Comment	Response
The authors report a series of new experiments and observations examining the etching rate of fission tracks compared to the surrounding crystal. Overall, though I disagree with some of the authors' methods and conclusions, I believe the science will be moved forward by reporting their data and analysis. That said, I have a number of comments and suggestions that I hope can be used to improve the work, with some nudges that I hope can move the discussion forward. I also strongly urge that the authors publish their raw (track-by-track) data in tabular format.	We are obliged to the reviewer for his thought-provoking comments. We will address differences of interpretation as we encounter them. Although we preferred to wait until the first reviews were in, we will now of course make our raw data available.
Part 2 is a very ingeniously constructed experiment to try to measure along-track etch rate, and is a valuable contribution. One very interesting result is that they document the possibility of an etching time deficit in step-etch experiments, where time is essentially lost by the etchant having to re-traverse the tracks; this is a new insight, as far as I can tell, and will be a necessary consideration going forward to experimentalists studying etching.	We thank the reviewer for pointing it out. Trilsch et al. (2023; Figure 8) and Jonckheere et al. (2024; Figure 6a) illustrate the resulting spread of effective etch times. We omitted to refer to them in the text, but intend to include them in our revised manuscript.
However, in terms of inferring whether along-track etching rates vary, the authors' interpretation is disputable. The residuals about the 20 s line in Fig. 3a appear to be clearly structured, with most points below the line at 5-10s and above it at 15s. Assuming the surface point is well-anchored as the authors claim, it appears that the data would be better fit by a concave-upwards curve, which would correspond to a diminishing along-track etch rate. The authors may disagree, and I would not hold up this paper if they do, but I would (again) strongly recommend that they publish the data for Fig. 3A so others can evaluate the possibility.	As each anchor point is the average of 363 to 514 measurements of the same feature (Dpar; same ion and orientation), its error is almost 20× less than that on a single measurement. A concave-upwards trend means that the diamonds at greater distance from the surface are larger than for a constant v_T. Larger means longer effective etch times and shorter access times, which in turn implies a greater average v_T between their surface and the crack intersections, i.e., a v_T increase towards the tip.
There are also some aspects of the experiment/data presentation that would be good to clarify. The authors use the variable "l", which usually means etched track length, for an extrapolated point based on the diameters of the etch figures. As shown in Fig. 2a, these extrapolated points appear to lie beyond where there is a visible track (though this may be in part due to the plunge of the tracks), and for a normal fission track that has an ending point the extrapolated length will always exceed the real etched one. Accordingly, I suggest using a different variable name (E, le, or el for extrapolated length?) to avoid this potential confusion.	Corrected.
It's a little surprising in Fig. 2a, b that all "etch figures" appear to be in focus at the same time, given that the interior figures are up to ~5 µm deeper than the surface ones; is this just the way it was, or are these images composited depth stacks or something like that?	Corrected: the images are compressed stacks; we omitted to mention it in the Figure caption.
It would also be good to specify exactly how dS and dD are measured, given that a circle is being used to measure a diamond-shaped feature. For dD, is the circle drawn so it is tangential to the diamond faces (which looks to be the case in Fig. 2b, but it's hard to be certain), or is it drawn so its equator ends at the diamond tips? In the surface tracks, the beginning of the track is above the midline of the circles, by perhaps subtly differing amounts; what criteria are used to locate dS circles?	d_S and d_D are the diameters of circles tangent to the outer edges of the diamonds, centred on their midpoints (d_D) and on the line connecting the tips of the elongated-hexagonal surface intersections (d_S). They are in that sense not true track dimensions. However, our calculations do not require them, but only their relative sizes (Figure 2). Circles are both practical and permit to eliminate variation of the measurement direction when attempting to measure across from one corner to the opposite. This is much clearer in the original images, but we will add an explanation as well. We believe this has made our v_T-measurements more precise than in earlier work.
It's noteworthy that the assumption is that etch figure size is a linear function of etching time,	The fact that the mean etch pit size increases less in relation to the immersion time

but dS increases at 0.088 μm/s in the first 20-s step, and 0.110 μm/s in the next two steps; this deserves some discussion and explanation.	during the initial 20-s etch, than during the subsequent 10-s immersions does not have to be due to different etch rates; it can also be due to what has been called an etch induction time (lag time; Jonckheere, 1995; Jonckheere and Van den haute, 1996, Figure 19). Other investigators made similar observations. We have speculated whether this was a chemical effect (surface contamination), due to polishing damage, or track formation, e.g., a shell of denser material surrounding the latent track, or annealing of defects in the surrounding apatite due to the heat from track formation. We have investigated without success if the latter case would permit to arrange tracks in chronological order. As we cannot even distinguish between rate and duration, we believe it would be an unhelpful distraction to speculate in our manuscript.
Finally, are the points A*, B*, and C* equivalent to l_M in Table 1? If so, it would be good to make this clearer by stating it directly in the figure and/or table caption.	Corrected.
Figures 4 and 7 need scale bars.	Corrected.
The unchanging angular frequency observed by the authors in Fig 5d is only slightly surprising (line 151-2). Ketcham (2003) observed no significant change in angular frequency until "accelerated length reduction," the faster relative shortening of tracks at high angle to the c-axis (Donelick et al. 1999) set in. This effect only appeared in pooled experiments with mean c-axis projected lengths of 12-13 μm. The authors' results for experiment 21-3 (their Figure S01d) are in the midst of this range (lc = 12.4 μm), and shows only a handful of these tracks, but considerably more on the "elliptical" trend. Evidently, accelerated length reduction is in a sufficiently incipient state in this experiment that there was not a significant loss of high-angle tracks.	This comment does not appear to necessitate a correction. We referred to the earlier observations of Ketcham (2003) on lines 152-153. Whether one considers the invariant angular distributions surprising or not is a matter of expectation. Even so, before the accelerated shortening of high-angle tracks, the length difference between those parallel and perpendicular to the ***c***-axis is ~30% (Figure S01d). One could expect that pure length bias would have a noticeable effect at that stage, considering that we measured horizontal confined tracks. We therefore suggested that it might not be the sole determining factor in confined-track selection (lines 163-165).
The intersection point distribution is also only somewhat surprising (line 154-155), but maybe not for the reasons the authors state. Any deviation from a uniform distribution of intersection points should mainly a function of the analyst avoiding tracks where a tip is obliterated or obscured by the etchant pathway. The overall pattern in Fig. 6a bears this out, with a flattening of the CDF as the intersection points approaches zero (the track tip). From an unannealed starting point, as the track shortens, this near-tip flattening zone should expand (in the normalized along-track coordinates used), as a higher proportion of the track is near a tip. There might be a hint of this in Fig. 6a, but it seems smaller than expected, given that mean track length has fallen almost 40% (from 16 to 10 μm). So, it appears there must be some countervailing bias. The most likely culprit is etching rates and track length; in a track that is longer and/or has slower etching, the far tip might be less likely to be fully etched, which would have the effect of expanding the size of the low-probability, near-tip region. This is fully consistent with the analysis in Ketcham and Tamer (2021).	This discussion does not appear to necessitate a correction. It shows how difficult it is to interpret rather straightforward data concerning a system that is supposed to be so well understood that it is used for retracing the thermal histories of geological samples. The reviewer lists two probable contributing factors that are hard to put numbers to: obscured track tips on the side of the host track intersection, and underetched tips on the opposite side, the latter dependent on the track length and etch rate. This is related to our interpretation of the 2-s difference of the median effective etch times of 21-2 and 21-3 in terms of their lengths (Figure 6c; lines 199-208). There is however a buffer to consider: the time to bridge the distance between the host track and confined track at the apatite etch rate v_R. Our simulations show that this takes on average ~⅕ of the immersion time for a standard etch protocol, and a track length of 16 μm and etch rate of 2 $\mu m \cdot s^{-1}$ (Thermo 2023; unpublished). In terms of fractions of the total immersion time, it is the most variable component. In other words, shorter tracks need on average less time than longer tracks to be etched to the point of selection; this time is however invested in accessing confined tracks at greater distance from the host track, while the confined-track properties are little affected; short and long tracks etch for a time commensurate with their lengths. Our modelling shows that, under stressful conditions (10 s etch), confined tracks are indeed more often intersected near their midpoints, and under more relaxed

	conditions (20 s) samples of non-horizontal confined tracks, etched with a standard protocol, are more often intersected near their upper endpoints. This is not the case for perfectly horizontal tracks although it is not a priori clear what to expect. It nevertheless demonstrates that various factors must be considered. Until we have a better understanding, we prefer to present our data but refrain from speculation.
In Figure 5b and line 247 (etc.), lP evidently means ***c***-axis-projected length of a single confined track, whereas in section 2 (Fig. 2c; line 74) it corresponds to surface-projected extrapolation of length inferred from etch figure sizes – an example of the confusion in variables that should be avoided.	Corrected.
Some would argue that change in median Dpar from 1.89 µm to 1.80 µm (line 215) is unlikely to be significant, given the limitations imposed by using visible light (wavelength 0.38-0.75 µm, leading to a measurement resolution limit of ~0.2 µm). I'm willing to believe there probably is a real change, but the uncertainty of the ratio of two numbers so close to each other, with substantial measurement imprecision, will be very high, which makes it very questionable to use that ratio for any sort of kinetics calculation; the interpretation (lines 209-228) seems like a stretch. At the very least, uncertainties should be propagated to show the range of possible values, but I have a feeling that it will demonstrate that the analysis is uninformative, and thus does not add anything to the paper. Additionally, the figure reference on line 212 should be to 6d, not 5d.	We explained in answer to a comment of reviewer #1 that the resolution, but not the precision, of microscopic measurements is limited by the Abbé rule. The dark borders of the diamond and triangle shapes in Figure 2a are about ~0.3 µm wide, but nothing prevented us aligning our circles with their outer edges for measurement. Skeletonization using common image software allows to extract a one-pixel centre line. Due to a few outliers, the median is a more robust central estimate than the mean in this case. The difference between the medians (1.89 µm and 1.80 µm) is 6 standard deviations of the median (0.015 µm; we will provide full data with our revised manuscript). Due to the small relative errors (<1%) on the median-Dpar, the activation energy increase due to annealing is still known with meaningful precision: (1.22 ± 27) eV per atom (1σ). The error on the pre-exponential constant for defect annealing k_D (equation 1) is accordingly large (23%), but that on the activation energy for annealing E_D is small (~1%), due to the favourable form of equation (2), and disregarding the unknown error on k_0. We believe that, given the equations, anyone interested can perform or model the error propagation, and we prefer not to interrupt our text with explicit statistical excursions.
To conclude my comments on sections 2 and 3, I've said in previous reviews of the authors' work that their v_T comparisons are somewhat compromised because their v_T measurement method presumes that it is unchanging along the track (which I don't believe their results demonstrate, including in this study).	Yes and no: a v_T-calculation for an individual track gives the average value between the surface and the crack (Figure 2), whether v_T is assumed (or is) constant along that section, or not. A different track gives the average v_T over a different section; Note that each calculation is independent, if not the underlying reality, due to section overlap. In principle, it is possible by differentiation to reconstruct v_T along the track, but not in practice because the calculation is unstable and too susceptible to noise. The assumption comes in when the data are fitted; as explained. In view of the data (Figure 3a) and other evidence (see text), it is appropriate to fit a linear trend.
That said, it's very reasonable to argue that the particular etching structure(s) proposed by Ketcham and Tamer (2021) may be inconsistent with some of the data presented here, but it should be kept in mind that the variable etching structures examined in that work were intentionally simple end-members, and other structures are possible, particularly in how etching	We do understand the limitations of models, and must own up to some modelling of our own. However, when a model conflicts with reliable data, the model should give way first. "*Even if experiment fails to confirm the model, the model remains valid for the condi-*

evolves with annealing.	*tions it was intended to describe*". (*after* Per Bak (1996) How nature works: The science of self-organized criticality). One wonders if the author drew comfort from his strange claim.
Figure 6b indicates that individual along-track etch rate, as determined by the authors, can be anything from 25 to >250 μm/min, or a range of a factor of 10 or more. They cite a prior work claiming that track orientation may be partly responsible, but even within a confined c-axis-angle range a factor of 4 variation is observed (Jonckheere et al. 2024, Fig. 5c). Is this remaining dispersion attributable to measurement uncertainties, or does it truly reflect a variation in etching rate, and if so what might be its cause?	We have no clear answer. The data in Figure 5c of Jonckheere et al. (2024) are limited and include three etch steps. The measured etch rates and the scatter increase at each step, due to finite rate at which the etchant advances through pre-etched track channels. We should thus focus on the initial step (15 s) for comparison with the present data (20 s). There is, in our opinion, a convincing angular dependence with minima (~25 $\mu m\cdot min^{-1}$) parallel and perpendicular to the ***c***-axis, that comes out in a quadratic fit. The v_T-values in the centre range from ~75 to ~125 $\mu m\cdot min^{-1}$. Part of that is doubtless due to the limited precision of the measurements as v_T is difficult to measure, but how much? We can prove that v_T-variation is real; there are substantial differences from track to track observable in images even without measuring. The data discussed above indicate that there is also a definite angular dependence. It appears that etching of damaged (track) and undamaged apatite is anisotropic with a similar angular dependence. However, that doesn't explain anything. One could, on the other hand, consider a correlation with the latent track diameters.
The large change in etch figure size between fossil and induced tracks in the Duluth samples, which the authors hypothesize is due to radiation damage, is interesting and unexpected (e.g., line 338). We have looked for such an effect in my group, and not found it. In our example with the highest dose, Tioga apatite, which has an AFT age of ~220 Ma and has ~35 ppm U (and thus a similar dose to Duluth, with ~10 ppm U and an AFT age of ~900 Ma, though we have not measured Th), we observed absolutely no difference in Dpar (2.50 μm induced, 2.48 μm fossil). It's possible that our preannealing (450 °C for 48 hours, twice the time the authors use) was nevertheless inadequate because of the higher resistance to annealing of Tioga apatite (with respect to fission tracks). We've also not seen an effect in Durango and Fish Canyon apatites; though the doses in these are much less, they imply that any radiation damage effect on etching will be limited to infrequently-encountered doses. Also worth noting is that Iwano et al. (2018) report unstoichiometrically high halogen contents (F ~ 5 wt%, where the maximum possible should be 3.88 wt%, and a standard deviation of 0.7 wt%, plus 0.6 wt% of Cl to boot), and so there's considerable uncertainty as to the possible OH content, and its homogeneity, which could also lead to varying etching properties in different aliquots. Figure 1 in Iwano et al (2018) shows appreciable zoning in both apatites. Accordingly, it would be good for the authors to do more to document and test their analysis. Measures might include reporting the individual Dpar data on a grain-by-grain basis, and perhaps seeing if there is a relationship between Dpar and track density; Iwano et al. (2018) report U concentrations varying by a factor of 3-5 from grain to grain, which might be a large enough range to detect a relationship between Dpar and (inferred) radiation damage, if there is one.	We were unaware of the reviewer's work on Dpar; if there is a citable document we would like to refer to it. It is important to be mindful of the visual effect of the images; although the contrast is striking, it is in part due to the higher fossil track densities and to the large Dper-"increase", in contrast to the 17% Dpar-difference. We can exclude chemical differences between the grains with fossil and induced tracks. Both mounts contained splits from the same aliquots of FC1 and AS3, provided by Hideki Iwano (we will include an acknowledgement). We confirm that there are Dpar-differences from grain to grain in the fossil-track mounts; that much is obvious from a superficial inspection. Whether this variation is correlated with the track densities is not so obvious; we will investigate this further. For now, we have too few data. We propose to refer to the reviewer's results (pers. com.?) and to moderate our conclusion.

Sincerely,

R. Jonckheere
F. Trilsch
Jie Liu
Pengfei Zhai
T. Nagel