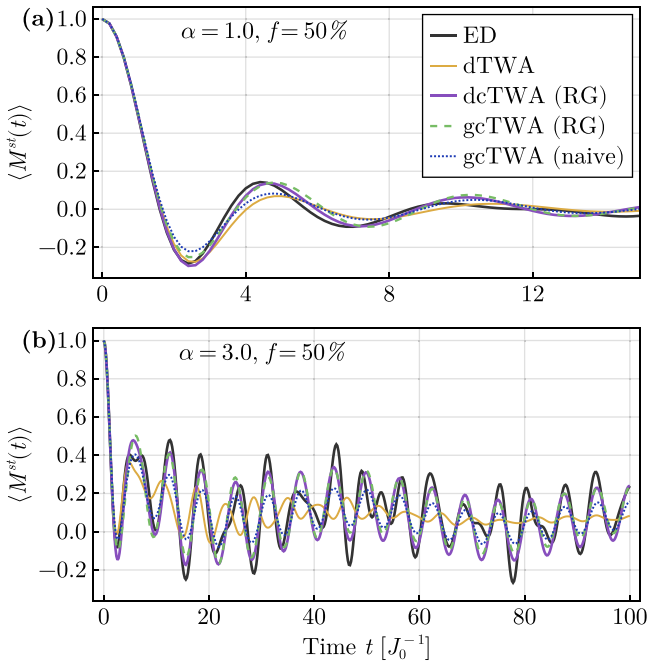

FIG. 4. Same as Fig. 3, but for filling fixed at $f = 50\%$ (weak disorder regime). Here, the methods using the RG-inspired clustering show some deviation from the exact results and for $\alpha = 1$ in panel (a) there are some differences between the Gaussian and discrete sampling schemes visible as well.

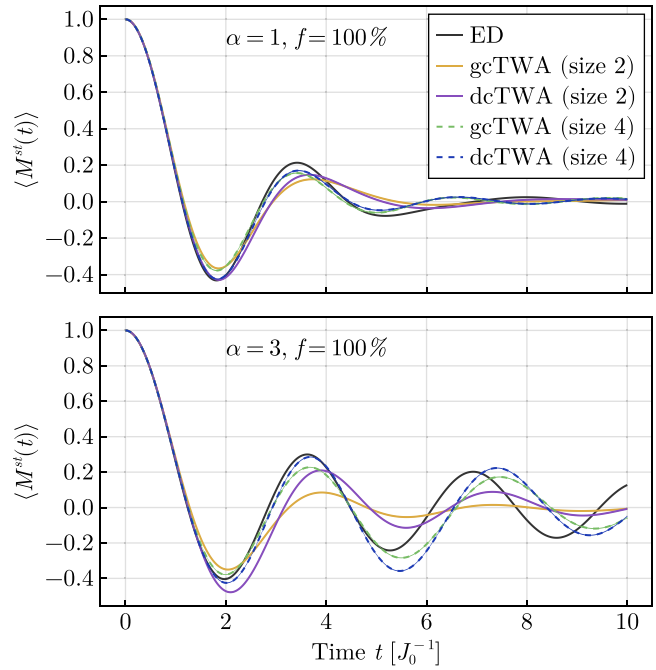

FIG. 5. Comparison of sampling schemes with different cluster sizes in a clean system (filling fraction $f = 100\%$). We compare TWA results for cluster sizes 2 (solid) and 4 (dashed) using the naive clustering and the different sampling schemes, Gaussian (lighter colors) and discrete (darker colors) to exact results [black (solid)]. Generally, cluster size 4 is more accurate than cluster size 2 and the discrete sampling scheme agrees with the exact results longer than the Gaussian sampling scheme. Other parameters are similar to Fig. 3.

to the oscillation frequency, it also underestimates the amplitude. In contrast, both cTWA variants using the RG clustering yield essentially exact results even at late times. This is a very strong indicator that the dynamics is strongly shaped by the presence of strongly interacting pairs of spins where interactions among pairs are weak [57]. With this pair model we can explain the observed curves qualitatively: It is known that dTWA is unable to correctly capture the dynamics of even a single pair and can just approximate the decay timescale (cf. Appendix B). If the two spins forming a strongly interacting pair are not part of the same cluster, then cTWA treats the interactions within the pair semiclassically similar to dTWA and thus faces the same problems. Consequently, using the naive clustering will result in a mixture of "correctly" and "incorrectly" chosen pairs and thus cTWA with this type of clustering provides only a slight improvement over dTWA. The RG clustering, in turn, ensures that all strongly interacting pairs are treated as clusters and thus the predictions match the exact dynamics much more closely. In turn, the high degree of agreement between cTWA with the RG clustering is also testament to the quality of pair approximation.

To further explore the efficacy of cTWA in regimes of weak disorder, we increase the filling fraction to $f = 50\%$ and repeat the analysis (cf. Fig. 4). In this regime, we do not expect the pair approximation to be accurate anymore. Indeed, for the long-range case $\alpha = 1$ we find all semiclassical methods to overestimate the oscillation frequency similarly. Both RG clustering-based methods predict the amplitudes almost exactly correct, while dTWA and cTWA with naive clustering again clearly underestimate it. In the more short-range case $\alpha = 3$, the picture is more complex. dTWA performs worst out of all the methods and does not resolve the oscillation well and cTWA with naive clustering again essentially underestimates the amplitude. Interestingly, in this case there is a clear difference on intermediate timescales $t \approx 8J_0$ between both cTWA methods with RG clustering but different choice of sampling. The discrete sampling captures the first oscillation slightly better both in position and in amplitude before at later times the prediction collapses onto the cTWA curve employing a Gaussian Wigner function. Generally, all cluster-based methods still capture the dynamics qualitatively but not quite quantitatively over the whole time shown here.

For a better comparison of the sampling schemes, it is instructive to examine a perfectly ordered regime by setting the filling factor to $f = 100\%$. In this setting, the RG and naive clustering schemes result in the same choice of clusters and we use this opportunity to check the convergence with increasing cluster size. Figure 5 shows the staggered magnetization results for systems with both long-range ($\alpha = 1.0$) and short-range ($\alpha = 3.0$) spin interactions and for cluster sizes 2 and 4. Similar to the weakly disordered case before, cluster size 2 is insufficient to capture the relaxation dynamics quantitatively. In the short-range case ($\alpha = 3$) gcTWA (cluster size 2) struggles to reproduce the oscillatory behavior, which is reflected better by dcTWA. This likely stems from the fact that this coherent dynamics comes about due to the discrete nature of the spin $\frac{1}{2}$ which is mimicked by the discrete sampling procedure [35]. Conversely, for the long-range system