PHY340 Data Analysis Feedback:

Group P07 doing Problem P6

# Data Analysis

The introduction to the data analysis is confused and misleading: you state that your equations 1 and 2 “allow the calculation of the coherence time *T*2, through the use of a Fourier Transform”—but in fact your equations 1 and 2 aare already in the time domain, and there is absolutely no need to use a Fou­rier transform to calculate *T*2 (one simply fits the equations to the data). A Fourier transform is needed to calculate the *line width* corresponding to this coherence time, not to calculate the coherence time itself.

Apart from this, your account of your data analysis is perfunctory in the extreme. You say that you wrote a Python script to identify the maximum and minimum intensities, but you do not explain how this script actually worked (it is of little interest to the reader to know that the code was written in Python: what the reader wants to know is the algorithm that you used). You state that “the script proved robust in dealing with noise”, but you do not explain how you tested this nor provide any evi­dence to justify it. Nor is noise the only issue: some sets of fringes, such as those shown in figure 1, have systematic trends over the whole sample. In this case, taking the maximum (from the central fringes) and the minimum (from the outer fringes) is clearly going to lead to an overestimate of the visibility. There is nothing in your two sentences to indicate that you considered such cases.

 *Figure 1: Fringes from the 0 mm RF run. There is a clear systematic trend in these data, and simply identifying the overall maximum and minim will overstate the visibility.*

Furthermore, the evidence of your figure 1(a) shows that in fact your algorithm, whatever it is, is *not* robust against noise as you claim. It is clear from this figure that your PL visibility values do not tend to zero at large *τ*, as they should for a Gaussian distribution, but in fact tail off to a constant value of about 0.05. The reason for this can be seen from figure 2 below: even where there are no visible frin­ges, noise gives the distribution a non-zero width, and any algorithm based on finding maxima and minima will yield a small but non-zero visibility.

*Figure 2: PL 0 mm data. There are no fringes here, but the noise band has non-zero width.*

This is something that needs to be taken into account in your fitting procedure, since any attempt to fit a pure Gaussian will be pulled away from the correct parameters by the non-zero tail. The simplest solution is to fit a Gaussian plus a constant,
instead of just a Gaussian. You have not done this, and as a consequence your Gaussian fit does not properly describe your data, and your *T*2 value is incorrect. This effect can be seen (using my data, which have the same feature as yours but to a lesser degree: your constant offset is about 0.05, mine is clearly less) in figure 3 below: note that neglecting the constant increases *T*2 by nearly 20%.


*Figure 3: fits to the PL data.
Solid line, fit to Gaussian plus constant, giving t0 = 369.9±2.0 ps and T2 = 179±5 ps;
dashed line, fit without constant, giving t0 = 371.1±4.6 ps and T2 = 213±9 ps.
The constant offset is 0.036±0.004.*

This brings up a key point: you can fit *any* function to *any* set of data, but if the function is not appro­priate the fit will not mean anything. Even though you have made no attempt to calculate error bars for your data points—which you should have done—it should be immediately apparent from looking at your figure 1a that *your PL fit is the wrong shape*. Clearly your data have a narrower central peak and wider wings than your fit. Interestingly, Makhonin at al. (2014), in your figure 1b, clearly *have* included a constant offset in their Gaussian fit, although their data do not seem to need one.

I think your visibility data are quite plausible, but I can only make this judgment because I have done the analysis myself and your data look similar to mine. Based on your own account, or lack thereof, I would have no confidence at all, because you have not supplied any useful information. Your expo­nential fit to the RF data appears reasonable, but your Gaussian fit is not, for the reasons discussed above. You yourselves do not provide *any* discussion of your results: Section 4 is missing altogether, and Section 5 quotes the values obtained by Makhonin et al. (2014) and not your own values—al­though, to be fair, this does appear to be a mistake, as the Δν values quoted in the same paragraph are based on your values of *T*2. There is no discussion of goodness of fit: ideally, you should quote a χ2, though in fact you cannot do this because there is no evidence that you made any attempt at all to determine the errors on your visibility values. You really should have done this, because the question of whether the models you have used are actually good descriptions of your data cannot be answered without error bars on the points.

I think that your report sells your analysis short: you probably *did* think about many of these issues, but there’s no evidence *in your report* that you did so, and your report is all I have to go by—just as in the real world the paper you write for an academic journal is all that the rest of the scientific com­mu­ni­ty has to go by. You need to think more carefully about the evidence that you need to present to back up your conclusions.

Average mark for this section: 30/50

# Data Presentation

The first obvious point to make is that if you are using colour to distinguish the lines in your plots, don’t hand in a black and white print-out! It is quite annoying to be told that the blue curve is PL and the red curve is RF, when what you actually have is two shades of grey. In general, it is good practice to ensure that your curves are distinguished by more than just colour (e.g. make one a dashed line), because some people *will* print out your work in greyscale (including print versions of journals) and a significant fraction of people (about 10% of men, 1% of women) are colour-blind to some degree and may not see the colours as you intend. In other respects your plot is well presented, although it is a pity that you did not find a plotting program that can render scientific notation properly (as 10−9 ra­ther than 1e−9), or simply opt to express your time axis in ns instead of seconds.

As noted above, there should be plots from the earlier stages of your work, to explain and justify your procedure and to illustrate potential systematic errors. You also need to make it more explicit that your figure 1b is taken directly from Makhonin et al. (2014), is not your work, and may involve a dif­ferent sample (though it’s almost certainly the same experimental setup). You do reference Makhonin et al. in the caption, but you don’t make it clear that the figure is taken directly from the reference. This is, of course, compounded by the fact that you subsequently claim their results as yours, but I am prepared to believe that this is a mistake (albeit a serious one—it amounts to scientific fraud—and one that should have been caught by any competent proof-reading).

It is good to see your results quoted with their uncertainties, although you do quote too many signi­ficant figures: the *T*2 values should be given as 241±14 ps and 662±39 ps (at most: 660±40 would be better), and the frequencies as 220±12 GHz and either 481±28 MHz or 0.48±0.03 GHz (either would be reasonable). Having presented Makhonin et al’s results in your figure 1b, you should make a com­parison between your results and theirs, though this is made more difficult by the fact that Makhonin et al. do not seem to have used the same Gaussian form for their fit as the one they gave you, see figure 4 below. I *think*, but cannot quite prove, that in the paper they actually fitted
i.e. without the factor of *π*/2 in the exponential.



*Figure 4: Fringe amplitude against time delay, from Makhonin et al., Nano Lett.* ***14*** *(2014) 6997 −7002 (their figure 3a). Note the presence of a constant offset of about 0.05 in their Gaus­sian fit, though their data do not obviously re­quire this (the last few points are below the fit). Super­imposed on this plot are functions of the form , with T*2 *= 154 ps (dark blue dashes) and 185 ps (dark red dashes): it is clear that T*2 *= 185 ps reproduces the “T*2 *= 154 ps” line in the paper, whereas T*2 *= 154 ps does not. This shows that the function fitted in this paper has a different form from that given in the instructions.*

Average mark for this section: 19.75/30

# Style

You do seem to know what is expected of a scientific report: your report has the correct overall struc­ture and appropriate section headings, and your abstract is a fair summary of the paper. It is also good that you have mastered the LaTeX word-processing system, which is much better than Word for scientific documents. How­ever, the whole report is seriously in need of careful proof-reading—in­deed, of any proof-reading at all. The most obvious sign of this is the complete absence of Section 4(!), but there are other signi­fi­cant problems: in the Fourier transform derivations on page 2, the derivation labelled “PL” actually re­fers to RF, and vice versa, and I have already mentioned the fact that in Section 5 you claim Makhonin et al’s fit results as yours. An entire paragraph at the bottom of page 1/top of page 2 is repeated ver­ba­tim. These are not things that require detailed inspection to spot.

There are also some less obvious issues. At the bottom of page 1, you say, “It is clear that resonance fluorescence will give a longer coherence time from these equations.” That is simply not true. Equa­tions 2 and 3 can both accept *any* non-zero value of *T*2: there is absolutely no requirement that the time constant of an exponential be larger than the standard deviation of a Gaussian. You did not think this through before writing. Again, during your Fourier derivation you state:

“At *ν* = 0, the visibility is at a maximum and *V*(*τ*) = 2*T*2. Solving for *T*2 gives

This is just nonsense: if *ν* = 0, then ! What you actually mean here is something like:

Equation 7 shows that *V* is maximal at *ν* = 0, at which point *V* = *V*max= 2*T*2. The FWHM is then given by the frequencies at which *V*(*ν*) = *V*max/2 = *T*2, i.e.
Solving for *ν* gives , or
The bandwidth is the difference between these two frequencies, i.e.

It is not clear whether you understood this but did not write it down clearly, or did not understand it. It should be noted that *every one* of your integrals is missing its integration variable: something like
***does not mean anything at all*** without the d*x* at the end. This is particularly critical in Fourier trans­forms, where you have two variables (*ω* and *τ*) and are integrating over only one of them.

Also, note that scipy.optimize is an entire package, not a single algorithm. I assume that you used the scipy.optimize function curve\_fit, since that’s what you were shown in the lectures, but the package contains other fitting routines besides this. In fact, it is more important to tell your reader what the algorithm is than what the actual package used was: there are many other least-squares fit programs available. You should say something like “the curve\_fit function from the Scientific Python package scipy.optimize. This is a non-linear least-squares fit using the Levenberg-Marquardt minimisation al­go­rithm (see, e.g., *Numerical Recipes in C* section 15.5).” With this information, a reader who does not use Python can identify an appropriate alternative, e.g. in Matlab.

This is all very unfortunate, since I think that your actual analysis was probably quite good, but you have thrown away at least 10% by doing a sloppy job of writing it up.

Average mark for this section: 12/20

Overall average mark: 61.75%