PNAS
www.pnas.org

Supplementary Information for

From Bloch Oscillations to Many Body Localization in Clean Interacting Systems

Evert P. L. van Nieuwenburg, Yuval Baum, Gil Refael

Evert P. L. van Nieuwenburg.
E-mail: evert@caltech.edu

This PDF file includes:

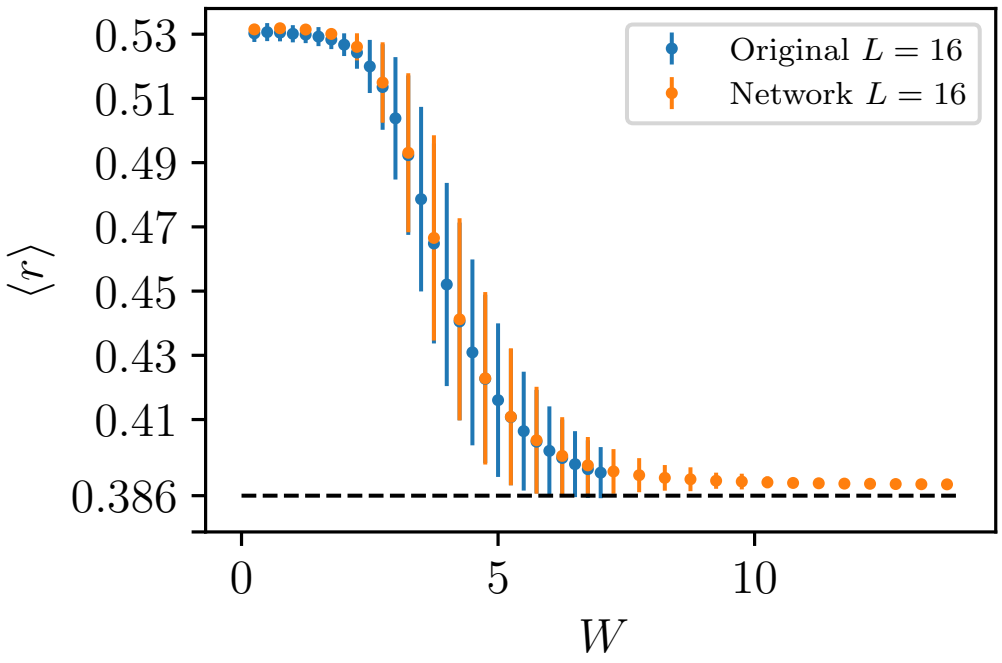

Fig. S1. The original $L = 16$ data and the machine learned map from the disorder realization h_1 through h_{16} to the resulting r-statistics. With the network we are able to generate considerably more realizations (10^6 versus 10^3) in a much shorter timespan, provided that the network is capable of learning and generalizing. The sigmoid output neuron rather than linear for optimizing the mean-squared-error ensures convergence of the output as a function of W. Error bars indicate the standard deviation over the number of realizations, and the black dashed horizontal line indicates the Poissonian r-value of $\ln 4 - 1$.

Supporting Information Text

Data augmentation using machine learning

The different disorder realizations we study in this manuscript differ only in the values for the on-site potentials. Given the on-site potentials, there exists a procedure that results in the value for the r-statistics. Namely, one builds the corresponding Hamiltonian matrix and diagonalizes it to obtain the eigenvalues ϵ_n. The r-statistics is obtained by looking at neighboring eigenvalue differences $\delta_n = \epsilon_{n+1} - \epsilon_n$ and computing the ratio $r = \langle \min(\delta_n, \delta_{n+1}) / \max(\delta_n, \delta_{n+1}) \rangle$ as discussed in the main text.

Here, however, we ask whether or not a more direct (approximate) map exists from the on-site potentials to r. Rather than trying to explicitly construct it, we attempt to train a neural network to perform this map for us. Hence we generate a large data-set of pairs $(\mathbf{h}, r)$, where $\mathbf{h}$ is a vector of the on-site potentials augmented with the value of W from which they were drawn, and r is the resulting r-statistics for this particular realization. These serve as the input and output respectively for the machine learning model.

Provided that such a mapping exists and that the network is capable of learning it, the resulting network can be used to generate more r-values by using it to predict on more realizations. This allows one to generate statistics much faster compared to running the full exact diagonalization. It must be noted that this procedure cannot take away the inherent statistical uncertainty due to the finite size of the system. Particularly, for disorder strengths near the transitions point, the exact r-values of systems with different realizations drawn from the same distribution, lie within a relatively large window. As the system becomes larger this window becomes smaller. For example, already by including a few hundreds of realizations, for $L = 16$, the error bars near the transition are dominated by the intrinsic finite size effect and cannot be improved by adding more realizations.

In Fig. S1 we demonstrate the above procedure for the $L = 16$ data, for which the data-set consists of $\sim$ 15k entries (25 values of W spread over $\sim$ 550 realizations). We split off 10% of the data as a validation set, and train a network with the following architecture. First, two convolutional layers with 32 filters and kernel sizes 6 and 3, followed by a maximum pooling of size 3. Then a convolutional layer with 64 filters and kernel size 2, followed by a global average pooling. Next, two fully connected sigmoid layers with 256 and 128 neurons respectively, and dropout 0.5. And finally an output layer with a single sigmoid neuron.

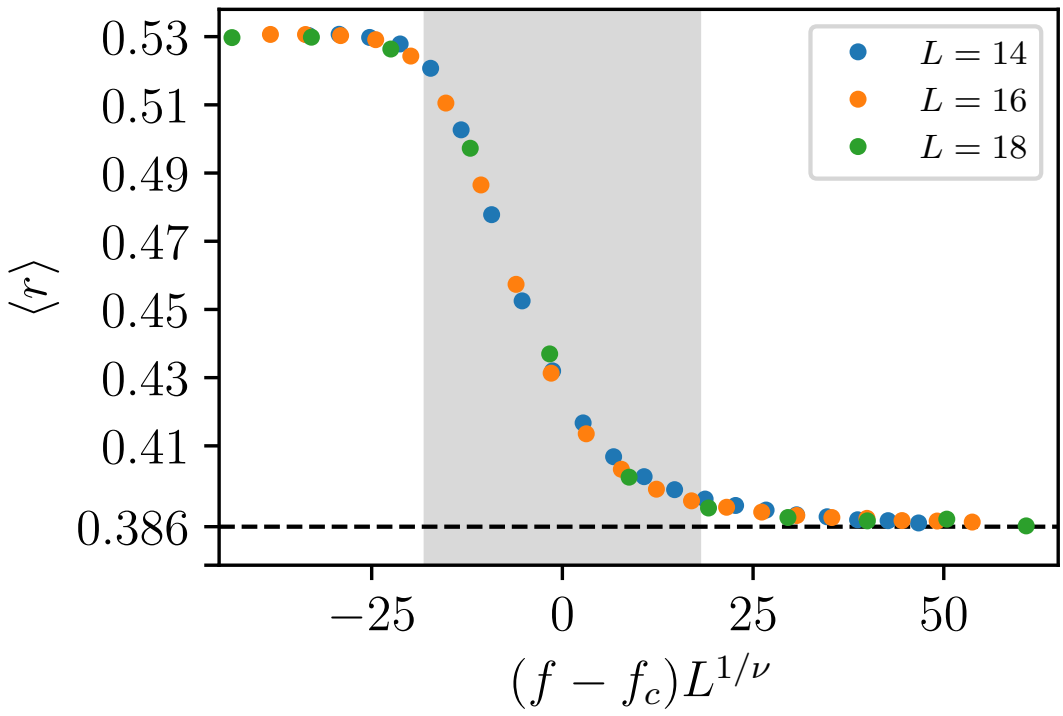

Fig. S2. Collapse of the $W = 0.5$ data for system sizes $L = 14, 16, 18$, as a function of the field strength f. The collapse is obtained by rescaling the fields according to $f \to (f - f_c)L^{1/\nu}$ with $f_c = 2.08$ and $\nu = 0.952$. The gray area indicates the width w that was used to make the curves collapse, and is the width at which the collapse is most stable against inclusion or removal of the $L = 12$ data.

We train the network with the Adam (1) optimizer to minimize the mean-squared-error loss function, and achieve a validation loss of $\sim 2 \cdot 10^{-5}$ in 100 epochs of batchsize 32. In our experiments, we have found no particular reason for the above network to work better than others, but we found that considerably simpler networks (e.g. just fully connected layers) converge much slower. For the purpose of extracting the mapping, our chosen network might be hard to interpret. It would be an interesting research direction however to see if the approximate mapping can be extracted from a network, or whether a single network can be trained on different system sizes to extract finite size behavior. Both would potentially allow predictions to be made on larger system sizes than trained on, although further investigation into this question is required.

Finite size scaling

In this appendix we discuss the transition from the ergodic to the non-ergodic phase as a function of the linear field f. To do so, we fix $W = 0.5$ and perform a finite size scaling analysis attempting to collapse the curves for different system sizes. We consider a universal function $g\left((f - f_c)\, L^{1/\nu}\right)$ for the r-statistics, and optimize the parameters f_c and ν so that the rescaled r-statistics curves for the different sizes collapse.

Each of the curves is first rescaled with proposed f_c and ν after which we use spline interpolation to numerically minimize the cost function $C(f_c, \nu) = \sum_{i<j} \int_x (y_i(x) - y_j(x))^2$, where i, j both run over system sizes $L = 12, 14, 16, 18$ and $y_i(x)$ represents the spline-interpolated data. The integration regime x is taken to be centered around the transition (i.e. $x = 0$) and has a width $2w$ that we vary to obtain statistics on f_c and ν. In the collapse including the system size $L = 12$ data, the $L = 12$ curve is consistently the most off. In the spirit of Ref. (2) we consider the width w for which the extracted parameters are least sensitive to the inclusion/removal of the $L = 12$ data. This results in the parameters $f_c = 2.08 \pm 0.10$ and $\nu = 0.952(5)$. The resulting collapse for this set of parameters is shown in Fig. S2.

Choice of gauge for numerics

We chose to work with a time-independent Hamiltonian for which the linear field is added via the dipole term, rather than as a time-dependent phase factor for the hopping. This interpretation brings with it the potential issue of having an infinite energy difference between the endpoints of our system as one scales up to the thermodynamic limit. The physics in these two gauges is evidently invariant, but since we consider (rather small) finite size systems the infinite energies are not a concern. Working in the time-independent gauge is numerically considerably more convenient, since the time evolution operator over a period T, i.e. $U(T)$, can be constructed by a single exponentiation through as $U(T) = \exp\left(-iH_{\text{static}}T\right)$. For the time-dependent case, one would have to compute the time ordered integral $U(T) = \mathcal{T} \exp\left(-i \int \mathrm{d}t H(t)\right)$ by breaking it down into many small $\mathrm{d}t$-sized steps and exponentiating $H(t)$ for each. The resulting differences in the spectra $\lambda_{i,\text{static}}$ and $\lambda_{i,\text{time}}$ are only of the order of

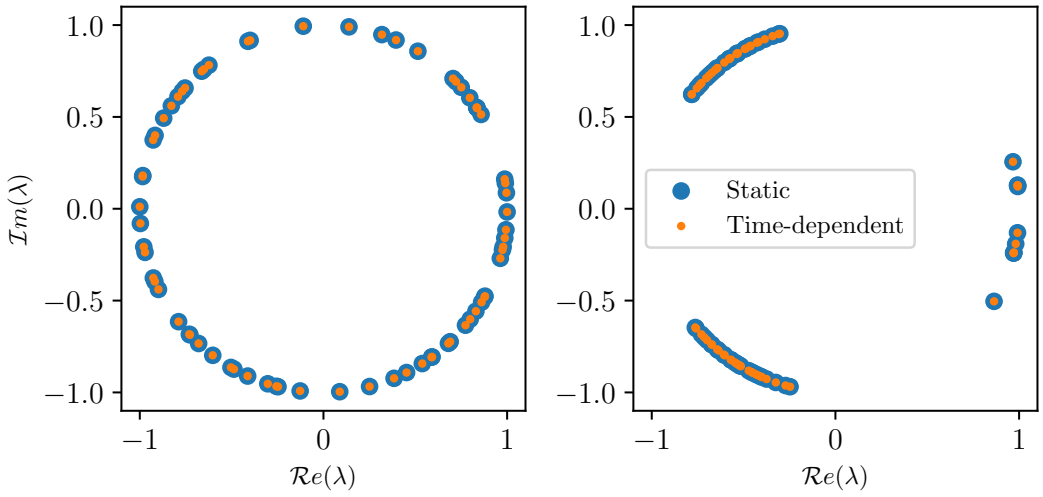

Fig. S3. The (real and imaginary parts of the) spectrum of $U(T)$ computed in the static gauge (larger blue dots) versus the spectrum of $U(T)$ computed in the time-dependent gauge (smaller orange dots), for field strengths $F = 0.5$ (left panel) and $F = 3.0$ (right panel). The difference between the spectra $\sum_i |\lambda_{i,\mathrm{static}} - \lambda_{i,\mathrm{time}}|$ is of order $\mathrm{d}t$ used to calculate the latter.

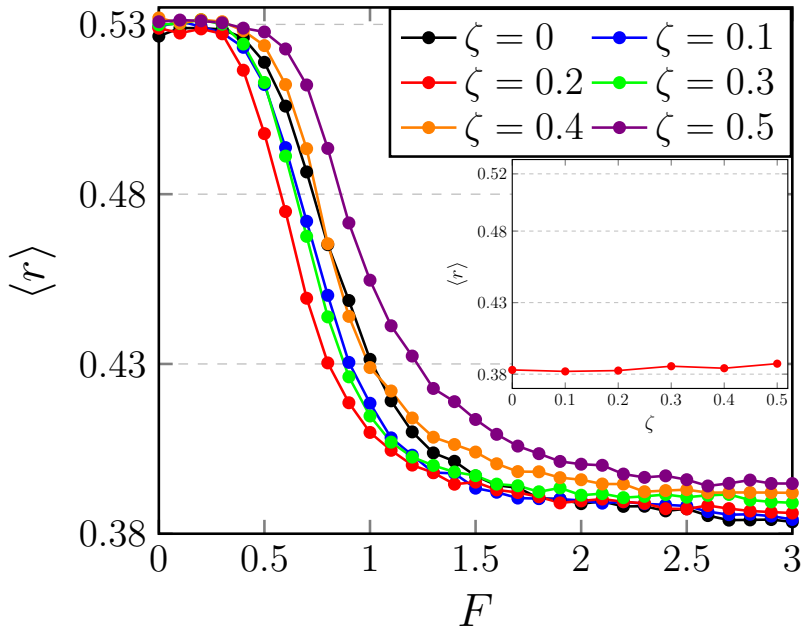

Fig. S4. The level statistics (r-index) as a function of the linear field for different values of the integrability-breaking strength, ζ. The calculation was done for a system of 14 sites (half-filled) with a fixed weak disorder $W = 0.2$ (averaged over 50 realizations),$t = 1/2$ and $U = 1$. Inset: the r-index of a clean system of 16 sites with fixed field $F = 3$ as a function of ζ.

$\mathcal{O}(\mathrm{d}t)$. An interesting phenomenon for future investigation is the observed clustering of the eigenvalues for field $F > F_c$, shown in Fig. S3.

Sensitivity to integrability-breaking terms

We now consider an extended version of Eq. 5 of the main text,

$$H = \sum_j t(c_j^\dagger c_{j+1} + h.c) - Fjn_j + h_j n_j + U n_j n_{j+1} \\ + \zeta \left(c_j^\dagger c_{j+2} + h.c + n_j n_{j+2}\right). \quad [1]$$

In the absence of both disorder and linear field, the above model is integrable for $\zeta = 0$. We show that also in the presence of the integrability-breaking terms, the application of linear field (with or without disorder) leads to a transition from a Wigner-Dyson level statistics (ergodic) to a Poisson level statistics (non-ergodic). While the value of the critical field depends on ζ and the disorder strength, the qualitative behavior is indifferent to these terms. In Fig. S4 we show the r-index as a function of the linear field strength. Different curves represent different values of ζ.

Time-evolution using the re-orthogonalized Lanczos algorithm

In this appendix we discuss algorithmic details of simulating the time-evolution of a wavefunction using a Krylov-subspace method. In particular, we have used the so-called Lanzcos algorithm with re-orthogonalization to obtain the results presented in Fig. 3 of the main text.

We wish to numerically perform the time-evolution of a wavefunction, i.e., to compute $|\psi(t_0 + t)\rangle = e^{-iHt}|\psi(t_0)\rangle$. To do this exactly would require the full diagonalization of the Hamiltonian H, which becomes impossible for large system sizes due to memory requirements. An improvement can be made by using a sparse matrix implementation of the Hamiltonian and iteratively simulating

$$|\psi(t_0 + \mathrm{d}t)\rangle = e^{-iH\mathrm{d}t}|\psi(t_0)\rangle \qquad [2]$$

for small time-steps $\mathrm{d}t$. A naive implementation of this iterative algorithm quickly accumulates numerical errors and becomes unstable, however, a more stable variant can be constructed using Krylov-subspaces (3). A Krylov-subspace of dimension m, $\mathcal{K}_m(H, |\psi\rangle)$, is defined as the span of the vectors $\left(|\psi\rangle, H|\psi\rangle, H^2|\psi\rangle, \ldots, H^{m-1}|\psi\rangle\right)$. The vector $|\psi(t_0 + \mathrm{d}t)$, after expanding the exponent on right-hand side of Eq. 2, is approximated well by a vector in this Krylov subspace.

The vectors in $\mathcal{K}_m(H, |\psi(t_0)\rangle)$ first need to be orthonormalized (discussed in more depth shortly), after which we store them as the columns of a new matrix Q_m of dimension $\mathcal{N} \times m$, where $\mathcal{N}$ is the size of the Hilbert space. After obtaining Q_m, we project the Hamiltonian into the spanned subspace to obtain $h_m = Q_m^\dagger H Q_m$. This is a much smaller $m \times m$ matrix that can be easily exponentiate, and allows us to compute

$$\begin{aligned} |\psi(t_0 + \mathrm{d}t)\rangle &= e^{-iH\mathrm{d}t}|\psi(t_0)\rangle \\ &\approx \text{the first column of } Q_m e^{-ih_m \mathrm{d}t}. \end{aligned} \qquad [3]$$

In all of the above, the Krylov subspace dimension m can either be systematically increased until convergence is obtained, or changed adaptively during the orthogonalization procedure described next.

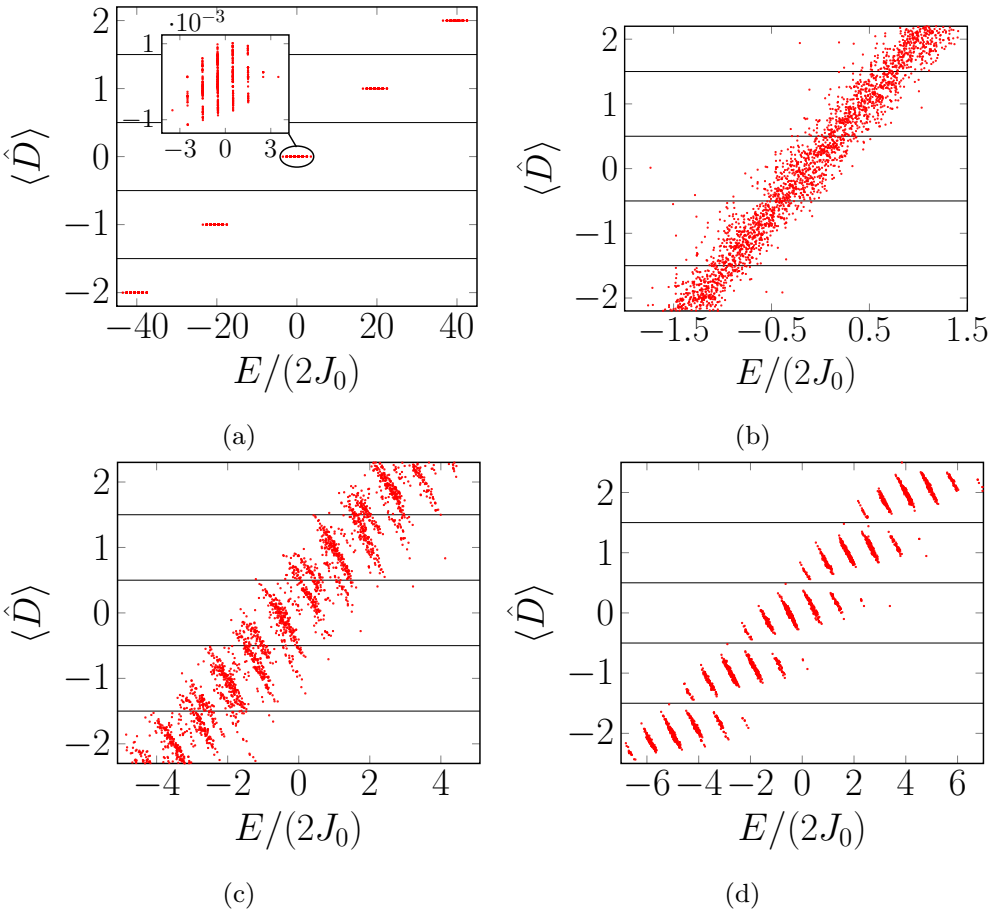

Fig. S5. Eigenstates dipole moment (expectation value) as a function of their energy for 16 sites half filled chain with $J_0 = 1/2, U = 1$ and different fields: (a) $F = 20$, (b) $F = 0.5$, (c) $F = 1.5$, (d) $F = 2.2$. For presentational reasons we show only the expectation value of the dipole moment and omit the fluctuations. The mean fluctuations are (a) $\sigma_F \approx 0.04$, (b) $\sigma_F \approx 3.2$, (c) $\sigma_F \approx 1$, (d) $\sigma_F \approx 0.6$. Above a critical field, the eigenstates in a given energy window have a well define dipole moment which restrict the dynamics.

The numerically most challenging step in this algorithm is obtaining the orthonormalized set of vectors for Q_m from $\mathcal{K}_m(H, |\psi(t)\rangle)$. A standard Gram-Schmidt procedure for orthonormalizing a set of vectors loses the orthogonality between successive vectors simply due to rounding errors (i.e. finite precision of floating point numbers). The modified Gram-Schmidt procedure does considerably better, but we have found it insufficient for our purpose. The set of vectors we wish to orthonormalize is a special set, in which each vector is generated from the previous one by application of a matrix. This means we can generate the vectors during the Gram-Schmidt procedure instead of having them given to use beforehand. This small but important difference leads to this algorithm often being called the Arnoldi method. The resulting projected matrix is in general an upper Hessenberg matrix (upper triangular plus the first lower off-diagonal). If the matrix is Hermitian as it is in our case, the projected matrix is therefore tri-diagonal. The Arnoldi algorithm with a Hermitian matrix is called the Lanczos algorithm, and provides an improvement in terms of computational effort.

Regardless of using modified Gram-Schmidt, Arnoldi or Lanczos, the orthogonality between successive vectors is gradually lost. A significant improvement, at computational cost of course, can be made by simply re-orthogonalizing the set of obtained (semi-)orthogonal vectors. It turns out that for the re-orthogonalization "twice is enough" for non-singular cases (4). For the numerics presented in Fig. 3 of the main text, we have checked the convergence of the curves with respect to the timestep dt and the Krylov-subspace dimension m. The values we have used are dt $= 0.02$ and $m = 15$.

1. Dipole moment analysis

We show in Fig. S5 the results of exact diagonalization of a half-filled fermionic system where each point represents an eigenstate in the space of energy and dipole-moment. As expected, in a given energy window and for large field (Fig. S5a) the many body wave functions have well-defined dipole moment. For a weak field however (Fig. S5b), this is not the case. The eigenstates in a given energy window span a range of dipole moments. Around the critical field (Fig. S5c,d), while the eigenstates in a given energy have a finite spread in the dipole moment, the different sectors become distinct and the integer part of the dipole moment behaves as a conserved quantity.

References

1. Kingma D, Ba J (2014) Adam: A Method for Stochastic Optimization. *ArXiv e-prints*.
2. Luitz DJ, Laflorencie N, Alet F (2015) Many-body localization edge in the random-field heisenberg chain. *Phys. Rev. B* 91(8):081103.
3. Moler C, Van Loan C (2003) Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review* 45(1):3–49.
4. Giraud L, Langou J, Rozložník M, Eshof Jvd (2005) Rounding error analysis of the classical gram-schmidt orthogonalization process. *Numerische Mathematik* 101(1):87–100.