The authors have significantly improved this manuscript since the first submission. In particular, the core claims are now supported in a quantitative manner, and the relative merits of the SPAD array versus other imaging devices is now clearly discussed and demonstrated.

I have one remaining significant concern with the manuscript, which is the definition of fidelity used, and the fidelity numbers presented using these definitions. The authors adopt methods for calculating infidelities from references 15, 16, 21, in which the infidelity is inferred from the probability of observing an void in the first of two images and an atom in the second. This method for inferring the infidelity is not general, and rests upon assumptions that are not valid here, such as the presence of only either zero or one atom in the tweezer (see for example appendix B5 of ref 16, in particular eq. B1).
As an example of the problems that are caused by ignoring this assumption, in the submitted manuscript, P0->1 is used as a proxy for the single atom infidelity, but does not incorporate the probability that two atoms initially occupying the tweezer are mistakenly identified as one. Visual inspection of the red curve in figure 5a indicates that the contributions from one, two, and three atoms are significantly overlapped, so claiming infidelities at the 1-2% level seems highly dubious here. Since the quantitative values claimed for fidelity are central to this work, I think that it is critical to provide a precise definition of what is meant by fidelity (which should include all ways in which a given atom number may be misidentified, such an initial double occupancy being mistaken for a single), and a clear derivation of how this definition is related to experimentally observed quantities.

We have re-analysed our fidelity data more extensively and revised our values.

In the subsequent analysis there are two key values of interest for each dataset. There is the limiting fidelity (F_1) that would be achieved in the regime where light assisted collisions were applied perfectly such that there is no multiple occupancy of the trap, and there is the selective fidelity ($\mathrm{F}_1'(\overline{N})$)) achieved within our dataset for a single atom including all possible misidentifications (i.e. also with multiple atoms). The former reflects the performance of the imaging system only, and is most useful for comparison with other experiments. The latter depends on the mean atom number loaded in the trap, and therefore will vary from experiments.

To calculate the limiting fidelity (F_1), a similar analysis to [16] is performed, where events with N >1 are excluded using a threshold. Therefore the number of infidelity events between 0 and 1 is normalised by the probability of there being a single atom or void in each dataset. The limiting fidelity is calculated in using the following expression:

$$\mathrm{F}_1 \ = \ 1 - \frac{P(01)}{P(0) + P(1)}$$

Which gives limiting fidelities of our 532 nm tweezer setup for the methods used in Figures 5, 6 and 7 of **98.9(3)%**, **98.9(6)%** and **99.8(2)%** respectively.

To evaluate the accuracy with which the atom number can be determined from the measured photon count distribution without parity projection, we define the selective fidelity ($\mathrm{F}_1'(\overline{N})$) . As the atom number is no longer constrained to be zero or one, there are additional possible infidelity events that must be considered, i.e. incorrectly discriminating between a single atom and multiple atoms. We generalise the single atom infidelity error to be when a single atom (N=1) is identified in the second frame, but not the first frame (N≠1). This has an associated probability

$$P(\overline{1}1) = P(01) + P(21) + P(m1)$$

where $\overline{1}$ means N≠1, and *m* refers to all events with N >2 for which we do not attempt to resolve atom number.

However $P(\overline{1}1)$ is not a suitable definition of infidelity, as throughout this paper we separate fidelity and loss into two distinct errors. Terms such as $P(21)$ include significant contributions from loss (i.e.

starting with two atoms in the first frame and decaying to one atom in the second), so these terms would double count events as both infidelity and loss errors. To prevent this double counting we replace $P(21)$ with a term based upon $P(12)$, as the two values should be similar in the absence of loss. Directly swapping the terms would underestimate the number of infidelity events, as the probability of seeing events with one atom in the first frame and two in the second are suppressed by loss. Therefore $P(12)$ is rescaled by the loss rate to prevent undercounting of infidelity events. We then neglect terms where the atom number increases by 2 or more between the first and second frame, as these events are highly unlikely. This leaves us with an expression for the selective fidelity of:

$$\mathrm{F}_1'(\bar{N}) \approx 1 - P(01) - \frac{P(12)}{1 - \mathrm{loss}}$$

This gives the selective fidelities of our datasets used in Figures 5, 6 and 7 of **97.0(2)%**, **97.9(6)%** and **99.1(2)%** respectively. The Sisyphus cooling data (Figure 5) highlights the effect of higher mean atom number of the measured fidelity, as that suffers the greatest decrease in fidelity. This value could however be improved by loading with a lower mean atom number than 1.2, hence the fidelity in the limit of light assisted collisions is the key value that we will use to show the viability of a long working distance tweezer for these experiments.

In response to the referee's comment we also revise how we calculate the fidelity with which can detect two atoms in the tweezer in Figure 5. Loss errors were calculated directly from the data as previously. We now consider the selective fidelity for two-atom detection $\mathrm{F}_2'(\bar{N})$ where the calculation of the fidelity is hampered by the small number of frames with N >2. Therefore the same threshold-based method was applied to the model instead, where the infidelity error can be obtained directly from the fraction of occurrences within thresholds B and C (see Fig. 5a) that were due to the model starting with two atoms. We checked that this model-based approach yielded values for F_1' that were almost identical with those obtained empirically. Using this new approach we now update the selective fidelity for two atoms to be **83(2)%**.

The improved values for the fidelity have been updated throughout the text, in Figures 5-7 and in Table 1. We have also added an accompanying explanation of the improved methods in Section 6 and the Appendix.