Optimization of Richardson extrapolation for quantum error mitigation

Michael Krebsbach [1,*] Björn Trauzettel [1,2] and Alessio Calzona [1,2]

[1]*Institute for Theoretical Physics and Astrophysics, University of Würzburg, 97074 Würzburg, Germany*
[2]*Würzburg-Dresden Cluster of Excellence ct.qmat, Würzburg, 01069 Dresden, Germany*

Quantum error mitigation is a key concept for the development of practical applications based on current noisy intermediate-scale quantum devices. One of the most promising methods is Richardson extrapolation to the zero noise limit. While its main idea is rather simple, the full potential of Richardson extrapolation has not been completely uncovered yet. We give an in-depth analysis of the relevant parameters of Richardson extrapolation and propose an optimized protocol for its implementation. This protocol allows for a precise control of the increase in statistical uncertainty and lays the foundation for a significant improvement of the mitigation performance achieved by increasing the number of nodes. Furthermore, we present a set of nodes that, on average, outperforms the linear, exponential, or Chebyshev nodes frequently used for Richardson extrapolation without requiring any additional resources.

I. INTRODUCTION

One of the main challenges for the development of quantum computers (QCs) is their high sensitivity to errors and noise. While fault tolerance can be theoretically achieved via quantum error correction (QEC) [1–5], its practical implementation comes with a large hardware overhead [6–8] that is out of reach for near-term quantum processors. For the time being, a temporary but precious alternative to QEC is represented by quantum error mitigation (QEM), a collection of techniques which can reduce the impact of errors at the cost of performing extra measurements and postprocessing of the data [9–11]. Differently from QEC, QEM has been successfully implemented in numerous experiments [12–15] and represents a key resource for the realization of practical applications in the present noisy intermediate-scale quantum era [16].

To date, a wide spectrum of QEM schemes have been proposed, each one with its own advantages and tradeoffs. Importantly, different techniques can be efficiently combined in order to achieve an even better performance [17–19]. Error mitigation schemes include probabilistic error cancellation [9,11], subspace expansion [20], symmetry verification [21–23], purification [24,25], and zero noise extrapolation [9–11]. The latter is one of the most popular approaches and relies on the possibility to tune the noise level λ of the quantum circuit above the smallest achievable value λ_0. By repeatedly running the circuit at different boosted noise strengths $\lambda_i \geqslant \lambda_0$, we can thus sample the noisy expectation value of a given observable as a function of λ and then try to estimate its zero noise limit. This operation can be done either by fitting the collected data points with a specific function, typically a single- or multiexponential [11,17], or by per-

forming a Richardson-like extrapolation [9,26]. The second approach has the advantage that it does not require specific assumptions on the λ dependence of the expectation values and can be used on a variety of noise models. Its effectiveness, however, depends nontrivially on the choice of the different boosted noise strengths λ_i [27,28].

In this paper, we perform a careful and in-depth analysis of the Richardson method for QEM, highlighting its most important parameters, and propose an optimized implementation. With previous methods, the set of noise levels has been determined by a trial and error procedure that terminates once an acceptably small statistical uncertainty is reached. Since for a fixed equidistant spacing of the noise levels the variance grows exponentially with their number [29], this was only feasible for no more than four noise levels, strongly limiting the mitigation capabilities of Richardson extrapolation. We resolve this issue by demonstrating that the statistical uncertainty can in fact be precisely controlled independently of the number of noise levels. This allows us to significantly improve the mitigation performance, especially for a large sampling budget. Moreover, we find a specific spacing of the noise levels λ_i (inspired by the Chebyshev nodes introduced in Ref. [30]) that minimizes, on average, the error on the zero noise estimation. These discoveries nicely complement the current literature, which mainly focuses on the linear spacing of the λ_i and a small number of noise levels [9,18]. We numerically verify our predictions on different error models, ranging from purely Markovian noise to non-Markovian noise, which can play a significant role in present-day quantum devices [31–33].

The paper is organized as follows. In Sec. II, we briefly review the Richardson approach to QEM. In Sec. III, we carefully discuss the three main parameters which characterize the extrapolation scheme and propose our optimized protocol, which we numerically validate in Sec. IV. We draw conclusions in Sec. V and move some technical details to the Appendices.

_____
*mkrebsbach@web.de

II. RICHARDSON EXTRAPOLATION

Assume that ρ_λ is the output state of an imperfect QC, where λ is the dimensionless noise parameter that indicates the level of noise in the QC. For this state the expectation value of an observable A is given by

$$E_\lambda = \text{tr}(\rho_\lambda A). \tag{1}$$

Since the QC is imperfect, the noise level λ cannot be decreased below some finite value $\lambda_0 > 0$ and the limit

$$\lim_{\lambda \to 0} E_\lambda = E^* \tag{2}$$

cannot be accessed experimentally. However, the noise parameter λ can be artificially increased, e.g., by rescaling coupling parameters [9], inserting identities into the circuit that are subject to the same noise [29,34–36], or gate Trotterization [37]. Finding the best way to change the noise levels is an active research area [37] and depends on the precise implementation of the quantum computer. In the following, we assume that we can arbitrarily choose the noise parameter $\lambda = \lambda_0 x$ where $x \in [1, \infty)$.

The idea behind Richardson extrapolation is to sample $E_{\lambda_0}(x) := E_\lambda$ at $n + 1$ distinct *nodes* $x_0, \ldots, x_n \in [1, \infty)$ and to find the unique n polynomial that fits these samples. This polynomial approximation of $E_{\lambda_0}(x)$ can be evaluated at $x = 0$, which gives an estimate R_n of the noiseless expectation value E^*. For simplicity, we assume that $1 = x_0 < x_1 < \cdots < x_n$.

This result of the extrapolation R_n can be directly calculated as a linear combination of the samples $E_{\lambda_0}(x_j)$:

$$R_n = \sum_{j=0}^{n} E_{\lambda_0}(x_j)\gamma_j, \tag{3}$$

where the factors

$$\gamma_j = \prod_{k \neq j} \frac{x_k}{x_k - x_j} \tag{4}$$

are the Lagrange polynomials at $x = 0$. They have the property that $\sum_{j=0}^{n} \gamma_j = 1$ and $\sum_{j=0}^{n} \gamma_j x_j^k = 0$ for $k = 1, \ldots, n$. If we write $E_{\lambda_0}(x)$ as

$$E_{\lambda_0}(x) = E^* + \sum_{k=1}^{\infty} a_k \lambda_0^k x^k \tag{5}$$

for some constant coefficients a_k, then the mitigated result reads

$$R_n = E^* + \sum_{k=n+1}^{\infty} a_k \lambda_0^k \sum_{j=0}^{n} \gamma_j x_j^k. \tag{6}$$

Compared to the result without mitigation $R_0 = E_{\lambda_0}(1)$, which corresponds to $n = 0$, Richardson extrapolation eliminates $a_1, \ldots, a_n$ as well as the first n orders of λ_0. Therefore, if λ_0 is sufficiently small then Richardson extrapolation is a promising candidate for QEM.

A general formula for the bias $\text{Bias}[\hat{R}_n] = R_n - E^*$ of this estimate comes from the theory of polynomial approximation. If $E_{\lambda_0}(x)$ is $n + 1$ times continuously differentiable on $[0, x_n]$

then there exists a $\xi \in [0, x_n]$ such that

$$\text{Bias}[\hat{R}_n] = (-1)^n E_{\lambda_0}^{(n+1)}(\xi) \frac{C_n}{(n+1)!}. \tag{7}$$

$E_{\lambda_0}^{(n+1)}(x)$ denotes the $(n + 1)$th derivative of $E_{\lambda_0}(x)$ and $C_n = \prod_{j=0}^{n} x_j$ (see, e.g., Refs. [30,38,39]).

Besides the error due to noise, additional sampling errors must be considered whenever a quantum-mechanical measurement is conducted. If for the noise level $\lambda_0 x_j$ the observable A is measured N_j times, then the sample mean $\hat{E}_{\lambda_0}(x_j)$ has a variance of

$$\text{Var}[\hat{E}_{\lambda_0}(x_j)] = \frac{\sigma_j^2}{N_j} \tag{8}$$

where $\sigma_j = \sqrt{\text{tr}(\rho_{\lambda_0 x_j} A^2) - E_{\lambda_0}(x_j)^2}$. The σ_j's depend on the observable A as well as on the expected value $E_{\lambda_0}(x_j)$ and the density matrix $\rho_{\lambda_0 x_j}$ of the quantum state. They are unknown quantities before the experiment has been conducted, but it is reasonable to assume the σ_j to be of the same order of magnitude. For simplicity, we assume that $\sigma_j = \sigma = \text{const}$ for all j in the following.

Since the result of Richardson extrapolation R_n is a linear combination of all $E_{\lambda_0}(x_j)$, the corresponding estimator $\hat{R}_n$ is thus also affected by sampling errors:

$$\text{Var}[\hat{R}_n] = \sum_{j=0}^{n} \gamma_j^2 \frac{\sigma^2}{N_j}. \tag{9}$$

As we will see later, this leads to an increase in variance compared to the unmitigated result with the same number of measurements $\text{Var}[\hat{R}_0] = \frac{\sigma^2}{\sum_{j=0}^{n} N_j}$.

In conclusion, Richardson extrapolation is able to reduce the bias of the estimate of E^*, however, only at the cost of increasing its variance. An example that illustrates these effects is shown in Fig. 1.

If we choose a linear spacing of the nodes

$$x_j = 1 + j, \tag{10}$$

then $\text{Var}[\hat{R}_n]$ grows exponentially with n in the limit $n \to \infty$ [29]. This poses the following questions.

(1) Is there a way to reduce this increase in $\text{Var}[\hat{R}_n]$?

(2) Is there a scaling method for x_j that optimizes the tradeoff between bias and variance?

(3) Is there an optimal n?

In the following section, we investigate these questions and how to efficiently execute this mitigation technique, i.e., how to choose the free parameters n, $\{N_j\}_{j=0,\ldots,n}$ and $\{x_j\}_{j=0,\ldots,n}$. For a fixed variance we search for those parameters that, given the same computational resources, minimize the bias of the estimator $\hat{R}_n$.

III. OPTIMIZED QEM PROTOCOL

In the literature, it has been noted that the numbers of measurements N_j for a given noise level $\lambda_0 x_j$ should be adjusted according to their weights γ_j in Eq. (9) [18,28]. This is also
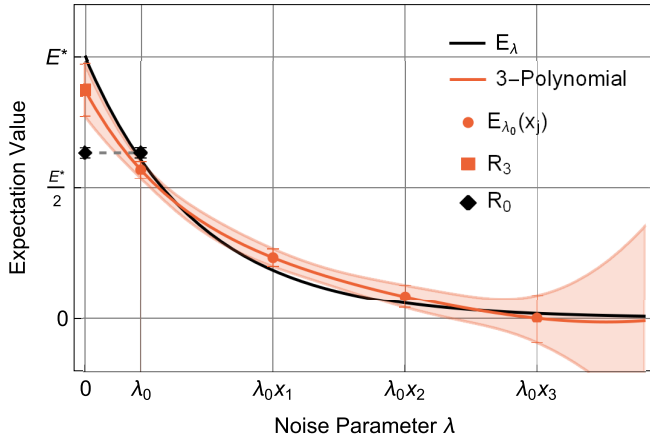
FIG. 1. Example for Richardson extrapolation with $n = 3$ for a noisy expectation value which decays exponentially $E_\lambda = E^* e^{-\lambda}$ and has units according to E^*. The black diamond is the result without mitigation R_0 and the red square is the mitigated result R_3, both obtained with the same sampling budget. The colored interval indicates the standard deviation of the extrapolation using Lagrange polynomials. The error bars are the standard deviations of the samples $E_{\lambda_0}(x_j)$ adjusted according to Eqs. (8) and (11). The nodes lie equidistant.

the starting point of our analysis in Sec. III A. After that, in Sec. III B, we turn to the spacing of the nodes x_j. Different strategies have been considered before, including equidistant and exponential nodes as well as nodes based on the extrema of the Chebyshev polynomials of the first kind. We compare these spacing methods to a new set of nodes which we show to be the optimal ansatz for QEM. In Sec. III C, we then focus on the number of nodes n. We show that, for certain noise functions and a proper choice of the nodes, increasing n can substantially improve mitigation. The complete procedure is summarized in Sec. III D.

A. Number of measurements N_j and measurement overhead Λ

A first simple step for the optimization of Richardson extrapolation is to reduce $\text{Var}[\hat{R}_n]$ by adjusting N_j while keeping the total sampling budget $\sum_{j=0}^{n} N_j = N_{\text{tot}}$ constant [18,28]. Whenever noninteger values for N_j appear in the following they should be rounded to the next nearest integer. A simple calculation shows that $\text{Var}[\hat{R}_n]$ is minimized by

$$N_j = N_{\text{tot}} \frac{|\gamma_j|}{\sum_{k=0}^{n} |\gamma_k|}. \tag{11}$$

Defining $\Lambda = \sum_{j=0}^{n} |\gamma_j|$ and $N_{\text{eff}} = N_{\text{tot}}/\Lambda^2$ the minimized variance of $\hat{R}_n$ can be written as

$$\text{Var}[\hat{R}_n] = \frac{\sigma^2}{N_{\text{tot}}} \Lambda^2 = \frac{\sigma^2}{N_{\text{eff}}}. \tag{12}$$

This way the estimator $\hat{R}_n$ that uses N_{tot} measurements has the same variance as the sample mean of a single expectation value that is measured N_{eff} times. $\Lambda^2 \geqslant 1$ is therefore the minimal overhead in measurements required to compensate for Richardson extrapolation. An example for this procedure is given in Fig. 1, where the standard deviation of

each sample $E_{\lambda_0}(x_j)$ is adjusted according to Eq. (11). In this case, the nodes are chosen such that $\Lambda = 5$.

The quantities N_{tot} and N_{eff} can be chosen freely at the beginning of an experiment. The first one is the sampling budget of the experiment while the second determines the variance of the estimator $\hat{R}_n$. The sampling overhead is thus given by $\Lambda^2 = \frac{N_{\text{tot}}}{N_{\text{eff}}}$. As we will see later, this overhead has a major influence on the range of noise that can be mitigated. After that, the nodes $\{x_j\}_{j=0,\dots,n}$ must be chosen in such a way that $\sum_{j=0}^{n} |\gamma_j| = \Lambda$. For each spacing method, the nodes usually only depend on x_1 [compare, e.g., Eqs. (15)–(18)]. It is therefore possible to first choose a spacing method and then numerically solve this equation to get the value of x_1 that leads to the chosen Λ. Importantly, this can be done independently of n. The variance $\text{Var}[\hat{R}_n]$ is therefore independent of the number of nodes n and it is not necessary to restrict considerations to small n.

Typically, a total sampling budget of the order 10^5 to 10^6 [12,40,41] or even up to 10^{10} measurements [19] is considered in the literature. If, for example, we have a budget of $N_{\text{tot}} = 10^6$ measurements and choose $N_{\text{eff}} = 1024$ then the sampling overhead is about $\Lambda^2 = 32^2$.

Now, we can compare different spacing methods in a *fair* way, i.e., for the same sampling budget N_{tot} as well as the same variance σ^2/N_{eff} of the estimator $\hat{R}_n$ of E^*. This way, we can search for those nodes that minimize the bias of $\hat{R}_n$.

B. Spacing of the nodes x_j

Taking a closer look at the bias in Eq. (7) and the variance in Eq. (12), we observe that the quantities that depend on the spacing of the nodes and can be optimized are

$$C_n = \prod_{j=0}^{n} x_j \tag{13}$$

and

$$\Lambda = \sum_{j=0}^{n} |\gamma_j| = \sum_{j=0}^{n} \left| \prod_{k \neq j}^{n} \frac{x_k}{x_k - x_j} \right|. \tag{14}$$

Note that $E_{\lambda_0}^{(n+1)}(\xi)$ depends on the nodes as well. Therefore, if $E_{\lambda_0}(x)$ was known, an optimal set of nodes could be constructed for this specific noise function. However, in this case there would be no need for Richardson extrapolation and we would directly fit $E_{\lambda_0}(x)$ to the samples. Therefore, we assume $E_{\lambda_0}(x)$ to be unknown and try to find the nodes that are intrinsically best suited for Richardson extrapolation independent of the noise function.

If all x_j are small, then C_n and the bias are small; however, at the same time, all x_j lie close together such that Λ and the variance get large. In contrast, if all x_j are far apart, then Λ is close to 1 but C_n gets large. The aim of this section is to determine nodes that find the best compromise between these effects.

In the literature the equidistant (linear, L), exponential (E), and Chebyshev nodes are frequently used. The latter requires a word of caution, as the expression *Chebyshev nodes* or *points* has been used both for the zeros and the extrema of the Chebyshev polynomials of the first kind. While the zeros are
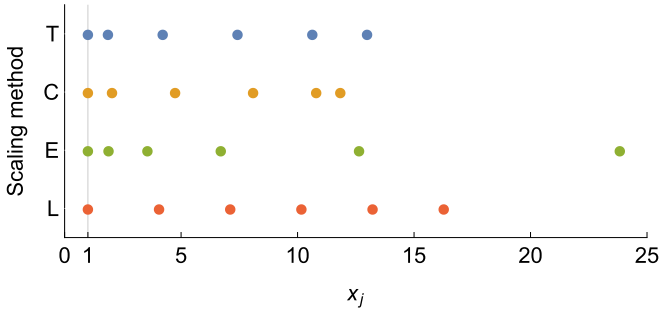
FIG. 2. Comparison of the different spacing methods for $n = 5$ and $\Lambda = 10$. The scaling methods are tilted Chebyshev (T), extremal Chebyshev (C), exponential (E), and equidistant (L).

well suited for *interpolation* [30,42], we are more interested in *extrapolation* and therefore focus on the extrema (C). The different spacing methods read

$$x_j^L = 1 + j(x_1 - 1), \tag{15}$$

$$x_j^E = x_1^j, \tag{16}$$

$$x_j^C = 1 + \frac{\sin^2\left(\frac{j}{n}\frac{\pi}{2}\right)}{\sin^2\left(\frac{1}{n}\frac{\pi}{2}\right)}(x_1 - 1), \tag{17}$$

where $j = 0, \ldots, n$.

Note, that, in contrast to Eq. (10), Eq. (15) has a variable distance between two neighboring nodes. By choosing N_{eff} and adjusting this distance to $\Lambda^2 = \frac{N_{\text{tot}}}{N_{\text{eff}}}$ as discussed in Sec. III A, the problem of variances increasing exponentially with n is resolved.

From Ref. [28], we know that x_j^C minimize x_n for a given Λ. However, they do not minimize the whole product C_n which is why we look at the following optimization problem.

For constant $\Lambda = \sum_{j=0}^{n}|\gamma_j|$ and $x_0 = 1$ find the nodes $x_0, \ldots, x_n$ that minimize C_n. This optimization problem is solved by the nodes

$$x_j^T = 1 + \frac{\sin^2\left(\frac{j}{n+1}\frac{\pi}{2}\right)}{\sin^2\left(\frac{1}{n+1}\frac{\pi}{2}\right)}(x_1 - 1). \tag{18}$$

A proof is outlined in Appendix A.

This result resembles the (extremal) Chebyshev nodes where n is increased by 1 and the last node is left out. This asymmetry makes sense since the goal is to extrapolate *to the left* and precision *to the right* is subordinate. We call the nodes *tilted Chebyshev nodes* (T). Figure 2 provides a visual comparison between the different spacing methods.

As a numerical verification of this result, we plot $\frac{(n+1)!}{C_n}$ for different spacing methods and numbers of nodes over Λ in Fig. 3. The larger this ratio, the smaller is the bias of $\hat{R}_n$ and the better the result of Richardson extrapolation. From this plot, we observe that the tilted Chebyshev nodes indeed lead to a lower C_n and consequently a higher ratio $\frac{(n+1)!}{C_n}$ than the other methods. While for $n \leqslant 3$ all scaling methods perform similarly, the tilted Chebyshev nodes perform best for larger n. For example for $n = 7$ they lead to a C_n that is about 1.25, 2, and 35 times smaller than that of the (extremal) Chebyshev, exponential, and linear nodes, respectively. Remember that
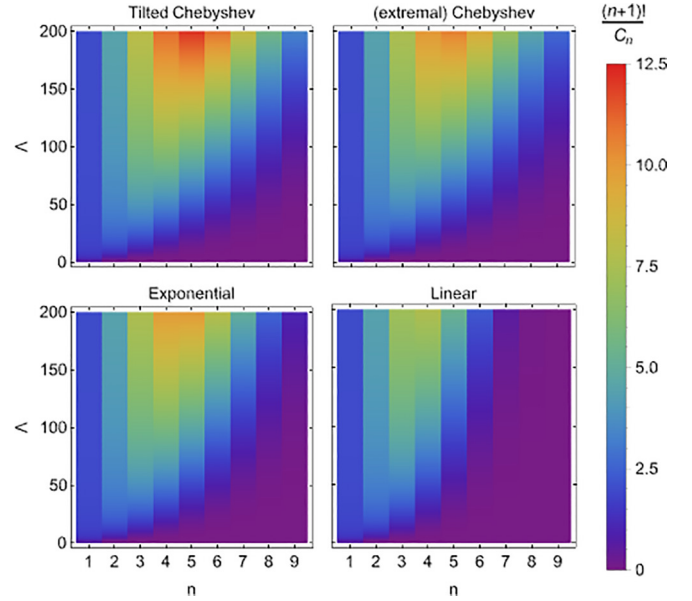

FIG. 3. The ratio $\frac{(n+1)!}{C_n}$ for different values of Λ and n for the spacing methods x_j^T, x_j^C, x_j^E, and x_j^L.

this result is independent of the noise model and the improvement does not require any additional resources.

Since we have determined N_j and x_j, the last remaining parameter of Richardson extrapolation is the number of nodes n.

C. Number of nodes n

While the variance $\text{Var}[\hat{R}_n]$ is independent of the number of nodes, the bias $\text{Bias}[\hat{R}_n]$ is not. It depends on $E_{\lambda_0}^{(n+1)}(\xi)$ as well as on $\frac{C_n}{(n+1)!}$ [see Eq. (7)]. Without knowing the whole function $E_{\lambda_0}(x)$, we cannot predict the derivative $E_{\lambda_0}^{(n+1)}(\xi)$ and thus cannot identify the best value for n.

Useful insight for the choice of n can be gained with the (strong) assumption that all $E_{\lambda_0}^{(n+1)}(\xi)$ are approximately the same for all spacing methods and all n. In this case, the n dependence of the bias is completely determined by $\frac{C_n}{(n+1)!}$. For a given spacing method, this quantity can be analyzed as a function of n and Λ. The results for the discussed spacing methods are shown in Fig. 3, where we plot the ratio $\frac{(n+1)!}{C_n}$.

Given the assumption that $|E_{\lambda_0}^{(n+1)}(\xi)|$ is approximately constant, Fig. 3 could be used to determine the optimal $\hat{n}$ once Λ is chosen. However, different derivatives have different properties in reality and ξ can be very different for different n such that the meaning of this *optimal* $\hat{n}$ should not be overestimated. For example, if $|E_{\lambda_0}^{(n+1)}(\xi)|$ decreases with n then this favors larger $n \geqslant \hat{n}$. While there is no optimal n in general, Fig. 3 might be able to suggest an interval from which we choose n in order to minimize the bias.

In practice, there might be additional factors limiting the choice of n. For example, if Λ is small then n should not be chosen too large; otherwise, x_n can get exceedingly large. Furthermore, this can lead to very low measurement numbers N_j which can be avoided for example by decreasing n or setting a lower bound on N_j. In contrast, if n is too small

while Λ is large then the x_j lie close to each other, which makes the procedure susceptible to errors in the amplification factors x_j.

In the next section we summarize the procedure before turning to numerical simulations of examples.

D. Summary of the procedure

We describe above that the relevant parameters of Richardson extrapolation are N_{eff}, the overhead Λ^2, and the number of nodes n. Since the spacing method cannot always be chosen at will we also include it into this list. We now summarize our protocol for Richardson extrapolation.

(1) Define a variance $\text{Var}[\hat{R}_n] = \sigma^2/N_{\text{eff}}$ that is acceptable and choose N_{eff} accordingly.

(2) Choose Λ^2 depending on the sampling budget ($N_{\text{tot}} = N_{\text{eff}}\Lambda^2$). The larger, the better [43].

(3) Determine the spacing of the nodes. As discussed in Sec. III B, the tilted Chebyshev nodes most likely lead to the best extrapolation result.

(4) Pick a value for n. Given Λ and the spacing method, $\hat{n}$ is the n maximizing $\frac{(n+1)!}{C_n}$ in Fig. 3 and can be viewed as guide for a good choice of n. If $E_{\lambda_0}(x)$ and its derivatives vary a lot, then $n \approx \hat{n}$ should be a good choice whereas if it has a small slope or is an exponential decay n should be chosen larger $n \geqslant \hat{n}$.

Once these choices have been made, the procedure is straightforward. First, numerically solve $\Lambda = \sum_{j=0}^{n} |\gamma_j|$ for x_1. Use this to calculate the nodes x_j, γ_j, and N_j. Measure $E_{\lambda_0}(x_j)$ each N_j times and calculate the estimate $R_n = \sum_{j=0}^{n} \gamma_j E_{\lambda_0}(x_j)$ of E^*. This estimate has a statistical error (standard deviation) of $\sigma/\sqrt{N_{\text{eff}}}$. This procedure is schematically represented in Fig. 4.

IV. EXAMPLES

In this section, we test the procedure for Richardson extrapolation on different noise models and investigate the dependency on n. First, we focus on Markovian noise and in a second step address non-Markovian noise. All estimators in this section have a variance of σ^2/N_{eff}. Note that the following figures focus on the bias and do not show the variance.

Markovian noise (M) can emerge for example when the quantum state interacts with a bath and the bath correlation functions decay much faster than typical time scales of the system [44]. It typically leads to an exponential decay of the expectation value [18]:

$$E_{\lambda_0}^M(x) = E^* e^{-\lambda_0 x}. \tag{19}$$

For $\Lambda = 8$ and 64 and $n = 5$, $|\text{Bias}[\hat{R}_n]|$ is plotted over λ_0 in Fig. 5. We can see that the bias depends most notably on the overhead Λ^2. It can therefore make sense to use more measurements and increase Λ^2 in order to reduce the bias. Furthermore, we can see that for $n = 5$ the equidistant spacing performs significantly worse than the other methods in accordance with the general results presented in Sec. III.

In order to investigate the dependency on n, we fix $\lambda_0 = 0.4$ and plot the resulting estimate for different n in Fig. 6.
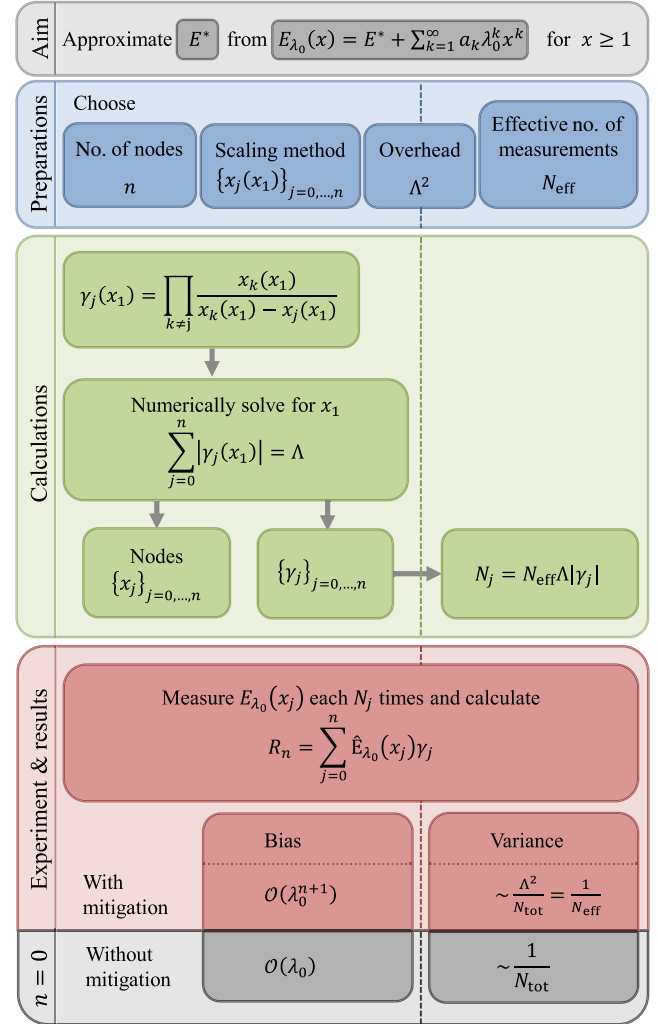

FIG. 4. Schematic representation of the procedure for Richardson extrapolation.

Remember that, for a given choice of Λ, each data point is calculated using the same sampling budget N_{tot}, independently of n. We can observe that the bias can be significantly reduced by increasing n for all spacing methods except for the equidistant nodes.

For this noise model, increasing n reduces the error by more than an order of magnitude. Interestingly, this is the case even if $\lambda_0 \geqslant 1$ (which is not shown here), such that this effect is not solely caused by a decrease of $|E_{\lambda_0}^{(n+1)}(\xi)|$ with n. It seems that increasing n is advantageous for any noise strength if the noise is given by an exponential decay. In practice, the limitations mentioned in Sec. III C render $n \gg 10$ impractical for many QEM applications.

If we know the explicit form of the noise function, we can directly fit it to the data points and there is no need for Richardson extrapolation. In the case that the noise is purely Markovian and can be approximated by a single or multiexponential decay, this has been done, e.g., in Refs. [11,17]. Another idea is to use Richardson extrapolation on the data points $\ln E_{\lambda_0}(x_j)$ for an estimate of $\ln E^*$ (assuming $\lim_{\lambda \to \infty} E_\lambda = 0$) which is equivalent to a poly-
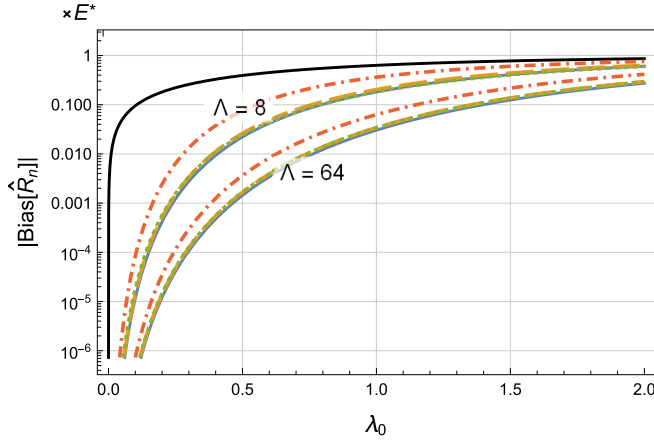
FIG. 5. Richardson extrapolation used on the Markovian noise model in Eq. (19). The bias is shown over the minimal noise strength λ_0 for $\Lambda = 8$ and 64 and $n = 5$. The spacing methods are tilted Chebyshev (blue, solid), extremal Chebyshev (yellow, dashed), exponential (green, dotted), and equidistant (red, dot-dashed). The black solid line shows the result without mitigation $|\text{Bias}[\hat{R}_0]|$. The y axis is in units of E^*. For the figures, we assume E^* to be nonzero and positive.

FIG. 7. Comparison of the bias of Richardson extrapolation on the noise model in Eq. (20) with $\eta = 0.1$ and $\lambda_0 = 0.4$ for different n. The overhead in measurements is $\Lambda^2 = 4^2$, 32^2, and 256^2, respectively. Note that the smaller the bias the better the QEM protocol works. The different spacing methods are tilted Chebyshev (blue, up-triangle), extremal Chebyshev (yellow, diamond), exponential (green, down-triangle), and equidistant (red, circle). The black square for $n = 0$ is the bias without mitigation $|\text{Bias}[\hat{R}_0]|$. The y axis is in units of E^*.

exponential extrapolation [29]. However, if the noise is a combination of different functions or is unknown altogether, then an important advantage of Richardson extrapolation is that polynomials are flexible enough to fit different kinds of functions.

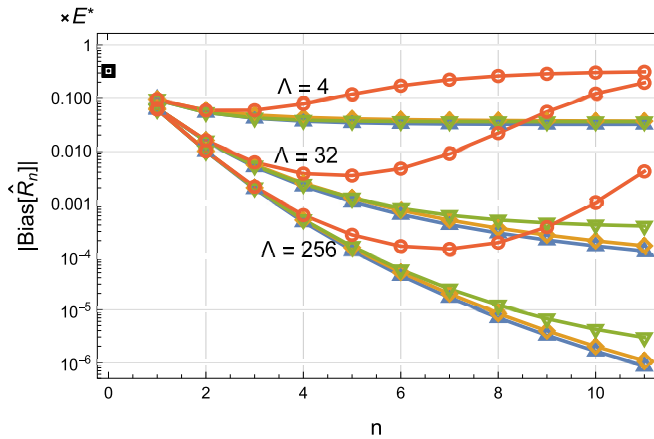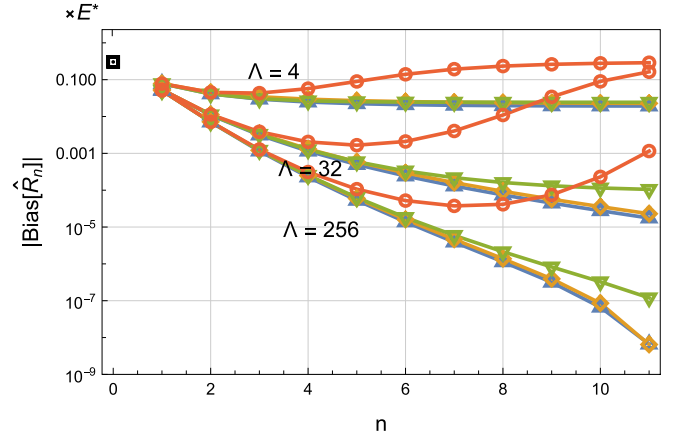To test this we now introduce non-Markovian (NM) noise components to the noise function. Therefore, we use the following model:

$$
E_{\lambda_0}^{\text{NM}}(x) = E^* \frac{e^{-(1-\eta)\lambda_0 x}}{\cos(2)} \left[\cos(\eta\lambda_0 x)\cos(\omega) + \frac{\eta\lambda_0 x}{\omega}\sin(\eta\lambda_0 x)\sin(\omega) \right] \quad (20)
$$

where $\omega = \sqrt{4 + (\eta\lambda_0 x)^2}$. The non-Markovian part of this noise stems from the coupling of a single-qubit system to an environment consisting of one qubit. The parameter η can be used to interpolate between purely Markovian noise ($\eta = 0$) and highly non-Markovian noise ($\eta = 1$). A more detailed description of this noise model can be found in Appendix B.

In Fig. 7, we consider the presence of a small non-Markovian component, $\eta = 0.1$, and plot the bias of $\hat{R}_n$ over n. In this case, the error mitigation behaves similarly to the purely Markovian case in Fig. 6. In particular, for all spacing methods except the equidistant nodes, the bias can be reduced significantly by increasing n. For these simple Markovian or near-Markovian noise models it seems that there is no disadvantage when increasing n.

In Fig. 8, we focus on the tilted Chebyshev nodes and investigate how mitigation for different n changes when interpolating between Markovian and non-Markovian noise. For $\eta \approx 0$, $n = 9$ (dashed lines) clearly leads to a lower bias than $n = 4$ (solid lines). For $\eta \to 1$ this is not true anymore, especially for small Λ. For the highly non-Markovian noise model, the derivatives $|E_{\lambda_0}^{(n+1)}(\xi)|$ do not decrease (on average) when increasing n. From the considerations in Sec. III C we therefore expect the mitigation result to worsen when $n \gg \hat{n}$ which is what can be observed in the plot. This confirms that, for a simple noise model, n should be chosen larger than $\hat{n}$ whereas for more complex noise n should be chosen in proximity

FIG. 6. Comparison of the bias of Richardson extrapolation on Markovian noise for fixed $\lambda_0 = 0.4$ and different n. The overhead in measurements is $\Lambda^2 = 4^2$, 32^2, and 256^2, respectively. Note that the smaller the bias the better the QEM protocol works. The different spacing methods are tilted Chebyshev (blue, up-triangle), extremal Chebyshev (yellow, diamond), exponential (green, down-triangle), and equidistant (red, circle). The black square for $n = 0$ is the bias without mitigation $|\text{Bias}[\hat{R}_0]|$. The y axis is in units of E^*.
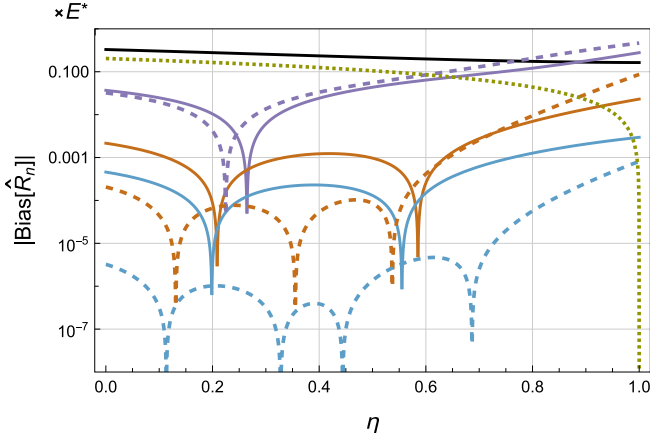
FIG. 8. The bias of mitigation for different levels of non-Markovianity. The noise is purely Markovian for $\eta = 0$ and highly non-Markovian for $\eta = 1$. There are three different values for Λ—4 (purple), 32 (brown), and 256 (blue)—and two for n: 4 (solid) and 9 (dashed). The black solid line is the result without mitigation $|\text{Bias}[\hat{R}_0]|$. The dotted green line is the fake-node approach with $S(x) = x^2$ for $\Lambda = 4$ and $n = 9$. For some η, there is a transition between over- and underestimation of the true expectation value E^* which is why we see "dips," where the bias briefly becomes zero. The y axis is in units of E^*.

to $\hat{n}$. The full dependency of the bias on n for the highly non-Markovian case $\eta = 0.9$ is shown in Appendix C.

Finally, we want to briefly comment on the idea of *fake nodes* introduced in Ref. [45]. From the Runge problem [27], it is well known that equidistant nodes are not well suited for interpolation. However, if specific hardware requirements force the nodes to be equidistant, their limitations can be bypassed by mapping them to so called *fake nodes* $\tilde{x}_j = S(x_j)$ for some invertible $S(x)$ with $S(0) = 0$ and $S(1) = 1$ and performing Richardson extrapolation of $E_{\lambda_0}[S^{-1}(\tilde{x})]$ on these fake nodes. By choosing for example the zeros of the Chebyshev polynomials as fake nodes the limitations of equidistant nodes can be avoided. While this proved to be effective for the interpolation of the Runge function, the map to the tilted Chebyshev nodes of the same Λ does not provide a benefit for the Richardson extrapolation of Eqs. (19) and (20). However, the same formalism can be used in a different way to improve the mitigation result in the completely non-Markovian case $\eta = 1$. Since then the function $E_{\lambda_0}^{\text{NM}}(x)$ is an even function in x, it makes sense to use only even polynomials for extrapolation. This idea has been mentioned in Ref. [9] and can be achieved by choosing $S(x) = x^2$, for which $E_{\lambda_0}^{\text{NM}}(x)$ is approximated in the space spanned by $\{1, x^2, x^4, \ldots, x^{2n}\}$ instead of $\{1, x, x^2, \ldots, x^n\}$. An example for this with $n = 9$ is plotted in Fig. 8. There the fake nodes $\tilde{x}_j$ are the tilted Chebyshev nodes with $\Lambda = 4$ and the corresponding real nodes are $x_j = \sqrt{\tilde{x}_j}$.

V. CONCLUSION

The goal of quantum error mitigation is to find an estimator of a noiseless expectation value from noisy observations. This estimator should have small bias and variance for as few

computational resources as possible. We analyze the interplay of these three factors in the case of Richardson extrapolation. It is a common perception that Richardson extrapolation is limited by its substantial increase in variance, especially if the number of nodes is increased. We propose a protocol in which the variance is independent of the number of nodes and can be tuned precisely. This allows us to increase the number of nodes n, which in turn can reduce the bias of the zero noise estimate significantly, especially for large sampling overheads. Furthermore, we show that equidistant nodes are not a good choice for Richardson extrapolation and propose the tilted Chebyshev nodes, which optimize the tradeoff between bias and variance and outperform other choices, especially for large n.

APPENDIX A: TILTED CHEBYSHEV NODES

The optimization problem described in Sec. III B is formalized as

$$\underset{x_0, \ldots, x_n \in [1, \infty)}{\text{minimize}} \quad C_n \tag{A1}$$

$$\text{subject to} \quad \sum_{j=0}^{n} |\gamma_j| = \Lambda \tag{A2}$$

$$\text{and} \quad x_0 = 1. \tag{A3}$$

It can be solved using Lagrange multipliers μ and μ_0. The solution has to satisfy

$$\nabla \left[C_n - \mu \left(\sum_{j=0}^{n} |\gamma_j| - \Lambda \right) - \mu_0 (x_0 - 1) \right] = 0 \tag{A4}$$

where ∇ is the gradient with respect to $(x_0, \ldots, x_n, \mu, \mu_0)^T$. Performing the derivatives, this leads to

$$C_n = \mu_0 x_0 \delta_{k,0} + \mu \varphi_k \tag{A5}$$

for $k = 0, \ldots, n$, where $\delta_{k,0}$ is the Kronecker delta and

$$\varphi_k := \sum_{j \neq k} \frac{1}{x_j - x_k} [(-1)^j x_j \gamma_j + (-1)^k x_k \gamma_k]. \tag{A6}$$

Notice that $\gamma_j = (-1)^j |\gamma_j|$. Summing this expression over $k = 0, \ldots, n$, this leads to

$$\sum_{k=0}^{n} C_n = (n+1) C_n = \mu_0 \tag{A7}$$

due to the antisymmetry under exchange of $j \leftrightarrow k$ and thus

$$\mu \varphi_k = \begin{cases} -n C_n & \text{if } k = 0 \\ C_n & \text{else.} \end{cases} \tag{A8}$$

What is left to do is to show that Eq. (18), i.e.,

$$x_j = 1 + \frac{\sin^2\left(\frac{j}{n+1}\frac{\pi}{2}\right)}{\sin^2\left(\frac{1}{n+1}\frac{\pi}{2}\right)}(x_1 - 1), \tag{A9}$$

solves this system of equations Eq. (A8). In order to do so the following equations are useful:

$$\alpha := \frac{\pi}{2(n+1)}, \tag{A10}$$

$$\frac{1}{x_m - x_j} = \frac{1}{x_1 - 1}\frac{\sin^2(\alpha)}{\sin^2(m\alpha) - \sin^2(j\alpha)} \tag{A11}$$

$$= \frac{1}{x_1 - 1}\frac{\sin^2(\alpha)}{\sin((m-j)\alpha)\sin((m+j)\alpha)},$$

$$\frac{\gamma_0}{C_n} = \left(\prod_{l=0}^{n}\frac{\sin^2(\alpha)}{\sin^2(l\alpha)}\frac{1}{x_1 - 1}\right), \tag{A12}$$

$$\frac{x_j\gamma_j}{C_n} = \frac{x_j}{C_n}\prod_{m\neq j}\frac{x_m}{x_m - x_j}$$

$$= \prod_{m\neq j}\frac{(x_1-1)^{-1}\sin^2(\alpha)}{\sin[(m-j)\alpha]\sin[(m+j)\alpha]}. \tag{A13}$$

In the last product, the sine appears in the denominator $2n$ times. The symmetries $\sin(x) = -\sin(-x)$ as well as $\sin[(n+1+k)\alpha] = \sin[(n+1-k)\alpha]$ can be used to bring all arguments into the interval $[0, \pi/2]$, which leads to an overall sign of $(-1)^j$. Then each argument $l\alpha$ with $l = 1, \ldots, n$ appears exactly twice except for $l = 2j, n+1-j$ (only once), and j (three times). Additionally, $l = n+1$ appears once. This leads to

$$\frac{x_j\gamma_j}{C_n} = (-1)^j\frac{\gamma_0}{C_n}\frac{\sin[(n+1-j)\alpha]\sin(2j\alpha)}{\sin[(n+1)\alpha]\sin(j\alpha)} \tag{A14}$$

$$= (-1)^j\frac{\gamma_0}{C_n}[2\cos^2(j\alpha) - \delta_{j,0}] \tag{A15}$$

such that for arbitrary j

$$x_j\gamma_j = (-1)^j x_0\gamma_0[2\cos^2(j\alpha) - \delta_{j,0}] \tag{A16}$$

Inserting all of this into φ_k, this gives

$$\varphi_k = \sin^2(\alpha)\frac{x_0\gamma_0}{x_1 - 1}\Omega_k, \tag{A17}$$

where $\Omega_k := \sum_{j\neq k}\frac{2\cos^2(j\alpha) + 2\cos^2(k\alpha) - \delta_{k,0} - \delta_{j,0}}{\sin^2(j\alpha) - \sin^2(k\alpha)}$. A last step is to show that

$$\Omega_k = \begin{cases} 2n(n+1) & \text{if } k = 0 \\ -2(n+1) & \text{else.} \end{cases} \tag{A18}$$

This last step has not been achieved analytically for arbitrary n. Only for $k = 0$ a proof can be found for example in Ref. [46] Chap. 9, Art. 69. However, for a given n this formula can be easily checked numerically, which we have done for $n \leqslant 1000$. For the purpose of zero noise extrapolation, we do not expect n to exceed 15. Therefore for $n \leqslant 1000$ and the Lagrange multiplier

$$\mu = \frac{-C_n(x_1 - 1)}{2\sin^2(\alpha)x_0\gamma_0(n+1)} \tag{A19}$$

we get

$$C_n = -\frac{\mu}{n}\varphi_0 = \mu\varphi_{k\neq 0}. \tag{A20}$$

This proves that Eq. (A9) solves Eq. (A8) and is thus a candidate for a minimum or a maximum of the optimization problem [Eqs. (A1)–(A3)]. Compared to other spacing methods, this choice has a lower C_n such that it is reasonable to assume that the tilted Chebyshev nodes are indeed a minimum of the optimization problem. It should be noted that there might be other minima (that respect the ordering $x_0 < \cdots < x_n$) leading to a lower minimum of C_n.

APPENDIX B: NON-MARKOVIAN NOISE MODEL

The non-Markovian noise model was derived using the following toy model. The system consists of a single qubit that is subjected to depolarization noise of strength $\lambda_M = (1 - \eta)\lambda$ and at the same time coupled to a second qubit, which simulates a highly non-Markovian environment, via the Hamiltonian $H = Z \otimes I + \lambda_{\text{NM}}X \otimes X + I \otimes Z$. Here $\lambda_{\text{NM}} = \eta\lambda$ is the non-Markovian coupling strength and I, X, and Z denote the identity and the x- and z-Pauli matrices, respectively. The dynamics of the two-qubit state ρ_λ is given by the equation

$$\frac{\partial}{\partial t}\rho_\lambda(t) = -i[H, \rho_\lambda] + \lambda_M\left(\frac{\text{tr}_2(\rho_\lambda)}{2} - \rho_\lambda\right). \tag{B1}$$

We used the initial condition $\rho_{\lambda_0}(0) = \frac{I+X}{2} \otimes \frac{I}{2}$. After time τ the expectation value $E_\lambda^{\text{NM}}(\tau) = \text{tr}[\rho_\lambda(\tau)A]$ of the observable $A = X \otimes I$ is given by

$$E_\lambda^{\text{NM}}(\tau) = e^{-\tau\lambda_M}\Bigg[\cos(\tau\lambda_{\text{NM}})\cos(\tau\omega) + \frac{\lambda_{\text{NM}}}{\omega}\sin(\tau\lambda_{\text{NM}})\sin(\tau\omega)\Bigg] \tag{B2}$$
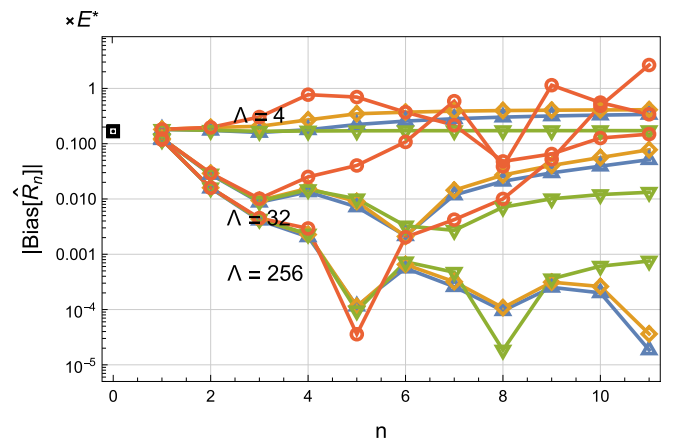

FIG. 9. Comparison of the bias of Richardson extrapolation on the noise model in Eq. (20) with $\eta = 0.9$ and $\lambda_0 = 0.4$ for different n. The overhead in measurements is $\Lambda^2 = 4^2$, 32^2, and 256^2, respectively. The different spacing methods are tilted Chebyshev (blue, up-triangle), extremal Chebyshev (yellow, diamond), exponential (green, down-triangle), and equidistant (red, circle). The black square for $n = 0$ is the bias without mitigation $|\text{Bias}[\hat{R}_0]|$. The y axis is in units of E^*.

where $\omega = \sqrt{4 + \lambda_{NM}}$. For simplicity we chose $\tau = 1$, which results in the noise model in Eq. (20).

APPENDIX C: HIGHLY NON-MARKOVIAN NOISE

In the highly non-Markovian case, the mitigation behavior can be more complex than for purely Markovian noise. In Sec. IV, we demonstrated that in this case the mitigation result does not necessarily improve with larger n. Analogously to Figs. 6 and 7, we plot the bias of $\hat{R}_n$ over n in Fig. 9, this time for a highly non-Markovian noise $\eta = 0.9$.

For this noise function, the mitigation is more chaotic and a small change in one of the parameters n or λ_0 can have a large influence on the mitigation result. We can see that, for small Λ, the mitigation result worsens for larger n. For $\Lambda = 4$ this is the case once $n > 3$ and for $\Lambda = 32$ for $n > 6$. This confirms that, for a complex noise model, n should not be chosen too large. From Fig. 3 we can estimate that $\hat{n}$ for $\Lambda = 4$, 32, and 256 are $\hat{n} = 1$, 2–3, and 5–6, respectively. For this specific noise model it seems that $n \approx 2\hat{n}$ is a good choice for n. The parameter Λ, on the other hand, should still be chosen as large as possible. The larger Λ, the larger n can be chosen and the more noise can be mitigated with Richardson extrapolation.

[1] P. W. Shor, Scheme for reducing decoherence in quantum computer memory, Phys. Rev. A **52**, R2493 (1995).

[2] A. R. Calderbank and P. W. Shor, Good quantum error-correcting codes exist, Phys. Rev. A **54**, 1098 (1996).

[3] A. M. Steane, Error Correcting Codes in Quantum Theory, Phys. Rev. Lett. **77**, 793 (1996).

[4] A. Yu. Kitaev, Fault-tolerant quantum computation by anyons, Ann. Phys. (NY) **303**, 2 (2003).

[5] B. M. Terhal, Quantum error correction for quantum memories, Rev. Mod. Phys. **87**, 307 (2015).

[6] S. J. Devitt, A. M. Stephens, W. J. Munro, and K. Nemoto, Requirements for fault-tolerant factoring on an atom-optics quantum computer, Nat. Commun. **4**, 2524 (2013).

[7] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, Phys. Rev. A **86**, 032324 (2012).

[8] J. O'Gorman and E. T. Campbell, Quantum computation with realistic magic-state factories, Phys. Rev. A **95**, 032338 (2017).

[9] K. Temme, S. Bravyi, and J. M. Gambetta, Error Mitigation for Short-Depth Quantum Circuits, Phys. Rev. Lett. **119**, 180509 (2017).

[10] Y. Li and S. C. Benjamin, Efficient Variational Quantum Simulator Incorporating Active Error Minimization, Phys. Rev. X **7**, 021050 (2017).

[11] S. Endo, S. C. Benjamin, and Y. Li, Practical Quantum Error Mitigation for Near-Future Applications, Phys. Rev. X **8**, 031027 (2018).

[12] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Error mitigation extends the computational reach of a noisy quantum processor, Nature (London) **567**, 491 (2019).

[13] R. Sagastizabal, X. Bonet-Monroig, M. Singh, M. A. Rol, C. C. Bultink, X. Fu, C. H. Price, V. P. Ostroukh, N. Muthusubramanian, A. Bruno, M. Beekman, N. Haider, T. E. O'Brien, and L. DiCarlo, Experimental error mitigation via symmetry verification in a variational quantum eigensolver, Phys. Rev. A **100**, 010302(R) (2019).

[14] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, S. Boixo, M. Broughton, B. B. Buckley, D. A. Buell, B. Burkett, N. Bushnell, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, S. Demura, A. Dunsworth, E. Farhi *et al.*, Hartree-fock on a superconducting qubit quantum computer, Science **369**, 1084 (2020).

[15] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, A. Bengtsson, S. Boixo, M. Broughton, B. B. Buckley, D. A. Buell, B. Burkett, N. Bushnell, Y. Chen, Z. Chen, Y.-A. Chen, B. Chiaro, R. Collins, S. J. Cotton, W. Courtney *et al.*, Observation of separated dynamics of charge and spin in the Fermi-Hubbard model, arXiv:2010.07965 (2020).

[16] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[17] Z. Cai, Multi-exponential error extrapolation and combining error mitigation techniques for NISQ applications, npj Quantum Inf. **7**, 80 (2021).

[18] Z. Cai, A practical framework for quantum error mitigation, arXiv:2110.05389 (2021).

[19] D. Bultrini, M. Hunter Gordon, P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, Unifying and benchmarking state-of-the-art quantum error mitigation techniques, arXiv:2107.13470 (2021).

[20] J. R. McClean, Z. Jiang, N. C. Rubin, R. Babbush, and H. Neven, Decoding quantum errors with subspace expansions, Nat. Commun. **11**, 636 (2020).

[21] S. McArdle, X. Yuan, and S. Benjamin, Error-Mitigated Digital Quantum Simulation, Phys. Rev. Lett. **122**, 180501 (2019).

[22] X. Bonet-Monroig, R. Sagastizabal, M. Singh, and T. E. O'Brien, Low-cost error mitigation by symmetry verification, Phys. Rev. A **98**, 062339 (2018).

[23] Z. Cai, Quantum error mitigation using symmetry expansion, Quantum **5**, 548 (2021).

[24] B. Koczor, Exponential Error Suppression for Near-Term Quantum Devices, Phys. Rev. X **11**, 031057 (2021).

[25] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, Ry. Babbush, and J. R. McClean, Virtual Distillation for Quantum Error Mitigation, Phys. Rev. X **11**, 041036 (2021).

[26] L. F. Richardson and J. A. Gaunt, The deferred approach to the limit, Phil. Trans. R. Soc. A **226**, 299 (1927).

[27] C. Runge, Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten, Zeitschrift für Mathematik und Physik **46**, 224 (1901).

[28] P. G. Hoel and A. Levine, Optimal spacing and weighting in polynomial prediction, Annal. Math. Stat. **35**, 1553 (1964).

[29] T. Giurgica-Tiron, Y. Hindy, R. LaRose, A. Mari, and W. J. Zeng, Digital zero noise extrapolation for quantum error mitigation, in Proceedings of the 2020 IEEE International

Conference on Quantum Computing and Engineering (IEEE, Piscataway, NJ, 2020), pp. 306–316.

[30] P. L. Chebyshev, Sur les questions de minima qui se rattachent a la représentation approximative des functions, *Mémoires de l'Académie Impériale des sciences de St.-Pétersbourg*, Sixième série, Sciences mathématiques et physiques, Tome VII (1859), pp. 199–291 [reprinted in Oeuvres de P. L. Tchebyschef, Vol. I (Imperial Academy of Sciences, St. Petersbourg, 1899), pp. 273–378].

[31] K. Goswami, C. Giarmatzi, C. Monterola, S. Shrapnel, J. Romero, and F. Costa, Experimental characterization of a non-Markovian quantum process, Phys. Rev. A **104**, 022432 (2021).

[32] J. Morris, F. A. Pollock, and K. Modi, Non-Markovian memory in IBMQX4, arXiv:1902.07980 (2019).

[33] G. A. L. White, C. D. Hill, F. A. Pollock, L. C. L. Hollenberg, and K. Modi, Demonstration of non-Markovian process characterisation and control on a quantum processor, Nat. Commun. **11**, 6301 (2020).

[34] E. F. Dumitrescu, A. J. McCaskey, G. Hagen, G. R. Jansen, T. D. Morris, T. Papenbrock, R. C. Pooser, D. J. Dean, and P. Lougovski, Cloud Quantum Computing of an Atomic Nucleus, Phys. Rev. Lett. **120**, 210501 (2018).

[35] A. He, B. Nachman, W. A. de Jong, and C. W. Bauer, Zero-noise extrapolation for quantum-gate error mitigation with identity insertions, Phys. Rev. A **102**, 012426 (2020).

[36] R. LaRose, A. Mari, S. Kaiser, P. J. Karalekas, A. A. Alves, P. Czarnik, M. El Mandouh, M. H. Gordon, Y. Hindy, A. Robertson, P. Thakre, M. Wahl, D. Samuel, R. Mistri, M.

Tremblay, N. Gardner, N. T. Stemen, N. Shammah, and W. J. Zeng, Mitiq: A software package for error mitigation on noisy quantum computers, Quantum **6**, 774 (2022).

[37] K. Schultz, R. LaRose, A. Mari, G. Quiroz, N. Shammah, B. D. Clader, and W. J. Zeng, Reducing the impact of time-correlated noise on zero-noise extrapolation, Phys. Rev. A **106**, 052406 (2022).

[38] P. J. Davis, *Interpolation and Approximation*, Introductions to Higher Mathematics (Blaisdell Publishing Company, Waltham, 1963), 1sT ed.

[39] R. W. Freund and R. W. Hoppe, *Stoer/Bulirsch: Numerische Mathematik* (Springer, Berlin, Heidelberg, 2007), Vol. 1.

[40] A. Mari, N. Shammah, and W. J. Zeng, Extending quantum probabilistic error cancellation by noise scaling, Phys. Rev. A **104**, 052607 (2021).

[41] P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, Error mitigation with Clifford quantum-circuit data, Quantum **5**, 592 (2021).

[42] L. Brutman, On the Lebesgue function for polynomial interpolation, SIAM J. Numer. Anal. **15**, 694 (1978).

[43] There is one exception to this rule. If errors on the nodes x_j play a significant role, it can make sense to choose a smaller Λ to reduce the effect of the errors on x_j for mitigation.

[44] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford, 2006).

[45] S. De Marchi, F. Marchetti, E. Perracchione, and E. Poggiali, Polynomial interpolation via mapped bases without resampling, J. Comput. Appl. Math. **364**, 112347 (2020).

[46] T. J. I'A. Bromwich, *An Introduction to the Theory of Infinite Series* (London Macmillan, London, 1908).