Reply to Comments by Prof Guunar Jeschke

The authors demonstrate convincingly that by using a high-power W-band spectrometer with an only weakly resonant shorted waveguide end instead of a microwave resonator, the problem of level-mixing related perturbation of large dipolar frequencies can be avoided. This problem makes extraction of distance distributions between Gd(III) labels precarious at distances shorter than 3 nm. Hence, the new approach solves an important problem for broader application of Gd(III)-Gd(III) distance measurements. Furthermore, the authors show convincingly that it can be advantageous to avoid excitation of the central transition of Gd(III) altogether in DEER experiments, which comes as a surprise. The manuscript is mostly clear and concise, the data is of high quality, and its analysis is mostly adequate. However, there are a few minor problems that require revision.

We would like to thank Prof Jeschke for his kind words and careful reading of the manuscript and his helpful comments and suggestions.

General:

1. **The text refers mostly to frequency offsets when discussing different excitation schemes, but the Figures use PnOn codes. Please give frequency offsets directly in figures.**

We are keen to keep the PnOn codes as they are used in all the tables and the text, and this scheme was only chosen after examining a number of alternatives. However, we have now added frequency offsets to all the figures.

2. **The claim on similar data quality as with nitroxide labels would have been much more convincing at intermediate distances, where nitroxide labels can convincingly resolve an asymmetry of the distance distribution that is related to flexibility of the ruler backbone. For that, the 2.1 nm may be too short and the 6 nm slightly too long (at the maximum dipolar evolution time achieved here).**

This is a very good suggestion, and (co-author) Prof Godt also specifically suggested we check for this asymmetry. Unfortunately, we do not observe this asymmetry in either ruler, very possibly for the reasons the referee suggests. However, we agree that this will be a good check at intermediate distances and we will add a statement to that effect.

3. **section 65: Gd(III)-Gd(III) RIDME was first demonstrated and the overtone problem noticed in 2014 (DOI: 10.1021/jz502129t)**

This was an unfortunate oversight and we have added this reference.

4. **section 70: "applying the wormlike chain model". I do not find this in either the results section, discussion, or Supplementary Material.**

We agree the intended reference (Dalaloyan 2015) for this statement was not clear and we have now added a sentence to clarify.

5. **section 155: Why did you measure Tm, T1 only at the maximum of the Gd(III) spectrum? These relaxation times are known to differ between central and satellite transitions and you focus on satellite transitions. Please make at least a remark that differences are to be expected and cite literature for that.**

We completely agree measuring T_M and T_1 at the offset frequencies used would have been both helpful and highly appropriate. It was an initial oversight. Unfortunately, immediately following the set of experiments described in the paper, the spectrometer was effectively rebuilt to incorporate a wideband AWG. The lab was then locked down due to Covid and we have not since been in a position to remeasure. We have measured T_M in the same system as a function of offset at Q-band and as expected it shows T_M becoming very significantly shorter with offset in line with results originally reported by Raitsmiring. We also now report in the SI, Q-band measurements with both pump and probe at offset frequencies (on one side of the central resonance). We cannot rule out that the substantial gain in sensitivity observed at W-band compared to Q-band is partly due to a difference in relaxation rates. We would point out that the excellent S/N observed at W-band is despite the expected reduction in T_M.

   We will make it clear that T_M was not measured at offset frequencies, and include a statement that we would expect T_M to be transition dependent for this model system along with relevant reference(s). We will indicate the change in T_M expected based on Q-band measurements.

6. **section 205: For systems other than Gd(III), Tm also can rarely be fitted with a single exponential function. Please rephrase to avoid the impression that this is a peculiar feature of Gd(III).**

    We agree this statement was not clear and we will rephrase.

    7. Section 210: Why did you use fixed exponents 1 and 2 for fitting? I do not see a good theoretical reason for that. Fit quality is hard to ascertain in Fig. S2a (please use a representation as in Figure 3a, black versus red line), However, looking closely I am not convinced that it is good. Please check if the fit residual is reasonably close to white noise.

The fit is actually extremely good ($R^2 = 0.9999$). If we look at the residual, we see a tiny bit of ESEEM at the very start, and then (at very small amplitude but well above the noise) we see a modulation that corresponds to the electron dipolar coupling. The residual is close to white noise at the end of the trace after the dipolar oscillation has decayed.

If we fit to the sum of two exponentials with different exponents, and let the exponents be free parameters, we get best fits where the exponents are found to be extremely close to 2 and 1. There is theoretical work by Salikhov that indicates that when Wt >>1 then the exponent is expected to be 1 and when Wt <<1 then you expect the exponent to be equal to 2 (where W is the rate of the dephasing process). We think it reasonable to suggest that the exponent 2 component is attributable to nuclear dipolar coupling and it is not unreasonable to believe there could be a very fast component. However, on reflection we agree that we should be more cautious in attributing a specific mechanism to this component.

The original basis for our speculation was we previously found (in unpublished results) that this same double exponent fit also offers excellent fits for large multi-spin systems with nitroxide spin labels (independent of sample concentration, but where the fast fluctuation depends on excitation bandwidth). But we haven't yet done similar experiments with Gd.We will use a similar representation as Fig 3a to make the data points clearer and give statistical

measures to indicate the fit is actually very good. We will add comments indicating why exponents 1 and 2 were chosen.

8. **Figure 3a: The fits for P1O1 and P2O2 do not compare favorably to the one shown by Dalaloyan et al. for the same compound (DOI: 10.1039/c5cp02602d). Do you have any idea why that may be the case?**

We would actually claim that the experimental data and fit for 6 nm-ruler are almost identical to the equivalent experimental data and fit of Dalaloyan at W-band (allowing for differences in modulation depth and S/N). It maybe the referee, was actually looking at the experimental data and fit for measurements made at Q-band, given in the same paper, for the same system. This does give a different modulation and, visually, a better fit. It also gives a different distance to that found at W-band. Previous computational modelling in Manukovsky 2017, that took into account pseudosecular terms, had shown that small distortions are still expected even at 5 nm distances, under experimental conditions comparable to P101, P202 (and for the Q-band data shown in Dalaoyan). One thus might expect Q-band data to be more susceptible to pseudosecular effects as the central transition is broader, and that has been our experience for other measurements with Gd spin-labelled systems (not shown). So we are more confident about the W-band data.

9. **section 265: "We do not expect any orientation selection and so the Pake pattern spectra reported in Fig. 6a show strong distortions". This sentence is unclear. Please rephrase.**

We were trying to convey that the distortions observed are not expected to be due to orientation selection (as might be the case for nitroxides). But we completely agree the current phrasing is unclear and we will change it.

10. **section 310: I recommend more cautious wording regarding the origin of the second component, as you do not have a comparison with a sample with only a single Gd(III).**

The current phrasing ("we speculate") was meant to convey a degree of caution, but as discussed above, we will change the wording so we do not attribute a specific mechanism to the second component.

11. **Figure S3: How can you have a rising background for P3O3? This needs to be commented.**

We had decided to include this background to be consistent with the methodology that was used in determining the background for all the other traces, together with a a comment in the caption stating this was unphysical. It is not completely clear why the background appears slightly different for this trace, compared to others. However, we agree that it casts unnecessary doubt on the results and we have now changed the fit to a slightly falling background.
We would stress this leads to almost exactly the same distance and distance distribution, but a slightly less perfect Pake pattern. We thus feel it does not change the underlying argument. We would also stress that the background in all the measurements is extremely low relative to previous measurements made on the same system (e.g. Dalolayan 2015). This reduces uncertainties in determining the distribution. The discrepancy probably arises because the time trace was not long enough to accurately determine the background as the oscillations have not completely died out. We will add explanatory comments.