We thank the reviewer for the in-depth feedback that has helped improve this paper. We are happy to make changes in response.

Several figures (8, 10, 11, 12, 14, 16) are not explained or even referenced in the text.
Text has been added to reference and explain all these figures.

How do you manage parallel file formats for distributed simulations with compressed chunks of data, where the size of each chunk might vary and unknown prior to the simulation?
The compression/checkpointing discussed in the paper is all in memory, and not the filesystem. After initial application startup, the only filesystem access is in reading shot data, and that is not a bottleneck in the setup we studied. We added text to the paper making this more explicit, along with justification. Due to this being in-memory, the distributed nature of the computation does not affect the discussion in the paper since each rank is only concerned with its own part of the domain.

Did you compare different compression algorithms, and can you comment on the computational overhead for (de)compressing the wavefield?
These questions were addressed in our previous publication on this topic [1]. We added text to the paper referring the interested reader to this.

Which fields do you compress (pressure, velocity, pressure gradient, …)? Do you apply different tolerances or compression strategies for different fields?
Since we are solving the acoustic isotropic equation, there is only one field present - the pressure field, and we compress it directly i.e. not the gradient. We have added text making this clearer.

Is it possible to a-priori ensure an absolute / relative tolerance after decompression?
This is the definition of atol (absolute tolerance), for ZFP, our chosen compressor for this work. We have added text making this clearer.

Can you elaborate more on the absolute tolerance atol used in the numerical examples? I think it would be better to somehow relate the tolerance to the maximum amplitude of the wavefield / resp. source. The absolute number is rather meaningless.
Figure 7 (PSNR vs atol) is intended to relate a relative error metric (PSNR) to the chosen absolute error tolerance.

When looking at Fig. 7, I am wondering if the frequency content of the (de)compressed snapshot is altered, which is why the error is growing in the first couple of time steps?
We have not studied the frequency content of the decompressed snapshot in this work. Partially because this was previously studied in [2]. This is a work we cite in our paper. Also, after changing the plot to PSNR (to address the reviewer's next point), we believe that the initial increase in error might not be significant in the bigger picture.

Furthermore, is the decreasing trend a simple result from the decaying amplitudes in the wavefield or is this normalized in some way? It would help to also show relative errors per time step.
We have changed the forward error plot to show PSNR to address this.

+++ Minor comments

page 2, line 22:
Referencing an equation long before it appears in the manuscript is bad style.
Reference removed

Furthermore, the equation is not called TTI. TTI refers to the medium / model parameterization or stress-strain relation, respectively, and should not be used as an acronym without introduction.
We now refer to it as “the acoustic anisotropic equation in a Tilted-Transverse Isotropic medium“

page 2, Table 1:
“Forward propagation” is misleading and should be “time steps” instead.
Text changed

Calling it “peak memory” is very misleading for gradient computations because no reasonable implementation would do that.
We understand that no reasonable implementation would do this because this is an unreasonable amount of memory, and practical implementations would employ compromises on runtime and accuracy to make it more reasonable. This is also the point we are trying to drive with that table - that we are proposing one new way of avoiding that exorbitant memory consumption.

page 3, line 48/49:
Either remove the reference to eq. (1) or state the equation here.
Reference removed.

page 4, line 81:
Why is there an asterisk after Louboutin?
Removed

page 5, Figure 2:
Add labels and annotation to make it easier to read. The horizontal axis could count multiples of single simulations.
Figure updated

page 9, line 191:
What density model are you using? I would still consider it an inverse crime if it were a scaled version of the velocity model or even just a homogeneous model.
We have used the Gardner relationship to derive the density model.

section 3:
This section seems a bit disconnected from the rest. I would recommend merging it with section 4. In particular, I don’t see a reason to introduce the subsections of section 4 already here with a single paragraph.
The part of section 3 in question has been merged into section 4.

page 11, Figures 4 and 5:

How does the absolute tolerance of 1e-4 relate to the pressure amplitude? Which other fields do you compress is this really an absolute tolerance and not a relative one?
We have added a note to the plot of the wavefield to highlight that the scale shown is of thresholded values, to aid visualisation. The low values in that plot might be contributing to the reason this question came up.
We have also added a histogram of pressure values to answer questions about the pressure amplitude.
The number we call absolute tolerance is the number we provide as input to ZFP in its field called "absolute tolerance" - which represents the absolute value of the error tolerance we are willing to incur as part of the lossy compression. We specify this at the point of compression. The actual errors incurred might, of course, be smaller. Figure A1 helps to compare actual errors incurred vs provided absolute tolerance values.

page 12, caption Figure 6:
There is a reference missing: "See figures A1 and ??"
Fixed

page 12, Figure 6:
How do you define the signal to noise ratio in this case?
This is defined in Section 3.2

page 15, Figure 10:
The same results are shown again in Fig. 19, so I would either show the entire x-axis here or remove the figure.
Figure updated.

page 15, Figure 11:
What error is shown on the y-axis? And why is it huge?
This figure shows the Taylor-series linearization error in the FWI objective function when using the calculated gradient for linearization. Since the absolute value of FWI objective function in the setup used is so high, these errors, which are actually miniscule on the relative scale, appear much larger. The focus of the figure was not supposed to be on the magnitude (absolute or relative) of these errors - but on the fact that the lossy gradient shows the same behaviour as the reference gradient (there are two curves in that plot - perfectly coinciding). Upon reflection, we feel that this figure is not serving its purpose and only creating confusion since the error shown here is very different in nature from all the other error plots, and is to be interpreted differently. Since this is such a tangent to the flow of the paper, we feel it's best to remove this figure.

page 17, line 269:
Please put atol in context to the maximum amplitude of the wavefields that are compressed.
We have added a histogram of pressure values to make this easier.

page 17, Figure 14:
Should this be "true model" instead of "true solution"? The figure is not referenced in the text.
Text updated to call it true model as well as reference the figure.

page 17, Figure 15:

Some of the recovered structure looks to be significantly smaller than a wavelength for the given frequency content. Could you comment on inverse crime? Are you inverting for density as well (see question on the density model above)?
The resolution of FWI is about 1/10th of the peak wavelength so the result is within the expected resolution

page 20, line 293:
What does atol > 4 mean?
Text updated to atol > 10^{-4}

[1] Kukreja, N., Hückelheim, J., Louboutin, M., Hovland, P. and Gorman, G., 2019, August. Combining checkpointing and data compression to accelerate adjoint-based optimization problems. In *European Conference on Parallel Processing*(pp. 87-100). Springer, Cham.

[2] Di, S., Tao, D., Liang, X., Li, S., Chen, Z. and Cappello, F., Z-checker (0.1. 3) Compression Assessment Report for EXAFEL.