \hat{\mathcal{O}}_j^A] \neq 0$. Since the aforementioned commutator is still fulfilled if a solution $\mathbf{g}^{\text{opt}}$ is scaled by a factor γ , the scale factor remains undetermined by the algorithm as well as the normalization constant c' (see Eq. (7)). To compare the ES of the variational solution and the exact ES, the universal ratios

$$\kappa_\alpha = \frac{\xi_\alpha - \xi_{\alpha_0}}{\xi_{\alpha_1} - \xi_{\alpha_0}} \quad (16)$$

are defined s.t. the undetermined scaling factor γ and the normalization constant c' are eliminated by division and subtraction, respectively. Our numerical simulations use ground states from exact diagonalization and exact time evolution. In quantum simulators based on trapped ions, the entanglement Hamiltonian was determined for a long-range transverse-field Ising model [32] and the ground state was prepared variationally. On digital quantum devices, the required time evolution can, in principle, be realized using a Trotter–Suzuki decomposition. Our convergence analysis in Appendix A shows that a modest maximum evolution time is sufficient for the models considered. Moreover, as demonstrated in Section 2, advanced quadrature schemes require only a small number of time points. Taken together, these observations indicate that the protocol can be implemented with a shallow Trotter–Suzuki decomposition.

1.5 Improvement of the cost function

Throughout the investigation of the algorithm, some difficulties, and thus, possibilities to improve the algorithm have been noticed. A major challenge is to determine the observation times t_n ,

such that the optimized parameters are stable to small changes in the t_n . This can be easily solved by reinterpreting the cost function. Recalling the cost function in Eq. 14, it is hard to compare numerical values of the cost function, since it is not normalized to the number of observables N_O and to the number of observation times N_T , which is easily fixed by dividing by these aforementioned quantities. Since the algorithm is based on monitoring observables, it is, in general, not enough to choose a few arbitrary discrete time points. Otherwise, the variational parameters $\mathbf{g}^{\text{opt}}$ will not be converged. Assuming equidistant time points i.e. a step size Δt for the observation times, the cost function can therefore, together with the aforementioned normalization, be rewritten as

$$\mathcal{C}(\mathbf{g}) = \frac{\Delta t}{T_{\max} N_O} \sum_{j=1}^{N_O} \sum_{n=1}^{N_T} \left(\langle \hat{\mathcal{O}}_j^A \rangle_{n\Delta t} - \langle \hat{\mathcal{O}}_j^A \rangle_0 \right)^2, \quad (17)$$

defining the maximum observation time $T_{\max} = N_T \Delta t$. To obtain a cost function, observing not at discrete time points but at all times, the discrete sum is replaced by an integral,

$$\begin{aligned} \mathcal{C}(\mathbf{g}) &\stackrel{\lim_{\Delta t \rightarrow 0}}{=} \frac{1}{T_{\max}} \int_0^{T_{\max}} \frac{1}{N_O} \sum_{j=1}^{N_O} \underbrace{\left(\langle \hat{\mathcal{O}}_j^A \rangle_t - \langle \hat{\mathcal{O}}_j^A \rangle_0 \right)^2}_{:=c(\mathbf{g},t)} dt \\ &= \frac{1}{T_{\max}} \int_0^{T_{\max}} c(\mathbf{g},t) dt, \end{aligned} \quad (18)$$

which allows numerical integration techniques to be applied. Note that there remains one degree of freedom to choose properly, namely the maximum observation time $T_{\max}$. The influence of $T_{\max}$ will be thoroughly discussed in Appendices A.2 and A.3.

1.6 Variational ansatz schemes

We mainly distinguish two types of variational Ansätze in this work. The first ansatz is the BW-like ansatz, denoted as $\hat{H}_A^{\text{BW}}$, which is used in Reference [29]. As the name suggests, it follows the BW theorem, i.e., it is a local deformation of the system Hamiltonian. The second ansatz, $\hat{H}_A^{\text{BWV}}$, is used to show a violation of the BW theorem in lattice models. In general, both Ansätze are given by a linear combination

$$\hat{H}_A^{\text{Var}}(\mathbf{g}) = \sum_i g_i \hat{h}_i, \quad (19)$$

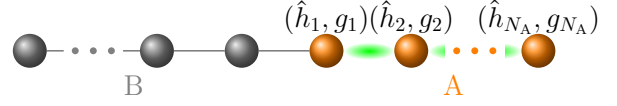


Figure 4: Schematic illustration of the variational ansatz $\hat{H}_A^{\text{BW}} = \sum_i g_i \hat{h}_i$. Each lattice site in the subsystem A is assigned a few-body quasi-local operator $\hat{h}_i$ together with a variational parameter g_i . Only interactions within the subsystem A are taken into account, as suggested by the green highlighting.

where g_i is a variational parameter and $\hat{h}_i$ is a quasi-local few-body operator, which will be referred to as a block.

1.6.1 BW-like ansatz

The BW theorem predicts that the EH is a spatially deformed version of the system Hamiltonian on a subsystem. That is, each lattice site i is assigned a block $\hat{h}_i$ together with one variational parameter g_i , as illustrated in Fig. 4. That means that the index i in Eq. (19) coincides with the i -th lattice site in the subsystem A. The blocks $\hat{h}_i$ are not local and act on more than one qubit. It is important to note that all interactions are restricted to be within subsystem A as well.

1.6.2 BW-violating ansatz

The BW-violating ansatz is not given by a spatially deformed Hamiltonian. Thus, each lattice site is assigned multiple blocks $\hat{h}_i$ and multiple variational parameters g_i and the index i in Eq. (19) does not coincide with the lattice site i . From now on, the dependence of the variational Ansätze on the variational parameters $\mathbf{g}$ will be omitted.

1.7 Transverse field Ising model

The Hamiltonian of the TFIM with N sites, open boundary conditions (OBC), nearest neighbour coupling strength J and transverse field strength Γ reads[52]

$$\hat{H} = -J \sum_{i=1}^{N-1} Z_i Z_{i+1} - \Gamma \sum_{i=1}^N X_i, \quad (20)$$

where X_i, Y_i, Z_i denotes the Pauli matrices on site i . The first term favors a ferromagnetic state for $J > 0$ and an antiferromagnetic state for $J < 0$ while the transverse field introduces fluctuations

s.t. an orientation along the x -axis is favored by the transverse term. It possesses a $\mathbb{Z}_2$ symmetry, where the Hamiltonian is invariant under flipping all spins, i.e.

$$Z_i \rightarrow -Z_i. \quad (21)$$

In the limit $J \gg \Gamma$, the ground state is two-fold degenerate and the system is fully polarized with all spins pointing either up or down

$$|\text{GS}\rangle = \bigotimes_{i=1}^N |\uparrow\rangle \quad \text{or} \quad |\text{GS}\rangle = \bigotimes_{i=1}^N |\downarrow\rangle, \quad (22)$$

breaking the $\mathbb{Z}_2$ symmetry spontaneously, whereas all spins are completely aligned in the x -direction in the limit $\Gamma \gg J$

$$|\text{GS}\rangle = \bigotimes_{i=1}^N \frac{1}{\sqrt{2}} (\underbrace{|\uparrow\rangle + |\downarrow\rangle}_{=|\rightarrow\rangle}), \quad (23)$$

exhibiting a paramagnetic behaviour. The TFIM has a quantum critical point at $J/\Gamma = 1$, separating the ordered ferromagnetic and the disordered paramagnetic phase [54]. From now on, $J = 1$ holds. In the case of the ansatz $\hat{H}_A^{\text{BW}}$, one block of site i is given by

$$\hat{h}_i = -\frac{1}{2} \sum_{j \in \langle j, i \rangle \cap A} Z_j Z_i - \Gamma X_i, \quad (24)$$

where $\langle j, i \rangle \cap A$ denotes nearest neighbour coupling only if i and j are in the subsystem A . The BW-violating ansatz for the TFIM is given by

$$\hat{H}_A^{\text{BWV}} = - \sum_{i=1}^{N_A-1} J_{i,i+1} Z_i Z_{i+1} - \Gamma \sum_{i=1}^{N_A} \Gamma_i X_i. \quad (25)$$

with $J_{i,i+1}$ and Γ_i as variational parameters. An analytical expression for the EH in the thermodynamic limit in case of a bipartition is given in Reference [51].

1.8 XXZ model

The Hamiltonian of the XXZ model with N lattice sites and OBC is defined as

$$\hat{H} = \sum_{i=1}^{N-1} (X_i X_{i+1} + Y_i Y_{i+1} + \Delta Z_i Z_{i+1}), \quad (26)$$

where Δ is the anisotropy. For $\Delta = 1$, the isotropic case, the Heisenberg model is recovered. The XXZ model is ferromagnetic for $\Delta < -1$, quantum critical for $-1 < \Delta \leq 1$, exhibiting a

Luttinger liquid phase, and antiferromagnetic for $\Delta > 1$ [52]. The phase transition at $\Delta = -1$ is of first order, s.t. the ferromagnetic state is exact for $\Delta < -1$, while the phase transition at $\Delta = 1$ is of second order [55]. Again in the ferromagnetic phase, the $\mathbb{Z}_2$ symmetry is spontaneously broken [52]. One block for the ansatz $\hat{H}_A^{\text{BW}}$ reads

$$\hat{h}_i = \frac{1}{2} \sum_{j \in \langle j, i \rangle \cap A} (X_i X_j + Y_i Y_j + \Delta Z_i Z_j), \quad (27)$$

while the variational ansatz $\hat{H}_A^{\text{BWV}}$ for the XXZ model reads

$$\hat{H}_A^{\text{BWV}} = \sum_{i=1}^{N_A-1} (J_{i,i+1}^{\text{XX}} (X_i X_{i+1} + Y_i Y_{i+1}) + J_{i,i+1}^Z \Delta Z_i Z_{i+1}). \quad (28)$$

2 Noise analysis and trainability

In exact numerical simulations, interpreting the cost function as an integral leads to a substantial improvement in convergence (see Appendix A.1). For devices in the noisy intermediate-scale quantum (NISQ) regime [56], the relevant question is whether this improvement persists in the presence of noise.

Our numerical simulations are implemented in Julia; technical details are provided in Appendix C. We provide a Julia package, with full code documentation in [57] and the scripts for the plots in [58].

2.1 Noisy integration

We model noise at the level of expectation values of the observables entering the cost function. Specifically, for an observable $\hat{\mathcal{O}}_j^A$ evaluated at time t , we assume fluctuations with variance

$$\sigma^2 = \text{Var}(\hat{\mathcal{O}}_j^A)_t \sigma_{\text{device}}^2, \quad (29)$$

where $\text{Var}(\hat{\mathcal{O}}_j^A)_t = \langle (\hat{\mathcal{O}}_j^A)^2 \rangle_t - \langle \hat{\mathcal{O}}_j^A \rangle_t^2$. Here σ_{device}^2 parametrizes the variance of Gaussian noise with zero mean that accounts for the cumulative impact of device noise, including readout error and gate errors. For pure sampling noise we set $\sigma_{\text{device}}^2 = 1/N_M$, where N_M is the number of measurements [31]. We perform multiple simulations and scan across σ_{device} .

We use the simple right point rule (RP) as in [29], the Tanh-sinh method (TS), the adaptive

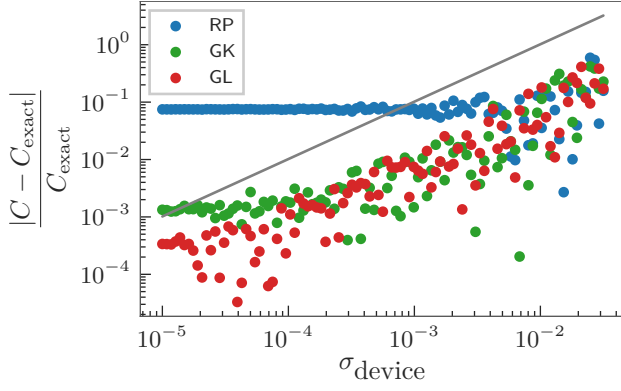


Figure 5: Deviation of the cost function value as a function of σ_{device} with five time points for the cost function evaluation. GK uses the order $n = 2$. The gray line is given by Eq. 29, where we use the maximum variance over all observables and time points to obtain a conservative upper bound for the uncertainty. The Tanh-sinh method requires at least 13 function evaluations and we therefore do not compare it here. We use the TFIM with $N = 8$, $N_A = 4$, $\Gamma = 1$, open boundary conditions with an unoptimized BW-like ansatz.

Gauss-Kronrod (GK) and the Gauss-Legendre (GL) quadratures. For a brief introduction to the integration methods, we refer to Appendix D. We compare the integrators during optimization in Fig. 5 for the TFIM.

For low to intermediate noise levels, the right-point rule exhibits a clear plateau, saturating at a $\sim 7\%$ error, while the Gauss-Kronrod and Gauss-Legendre quadratures converge to values at the noise floor down to $\sigma_{\text{device}} = 10^{-5}$, with only five function evaluations. In the very large noise regime, all methods yield comparable results. Thus improved quadrature schemes are uniformly preferable: in the presence of strong noise they perform on par with simpler rules, while for low to intermediate noise they provide substantially improved accuracy at the same evaluation cost.

The importance of a fast converging integration method in the presence of noise is confirmed in Fig. 6, where we investigate the dependence on the number of function evaluations. The right point rule reaches the precision of the Gauss-Legendre and Gauss-Kronrod quadrature with five time points at ~ 250 time points. Thus, the right point rule requires ~ 50 times more time steps, which in turn requires ~ 50 times more measurements to reach the same precision.

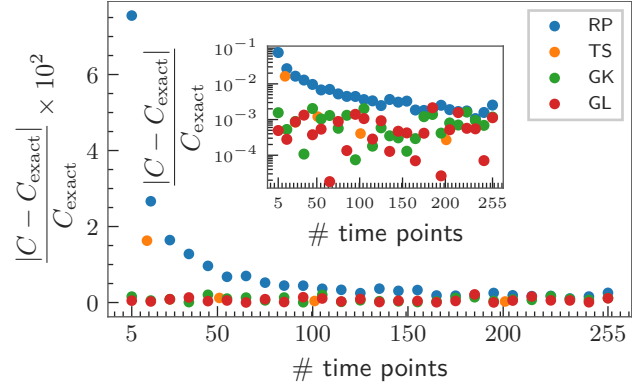


Figure 6: Deviation of the cost function value as a function of the the number of time points for the cost function evaluation. Here, the number of measurements per time point is fixed to $N_M = 10^8$. We use the TFIM with $N = 8$, $N_A = 4$, $\Gamma = 1$, open boundary conditions with an unoptimized BW-like ansatz.

2.2 Analysis of trainability

The barren plateau problem refers to the concentration of the loss function landscape in variational quantum algorithms, where gradients become exponentially small in system size N , rendering optimization ineffective [59]. It is characterized by a vanishing mean gradient as well as an exponentially suppressed variance, indicating that gradients concentrate around zero for almost all parameter choices. We probe this behavior numerically by analyzing the mean and variance of the gradient sampled over random parameters. This statistical characterization provides a direct measure of trainability and also allows us to assess whether increasing expressivity leads to further gradient concentration in the present setting.

Specifically, we compute 680 gradients sampled uniformly from the interval $[0, N]$ and analyze the mean and variance of their ℓ_∞ -norm for both the BW-like and the (more expressive) BW-violating ansatz. We use the TFIM with a varying subsystem chain length N , $N_A = N/2$, $\Gamma = 1$, OBC, $T_{\text{max}} = 1$ and 51 integration steps using Tanh-sinh quadrature.

In Fig. 7 we show the mean and variance of the gradient in a semi-logarithmic plot. While both quantities decay as a function of system size, this decay is far less severe than the barren-plateau behavior associated with the 2-design properties of random circuits [60]. For larger qubit numbers the gradient of the BWV ansatz begins to flatten, indicating a sub-exponential decay.

Increased expressivity is often associated with

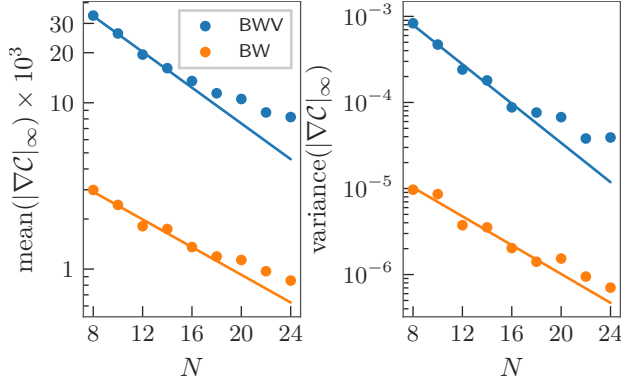


Figure 7: (a) Mean and (b) variance of the ℓ_∞ -norm of the gradient over 680 samples for the BW-like (BW) and the BW-violating (BWV) ansatz vs the number of spins in the composite system N . We use the TFIM with $N_A = N/2$, $\Gamma = 1$ and open boundary conditions. Straight lines are exponential fits using system sizes 8, 10, 12, and 14.

reduced trainability due to loss-function concentration over the enlarged variational manifold [61, 62]. This behavior is not observed in this case. Although the BWV ansatz introduces additional variational parameters, the enlargement of the variational manifold does not lead to a decrease in the gradient. Instead, both mean and variance of the gradient are significantly enhanced, indicating a reduction of loss-function concentration relative to the BW ansatz. This demonstrates that increasing the dimensionality of the variational manifold does not necessarily exacerbate barren-plateau behavior when the added directions are highly relevant for the optimization problem.

3 Results

3.1 Violation of the BW theorem

The convergence properties of different ansatz schemes are analyzed in Appendix A. In the following we investigate the variational accuracy of the converged solutions across the different ansatz schemes more systematically by comparing to a cost function that directly measures the commutator between the ansatz Hamiltonian and the exact reduced density matrix. We investigate the TFIM with $\Gamma = 1$ using OBC and PBC for the $\hat{H}_A^{\text{BW}}$ and the $\hat{H}_A^{\text{BWV}}$ ansatz, with $N = 10$, $N_A = 5$ and $T_{\text{max}} = 1$. For the same analysis of the XXZ model, we refer to Appendix B. We use numerically exact calculations without noise. We

perform a total of 50 minimizations, drawing every initial parameter uniformly from the interval [2, 6]. Outliers are filtered out based on the value of the cost function and the ratios of the optimal parameters. The optimal parameters from the QCFL are compared to the optimal parameters from the cost function

$$\mathcal{C}^{\text{Comm}}(\mathbf{g}) = \frac{\|\hat{\rho}_A, \hat{H}_A^{\text{Var}}(\mathbf{g})\|_F}{2\|\hat{\rho}_A\|_F\|\hat{H}_A^{\text{Var}}(\mathbf{g})\|_F}, \quad (30)$$

which measures the commutativity of the exact reduced density matrix and the variational ansatz. The notation $\|X\|_F$ with some $n \times m$ matrix X denotes the Frobenius norm of X defined as [63]

$$\|X\|_F = \left(\sum_{i=1}^n \sum_{j=1}^m |x_{ij}|^2 \right)^{\frac{1}{2}}. \quad (31)$$

Here, $n = m = 2^{N_A}$ holds. No filtering of outliers is done with the results of the aforementioned commutator as a cost function $\mathcal{C}^{\text{Comm}}(\mathbf{g})$. We plot the ratios of the parameters to eliminate the undetermined scale factor in the found solution (see Eq. (16)).

If the ansatz $\hat{H}_A^{\text{BW}}$ is used in the case of OBC, only the ratio $g_2^{\text{opt}}/g_1^{\text{opt}}$ obtained with the QCFL shows good agreement with the ratio obtained via the commutator (Fig. 8(a)), whereas the third, fourth, and fifth ratios deviate significantly. A linear rise near the entanglement cut can be observed and bending in accordance with the second CFT extension (Eq. (9)) becomes apparent. In the case of PBC, the optimal parameters exhibit large deviations from the parameters obtained via the commutator and do not follow the first CFT extension (Eq. 8) but rather a triangular form, while the parameters yielded with the commutator do show the behaviour predicted by the first CFT extension (Fig. 8(b)). In both cases, especially in the case of PBC, the standard deviation is very small, indicating the minimizer almost always finds the same solution. This observation is supported by the cost function value at its minimum agreeing in almost all runs, which takes the value of $\mathcal{C}(\mathbf{g}^{\text{opt}}) \approx 1.08(2) \times 10^{-5}$ and $\mathcal{C}(\mathbf{g}^{\text{opt}}) \approx 3.95 \times 10^{-4}$ for OBC and PBC, respectively. On the other hand, if the ansatz $\hat{H}_A^{\text{BWV}}$ is used, the optimal parameters obtained via the QCFL match perfectly with the parameters obtained via the commutator for both, OBC and

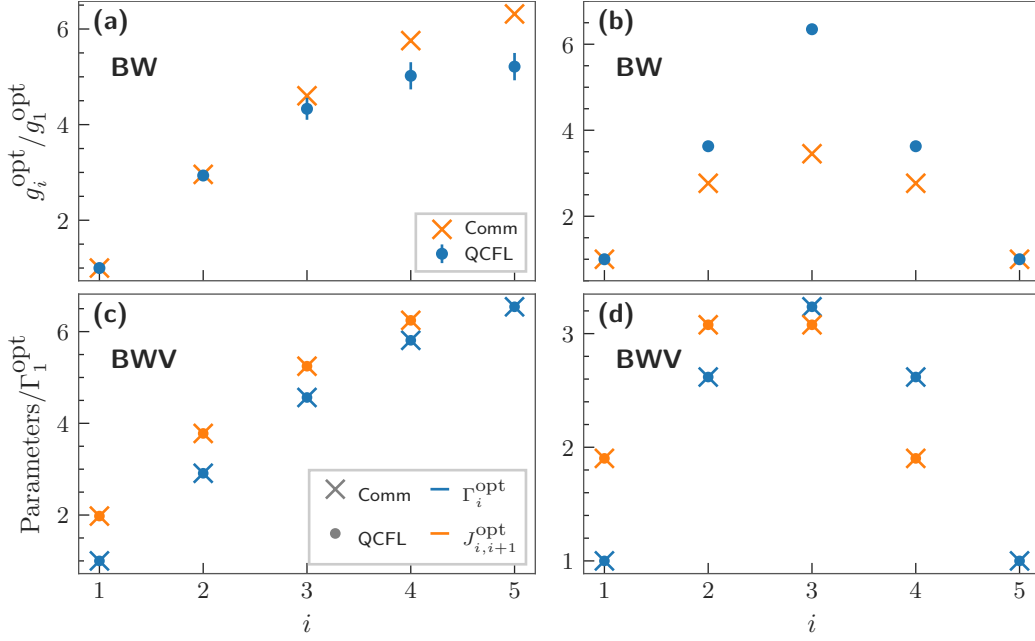


Figure 8: 50 runs with different random initial parameters for the TFIM with the QCFL and the commutator as a cost function (Comm). Outliers from the QCFL were filtered out. The mean value from the 50 runs are depicted. (a),(b) Optimal parameters for the BW-like ansatz with OBC and PBC, respectively. The error bars for the QCFL are given by the standard deviation over the 50 runs after filtering out. (c),(d) Optimal parameters for the BW-violating ansatz with OBC and PBC, respectively. We use the TFIM with $N = 10$, $N_A = 5$, and $\Gamma = 1$.

PBC (see Fig. 8(c),(d)). In addition to the perfect match, the standard deviation of each parameter obtained via the QCFL is in the vicinity of 10^{-14} , indicating that there is one good solution, while the standard deviation of the ratios obtained via the commutator are not larger than 6×10^{-9} , which is why no error bars are depicted. The minimum of the cost function is numerically zero for both OBC and PBC, underlining the fact that the ansatz $\hat{H}_A^{\text{BWV}}$ works much better. This means that the BW theorem does not deliver an accurate description for the EH for the TFIM on a finite size chain. Additionally, the only corrections to the BW theorem are that there is not just one parameter per lattice, but two. Further corrections such as long-range interactions or higher-body interactions are not needed, since the cost function is already numerically zero at its minimum.

3.2 TFIM and XXZ model across the phase diagram

This section investigates the TFIM and XXZ model across the respective phase diagrams to analyze how the algorithm performs when the systems are not critical and whether the algorithm

can indicate the critical points or specific phases in the presence of measurement uncertainty.

Note that the statistical error of the cost function is governed by the measurement uncertainty of the observables, modeled as noise σ_{j,t_n} in each measured expectation value $\langle \hat{O}_j^A \rangle_{t_n}$,

$$\begin{aligned} C^{\text{noisy}} &= \frac{1}{T_{\text{max}}} \int_0^{T_{\text{max}}} \sum_{j=1}^{N_O} \frac{1}{N_O} \left(\Delta \langle \hat{O}_j^A \rangle_t + \sigma_{j,t} \right)^2 dt \\ &= \frac{1}{T_{\text{max}}} \int_0^{T_{\text{max}}} \sum_{j=1}^{N_O} \frac{1}{N_O} \left(\Delta \langle \hat{O}_j^A \rangle_t^2 \right. \\ &\quad \left. + 2\Delta \langle \hat{O}_j^A \rangle_t \sigma_{j,t} + \sigma_{j,t}^2 \right) dt, \end{aligned} \quad (32)$$

where $\Delta \langle \hat{O}_j^A \rangle_t \equiv \langle \hat{O}_j^A \rangle_t - \langle \hat{O}_j^A \rangle_0$. We use this scaling to realize a controlled zero-noise extrapolation of the cost function. For each point in the phase diagram, we extrapolate the measured cost to the zero-noise limit using data obtained at different N_M . Details of the extrapolation procedure are provided in Appendix H.

The noisy simulations use the Gauss-Legendre quadrature with five time points. We compare with exact numerics using the fully converged solution with the Tanh-sinh method. The ansatz $\hat{H}_A^{\text{BWV}}$ is used without any other corrections for both models and the maximum integration time

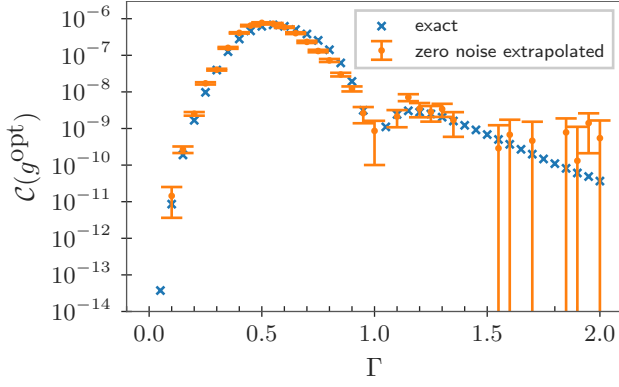


Figure 9: Minimum of the cost function in dependence on the transverse field strength Γ for the TFIM, when the ansatz $\hat{H}_A^{\text{BWV}}$ is used. We perform simulations without noise (exact) and noisy simulations with zero-noise extrapolations. Values below double precision and extrapolated values below zero are discarded. The error bars stem from the uncertainty of the fit parameters. We use the TFIM with $N = 8$, $N_A = 4$ and open boundary conditions.

is set to $T_{\text{max}} = 2$.

Figure 9 shows the minimum of the cost function for varying the transverse field strength Γ in the TFIM. At $\Gamma = 0$ the minimum of the cost function is zero, consistent with the fact that the ground state is unentangled. A second notable point appears at $\Gamma = 1$, the quantum critical point of the model, where the cost function is zero as well. For intermediate values of Γ the cost function remains low but takes finite values. The zero-noise extrapolated values are in close agreement with the exact values. We use a reasonable number of measurements ($\sim 10^9$), which is sufficient to extract small values of the cost function. This number of measurements is within reach of superconducting quantum processors, see [64, 65].

In the case of the XXZ model, see Fig. 10, the first order phase transition at $\Delta = -1$ is clearly visible. Here, due to the smaller cost function, we require two orders of magnitude more measurements in the zero-noise extrapolation ($\sim 10^{11}$). For $\Delta < -1$ the ground state of the XXZ model is a simple product state with all spins pointing in the same direction, and thus, no entanglement is present in the composite system. For the system size shown here the second order phase transition at $\Delta = 1$ is not visible, but can be seen as a cusp-like behaviour of the cost function for larger system sizes, see Appendix I. For $\Delta = 0$ the sys-

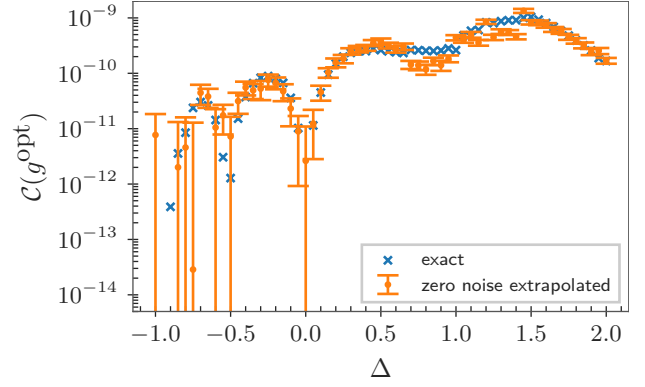


Figure 10: Minimum of the cost function in dependence on the anisotropy Δ for the XXZ model, when the ansatz $\hat{H}_A^{\text{BWV}}$ is used. We perform simulations without noise (exact) and noisy simulations with zero-noise extrapolations. Values below double precision and extrapolated values below zero are discarded. The error bars stem from the uncertainty of the fit parameter. We use the XXZ model with $N = 8$, $N_A = 4$ and open boundary conditions.

tem is at its self-dual point. In terms of Jordan-Wigner fermions it is non-interacting, and the EH is captured perfectly by the BWV ansatz. The cost function values show a cusp-like behavior at $\Delta = -0.5$, although there is no phase transition. This point is known to be algebraically special in the integrable structure of the XXZ chain.

To conclude, the algorithm delivers indications for the quantum phase transitions and classical states. Additionally, the accuracy of the BW-violating ansatz varies across the phase diagram.

3.3 Violation of the BW theorem in the XXZ model in the thermodynamic limit

In this section we discuss the reliability of extrapolating the obtained results into the thermodynamic limit (TDL). We focus on the XXZ model at $\Delta = -0.5$, where violations of the BW theorem have been observed in Ref. [66]. For simplicity we neglect long-range corrections. The subsystem chain length ranges from $N_A = 4$ to $N_A = 7$, and $T_{\text{max}} = 5$ is used. To measure the deviation from the BW theorem, the quantity

$$\theta_i = \frac{J_{i,i+1}^{\text{XX,opt}}}{J_{i,i+1}^{\text{Z,opt}}} - 1 \quad (33)$$

is defined, which will be referred to as the discrepancy. The procedure to obtain the parameters in the TDL and an example plot to show how the

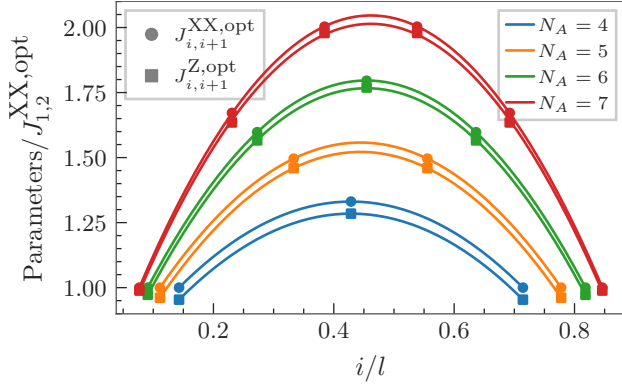


Figure 11: Optimal parameters $J_{i,i+1}^{XX,opt}$ and $J_{i,i+1}^{Z,opt}$ normalized to $J_{1,2}^{XX,opt}$ extrapolated into the TDL vs. the lattice site i in units of the subsystem chain length l for each number of sites N_A in the subsystem A. We use $\Delta = -0.5$.

Table 1: The mean value of the discrepancies as defined in Eq. (33) in the TDL for each subsystem size N_A .

N_A	$\bar{\theta}_i$
4	0.0443 ± 0.0002
5	0.0326 ± 0.0006
6	0.0224 ± 0.0004
7	0.0154 ± 0.0006

fit is done is given in Appendix G. The extrapolated parameters normalized to $J_{1,2}^{XX,opt}$ are given in Fig. 11, where the solid lines are there to guide the eye and take the quadratic form $\propto i \frac{N_A-i}{N_A}$, as suggested by the conformal extension H_A^{CFT3} (see Eq. 10), although the CFT extensions apply only to the BW theorem. No error bars are given, since the propagated estimated uncertainties of the parameters of the fit (see Eq. (48)) are not larger than 2×10^{-3} . The gap between the x - and y -couplings, and the z -couplings can be seen. Table 1 lists the mean value of the discrepancies, as defined in Eq. (33), over the lattice sites i for all subsystem lattice sizes N_A in the TDL. The errors given in Table 1 are the propagated errors, stemming from the estimated uncertainties of the fit.

We conclude that a clear discrepancy can be observed in the TDL. However, instead of the discrepancies of ≈ 0.1 from Reference [66], the discrepancies found with the algorithm of this work are approximately two times lower.

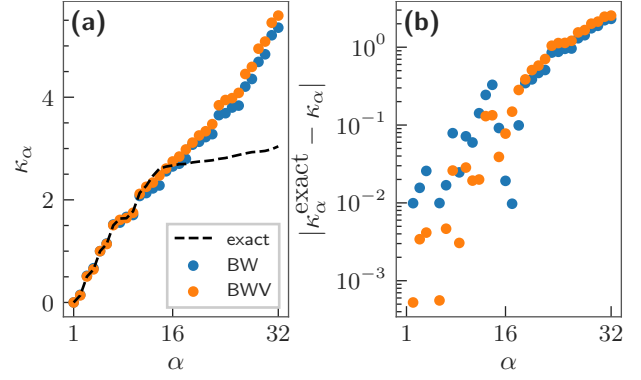


Figure 12: (a) Universal ratios and (b) deviations from the exact universal ratios for the BW-like ansatz $\hat{H}_A^{BW}$ (BW) and BW-violating ansatz $\hat{H}_A^{BWV}$ (BWV) in the TFIM. We use $N = 10$, $N_A = 5$, $\Gamma = 1$ and open boundary conditions.

3.4 Comparison of the Entanglement spectra

This section compares the universal ratios (see Eq. (16)) of the variational solutions of the previous sections to the exact ratios. We choose $\alpha_0 = 1$ and $\alpha_1 = 5$. The exact ES is computed through exact diagonalization of the exact EH, given by $\hat{H}_A = -\ln(\hat{\rho}_A)$.

In the case of the TFIM, the low-lying universal ratios match the exact universal ratios better if the BW-violating ansatz is used, as expected (see Fig. 12 (a),(b)). In contrast, in the higher part of the spectrum, the universal ratios obtained via variation show significant deviations from the exact universal ratios (Fig. 12(b)). One source for this discrepancy is numerical precision: the reduced density matrix contains (eigen-)values, which are so small s.t. they cannot be accurately captured with double precision. Since the ES (and thus the universal ratios) are obtained by taking the logarithm of the reduced density matrix, the lowest eigenvalues of the reduced density matrix are mapped to the highest universal ratios. Thus, to compare the universal ratios, we focus on the first few universal ratios. The first ten universal ratios exhibit a mean absolute deviation from the exact universal ratios of $\overline{\Delta\kappa_\alpha^{BW}} = 0.0253$ and $\overline{\Delta\kappa_\alpha^{BWV}} = 0.0071$ for the ansatz $\hat{H}_A^{BW}$ and $\hat{H}_A^{BWV}$, respectively. That is, the low-lying spectrum (here, the first ten universal ratios) is reconstructed more than three times more accurately on average if the BW-violating ansatz is used.

The universal ratios for the XXZ model are shown in Fig. 13(a). Here we include the BW-violating ansatz $\hat{H}_A^{BWV}$ with long range corrections (see

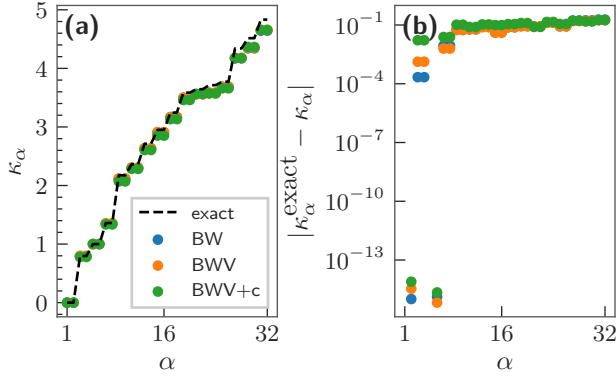


Figure 13: (a) Universal ratios and (b) deviations from the exact universal ratios for the BW-like ansatz $\hat{H}_A^{\text{BW}}$ (BW), BW-violating ansatz $\hat{H}_A^{\text{BWV}}$ (BWV) and the BW-violating ansatz with all long-range corrections (BWV+c) in the XXZ model. We use $N = 10$, $N_A = 5$, $\Delta = -0.5$ and open boundary conditions.

Appendix F). As can be seen in Fig. 13(b), the low-lying spectrum is not significantly better reconstructed if the ansatz $\hat{H}_A^{\text{BWV}}$ or the ansatz $\hat{H}_A^{\text{BWV}}$ with all long-range corrections is used.

To understand this behaviour, we investigate two additional measures, which involve the variational reduced density matrix on subsystem A

$$\hat{\rho}_A^{\text{Var}} = \frac{1}{\text{Tr} [e^{-\hat{H}_A^{\text{Var}}(g^{\text{opt}})}]} e^{-\hat{H}_A^{\text{Var}}(g^{\text{opt}})}. \quad (34)$$

The first measure is the trace distance

$$\mathcal{T}(\hat{\rho}_A, \hat{\rho}_A^{\text{Var}}) = \frac{1}{2} \text{Tr} \left[\sqrt{(\hat{\rho}_A - \hat{\rho}_A^{\text{Var}})^2} \right], \quad (35)$$

which measures how close two quantum states are and ranges from 0 (identical states) to 1 (maximally distant states) [48]. The second measure utilizes the commutator of $\hat{\rho}_A$ and $\hat{\rho}_A^{\text{Var}}$

$$\mathcal{F}(\hat{\rho}_A, \hat{\rho}_A^{\text{Var}}) = \frac{\|[\hat{\rho}_A, \hat{\rho}_A^{\text{Var}}]\|_{\text{F}}}{2\|\hat{\rho}_A\|_{\text{F}}\|\hat{\rho}_A^{\text{Var}}\|_{\text{F}}}, \quad (36)$$

which, ranges from 0 (completely commuting) to 1 (maximally non-commutative) [67]. This measure is included, since the cost function is based on the commutativity of the variational ansatz and the exact reduced density matrix. Table 2 lists the mean of the absolute deviations of the first ten universal ratios from the exact universal

ratios $\overline{\Delta\kappa_\alpha}$, the trace distance $\mathcal{T}(\hat{\rho}_A, \hat{\rho}_A^{\text{Var}})$, the norm of the commutator $\mathcal{F}(\hat{\rho}_A, \hat{\rho}_A^{\text{Var}})$ and the cost function value at its minimum $\mathcal{C}(g^{\text{opt}})$. It can be seen that $\mathcal{F}(\hat{\rho}_A, \hat{\rho}_A^{\text{Var}})$ is lower if the ansatz $\hat{H}_A^{\text{BWV}}$ is used and the lowest if the long-range corrections are included. This is in agreement with the cost function value, which shows the same trend. This observation makes sense, since the cost function is based on the commutativity of the exact reduced density matrix and the variational ansatz. However, the trace distance is the highest for the ansatz $\hat{H}_A^{\text{BWV}}$ with long-range interactions included and the lowest for the BW-like ansatz $\hat{H}_A^{\text{BW}}$. Thus, the universal ratios are *not* reconstructed more accurately with the ansatz $\hat{H}_A^{\text{BWV}}$ with long-range interactions, since the trace distance is a measure for how close two quantum states are. We conclude that a cost function based on the commutator between the exact and the variational reduced density matrices is not necessarily sufficient to minimize the trace distance between the variational EH and the exact EH. A zero commutator only implies a shared eigenbasis of the operators and not the same eigenvalues. Any function of the exact EH also leads to stationarity in the dynamics of the subsystem. While this clearly restricts the applicability of the algorithm, we expect that certain properties remain invariant, even if the variational algorithm captures only the eigenbasis correctly, such as degeneracies.

3.5 Topological Entanglement Spectrum

To investigate degeneracies in the ES, we consider the spin-1 chain model

$$\hat{H} = J \sum_i \hat{S}_i \cdot \hat{S}_{i+1} + B_x \sum_i \hat{S}_i^x + U_{zz} \sum_i (\hat{S}_i^z)^2, \quad (37)$$

where $\hat{S}_i^\alpha$ with $\alpha \in \{x, y, z\}$ are the components of the spin-1 operator $\hat{S}_i$ on the i -th lattice site. Pollmann *et al.* [36] showed that this model has a topological Haldane phase, where the ES has symmetry protected degeneracies in case of a semi-infinite half-chain. We consider a composite system of $N = 8$ sites with PBC, $J = 1$, $B_x = 0$ and $U_{zz} = 0$. The subsystem consists of $N_A = 4$ sites and we use the variational ansatz

Table 2: Measures to compare the accuracy of the variational Ansätze, BW-like ansatz (BW), BW-violating ansatz (BWV) and the BW-violating ansatz with all long-range corrections (BWV+c) for the XXZ model. We use $N = 10$, $N_A = 5$, $\Delta = -0.5$ and open boundary conditions.

ansatz	$\overline{\Delta\kappa_\alpha}$	$\mathcal{T}(\hat{\rho}_A, \hat{\rho}_A^{\text{Var}})$	$\mathcal{F}(\hat{\rho}_A, \hat{\rho}_A^{\text{Var}})$	$\mathcal{C}(\mathbf{g}^{\text{opt}})$
BW	0.013 09	0.007 51	$4.173\,79 \times 10^{-4}$	$3.550\,22 \times 10^{-7}$
BWV	0.012 56	0.019 83	$9.321\,76 \times 10^{-5}$	$5.768\,21 \times 10^{-10}$
BWV+c	0.028 87	0.030 30	$1.373\,17 \times 10^{-6}$	$1.945\,40 \times 10^{-15}$

$$\hat{H}_A^{\text{Var}} = J \sum_{i=1}^{N_A-1} (J_{i,i+1}^x S_i^x S_{i+1}^x + J_{i,i+1}^y S_i^y S_{i+1}^y + J_{i,i+1}^z S_i^z S_{i+1}^z) + B_x \sum_{i=1}^{N_A} B_i^x S_i^x + U_{zz} \sum_{i=1}^{N_A} U_i^{zz} (S_i^z)^2, \quad (38)$$

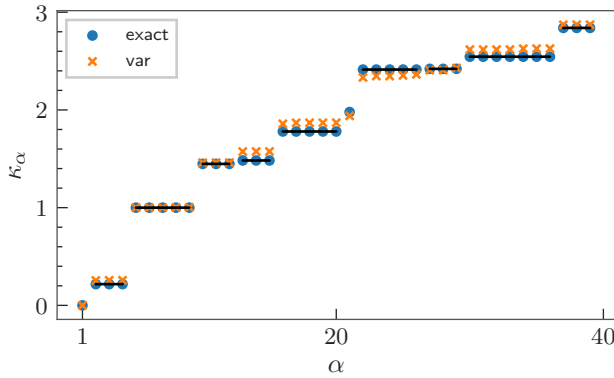


Figure 14: First forty variational (var) and exact universal ratios with $N = 8$, PBC, $J = 1$, $B_x = 0$, $U_{zz} = 0$ and $N_A = 4$. The cost function value at the minimum is $\mathcal{C}(\mathbf{g}^{\text{opt}}) \approx 5.29 \times 10^{-5}$. The black lines indicate which eigenvalues are degenerate.

where $\{J_{i,i+1}^x, J_{i,i+1}^y, J_{i,i+1}^z, B_i^x, U_i^{zz}\}$ are the variational parameters. Figure 14 shows the first forty values of the exact and variational ES. The cost function converged to $\mathcal{C}(\mathbf{g}^{\text{opt}}) \approx 5.29 \times 10^{-5}$. Although the cost function is still finite, the variational algorithm can reproduce the degeneracies in the ES. We do not find the perfect double degeneracy in the ES, as predicted for the system in the TDL [36], due to finite size effects. We conclude that degeneracies can be obtained with the variational algorithm, even if the minimum of the cost function is finite.

4 Conclusion and Outlook

In summary, the discretized Bisognano–Wichmann Ansatz provides a useful starting point for learning lattice entanglement Hamiltonians, but allowing independent

couplings per site already improves the variational description markedly, while in anisotropic Heisenberg chains a set of long-range corrections is still necessary. A system size analysis of the gradient demonstrates that the BW-violating ansatz has improved trainability despite a greater expressivity compared to the BW ansatz. This is a counterexample to what is often expected from more expressive circuits. Recasting the cost function as a time integral and evaluating it with iterative quadrature lowers the sampling effort by orders of magnitude in the absence of noise, and still by a factor of 50 in the presence of noise, a saving that will be indispensable when the protocol is run on noisy quantum hardware. The cost itself carries physical information, as it shows pronounced minima at phase transitions and remains finite otherwise, so that the cost function itself emerges as a simple experimental probe. Because the optimization enforces commutativity with the reduced density matrix, the learned operator reproduces the eigenbasis of the exact entanglement Hamiltonian and therefore captures degeneracies and gaps in the spectrum, even though individual eigenvalues are obtained only qualitatively.

An immediate outlook is to implement the iterative integration procedure and the BW violating Ansatz on analogue or digital simulators to demonstrate the improvement resulting from the integral reformulation of the cost function. Finally, the hybrid loop is not tied to ground-state physics. Applying it to excited, driven and Floquet states, where more expressive Ansatz schemes are necessary, should open the way to entanglement spectroscopy of excited states and non-equilibrium phases on current devices.

Acknowledgements

We thank Peter Zoller, Christian Kokail and Bhuvanesh Sundar for helpful comments. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)-Project No. FA 1884/5-1. Q-Neko project has received funding from the European Union’s Horizon Europe research and innovation programme under Grant Agreement No. 101241875. This work was also performed for Council for Science, Technology and Innovation (CSTI), Cross-ministerial Strategic Innovation Promotion Program (SIP), “Promoting the application of advanced quantum technology platforms to social issues”(Funding agency: QST)

References

- [1] Richard P. Feynman. “Simulating physics with computers”. *Int. J. Theor. Phys.* **21**, 467–488 (1982).
- [2] Seth Lloyd. “Universal Quantum Simulators”. *Science* **273**, 1073–1078 (1996).
- [3] Andrew J. Daley, Immanuel Bloch, Christian Kokail, Stuart Flannigan, Natalie Pearson, Matthias Troyer, and Peter Zoller. “Practical quantum advantage in quantum simulation”. *Nature* **607**, 667–676 (2022).
- [4] Benedikt Fauseweh. “Quantum many-body simulations on digital quantum computers: State-of-the-art and future challenges”. *Nature Communications* **15**, 2123 (2024).
- [5] C. Monroe, W. C. Campbell, L.-M. Duan, Z.-X. Gong, A. V. Gorshkov, P. W. Hess, R. Islam, K. Kim, N. M. Linke, G. Pagano, P. Richerme, C. Senko, and N. Y. Yao. “Programmable quantum simulations of spin systems with trapped ions”. *Rev. Mod. Phys.* **93**, 025001 (2021).
- [6] J. Eisert, M. Friesdorf, and C. Gogolin. “Quantum many-body systems out of equilibrium”. *Nature Physics* **11**, 124–130 (2015).
- [7] J. Zhang, G. Pagano, P. W. Hess, A. Kyprianidis, P. Becker, H. Kaplan, A. V. Gorshkov, Z.-X. Gong, and C. Monroe. “Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator”. *Nature* **551**, 601–604 (2017).
- [8] Lata Kh Joshi, Andreas Elben, Amit Vikram, Benoît Vermersch, Victor Galitski, and Peter Zoller. “Probing many-body quantum chaos with quantum simulators”. *Phys. Rev. X* **12**, 011018 (2022).
- [9] Hannes Bernien, Sylvain Schwartz, Alexander Keesling, Harry Levine, Ahmed Omran, Hannes Pichler, Soonwon Choi, Alexander S. Zibrov, Manuel Endres, Markus Greiner, Vladan Vuletić, and Mikhail D. Lukin. “Probing many-body dynamics on a 51-atom quantum simulator”. *Nature* **551**, 579–584 (2017).
- [10] Christian Gross and Immanuel Bloch. “Quantum simulations with ultracold atoms in optical lattices”. *Science* **357**, 995–1001 (2017).
- [11] Immanuel Bloch, Jean Dalibard, and Sylvain Nascimbène. “Quantum simulations with ultracold quantum gases”. *Nature Physics* **8**, 267–276 (2012).
- [12] R. Blatt and C. F. Roos. “Quantum simulations with trapped ions”. *Nature Physics* **8**, 277–284 (2012).
- [13] Andrew A. Houck, Hakan E. Türeci, and Jens Koch. “On-chip quantum simulation with superconducting circuits”. *Nature Physics* **8**, 292–299 (2012).
- [14] Alán Aspuru-Guzik and Philip Walther. “Photonic quantum simulators”. *Nature Physics* **8**, 285–291 (2012).
- [15] G. Semeghini, H. Levine, A. Keesling, S. Ebadi, T. T. Wang, D. Bluvstein, R. Verresen, H. Pichler, M. Kalinowski, R. Samajdar, A. Omran, S. Sachdev, A. Vishwanath, M. Greiner, V. Vuletić, and M. D. Lukin. “Probing topological spin liquids on a programmable quantum simulator”. *Science* **374**, 1242–1247 (2021).
- [16] M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles. “Variational quantum algorithms”. *Nature Reviews Physics* **3**, 625–644 (2021).

- [17] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O’Brien. “A variational eigenvalue solver on a photonic quantum processor”. *Nature Communications* **5**, 4213 (2014).
- [18] P. J. J. O’Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V. Coveney, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis. “Scalable quantum simulation of molecular energies”. *Phys. Rev. X* **6**, 031007 (2016).
- [19] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta. “Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets”. *Nature* **549**, 242–246 (2017).
- [20] Kevin Lively, Tim Bode, Jochen Szangolies, Jian-Xin Zhu, and Benedikt Fauseweh. “Noise robust detection of quantum phase transitions”. *Phys. Rev. Res.* **6**, 043254 (2024).
- [21] Oscar Higgott, Daochen Wang, and Stephen Brierley. “Variational Quantum Computation of Excited States”. *Quantum* **3**, 156 (2019).
- [22] Ken M. Nakanishi, Kosuke Mitarai, and Keisuke Fujii. “Subspace-search variational quantum eigensolver for excited states”. *Phys. Rev. Res.* **1**, 033062 (2019).
- [23] Yasar Y. Atas, Jinglei Zhang, Randy Lewis, Amin Jahanpour, Jan F. Haase, and Christine A. Muschik. “Su(2) hadrons on a quantum computer via a variational approach”. *Nature Communications* **12**, 6499 (2021).
- [24] Joe Gibbs, Kaitlin Gili, Zoë Holmes, Benjamin Commeau, Andrew Arrasmith, Lukasz Cincio, Patrick J. Coles, and Andrew Sornborger. “Long-time simulations for fixed input states on quantum hardware”. *npj Quantum Information* **8**, 135 (2022).
- [25] Noah F. Berthusen, Thaís V. Trevisan, Thomas Iadecola, and Peter P. Orth. “Quantum dynamics simulations beyond the coherence time on noisy intermediate-scale quantum hardware by variational trotter compression”. *Phys. Rev. Res.* **4**, 023097 (2022).
- [26] Stefano Barison, Filippo Vicentini, and Giuseppe Carleo. “An efficient quantum algorithm for the time evolution of parameterized circuits”. *Quantum* **5**, 512 (2021).
- [27] Benedikt Fauseweh and Jian-Xin Zhu. “Quantum computing Floquet energy spectra”. *Quantum* **7**, 1063 (2023).
- [28] Abhishek Kumar, Karunya Shirali, Nicholas J. Mayhall, Sophia E. Economou, and Edwin Barnes. “Floquet-adapt-vqe: A quantum algorithm to simulate non-equilibrium physics in periodically driven systems” (2025). [arXiv:2503.11613](https://arxiv.org/abs/2503.11613).
- [29] Christian Kokail, Bhuvanesh Sundar, Torsten V. Zache, Andreas Elben, Benoît Vermersch, Marcello Dalmonte, Rick van Bijnen, and Peter Zoller. “Quantum variational learning of the entanglement hamiltonian”. *Phys. Rev. Lett.* **127**, 170501 (2021).
- [30] Torsten V. Zache, Christian Kokail, Bhuvanesh Sundar, and Peter Zoller. “Entanglement Spectroscopy and probing the Li-Haldane Conjecture in Topological Quantum Matter”. *Quantum* **6**, 702 (2022).
- [31] Christian Kokail, Rick van Bijnen, Andreas Elben, Benoît Vermersch, and Peter Zoller. “Entanglement hamiltonian tomography in quantum simulation”. *Nature Physics* **17**, 936–942 (2021).
- [32] Manoj K. Joshi, Christian Kokail, Rick van Bijnen, Florian Kranzl, Torsten V. Zache, Rainer Blatt, Christian F. Roos, and Peter Zoller. “Exploring large-scale entanglement in quantum simulation”. *Nature* **624**, 539–544 (2023).
- [33] Quentin Redon, Qi Liu, Jean-Baptiste Bouhiron, Nehal Mittal, Aurélien Fabre, Raphael Lopes, and Sylvain Nascimbene. “Realizing the entanglement hamiltonian of a topological quantum hall system”. *Nature Communications* **15**, 10086 (2024).
- [34] Hui Li and F. D. M. Haldane. “Entanglement spectrum as a generalization of entanglement entropy: Identification of topological order in non-abelian fractional quantum hall effect states”. *Phys. Rev. Lett.* **101**, 010504 (2008).

- [35] Lukasz Fidkowski. “Entanglement spectrum of topological insulators and superconductors”. *Phys. Rev. Lett.* **104**, 130502 (2010).
- [36] Frank Pollmann, Ari M. Turner, Erez Berg, and Masaki Oshikawa. “Entanglement spectrum of a topological phase in one dimension”. *Phys. Rev. B* **81**, 064439 (2010).
- [37] Rui-Zhen Huang, Long Zhang, Andreas M. Läuchli, Jutho Haegeman, Frank Verstraete, and Laurens Vanderstraeten. “Emergent conformal boundaries from finite-entanglement scaling in matrix product states”. *Phys. Rev. Lett.* **132**, 086503 (2024).
- [38] Maksym Serbyn, Alexios A. Michailidis, Dmitry A. Abanin, and Z. Papić. “Power-law entanglement spectrum in many-body localized phases”. *Phys. Rev. Lett.* **117**, 160601 (2016).
- [39] G. De Chiara, L. Lepori, M. Lewenstein, and A. Sanpera. “Entanglement spectrum, critical exponents, and order parameters in quantum spin chains”. *Phys. Rev. Lett.* **109**, 237208 (2012).
- [40] L. Lepori, G. De Chiara, and A. Sanpera. “Scaling of the entanglement spectrum near quantum phase transitions”. *Phys. Rev. B* **87**, 235107 (2013).
- [41] J. T. Schneider, S. J. Thomson, and L. Sanchez-Palencia. “Entanglement spectrum and quantum phase diagram of the long-range xxz chain”. *Phys. Rev. B* **106**, 014306 (2022).
- [42] Kenny Choo, Curt W. von Keyserlingk, Nicolas Regnault, and Titus Neupert. “Measurement of the entanglement spectrum of a symmetry-protected topological state using the ibm quantum computer”. *Phys. Rev. Lett.* **121**, 086808 (2018).
- [43] Joseph J. Bisognano and Eyvind H. Wichmann. “On the duality condition for a hermitian scalar field”. *Journal of Mathematical Physics* **16**, 985–1007 (1975).
- [44] Joseph J. Bisognano and Eyvind H. Wichmann. “On the duality condition for quantum fields”. *Journal of Mathematical Physics* **17**, 303–321 (1976).
- [45] M. Dalmonte, B. Vermersch, and P. Zoller. “Quantum simulation and spectroscopy of entanglement hamiltonians”. *Nature Physics* **14**, 827–831 (2018).
- [46] Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. “Quantum entanglement”. *Rev. Mod. Phys.* **81**, 865–942 (2009).
- [47] Nicolas Regnault. “Entanglement spectroscopy and its application to the quantum Hall effects”. In *Topological Aspects of Condensed Matter Physics: Lecture Notes of the Les Houches Summer School: Volume 103, August 2014*. Oxford University Press (2017).
- [48] Michael A. Nielsen and Isaac L. Chuang. “Quantum computation and quantum information: 10th anniversary edition”. Cambridge University Press. (2010).
- [49] David Lay, Steven Lay, and Judi McDonald. “Linear Algebra and Its Applications Global Edition”. Page 576. Pearson Deutschland. (2016). url: <https://elibrary.pearson.de/book/99.150005/9781292092249>.
- [50] Federico Rottoli, Michele Fossati, and Pasquale Calabrese. “Entanglement hamiltonian in the non-hermitian ssh model”. *Journal of Statistical Mechanics: Theory and Experiment* **2024**, 063102 (2024).
- [51] Marcello Dalmonte, Viktor Eisler, Marco Falconi, and Benoît Vermersch. “Entanglement hamiltonians: From field theory to lattice models and experiments”. *Annalen der Physik* **534**, 2200064 (2022).
- [52] G. Giudici, T. Mendes-Santos, P. Calabrese, and M. Dalmonte. “Entanglement hamiltonians of lattice models via the bisognano-wichmann theorem”. *Phys. Rev. B* **98**, 134403 (2018).
- [53] T. Mendes-Santos, G. Giudici, M. Dalmonte, and M. A. Rajabpour. “Entanglement hamiltonian of quantum critical chains and conformal field theories”. *Phys. Rev. B* **100**, 155122 (2019).
- [54] Benedikt Fauseweh and Götz S. Uhrig. “Multiparticle spectral properties in the transverse field ising model by continuous unitary transformations”. *Phys. Rev. B* **87**, 184406 (2013).
- [55] Ming-Ming Du, Da-Jian Zhang, Zhao-Yi Zhou, and D. M. Tong. “Visualizing quantum phase transitions in the xxz model via the quantum steering ellipsoid”. *Phys. Rev. A* **104**, 012418 (2021).
- [56] John Preskill. “Quantum Computing in the NISQ era and beyond”. *Quantum* **2**, 79 (2018).
- [57] Yanick S. Kind. “QuantVarEntHam.jl”. *Zenodo* (2026).

- [58] Yanick S. Kind and Benedikt Fauseweh. “Data scripts for exploring variational entanglement hamiltonians”. [Zenodo](#) (2026).
- [59] Martín Larocca, Supanut Thanasilp, Samson Wang, Kunal Sharma, Jacob Biamonte, Patrick J. Coles, Lukasz Cincio, Jarrod R. McClean, Zoë Holmes, and M. Cerezo. “Barren plateaus in variational quantum computing”. [Nature Reviews Physics](#) **7**, 174–189 (2025).
- [60] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. “Barren plateaus in quantum neural network training landscapes”. [Nature Communications](#) **9**, 4812 (2018).
- [61] Zoë Holmes, Kunal Sharma, M. Cerezo, and Patrick J. Coles. “Connecting ansatz expressibility to gradient magnitudes and barren plateaus”. [PRX Quantum](#) **3**, 010313 (2022).
- [62] Andrew Arrasmith, Zoë Holmes, M Cerezo, and Patrick J Coles. “Equivalence of quantum barren plateaus to cost concentration and narrow gorges”. [Quantum Science and Technology](#) **7**, 045015 (2022).
- [63] Gene H. Golub and Charles F. Van Loan. “Matrix computations - 4th edition”. Johns Hopkins University Press. Philadelphia, PA (2013).
- [64] Roland C. Farrell, Marc Illa, Anthony N. Ciavarella, and Martin J. Savage. “Quantum simulations of hadron dynamics in the schwinger model using 112 qubits”. [Phys. Rev. D](#) **109**, 114510 (2024).
- [65] Google Quantum AI and Collaborators. “Constructive interference at the edge of quantum ergodic dynamics” (2025). [arXiv:2506.10191](#).
- [66] Bernard Nienhuis, Massimo Campostrini, and Pasquale Calabrese. “Entanglement, combinatorics and finite-size effects in spin chains”. [Journal of Statistical Mechanics: Theory and Experiment](#) **2009**, P02063 (2009).
- [67] Man-Duen Choi. “Almost commuting matrices need not be nearly commuting”. [Proceedings of the American Mathematical Society](#) **102**, 529–533 (1988).
- [68] Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B Shah. “Julia: A fresh approach to numerical computing”. [SIAM review](#) **59**, 65–98 (2017).
- [69] Patrick Kofod Mogensen and Asbjørn Nilsen Riseth. “Optim: A mathematical optimization package for Julia”. [Journal of Open Source Software](#) **3**, 615 (2018).
- [70] “Chainrules.jl”. <https://github.com/JuliaDiff/ChainRules.jl> (2024).
- [71] Jutho Haegeman. “Krylovkit.jl”. <https://github.com/Jutho/KrylovKit.jl> (2024).
- [72] Hidetosi Takahasi and Masatake Mori. “Double exponential formulas for numerical integration”. [Publications of the Research Institute for Mathematical Sciences](#) **9**, 721–741 (1973).
- [73] Kazuo Murota and Takayasu Matsuo. “Double-exponential transformation: A quick review of a japanese tradition” (2023). url: <https://doi.org/10.48550/arXiv.2301.01920>.
- [74] Masaaki Nakamura. “DoubleExponentialFormulas”. <https://machakann.github.io/DoubleExponentialFormulas.jl/stable/> (2021).
- [75] David H. Bailey, Karthik Jeyabalan, and Xiaoye S. Li. “A comparison of three high-precision quadrature schemes”. [Experimental Mathematics](#) **14**, 317 – 329 (2005).
- [76] Fredrik Johansson and Marc Mezzarobba. “Fast and rigorous arbitrary-precision computation of gauss-legendre quadrature nodes and weights” (2018). [arXiv:1802.03948](#).
- [77] JuliaApproximation. “FastGaussQuadrature.jl”. <https://github.com/JuliaApproximation/FastGaussQuadrature.jl> (2014).
- [78] Steven G. Johnson. “QuadGK.jl: Gauss–Kronrod integration in Julia”. <https://github.com/JuliaMath/QuadGK.jl> (2013).
- [79] Dirk P. Laurie. “Calculation of gauss-kronrod quadrature rules”. [Mathematics of Computation](#) **66**, 1133–1145 (1997).

A Convergence analysis

In this section we analyze the dependence of the cost function of the BW-like Ansatz on the number of observed time points, the maximum observation time and we discuss the convergence properties of the BW-violating Ansatz. Throughout this section, the TFIM with $N = 8$, $N_A = 4$, $\Gamma = 1$ and OBC is used.

A.1 Cost function and convergence of the midpoint rule for the BW-like Ansatz

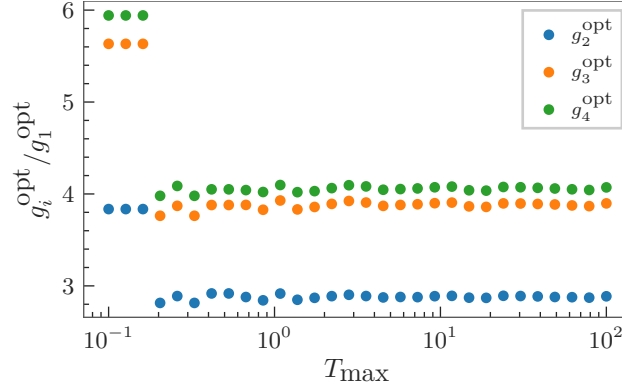


Figure 15: Optimal parameters normalized to g_1^{opt} vs. T_{max} for the Ansatz $\hat{H}_A^{\text{BW}}$. For each T_{max} the same initial parameters have been used. We use $N = 8$, $N_A = 4$, $\Gamma = 1$ and open boundary conditions.

Table 3 contains the normalized optimal parameters $\mathbf{g}^{\text{opt}}/g_1^{\text{opt}}$ rounded to 15 decimal places obtained via the midpoint rule with different Δt and the Tanh-sinh quadrature (see appendix D) and the corresponding cost function value $\mathcal{C}(\mathbf{g}^{\text{opt}})$. Here, the BW-like Ansatz $\hat{H}_A^{\text{BW}}$ is used. The maximum observation time is $T_{\text{max}} = 1$. All runs are initialized with the same initial parameters.

Here, the fact that the cost function value at the minimum $\mathcal{C}(\mathbf{g}^{\text{opt}})$ is not numerically zero is crucial.

Table 3: Normalized optimal parameters $\mathbf{g}^{\text{opt}}/g_1^{\text{opt}}$ for the Ansatz $\hat{H}_A^{\text{BW}}$ obtained with the midpoint rule with different Δt and the Tanh-sinh quadrature (101 evaluations). All runs were initialized with the same initial parameters. We use the TFIM with $N = 8$, $N_A = 4$, $\Gamma = 1$ and open boundary conditions.

method	$g_2^{\text{opt}}/g_1^{\text{opt}}$	$g_3^{\text{opt}}/g_1^{\text{opt}}$	$g_4^{\text{opt}}/g_1^{\text{opt}}$	$\mathcal{C}(\mathbf{g}^{\text{opt}})$
$\Delta t = 10^{-1}$	3.824 039 596 427	5.608 777 483 165	5.921 893 959 512	2.5110×10^{-5}
$\Delta t = 10^{-2}$	3.835 351 726 582	5.632 970 094 804	5.941 793 208 331	2.0946×10^{-5}
$\Delta t = 10^{-3}$	3.835 421 398 527	5.633 117 585 462	5.941 909 050 130	2.0948×10^{-5}
$\Delta t = 10^{-4}$	3.835 422 093 758	5.633 119 057 108	5.941 910 205 525	2.0948×10^{-5}
$\Delta t = 10^{-5}$	3.835 422 100 710	5.633 119 071 825	5.941 910 217 079	2.0948×10^{-5}
$\Delta t = 10^{-6}$	3.835 422 100 782	5.633 119 071 975	5.941 910 217 197	2.0948×10^{-5}
Tanh-sinh	3.835 422 100 780	5.633 119 071 973	5.941 910 217 195	2.0948×10^{-5}

It can be seen that the optimal parameters obtained with the midpoint rule approach the optimal parameters calculated with the Tanh-sinh quadrature asymptotically from below as the time step size Δt decreases, except for $\Delta t = 10^{-6}$, where the ratios exceed those obtained with the Tanh-sinh quadrature. Note that for $\Delta t = 10^{-2}$ the integrand is evaluated 100 times but only two decimal places match the results of the Tanh-sinh quadrature, while for $\Delta t = 10^{-6}$ eleven decimal places are in agreement with the Tanh-sinh quadrature but the integrand is evaluated at 100000 time points. The Tanh-sinh quadrature is more efficient, since it evaluates the integrand only 101 times at the minimum of the cost function and gives highly accurate results. These results demonstrate that, to get accurate

Table 4: Optimal parameters normalized to Γ_1^{opt} for the Ansatz $\hat{H}_A^{\text{BWV}}$ and its corresponding minimum of the cost function for different time steps Δt for the midpoint rule. All runs were initialized with the same initial parameters. We use the TFIM with $N = 8$, $N_A = 4$, $\Gamma = 1$ and open boundary conditions.

method	$J_{1,2}^{\text{opt}}/\Gamma_1^{\text{opt}}$	$\Gamma_2^{\text{opt}}/\Gamma_1^{\text{opt}}$	$J_{2,3}^{\text{opt}}/\Gamma_1^{\text{opt}}$
$\Delta t = 10^{-1}$	1.965 946 199 368	2.864 944 458 809	3.666 380 470 827
$\Delta t = 10^{-2}$	1.965 946 199 368	2.864 944 458 809	3.666 380 470 827
$\Delta t = 10^{-3}$	1.965 946 199 368	2.864 944 458 809	3.666 380 470 827
$\Delta t = 10^{-4}$	1.965 946 199 368	2.864 944 458 809	3.666 380 470 827
$\Delta t = 10^{-5}$	1.965 946 199 368	2.864 944 458 809	3.666 380 470 827
Tanh-sinh	1.965 946 199 368	2.864 944 458 809	3.666 380 470 827

$\Gamma_3^{\text{opt}}/\Gamma_1^{\text{opt}}$	$J_{3,4}^{\text{opt}}/\Gamma_1^{\text{opt}}$	$\Gamma_4^{\text{opt}}/\Gamma_1^{\text{opt}}$	$\mathcal{C}(\mathbf{g}^{\text{opt}})$
4.342 962 293 250	4.871 649 743 586	5.234 439 004 803	0
4.342 962 293 250	4.871 649 743 585	5.234 439 004 803	0
4.342 962 293 250	4.871 649 743 585	5.234 439 004 803	0
4.342 962 293 250	4.871 649 743 585	5.234 439 004 803	0
4.342 962 293 250	4.871 649 743 586	5.234 439 004 803	0
4.342 962 293 250	4.871 649 743 586	5.234 439 004 803	0

results with the BW-like Ansatz, it is not sufficient to observe the observables at a few arbitrary time points, underlining that the cost function should not be treated as a discrete sum over a few time points but rather as an integral over the time domain.

A.2 Influence of the maximum observation time for the BW-like Ansatz

Figure 15 shows the ratios of the optimal parameters vs. T_{max} , for the BW-like Ansatz $\hat{H}_A^{\text{BW}}$. Here, each run was initialized with the parameters $\mathbf{g}_{\text{init}} = (6 \ 12 \ 15 \ 17)^T$. The cost function is evaluated with the Tanh-sinh quadrature. It can be seen that the optimal parameters exhibit oscillations and are not converged for the displayed T_{max} . This observation is a fundamental consequence of the finite value for the cost function $\mathcal{C}(\mathbf{g}^{\text{opt}})$, as the system still shows dynamics on the order of the cost function value. The parameters will not converge, if the variational Ansatz captures the EH only poorly. Thus T_{max} will always have an influence on the optimal parameters so that no convergence will be reached if the BW-like Ansatz is used. Additionally we see for very small T_{max} a jump in the parameters. This is an artificial solution, that occurs if the typical oscillation of the cost function can not be captured by the time window anymore.

A.3 Convergence properties of the BW-violating Ansatz

The previous discussions on convergence were based on the BW-like Ansatz $\hat{H}_A^{\text{BW}}$, where the minimum of the cost function value was finite. In this subsection we investigate the convergence properties of the $\hat{H}_A^{\text{BWV}}$ Ansatz. We choose $T_{\text{max}} = 1$ and the initial parameters $\mathbf{g}_{\text{init}} = (3 \ 5 \ 8 \ 10 \ 12 \ 14 \ 15)^T$ were used for all runs. Note that the index i of a parameter g_i is not directly related to the i -th lattice site for the Ansatz $\hat{H}_A^{\text{BWV}}$.

Table 4 shows the optimal parameters normalized to Γ_1^{opt} for all different time steps Δt . It can be seen that all normalized parameters agree up to twelve decimal places. The second study concerns the influence of the maximum observation time T_{max} . Here, the Tanh-sinh quadrature is used and each run uses a different T_{max} . Each run is initialized with the same initial parameters as in the first study. Table 5 shows the optimal parameters normalized to Γ_1^{opt} for all different T_{max} . All normalized

Table 5: Optimal parameters normalized to Γ_1^{opt} for the Ansatz $\hat{H}_A^{\text{BWV}}$ and its corresponding minimum of the cost function for different T_{max} . The Tanh-sinh quadrature was used to evaluate the cost function. All runs were initialized with the same initial parameters. We use the TFIM with $N = 8$, $N_A = 4$, $\Gamma = 1$ and open boundary conditions.

T_{max}	$J_{1,2}^{\text{opt}}/\Gamma_1^{\text{opt}}$	$\Gamma_2^{\text{opt}}/\Gamma_1^{\text{opt}}$	$J_{2,3}^{\text{opt}}/\Gamma_1^{\text{opt}}$
0.1	1.965 946 199 360	2.864 944 458 793	3.666 380 470 804
1	1.965 946 199 368	2.864 944 458 809	3.666 380 470 827
10	1.965 946 199 368	2.864 944 458 809	3.666 380 470 827
100	1.965 946 199 368	2.864 944 458 809	3.666 380 470 827
$\Gamma_3^{\text{opt}}/\Gamma_1^{\text{opt}}$	$J_{3,4}^{\text{opt}}/\Gamma_1^{\text{opt}}$	$\Gamma_4^{\text{opt}}/\Gamma_1^{\text{opt}}$	$\mathcal{C}(\mathbf{g}^{\text{opt}})$
4.342 962 293 221	4.871 649 743 552	5.234 439 004 767	0
4.342 962 293 250	4.871 649 743 586	5.234 439 004 803	0
4.342 962 293 250	4.871 649 743 586	5.234 439 004 803	0
4.342 962 293 250	4.871 649 743 586	5.234 439 004 803	0

parameters agree up to 13 decimal places, except for $T_{\text{max}} = 0.1$, where only nine decimal places of the optimal ratios agree with the optimal ratios of the runs with higher T_{max} .

Both results demonstrate a robust behaviour with respect to the number of observation times and the maximum observation time.

The results are converged for all Δt and almost all T_{max} if a good Ansatz is chosen, i.e. an Ansatz that leads to a cost function that is numerically zero at its minimum, which is the case for the BW-violating Ansatz for the TFIM with $\Gamma = 1$. The mismatch between the convergence properties of the $\hat{H}_A^{\text{BW}}$ and the $\hat{H}_A^{\text{BWV}}$ Ansatz can be explained by the values of the cost function. If the cost function reaches numerically zero for an Ansatz, the subsystem becomes constant in time and no dynamics is present at the optimal solution $\mathbf{g}^{\text{opt}}$. In this case Δt and T_{max} become irrelevant. On the other hand, if an Ansatz does not capture the EH accurately, the cost function will be finite at the minimum. Thus, even at the minimum, dynamics will be present s.t. it is not irrelevant how often or how long the system is sampled.

Since it is not known a priori how accurate an Ansatz is, we conclude that it is best practice to evaluate the cost function with the Tanh-sinh quadrature to reduce the number of required measurements. Note, one parameter can be fixed throughout the optimization if the Ansatz is good and the cost function drops to numerically zero. Fixing one parameter effectively fixes the “entanglement energy scale”, which has an influence on the time scale.

B Violation of the BW-theorem in the XXZ model

Here we compare the BW-like and BW-violating Ansatz in the XXZ model with $N = 10$, $N_A = 5$, $T_{\text{max}} = 1$ and $\Delta = -0.5$. The analysis is the same as for the TFIM (see Section 3.1). The optimal parameters exhibit huge standard deviations in the case of the XXZ model with OBC and the BW-like Ansatz $\hat{H}_A^{\text{BW}}$ (Fig. 16(a)). The differences of the optimal ratios obtained via the QCFL and the commutator show huge deviations, where as the deviations are smaller in the case of PBC (Fig. 16(b)). Again, the linear rise near the entanglement cut and bending at the right border become apparent in the case of OBC, while the symmetric behaviour can be observed if the composite system obeys PBC. However, the large error bars indicate that there many suboptimal or local minima, which are found by the optimizer. The minimum of the cost function is not numerically zero, namely $\mathcal{C}(\mathbf{g}^{\text{opt}}) \approx 3.54(10) \times 10^{-7}$ (OBC) and $\mathcal{C}(\mathbf{g}^{\text{opt}}) \approx 4.35(4) \times 10^{-6}$ (PBC). Figure 16(c),(d) show the optimal parameters after filtering obtained with the Ansatz $\hat{H}_A^{\text{BWV}}$ for OBC and PBC, respectively. The parameters obtained with the QCFL and the commutator show good agreement and exhibit very

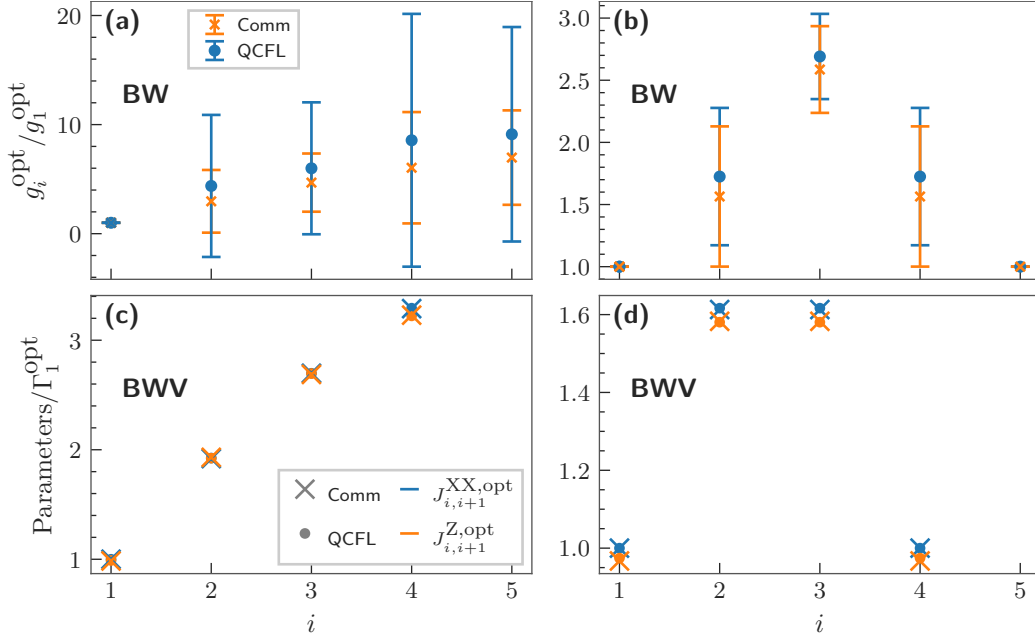


Figure 16: 50 runs with different random initial parameters for the XXZ model with the QCFL and the commutator as a cost function (Comm). Outliers from the QCFL were filtered out. The mean value from the 50 runs are depicted. (a),(b) Optimal parameters for the BW-like Ansatz with OBC and PBC, respectively. The errorbars are given by the standard deviation over the 50 runs (outliers from the QCFL were filtered out). (c),(d) Optimal parameters for the BW-violating Ansatz with OBC and PBC, respectively. We use $N = 10$, $N_A = 5$, $\Delta = -0.5$ and open boundary conditions.

small standard deviations (see Fig. 16(c),(d)). Remarkably, in the case of OBC, the parameters only deviate slightly from the BW theorem. That is, the ratios of the coupling in the x - and y -direction, and the coupling in z -direction, $J_{i,i+1}^{XX, \text{opt}}/J_{i,i+1}^{Z, \text{opt}}$, is 1.0219 ± 0.0005 at most, which almost agrees with the BW theorem, predicting a value of 1. The deviation of the parameters from the BW theorem in the case of PBC is similar. The cost function value at the optimum takes the value of $\mathcal{C}(\mathbf{g}^{\text{opt}}) \approx 1.12(27) \times 10^{-9}$ (OBC) and $\mathcal{C}(\mathbf{g}^{\text{opt}}) \approx 2.51(55) \times 10^{-8}$ (PBC). The deviation from the BW theorem is not negligible, underlined by the drop in the cost function value by approximately three (OBC) and two (PBC) orders of magnitude if the BW-violating Ansatz is used. Although the cost function is noticeably lower with the Ansatz $\hat{H}_A^{\text{BWV}}$ and parameters obtained via the QCFL and the commutator match very well, the minimum of the cost function is still finite.

C Methods

We use the Julia programming language[68]. For the optimization of the cost function, we chose the implementation of the LBFGS algorithm from the package `Optim.jl`[69]. We compute the time evolution operator by matrix exponentiation. The gradient is computed with an exact expression, utilizing the Fréchet derivative of the time evolution operator, i.e. an matrix exponential. For its computation we adapted the implementation from the package `ChainRules.jl`[70]. We chose the infinity norm of the gradient as a convergence criterium, which is referred to as ∇_{tol} . In order to stop a minimization, $\nabla_{\text{tol}} \leq 10^{-16}$ must hold if not mentioned otherwise. The monitored observables are $\{Z_i Z_{i+1} | 1 \leq i < N_A\}$. To obtain the ground state, we use exact diagonalization if the composite system Hilbert space dimension is $\dim \mathcal{H} \leq 1024$ and the Lanczos algorithm from the package `KrylovKit.jl`[71] otherwise.

D Numerical integration methods

Here we briefly introduce the used integration methods. Note that the integrand changes with each new parameters $\mathbf{g}$, and thus, the integral can be seen as a blackbox. This is why an adaptive or iterative integration method is a good choice to achieve convergence. Additionally, the integrand is oscillatory, which poses another challenge.

D.1 Right point rule

With the right point rule, the integral over the interval $[a, b]$ is evaluated at n points, equidistantly separated by a step size $\Delta t = (b - a)/n$ starting at $a + \Delta t$

$$I = \int_a^b f(t)dt \approx \sum_{i=1}^n f(a + i\Delta t)\Delta t. \quad (39)$$

The right point rule converges linearly in n .

D.2 Mid point rule

The mid point rule is, like the right point rule, a rectangular integration method with the difference that the evaluation points are shifted to the left by $\Delta t/2$, s.t. the approximation reads

$$I = \int_a^b f(t)dt \approx \sum_{i=1}^n f\left(a + \left(i - \frac{1}{2}\right)\Delta t\right)\Delta t. \quad (40)$$

The mid point rule converges quadratically in n .

D.2.1 Tanh-sinh method

The Tanh-sinh quadrature is one quadrature formula of a whole family, the Double Exponential Formulas[72] (DE Formulas).

Starting from an integral over the interval $[-1, 1]$

$$I = \int_{-1}^1 f(t)dt, \quad (41)$$

the DE Formulas utilize a variable transformation $x = \Phi(u)$ mapping the boundaries to infinity, i.e., $\phi(-\infty) = -1$ and $\phi(\infty) = 1$. That is, the integral reads

$$I = \int_{-\infty}^{\infty} f(\Phi(u))\Phi'(u) du. \quad (42)$$

Applying the trapezoidal rule with a step size h to Eq. (42) and truncating the number of evaluation points to a finite value yields

$$I_h = h \sum_{j=-M}^M f(\Phi(jh))\Phi'(jh) \quad (43)$$

The error stems from discretization and the truncation of the infinite sum [73]. The best balance between the discretization error and the truncation error is achieved by a variable transformation for which the integrand decays double exponentially, giving the DE Formulas their name. The double exponential decay is achieved by the variable transformation

$$\Phi(u) = \tanh\left(\frac{\pi}{2} \sinh(u)\right), \quad (44)$$

giving the Tanh-sinh quadrature its name. We adapted the implementation from the package `DoubleExponentialFormulas.jl`[74] and added a reliable error estimation scheme presented in Reference [75].

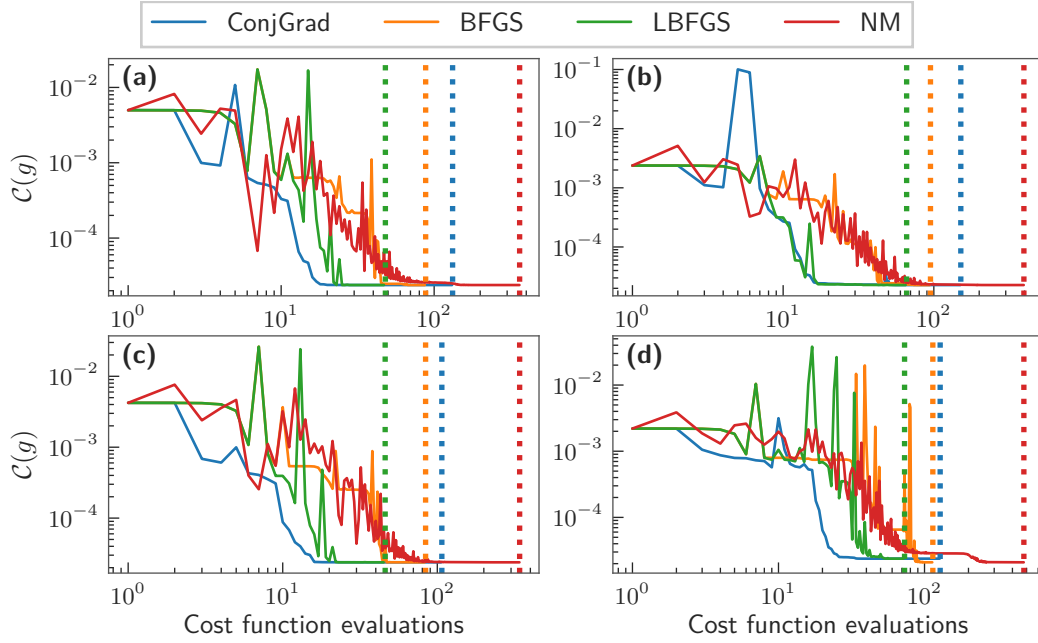


Figure 17: Cost function value vs. number of cost function evaluations in one minimization run. Random initial parameters are sampled and the cost function is minimized with the different algorithms. This procedure is done four times (run one, two, three and four in (a),(b),(c) and (d), respectively). The vertical dotted lines indicate the points, where the optimizers are converged. We use the TFIM with $N = 8$, $N_A = 4$, $\Gamma = 1$, open boundary conditions and the BW-like Ansatz $\hat{H}_A^{\text{BW}}$.

D.3 Gauß-Legendre quadrature

The n -point Gauß-Legendre quadrature approximates the integral

$$\int_{-1}^1 f(t)dt = \sum_{i=0}^{n-1} w_i f(t_i), \quad (45)$$

where t_i are the roots of the n -th Legendre polynome P_n and the weights are given by[76]

$$w_i = \frac{2}{(1 - t_i^2)(P'_n(t_i))^2}. \quad (46)$$

We use the implementation of the package `FastGaussQuadrature.jl`[77].

D.4 Adaptive Gauß-Kronrod quadrature

The Gauß-Kronrod quadrature is an extension of the standard Gauß-quadratures. Starting from a n' -point Gauß-quadrature, they are extended with $n' + 1$ additional points to a $2n' + 1$ quadrature scheme. This has the advantage that the Gauß-Kronrod quadrature has a good error estimate by computing the difference of the standard Gauß quadrature and its Kronrod extension, where in contrast a standalone Gauß-quadrature cannot reuse already computed function values to estimate an error s.t. it would need to compute more function values in order to get an error estimate. For more details, we refer to the documentation of the used package `Quadgk.jl`[78] and the seminal work by Laurie[79].

E Optimizer benchmark

This section shows that the used method for optimization is the best methods for this kind of problem among all the methods, which are used for the comparison. The model used is the TFIM with $N = 8$, $N_A = 4$, OBC and $\Gamma = 1$ with the variational Ansatz $\hat{H}_A^{\text{BW}}$.

Table 6: Runtime of the algorithms for the four sets of random initial parameters (i.e. four different runs). We use the TFIM with $N = 8$, $N_A = 4$, $\Gamma = 1$, open boundary conditions and the BW-like Ansatz $\hat{H}_A^{\text{BW}}$.

algorithm	runtime / s			
	run 1	run 2	run 3	run 4
ConjGrad	2.400	3.088	1.898	2.271
BFGS	1.727	1.914	1.625	2.494
LBFGS	0.940	1.372	0.909	1.377
NM	1.873	2.161	1.779	2.505

The algorithms used for comparison are Conjugate Gradient (ConjGrad), BFGS and Nelder-Mead (NM), which are all implemented in the package `Optim.jl` as well. The maximum integration time is set to $T_{\text{max}} = 1$. The first three algorithms are gradient-based, whereas the Nelder-Mead algorithm is a direct search method and therefore does not require information about the gradient or Hessian. Initial parameters are chosen at random and the cost function is minimized with all different algorithms. The convergence criterium for the gradient-based methods is set to $\nabla_{\text{tol}} = 10^{-16}$. NM uses a so called simplex, which consists of multiple points in the parameter space. The convergence criterium here is the standard deviation of the cost function value at these points in the current simplex, since the cost function value at the points of the simplex should be equal in the vicinity of the minimum. In this test, the standard deviation is required to be less than or equal to 10^{-16} , too. The cost function value vs. the number of cost function evaluations in one minimization run is shown in Fig. 17. For all four different initial parameters, the LBFGS algorithm needed the fewest cost function evaluations, as indicated by the vertical dotted lines. However, in the fourth run (Fig. 17(d)), the BFGS and Nelder-Mead algorithms found a slightly lower minimum ($\approx 3 \times 10^{-5}$ smaller) than the BFGS and Conjugate Gradient algorithms. Table 6 lists the time it took to minimize the cost function among all runs for all different algorithms. The minimization was repeated ten times for one set of initial parameters and the smallest time over these ten repetitions is listed. In all runs, the LBFGS algorithm won again in terms of runtime. To conclude, even tho the LBFGS algorithm did not find the lowest minimum in the fourth run, the efficiency is very convincing. Additionally, this happened only one out of four times with random initial parameters. With a good initial guess this should not happen. The Gradient descent, ADAM and Simulated annealing algorithm have also been tested but not listed, because their performance were much worse than the algorithms included in the benchmark. It needs to be mentioned that a model with only 4 parameters has been used. To get deeper insights into the performance of the optimization algorithms, a model with significantly more parameters could prove helpful.

F Long range corrections in the XXZ model

We investigate if the cost function for the XXZ can be further reduced by including long-range corrections to the variational entanglement Hamiltonian. The variational Ansatz, together with the corrections $\hat{H}_A^c$, reads

$$\hat{H}_A^{\text{Var}} = \hat{H}_A^{\text{BWV}} + \hat{H}_A^c = \sum_{r=1}^{r_{\text{max}}} \sum_{i=1}^{N_A-r} \left(J_{i,i+r}^{\text{XX}} (X_i X_{i+r} + Y_i Y_{i+r}) + J_{i,i+r}^Z \Delta Z_i Z_{i+r} \right), \quad (47)$$

where $\{J_{i,i+r}^{\text{XX}}, J_{i,i+r}^Z\}$ act as variational parameters. The quantity r_{max} determines the maximum range of interaction. Every term beyond $r = 1$ is a part of the long-range interactions, and thus, part of the corrections. The optimal parameters from Section B of the BW-violating Ansatz $\hat{H}_A^{\text{BWV}}$ are used for initialization of the parameters for $r > 1$. All long-range couplings (beyond $r = 1$) are initialized to

Table 7: Optimal parameters $\{J_{i,i+r}^{\text{XX}}, J_{i,i+r}^{\text{Z}}\}$ including all long-range interactions for $r_{\text{max}} = 4$ rounded to four decimal places. Here we use $N = 10$, $N_A = 5$, $\Delta = -0.5$ and open boundary conditions.

	$r = 1$	$r = 2$	$r = 3$	$r = 4$
$J_{1,1+r}^{\text{XX}}$	0.8435	0.0056	0.0013	0.0032
$J_{1,1+r}^{\text{Z}}$	0.8821	0.0482	0.0447	0.0474
$J_{2,2+r}^{\text{XX}}$	1.6238	-0.0081	0.0027	
$J_{2,2+r}^{\text{Z}}$	1.6379	0.0111	0.0265	
$J_{3,3+r}^{\text{XX}}$	2.3228	0.0673		
$J_{3,3+r}^{\text{Z}}$	2.3268	0.0975		
$J_{4,4+r}^{\text{XX}}$	2.8433			
$J_{4,4+r}^{\text{Z}}$	2.6806			

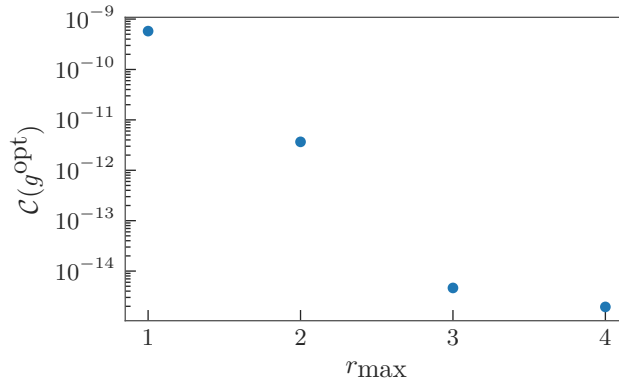


Figure 18: Cost function $\mathcal{C}(g^{\text{opt}})$ at the found solution vs. the maximum range of interaction r_{max} included in the correction term. Here we use $N = 10$, $N_A = 5$, $\Delta = -0.5$ and open boundary conditions.

zero, as these are expected to be small. Table 7 lists the optimal parameters for a run with $r_{\text{max}} = 4$, i.e., all long-range terms included. It can be seen that the corrections are at least one magnitude smaller in comparison to the parameters for $r = 1$. The parameters show a decay with the range of interaction r . We observe in Fig. 18, that the variational Ansatz containing long-range corrections indeed reduces the value of the cost function, demonstrating that such corrections to the BW ansatz are crucial to represent the EH. Note that the norm of the gradient was very low in case of $r_{\text{max}} = 3$ and $r_{\text{max}} = 4$, and we terminated the optimization after 100000 iterations.

G Extrapolation into the thermodynamic limit

To obtain the couplings $J_{i,i+1}^{\text{XX,opt}}$ and $J_{i,i+1}^{\text{Z,opt}}$ in the TDL, the procedure is the following:

1. Extract the ground state of the system Hamiltonian for different lattice sizes N (up to $N = 29$ could have been achieved with the Lancos algorithm).
2. Construct the RDM with the ground state obtained in step one for a subsystem chain length N_A .
3. Run the algorithm with the BW-violating Ansatz $\hat{H}_A^{\text{BWV}}$ with $J_{i,i+1}^{\text{XX}}$ and $J_{i,i+1}^{\text{Z}}$ as variational parameters for the different RDMs for each composite system size N from step 2 for a subsystem chain length N_A .
4. Plot the ratios of the obtained parameters $J_{i,i+1}^{\text{XX,opt}}/J_{1,2}^{\text{XX,opt}}$ and $J_{i,i+1}^{\text{Z,opt}}/J_{1,2}^{\text{XX,opt}}$ vs. $1/N^2$ and extrapolate for $1/N \rightarrow 0$, i.e., into the TDL.

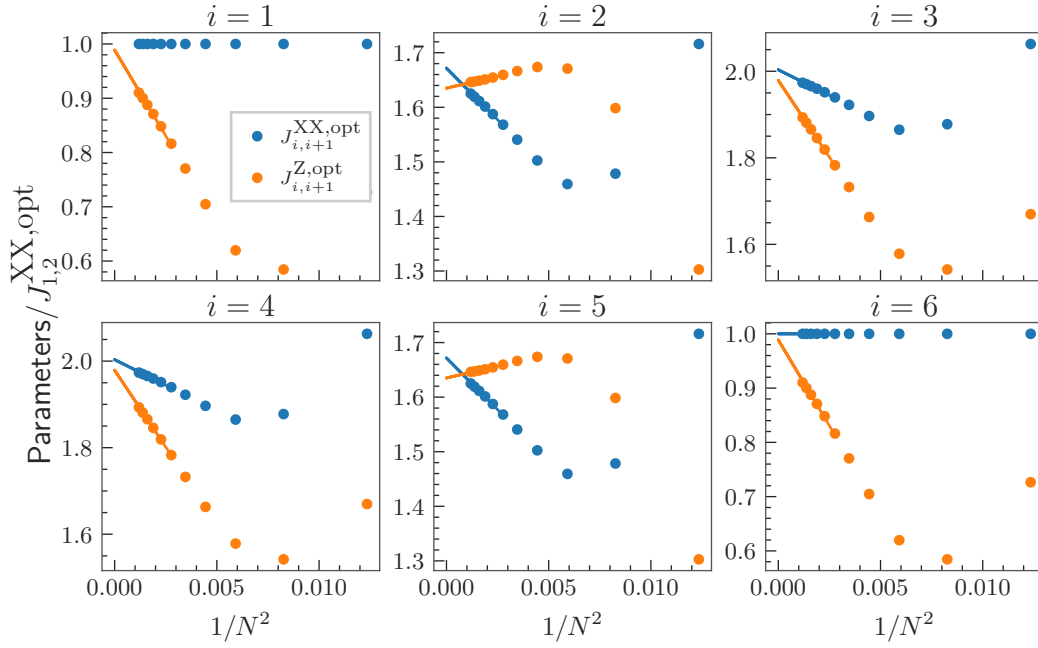


Figure 19: Optimal parameters $J_{i,i+1}^{XX,\text{opt}}$ and $J_{i,i+1}^{Z,\text{opt}}$ normalized to $J_{1,2}^{XX,\text{opt}}$ vs. $1/N^2$ for each lattice site i in the subsystem A for $N_A = 7$, $\Delta = -0.5$ and PBC. The solid lines are linear fits.

5. Repeat step two to four for different subsystem chain lengths N_A .

Figure 19 shows the obtained optimal parameters normalized to $J_{1,2}^{XX,\text{opt}}$ vs. $1/N^2$ for $N_A = 7$. The index i indicates the lattice site and the solid lines are the corresponding fits. For brevity, the plots for other N_A are omitted. A linear fit was used

$$\frac{J_{i,i+1}^{\Omega,\text{opt}}}{J_{1,2}^{XX,\text{opt}}}(N) = p_1 \frac{1}{N^2} + p_2, \quad \Omega = \text{XX}, \text{Z} \quad (48)$$

where p_1 and p_2 act as parameters for the fit. The optimal parameters normalized to $J_{1,2}^{XX,\text{opt}}$ extrapolated in to the TDL ($N \rightarrow \infty$) then are

$$\frac{J_{i,i+1}^{\Omega,\text{opt}}}{J_{1,2}^{XX,\text{opt}}}(N \rightarrow \infty) = p_2, \quad \Omega = \text{XX}, \text{Z}. \quad (49)$$

H Zero-noise extrapolation

For each point in the phase diagram and number of measurements $N_M \in [10^3, 10^9]$ (TFIM) and $N_M \in [10^3, 10^{11}]$ (XXZ model), we sampled the cost function ten times at the optimal parameters obtained with optimizations without noise and averaged over these ten samples. We used the Gauß-Legendre quadrature with five time points and $T_{\text{max}} = 2$. Our fit model is based on the fact that the cost function is a second order polynomial in $1/\sqrt{N_M}$ in the presence of noise

$$C(N_M; \mathbf{g}_{\text{opt}}) = a \frac{1}{N_M} + b \frac{1}{\sqrt{N_M}} + c. \quad (50)$$

Since we want to fit over several orders of magnitude, we first need to take the logarithm of our data to prevent that the higher data points dominate the errors in the least-square algorithm. Our final fit model $y(x)$ is

$$y(x) = \log_{10}(a10^{2x} + b10^x + c) \quad (51)$$

with $x = \log_{10}(1/\sqrt{N_M})$. The extrapolated value, i.e. $N_M \rightarrow \infty$, then is the fit parameter c .

I XXZ model across the phase diagram for larger systems

Here, in Fig. 20, we show the minimum of the cost function in dependence on the anisotropy Δ for different system sizes in the XXZ model without noise. $N = 2N_A$ with OBC always holds and we use the BW-violating Ansatz. All values below double precision have been omitted.

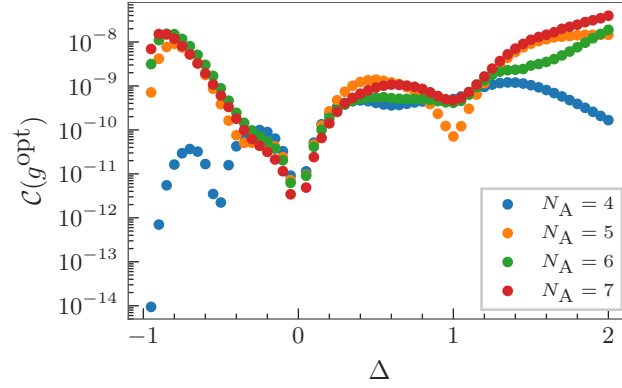


Figure 20: Minimum of the cost function for the XXZ model with varying anisotropy Δ for different system sizes. $N = 2N_A$ with OBC always holds and we use the BW-violating Ansatz $\hat{H}_A^{\text{BWV}}$.