“ This is a very solid, understandable and novel paper in the field of non-thermalizing systems and in particular in the subfield of Hilbert space (or, in the classical case, configuration space) fragmentation. It presents a solid demonstration that the strong Hilbert space fragmentation can also exist in higher than one dimension. Additional beauty lies in the fact that the class of models introduced in the paper is relatively simple, but nevertheless quite generic with respect to the considered lattice.”

We thank the referee for the positive assessment of our work, and we very much appreciate their detailed report providing very valuable comments. We have implemented all of them in the updated version of the manuscript.

“My main issue is the following. The main result of the paper would only be rendered interesting if the considered models are non-integrable. Otherwise, the absence of thermalization can be explained by emergent local conserved quantities (in the bulk). The authors neither provide any proof that local conserved quantities are absent in an infinite system, nor do they prove non-integrability numerically by e.g. studying the level spacing statistics. I believe that the absence of local conserved quantities might be shown analogously to Appendix H of https://arxiv.org/abs/2107.09690, generalized to arbitrary graphs and higher spins, although special care should be taken when dealing with higher-spin operators. ”

We assume that by “intgrable”, the referee means a model exhibiting a (possibly complete) set of strictly local conserved quantities, akin to e.g. commuting projector models; indeed, other notions of integrability, such as Bethe Ansatz, do not seem relevant, since we consider systems in arbitrary dimensions and can break translation invariance in ways that do not change any of our conclusions.

First, let us point out that in the one-dimensional version of the strongly fragmented models we study, a full set of conserved quantities labeling the fragmented sectors is known (see ArXiv 1904.04266 and PRB 101.125126). In that case, the conserved quantities are indeed non-local; instead they correspond to string-like operators. Moreover, they are not complete and allow for sectors of the Hilbert space that are exponentially large. It can be shown that at least some of these exhibit thermalizing dynamics within the sector. We expect similar conclusions to hold in the more general models we consider here.

Moreover, as the referee suggests, we can prove directly that our models do not have strictly local conserved quantities. In the following we prove that this is indeed the case for cubic lattices in any spatial dimension d, and for any spin representation. In this setup the arguments presented in Appendix H of SciPostPhys 13.4.098 (in the following cited as Ref. [1]) can be easily generalized. The proof works by contradiction. We assume that there exists a conserved local operator Q, with local and bounded support $\mathrm{supp}(Q)$ contained within a bounded region $\mathcal{R}_Q$. Here, $\mathrm{supp}(Q)$ is defined as the set of sites where Q acts non-trivially. This operator can be written as

$$Q = \sum_{\{\mu\}} c_\mu \bigotimes_{\mathbf{r}\in\mathcal{R}_Q} \lambda_{\mathbf{r}}^{\mu_{\mathbf{r}}},$$

where $\mu \in \{1, \dots, (2S+1)^2\}$ and $\lambda_{\mathbf{r}}^{\mu_{\mathbf{r}}}$ are the generalized Gell-Mann matrices —including the identity $\mathbb{1}$— forming an orthogonal basis for the space of $(2S+1)\times(2S+1)$ complex matrices, and c_μ are arbitrary complex coefficients.

Let us represent all sites that lie inside $\mathcal{R}_Q$ as having coordinates $\mathbf{r} = (x_1, x_2, \dots, x_d)$. W.l.o.g. pick the minimum value of the set of first components $x_{1,\min} = \min\{x_1\}$ and only look at sites with $\mathbf{r}_{\text{face}} = (x_{1,\min}, x_2, \dots, x_d) \in \mathrm{supp}(Q)$, i.e. one particular “face” of the region $\mathrm{supp}(Q)$ traces. Utilizing the +-shaped pattern of the local regions our Hamiltonian acts upon via local terms $h_{\mathbf{r}}$, we can choose a particular site $\mathbf{r}' = (x_{1,\min}, x_2', \dots, x_d') \in \mathrm{supp}(Q)$ and centering $h_{\mathbf{r}}$ at $(x_{1,\min}-1, x_2', \dots, x_d')$; this is one site away from the boundary of $\mathrm{supp}(Q)$. Hence, $h_{\mathbf{r}}$ only acts on a single site of $\mathrm{supp}(Q)$. By the form of the operator and the fact that $[Q,H]=0$, for any non-zero c_μ we must have $\lambda_{\mathbf{r}'}^{\mu_{\mathbf{r}'}} = \mathbb{1}$. However, this is a contradiction to $\mathbf{r}'$ belonging to the support of Q. Therefore, the initial assumption about the existence of a local conserved operator Q is wrong. We added a footnote to the main text to point this out.

1. *Minor issues/typos/comments/suggestions*

1. We have added a line highlighting the diffusive behavior mentioned in the text to Fig. 2(a).
2. We have clarified the definition of the boundary correlation as appearing in the insert Fig. 2(b).
3. As pointed out by the referee, the data points for the latest available time were absent from the insert Fig. 2(b). This was indeed a mistake in the legend displayed, which we corrected.
4. We have fixed several minor mistakes in Lemma 1 and Theorem 1 including those pointed out by the referee:

- $F_v(t-1, t'+1)$ has been replaced with $F_v(t+1, t'-1)$, in the line starting with "In the situation above ...".
- $\Delta m_v(t, t') \equiv m_v(t) - m_v(t'+1) = z_v$ in the line starting with " Note that between the two firings ..." now reads $\Delta m_v(t, t') \equiv m_v(t) - m_v(t') = z_v$.
- There was one unnumbered equation, which now has the correct numbering, and the prefactor of $F_v(t, t')$ was replaced from 4 to z_v.
- v' in the sentence "Between these two times ..." now correctly reads v_1.
- In the sentence "We thus end up with an infinite regress ...", the word "one" was missing, which is now added.

5. Per the suggestion of the referee, we checked our use of the neighborhood set and corrected some inconsistencies thereof.

6. We have replaced $s_{v'} = 0$ by $m_{v'} = 0$ in the first sentence of the proof of Theorem 1.

7. We believe that the updated version of the manuscript, which now includes Appendices C and D in Section 3.3., makes clear the difference between strong fragmentation of the configuration space, and the presence of exponentially many frozen states, which can also appear in weakly fragmented systems. We do so by first showing that the systems we considered do showcase strong fragmentation, and only then discussing the scaling of the number of frozen states.

8. We further want to thank the referee for pointing out a simple lower bound on the number of frozen states. While it is only valid for spin 2, we have included it in the main text as well as in Fig.4.

9. We have corrected and updated the notation regarding the use of **v** and **r** in several parts of the text.