A New Proof of Many-Body Localization in Finite Spin Chains

Absence of Normal Heat Conduction in Strongly Disordered Interacting Quantum Chains
Authors: Wojciech De Roeck, Lydia Giacomin, François Huveneers, Oskar Prośniak
arXiv:2408.04338

Recommended with a Commentary by Chris R. Laumann, Boston University and Max-Planck-Institute for Physics of Complex Systems

Anderson localization [1] is usually understood as a single particle (or linear wave) phenomenon in which static disorder localizes single particle wavefunctions. Particles placed in such an environment remain where they are placed for all time, or at least, until they interact with some kind of bath to get the energy they need to make a transition to another localized wavefunction. Although oft forgotten due to the rich fields which grew around non-interacting localization, Anderson's original paper already speculated about whether interacting particles could also remain strictly localized. If such a 'many-body localized' (MBL) phase existed, the lack of transport would undermine the usual ergodic hypothesis and statistical mechanics would not apply. These considerations were elaborated some fifty years later when [2, 3] presented extensive arguments in favor of the stability of localization in systems of strongly disordered fermions perturbed by weak short-range interactions.

The study of MBL was fundamentally shifted shortly thereafter when [4] observed that localization, a low temperature phenomena in the context of materials, could be studied even at infinite temperature in strongly disordered, strongly interacting spin chains. Inspired by ideas from random matrix theory, they took as a numerically observable proxy for the localization transition the transition in the level repulsion between the many-body eigenstates. Practically, this shift in perspective enabled modern computers to be brought to bear on MBL and led to an explosion of increasingly detailed numerical work. Conceptually, it led to a new understanding of MBL from the perspective of entanglement, many-body eigenstates and the breakdown of eigenstate thermalization (for reviews, see [5]). Many-body localization could now be studied without reference to particles (or waves, for that matter).

How should we understand the many-body localized phase then? One modern mathematical perspective is that a local Hamiltonian $H = \sum_x H_x$ acting on a spin chain of length L is localized if it can be diagonalized by a quasi-local unitary U with a decay length ξ independent of L. This is a very strong form of localization as it implies that H can be written in terms of a collection of exactly conserved commuting l-bits or quasi-local integrals

of motion and that the many-body eigenstates can be labeled by the values of those exactly conserved quantities. Indeed, the l-bit model plays a starring role in the phenomenology of MBL and associated non-equilibrium phases [6, 7].

Returning to numerical studies, the good news for MBL is that numerical studies of disordered spin chains always show a crossover to localized behavior as the disorder strength W is increased. The bad news is that the crossover has many inconsistencies with theoretical ideas on the thermodynamic ($L \to \infty$) transition [8]. Not least of these is that it drifts to larger disorder W as the system size grows [9]. Indeed, even with sophisticated extrapolation schemes, numerics alone can never determine the fate of localization in the thermodynamic limit.

Thus, it was essential when, in 2016, John Imbrie presented a proof of the existence of the MBL phase [10] in disordered 1D spin chains with weak (but finite) off-diagonal perturbations. More precisely, Imbrie argued that the MBL phase exists subject to a technical assumption on spectral statistics (called 'Limited Level Attraction'), which few doubt but even fewer can prove. Even accepting this limitation, Imbrie's proof is complicated and many technical parts are challenging to understand (for non-rigorously trained plebian physicists) and verify (even for rigorously trained mathematicians!) This perhaps underlies the recent reemergence of the controversy regarding whether or not MBL exists; see the commentary in a previous journal club and references therein [11].

The paper highlighted here [12] presents a newly expanded and detailed proof of the first essential building block in the Imbrie paper. To wit: they prove that short-range interacting spin chains of length L are many-body localized with a probability which is at least e^{-cL} so long as the off-diagonal perturbations are sufficiently weak.

At first glance, this is much weaker than the desired result, in which the probability of being MBL should approach 1 rather than 0 as L increases. However, as the title suggests, it is already sufficient to rigorously show the existence of a subdiffusive phase for heat transport in interacting 1D spin chains. This follows because even exponentially rare MBL subregions can be sufficient bottlenecks for transport that the average thermal resistance scales superlinearly with system size. Another feature is that the result holds without the caveat of the Limited Level Attraction assumption.

Perhaps more importantly, the result paves the way to cleaning up and verifying the second major step in Imbrie's approach to proving the existence of the MBL phase. To understand this, we need to understand a bit more about the strategy behind both steps in the proof.

In this first step, the lower bound on the probability that a finite segment of length L is MBL is exponentially small because the authors actually lower bound the probability of a much stronger condition: that H has absolutely *no resonances*. More precisely, to show that H can be diagonalized by a quasi-local unitary, the authors analyze an interated Schrieffer-Wolf-like diagonalization procedure which, at each iteration k, attempts to rotate away off-diagonal terms on an exponentially growing length scale L_k. This procedure is guaranteed to succeed if every rotation is sufficiently small – or, equivalently, if every perturbative energy denominator which shows up between any two states at any scale is big enough (*non-resonant*). The heart of the proof is thus to bound from below the probability that all of these infinitely many denominators are indeed non-resonant. This requires some serious combinatorial pyrotechnics which the highlighted paper [12] unpacks and elaborates signifi-

cantly compared to [10]. At the end of the day, the result can be thought of as guaranteeing that the iterated diagonalization is in a purely perturbative regime.

On the other hand, since local disorder fluctuations can produce local resonances with a finite probability, the probability of non-resonance must also be upper bounded by an exponential decaying with system size. Even in a putative MBL phase, iterative diagonalizaton should, in fact, encounter local resonances. The second step of the MBL proof is to demonstrate that these resonances need not delocalize the system. Heuristically, what the exponential *lower* bound provides is that local resonances (say at short scales) will typically be separated by relatively long non-resonant regions (if we choose a sufficiently strong disorder to make c very small). These localized 'collars' bound the influence of the short length scale resonances and effectively merge into larger MBL regions which hide the resonance. At the next larger length scale, any new resonances which emerge can again be collared by the merged MBL regions from the previous scale, and so on. Rigorously analyzing this scale dependent procedure for collaring resonances is the key to obtaining a full proof of MBL. It is also where Imbrie introduces the Limited Level Attraction hypothesis in order control the worst-case behavior of resonant regions.

The highlighted paper does not attempt to expand and verify this second stage of analysis but it appears increasingly likely that this will hold up to scrutiny as well. From a conceptual point of view, this would be a particularly satisfying response to Anderson's long-ago speculations about the strict breakdown of ergodicity enabled by quantum mechanics and disorder. It would also put on a more rigorous footing many of the ideas about non-equilibrium orders and dynamics that build on MBL.

References

[1] P. W. Anderson, Physical Review **109**, 1492 (1958).

[2] D. M. Basko, I. L. Aleiner, and B. L. Altshuler, Annals of Physics **321**, 1126 (2006).

[3] I. Gornyi, A. Mirlin, and D. Polyakov, Physical Review Letters **95**, 206603 (2005).

[4] V. Oganesyan and D. A. Huse, Physical Review B **75** (2007).

[5] R. Nandkishore and D. A. Huse, Annual Review of Condensed Matter Physics **6**, 15 (2015); D. A. Abanin, E. Altman, I. Bloch, and M. Serbyn, Reviews of Modern Physics **91**, 021001 (2019).

[6] D. A. Huse, R. Nandkishore, and V. Oganesyan, Physical Review B **90**, 174202 (2014).

[7] M. Serbyn, Z. Papić, and D. A. Abanin, Physical Review Letters **111**, 127201 (2013).

[8] A. Chandran, C. R. Laumann, and V. Oganesyan, Finite size scaling bounds on many-body localized phase transitions (2015), arXiv:1509.04285 .

[9] P. Sierant, M. Lewenstein, A. Scardicchio, L. Vidmar, and J. Zakrzewski, Many-Body Localization in the Age of Classical Computing (2024), arXiv:2403.07111 .

[10] J. Z. Imbrie, Journal of Statistical Physics **163**, 998 (2016).

[11] D. Sels and A. Polkovnikov, Journal Club for Condensed Matter Physics (2023).

[12] W. De Roeck, L. Giacomin, F. Huveneers, and O. Prosniak, Absence of Normal Heat Conduction in Strongly Disordered Interacting Quantum Chains (2024), arXiv:2408.04338 .

A Sketch of Some Diagrams In the addenda to the commentary, we will sketch a few ingredients of the iterated Schrieffer-Wolf scheme and the diagrams which the manuscript controls. We view this as a pedagogical service for physicists who may be less familiar with the super-convergent style of iterated perturbation theory popular in this part of mathematical physics. Also, the visual sketches may help indicate the geometry (in real space and hierarchical operator space) of the diagrams which the manuscript analyzes axiomatically.

Let us build this up step by step beginning with the concrete Hamiltonian,

$$H = \sum_x Z_x Z_{x+1} + h_x Z_x + \gamma X_x \tag{1}$$

which represents an Ising model in a random longitudinal field and small, uniform transverse field.

Step 1. We call $H^{(0)} \equiv H$ the Hamiltonian at scale $k = 0$. Separate it into diagonal $E^{(0)}$ and off-diagonal $\gamma_0 V^{(0)}$ parts:

$$H^{(0)} = E^{(0)} + \gamma_0 V^{(0)} \tag{2}$$

$$= \sum_x \underbrace{Z_x Z_{x+1} + h_x Z_x}_{E_x^{(0)}} + \underbrace{\gamma X_x}_{\gamma_0 V_x^{(0)}} \tag{3}$$

We introduce the coupling γ_0 to formally keep track of the order in the SW expansion at this scale, but here it is simply $\gamma_0 = \gamma$, the initial transverse field strength.

Step 2. Construct a generator $A^{(1)}$ designed to cancel $V^{(0)}$ to leading order

$$e^{\gamma_0 \hat{A}^{(1)}}(H^{(0)}) = E^{(0)} + \gamma_0 \underbrace{(V^{(0)} + \hat{A}^{(1)}(E^{(0)}))}_{\equiv 0} + \cdots \tag{4}$$

The hat on $A^{(1)}$ indicates the adjoint/commutator superoperator, $\hat{A}(X) \equiv [A, X]$. This demand can be solved locally for each term $V_x^{(0)}$:

$$A_x^{(1)} = V_x^{(0)} \frac{1}{\Delta_x E^{(0)}} \tag{5}$$

where, in a slight abuse of notation, we write the energy denominator $\Delta_x E^{(0)}$ as a diagonal operator given by the energy difference between the states connected by the off-diagonal numerator $V_x^{(0)}$. In the Ising model example,

$$A_x^{(1)} = X_x \frac{1}{-2Z_x(h_x + Z_{x-1} + Z_{x+1})} \tag{6}$$

Let us begin to introduce a diagrammatic notation by representing the terms $V_x^{(0)}$ and $A_x^{(1)}$:

$V_x^{(0)} =$  (7)

$A_x^{(1)} =$  (8)

Each diagram (here just a single node) represents a term arising at scale $k = 0$ with the label x. The shading indicates that the associated term carries an energy denominator.

Step 3. Construct the rotated Hamiltonian,

$$H^{(1)} = e^{\gamma_0 \hat{A}^{(1)}}(E^{(0)} + \gamma_0 V^{(0)}) \tag{9}$$

$$= E^{(0)} + \sum_{n=2}^{\infty} \gamma_0^n \frac{n-1}{n!} (\hat{A}^{(1)})^{n-1}(V^{(0)}) \tag{10}$$

where we have formally expanded and exploited the cancellation $\hat{A}^{(1)}(E^{(0)}) = -V^{(0)}$.

We further expand all of the generated terms locally,

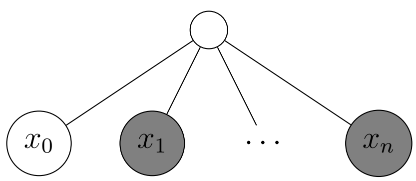
$$H^{(1)} = E^{(0)} + \sum_{n=2}^{\infty} \sum_{x_0, x_1, \cdots, x_n} \underbrace{\gamma_0^n \frac{n-1}{n!} \hat{A}^{(1)}_{x_n} \cdots \hat{A}^{(1)}_{x_1}(V^{(0)}_{x_0})}_{g_1 \text{ at scale } 1} \tag{11}$$

Each of the terms can be labeled by a diagram at scale $k = 1$,

$$H^{(1)} = E^{(0)} + \sum_{g_1} \textcircled{g_1} \qquad\qquad g_1 = (x_0, x_1, \cdots, x_n) \tag{12}$$

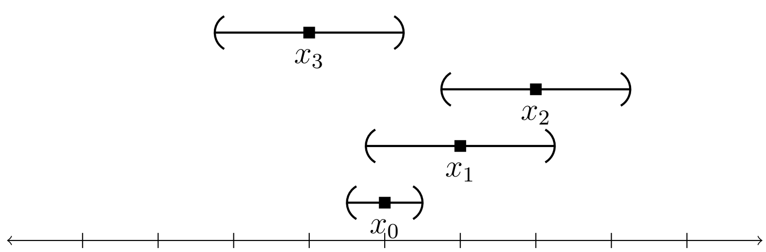
Ay given diagram g_1 at scale $k = 1$ can be expanded in terms of subdiagrams at scale $k = 0$,

$$\textcircled{g_1} =$$



$$= \gamma_0^n \frac{n-1}{n!} \hat{A}^{(1)}_{x_n} \cdots \hat{A}^{(1)}_{x_1}(V^{(0)}_{x_0}) \tag{13}$$

The tree diagram for g_1 encodes the hierarchical organization of the terms between scales. The diagrams also have spatial structure on the chain, which we can roughly summarize by the example below for a third order term,



Here, the diagram has a collection of 4 active spins (indicated by filled dots) which it flips. The operator support of the diagram is given by the union of the intervals indicated by each of the line segments associated to the subdiagrams. This is somewhat larger than the bare active spins because of the energy denominators. Notice that the diagrams must remain reasonably local, as the support of each subdiagram x_i must intersect the support of the previous subdiagrams.

Step 4. Now, we repeat the SW transformation by separating $H^{(1)}$,

$$H^{(1)} = \underbrace{E^{(0)} + \sum_{\substack{g_1 \text{ diagonal} \\ g_1 \text{ not too small}}} \circled{g_1}}_{E^{(1)}} + \underbrace{\sum_{\substack{g_1 \text{ off-diagonal} \\ g_1 \text{ not too small}}} \circled{g_1}}_{\gamma_1 V^{(1)}} + \underbrace{\sum_{g_1 \text{ rest of 'em}} \circled{g_1}}_{\gamma_1^2 R^{(1)}} \tag{14}$$

Naively, iterated SW simply separates $H^{(1)}$ into diagonal and off-diagonal, but we have some choice to leave out terms that are sufficiently small (formally, of order γ_1^2 or higher) in a residual piece of the Hamiltonian. Even if we tried to rotate them away at this stage, terms of a similar size would be regenerated by the first order rotation. So, if it helps the analysis, they might as well be dealt with later. Let us defer for a moment discussion of this choice.

Step 5. Construct the generator at scale $k = 2$ designed to cancel $\gamma_1 V^{(1)}$ to leading order,

$$A^{(2)}_{g_1} = V^{(1)}_{g_1} \frac{1}{\Delta_{g_1} E^{(1)}} = V^{(1)}_{g_1} \frac{1}{\Delta_{g_1} E^{(0)} + \underbrace{\Delta_{g_1} \sum_{g' \text{ diag}} \circled{g'}}_{O(\gamma_1)}} \tag{15}$$

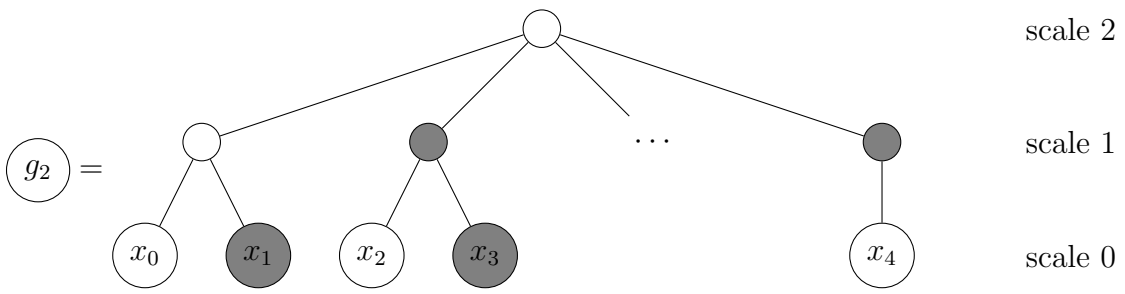
The energy shift (underbraced) in the denominator may be thought of as a real self-energy. These are often dropped in approximate treatments as not being important to the statistics of resonance, but the rigorous treatment requires controlling their influence. This is handled in the manuscript by formally expanding the denominator and generating yet another diagrammatic structure (called 'triads') to keep track of the diagonal terms this generates. For this commentary, let us follow physics precedent and drop these corrections entirely.

Step 6. Rotate to find the scale $k = 2$ Hamiltonian,

$$H^{(2)} = e^{\gamma_1 \hat{A}^{(2)}}(H^{(1)}) \tag{16}$$

$$= E^{(0)} + \sum_{n=2}^{\infty} \gamma_1^n \frac{n-1}{n!} (\hat{A}^{(2)})^{n-1}(V^{(1)}) + \sum_{n=0}^{\infty} \frac{\gamma_1^{n+2}}{n!} (\hat{A}^{(2)})^n (R^{(1)}) \tag{17}$$

As at the first scale, each of these terms can be expanded in terms of scale 2 diagrams built out of local subdiagrams arising at the previous scale:



(18)

Again, this graphical representation emphasizes the hierarchical structure of the diagram corresponding to g_2 at scale 2 in terms of the lower scale subdiagrams. It leaves implicit the spatial structure of the diagram, such as its interval of support and active spins on the spin chain.

Step 8 and higher. Rinse, lather and repeat to construct all higher scales k.

Some Comments. We have shaded diagrams which come with energy denominators (ie. $A = V/\Delta E$ terms) as these are where all of the danger lies. The goal of the proof is to lower bound the probability that every shaded node, at every scale k, is small – ie. that every A_g is *non-resonant* for γ small enough (but independent of system size L). In this event, it is relatively straightforward to show that the diagonalizing unitary,

$$U = \lim_{k\to\infty} e^{\gamma_k A^{(k+1)}} \cdots e^{\gamma_1 A^{(2)}} e^{\gamma_0 A^{(1)}} \tag{19}$$

converges and is quasi-local.

The denominators which arise depend on the choice of what gets kept in $V^{(k)}$ and what goes into the residual $R^{(k)}$ at each scale k; this is crucially important for proving the result. Essentially, one should not try to rotate away a small term with a very large support ℓ at too early a scale, because it is hard to guarantee it does not access a very small effective level spacing $\sim 2^{-\ell}$.

The actual choice of which terms are shunted into the residual terms R is inductive and somewhat complicated. Technically, this is governed by a constraint on the renormalized order of the diagrams g (called $|g|$). However, the essential idea is to limit the growth of the support of the diagrams in $V^{(k)}$ to a scale dependent length L_k.

We can estimate how L_k needs to scale as follows. The denominators at scale $k+1$, $A_g^{(k+1)} = V_g^{(k)}/\Delta_g E^{(k)}$ access up to L_k spin flips, producing a typical smallest energy $\epsilon_k \sim 2^{-L_k}$. Using this as a worst case estimate for energy denominators, and recalling that the iterated SW scheme is formally quadratic, the coupling at scale $k+1$ scales as

$$\gamma_{k+1} \sim \gamma_k \frac{\gamma_k}{\epsilon_k} = \gamma_k^2 2^{L_k} \tag{20}$$

Solving this recursion, we find

$$\ln \gamma_{k+1} = 2^{k+1} \ln \gamma_0 + (2^k L_0 + 2^{k-1} L_1 + \cdots + L_k) \ln 2 \tag{21}$$

For any initial coupling γ_0, no matter how small, the coupling γ_{k+1} diverges with k unless the sum in brackets converges to something much less than 2^{k+1}. The proof proceeds by taking $L_k = (1+\beta)^k$ with β close to 1. So long as β is close enough to 1, most of the diagrams which are generated at scale k can be dealt with by the estimate we made just now and only a small fraction (though still exponentially large number!) need to be dealt with by more sophisticated techniques.

It turns out that those final dangerous diagrams cannot be estimately inductively as in the previous paragraph. Rather, one must return to the full expansion of g at scale k in terms of a product of contributions at the earlier scales to see that the probability of resonance is

actually small, which involves joint probabilities of denominators across scales. Thus, much of the technical meat of the proof consists of proving that the probability of non-resonance is bounded below both for the bulk of diagrams which can be handled inductively and those which cannot.