Matrix product state approximations to quantum states of low energy variance

Kshiti Sneh Rai[1,2], J. Ignacio Cirac[2], and Álvaro M. Alhambra[3,2]

[1]Instituut-Lorentz, Niels Bohrweg 2, Leiden, NL-2333 CA, The Netherlands

[2]Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Straße 1, D-85748 Garching, Germany

[3] Instituto de Física Teórica UAM/CSIC, C/ Nicolás Cabrera 13-15, Cantoblanco, 28049 Madrid, Spain

We show how to efficiently simulate pure quantum states in one dimensional systems that have both finite energy density and vanishingly small energy fluctuations. We do so by studying the performance of a tensor network algorithm that produces matrix product states whose energy variance decreases as the bond dimension increases. Our results imply that variances as small as $\propto 1/\log N$ can be achieved with polynomial bond dimension. With this, we prove that there exist states with a very narrow support in the bulk of the spectrum that still have moderate entanglement entropy, in contrast with typical eigenstates that display a volume law. Our main technical tool is the Berry-Esseen theorem for spin systems, a strengthening of the central limit theorem for the energy distribution of product states. We also give a simpler proof of that theorem, together with slight improvements in the error scaling, which should be of independent interest.

1 Introduction

It is widely established that entanglement is one of the most important concepts in the study of quantum many-body systems. The main reason why is that the character of entanglement in a system can typically be connected to fundamental physical properties. For instance, an area law in the low-energy states is associated with absence of criticality, localized correlations, and tensor network approximations [1, 2]. On

Kshiti Sneh Rai: ksrai@lorentz.leidenuniv.nl

Álvaro M. Alhambra: alvaro.alhambra@csic.es

the other hand, a larger amount of entanglement in the ground state can be associated to the appearance of quantum phase transitions [3].

The entanglement properties in the bulk of the spectrum, beyond the low energy sector, are also of crucial importance. For instance, the energy eigenstates of finite energy density (with zero energy variance) most often have large amounts of entanglement, compatible with their local marginals resembling Gibbs states as per the Eigenstate Thermalization Hypothesis (ETH) [4, 5]. In contrast, product states are also in the bulk of the spectrum, but their lack of entanglement comes hand in hand with a larger energy variance.

Eigenstates and product states are the two extreme situations of either no energy variance or no entanglement. However, is this a fundamental trade-off? Or, alternatively, are there states that have both low entanglement and small energy variance? Here, we study this intermediate regime by answering the following question: what is the entanglement generated when narrowing down the variance of an initial quantum state? We do this by rigorously analyzing the performance of a matrix product state algorithm inspired by that in [6] that decreases the energy variance of any initial product state, while only increasing the bond dimension in a controlled manner. In [6], a heuristic analysis of the algorithm was given, with a similar expression for the performance which suggested that small variances were possible in arbitrary Hamiltonians. They also numerically showed that for specific ones obeying the ETH, arbitrarily small variances for the final state can be achieved. This analysis, however, did not consider the most general cases. In fact, we show that some simple counterexamples are not able to reach arbitrar-

ily small variances, and that these are related to cases in which the distance of the wavefunction from a Gaussian is maximal. Our rigorous results explicitly show how in general the larger the deviation, the more limited one is in achieving smaller variances.

As a main result, we show that vanishingly small energy variances δ^2, down to $\delta^2 \propto 1/\log(N)$, can be achieved with polynomial bond dimension, or $\log N$ entanglement entropy, in many models of interest. We demonstrate the accuracy of our results by discussing particular cases in which the bounds are tight (Sec. 3), corresponding to examples where the support of the initial product state limits how much the variance can be decreased. These models also include the counterexamples to the heuristic results in [6].

Our method for studying the energy distribution at the output of the algorithm is based on the Berry-Esseen theorem [7, 8] for spin systems. This result quantifies how much the energy distribution of product states resembles a Gaussian. We give a short proof of it based on the cluster expansion result from [9], improving previous ones by a polylogarithmic factor [10]. This renders the bound optimal, and should be of independent interest. In our proofs, we use that approximate Gaussianity to estimate the energy average and variance of the state after a *filtering operator* has been applied to it. This filter is designed to narrow down the energy fluctuations, while being expressible as a tensor network with a bond dimension that can be easily upper bounded.

An additional motivation for this scheme is that it is expected that, under the ETH, any state with a low enough energy variance will also locally resemble a Gibbs state [11, 6]. This idea is the basis of heuristic quantum algorithms for measuring local thermal expectation values [12, 13], in a way that is potentially amenable to near term quantum simulators [14]. With our bounds on the bond dimension of the MPS scheme, we rigorously analyze the classical simulability of those algorithms. In that sense, we narrow down the situations in which such a scheme will be computing quantities beyond the reach of classical computers. In doing so, we also offer efficiency guarantees for related tensor network algorithms [15, 12, 16].

The article is arranged as follows. We introduce the setting in Sec. 2, and analyze the energy distribution of product states in Sec. 3. The energy filter is introduced in Sec. 4, and in Sec. 5 we upper bound its bond dimension. We specialize to ETH Hamiltonians in Sec. 6 and conclude. The more technical proofs are placed in the appendices.

2 Preliminaries

We focus on systems of N particles described by a local Hamiltonian

$$H = \sum_{i=1}^{N} h_i = \sum_j E_j \left| E_j \right\rangle\!\left\langle E_j \right|, \tag{1}$$

where each of the Hamiltonian terms is such that $\|h_i\| \leq 1$, and overlaps with a small $\mathcal{O}(1)$ number of other qubits. If the Hamiltonian is in 1D (which we assume in Sec. 5), this refers to adjacent sites on the chain. The spectrum $\{E_j\}$ is thus confined to the range $[-N, N]$. The initial states that we consider are product among all particles

$$\left| p \right\rangle = \bigotimes_i \left| p_i \right\rangle = \sum_{E_j} b_j \left| E_j \right\rangle, \tag{2}$$

so that b_j are the coefficients of the wavefunction in the energy eigenbasis. Their average and variance are

$$E = \left\langle p \right| H \left| p \right\rangle = \sum_{E_j} |b_j|^2 E_j, \tag{3}$$

$$\sigma^2 = \left\langle p \right| (H - E)^2 \left| p \right\rangle = \sum_{E_j} |b_j|^2 (E_j - E)^2. \tag{4}$$

A mild but important assumption throughout this work is the following.

Assumption 1. *The ratio $\frac{\sigma}{\sqrt{N}} \equiv s$ is $\Omega(1)$ for all N.*

The symbol $\Omega(1)$ means that the quantity is at least as large as a constant factor, independent of N. Also, it can be easily shown that, for product states, $\sigma = \mathcal{O}\left(\sqrt{N}\right)$. Thus, we are only assuming that the variance scales as fast as possibly allowed, which is a generic property of product states. With this, we rule out trivial situations in which e.g. $\left| p \right\rangle$ is already an eigenstate of the Hamiltonian or fine-tuned situations in which the energy variance of a large region in the system is made artificially small. Our goal will be to reduce the energy variance of these product states

by applying a filtering operator, while controlling the amount of entanglement generated in the filtering process.

3 Gaussian energy distribution

Due to the lack of correlations among all the sites, the energy distribution of a product state shares features with the distribution of N independent random variables. Along these lines, the central limit theorem [17] and the Chernoff-Höffding bound [18, 19] have previously been shown.

Here, we focus on a closely related but stronger result: the Berry-Esseen theorem [7, 8]. Originally, this showed that the cumulative distribution function of N random variables is close to that of a Gaussian with the same average and variance, up to an error $\mathcal{O}(N^{-1/2})$. To introduce it in our context, let us define the following.

Definition 1. *Consider the cumulative function*

$$J(x) := \sum_{E_j \leq x} |b_j|^2, \qquad (5)$$

and the corresponding Gaussian cumulative function

$$G(x) := \int_{-\infty}^{x} \frac{dt}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-E)^2}{2\sigma^2}}. \qquad (6)$$

The Berry-Esseen error ζ_N is

$$\sup_x |J(x) - G(x)| \equiv \zeta_N. \qquad (7)$$

The generalization of this definition to arbitrary states, and in particular mixed ones, is straightforward. The scaling of the error ζ_N with N quantifies how much does the energy distribution deviate from the normal distribution. Our first technical result is the following.

Lemma 1. *Let $|p\rangle$ be a product state obeying Assumption 1. Then,*

$$\zeta_N = \mathcal{O}(N^{-1/2}). \qquad (8)$$

The proof is shown in Appendix A. It follows straightforwardly from the original Esseen's inequality, together with the cluster expansion results from [9]. Notice that this Lemma holds for all Hamiltonians that are few-body local, even beyond 1D.

Lemma 1 can be seen as a qualitative strengthening of [17], where it was shown that in 1D

$\lim_{N\to\infty} \zeta_N = 0$. It also gives a poly-logarithmic improvement on the best previous bound [10, 20] which showed that $\zeta_N = \mathcal{O}\left(\frac{\log^{2D} N}{\sqrt{N}}\right)$ in a D dimensional lattice. That result, however, also holds for the more general class of states with exponential decay of correlations, for which we expect the logarithmic factor is necessary [21].

The scaling of Eq. (8) is tight up to constant factors. For a specific example matching the bound, consider N independent random coin tosses with equal probability, so that the probability of k tails is (see Fig. 1)

$$p_k = \frac{1}{2^N} \binom{N}{k}. \qquad (9)$$

Let N be odd. Then, for $\frac{1}{2}N \geq x \geq \frac{1}{2}(N-1)$,

$$|J(x) - G(x)| = \left| \frac{1}{2} - \int_{-\infty}^{x} \frac{dt}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\frac{1}{2}N)^2}{2\sigma^2}} \right|$$

$$(10)$$

$$= \left| \int_{0}^{\frac{1}{2}N-x} \frac{dt}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\frac{1}{2}N)^2}{2\sigma^2}} \right|$$

$$\geq \frac{1}{\sqrt{\pi}} \left(\frac{\sqrt{2}\left(\frac{1}{2}N-x\right)}{\sigma} \right.$$

$$\left. -\frac{\left(\frac{1}{2}N-x\right)^3}{\sqrt{2}\sigma^3} \right).$$

Choosing e.g. $x = \frac{1}{2}(N - \frac{1}{2})$ means the lower bound is $\Omega(\sigma^{-1}) = \Omega(N^{-1/2})$.

Importantly, there are product states on spin systems with exactly Eq. (9) as the energy distribution. A trivial example is the state $\bigotimes_i^N |+\rangle_i$ with a non-interacting Hamiltonian $H = \sum_i \sigma_{Z,i}$. Additionally, it also appears in certain Hamiltonians with strong interactions, such as the model considered in [22], in which the product state has support on a $\mathcal{O}(N)$ number of eigenstates called *quantum scars* [23, 24]. These models are counterexamples to the generic case assumed in [6]. In Sec. 6, we restrict to ETH Hamiltonians, and explain how in those generic cases these counterexamples are not present.
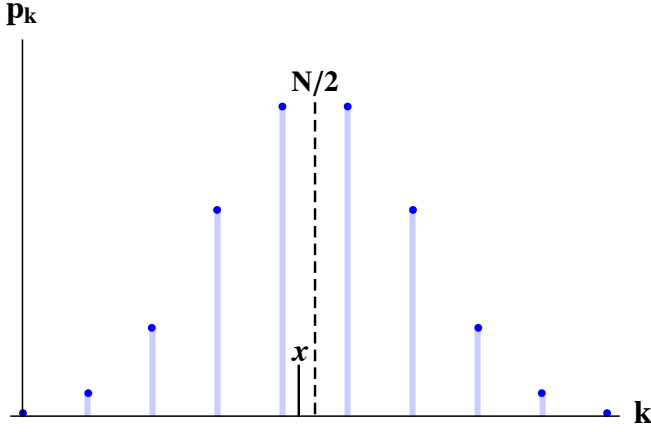
Figure 1: Representation of the binomial distribution p_k in Eq. (9) for an odd value of N. For a non-interacting Hamiltonian with this distribution, the average energy is at $N/2$, where there is no eigenstate. The smallest standard deviation one can reach by filtering this distribution is limited to the difference between eigenergies at $k = \frac{N-1}{2}$ and $k = \frac{N+1}{2}$. We also illustrate a possible choice of x in the derivation of Eq. (10).
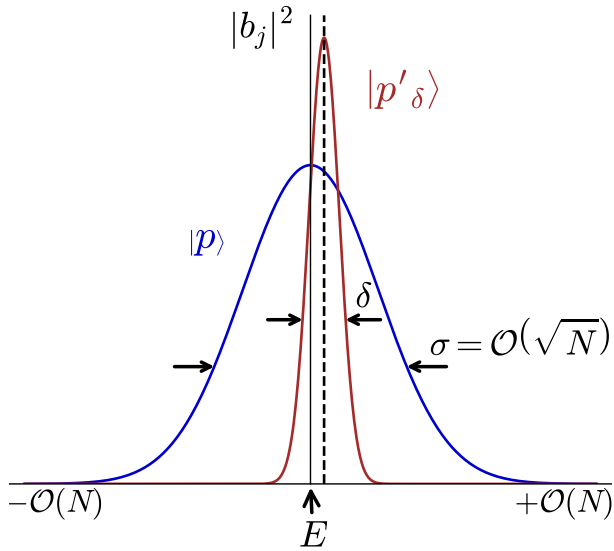

Figure 2: Illustration of the energy distribution of the initial product state, with standard deviation σ, and of the filtered distribution, with a smaller standard deviation δ and potentially a slightly shifted average (dashed line).

4   Energy filter

We now define an operator which, when applied to any quantum state (in this case, our initial product state), will decrease its energy variance. To do this, we first consider the *cosine filter* [25, 6]

$$P_M(E) = \cos^M\left(\frac{H-E}{N}\right). \qquad (11)$$

For an illustration of the effect of the filter on the energy distribution of the product state, see Fig. 2.

Since the eigenvalues of $\frac{H-E}{N}$ are within $[-1, 1]$, the cosine function acts as an approximate projector around the energies close to E when M is very large. An advantage of this operator is that using the binomial expansion we can write

$$\cos^M\left(\frac{H-E}{N}\right) = \frac{1}{2^M}\sum_{m=-\frac{M}{2}}^{\frac{M}{2}} c_m e^{i2m(H-E)/N}. \qquad (12)$$

where $c_m = \binom{M}{\frac{M}{2}-m}$. There is some freedom as to how to define the denominator inside the cosine filter [16], which does not change the final result significantly, so for simplicity we choose N as in Eq. (11).

Eq. (12) shows that when the full operator $P_M(E)$ is applied to a trial state, the result can be written as a superposition of time evolution operators. Additionally, the amount of complex exponentials can be significantly reduced for a small cost in precision.

Lemma 2. *Let the approximate cosine function $g_y(x)$ be*

$$g_y(x) := \frac{1}{2^M}\sum_{m=-y\sqrt{M}}^{y\sqrt{M}} \binom{M}{\frac{M}{2}-m} e^{i2mx}. \qquad (13)$$

This is such that, $\forall x \in [-1, 1]$,

$$|\cos^M(x) - g_y(x)| \leq e^{-\frac{y^2}{2}}. \qquad (14)$$

This is obtained from the Chernoff bound on the binomial coefficients, which is exponentially decaying in y^2. For a significant reduction in the number of terms, we need y to be $o(\sqrt{M})$.

Lemma 2 allows us to define our filtered state as

$$|p_{M,y}\rangle = \frac{g_y\left(\frac{H-E}{N}\right)|p\rangle}{\left\|g_y\left(\frac{H-E}{N}\right)|p\rangle\right\|}, \qquad (15)$$

which leads to the main technical result of this section.

Lemma 3. *Let $M = \mathcal{O}\left(\frac{N}{\zeta_N^2}\right)$ and $M = \Omega(N)$, and $y = \Omega\left(\sqrt{\log\left(M^3/N^2\right)}\right)$. The energy average μ and variance δ^2 of $|p_{M,y}\rangle$ are bounded as*

$$|\mu - E| = \mathcal{O}\left(\frac{N}{\sqrt{M}}\right), \qquad (16)$$

$$\delta^2 = \mathcal{O}\left(\frac{N^2}{M}\right). \qquad (17)$$

This lemma only applies for large enough values of M. On the other hand, for $M = \mathcal{O}(N)$, the best bounds one can achieve are

$$|\mu - E| = \mathcal{O}\left(\sqrt{N}\right), \qquad (18)$$

$$\delta^2 = \mathcal{O}\left(N\right), \qquad (19)$$

which means that the filter does not change the variance in a substantial way - at most up to a constant factor.

For large M, the lemma shows that the filter $g_y\left(\frac{H-E}{N}\right)$ decreases the energy variance of the state without changing the average energy too much. The proof is shown in App. B. It is based on the Berry-Esseen error from Definition 1, which we use to approximate the expression of the variance δ^2 as a Gaussian integral, with a considerable amount of error terms that need to be estimated. The overall argument is a way of segmenting energy sums in pieces, so that via the approximation of Lemma 1 they can be turned into pieces of a tractable Gaussian integral. A similar argument previously appeared in a different context in [26]. We note that this holds for Hamiltonians that are local, independent of the dimension. The upper bound on M in terms of ζ_N^{-1} is due to the fact that the deviation from Gaussian limits how much we can resolve the energy distribution after the filter is applied: if the deviation is large, the filter may act in a more uncontrolled way, and the bound may not hold.

In many practical scenarios we expect the wavefunction $|p\rangle$ to be symmetric about the energy average, in which case μ and E will be identical. On the other hand, the scaling of the bound on δ is tight up to constant factors: if the wavefunction coefficients b_j are taken to be an exact Gaussian, one can explicitly calculate $\delta \simeq \frac{N}{\sqrt{2M}}$ (see Eq. (15) in [6]).

The result is purposely stated in terms of the general error from Def. 1. Lemma 1, however, allows us to establish that choosing $M \propto N^2$, the bound on the variance is

$$\delta^2 = \mathcal{O}(1). \qquad (20)$$

This means that for any initial product state with standard deviation of $\mathcal{O}(\sqrt{N})$, the cosine filter can be applied to reduce the standard deviation down to $\mathcal{O}(1)$, and shifting the average by

at most that amount. Note that a similar variance bound (Eq. (17)) was estimated in [6]. In this section, we have provided a rigorous proof for it, and showed that it is only valid for particular range of values of M in the cosine filter, which depend on ζ_N, quantity appearing from the Berry-Esseen theorem. Intuitively, this is because, in order to estimate the variance δ, we need a high resolution of the energy populations $|b_j|^2$, which is only possible if ζ_N is small enough. Larger values for M are responsible for filtering a narrower range of energies by implementing an approximate eigenstate projector around energy E. As a result, the norm of the final state becomes exceedingly smaller with increasing M. However, to retain the Gaussian approximation from the Berry-Esseen theorem (Lemma (1)), we require the norm of the filtered state to be above a certain constant ($= \mathcal{O}(1)$), which forces an upper bound on M. The examples from Sec. 3 show that this is the smallest variance one can reach in general (see Fig. 1). However, if the factor ζ_N would decrease more quickly than $\mathcal{O}(N^{-1/2})$, the range of values which M can take becomes larger, and lower variances can be achieved, as we discuss in Sec. 6.

5 Low variance state using MPS

In this section we prove our main result on the approximability of the low variance filtered states with matrix product states. To do this, we study the performance of the cosine filter algorithm for preparing arbitrarily low energy variance states in the finite energy density regime of a 1D local Hamiltonian. We reiterate that producing these states is interesting for the following reasons: i) understanding the fundamental trade-offs between entanglement and energy variance in the bulk of the spectrum, ii) producing the simplest possible states that can approximate thermal properties locally, and iii) preparing initial states for filtering quantum algorithms, designed for constructing low-variance states inaccessible using tensor network techniques.

The main idea is that it is possible to construct a matrix product operator g_y^D that approximates $g_y\left(\frac{H-E}{N}\right)$ in one spatial dimension. When applied to our initial product state $|p\rangle$, this implies that one can efficiently find an MPS approximation to $|p_{M,y}\rangle$ with a close enough average and

variance in energy, while also having a small bond dimension. This is contained in our main result Theorem 1.

The starting point is the following lemma from [27], which shows there is a Matrix Product Operator (MPO) [28] approximation to the matrix exponential of H.

Lemma 4. *[27] Let H be a 1D local Hamiltonian. There is an algorithm that outputs an MPO approximation T_t of e^{-itH} such that*

$$\left\| T_t - e^{-itH} \right\| \le \epsilon, \tag{21}$$

with bond dimension upper bounded by

$$D \le e^{\mathcal{O}(t) + \mathcal{O}\left(\sqrt{t \log(N/\epsilon)}\right)}. \tag{22}$$

The idea here is that g_y can be written as a sum of complex exponentials as per Eq. (12). Thus, we can approximate each term in the filter with Lemma 4, and add them to obtain our approximate filter

$$g_y^D = \frac{1}{2^M} \sum_{m=-y\sqrt{M}}^{m=+y\sqrt{M}} \binom{M}{\frac{M}{2} - m} T_{2m/N}. \tag{23}$$

When applied to the initial product state, this yields the final MPS with bounded variance and bond dimension. The precise result is as follows.

Theorem 1. *Let H be one-dimensional, and let $|p\rangle$ satisfy Assumptions 1. The MPS $|p'_\delta\rangle = \frac{g_y^D |p\rangle}{||g_y^D |p\rangle||}$, under the conditions of Lemma 3, has the properties*

$$|\langle p'_\delta | H | p'_\delta \rangle - E| = \mathcal{O}\left(\frac{N}{\sqrt{M}}\right) \tag{24}$$

$$\langle p'_\delta | H^2 | p'_\delta \rangle - (\langle p'_\delta | H | p'_\delta \rangle)^2 \le 2\delta^2 = \mathcal{O}\left(\frac{N^2}{M}\right). \tag{25}$$

Moreover, the bond dimension of g_y^D (and thus of $|p'_\delta\rangle$) is bounded by

$$D \le \frac{N \log N}{\delta} e^{\mathcal{O}\left(\frac{\sqrt{\log(N/\delta)}}{\delta}\right) + \mathcal{O}\left(\frac{\log^{\frac{3}{4}}(N/\delta)}{\sqrt{\delta}}\right)}. \tag{26}$$

Proof. First we have that, by Lemma 4 and the triangle inequality,

$$\left\| g_y\left(\frac{H-E}{N}\right) - g_y^D \right\| \le 2^{-M} \sum_{|m| \le y\sqrt{M}} \binom{M}{\frac{M}{2} - m} \epsilon$$

$$\le \epsilon, \tag{27}$$

where we consider the approximation error ϵ coming from Eq. (21). Thus, for the states

$$|| |p_{M,y}\rangle - |p'_\delta\rangle || = \left\| \frac{g_y\left(\frac{H-E}{N}\right)|p\rangle}{\left\| g_y\left(\frac{H-E}{N}\right)|p\rangle \right\|} - \frac{g_y^D |p\rangle}{\left\| g_y^D |p\rangle \right\|} \right\| \tag{28}$$

$$\le 2 \frac{\left\| g_y\left(\frac{H-E}{N}\right) - g_y^D \right\|}{\left\| g_y\left(\frac{H-E}{N}\right)|p\rangle \right\|} \tag{29}$$

$$\le \frac{2\epsilon}{\left\| g_y\left(\frac{H-E}{N}\right)|p\rangle \right\|}. \tag{30}$$

The factor $\left\| g_y\left(\frac{H-E}{N}\right)|p\rangle \right\|$ is lower bounded in Eq. (110) as,

$$\left\| g_y\left(\frac{H-E}{N}\right)|p\rangle \right\| = \Omega\left(\frac{N^{1/2}}{M^{1/4}}\right). \tag{31}$$

Now, notice that for $|p_{M,y}\rangle$ and $|p'_\delta\rangle$ to have the same average and variance up to the errors of Eq. (24) and (25) it is enough that,

$$|| |p_{M,y}\rangle - |p'_\delta\rangle || \le \frac{\delta^2}{2N^2}. \tag{32}$$

Given Eq. (32), the error we require from Lemma 4 is $\epsilon = \mathcal{O}\left(\frac{\delta^2}{N^{3/2} M^{1/4}}\right)$. With this, notice that the leading error for the average is given by Eq. (16).

To estimate the total bond dimension of g_y^D, notice that we have $2y\sqrt{M}$ terms, and that, considering Lemma 3, the longest t to consider is $y\sqrt{M}/N = \mathcal{O}(y\delta^{-1})$. Using Lemma 4 and simplifying the powers in the logarithm, we can upper bound

$$D \le \frac{2yN}{\delta} e^{\mathcal{O}(y\delta^{-1}) + \mathcal{O}\left(\sqrt{y\delta^{-1} \log\left(\frac{N}{\delta}\right)}\right)} \tag{33}$$

Finally, choosing $y^2 = 6 \log \frac{N}{\delta}$ as allowed by Lemma 3, we obtain

$$D \le \frac{N \log N}{\delta} e^{\mathcal{O}\left(\frac{\sqrt{\log(N/\delta)}}{\delta}\right) + \mathcal{O}\left(\frac{\log^{\frac{3}{4}}(N/\delta)}{\sqrt{\delta}}\right)}. \tag{34}$$

$\square$

Additionally, the run-time of the algorithm that constructs $|p'_\delta\rangle$ is simply the bond dimension multiplied by a factor of N, to account for the cost of manipulating the tensors of the MPO [27]. A similar estimate of the bond dimension

was given in [6], but their argument omitted some key steps in the analysis, such as the effect of using the Chernoff bound, and the explicit distance of the wavefunction to a Gaussian, which appears through Lemma 3.

The most important particular case of this result is if we aim for a variance $\delta = \mathcal{O}(1)$, in which case we have the following.

Corollary 1. *There exists an MPS $|p'\rangle$ of quasi-linear bond dimension $D \leq N e^{\mathcal{O}(\log^{\frac{3}{4}}(N))}$ such that*

$$|\langle p'|H|p'\rangle - E| = \mathcal{O}(1) \qquad (35)$$
$$\langle p'|H^2|p'\rangle - (\langle p'|H|p'\rangle)^2 = \mathcal{O}(1). \qquad (36)$$

This is achieved by choosing $M \propto N^2$ in the filter, given Theorem 1 together with Lemma 3. This is the smallest variance that can be achieved in general, as illustrated by the examples with the binomial energy distribution in Eq. 9.

If instead of a product state our initial state has exponential decay of correlations, the corresponding standard deviation one can prove to reach in general is $\delta = \mathcal{O}(\log^{2D} N)$ as per the result in [20]. Additionally, if the Berry-Esseen error ζ_N is much smaller than the upper bound of Lemma 1 we can also reach smaller variances, down to $\zeta_N^2 \times N$, at an additional cost in the bond dimension. In particular, we expect that in practice it is most often possible to reach $\delta^2 \propto 1/\log N$ while still keeping the bond dimension from Eq. (26) polynomial in N.

Theorem 1 is restricted to one dimension. For higher dimensions, the existence of a similar PEPO is guaranteed by the results of [29], which in our setting yields a bond dimension

$$D = \left(\frac{N}{\delta}\right)^{O\left(\frac{\sqrt{\log N}}{\delta}\right)}. \qquad (37)$$

This, however, is only polynomial for larger variances $\delta^2 = \Omega(\log N)$ and in no case guarantees a provably efficient approximation algorithm, considering the difficulty of contracting PEPS [30].

The upper bounds on the bond dimension also guarantee bounds on the entanglement entropy of the state $|p'_\delta\rangle$. Defining $S(\rho'_A) = -\text{tr}[\rho'_A \log \rho'_A]$ with ρ'_A the marginal on region A, we have that $S(\rho'_A) \leq |\partial_A| \log D$, with $|\partial_A|$ the number of bonds between region A and its complement. In

fact, this upper bound applies to all Rényi entanglement entropies $S_\alpha(\rho'_A) = \frac{1}{1-\alpha} \log \text{tr}[\rho'^\alpha_A]$ with $\alpha > 0$.

Also notice that, by the Alicki-Fannes inequality together with Eq. (32), the upper bound on the entanglement entropy also holds for the state $|p_{M,y}\rangle$, in which we have applied the exact filter $g_y\left(\frac{H-E}{N}\right)$. Overall, we can conclude that in 1D, both $|p'_\delta\rangle$ and $|p_{M,y}\rangle$ (with its corresponding marginal ρ_A) have an entanglement entropy on a region A bounded by

$$S(\rho'_A), S(\rho_A) \leq \mathcal{O}\left(\log(N/\delta)\right) + \qquad (38)$$
$$+ \mathcal{O}\left(\frac{\sqrt{\log(N/\delta)}}{\delta}\right) + \mathcal{O}\left(\frac{\log^{\frac{3}{4}}(N/\delta)}{\sqrt{\delta}}\right).$$

For instance, for $\delta = \Omega(1)$ we obtain the bound $S \leq \mathcal{O}(\log N)$, which is significantly smaller than the largest possible volume law scaling $S(\rho_A) \propto N$.

6 ETH Hamiltonians

So far, we have expressed many of our results, and in particular the range of δ achievable in Lemma 3, without lower bounding the value of ζ_N. Given the examples of Sec. 3, the error $\mathcal{O}(N^{-1/2})$ in Lemma 1 is the strongest general upper bound on ζ_N. However, in most systems of interest much more favourable scalings should hold, as we now illustrate.

Let $S_2(E_j) = -\log \text{tr}\left\{\rho^2_{A,j}\right\}$ be the Rényi-2 entanglement entropy of eigenstate $|E_j\rangle$ for an arbitrary bipartition into subsystems A, B with $|A|, |B| \propto N$. It was shown in [31] that, given that $|p\rangle$ is a product state,

$$|b_j|^2 \leq e^{-\frac{S_2(E_j)}{2}}. \qquad (39)$$

When the entanglement in the energy eigenstates is a volume law, as is generically expected [32, 33], $S_2(E_j) \propto N$ and individual populations $|b_j|^2$ are exponentially small. This means that counterexamples such as Eq. (9), in which the individual populations are as large as $\propto N^{-1/2}$, are ruled out. In these rather ubiquitous Hamiltonians, the product states thus have support in exponentially many eigenstates, and with exponentially small populations. This means that $J(x)$ in Def. 1 may be exponentially close to the smooth function $G(x)$, and in particular, we expect that
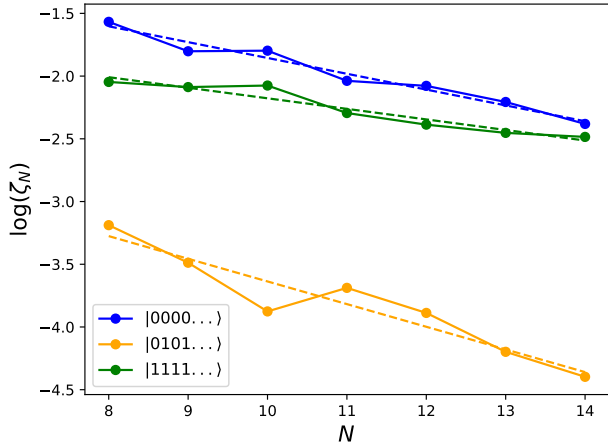
Figure 3: Illustration of the scaling of the logarithm of Berry-Esseen error ζ_N with the size of the system N for the Mixed Field Ising Hamiltonian as defined in Eq. (40) with the parameters $(J, g, h) = (1, -1.05, 0.5)$. An approximately linear decrease in $\log \zeta_N$ is observed even for finite system sizes (8 to 14 sites).

the Berry-Esseen error is exponentially small in system size $\zeta_N = e^{-\Omega(N)}$. This is consistent with the numerical results in [6]. In Fig. 3 we plot the scaling of $\log \zeta_N$ with N for the mixed field Ising model Hamiltonian defined as follows,

$$H = \sum_{i=1}^{N-1} J\sigma_i^z \sigma_{i+1}^z + g\sigma_i^x + h\sigma_i^z, \qquad (40)$$

where we chose the values of $(J, g, h) = (1, -1.05, 0.5)$ for the non-integrable regime where ETH is expected to hold. For various choices of product states, we observe approximately linearly decreasing values for $\log \zeta_N$ with increasing N even for finite system sizes, consistent with an exponential decay of ζ_N as theoretically expected from Eq. (39).

With such a favourable scaling, Lemma 3 allows one to increase the filter parameter up to $M \propto e^{\Omega(N \log N)}$ in order to achieve vanishingly small variances given Eq. (17). However in this case, the bond dimension (Eq. (26)) restricts the minimum possible variance. In particular, a variance of $\delta^2 \propto 1/\log N$ can be reached, which converges to zero in the thermodynamic limit, while still guaranteeing a polynomial bond dimension in Eq. (26).

It is also in these ETH Hamiltonians that we expect that states of low energy variance will have small subsystems resembling those of the Gibbs distribution. Considering $D \propto \text{poly}(N)$,

and a corresponding entanglement entropy $\propto \log D$, subsystems of $\mathcal{O}(\log D)$ particles will display large amounts of entanglement entropy. Thus, it is possible that the filtered state displays approximately thermal marginals of up to that size. Along these lines, and depending on the observables considered, previous numerical analyses suggest one might need up to $\delta^2 \propto 1/\log^2 N$ [6, 12] (see in particular page 6 of [6]). Additionally, the analytical results from [11] suggesting that for more general observables, variances vanishing as $\delta \propto 1/\text{poly}(N)$ might be needed.

7 Conclusion

We have shown that for arbitrary systems evolving under local Hamiltonians, it is possible to efficiently construct states with variance $\delta = \Omega(1)$ via MPS representations, while shifting the average at most by a comparable amount. This is the smallest possible variance one could achieve, due to existing counterexamples. However, we expect that significantly lower values of the variance can be reached in most cases of interest, with the complexity of the algorithm increasing accordingly. In particular, in many models, $\delta = \Omega(1/\sqrt{\log N})$ can still be achieved efficiently.

With our main results, we provide rigorous analytical bounds on the classical simulability of the algorithm proposed in [6, 12], which has recently been implemented with tensor networks [16]. We expect that this type of scheme will be important in near-term quantum simulation experiments of equilibrium and non-equilibrium properties of quantum many-body systems. Specifically, we show that to perform calculations beyond the reach of classical computers, these will have to go beyond one dimension or reach very small energy variances of at least $\delta = o(1/\sqrt{\log N})$. These low energy-variance and MPS representable states are easy to prepare with quantum circuits. Thus they can serve as starting points of a filtering quantum algorithm, that will further reduce their variance and address quantities that are difficult to get with tensor network techniques. By starting with states with relatively low variance, one can conserve quantum computational resources. We hope that this further motivates experimental efforts on quantum simulation, such as [34].

The key to our method of analyzing the filtered

state is the knowledge of the gaussianity of the wavefunction granted by the Berry-Esseen theorem. This allows us to, for instance, lower bound the norm of a product state to which an operator has been applied to (in this case, the filter g_y). It would be interesting to see if our proof techniques can be applied to other types of operators, such as the Chebyshev filter proposed in [6]. We expect that this type of argument will have further applications in the study of classical and quantum algorithms for many body systems at finite energy density.

Acknowledgements

References

[1] F. Verstraete, M. M. Wolf, D. Perez-Garcia, and J. I. Cirac. Criticality, the area law, and the computational power of projected entangled pair states. *Physical Review Letters*, 96(22), jun 2006. DOI: 10.1103/physrevlett.96.220601. URL https://doi.org/10.1103%2Fphysrevlett.96.220601.

[2] J. Eisert, M. Cramer, and M. B. Plenio. Colloquium: Area laws for the entanglement entropy. *Reviews of Modern Physics*, 82(1): 277–306, feb 2010. DOI: 10.1103/revmodphys.82.277. URL https://doi.org/10.1103%2Frevmodphys.82.277.

[3] G. Vidal, J. I. Latorre, E. Rico, and A. Kitaev. Entanglement in quantum critical phenomena. *Phys. Rev. Lett.*, 90:227902, Jun 2003. DOI: 10.1103/PhysRevLett.90.227902. URL https://link.aps.org/doi/10.1103/PhysRevLett.90.227902.

[4] Mark Srednicki. Chaos and quantum thermalization. *Physical Review E*, 50(2): 888–901, aug 1994. DOI: 10.1103/physreve.50.888. URL https://doi.org/10.1103%2Fphysreve.50.888.

[5] J. M. Deutsch. Quantum statistical mechanics in a closed system. *Phys. Rev. A*, 43: 2046–2049, Feb 1991. DOI: 10.1103/PhysRevA.43.2046. URL https://link.aps.org/doi/10.1103/PhysRevA.43.2046.

[6] Mari Carmen Bañuls, David A. Huse, and J. Ignacio Cirac. Entanglement and its relation to energy variance for local one-dimensional hamiltonians. *Physical Review B*, 101(14), apr 2020. DOI: 10.1103/physrevb.101.144305. URL https://doi.org/10.1103%2Fphysrevb.101.144305.

[7] Andrew C Berry. The accuracy of the gaussian approximation to the sum of independent variates. *Transactions of the american mathematical society*, 49(1):122–136, 1941.

[8] C.G. Esseen. *On the Liapounoff Limit of Error in the Theory of Probability*. Arkiv för matematik, astronomi och fysik. Almqvist & Wiksell, 1942. URL https://books.google.es/books?id=VjXgPgAACAAJ.

[9] Dominik S. Wild and Álvaro M. Alhambra. Classical simulation of short-time quantum dynamics. *PRX Quantum*, 4: 020340, Jun 2023. DOI: 10.1103/PRXQuantum.4.020340. URL https://link.aps.org/doi/10.1103/PRXQuantum.4.020340.

[10] Fernando G. S. L. Brandao and Marcus Cramer. Equivalence of statistical mechanical ensembles for non-critical quantum systems, 2015. URL https://arxiv.org/abs/1502.03263.

[11] Anatoly Dymarsky and Hong Liu. New char-

acteristic of quantum many-body chaotic systems. *Phys. Rev. E*, 99:010102, Jan 2019. DOI: 10.1103/PhysRevE.99.010102. URL https://link.aps.org/doi/10.1103/PhysRevE.99.010102.

[12] Sirui Lu, Mari Carmen Bañuls, and J. Ignacio Cirac. Algorithms for quantum simulation at finite energies. *PRX Quantum*, 2:020321, May 2021. DOI: 10.1103/PRXQuantum.2.020321. URL https://link.aps.org/doi/10.1103/PRXQuantum.2.020321.

[13] Alexander Schuckert, Annabelle Bohrdt, Eleanor Crane, and Michael Knap. Probing finite-temperature observables in quantum simulators of spin systems with short-time dynamics. *Physical Review B*, 107 (14), apr 2023. DOI: 10.1103/physrevb.107.l140410. URL https://doi.org/10.1103%2Fphysrevb.107.l140410.

[14] Andrew J. Daley, Immanuel Bloch, Christian Kokail, Stuart Flannigan, Natalie Pearson, Matthias Troyer, and Peter Zoller. Practical quantum advantage in quantum simulation. *Nature*, 607 (7920):667–676, Jul 2022. ISSN 1476-4687. DOI: 10.1038/s41586-022-04940-6. URL https://doi.org/10.1038/s41586-022-04940-6.

[15] Aslı Çakan, J. Ignacio Cirac, and Mari Carmen Bañuls. Approximating the long time average of the density operator: Diagonal ensemble. *Physical Review B*, 103(11), March 2021. ISSN 2469-9969. DOI: 10.1103/physrevb.103.115113. URL http://dx.doi.org/10.1103/PhysRevB.103.115113.

[16] Yilun Yang, J. Ignacio Cirac, and Mari Carmen Bañuls. Classical algorithms for many-body quantum systems at finite energies. *Phys. Rev. B*, 106:024307, Jul 2022. DOI: 10.1103/PhysRevB.106.024307. URL https://link.aps.org/doi/10.1103/PhysRevB.106.024307.

[17] Michael Hartmann, Günter mahler, and Ortwin Hess. Gaussian quantum fluctuations in interacting many particle systems. *Letters in Mathematical Physics*, 68(2):103–112, may 2004. DOI: 10.1023/b:math.0000043321.00896.86. URL https://doi.org/10.1023%2Fb%3Amath.0000043321.00896.86.

[18] Anurag Anshu. Concentration bounds for quantum states with finite correlation length on quantum spin lattice systems. *New Journal of Physics*, 18(8):083011, aug 2016. DOI: 10.1088/1367-2630/18/8/083011. URL https://doi.org/10.1088%2F1367-2630%2F18%2F8%2F083011.

[19] Tomotaka Kuwahara. Connecting the probability distributions of different operators and generalization of the chernoff–hoeffding inequality. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(11):113103, nov 2016. DOI: 10.1088/1742-5468/2016/11/113103. URL https://doi.org/10.1088%2F1742-5468%2F2016%2F11%2F113103.

[20] Fernando G. S. L. Brandão, Marcus Cramer, and Mădălin Guţă. Berry-Esseen theorem for quantum lattice systems and the equivalence of statistical mechanical ensembles. http://quantum-lab.org/qip2015/talks/125-Brandao.pdf, 2015. URL http://quantum-lab.org/qip2015/talks/125-Brandao.pdf.

[21] A. N. Tikhomirov. On the convergence rate in the central limit theorem for weakly dependent random variables. *Theory of Probability & Its Applications*, 25(4):790–809, 1981. DOI: 10.1137/1125092. URL https://doi.org/10.1137/1125092.

[22] Michael Schecter and Thomas Iadecola. Weak ergodicity breaking and quantum many-body scars in spin-1 XY magnets. *Physical Review Letters*, 123(14), October 2019. ISSN 1079-7114. DOI: 10.1103/physrevlett.123.147201. URL http://dx.doi.org/10.1103/PhysRevLett.123.147201.

[23] C. J. Turner, A. A. Michailidis, D. A. Abanin, M. Serbyn, and Z. Papić. Weak ergodicity breaking from quantum many-body scars. *Nature Physics*, 14(7):745–749, may 2018. DOI: 10.1038/s41567-018-0137-5. URL https://doi.org/10.1038%2Fs41567-018-0137-5.

[24] Álvaro M. Alhambra, Anurag Anshu, and Henrik Wilming. Revivals imply quantum many-body scars. *Phys. Rev. B*, 101:205107, May 2020. DOI: 10.1103/PhysRevB.101.205107. URL

https://link.aps.org/doi/10.1103/PhysRevB.101.205107.

[25] Yimin Ge, Jordi Tura, and J. Ignacio Cirac. Faster ground state preparation and high-precision ground energy estimation with fewer qubits. *Journal of Mathematical Physics*, 60(2):022202, 02 2019. ISSN 0022-2488. DOI: 10.1063/1.5027484. URL https://doi.org/10.1063/1.5027484.

[26] Karen V. Hovhannisyan, Mathias R. Jørgensen, Gabriel T. Landi, Álvaro M. Alhambra, Jonatan B. Brask, and Martí Perarnau-Llobet. Optimal quantum thermometry with coarse-grained measurements. *PRX Quantum*, 2:020322, May 2021. DOI: 10.1103/PRXQuantum.2.020322. URL https://link.aps.org/doi/10.1103/PRXQuantum.2.020322.

[27] Tomotaka Kuwahara, Álvaro M. Alhambra, and Anurag Anshu. Improved thermal area law and quasilinear time algorithm for quantum gibbs states. *Physical Review X*, 11(1), mar 2021. DOI: 10.1103/physrevx.11.011047. URL https://doi.org/10.1103%2Fphysrevx.11.011047.

[28] B Pirvu, V Murg, J I Cirac, and F Verstraete. Matrix product operator representations. *New Journal of Physics*, 12(2):025012, feb 2010. DOI: 10.1088/1367-2630/12/2/025012. URL https://doi.org/10.1088%2F1367-2630%2F12%2F2%2F025012.

[29] Andras Molnar, Norbert Schuch, Frank Verstraete, and J. Ignacio Cirac. Approximating gibbs states of local hamiltonians efficiently with projected entangled pair states. *Physical Review B*, 91(4), jan 2015. DOI: 10.1103/physrevb.91.045138. URL https://doi.org/10.1103%2Fphysrevb.91.045138.

[30] Norbert Schuch, Michael M. Wolf, Frank Verstraete, and J. Ignacio Cirac. Computational complexity of projected entangled pair states. *Physical Review Letters*, 98(14), apr 2007. DOI: 10.1103/physrevlett.98.140506. URL https://doi.org/10.1103%2Fphysrevlett.98.140506.

[31] H. Wilming, M. Goihl, I. Roth, and J. Eisert. Entanglement-ergodic quantum systems equilibrate exponentially well. *Physical Review Letters*, 123(20), nov 2019. DOI: 10.1103/physrevlett.123.200604. URL https://doi.org/10.1103%2Fphysrevlett.123.200604.

[32] Don N. Page. Average entropy of a subsystem. *Phys. Rev. Lett.*, 71:1291–1294, Aug 1993. DOI: 10.1103/PhysRevLett.71.1291. URL https://link.aps.org/doi/10.1103/PhysRevLett.71.1291.

[33] Lev Vidmar and Marcos Rigol. Entanglement entropy of eigenstates of quantum chaotic hamiltonians. *Physical Review Letters*, 119(22), nov 2017. DOI: 10.1103/physrevlett.119.220603. URL https://doi.org/10.1103%2Fphysrevlett.119.220603.

[34] Kévin Hémery, Khaldoon Ghanem, Eleanor Crane, Sara L. Campbell, Joan M. Dreiling, Caroline Figgatt, Cameron Foltz, John P. Gaebler, Jacob Johansen, Michael Mills, Steven A. Moses, Juan M. Pino, Anthony Ransford, Mary Rowe, Peter Siegfried, Russell P. Stutz, Henrik Dreyer, Alexander Schuckert, and Ramil Nigmatullin. Measuring the loschmidt amplitude for finite-energy properties of the fermi-hubbard model on an ion-trap quantum computer, 2023. URL https://doi.org/10.48550/arXiv.2309.10552.

[35] William Feller. *An introduction to probability theory and its applications, Volume 2*, volume 81. John Wiley & Sons, 1991.

A   Proof of Lemma 1

We study the characteristic function of the system's energy (assuming $\langle H \rangle = 0$)

$$\phi(t') \equiv \left\langle e^{\frac{it'H}{\sigma}} \right\rangle = \left\langle e^{iHt} \right\rangle. \tag{41}$$

The bound on the error follows from Esseen's inequality [35],

$$\pi \zeta_N \leq \frac{C}{T} + \int_0^T \frac{|\phi(t') - e^{-\frac{(t')^2}{2}}|}{t'} \mathrm{d}t', \tag{42}$$

where $C \leq \frac{18}{\sqrt{2\pi}}$. To estimate this, we just need the following Lemma bounding how close the characteristic function is to a Gaussian for small t'.

Lemma 5. *Let $|p\rangle$ be a product state such that $\langle H \rangle \equiv \langle p| H |p\rangle = 0$ and $\langle H^2 \rangle = \sigma^2$. Moreover, let $t' \leq \frac{t^* \sigma}{2}$ with $t^* = \mathcal{O}(1)$. Then,*

$$\left| \log \phi(t') - \left(\frac{-(t')^2}{2} \right) \right| \leq 2 \frac{N}{\sigma^3} \left(\frac{t'}{t^*} \right)^3. \tag{43}$$

This follows straightforwardly from Theorem 10 in [9], where t^* is defined as an $\mathcal{O}(1)$ number that depends on the connectivity of the Hamiltonian. As a consequence, we have that, for $t' \leq \frac{t^* \sigma}{2}$,

$$-2 \frac{N}{\sigma^3} \left(\frac{t'}{t^*} \right)^3 \leq \log \phi(t') + \left(\frac{(t')^2}{2} \right) \leq 2 \frac{N}{\sigma^3} \left(\frac{t'}{t^*} \right)^3, \tag{44}$$

Subtracting the term $\frac{(t')^2}{2}$ from the whole equation, and then exponentiating, we get

$$e^{-\frac{(t')^2}{2}} \left(1 - \mathcal{O}\left(\frac{N}{\sigma^3} \left(\frac{t'}{t^*} \right)^3 \right) \right) \leq \phi(t') \leq e^{-\frac{(t')^2}{2}} \left(1 + \mathcal{O}\left(\frac{N}{\sigma^3} \left(\frac{t'}{t^*} \right)^3 \right) \right) \tag{45}$$

Subtracting $e^{-\frac{(t')^2}{2}}$ from both sides,

$$|\phi(t') - e^{-\frac{t'^2}{2}}| \leq e^{\frac{-(t')^2}{2}} \mathcal{O}\left(\frac{N}{\sigma^3} \left(\frac{t'}{t^*} \right)^3 \right). \tag{46}$$

Plugging this bound in Eq. (42) and integrating yields,

$$\pi \zeta_N \leq \frac{C}{T} + \int_0^T e^{\frac{-(t')^2}{2}} \mathcal{O}\left(\frac{N}{\sigma^3} \frac{(t')^2}{(t^*)^3} \right) dt' \tag{47}$$

Integrating and choosing $T = \frac{\sigma t^*}{2}$ and $t^* = \mathcal{O}(1)$,

$$\pi \zeta_N \leq \frac{2C}{\sigma t^*} + \frac{3N}{(\sigma t^*)^3} = \mathcal{O}\left(\frac{1}{\sqrt{N}} \right), \tag{48}$$

using Assumption 1, the result follows.

B   Proof of Lemma 3

To have a truncated number of terms in the final bond dimension calculation we use the approximate cosine filter, $g_y \left(\frac{H-E}{N} \right) = \frac{1}{2^M} \sum_{m=-y\sqrt{M}}^{y\sqrt{M}} \binom{M}{\frac{M}{2}-m} e^{i2m\left(\frac{H-E}{N} \right)}$, where E is the average energy around which we filter. Applying this operator to the initial state $|p\rangle$, the normalized filtered state $|p_{M,y}\rangle$ can be written as

$$|p_{M,y}\rangle = \frac{g_y \left(\frac{H-E}{N} \right) |p\rangle}{\left\| g_y \left(\frac{H-E}{N} \right) |p\rangle \right\|}. \tag{49}$$

Let the variance of the initial state $|p\rangle$ be $\sigma = \mathcal{O}(\sqrt{N})$. Given the Hamiltonian operator H, the new variance is defined as

$$\delta^2 = \langle p_{M,y}| H^2 |p_{M,y}\rangle - \langle p_{M,y}| H |p_{M,y}\rangle^2, \tag{50}$$

where $\mu = \langle p_{M,y}| H |p_{M,y}\rangle$ is the average of the filtered state.

The average is not expected to change significantly from E on application of the filter operator since both the filter and the initial state are taken to be centered around the energy E and the filter is symmetric about the average. We first prove this by showing that $|\mu - E| = \mathcal{O}\left(\frac{N}{\sqrt{M}}\right)$, and then upper bound the variance in a similar manner. Without loss of generality, we choose $E = 0$. We begin by writing the initial state $|p\rangle$ in the energy eigenbasis of Hamiltonian H,

$$|p\rangle = \sum_{E_j} b_j |E_j\rangle, \tag{51}$$

where E_j's represent the eigenvalues. We now calculate the average μ by substituting $|p\rangle$ from the above equation

$$\mu = \langle p_{M,y}| H |p_{M,y}\rangle = \frac{\langle p| g_y^2\left(\frac{H}{N}\right) H |p\rangle}{\left\| g_y\left(\frac{H-E}{N}\right) |p\rangle \right\|^2} = \frac{\sum_{E_j} |b_j|^2 g_y^2\left(\frac{E_j}{N}\right) E_j}{\sum_{E_j} |b_j|^2 g_y^2\left(\frac{E_j}{N}\right)}. \tag{52}$$

Recall that the approximate cosine filter was introduced (Lemma 2) to decrease the extensive number of terms in the time evolution representation of the original cosine filter. The approximate cosine filter is related to the original filter in terms of the following inequalities

$$\cos^{2M}\left(\frac{H}{N}\right) - 4e^{-\frac{y^2}{2}} \leq g_y^2\left(\frac{H}{N}\right) \leq \cos^{2M}\left(\frac{H}{N}\right) + 4e^{-y^2}. \tag{53}$$

In the upcoming sections, we will use the setting introduced above to bound the average and variance of the final filtered state.

B.1 Average energy

The distance from the mean μ can be bounded as follows by substituting equation (53) into equation (52),

$$|\mu| \leq \left| \frac{\sum_{E_j} |b_j|^2 \cos^{2M}\left(\frac{E_j}{N}\right) E_j + 4e^{-y^2} \sum_{E_j} |b_j|^2 |E_j|}{\sum_{E_j} |b_j|^2 \cos^{2M}\left(\frac{E_j}{N}\right) - 4e^{-\frac{y^2}{2}} \sum_{E_j} |b_j|^2} \right| \leq \left| \frac{\sum_{E_j} |b_j|^2 \cos^{2M}\left(\frac{E_j}{N}\right) E_j + 4e^{-y^2} \sum_{E_j} |b_j|^2 |E_j|}{\sum_{E_j} |b_j|^2 \cos^{2M}\left(\frac{E_j}{N}\right) - 4e^{-\frac{y^2}{2}}} \right| \tag{54}$$

$$:= \left| \frac{\mathcal{M}_{1(+)} + \mathcal{M}_{1(-)} + \mathcal{E}_1}{\mathcal{M}_2} \right|,$$

where the second inequality uses that the initial average energy is 0 and $\sum_{E_j} |b_j|^2 = 1$, and $\mathcal{M}_{1(+)}$, $\mathcal{M}_{1(-)}$, $\mathcal{E}_1$, and $\mathcal{M}_2$ are defined as follows

$$\mathcal{M}_{1(+)} = \sum_{E_j > 0} |b_j|^2 \cos^{2M}\left(\frac{E_j}{N}\right) E_j, \tag{55}$$

$$\mathcal{M}_{1(-)} = \sum_{E_j < 0} |b_j|^2 \cos^{2M}\left(\frac{E_j}{N}\right) E_j, \tag{56}$$

$$\mathcal{E}_1 = 4e^{-y^2} \sum_{E_j} |b_j|^2 |E_j|, \tag{57}$$

$$\mathcal{M}_2 = \sum_{E_j} |b_j|^2 \cos^{2M}\left(\frac{E_j}{N}\right) - 4e^{-\frac{y^2}{2}}. \tag{58}$$

First we focus on simplifying the numerator of Eq. (54). We use what we term the *segmentation method* (motivated from a similar procedure used in [26]). As such, the initial steps of the proof mirror those of Appendix E in [26], although here the object we are estimating yields a variety of additional technical complications, due to the presence of the filter, and require more careful error estimates. The start of the method is to divide the energy range $E_{\max} \sim \mathcal{O}(N)$ into R_1 number of segments of width $\Lambda_1 > 0$ each, so that

$$\mathcal{M}_{1(+)} = \sum_{E_j > 0} |b_j|^2 \cos^{2M}\left(\frac{E_j}{N}\right) E_j \leq \sum_{l=0}^{R_1 - 1} \cos^{2M}\left(\frac{\Lambda_1 l}{N}\right)(\Lambda_1(l+1)) \sum_{\Lambda_1 l < E_j < \Lambda_1(l+1)} |b_j|^2. \tag{59}$$

Using Definition 1, the coefficients b_j can be approximated using a Gaussian integral

$$\sum_{\Lambda_1 l < E_j < \Lambda_1(l+1)} |b_j|^2 \leq \int_{\Lambda_1 l}^{\Lambda_1(l+1)} \frac{dt}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2\sigma^2}} + \zeta_N. \tag{60}$$

Since by definition $\mathcal{M}_{1(-)} \leq 0$, the terms in the numerator can be bounded as

$$\mathcal{M}_{1(+)} \leq \sum_{l=0}^{R_1 - 1} \cos^{2M}\left(\frac{\Lambda_1 l}{N}\right)(\Lambda_1(l+1)) \left[\int_{\Lambda_1 l}^{\Lambda_1(l+1)} \frac{dt}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2\sigma^2}} + \zeta_N\right], \tag{61}$$

$$-\mathcal{M}_{1(-)} \leq \sum_{l=0}^{R_1 - 1} \cos^{2M}\left(\frac{\Lambda_1 l}{N}\right)(\Lambda_1(l+1)) \left[\int_{-\Lambda_1(l+1)}^{-\Lambda_1 l} \frac{dt}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2\sigma^2}} + \zeta_N\right], \tag{62}$$

Let $\tilde{t} = \frac{t}{\sqrt{2}\sigma}$, $d\tilde{t} = \frac{dt}{\sqrt{2}\sigma}$, and $\tilde{\Lambda}_1 = \frac{\Lambda_1}{\sqrt{2}\sigma}$, then $\mathcal{M}_{1(+)}$ can be bounded as

$$\mathcal{M}_{1(+)} \leq \frac{\sqrt{2}\sigma}{\sqrt{\pi}} \sum_{l=0}^{R_1 - 1} \int_{\tilde{\Lambda}_1 l}^{\tilde{\Lambda}_1(l+1)} d\tilde{t} \cos^{2M}\left(\sqrt{2}\sigma\left(\frac{\tilde{t} - \tilde{\Lambda}_1}{N}\right)\right)(\tilde{t} + \tilde{\Lambda}_1)e^{-\tilde{t}^2} +$$
$$+ \sqrt{2}\sigma\zeta_N \sum_{l=0}^{R_1 - 1} \cos^{2M}\left(\frac{\sqrt{2}\sigma\tilde{\Lambda}_1 l}{N}\right)\tilde{\Lambda}_1(l+1) \tag{63}$$

$$\Rightarrow \mathcal{M}_{1(+)} \leq \frac{\sqrt{2}\sigma}{\sqrt{\pi}} \int_{0}^{\tilde{\Lambda}_1 R_1} d\tilde{t} \cos^{2M}\left(\sqrt{2}\sigma\left(\frac{\tilde{t} - \tilde{\Lambda}_1}{N}\right)\right)(\tilde{t} + \tilde{\Lambda}_1)e^{-\tilde{t}^2} +$$
$$+ \sqrt{2}\sigma\zeta_N \sum_{l=0}^{R_1 - 1} \cos^{2M}\left(\frac{\sqrt{2}\sigma\tilde{\Lambda}_1 l}{N}\right)(\tilde{\Lambda}_1(l+1)). \tag{64}$$

Similarly

$$-\mathcal{M}_{1(-)} \leq -\frac{\sqrt{2}\sigma}{\sqrt{\pi}} \int_{-\tilde{\Lambda}_1 R_1}^{0} d\tilde{t} \cos^{2M}\left(\sqrt{2}\sigma\left(\frac{\tilde{t} + \tilde{\Lambda}_1}{N}\right)\right)(\tilde{t} - \tilde{\Lambda}_1)e^{-\tilde{t}^2} +$$
$$+ \sqrt{2}\sigma\zeta_N \sum_{l=0}^{R_1 - 1} \cos^{2M}\left(\frac{\sqrt{2}\sigma\tilde{\Lambda}_1 l}{N}\right)(\tilde{\Lambda}_1(l+1)). \tag{65}$$

We get the following upper bound on the $|\mathcal{M}_{1(+)} + \mathcal{M}_{1(-)}|$,

$$|\mathcal{M}_{1(+)} + \mathcal{M}_{1(-)}| \leq \frac{2\sqrt{2}\sigma}{\sqrt{\pi}} \int\limits_0^{\tilde{\Lambda}_1 R_1} d\tilde{t} e^{-\tilde{t}^2} \left[\cos^{2M}\left(\sqrt{2}\sigma\left(\frac{\tilde{t} - \tilde{\Lambda}_1}{N}\right)\right)\tilde{\Lambda}_1\right] + \qquad (66)$$
$$+ 2\sqrt{2}\sigma\zeta_N \sum_{l=0}^{R_1-1} \cos^{2M}\left(\frac{\sqrt{2}\sigma\tilde{\Lambda}_1 l}{N}\right)(\tilde{\Lambda}_1(l+1)).$$

Additionally, the error term $\mathcal{E}_1$ can be bounded similarly as

$$\mathcal{E}_1 \leq 4e^{-y^2} \sum_{l=0}^{R_1-1} \Lambda_1(l+1) \sum_{\Lambda_1 l \leq E_j \leq \Lambda_1(l+1)} |b_j|^2 \qquad (67)$$

$$\leq 4e^{-y^2} \left[\frac{\sqrt{2}\sigma}{\sqrt{\pi}} \int_0^{\tilde{\Lambda}_1 R_1} d\tilde{t} e^{-\tilde{t}^2}(\tilde{t} + \tilde{\Lambda}_1) + \sqrt{2}\sigma\zeta_N \sum_{l=0}^{R_1-1} \tilde{\Lambda}_1(l+1)\right] \qquad (68)$$

$$\leq 2\sqrt{2}\sigma e^{-y^2} \left[\frac{(1 + \sqrt{\pi}\tilde{\Lambda}_1)}{\sqrt{\pi}} + \zeta_N \tilde{\Lambda}_1(R_1^2 + 3R_1)\right] \qquad (69)$$

$$\leq 2\sqrt{2}\sigma e^{-y^2} \left[\frac{(1 + \sqrt{\pi}\tilde{\Lambda}_1)}{\sqrt{\pi}} + 4\zeta_N\sqrt{N}R_1\right]. \qquad (70)$$

Applying a similar method to lower bound the first term in the denominator of Eq. (54),

$$\sum_{E_j} |b_j|^2 \cos^{2M}\left(\frac{E_j}{N}\right) \geq 2 \sum_{l=0}^{R_2-1} \cos^{2M}\left(\frac{\Lambda_2(l+1)}{N}\right) \sum_{\Lambda_2 l < E_j < \Lambda_2(l+1)} |b_j|^2 \qquad (71)$$

$$\geq 2 \sum_{l=0}^{R_2-1} \cos^{2M}\left(\frac{\Lambda_2(l+1)}{N}\right)\left[\int\limits_{\Lambda_2 l}^{\Lambda_2(l+1)} \frac{dt}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2\sigma^2}} - \zeta_N\right] \qquad (72)$$

$$\geq \frac{2}{\sqrt{\pi}} \sum_{l=0}^{R_2-1} \int\limits_{\tilde{\Lambda}_2 l}^{\tilde{\Lambda}_2(l+1)} d\tilde{t} \cos^{2M}\left(\sqrt{2}\sigma\left(\frac{\tilde{t} + \tilde{\Lambda}_2}{N}\right)\right) e^{-\tilde{t}^2}$$
$$- 2\zeta_N \sum_{l=0}^{R_2-1} \cos^{2M}\left(\frac{\sqrt{2}\sigma\tilde{\Lambda}_2(l+1)}{N}\right). \qquad (73)$$

We get the following lower bound on the denominator

$$\mathcal{M}_2 \geq \frac{2}{\sqrt{\pi}} \int\limits_0^{\tilde{\Lambda}_2 R_2} d\tilde{t} \cos^{2M}\left(\sqrt{2}\sigma\left(\frac{\tilde{t} + \tilde{\Lambda}_2}{N}\right)\right) e^{-\tilde{t}^2} - 2\zeta_N \sum_{l=0}^{R_2-1} \cos^{2M}\left(\frac{\sqrt{2}\sigma\tilde{\Lambda}_2(l+1)}{N}\right) - 4e^{-\frac{y^2}{2}}. \quad (74)$$

Combining the bounds on the numerator (Eq. (66) and Eq. (70)) and the denominator (Eq. (74))

$$|\mu| \leq \frac{\sqrt{2}\sigma\left[\frac{1}{\sqrt{\pi}} \int\limits_0^{\tilde{\Lambda}_1 R_1} d\tilde{t} e^{-\tilde{t}^2}\left[\cos^{2M}\left(\sqrt{2}\sigma\left(\frac{\tilde{t} - \tilde{\Lambda}_1}{N}\right)\right)\tilde{\Lambda}_1\right] + \mathcal{U}\right]}{\frac{1}{\sqrt{\pi}} \int\limits_0^{\tilde{\Lambda}_2 R_2} d\tilde{t} \cos^{2M}\left(\sqrt{2}\sigma\left(\frac{\tilde{t} + \tilde{\Lambda}_2}{N}\right)\right) e^{-\tilde{t}^2} - \zeta_N \sum\limits_{l=0}^{R_2-1} \cos^{2M}\left(\frac{\sqrt{2}\sigma\tilde{\Lambda}_2(l+1)}{N}\right) - 2e^{-\frac{y^2}{2}}}, \qquad (75)$$

where

$$\mathcal{U} = \zeta_N \sum_{l=0}^{R_1-1} \cos^{2M}\left(\frac{\sqrt{2}\sigma\tilde{\Lambda}_1 l}{N}\right)(\tilde{\Lambda}_1(l+1)) + 2e^{-y^2}\left[\frac{(1 + \sqrt{\pi}\tilde{\Lambda}_1)}{\sqrt{\pi}} + 4\zeta_N\sqrt{N}R_1\right]. \qquad (76)$$

Note the following upper and lower bounds on $\cos^a(x)$ for $a > 0$,

$$e^{-ax^2} \leq \cos^a(x) \leq e^{-\frac{ax^2}{2}}. \tag{77}$$

We can use these to replace the cosine power function in the numerator and denominator to obtain

$$|\mu| \leq \frac{\frac{\sqrt{2}\sigma\tilde{\Lambda}_1}{\sqrt{\pi}} \int\limits_0^{\tilde{\Lambda}_1 R_1} d\tilde{t}\, e^{-2M\sigma^2(\frac{\tilde{t}-\tilde{\Lambda}_1}{N})^2} e^{-\tilde{t}^2} + \sqrt{2}\sigma\zeta_N \sum\limits_{l=0}^{R_1-1} e^{-2M\sigma^2\frac{\tilde{\Lambda}_1^2 l^2}{N^2}}(\tilde{\Lambda}_1(l+1)) + 2\sqrt{2}\sigma e^{-y^2}\left[\frac{(1+\sqrt{\pi}\tilde{\Lambda}_1)}{\sqrt{\pi}} + 4\zeta_N\sqrt{N}R_1\right]}{\frac{1}{\sqrt{\pi}}\int\limits_0^{\tilde{\Lambda}_2 R_2} d\tilde{t}\, e^{-4M\sigma^2(\frac{\tilde{t}+\tilde{\Lambda}_2}{N})^2} e^{-\tilde{t}^2} - \zeta_N \sum\limits_{l=0}^{R_2-1} e^{-4M\sigma^2(\frac{\tilde{\Lambda}_2(l+1)}{N})^2} - 4e^{-\frac{y^2}{2}}}. \tag{78}$$

Now using Assumption 1 by substituting $\sigma = s\sqrt{N}$,

$$|\mu| \leq \frac{\sqrt{2N}s\left[\frac{\tilde{\Lambda}_1}{\sqrt{\pi}} \int\limits_0^{\tilde{\Lambda}_1 R_1} d\tilde{t}\, e^{-\frac{2s^2 M}{N}(\tilde{t}-\tilde{\Lambda}_1)^2} e^{-\tilde{t}^2} + \zeta_N \sum\limits_{l=0}^{R_1-1} e^{-\frac{2s^2 M}{N}\tilde{\Lambda}_1^2 l^2}(\tilde{\Lambda}_1(l+1)) + 2e^{-y^2}\left[\frac{(1+\sqrt{\pi}\tilde{\Lambda}_1)}{\sqrt{\pi}} + 4\zeta_N\sqrt{N}R_1\right]\right]}{\frac{1}{\sqrt{\pi}}\int\limits_0^{\tilde{\Lambda}_2 R_2} d\tilde{t}\, e^{-\frac{4s^2 M}{N}(\tilde{t}+\tilde{\Lambda}_2)^2} e^{-\tilde{t}^2} - \zeta_N \sum\limits_{l=0}^{R_2-1} e^{-\frac{4s^2 M}{N}(\tilde{\Lambda}_2(l+1))^2} - 4e^{-\frac{y^2}{2}}} := \frac{\mathcal{M}_1'}{\mathcal{M}_2'}. \tag{79}$$

The first term of $\mathcal{M}_1'$ can be bounded as

$$
\begin{aligned}
\frac{\sqrt{2N}s\tilde{\Lambda}_1}{\sqrt{\pi}} & \int\limits_0^{\tilde{\Lambda}_1 R_1} d\tilde{t}\, e^{-\frac{2s^2 M}{N}(\tilde{t}-\tilde{\Lambda}_1)^2} e^{-\tilde{t}^2} \\
& \leq \frac{\sqrt{N}s\tilde{\Lambda}_1 e^{-\frac{\tilde{\Lambda}_1^2}{(1+\frac{N}{2s^2 M})}}}{\sqrt{2}\sqrt{1+\frac{2s^2 M}{N}}}\left[\operatorname{erf}\left(\tilde{\Lambda}_1 R_1\sqrt{1+\frac{2s^2 M}{N}} - \frac{2s^2 M\tilde{\Lambda}_1}{N}\frac{1}{\sqrt{1+\frac{2s^2 M}{N}}}\right) \right. \\
& \left. + \operatorname{erf}\left(\frac{2s^2 M\tilde{\Lambda}_1}{N}\frac{1}{\sqrt{1+\frac{2s^2 M}{N}}}\right)\right] \\
& \leq \frac{\sqrt{2N}s\tilde{\Lambda}_1 e^{-\frac{\tilde{\Lambda}_1^2}{1+\frac{N}{2s^2 M}}}}{\sqrt{1+\frac{2s^2 M}{N}}}.
\end{aligned}
\tag{80}
$$

In the second inequality, we have used that the error function is bounded by 1. To bound the discrete sum error terms, we use the *Euler-Maclaurin formula* to change summations to integrals. It is given as follows

$$\sum_{i=a}^b f(i) - \int_a^b f(x)dx = \frac{f(a)+f(b)}{2} + \sum_{k=1}^{\lfloor \frac{p}{2} \rfloor} \frac{B_{2k}}{(2k)!}\left(f^{(2k-1)}(b) - f^{(2k-1)}(a)\right) + \mathcal{R}_p, \tag{81}$$

where B_n denotes the nth Bernoulli number, and the remainder term $\mathcal{R}_p$ is bounded as

$$|\mathcal{R}_p| \leq \frac{2\zeta(p)}{(2\pi)^p}\int_a^b |f^{(p)}(x)|dx. \tag{82}$$

The choice of p is up to us. Taking $p = 2$, the error term of $\mathcal{M}'_1$ can be bounded as

$$\sum_{l=0}^{R_1-1} e^{-\frac{2s^2 M}{N}\tilde{\Lambda}_1^2 l^2}(\tilde{\Lambda}_1(l+1)) \leq \int_0^{R_1} dl e^{-\frac{2s^2 M}{N}\tilde{\Lambda}_1^2 l^2}(\tilde{\Lambda}_1(l+1)) + \frac{\tilde{\Lambda}_1 + \tilde{\Lambda}_1 R_1 e^{-\frac{2s^2 M}{N}\tilde{\Lambda}_1^2(R_1-1)^2}}{2} + \tag{83}$$
$$+ \frac{B_2}{2}\tilde{\Lambda}_1 \left[e^{-\frac{2s^2 M}{N}\tilde{\Lambda}_1^2(R_1-1)^2}\left(1 - \frac{4s^2 M}{N}\tilde{\Lambda}_1^2 R_1(R_1-1)\right) - 1\right] + \mathcal{R}_2.$$

The integral on the RHS can be bounded as follows

$$\int_0^{R_1} dl e^{-\frac{2s^2 M}{N}\tilde{\Lambda}_1^2 l^2}(\tilde{\Lambda}_1(l+1)) \leq \frac{N}{4\tilde{\Lambda}_1 M s^2}(1 - e^{-\frac{2s^2 M}{N}\tilde{\Lambda}_1^2 R_1^2}) + \frac{\sqrt{\pi}}{2\sqrt{2}s}\sqrt{\frac{N}{M}}\,\mathrm{erf}\left(2\tilde{\Lambda}_1 R_1 s\sqrt{\frac{M}{N}}\right), \tag{84}$$

$$\leq \frac{N}{4\tilde{\Lambda}_1 M s^2} + \frac{\sqrt{\pi}}{2\sqrt{2}s}\sqrt{\frac{N}{M}}. \tag{85}$$

The the sum in Eq. (83) can be bounded as

$$\sum_{l=0}^{R_1-1} e^{-\frac{2s^2 M}{N}\tilde{\Lambda}_1^2 l^2}(\tilde{\Lambda}_1(l+1)) \leq \frac{N}{4\tilde{\Lambda}_1 M s^2} + \frac{\sqrt{\pi}}{2\sqrt{2}s}\sqrt{\frac{N}{M}} + \frac{\tilde{\Lambda}_1 + \tilde{\Lambda}_1 R_1 e^{-\frac{2s^2 M}{N}\tilde{\Lambda}_1^2(R_1-1)^2}}{2} + \tag{86}$$
$$+ \frac{\tilde{\Lambda}_1 e^{-\frac{2s^2 M}{N}\tilde{\Lambda}_1^2(R_1-1)^2}}{12} + \mathcal{R}_2,$$

where we substitute $B_2 = \frac{1}{6}$ for the second Bernoulli number and $\mathcal{R}_2$ is the remainder term from the approximation, bounded as follows

$$\mathcal{R}_2 \leq \frac{\tilde{\Lambda}_1\zeta(2)}{2\pi^2}\left(5 + \sqrt{\frac{8\pi s^2\tilde{\Lambda}_1^2 M}{N}}\right). \tag{87}$$

This means that

$$\sum_{l=0}^{R_1-1} e^{-\frac{2s^2 M}{N}\tilde{\Lambda}_1^2 l^2}(\tilde{\Lambda}_1(l+1)) \leq \frac{N}{4\tilde{\Lambda}_1 M s^2} + \frac{\sqrt{\pi}}{2\sqrt{2}s}\sqrt{\frac{N}{M}} + \frac{\tilde{\Lambda}_1 + \tilde{\Lambda}_1(R_1 + \frac{1}{6})e^{-\frac{2s^2 M}{N}\tilde{\Lambda}_1^2(R_1-1)^2}}{2} + \tag{88}$$
$$+ \frac{\tilde{\Lambda}_1\zeta(2)}{2\pi^2}\left(5 + \sqrt{\frac{8\pi s^2\tilde{\Lambda}_1^2 M}{N}}\right).$$

The value of the constant C_1 is arbitrary. To achieve the best possible upper bound on the variance, we find that the best choice is $C_1 := \frac{2s^2 M}{N}\tilde{\Lambda}_1^2$. Then we can bound $\mathcal{M}'_1$ by adding Eq. (80) and Eq. (88) as

$$\mathcal{M}'_1 \leq \frac{\sqrt{2N}s\tilde{\Lambda}_1 e^{-\frac{\tilde{\Lambda}_1^2}{(1+\frac{N}{2s^2 M})}}}{\sqrt{1 + \frac{2s^2 M}{N}}} + \sqrt{2N}s\zeta_N\left[\frac{\tilde{\Lambda}_1}{2C_1}\left(1 + \sqrt{\pi C_1}\right) + \frac{\tilde{\Lambda}_1 + \tilde{\Lambda}_1(R_1 + \frac{1}{6})e^{-C_1(R_1-1)^2}}{2} + \right.$$
$$\left. + \frac{\tilde{\Lambda}_1\zeta(2)}{2\pi^2}\left(5 + 2\sqrt{\pi C_1}\right)\right] \tag{89}$$
$$+ 2\sqrt{2N}s e^{-y^2}\zeta_N\sqrt{N}R_1 + 2s\sqrt{N}e^{-y^2}\frac{(1 + \sqrt{\pi}\tilde{\Lambda}_1)}{\sqrt{\pi}}.$$

Using $\tilde{\Lambda}_1 = \frac{\Lambda_1}{s\sqrt{2N}}$, we can write the bound as follow

$$\mathcal{M}_1' \leq \sqrt{\frac{N}{2s^2M}}\frac{\Lambda_1 e^{-\frac{1}{2s^2}\frac{\Lambda_1^2}{N}\frac{1}{1+\frac{N}{2s^2M}}}}{\sqrt{1+\frac{N}{2s^2M}}} + \zeta_N\left[\frac{\Lambda_1}{2C_1}\left(1+\sqrt{\pi C_1}\right) + \frac{\Lambda_1 + \Lambda_1(R_1+\frac{1}{6})e^{-C_1(R_1-1)^2}}{2} + \right.$$

$$\left. + \frac{\Lambda_1 \zeta(2)}{2\pi^2}\left(5 + 2\sqrt{\pi C_1}\right)\right] + \tag{90}$$

$$+ 2\sqrt{2N}se^{-y^2}\zeta_N\sqrt{N}R_1 + 2s\sqrt{N}e^{-y^2}\frac{\left(1+\sqrt{\pi}\tilde{\Lambda}_1\right)}{\sqrt{\pi}}.$$

Note that $e^{-\frac{1}{2s^2}\frac{\Lambda_1^2}{N}\frac{1}{1+\frac{N}{2s^2M}}} \leq 1$ and $(1+\frac{N}{2s^2M})^{-1/2} \leq 1$ for any value of the free parameters and this is particularly tight approximation for $\frac{N}{2s^2M} < 1$. Thus, choosing Λ_1 such that $C_1 = \mathcal{O}(1)$, we get the following overall upper bounds on the numerator of the average

$$\mathcal{M}_1' \leq \frac{1}{s}\sqrt{\frac{N}{2M}}\Lambda_1 + 2\sqrt{2}se^{-y^2}N\zeta_N R_1 + 2s\sqrt{N}e^{-y^2}\frac{\left(1+\sqrt{\pi}\tilde{\Lambda}_1\right)}{\sqrt{\pi}} + \mathcal{O}(\zeta_N\Lambda_1). \tag{91}$$

For the choice of y, we require the following condition, such that the 2nd term is smaller than the leading term

$$e^{-y^2} \leq \frac{1}{4s^2}\sqrt{\frac{\Lambda_1^2}{2MN\zeta_N^2 R_1^2}}. \tag{92}$$

First substituting $R = \frac{N}{\Lambda_1}$ and then $\Lambda_1 = \sqrt{C_1}\frac{N}{\sqrt{M}}$, we get the bound

$$y = \Omega\left(\sqrt{\log\frac{M^3\zeta_N^2}{N}}\right). \tag{93}$$

We now use Lemma (1) to find the worst case bound on y which will be applicable for all values of ζ_N,

$$y = \Omega\left(\sqrt{\log\frac{M^3}{N^2}}\right). \tag{94}$$

The parameter y dictates how many terms in Eq. (13) are to be included in the later calculation of the bond dimension. Using that $C_1 = \mathcal{O}(1)$, we get the scalings $\Lambda_1 \propto \frac{N}{\sqrt{M}}$ and $R_1 \propto \sqrt{M}$. Considering these, we choose $y = \Omega\left(\sqrt{\log\frac{M^3\zeta_N^2}{N}}\right)$, to get the following upper bound on the numerator

$$\mathcal{M}_1' \leq \frac{1}{s}\sqrt{\frac{N}{2M}}\Lambda_1 + \mathcal{O}(\zeta_N\Lambda_1). \tag{95}$$

We now simplify the denominator of Eq. (79). Define the variable

$$C_2 := \frac{4s^2M}{N}\tilde{\Lambda}_2^2 = \frac{2M\Lambda_2^2}{N^2}, \tag{96}$$

Using this value of C_2 and the Euler-Maclaurin formula from the equation (81), the integral in the first term of denominator $\mathcal{M}_2'$ can be lower bounded as

$$\frac{1}{\sqrt{\pi}} \int_0^{\tilde{\Lambda}_2 R_2} d\tilde{t}\, e^{-\frac{4s^2 M}{N}(\tilde{t}+\tilde{\Lambda}_2)^2} e^{-\tilde{t}^2} \geq \frac{1}{2} \frac{e^{-\frac{4s^2 M}{N}\frac{\tilde{\Lambda}_2^2}{1+\frac{4s^2 M}{N}}}}{\sqrt{1+\frac{4s^2 M}{N}}} \left[\mathrm{erf}\left(\tilde{\Lambda}_2 R_2 \sqrt{1+\frac{4s^2 M}{N}} + \frac{\tilde{\Lambda}_2}{\sqrt{\frac{N}{4s^2 M}}\sqrt{1+\frac{N}{4s^2 M}}} \right) \right.$$
$$\left. - \mathrm{erf}\left(\frac{\tilde{\Lambda}_2}{\sqrt{\frac{N}{4s^2 M}}\sqrt{1+\frac{N}{4s^2 M}}} \right) \right] \tag{97}$$

Since we have the freedom of choosing $\tilde{\Lambda}_2$, we take $C_2 = \frac{3}{4}$ (or $\tilde{\Lambda}_2 = \frac{3N}{16 s^2 M}$). Then the integral is bounded as

$$\frac{1}{\sqrt{\pi}} \int_0^{\tilde{\Lambda}_2 R_2} d\tilde{t}\, e^{-\frac{4s^2 M}{N}(\tilde{t}+\tilde{\Lambda}_2)^2} e^{-\tilde{t}^2} \geq \frac{1}{2} \frac{\tilde{\Lambda}_2 e^{-\frac{3\tilde{\Lambda}_2^2}{4\tilde{\Lambda}_2^2+3}}}{\sqrt{\tilde{\Lambda}_2^2+\frac{3}{4}}} \left[\mathrm{erf}\left(R_2 \sqrt{\tilde{\Lambda}_2^2+\frac{3}{4}} + \frac{3}{4\sqrt{\tilde{\Lambda}_2^2+\frac{3}{4}}} \right) - \mathrm{erf}\left(\frac{3}{4\sqrt{\tilde{\Lambda}_2^2+\frac{3}{4}}} \right) \right] \tag{98}$$

$$\Rightarrow \frac{1}{\sqrt{\pi}} \int_0^{\tilde{\Lambda}_2 R_2} d\tilde{t}\, e^{-\frac{4s^2 M}{N}(\tilde{t}+\tilde{\Lambda}_2)^2} e^{-\tilde{t}^2} \geq \frac{1}{2} \frac{\tilde{\Lambda}_2 e^{-\frac{3\tilde{\Lambda}_2^2}{4\tilde{\Lambda}_2^2+3}}}{\sqrt{\tilde{\Lambda}_2^2+\frac{3}{4}}} \left[\mathrm{erf}\left(R_2 \sqrt{\frac{3}{4}} \right) - \mathrm{erf}\left(\frac{3}{4\sqrt{\tilde{\Lambda}_2^2+\frac{3}{4}}} \right) \right]. \tag{99}$$

Considering the power series approximation of $\mathrm{erf}(x)$ for $x = \Omega(1)$ we get the following lower bound

$$\mathrm{erf}(x) \geq 1 - \frac{1}{\sqrt{\pi}}\frac{e^{-x^2}}{x}. \tag{100}$$

Substituting this bound in the first term of Eq. (99),

$$\frac{1}{\sqrt{\pi}} \int_0^{\tilde{\Lambda}_2 R_2} d\tilde{t}\, e^{-\frac{4s^2 M}{N}(\tilde{t}+\tilde{\Lambda}_2)^2} e^{-\tilde{t}^2} \geq \frac{1}{2} \frac{\tilde{\Lambda}_2 e^{-\frac{3\tilde{\Lambda}_2^2}{4\tilde{\Lambda}_2^2+3}}}{\sqrt{\tilde{\Lambda}_2^2+\frac{3}{4}}} \left[1 - \frac{2}{\sqrt{3\pi}}\frac{e^{-\frac{3}{4}R_2^2}}{R_2} - \mathrm{erf}\left(\frac{3}{4\sqrt{\tilde{\Lambda}_2^2+\frac{3}{4}}} \right) \right]. \tag{101}$$

The second term of $\mathcal{M}_2'$ in Eq. (80) can be bounded as

$$\zeta_N \sum_{l=0}^{R_2-1} e^{-\frac{4s^2 M}{N}(\tilde{\Lambda}_2(l+1))^2} \leq \zeta_N \int_0^{R_2-1} dl\, e^{-\frac{4s^2 M}{N}(\tilde{\Lambda}_2(l+1))^2} + \zeta_N \frac{e^{-\frac{4s^2 M}{N}\tilde{\Lambda}_2^2} + e^{-\frac{4s^2 M}{N}\tilde{\Lambda}_2^2 R_2^2}}{2} +$$
$$+\zeta_N \frac{2s^2 M}{3N}\tilde{\Lambda}_2^2 \left(e^{-\frac{4s^2 M}{N}\tilde{\Lambda}_2^2} - R_2 e^{-\frac{4s^2 M}{N}\tilde{\Lambda}_2^2 R_2^2} \right) +$$
$$+4\sqrt{\pi}\zeta_N s \sqrt{\frac{M}{N}}\tilde{\Lambda}_2 \left[\mathrm{erf}\left(\sqrt{\frac{4s^2 M}{N}}\tilde{\Lambda}_2 R_2 \right) - \mathrm{erf}\left(\sqrt{\frac{4s^2 M}{N}}\tilde{\Lambda}_2 \right) \right] +$$
$$+\frac{8s^2 M}{N}\zeta_N \tilde{\Lambda}_2^2 \left[e^{-\frac{4s^2 M}{N}\tilde{\Lambda}_2^2} - R_2 e^{-\frac{4s^2 M}{N}\tilde{\Lambda}_2^2 R_2^2} \right]. \tag{102}$$

$$\zeta_N \sum_{l=0}^{R_2-1} e^{-\frac{4s^2 M}{N}(\tilde{\Lambda}_2(l+1))^2} \leq \zeta_N \sqrt{\pi}\sqrt{\frac{N}{16 s^2 M}}\frac{1}{\tilde{\Lambda}_2}\left[\mathrm{erf}\left(\sqrt{\frac{4s^2 M}{N}}\tilde{\Lambda}_2 R_2 \right) - \mathrm{erf}\left(\sqrt{\frac{4s^2 M}{N}}\tilde{\Lambda}_2 \right) \right] +$$
$$+ \zeta_N e^{-\frac{4s^2 M}{N}\tilde{\Lambda}_2^2} + \zeta_N \frac{2s^2 M}{3N}\tilde{\Lambda}_2^2 e^{-\frac{4s^2 M}{N}\tilde{\Lambda}_2^2}$$
$$+ \sqrt{\pi}\zeta_N \sqrt{\frac{16 s^2 M}{N}}\tilde{\Lambda}_2 \, \mathrm{erf}\left(\sqrt{\frac{4s^2 M}{N}}\tilde{\Lambda}_2 R_2 \right) + \frac{8s^2 M}{N}\zeta_N \tilde{\Lambda}_2^2 e^{-\frac{4s^2 M}{N}\tilde{\Lambda}_2^2}. \tag{103}$$

In terms of C_2, the error term is

$$\zeta_N \sum_{l=0}^{R_2-1} e^{-\frac{4s^2 M}{N}(\tilde{\Lambda}_2(l+1))^2} \leq \zeta_N \sqrt{\pi} \frac{1}{2\sqrt{C_2}} \left[\mathrm{erf}\left(\sqrt{C_2}R_2\right) - \mathrm{erf}(\sqrt{C_2}) \right] + \zeta_N e^{-C_2} + \zeta_N \frac{C_2}{6} e^{-C_2} + \tag{104}$$
$$+ 2\sqrt{\pi}\zeta_N \sqrt{C_2}\,\mathrm{erf}(\sqrt{C_2}R_2) + 2C_2\zeta_N e^{-C_2}.$$

Substituting with the choice $C_2 = \frac{3}{4}$,

$$\zeta_N \sum_{l=0}^{R_2-1} e^{-\frac{4s^2 M}{N}(\tilde{\Lambda}_2(l+1))^2} \leq 3.7\zeta_N \,\mathrm{erf}(0.86R_2) + 1.3\zeta_N. \tag{105}$$

Combining everything, the lower bound on $\mathcal{M}_2'$ can be written as

$$\mathcal{M}_2' \geq \frac{1}{2} \frac{\tilde{\Lambda}_2 e^{-\frac{3\tilde{\Lambda}_2^2}{4\tilde{\Lambda}_2^2+3}}}{\sqrt{\tilde{\Lambda}_2^2 + \frac{3}{4}}} \left[1 - \frac{2}{\sqrt{3\pi}} \frac{e^{-\frac{3}{4}R_2^2}}{R_2} - \mathrm{erf}\left(\frac{3}{4\sqrt{\tilde{\Lambda}_2^2 + \frac{3}{4}}}\right) \right] - 3.7\zeta_N \,\mathrm{erf}(0.86R_2) - 1.3\zeta_N - 4e^{-\frac{y^2}{2}}. \tag{106}$$

Replacing $\frac{\tilde{\Lambda}_2^2}{\tilde{\Lambda}_2^2 + \frac{3}{4}} = \frac{\frac{N}{4s^2 M}}{\frac{N}{4s^2 M}+1}$,

$$\mathcal{M}_2' \geq \frac{1}{2} \sqrt{\frac{\frac{N}{4s^2 M}}{\frac{N}{4s^2 M}+1}} e^{-\frac{3}{4}\frac{\frac{N}{4s^2 M}}{\frac{N}{4s^2 M}+1}} \left[1 - \frac{2}{\sqrt{3\pi}} \frac{e^{-\frac{3}{4}R_2^2}}{R_2} - \mathrm{erf}\left(\frac{3}{4\sqrt{\tilde{\Lambda}_2^2 + \frac{3}{4}}}\right) \right] - 5\zeta_N - 4e^{-\frac{y^2}{2}}. \tag{107}$$

Again assuming that $\frac{N}{4s^2 M} \leq 1$, and using $\left(1 + \frac{N}{4s^2 M}\right)^{-\frac{1}{2}} \geq \left(1 - \frac{N}{8s^2 M}\right) \geq \frac{1}{2}$, and $\frac{\frac{N}{4s^2 M}}{\frac{N}{4s^2 M}+1} \leq \frac{N}{4s^2 M}$ in the exponent

$$\mathcal{M}_2' \geq \frac{1}{4} \sqrt{\frac{N}{4s^2 M}} e^{-\frac{3}{4}\frac{N}{4s^2 M}} \left[1 - \frac{2}{\sqrt{3\pi}} \frac{e^{-\frac{3}{4}R_2^2}}{R_2} - \mathrm{erf}\left(\sqrt{\frac{3}{4}}\right) \right] - 5\zeta_N - 4e^{-\frac{y^2}{2}}. \tag{108}$$

Now, since $e^{-\frac{3}{4}\frac{N}{4s^2 M}} \geq \left(1 - \frac{3}{4}\frac{N}{4s^2 M}\right)$, and $\mathrm{erf}(0.86R_2) \leq 1$,

$$\mathcal{M}_2' \geq \frac{1}{2} \sqrt{\frac{N}{4s^2 M}} \left(1 - \frac{3}{4}\frac{N}{4s^2 M}\right) \left[0.11 - 0.33\frac{e^{-\frac{3}{4}R_2^2}}{R_2}\right] - 5\zeta_N - 4e^{-\frac{y^2}{2}}. \tag{109}$$

The final bound on $\mathcal{M}_2'$ is the following

$$\mathcal{M}_2' \geq 0.05\sqrt{\frac{N}{4s^2 M}} - \mathcal{O}\left(\left(\frac{N}{M}\right)^{\frac{3}{2}}\right) - \mathcal{O}(\zeta_N) - \mathcal{O}(e^{-\frac{y^2}{2}}) - \mathcal{O}\left(\frac{e^{-\frac{R_2^2}{2}}}{R_2}\sqrt{\frac{N}{M}}\right). \tag{110}$$

Considering the bound on y from Eq. (94), the overall average for the case when $M = \mathcal{O}\left(\frac{N}{\zeta_N^2}\right)$ and $M = \Omega(N)$, is

$$|\mu| \leq \frac{\sqrt{\frac{N}{2s^2 M}}\Lambda_1 + \mathcal{O}(\zeta_N\Lambda_1)}{0.05\sqrt{\frac{N}{4s^2 M}} - \mathcal{O}\left(\left(\frac{N}{M}\right)^{\frac{3}{2}}\right) - \mathcal{O}(\zeta_N) - \mathcal{O}\left(\frac{e^{-\frac{R_2^2}{2}}}{R_2}\sqrt{\frac{N}{M}}\right)}, \tag{111}$$

$$|\mu| \leq 20\sqrt{2}\Lambda_1 \left[1 + \mathcal{O}\left(\sqrt{\frac{M}{N}}\zeta_N\right) + \mathcal{O}\left(\frac{N}{M}\right) + \mathcal{O}\left(\frac{e^{-\frac{R_2^2}{2}}}{R_2}\right)\right], \tag{112}$$

$$|\mu| \leq 30\Lambda_1 \left[1 + \mathcal{O}\left(\frac{N}{M}\right)\right] \Rightarrow |\mu| = \mathcal{O}\left(\frac{N}{\sqrt{M}}\right). \tag{113}$$

B.2 Upper bound on variance

We now find an upper bound on the variance δ^2 with a similar method. First, consider

$$\delta^2 = \langle p_{M,y}| H^2 |p_{M,y}\rangle - \langle p_{M,y}| H |p_{M,y}\rangle^2 \leq \langle p_{M,y}| H^2 |p_{M,y}\rangle \tag{114}$$

$$\leq \frac{\sum\limits_{E_j \leq 0} |b_j|^2 g_y^2\left(\frac{E_j}{N}\right)(E_j)^2 + \sum\limits_{E_j > 0} |b_j|^2 g_y^2\left(\frac{E_j}{N}\right)(E_j)^2}{\sum\limits_{E_j \leq 0} |b_j|^2 g_y^2\left(\frac{E_j}{N}\right) + \sum\limits_{E_j > 0} |b_j|^2 g_y^2\left(\frac{E_j}{N}\right)} := \frac{\mathcal{V}_{1(-)} + \mathcal{V}_{1(+)}}{\mathcal{V}_{2(-)} + \mathcal{V}_{2(+)}}. \tag{115}$$

We begin by bounding the positive energy contributions, $\mathcal{V}_{1(+)}$ and $\mathcal{V}_{2(+)}$. The corresponding negative energy contributions follow directly, because the functions inside the sum are even. Using the inequalities from Eq. (53), we can rewrite the expression for $\mathcal{V}_{1(+)}$ and $\mathcal{V}_{2(+)}$ in terms of cosine power operators

$$\mathcal{V}_{1(+)} \leq \sum_{E_j > 0} |b_j|^2 \cos^{2M}\left(\frac{E_j}{N}\right)(E_j)^2 + 4e^{-y^2} \sum_{E_j > 0} |b_j|^2 (E_j)^2 \tag{116}$$

$$\leq \sum_{E_j > 0} |b_j|^2 \cos^{2M}\left(\frac{E_j}{N}\right)(E_j)^2 + 4e^{-y^2}\sigma^2 =: \mathcal{I}_1, \tag{117}$$

where for the second term we use that $\sum_{E_j > 0} |b_j|^2 (E_j)^2 \leq \sigma^2$. Similarly, the denominator can be lower bounded as

$$\mathcal{V}_{2(+)} \leq \sum_{E_j > 0} |b_j|^2 \cos^{2M}\left(\frac{E_j}{N}\right) - 4e^{-\frac{y^2}{2}} \sum_{E_j > 0} |b_j|^2 \tag{118}$$

$$\leq \sum_{E_j > 0} |b_j|^2 \cos^{2M}\left(\frac{E_j}{N}\right) - 4e^{-\frac{y^2}{2}} =: \mathcal{I}_2 \tag{119}$$

where the second inequality is obtained by upper bounding the cumulative distribution of $|b_j|^2$ by 1. First we focus on simplifying $\mathcal{I}_1$. We again apply the segmentation method, in a similar manner as for the average in Eq. (59). Dividing the total energy range $[-N, N]$ into R_1 segments of width $\Lambda_1 > 0$ each, the first term of $\mathcal{I}_1$ is then bounded as

$$\sum_{E_j > 0} |b_j|^2 \cos^{2M}\left(\frac{E_j}{N}\right)(E_j)^2 \leq \sum_{l=0}^{R_1-1} \cos^{2M}\left(\frac{\Lambda_1 l}{N}\right)(\Lambda_1(l+1))^2 \sum_{\Lambda_1 l < E_j < \Lambda_1(l+1)} |b_j|^2, \tag{120}$$

$$\leq \sum_{l=0}^{R_1-1} \cos^{2M}\left(\frac{\Lambda_1 l}{N}\right)(\Lambda_1(l+1))^2 \left[\int\limits_{\Lambda_1 l}^{\Lambda_1(l+1)} \frac{dt}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2\sigma^2}} + \zeta_N\right]. \tag{121}$$

Let $\tilde{t} = \frac{t}{\sqrt{2}\sigma}, d\tilde{t} = \frac{dt}{\sqrt{2}\sigma}$, and $\tilde{\Lambda}_1 = \frac{\Lambda_1}{\sqrt{2}\sigma}$, then LHS is bounded as

$$\sum_{E_j > 0} |b_j|^2 \cos^{2M}\left(\frac{E_j}{N}\right)(E_j)^2 \leq \frac{1}{\sqrt{\pi}} \sum_{l=0}^{R_1-1} \int\limits_{\tilde{\Lambda}_1 l}^{\tilde{\Lambda}_1(l+1)} d\tilde{t} \cos^{2M}\left(\sqrt{2}\sigma\left(\frac{\tilde{t} - \tilde{\Lambda}_1}{N}\right)\right)(\sqrt{2}\sigma(\tilde{t} + \tilde{\Lambda}_1))^2 e^{-\tilde{t}^2}$$

$$+ 2\sigma^2 \zeta_N \sum_{l=0}^{R_1-1} \cos^{2M}\left(\frac{\sqrt{2}\sigma\tilde{\Lambda}_1 l}{N}\right)(\tilde{\Lambda}_1(l+1))^2. \tag{122}$$

We get the following upper bound on $\mathcal{I}_1$,

$$\mathcal{I}_1 \leq \frac{1}{\sqrt{\pi}} \int\limits_0^{\tilde{\Lambda}_1 R_1} d\tilde{t} \cos^{2M}\left(\sqrt{2}\sigma\left(\frac{\tilde{t} - \tilde{\Lambda}_1}{N}\right)\right)(\sqrt{2}\sigma(\tilde{t} + \tilde{\Lambda}_1))^2 e^{-\tilde{t}^2}$$

$$+ 2\sigma^2 \zeta_N \sum_{l=0}^{R_1-1} \cos^{2M}\left(\frac{\sqrt{2}\sigma\tilde{\Lambda}_1 l}{N}\right)(\tilde{\Lambda}_1(l+1))^2 + 4e^{-y^2}\sigma^2. \tag{123}$$

Using inequalities from Eq. ([77](#)) to replace cosines by exponentials

$$
\mathcal{I}_1 \leq \frac{1}{\sqrt{\pi}} \int_0^{\tilde{\Lambda}_1 R_1} d\tilde{t}\, e^{-2M\sigma^2(\frac{\tilde{t}-\tilde{\Lambda}_1}{N})^2} (\sqrt{2}\sigma(\tilde{t}+\tilde{\Lambda}_1))^2 e^{-\tilde{t}^2} + 2\sigma^2 \zeta_N \sum_{l=0}^{R_1-1} e^{-4M\sigma^2 \frac{\tilde{\Lambda}_1^2 l^2}{N^2}} (\tilde{\Lambda}_1(l+1))^2 + 4e^{-y^2}\sigma^2 .
$$

(124)

Now consider $\sigma = s\sqrt{N}$ so that

$$
\mathcal{I}_1 \leq \frac{2s^2 N}{\sqrt{\pi}} \int_0^{\tilde{\Lambda}_1 R_1} d\tilde{t}\, e^{-\frac{2s^2 M}{N}(\tilde{t}-\tilde{\Lambda}_1)^2} (\tilde{t}+\tilde{\Lambda}_1)^2 e^{-\tilde{t}^2} + 2s^2 N \zeta_N \sum_{l=0}^{R_1-1} e^{-\frac{4s^2 M}{N}\tilde{\Lambda}_1^2 l^2} (\tilde{\Lambda}_1(l+1))^2 + 4s^2 N e^{-y^2} .
$$

(125)

Using the Euler-Maclaurin formula from equation ([81](#)), the integral in the first term is bounded as

$$
\begin{aligned}
\frac{2s^2 N}{\sqrt{\pi}} \int_0^{\tilde{\Lambda}_1 R_1} d\tilde{t}\, e^{-\frac{2s^2 M}{N}(\tilde{t}-\tilde{\Lambda}_1)^2} (\tilde{t}+\tilde{\Lambda}_1)^2 e^{-\tilde{t}^2} \leq & \frac{2s^2 N}{\sqrt{\pi}} \frac{e^{-\frac{2s^2 M}{N}\tilde{\Lambda}_1^2}}{4(1+\frac{2s^2 M}{N})^{\frac{5}{2}}} \Bigg[4\tilde{\Lambda}_1 \sqrt{1+\frac{2s^2 M}{N}} \left(1+\frac{3s^2 M}{N}\right) + \\
& -2e^{-\tilde{\Lambda}_1^2 R_1^2(1+\frac{2s^2 M}{N})} e^{\tilde{\Lambda}_1^2 R_1 \frac{4s^2 M}{N}} \tilde{\Lambda}_1 \sqrt{1+\frac{2s^2 M}{N}} \left(R_1\left(1+\frac{2s^2 M}{N}\right) + 2\left(1+\frac{3s^2 M}{N}\right)\right) + \\
& +e^{\frac{1}{1+\frac{2s^2 M}{N}}(\frac{2s^2 M}{N}\tilde{\Lambda}_1)^2} \sqrt{\pi} \left(\left(1+\frac{2s^2 M}{N}\right) + 2\tilde{\Lambda}_1^2\left(1+\frac{4s^2 M}{N}\right)^2\right) \Bigg[\mathrm{erf}\left(\frac{\frac{2s^2 M}{N}\tilde{\Lambda}_1}{\sqrt{1+\frac{2s^2 M}{N}}}\right) + \\
& + \mathrm{erf}\left(\frac{\tilde{\Lambda}_1 R_1(1+\frac{2s^2 M}{N}) - \frac{2s^2 M}{N}\tilde{\Lambda}_1}{\sqrt{1+\frac{2s^2 M}{N}}}\right) \Bigg] \Bigg],
\end{aligned}
$$

(126)

$$
\begin{aligned}
\frac{2s^2 N}{\sqrt{\pi}} \int_0^{\tilde{\Lambda}_1 R_1} d\tilde{t}\, e^{-\frac{2s^2 M}{N}(\tilde{t}-\tilde{\Lambda}_1)^2} (\tilde{t}+\tilde{\Lambda}_1)^2 e^{-\tilde{t}^2} \leq & \frac{2s^2 N}{\sqrt{\pi}} \frac{e^{-\frac{2s^2 M}{N}\tilde{\Lambda}_1^2}}{4(1+\frac{2s^2 M}{N})^{\frac{5}{2}}} \Bigg[4\tilde{\Lambda}_1 \sqrt{1+\frac{2s^2 M}{N}} \left(1+\frac{3s^2 M}{N}\right) \\
& - 2e^{-\tilde{\Lambda}_1^2 R_1^2(1+\frac{2s^2 M}{N})} e^{\tilde{\Lambda}_1^2 R_1 \frac{4s^2 M}{N}} \tilde{\Lambda}_1 R_1 \left(1+\frac{2s^2 M}{N}\right)^{\frac{3}{2}} \\
& + 2e^{\frac{1}{1+\frac{2s^2 M}{N}}(\frac{2s^2 M}{N}\tilde{\Lambda}_1)^2} \sqrt{\pi} \left(\left(1+\frac{2s^2 M}{N}\right) \right. \\
& \left. + 2\tilde{\Lambda}_1^2\left(1+\frac{4s^2 M}{N}\right)^2\right) \Bigg] .
\end{aligned}
$$

(127)

Substituting $C_1 = \frac{2s^2 M}{N}\tilde{\Lambda}_1^2$

$$
\frac{2s^2 N}{\sqrt{\pi}} \int_0^{\tilde{\Lambda}_1 R_1} d\tilde{t}\, e^{-\frac{2s^2 M}{N}(\tilde{t}-\tilde{\Lambda}_1)^2} (\tilde{t}+\tilde{\Lambda}_1)^2 e^{-\tilde{t}^2} \leq \frac{3s^2 N}{\sqrt{\pi}} \frac{\tilde{\Lambda}_1^3 e^{-C_1}}{\tilde{\Lambda}_1^2 + C_1} + s^2 N \frac{\tilde{\Lambda}_1^3 e^{-\frac{\tilde{\Lambda}_1^2 C_1}{\tilde{\Lambda}_1^2 + C_1}}}{(\tilde{\Lambda}_1^2 + C_1)^{\frac{3}{2}}} + 8s^2 N \frac{\tilde{\Lambda}_1^3 e^{-\frac{\tilde{\Lambda}_1^2 C_1}{\tilde{\Lambda}_1^2 + C_1}}}{\sqrt{\tilde{\Lambda}_1^2 + C_1}} .
$$

(128)

The error term of Eq. (125) is bounded as

$$2s^2 N \zeta_N \sum_{l=0}^{R_1-1} e^{-\frac{4s^2 M}{N}\tilde{\Lambda}_1^2 l^2} (\tilde{\Lambda}_1(l+1))^2 \leq \frac{s^2(8\sqrt{C_1} + \sqrt{2\pi}(1+4C_1))}{8C_1\sqrt{C_1}} N\zeta_N \tilde{\Lambda}_1^2$$
$$+ s^2 N \zeta_N \tilde{\Lambda}_1^2 \left(e^{-2C_1(R_1-1)^2}R_1^2 + 1\right)$$
$$+ \frac{s^2 N \zeta_N}{6}(2e^{-\frac{4s^2 M}{N}\tilde{\Lambda}_1^2(R_1-1)^2}\tilde{\Lambda}_1^2 R_1)$$
$$+ 2s^2 N \zeta_N \tilde{\Lambda}_1^2 \frac{\zeta(2)}{\pi^2}\left[5 + \sqrt{\frac{\pi}{8C_1}}(5+4C_1)\right]. \tag{129}$$

Combining Eq. (128) and Eq. (129), we get the following bound on $\mathcal{I}_1$,

$$\mathcal{I}_1 \leq \frac{3s^2 N}{\sqrt{\pi}} \frac{\tilde{\Lambda}_1^3 e^{-C_1}}{\tilde{\Lambda}_1^2 + C_1} + s^2 N \frac{\tilde{\Lambda}_1^3 e^{-\frac{\tilde{\Lambda}_1^2 C_1}{\tilde{\Lambda}_1^2 + C_1}}}{(\tilde{\Lambda}_1^2 + C_1)^{\frac{3}{2}}} + 8s^2 N \frac{\tilde{\Lambda}_1^3 e^{-\frac{\tilde{\Lambda}_1^2 C_1}{\tilde{\Lambda}_1^2 + C_1}}}{\sqrt{\tilde{\Lambda}_1^2 + C_1}} +$$
$$+ \frac{s^2(8\sqrt{C_1} + \sqrt{2\pi}(1+4C_1))}{8C_1\sqrt{C_1}} N\zeta_N \tilde{\Lambda}_1^2 + s^2 N \zeta_N \tilde{\Lambda}_1^2 \left(e^{-2C_1(R_1-1)^2}R_1^2 + 1\right) +$$
$$+ \frac{s^2 N \zeta_N}{6}(2e^{-\frac{4s^2 M}{N}\tilde{\Lambda}_1^2(R_1-1)^2}\tilde{\Lambda}_1^2 R_1) + 2s^2 N \zeta_N \tilde{\Lambda}_1^2 \frac{\zeta(2)}{\pi^2}\left[5 + \sqrt{\frac{\pi}{8C_1}}(5+4C_1)\right] + 4s^2 N e^{-y^2}. \tag{130}$$

To simplify the above form note that

$$e^{-\frac{\tilde{\Lambda}_1^2 C_1}{\tilde{\Lambda}_1^2 + C_1}} \leq 1 \qquad \text{and} \qquad \frac{\tilde{\Lambda}_1}{\sqrt{\tilde{\Lambda}_1^2 + C_1}} = \sqrt{\frac{1}{1 + \frac{2s^2 M}{N}}} \leq 1, \tag{131}$$

for any scaling of M. Additionally, for $\frac{N}{2s^2 M} < 1$ there is the tighter bound

$$\frac{\tilde{\Lambda}_1}{\sqrt{\tilde{\Lambda}_1^2 + C_1}} = \sqrt{\frac{1}{1 + \frac{2s^2 M}{N}}} \leq \sqrt{\frac{N}{2s^2 M}}. \tag{132}$$

So, for $\frac{N}{2s^2 M} < 1$ the numerator is bounded as

$$\mathcal{I}_1 \leq \frac{3s^2 N}{\sqrt{\pi}} \frac{\tilde{\Lambda}_1^3}{\tilde{\Lambda}_1^2 + C_1} + s^2 N \frac{\tilde{\Lambda}_1^3}{(\tilde{\Lambda}_1^2 + C_1)^{\frac{3}{2}}} + 8s^2 N \frac{\tilde{\Lambda}_1^3}{\sqrt{\tilde{\Lambda}_1^2 + C_1}} +$$
$$+ \frac{s^2(8\sqrt{C_1} + \sqrt{2\pi}(1+4C_1))}{8C_1\sqrt{C_1}} N\zeta_N \tilde{\Lambda}_1^2 + s^2 N \zeta_N \tilde{\Lambda}_1^2 \left(e^{-2C_1(R_1-1)^2}R_1^2 + 1\right) +$$
$$+ \frac{s^2 N \zeta_N}{6}(2e^{-\frac{4s^2 M}{N}\tilde{\Lambda}_1^2(R_1-1)^2}\tilde{\Lambda}_1^2 R_1) + 2s^2 N \zeta_N \tilde{\Lambda}_1^2 \frac{\zeta(2)}{\pi^2}\left[5 + \sqrt{\frac{\pi}{8C_1}}(5+4C_1)\right] + 4s^2 N e^{-y^2}. \tag{133}$$

Now applying the following inequalities to the first three terms

$$\frac{\tilde{\Lambda}_1^3}{\tilde{\Lambda}_1^2 + C_1} \leq \frac{N}{2s^2 M}\tilde{\Lambda}_1, \tag{134}$$

$$\frac{\tilde{\Lambda}_1^3}{(\tilde{\Lambda}_1^2 + C_1)^{\frac{3}{2}}} \leq \left(\frac{N}{2s^2 M}\right)^{\frac{3}{2}}, \tag{135}$$

$$\frac{\tilde{\Lambda}_1^3}{\sqrt{\tilde{\Lambda}_1^2 + C_1}} \leq \tilde{\Lambda}_1^2 \sqrt{\frac{N}{2s^2 M}}. \tag{136}$$

Consider the following,

    (i) $C_1 = \mathcal{O}(1)$ which means that $\tilde{\Lambda}_1 = \mathcal{O}(1)$ for $M = \Omega(N)$,

    (ii) $y = \Omega\left(\sqrt{\log \frac{M^3}{N^2}}\right)$ from Eq. (94),

    (iii) $\tilde{\Lambda}_1 R_1 \propto \sqrt{N}$, then

$$\mathcal{I}_1 \leq \frac{3}{2\sqrt{\pi}}\tilde{\Lambda}_1\frac{N^2}{M} + \frac{1}{2\sqrt{2}s}\frac{N^{\frac{5}{2}}}{M^{\frac{3}{2}}} + 4\sqrt{2}s\tilde{\Lambda}_1^2\frac{N^{\frac{3}{2}}}{M^{\frac{1}{2}}} + \mathcal{O}(N\zeta_N\tilde{\Lambda}_1^2) + \mathcal{O}(N^2\zeta_N e^{-2C_1(R_1-1)^2}). \qquad (137)$$

Substituting $\tilde{\Lambda}_1 = \sqrt{C_1}\sqrt{\frac{N}{2s^2 M}}$,

$$\mathcal{I}_1 \leq \left(\frac{3\sqrt{C_1}}{2s\sqrt{2\pi}} + \frac{1}{2\sqrt{2}s} + \frac{2\sqrt{2}C_1}{s}\right)\frac{N^{\frac{5}{2}}}{M^{\frac{3}{2}}} + \mathcal{O}\left(\frac{N^2}{M}\zeta_N\right) + \mathcal{O}(N^2\zeta_N e^{-2C_1(R_1-1)^2}). \qquad (138)$$

We have obtained an upper bound on $\mathcal{I}_1$ which upper bounds $\mathcal{V}_{1(+)}$ and $\mathcal{V}_{1(-)}$ (numerator of Eq. (115)). For the denominator, note that $\mathcal{M}'_2$ in Eq. (110) from the average calculation is also a lower bound on the denominator terms $\mathcal{V}_{2(+)}$ and $\mathcal{V}_{2(-)}$ of the variance. Combining all the expressions, we can obtain the upper bound on the variance for $M = \mathcal{O}\left(\frac{N}{\zeta_N^2}\right)$ and $M = \Omega(N)$ as follows,

$$\delta^2 \leq \frac{\left(\frac{3\sqrt{C_1}}{2s\sqrt{2\pi}} + \frac{1}{2\sqrt{2}s} + \frac{2\sqrt{2}C_1}{s}\right)\frac{N^{\frac{5}{2}}}{M^{\frac{3}{2}}}\left[1 + \mathcal{O}\left(\sqrt{\frac{M}{N}}\zeta_N\right) + \mathcal{O}\left(\frac{M^{\frac{3}{2}}}{N^{\frac{1}{2}}}\zeta_N e^{-2C_1(R_1-1)^2}\right) + \mathcal{O}\left(\frac{M^{\frac{3}{2}}}{N^{\frac{3}{2}}}\zeta_N R_1 e^{-\frac{2s^2 M}{N}R_1}\right)\right]}{0.1\sqrt{\frac{N}{4s^2 M}}\left[1 - \mathcal{O}\left(\frac{N}{M}\right) - \mathcal{O}\left(\zeta_N\sqrt{\frac{M}{N}}\right) - \mathcal{O}\left(\frac{e^{-\frac{R_2^2}{2}}}{R_2}\right)\right]}, \qquad (139)$$

$$\Rightarrow \delta^2 \leq \eta\frac{N^2}{M}\left[1 + \mathcal{O}\left(\frac{N}{M}\right)\right], \qquad (140)$$

where we define the constant

$$\eta := \frac{15\sqrt{C_1}}{\sqrt{2\pi}} + \frac{5}{\sqrt{2}} + 20\sqrt{2}C_1. \qquad (141)$$