## Answer to Report 2

We thank the referee Marin Bukov for the positive comments on our manuscript, and for the list of suggestions and questions, which we address in what follows.

> *The authors state: "Moreover, the final annealed state $\psi_T$ — resulting from the exact QA time evolution (with $\tau \gg 1$) and thus expected to yield a large overlap with classical solutions — is often a low-entanglement state": is this true also when the ground state of Hz is largely degenerate? This is the case, e.g., in frustrated models; when a superposition of a large number of degenerate states is considered, the resulting state may happen to be a quantum spin liquid — a class of topological states that possess high entanglement.*

This is indeed a crucial aspect: for a classical optimization problem $\widehat{H}_z$ with a large number of classical solutions (ground-state configurations), the exact QA evolution could in principle converge to a highly entagled state with non-vanishing overlap with a large subset (or possibly, all) of the classical solutions. In this case, an MPS simulation with small bond dimension $\chi$ would not accurately describe the actual quantum evolution.

In the general case of a classical $\widehat{H}_z$ with a largely-degenerate ground state, the study of the entanglement properties of the annealed state — e.g. by varying annealing hyperparameters — represents an interesting direction of future research. For what concerns the class of problems addressed in our paper, we verified numerically that the final annealed state has low entanglement, see Appendix E. More specifically, in Fig. 22, we plot the half system entanglement entropy of the final annealed state $S_{N/2}(1)$, for dQA *with* and *without* Trotterization via ED, compared with MPS results (which refer, as always, to the Trotterized case).

In particular, the upper right plot refers to the case of a perceptron with $N = 18$ and $N_\xi = 3\,(\alpha \simeq 0.17)$, thus having (on average) a large number of classical solutions (deep in the SAT regime). Nevertheless, the "true" QA *without* Trotterization (compare with Fig. 21) converges to low entanglement states for all values of $\delta t$.

On a side note, a large increase in $S_{N/2}(1)$ in the regime of large $\delta t \sim \mathcal{O}(1)$ is observed for dQA *with* Trotterization, this behavior being linked to large Trotter errors, as discussed in the manuscript, which MPS simulations can partially overcome.

> *Fig 7b: the $P = 1000$ data point at $dt = 1.7$ (orange square) seems to be an outlier; did the variational algorithm get stuck in a local minimum, or what is the reason for this behavior?*

Thanks for pointing out this issue. In this Figure, we are performing an average over five different training sets, and the plot is in log-scale for the residual energy density ($y$-axis). As a consequence, even if for a single training set the energy density jumps to large values, this will result in an outlier point, as it happened in this case. In our opinion, for such large values of $\delta t$, we enter in an "instability region", where our MPS techniques may no longer provide better results compared to ED. This is more evident for even larger values of $\delta t > 2$, where our MPS simulations do not converge to high-quality final states, with a complete degradation in performance. This fact might either be linked to a similar degradation in performance of a digitized Quantum Annealing *without* Trotterization (see Section 3.3) for such values of $\delta t$, or simply to technical limitations of our MPS methods.

*Fig 8: what happens deeper in the UNSAT regime? Note that if exact GS cannot be reached the algorithm may still be useful in practice since in many practical cases one requires finding a single "good" solution.*

Thanks to your suggestion, we perform a numerical experiment running our MPS algorithm for the perceptron model with parameters $N = 21$, $P = 1000$, $\delta t = 1.0$ and $N_\xi = 25$ ($\alpha \simeq 1.2 > \alpha_c \simeq 0.83$), thus being in the UNSAT phase. As a benchmark, we also compute the optimal classical configuration by enumeration. We run the simulation for three random realizations of the training set $\{\boldsymbol{\xi}^\mu\}$, finding an average final energy density $\varepsilon(1) \approx 1.1 \cdot 10^{-4}$. We conclude that our algorithm is effective also in the UNSAT regime: our MPS simulations yield a final annealed wavefunction with a large overlap with "good" solutions, i.e. close to the global minimum of the energy.

*The authors benchmark their algorithms against system sizes within the scope of ED. This is meaningful, if one wants to compare against quantum annealing. However, for N ∼100 there are developed tools to easily find the GS of any two-body Hz, see e.g., http://spinglass.uni-bonn.de/ . If possible, it would be nice to demonstrate one instance of a system size in the Hopfield model where the proposed algorithm outperforms maxcut (even if it doesn't find the exact GS, or it's not feasible to verify that the GS has been reached). This would require system sizes of $N > 120$ sites or so I guess.*

Thanks to your advice, we perform a numerical experiment running our MPS algorithm for the Hopfield model, with parameters $N = 150$, $N_\xi = 15$, $P = 100$ and $\delta t = 0.5$. We consider a single random training set $\{\boldsymbol{\xi}^\mu\}$ and we run the online solver http://spinglass.uni-bonn.de/ with the same couplings $J_{ij}$. With the MPS QA approach, we get $\tilde{\varepsilon}(1) = \langle\psi(1)|\widehat{H}_z|\psi(1)\rangle / N \simeq -1.11$, whereas the classical solver yields $\simeq -1.16$. Thus, this implementation of QA seems to be not enough to outperform state-of-the-art classical methods. However, let us point out that one could significantly improve our result by *a*) running a simulation with more Trotter steps (for instance $P = 1000$), although this would require large computational time, and *b*) performing a QAOA optimization. Furthermore, it might be necessary to push the simulation to even larger system sizes, to try to outperform classical methods.

*Related to the above, how much can one hope to push the system size in practice with present-day state-of-the-art classical resources? The variational optimization poses some restriction on the system size N due to the extra iteration loops.*

We performed MPS simulations with a Python code, running on a single node of a cluster (8 cores). We reached systems of size $N \approx 100$. We believe that a further optimization of the code, involving for instance multi-node parallelization, can eventually boost the performances allowing the simulation of systems up to $N \approx 1000$.

*"The rest of the manuscript is organized as follows": I would make this a separate paragraph so it can be easily noticed.*

Thanks, fixed.

*Algorithm* 1*: refer to the corresponding Eqn numbers to easily locate the definition of the used quantities (e.g.,* $\tilde{U}_{k,p}$*, etc.).*

Done, with a minor modification for improved clarity.

*A good analysis of the computational scaling of the algorithm is provided; maybe add a separate table with the cost of the different parts of Algo 1, so the scaling can be easily located.*

In the new version of the manuscript, we included a table with the computational cost of each step of our Algorithm.

*Results section: it will be helpful to explicitly label each figure caption so that the model being studied is immediately visible.*

Thanks, we did so in all figure captions, also in the Appendices.

*"Suffering sensible noise": do the authors mean sensitivity to noise?*

Yes, we rephrased the concept in better words.

*"For the models in exam"* $\rightarrow$ *for the models we examine.*

Thanks, fixed.