1) Although the paper is based on experiments described in Sec. 2, there is not much experimental details. For instance, there is no overview of the experimental setup. Referring to a previous paper does not help, since no detailed overview of the setup is provided in that paper. Also, powers used for the trapping and imaging are not discussed, although the power seemed to be limiting.

We agree with this point and happily include more information on the experimental setup. This is included in the new "Appendix A. Experimental details". We include specifics about the experimental setup and the optics between SLM and tweezers. We also give numbers on the powers used for trapping and imaging. In "Appendix B. Data analysis and fidelities", we present details on the analysis of the success probabilities. We hope that this takes away this concern of the referee.

We have not included another figure, because it would resemble strongly Fig. 6 of Urech, *et al.*, PhysRevResearch **4** 023245 (2022). We did also include a reference to the Ph.D. thesis of A. Urech, because Chapter 4 provides more information on the experimental setup, including pictures and CAD drawings.

2) The average filling fraction for 16 atoms is 0.988, which is a big achievement. However, it is not discussed at all, what limits this filling fraction. Is this a fundamental limit of the technique, is this caused by the limited power, or are there other technical reasons, why it is this number.

A good point that indeed deserved more clarification. The previously reported values are the raw measurements including both the imaging survival and the success probability of the rearrangement step. This number was dominated by the survival probability of the single image before rearrangement. We do not believe that there is a fundamental limit to the technique or that other technical reasons were relevant at this point, although deeper traps would have increased the imaging survival as shown in the added correction data.

In Appendix B, we have included fidelities for the 4x4 and a 6x6 array under the same measuring conditions as the data presented in the main text and separated the imaging and rearrangement success.

We initially followed a commonly used correction formula discussed extensively in the supplementary information of Madjarov, *et al.*, Nature Physics **16**, 857-861 (2020). This gave us the result that our imaging survivals were over 99% in both the 130uK and 270uK patterns. However, we now believe the formula in above reference is based on an unjustified assumption, and we present a derivation for the correction formulas used in the new version of this paper. This has led us to quote the imaging survival in the first paragraph of Section 2 as "99%", rather than "over 99%".

Correcting for the imaging survival the high success probability of the rearrangement procedure becomes clear. In the main text, we have kept the same uncorrected number and added the following sentences:

> "The detection probability is mainly limited by the survival rate of the atoms during the first image. Correcting for this survival, we obtain a success probability of 0.998(+0.002 -0.006) per atom for the rearrangement, see Appendix B."

We note that due to the uncertainty on the imaging survival, the propagated standard deviation is large enough to include unphysical values of the survival rate above 1. We therefore put an upper limit to the error bar and employ asymmetric errors.

3) The final sentence of Sec. 2 concludes "This corresponds to a success rate of ..". It is not clear, where this number (0.991) is based on, and how it scales with the number of atoms (16) involved. And again, what limits this number? And how why does it make the method suitable for adjusting geometries on the same atomic sample?

The number 0.991 was based on a compound imaging + rearrangement success. With the analysis presented in Appendix B, we now separate the infidelities of these two processes, and we have included a more detailed analysis of the success probability of each rearrangement. Under the assumption that the rearrangement losses are the same for each rearrangement cycle, this results in a success rate of 0.997(2). This is the same as the success rate presented for the 6x6 -> 4x4 case, and we believe this is a realistic limit of the current implementation of the method. This number should not depend on the number of atoms, as long as the trap depth does not change.

In the paper the final sentence of Sec. 2 is now rewritten as:

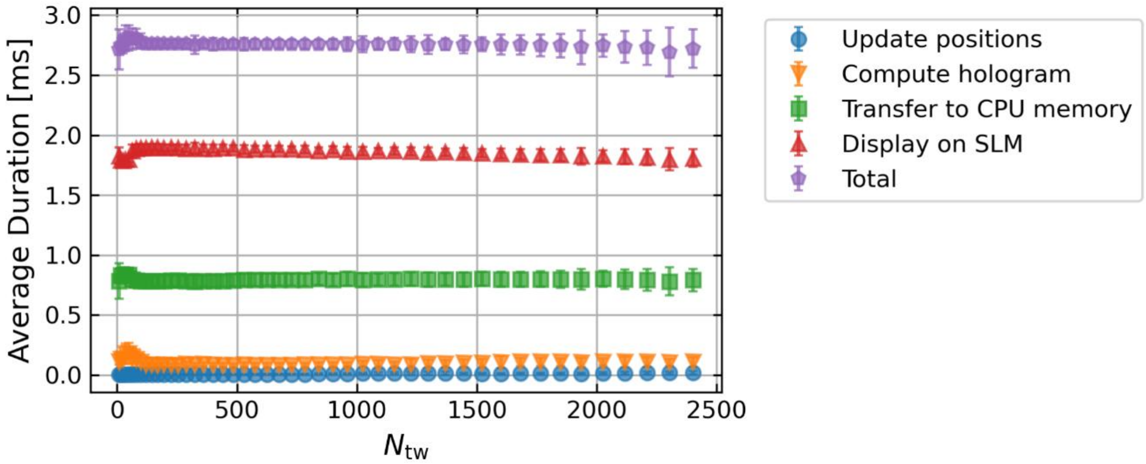
> "Assuming the same success probability for each rearrangement and correcting for losses induced by imaging in the different geometries, this results in a rearrangement success of 0.997(2) per atom per cycle."

The idea behind the last sentence was to highlight the capability of moving atoms to multiple geometries with limited losses. We believe this point is clear already from the data and removed the last sentence of the paragraph.

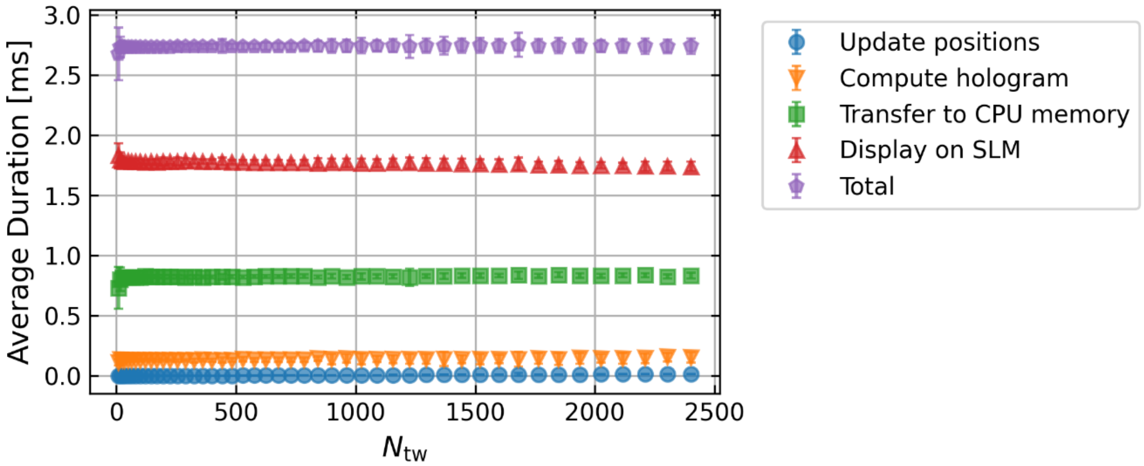
4) In Sec. 4 it is concluded that the method is nearly independent of Ntw. Figure 6 is an illustration of it. However, there are certain systematic effects visible that are not discussed. For instance, apparently the computation of the hologram takes more time for a small number of Ntw. Also, the display on SLM seems to go faster for more Ntw. And the error bar in the total time increases significantly for more than 2000 Ntw. These effect are all small, but require some interpretation from the authors to get a better feeling on the scaling with Ntw.

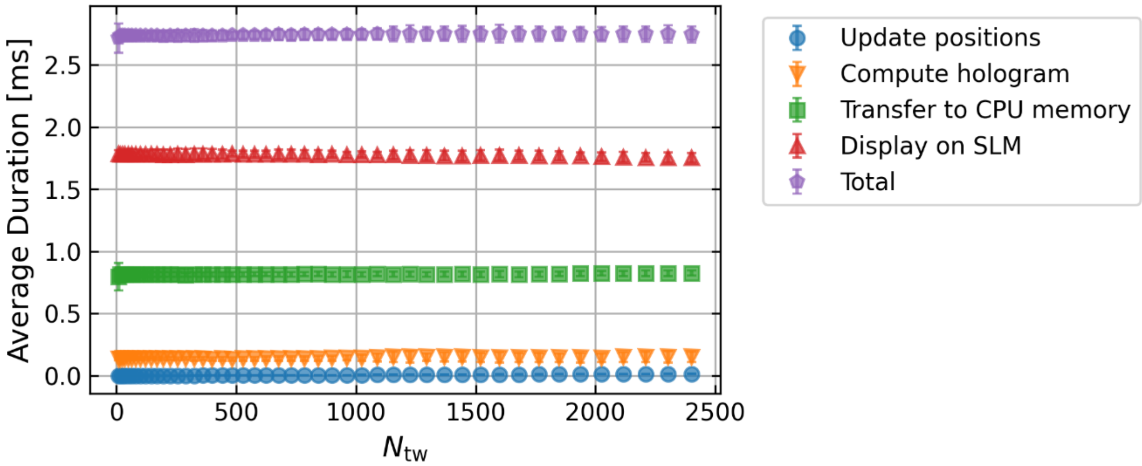
We highly appreciate these remarks. They have let us to repeat the measurement several times. We repeated twice the same order of low Ntw to high Ntw, which is the same as done in the paper, and twice in reversed order. See below the relevant plots.

Original figure in the paper:



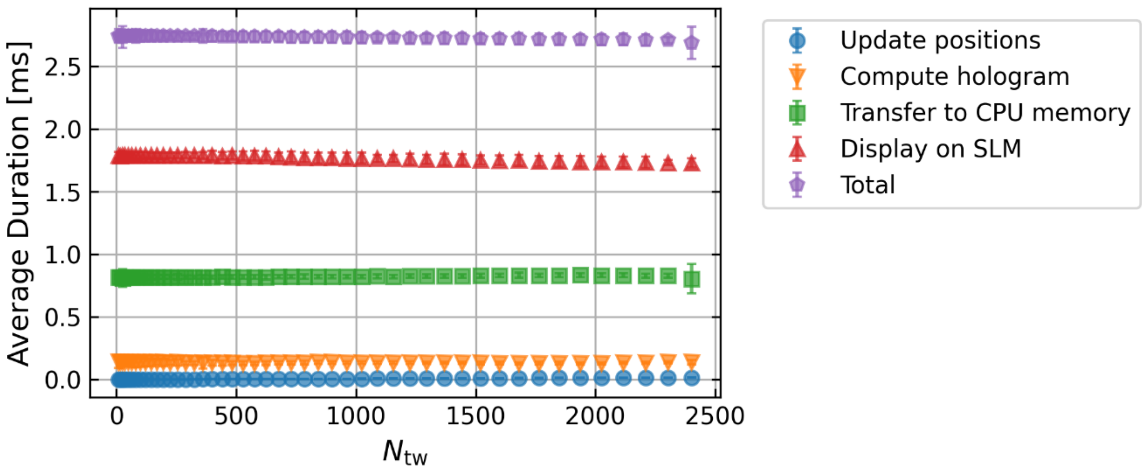
Retakes of same measurement as in the first version of the paper in ascending order of Ntw:

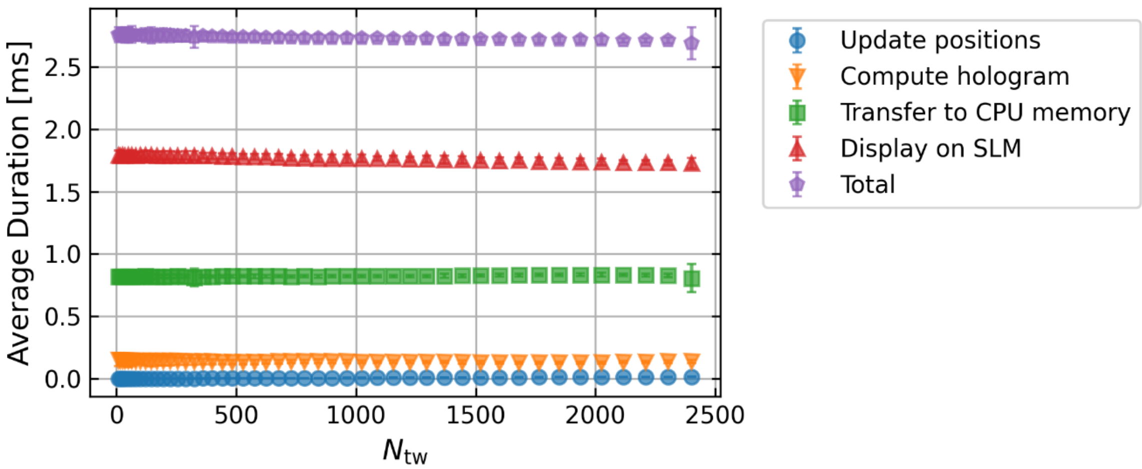




In both cases, the effect at low Ntw is much less pronounced. Only a small start-up effect is observed. We do not know the exact reasoning, but believe the improvement over the data presented in the first version of the paper may be due to a better isolation of this benchmarking process on the operating system. We note that the software of the SLM requires us to run Microsoft Windows OS. Scheduling precision of processes in the OS is on the order of milliseconds. In these retakes, we took care to log in remotely to the computer at night, when no other processes are running and close this connection right away, to limit the amount of processes running at the same time, which can cause unwanted delays. We further note that the data from the paper and the retakes are taken several months apart. In general, we believe that we can precisely measure the timings in our benchmarking process but have little control over the accuracy over such long time scales due to e.g. environmental changes or processes out of our control such as the OS.

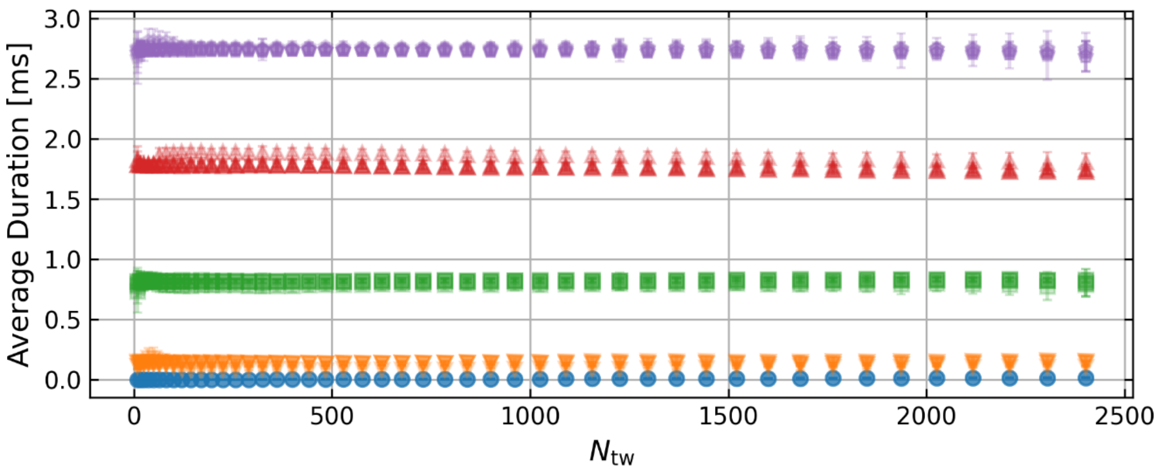
We observe a small but seemingly significant speed up on the display on SLM. We hypothesized that this was due to heating of something inside the SLM and reversed the order of the measurement, starting with high Ntw:





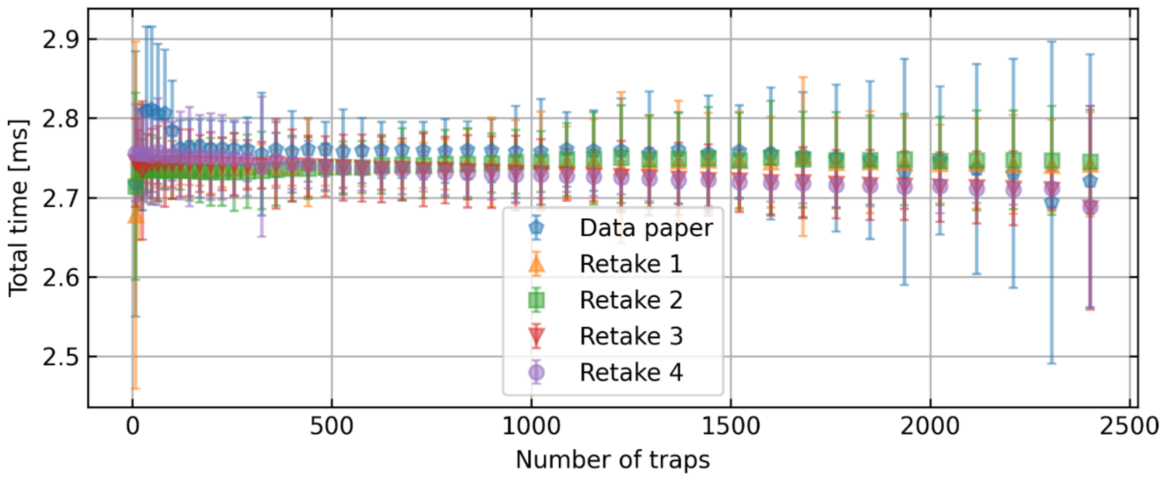
The start-up effects now give rise to an increased uncertainty at the highest Ntw value. The same speed-up of the SLM display for higher Ntw was observed. We do not know why this is the case, but emphasize that this is an effect of the SLM and not inherent to the method presented in the paper.

Plotting all runs on top of each other, we see:

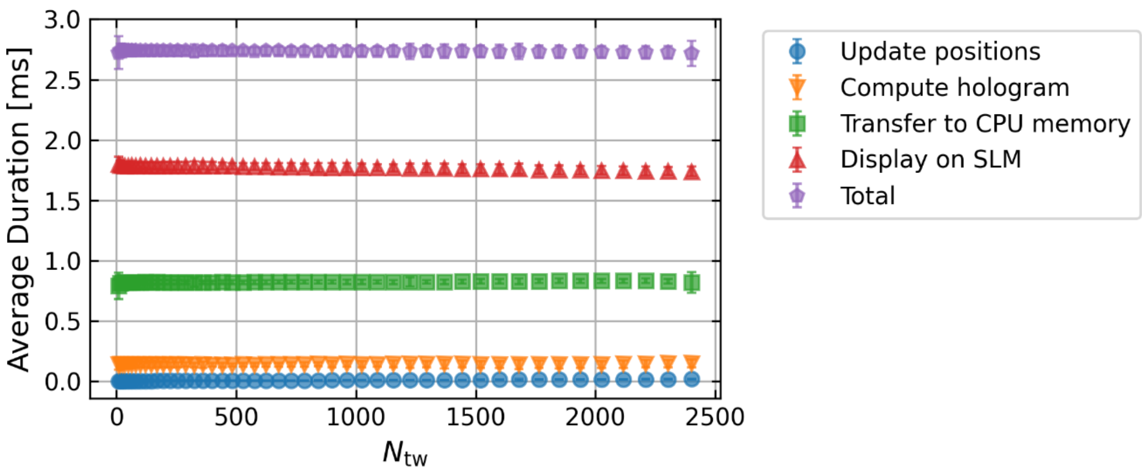


We have not been able to reproduce the same artifacts as the data in the paper. It should be noted that all four retakes agree very well, while the original data is a bit slower in the display (red) and faster in the transfer from GPU to CPU (green) and calculation (blue). We do not know why this is the case.

The total duration per cycle seems to be well reproduced:



In the end, we decided to treat the unreproduced effects as artifacts in the first measurement and rather present the average over the four retake measurements:



With the exception of the start-up effects at the lowest and highest Ntw, the timings stay nearly constant with the small exception of the SLM that we cannot explain. In the paper, we report the mean value for all timings with the RMS error of the mean for all points:

Update positions	0.006(5) ms
Compute Hologram	0.137(3) ms
Transfer to CPU memory	0.821(6) ms
Display on SLM	1.772(15) ms
Total	2.736 (6) ms

Furthermore, we stressed the point that these values do not capture environmental changes and are closely connected to the used hardware by adding as a last sentence in paragraph 3 of Section 4: " We note that the measured timing may vary with different hardware."

5) The final paragraph in Sec. 4 is very qualitative, and not quantitative. The values for 36 tweezers are extrapolated to many thousands of atoms, but unclear is how the scaling can be trusted. What kind of powers are needed for 1000 atoms? What success probability is allowed for scaling up to 1000 atoms? When will multiple rearrangements cycles becomes necessary? Of course, as with any extrapolation there are uncertainties, but to get a better feeling on how the method will work for 1000 atoms, more information is required.

We see the point of concern here and we have quantified the argument. We have included an example for 1000 atoms with our measured rearrangement success rate. We also quantify the main result we take from Z. Zhang, *et al.*, ArXiv: 2412.09780 where they mention a 1/sqrt(N) scaling of the required imaging + rearrangement success probability for square arrays to make repeated rearrangements efficient. With the reported compound probability of 99%, this allows us to have thousands of tweezers.

We will not be able to obtain this number of tweezers with 813-nm light while having sufficiently deep traps. We included this in the text, but also stress that for other tweezer machines that can achieve such large arrays and still obtain survival rates over 99%, we do not see any reason why the proposed rearrangement method would not work.

6) The abstract and the conclusion mention a specific number for the update time, namely 2.72(2) ms. This number seems to be very depending on the equipment that the authors use, and will be different from setup to setup. Given the very concrete value that the authors use, it looks like a generic result for the technique. Perhaps toning this statement a bit down to a few ms, creates a better impression of the capabilities of the method.

We agree that the precise number versus the dependency on the equipment can be misleading. In fact, while we think that with improved hardware and software (e.g. synchronizing the retrieval of the next hologram from the GPU and the display of the current hologram on the SLM) the method can be much faster, supported by the low computation times.

In the abstract, we have addressed this point by adopting the phrasing of "a few ms" with the addition that this is limited by technology.

In the conclusion, we feel this was already reflected by the sentences: "We could update the tweezer positions every 2.76(2)ms, with several options for further speedup. This number does not vary significantly for arrays of up to at least 2400 tweezers and is mostly limited by technological restrictions."