Authors' Comments

We again thank the editor, reviewers, and community members for their commentary on our manuscript, "Chemical abrasion: The mechanics of zircon dissolution" that has ultimately helped to strengthen this contribution. We made the revisions that we outlined in our response-to-the reviewers document, and we address these revisions point-by-point in the comments below. Reviewer comments are in black text, and our responses are in blue text.

Reviewer 1 – Fernando Corfu

Overall the paper is well prepared and structured. I find, however, that the present version could be improved by a good weeding of unnecessary words and sentences. Some expressions that the authors like to use frequently should be reconsidered. One example is 'compositional zone'. I have made some suggestions in the file and discuss specific points below.

We have tightened the language in the manuscript and removed the use of the phrase "compositional zone." Overall, the manuscript is >1500 words shorter. We address more specific wording suggestions below.

Although the results are very relevant for zircon U-Pb geochronology, the studied zircons have not themselves been dated, and so the specific data do not have a direct connection with the U-Pb behavior of the samples. Consequently, the discussion of implications for U-Pb dating presented in the paper is very generalized and in part trivial. I suggest to concentrate on the main substance of the study and avoid meandering off in inefficient discussions.... Section '4.2.1 Zircon U-Pb ages' lines 648 – 726: I find the discussion on this topic very generalized and superfluous. The experiments in this paper demonstrate the great variability between zircon of different geneses, compositions and ages. Yet many of the reflections made here focus on some idealized magmatic zircon in young systems. Because of such a simplification the discussion is trivialized and almost meaningless. Clearly, the lessons from the present experiments must be considered separately for each set of zircon used for geochronology. I recommend removing this section, it just detracts from the paper.

We thank this reviewer for these comments. In a revised contribution we would refocus the manuscript on the *textural evidence* for the mechanics of zircon dissolution more strongly since U-Pb data are not included in this piece. We have shortened the U-Pb discussion section (new section 4.2) significantly to avoid over generalization and more strongly emphasize that the effectiveness of any chemical abrasion protocol will

ultimately be sample-dependent (lines 33, 623-628, 733). However, since many ID-TIMS zircon studies do target magmatic zircon, we feel that a brief discussion about implications for the dissolution of an idealized concentrically zoned, magmatic grain is worthwhile. Since no other study in the literature to date has systematically documented how effective chemical abrasion is at dissolving inclusions, the section on inclusions is also important to eliminate.

Line 64; '... poorly understood, and several outstanding questions remain. Do most zircon crystals predominantly dissolve from rim to core?' Everyone who has done chemical abrasion will have noted very rapidly that zoned and metamict zircons do not dissolve that way. That can hardly be described as a poorly understood fundamental question.

We have rephrased our study motivation in the introduction and eliminated the phrases that the reviewer cites (lines 69-87). The reviewer is correct that many practitioners of chemical abrasion are aware that many zircon crystals dissolve heterogeneously. However, apart from a small number of SE images of chemically abraded zircon crystals that indicate that acid sometimes reaches grain interiors (Mundil et al., 2004; Mattinson, 2005; 2011), the mechanics of zircon dissolution are poorly documented in the literature. In the new wording, we emphasize that no study to date has systematically documented zircon dissolution textures given a range of zircon types and leaching conditions, nor leveraged such findings to gain a mechanistic understanding of the microstructural processes that occur during partial dissolution in HF.

Fig.2: The explanations are unclear. Are the grains shown in each of the four panels representing each of the sample? If yes, why does the second panel contain zircon from two different samples (blue and yellow circles). If not, what is the distinction? Overall I find this figure quite useless, not even as a decoration.

We have kept this figure at the suggestion of other reviewers in part to highlight the color change that occurs during chemical abrasion. However, the labeling has been updated to better illustrate how samples are distributed in b).

Fig.3: Not sure about using the term 'metamict' once zircon has been annealed. Can be confusing.

We have changed the wording from a 'metamict' rim to a 'high-damage, CL black' rim.

209: 'typical magmatic growth patterns' It is true that many magmatic zircons have oscillatory zoning, but the same pattern can in part be seen in metamorphic zircon. Better to use descriptive terms.

We have changed the wording to more descriptive terms (lines 222-225).

287: ' This could imply that increasing the duration of the leaching step results in a more crystalline zircon residue due to the progressive dissolution of higher damage domains.' Why the 'could'? It seems to be the most logical explanation. And then the following sentence; 'We note, however, that only a small number of AS3 crystals survived 12 h of chemical abrasion, and only a small number of Raman analyses were made. We recommend further study to better evaluate this possibility.' Sounds rather trivial.

We have removed the word 'could' and the lines that the reviewer suggests.

316: 'As evidenced by our SE images and discussed further below, μCT does not capture radiation damage zoning that does not result in a strong density contrast such as variations in radiation damage below the ~1×1018 α/g threshold.' Suggest rephrasing to avoid the double negatives.

We have rephrased this sentence to eliminate the double negative (lines 317-320).

Fig. 9, caption: 'interior compositional zone'. Compositional zone? As opposed to what?

We eliminated the use of the term 'compositional zone.'

367: ' All observed dumbbells are oriented parallel to the grain's c-axis'. That is a surprising statement. Looking at the figures I would have assumed that they are all normal to the elongation of the crystal (= c-axis). Please elaborate to avoid confusion

The reviewer is correct. Our original description is inaccurate. We have rephrased this line accordingly (line 370).

Fig 13 caption: 'The yellow arrow highlights the grain's shell-like appearance because of significant dissolution in the grain's interior.' I see a highly resorbed grain, not much left of the shell.

We feel that our description of the grain's "shell-like" appearance is still appropriate.

'images of dog-chewed zircon residues ' maybe a bit too colloquial?

We have eliminated the colloquial wording (line 406).

419: 'In a visual game of connect-the-dots ...' ??

We have rephrased this line (lines 414-416).

422: 'We see dumbbell-like fracture patterns again in sample Zr36 (Fig. 13b-III) where crosscutting fractures connect different oscillatory zones removed by dissolution to one another and to the grain surface'. Rather convolute sentence, hard to understand. Please rephrase.

We have rephrased this line (lines 413-418).

483: '... The long axes of deep, octahedral etch pits on (100) align with the crystal's c-axis...' I wonder why they are called octahedral. Those in Fig 16a look prismatic to me.

This line actually refers to Fig. 17. The figure reference number was inadvertently mislabeled and is now corrected. We have also changed the "octahedral" description to "prismatic."

Fig. 19, caption: ' Projection on (100) looking down the a-axis. The c-axis is vertical to the page, and the a2-axis is horizontal' Confusing: should it not be: 'Projection on (100)

looking down the c-axis' ? The same for (b): if the pane is 001 then the view must be parallel to a? Or not?

We have simplified the figure caption; however, the description "Projection on (100) looking down the a-axis is correct, since (100) is orthogonal to the a-axis.

640: '... suggest that a crystal's bulk radiation damage also plays an important role.' That seems a rather trivial discovery. What else could one expect?

We feel it is necessary to describe our findings even if the result is the expected outcome.

641: 'Crystal morphology plays a lesser role in that crystals with very high aspect ratios dissolve more slowly than more equant grains' ??? That is not apparent in Fig. 18e, and would seem to contradict the higher solubility along the c-axis than along a.

We have added a trendline to Fig. 18e to highlight the trend to which we refer. While there is higher solubility along the c-axis, the longer the c-axis is, the longer it will take the crystal to dissolve.

778: '...chemically abraded residues are more crystalline than their annealed counter parts...': Isn't that the logical relationships since CA removed the less crystalline domains? And: '... radiation damage is annealed hydrothermally during HF leaching...'. Speculative?

It is the logical relationship that chemically abraded residues are more crystalline than their annealed counterparts. We have more strongly emphasize this point (lines 284-287, 716-719).

We have removed the section "Implications for Radiation Damage Annealing Models" which discusses hydrothermal annealing of radiation damage. There are a few studies that report structural recovering during hydrothermal treatment (Rizvanova et al. 2000; Geisler et al. 2001b, 2002, 2003) that we cite in or draft manuscript, and this remains our preferred interpretation to explain some characteristics of our Raman data (i.e. changes in the relationship between the two Raman bands post-leaching). However, we recognize that this discussion needs additional data to support it and detracts from the main purpose of this paper.

867: 'Increasing the leaching temperature from 180 °C to 210 °C or increasing the leaching duration leads to the development of more extensive dissolution networks in higher damage grains ' This is a rather trivial conclusion. Something CA-users did not observe before? .... 870: 'More crystalline zircon samples lack fracturing related to

radiation damage zoning.' Another trivial statement. The conclusions should focus on the important aspects of the research, not on trivialities.

Again, while these conclusions may be the expected outcome, we feel it is still important describe these findings in the conclusion.

Reviewer #2 – Nasdala Lutz

The main shortcoming of the present research is perhaps that there is no chemical information on the samples studied. The present paper presents results of rather sophisticated studies including micro-CT and precise dimension estimates, but the most basic information remains hidden. There are quite a few journals that would refuse to publish a study on samples whose chemical compositions are unknown; for good reasons, as I think.

In particular, there are no U and Th concentrations available that, along with U–Pb ages, could have been used to calculate realistic time-integrated alpha doses. Instead, authors fiddle with equation #2 of Váczi and Nasdala (2017) in attempting to estimate self-irradiation doses from Raman FWHMs. Which is based on a serious fallacy of thinking: The equation of Váczi and Nasdala (2017) – that by the way has never been proposed as a "calibration", as present authors claim in line 223 – refers solely to Sri Lankan gem zircon with its particular damage-annealing history, whereas other zircon populations may have appreciably different FWHM-dose relationships.

In contrast, comparing alpha doses that were calculated from U, Th and age, with the observed FWHM values, would have opened up valuable opportunities for characterising the initial samples and their annealing histories. This chance was missed. Also, because of the unavailability of chemical compositions, it cannot be evaluated whether HF etching has removed completely some interiors of the grains but has left the rest fully unchanged, or did un-dissolved remnants experience some chemical leaching?

We appreciate the reviewer's concern regarding the absence of geochemical data for these samples. Our plan is to pair this microstructural investigation with a forthcoming study (the manuscript is currently in preparation) that presents geochemical and geochronological data for three of the zircon samples studied. This current contribution is very long (19 figures plus supplementary tables and videos); efforts to include both microtextural and geochemical data in a single contribution would produce a paper of unreasonable length. As such, the dataset has been split. In the second forthcoming contribution, we will compare alpha dose estimates based on zircon age and actinide concentrations with the Raman-derived alpha dose estimates presented here as suggested by the reviewer; the opportunity is delayed rather than missed. Further, because alpha dose estimates based on U and Th concentrations and U-Pb crystallization ages do not consider radiation damage annealing, they are not necessarily accurate representations of present-day damage in the zircon samples analyzed. Directly measuring the present-day radiation damage densities with Raman is more relevant to the present solubility experiments.

We intend only to use the v_3SiO_4-alpha dose relationship presented by Váczi and Nasdala (2017) for Sri Lankan zircon to broadly characterize damage in these samples as low-to-intermediate (KR18-04 and BOM2A) and intermediate-to-high (SAM-47 and AS3) and to demonstrate approximate magnitudes of inter- and intra-crystalline variations in damage. The v_3SiO_4-alpha dose relationship presented by Palenik et al. (2003) (their Fig. 5) for Sri Lankan zircon (which assumes an equivalent damage accumulation interval for the Sri Lankan zircon of 375 Ma to correct for geological

annealing) nicely fits the dataset of unannealed zircon presented by Nasdala et al. (2001), suggesting that the relationship is broadly appropriate for zircon from a wide range of geological environments. We have added caveats in the revisions that better clarifies that the derived alpha dose estimates are equivalent alpha doses based specifically on the v_3SiO$_4$-alpha dose relationship established for Sri Lankan zircon (lines 240-247). However, since we do not establish a firm alpha dose–dissolution mathematical model or the like, we do not think that using Raman-derived alpha dose damage equivalents is inappropriate.

Present authors rely on the results of Palenik et al. (2003) and Váczi and Nasdala (2017) who both have claimed that the FWHM of the n$_3$(SiO$_4$) Raman band of Sri Lankan zircon has a maximum ("saturation") value of about 35 cm^{-1}. On the other hand, sample AS3 yielded FWHMs of about 48 cm^{-1} (Fig. 5a, Table S1). This apparent contradiction should be discussed, and actually such extremely broadened Raman spectra should be shown (at least in a supplementary figure). It has been suspected that Sri Lankan zircon of elevated degrees of radiation damage have experienced preferred annealing, which might explain the "saturation" as a particular feature of Sri Lankan zircon. In conclusion, it is not surprising that other zircon, not affected by the particular "Sri Lankan" annealing history, may indeed show further band broadening. This has been rarely observed thus far but is valuable; so it should been shown and discussed.

The reviewer is correct that previous studies have shown that the maximum v_3(SiO$_4$) width is ~35 cm^{-1}. The maximum FWHM value recorded for the v_3(SiO$_4$) band for AS3 is ~35 cm^{-1}, consistent with previous studies. The 48 cm^{-1} width that the reviewer cites is for the E$_g$ band and not v_3. As demonstrated in Fig. 5, Fig. 6, and in previous studies, the E$_g$ band is always broader than the v_3 band (Härtel et al., 2020; Anderson et al., 2017).

Line 183, it is not really of relevance for the reader to learn about the equipment of the Princeton Raman system; instead, it should merely be stated which particular laser was actually used in the present study. Presumably red, as the green laser tends to induce Er^{3+}-related photoluminescence obscuring the Raman spectrum? The laser power at the sample surface (so not laser output but power measured behind the objective) needs to be reported.

We have rephrased this line in the revised text (lines 193-196). Both red and green laser wavelengths are reported because both lasers were used for Raman measurements. Most measurements were made using the red laser, since the green laser often produced fluorescent artifacts as the reviewer suggests. The laser power to the samples surface was 8.5 to 17 mW and ~7.5 to to 30 mW for the red and green lasers, respectively.

- Line 185: I wished Horiba would stop implementing the main silicon band in their auto-calibration procedure. However, if authors believe the friendly Horiba service person who did so was right, they should provide a published reference for the 520.7 cm^{-1} value they used for calibration. Other authors have used 520.5 or 521.0 cm^{-1}, and none of these values has ever been appropriately supported. Note that the oriented Si wafer is provided for alignment purposes but pretty unsuitable for wavenumber calibration: Visible light penetrates only a few µm into Si, which is why increasing the laser power, or changing the laser colour, changes absorption-induced heating and causes a shift of the Raman band. Why not using a truly known signal instead? The Rayleigh line has a Raman shift of 0 cm^{-1}; a perfect calibration means.

We used the Si 520.7 cm^{-1} band for instrument calibration because as the reviewer points out it is what the instrument manufacturer recommends, and it is the calibration standard supplied by Horiba for the LabSpec6 software's autocalibration function. Itoh and Shirono (2020, Journal of Raman Spectroscopy) demonstrated that a Si wafer has a Raman peak position of 520.45 ± 0.28 cm^{-1} which is within uncertainty of the value recommended by Horiba. We also monitored the accuracy of measured peak positions daily by measuring a reference spectrum for quartz (Krishnam, 1945 Nature). We have added both references to the revised manuscript (lines 197-200).

- Line 189–190, quoting ~2 µm spatial resolution upon using a "400 to 100 µm confocal pin hole" is too general and imprecise. With the confocal hole set to 400 µm, the spatial resolution will be much poorer than 2 µm lateral, and with a 100 µm confocal hole, the LabRAM Evolution should – provided the beam alignment is well done – reach its maximum performance close to the Rayleigh or Sparrow criterions, well below 1 µm (see detailed study by Kim et al., 2020, Current Applied Physics 20:71–77). The above resolution estimate does of course only apply if the Olympus 100× / n.a. 0.90 objective was used (this essential information is missing as well).

The ~2 µm spatial resolution sited is meant to be generalized because measurements were made using a range of conditions including confocal pin hole size and objective lenses. Most measurements were made using an Olympus 100x/0.9na lens or Mitutoyo 50x or 20x long working distance objective lens. We updated our spatial resolution to <1 µm to ~5 µm to better reflect the range of spatial resolutions expected and report the objective lenses (lines 199-203). We do not attempt to target or correlate our measurements across specific, fine-scale zones within a crystal. Rather, we use Raman measurements to broadly characterize the range of radiation damage densities in these samples, so precise spatial resolution is not critical.

- Page 8, caption of Fig. 3a: The wording "SEM images..." is too imprecise. State explicitly which images are CL, BSE, SE.

Each image in the figure has a symbol indicating whether the image is a CL or a BSE image. The figure includes a symbol legend. This was done to help shorten the text in the figure caption.

- Lines 218–219, the expression "Si–O asymmetric stretching band" is a contradiction in itself. To vibrate asymmetrically, it takes at least two bonds, whereas the stretching vibration of a single Si–O bond does not have any symmetry. Should be reworded to "asymmetric SiO_4 stretching".

We made the suggested correction (line 233-234).

- Fig. 18, in subfigures a–d the main x-axis label is missing.

The x-axis labels should simply be "Leaching Temperature." We added the labels to revised Figure 18.

- Lines 768–769, "...suggest that annealing or radiation damage is not simply the inverse of damage accumulation"; and again lines 891–892, "Raman analyses of annealed grains suggest that dry annealing is not the inverse of radiation damage accumulation". These statements are presented as if they were a result of the present study, which is decidedly not the case. In contrast, it has been emphasised may times already during the last 20 years that annealing does not inverse the damage accumulation process (mismatch of Raman FWHMs and Raman shifts, mismatch of unit-cell parameters a and c, etc.). Something like "our observations confirm again that..." would be more appropriate and more honest.... - Same in lines 250–252, "We note that none of the samples have achieved complete structural recovery after annealing at 900 °C for 48 h, since all measured peak widths are broader than that of synthetic zircon". This is not a new finding but merely another well-known fact, here merely reconfirmed again. People who wish to achieve near-complete annealing of zircon do this at 1400°C for four days or so.

Our intention was not to imply by that we were the first to draw these conclusions, but simply to describe the basic implications of the Raman data presented. We either rephrased (lines 274-277) or removed these lines.

By the way, Raman analyses do not suggest anything; only the results of Raman analyses suggest something (wording in line 891 should be improved).

This line has been removed.

-Line 800, a typo? Instead of "verses" there should be "versus".... - Line 983, the second author is Dr Chutimun (first name) Chanmuang N. (family name), so her name should be abbreviated "Chanmuang N., C."; same in line 1012. Similarly, in line 1050 the second author should be quoted as "Van den haute, P."

We have either corrected the Van den haute reference (line 883). The other reference was removed.

- Table S1, the "empirical correction" of Raman shift values for band downshifts that are claimed to be caused by laser-induced heating remains fully vague (unclear). Something like this has never been observed and reported before; so at least some comprehensible documentation is needed. How about simply decreasing the laser-power density, to reduce laser-induced heating to a negligible extent??

We noticed that $v_3(SiO_4)$ and E_g peak positions were slightly down-shifted in carbon-coated samples compared to uncoated samples, so we tested how different laser-power densities affected Raman FWHM and Raman shift values in carbon-coated samples. We found a correlation between Raman shifts and laser power density. As such an empirically-derived correction was applied to affected Raman shift values measured at too high of laser-power densities. There was no correlation between FWHM values and laser power. We recommend that all Raman analyses be made prior to carbon coating and SEM imaging to avoid this problem. Unfortunately, the carbon coat could not be removed from our fragile samples. Because of the additional uncertainty involved with these Raman shift corrections, we do not base any of the conclusions presented in this study based on Raman shift values, nor do we present Raman shifts in any figures in the main text. We have added additional text to the footnote in Table S1 to clarify this.

- Table S1, in view of the fact that many authors present and discuss uncorrected FWHMs, applying FWHM correction in the present article is meritorious. However, a correction assuming a FWHM of the IPF (instrumental profile function) of 1.5 cm^{-1} (I assume this is the meaning of the unitless "1.5" in footnote c?) most likely has led to an overcorrection. The FWHM of the IPF of the LabRAM Evolution with its 800 mm focal length, in case the 1800 grooves per millimetre diffraction grating is used, is about 0.8 cm^{-1} in the red and about 1 cm^{-1} in the green range of the electromagnetic spectrum. Authors should test this – and hence the experimental band broadening of their system

– by measuring the width of the Rayleigh line, Kr lamp or similar emission lines, or even better narrow Raman bands of known widths (unstressed diamond, Ba nitride).

Unfortunately I was unable to find Raman shifts and corrected FWHMs of synthetic zircon as measured by present authors. However, data points in Figs. 5 and 6 imply that synthetic zircon would have a FWHM of the n$_3$(SiO$_4$) band of about 1 cm^{-1}, which is well below the real value of about 1.7–1.8 cm^{-1}. This seems to support my above suspicion of FWHM overcorrection. Let me assume authors have fitted an FWHM of 2.1 cm^{-1}. This value, corrected for a (too large) 1.5 cm^{-1} IPF, yields an overcorrected FWHM of 1.0 cm^{-1}. In contrast, a fitted FWHM of 2.1 cm^{-1}, corrected for a 0.8 cm^{-1} IPF, yields a quite realistic FWHM of 1.8 cm^{-1}. Anyway, data for synthetic zircon should be added to Table S1.

Measured (uncorrected) v_3(SiO$_4$) and E$_g$ FWHM values for synthetic zircon are 2.1 cm^{-1} and 4.0 cm^{-1}, respectively. Corrected for a 1.5 cm^{-1} bandpass, these values are 1.0 and 3.4 cm^{-1}, respectively. These values have been added to Table S1. It is possible that these values are overcorrected as the reviewer argues, however, 1.5 cm^{-1} is the measured bandpass for the instrument that we used.

- Table S1: Reporting FWHM values with two decimal digits fakes an unrealistic accuracy. As the uncertainty is 10% for narrow FWHMs, reporting 4.4 instead of 4.38 cm^{-1} will do the job.

All Raman FWHM values are not reported with one decimal digit in Table S1 and new Table 2.

Editorial comment – Daniel Condon

The paper mixes a set of quantitative data that is linked to observations and qualitative data derived from it. As a result, the description of the data is rather wordy and at times difficult to follow – wonder if some form of tabulation wouldn't help? Also, lots of the discussion/generalisation seems reasonable, its worth making it clear that this if for four samples and a wider range of samples need to be characterised, and perhaps a wider range of parameters (i.e., the annealing in addition to the leaching that his manuscript focusses on).

We have re-organized and refocused sections as described elsewhere to improve manuscript readability. We have added a new table that summarizes key sample information including age and degree of radiation damage (Table 1) and text that emphasizes that dissolution mechanics are strongly sample-dependent (lines 33, 623-628, 733). In the conclusion we have added lines recommending that future studies should evaluate a wider range of zircon samples (lines 735-738). We address the comment regarding the annealing conditions below.

Fractures – the use of this term implies that planar features are a result of stress that is applied to the material. There is a good argument to be made for this process and it seems likely to be a common occurrence in some samples. My comment is that calling all planar features fractures implies a certain causative process (differential stress). Are all planar features fractures? Maybe they are.

Yes, all planar features are interpreted as fractures, and we attribute their formation to differential stress caused by volume expansion/reduction related to radiation damage accumulation/annealing or inclusions. Differential stresses may be caused either by geological processes or during thermal annealing in the laboratory. We have stated this more clearly in revised text (lines 538-542, 722-727). Non-planar features are not interpreted as fractures, and we refer to them in the text simply as acid paths.

Rim to core dissolution. A lot of different mechanisms and processes are discussed and the authors do a good job of introducing these and tracking them through the discussions. Whilst the images/analyses support that rim to core dissolution is not typical the authors do state (line 680) that there is a progressive rim to core dissolution

The reviewer is correct that rim-to-core dissolution is not typical. We only expect rim-to-core dissolution (with the caveat along the crystallographic c-axis) in highly crystalline grains only in the event that there are no acid paths to the grain interior (i.e. no fractures, surface reaching inclusions, or overlapping defects that reach the grain surface). We clarify this in revision (lines 643-655).

Much of the focus in the discussion, and in the community, is around the leaching temperature as being the thing that is most significant. Perhaps it is but what above the duration of the annealing, or the rate of cooling? This paper focusses on samples that have bene annealed for 48 hours but it should be acknowledged in the manuscript that practitioners quote a range of annealing durations, typically 48 or 60 hours. I assume this is the time between turning the furnace on and off, and often cooling can take several hours although the rate of cooling can be increased by opening the furnace.... and "How does the duration/temperature of the annealing impact the crystal structure

and precondition it for reaction to leaching?" ... Line 251 – how does this look for 60 hours? Is there any published data for this?

We selected our annealing temperature (900 °C) based on the recommendations of Huysken et al. (2016) who showed that increasing the annealing temperature can lower the solubility of domains affected by residual Pb-loss. Annealing studies of radiation damage in zircon demonstrate that annealing is strongly dependent on temperature and weakly dependent on heating duration (Zhang et al., 2000; Ginster et al., 2019). The difference between 48 and 60 h or oven cooling times is expected to have negligent effect on the crystallinity of zircon based on these previous studies (Ginster et al., 2019, their Figure 1). We have added these points to the revised manuscript (lines 128-136).

The zircon crystals studied have not been analysed for high-precision U-Pb – but data does exist for the samples (AS3 – Schoene et al., GCA, 2005, coherent U-Pb; SAM-47 – no U-Pb data published? KR18-04 – MacLennan et al., Sci Adv, 2020 – overdispersion, Pb loss? BOM2A, Basu et al., 2020, single population). Have these analyses been conducted at experimental conditions analogous to those deployed in this study? I appreciate that the precision/resolution may not be at the level to preclude Pb-loss but presenting the data might help frame the discussion around the implications for zircon U-Pb systematics and age interpretations

There is currently no published U-Pb data for SAM-47, and not all published U-Pb analyses were conducted at analogous experimental conditions. As such it would be difficult to extrapolate our textural findings to how well any specific leaching condition eliminates Pb-loss for these samples. As such, we have shortened the U-Pb discussion section in revisions (Section 4.2) and refrain from prescribing recommended leaching conditions. We instead refocus the paper on describing dissolution textures. The goal of this manuscript is to lay the mechanistic groundwork for a second contribution that will include geochronological and geochemical data for three of the four samples and focus on how different chemical abrasion conditions affect zircon U-Pb and trace element systematics. This second piece will link the textural results presented here to both the literature data cited and the new U-Pb data. Efforts to include textural, geochronological, and geochemical data in a single contribution proved unwieldy.

The paper seems focussed around leaching mechanisms/processes applied to whole crystals – however the process will often be applied to fragments and/or grains that has been polished for CL, on both cases exposing the interior of the grains. Would be useful to mention for the non-practitioners that not all zircons will come as complete crystals.

We have added this point to the discussion (line 630-634).

It is a long paper and much of the qualitative observational data based upon examination of many observations from the four samples, which a subset of representative images presented. One issue is around readability – could some of the generalised observational data/interpretations be tabulated to make it more accessible? Personally, I found it challenging, going back and forth to try and compare what is said for the different samples and leaching temperatures. I felt the use of tables may be helpful for compiling this qualitative information and making it more readily accessible.

We have shortened the manuscript considerably by >1500 words, reorganized, and refocused the results and discussion sections as described elsewhere to improve manuscript readability. In the results section, we made "3.3 µCT imaging of radiation damage zoning" a stand-alone section. Discussion sections 4.1.1 and 4.1.2 have been reorganized to better summarize key findings from the results section and their implications for the dissolution of higher and lower damage grains. Discussion section "Implications for radiation damage annealing models" has been removed entirely. We added a table that summarize basic sample information (Table 1) and Raman results (Table 2).

Conclusion section – is it possible to draw out the observations and how they record a progression of processes?

This is an excellent suggestion. We have retooled the conclusion to better reflect a progression of processes (lines 700-742)

Bowring and Schmitz, not Bowring and Schmidtz - We corrected this typo.

also mention the rare occurrences of reverse discordance seen in some samples? - We can add this point to the introduction. (lines 43, 56)

thermal annealing instead of laboratory annealing – We have changed all occurrences of laboratory annealing to thermal annealing

remove more soluble – We have corrected this.

bias, yes, but more realistically this should be considered an additional source of uncertainty in the assigned age – We can add this point to the introduction (line 58-60).

also prompted the community to question/explore a range of interpretative frameworks for such datasets – We have added this point to the introduction (line 66-68).

Line 96 – could the sample information be tabulated?

We added a new table with basic sample information including U-Pb age and alpha dose estimates (Table 1).

Line 146 – what portion (percentage of grains) didn't survive the leaching and was their anything distinctive about those grains? Did any grains break apart?

AS3 and SAM-47 residues are *extremely* fragile. Many of the AS3 and SAM-47 grains broke in the process of transferring the crystals from the Teflon microcap to the tape or disintegrated entirely. Touching the residue with a tweezer tip was sometimes all that it took for a grain to disintegrate. Examples of several broken crystals are shown in the SE and μCT collages. The only distinctive characteristic of broken grains was either large, pre-existing fractures visible in μCT images or more generally samples with high initial radiation damage (i.e. all AS3 and SAM-47 grains). A large percentage of grains were also dropped or lost during pipetting. Lab notes do not distinguish between disintegrated and misplaced, so calculating a percentage of broken is not possible. These are important points to make that have added them to the results section (lines 326-328, 388-389, 424-425, 462-463).

Figure 2 A is reflected light, what is the light source for B? I assume AS3 is top left etc., for Panel B but there is space to add label, or state this in the figure caption. The images are low resolution – will higher resolution version be submitted as a supplement? Also, the top right panel indicated the images contain residues that have been leached for varying times (4 and 12 hours) – how does the reader distinguish these different grains?

Both A and B are reflected light images. We have modified the figure to better illustrate which samples are shown (AS3, SAM-47, KR18-04, and BOM2A) in Fig. 2b. Labeling the leaching condition for each individual crystal, however, isn't practical given the restricted space, nor can dissolution features be seen at the image's low magnification. We do not have photomicrographs at higher resolution – The SEM images in Figures 9 – 17 do much better at showing high resolution images of dissolution textures.

Line 262 – were replicate raman determinations made on any of the crystals to assess variation within a crystal?

Yes, replicate analyses were made on some zircon crystals. Most grains exhibit intracrystalline variations to some extent as discussed in the results section. Graphically, intracrystalline variations are best illustrated in Figure 5a by the core and rim measurements made for SAM-47. The sample names in Table S1 in the Supplementary

Material indicates which samples have replicate analyses. We have added text to better clarify (lines 237-238).

Figure 8 – what is the 2D nature of the 3D rendering in figure 8?

The 2D image is a single slice of the image stack used to render the 3D grain in B. We have changed the figure caption to clarify.

376 – remove interestingly.

We have made the suggested change.

678 – ID-TIMs analyses represent an integrated analyses of the residue post-leaching – this could be core-rim, or more core, or more rim – seems that is will be sample and duration dependent?

Yes. This is a key point. We added this point to the text (lines 648-655).

687 – "hot leaching" (210C)? Or should this be hotter leaching? Or longer leaching?... 702 – yes, but in lower radiation damaged samples it may also impact an age-bias towards the core/older material...723 – and could this be a mechanism that results in the rare cases of reverse discordance we see in CA ID-ITMS data for some old zircons? Could we be seeing this in samples but not at a resolvable level, where the zonation is favourable?

That is indeed possible on both points. However, including these points in the text may be too speculative since we do not present complimentary geochronological data in this study. Based on the suggestions of some of the other reviewers we have removed these lines from the text.

741 – yes but sometimes the inclusion rich zircon may be the ones we want to date...

This is true. We added this point to text (lines 664-665).

781 – is hydrothermal annealing a thing? Can you provide a reference? Looking at the literature I couldn't find anything in materials science – do you mean hydrothermal treatment? This sounds odd – how robust are the few data this discussion is based up?

There are a few studies that report structural recovering during hydrothermal treatment (Rizvanova et al. 2000; Geisler et al. 2001b, 2002, 2003), and this remains our preferred interpretation to explain some characteristics of our Raman data (i.e. changes in the

relationship between the two Raman bands post-leaching). However, we recognize that this discussion needs additional data to support it and detracts from the main purpose of this paper. We have removed this section in revisions.

784 - Widmann. We corrected the typo.

865 – Okay – then what impact does the annealing temperature/duration/cool down rate have on the formation of micro fractures? All the discussion is around varying the leaching parameters but should we also be considering the annealing step?

To the best of our knowledge previous radiation damage annealing studies have not focused on how different temperature conditions affect zircon micro fracturing. We have added text to the conclusion section suggesting that future studies should evaluate how different annealing conditions affect micro fracturing in zircon and the rate of zircon dissolution (line 740-742).

839 – four samples covering a range of ages and radiation damage accumulation.

We have made the suggested correction (700-702).

888 - ... removal of excess closed system material AND potential age bias in lower radiation damaged materials?

We have removed these lines.

Community Comment 1 – Charles Magee

Section 2.1 Firstly, the introduction says that the second described sample- SAM-47- is from the Corunna Downs granitoid complex. The Australian Stratigraphic Unit database describes this term as informal. See:

https://asud.ga.gov.au/search-stratigraphic-units/results/34394 It has been replaced by the Corunna Downs Granitic Complex:

https://asud.ga.gov.au/search-stratigraphic-units/results/72996

Note, however, that this term also is obsolete. As shown in the links above, Pilbara Granitic Complexes are not stratigraphic units, but are geological provinces, so neither of these descriptors is particularly informative. More importantly, the SAM-47 description is the only description that does not have any references associated with it,

making it difficult to understand the geologic background of this sample, or any associated information that would allow readers to interpret the results.

Furthermore, the latitude and longitude (89°59′55.97″, 100°08′2.38″) given are in the

Arctic Ocean near the north pole, and are not in the Pilbara craton.

In summary, it would be helpful if the authors could more accurately locate the sample site, and relate it to a local, named stratigraphic unit, and provide appropriate reference(s) to previous work that provides geological context.

We thank the reviewer for catching these mistakes. The correct GPS coordinates are (-21°24′29.01″, 119°46′21.03″) The sample comes from the Emu Pools Supersuite stratigraphic unit. Appropriate literature references about the region's geological setting have been added (lines 104-107).

Section 3.2.1. The use of a synthetic zircon (which is not described in the samples section of the methods) may not be the most appropriate measure of a full annealing natural zircon. Lattice strain can be caused by factors other than radiation damage, such as the incorporation of variably incompatible trace elements into the lattice structure. If the synthetic zircon is pure ZrSiO4, instead of being grown with levels of P, Y, REE, and other trace elements typical of zircons from basic to felsic host rocks, then the ability of chemical abrasion to repair lattice strain may be underestimated due to the lack compositionally related lattice strain in the chemically pure synthetic crystal.

Similarly, we can't tell from the Raman data whether the narrower peak widths of KR18-04 and BOM2A are due to damage or composition, although the narrower peaks

for the younger zircons, and excellent choice of one mafic and one felsic zircon from both the 'old' and 'young' groups does suggest irradiation is important.

We added text to the Raman methods section describing the synthetic zircon used as a loose analog for undamaged zircon (lines 207-209) and report Raman results for it in Table S1. The reviewer is correct that synthetic and natural zircon crystals have different types of intrinsic defects, some of which are related to composition. However, composition has not been shown to significantly influence the zircon Raman spectrum (Nasdala, 1995), and synthetic zircon is used in the literature as an analog for undamaged zircon in Raman FWHM-alpha dose models (Vaczi and Nasdala, 2017), so we feel that including the synthetic zircon in Fig. 5, 6, and 7 as a visual reference point is useful and appropriate. We updated the wording to acknowledge that the slight differences in FWHM values between annealed BOM2A and KR18-04 samples and

synthetic zircon could reflect either residual radiation damage in the annealed samples or other intrinsic differences between synthetic and natural samples (lines 266-270).

On a related note, when estimating the accumulated lattice strain, a U/Th/He age or a fission track age may be more appropriate than a crystallization age, depending on the ability of moderate-to-high temperature zircon to self-anneal radiation damage over geological time. The lack of location data for SAM-47 (see above) makes estimating this difficult, but to use a well-studied East Pilbara Archean example, the Owen's Gully Diorite has a crystallization age of 3467 Ma (Stern et al. 2009), but a helium age of only about ~750 Ma (Magee et al. 2017).

The reviewer is correct that a zircon's crystallization age is not equivalent to its radiation damage accumulation interval, since radiation damage anneals at relatively low temperatures over geologic time. Depending on a sample's thermal history, a zircon (U-Th)/He age or other thermochronometric data often provides better insight into how long a sample has accumulated radiation damage. However, we do not use U-Pb crystallization ages to calculate samples' radiation damage. All equivalent alpha dose estimates cited in this study are instead based on Raman v_3 FWHM peak width measurements of accumulated damage. A second, forthcoming contribution which includes U-Pb isotopic data for three of the four zircon samples will more extensively discuss what is known about each sample's thermal history since it is pertinent to isotopic systematics. We appreciate the suggested reference, and we will include it in the forthcoming manuscript.

Figure 2. Finally, it might be worth specifically pointing out that 2b is a colour photomicrograph, as the annealing out of radiation damaged colour centres is an important but sometimes overlooked part of CA. This illustration is so dramatic that

readers might not appreciate that the second image is a colour image in which all the colour has been annealed out of the zircons, leaving them almost colourless.

We have added text to the figure caption emphasizing that Figure 2b is in color, and we highlight the significance of the color change.

Community Comment 2 – Magdalena Huyskens

There are quite a few qualitative statements that can and should be backed up with statistical analyses. For example, claiming changes in the slopes between $v3(SiO4)$ and Eg peak after annealing and leaching (lines 254-256) and reporting the average changes

and range for the peak width in a raman spectrum for the individual samples and temperature steps.

In the revised manuscript we have included slopes calculated using a simple least-squares linear regression for all unannealed, annealed, and chemically abraded samples in Fig. 5 and Fig. 6 to support our earlier statements. We have also include a new table (Table 2) that summarizes the minimum and maximum $v_3(SiO_4)$ FWHM values for each sample. A similar table could be generated for E_g FWHM values, however, we feel it would be a bit redundant. All measured $v_3(SiO_4)$ and E_g peak position and FWHM values are included in the Table S1.

For the estimates of volume loss, the method needs to be described in more detail. Right now, I am not sure if it is including interior dissolution features or not.

Volume estimates for these two samples were made using the Dragonfly software's "Upper OTSU" segmentation function. This function differentiates zircon from inclusions, dissolution features, and background (tape/air) based on grayscale intensity. Total volume is calculated by adding the number of selected high-intensity zircon voxels together. The volume of one voxel is ~4.25 μm^3. The volume loss estimates do include interior dissolution features. However, we note that some of the finer-scale dissolution features and inclusions are missed due the dataset's spatial resolution and beam hardening effects. We have added these details to the methods section (lines 185-190)

One of the findings is that some compositional zones are preferentially dissolved. Do you have any compositional data for these? It would been great to know what the difference in composition is and if the solubility is solely based on radiation damage, or some composition otherwise is more soluble.

We agree with the reviewer that the geochemistry of these zones is an important piece of the puzzle. Unfortunately, we do not have compositional data for these specific samples. This study was designed to evaluate the microstructural changes that occur in zircon during chemical abrasion, so geochemical changes are outside the scope of the current investigation. This contribution is intended to lay the groundwork for a second step-leaching study of 3 of the 4 zircon samples that evaluates the geochemistry and geochronology of the dissolved zones.

The one recommendation that is put forward for the CA technique is to increase the temperature: "In most samples regardless of initial damage content, we find that chemical abrasion at 210 °C is more effective at mining out soluble zones from crystal

interiors. Based on our mechanistic blueprint, we predict that hotter leaching temperatures are thus more likely to better mitigate Pb-loss in geochronological datasets." I do not find this supported in the data. Yes, the solubility increases with increasing temperature. Increasing it a little bit more or increasing the time will

completely dissolve the entire zircon grain. Since there is no U-Pb data associated with this study, the mitigation of Pb loss is speculative. In addition, there are many zircons that will completely dissolve with such a treatment, in which case no U-Pb date could be collected.

The reviewer is correct that without geochronological data this interpretation is speculative. We have removed this recommendation in revisions. We will note, however, that this interpretation is supported by our step-leaching investigation.

Minor comments:

Misspelling of reference "Bowring and Schmidtz, 2003" (multiple times in the introduction)

We have corrected the misspelling.

Figure 2: Is there any way to track which conditions were used for which grains? It would be helpful to get an idea overall how the different samples are behaving under the different conditions. In addition, it would be nice to have images of the zircons before annealing. There are often colour changes associated with this step.

We modified Figure 2 to better illustrate which samples are shown (AS3, SAM-47, KR18-04, and BOM2A) in Fig. 2b. Labeling the leaching condition for each individual crystal, however, isn't practical given the restricted space, nor can dissolution features be seen at the image's low magnification. Figures 9 – 17 do a much better job providing an overview of how the different samples behave given different leaching conditions. Annealing did induce color changes, but unfortunately, we do not have images of grains prior to annealing.

Figure 5: All panels should have the sample name in them, at roughly the same position. It is confusing that c) is within a). Use the same font type. If one panel has the alpha dose, the other one should have this too..... L 254-260: "We note that relationship between the $v3(SiO4)$ and Eg peak widths is steeper after annealing in each of the four samples, since the two Raman peaks have different temperature sensitivities (Hartel et al., 2021). This observation suggests that laboratory annealing is not simply the inverse of radiation damage accumulation. As such, we caution against using the Váczi and

Nasdala (2017) calibration to derive alpha dose estimates from v3(SiO4) peak widths for annealed or chemically abraded samples and omit alpha dose axes from Figures 5b and 6b." This is inconsistent. The alpha dose is noted for the annealed samples, but not the partially leached ones.

We added sample names to the bottom panels of the Figure 5 and Figure 6. The alpha dose axes in the figures are intended for the unannealed samples *only*. We modified the axis label to better stress this point. We do not feel it is appropriate to use the alpha dose scale for either the annealed or chemically abraded samples as emphasized in the text, so we will not add the axis to Fig. 5b or Fig. 6b. We will need to leave panel c) within a) because there isn't another good space to place it without vastly expanding the size of the figure.

Figure 7: The choice of color for the 180 °C for 12 h for AS3 & SAM-47 is odd, since it fits the color scheme of samples KR18-04 & BOM2A.

The color scheme in this figure is based on leaching condition as opposed to sample which we feel is more appropriate for easier cross-sample comparison of different leaching conditions. All 12 h leaching experiments are assigned a shade of teal: 180 C, 12 h experiments are dark teal, and 210 C 12 h experiments are light teal. We have changed the legend in the figure to better reflect this.

Lines 274-281: "Notably, SAM-47 and BOM2A residues each have at least one data point with a narrower v3(SiO4) and Eg peak width than their solely annealed counterparts suggesting that some residues have a higher degree of crystallinity. Further, we find that the residue datapoints for these two samples largely plot below (at lower v3 for a given Eg) the annealed datapoints indicating a change in the relationship between the v3(SiO4) and Eg peaks. Taken together, these observations could suggest that additional structural recovery occurs in some zircon samples during HF leaching even after dry annealing at significantly higher temperatures."

Is there any reason that this observation can't just be explained by the removal of more damaged zones that were not annealed during the high temperature annealing step? Structural recovery during HF leaching seems impossible to me and would need some further explanation.... ... Line 777- 780: "There is also an apparent change in the relationship between the widths of the v3(SiO4) and Eg peak after partial dissolution in HF acid in some samples, and a small number of Raman analyses for chemically abraded residues are more crystalline than their annealed counter parts" Same as for Lines 274-281. Does this not just mean that the parts that were not annealed due to larger radiation damage are still present before leaching?

Dissolution does remove more damaged zones, but dissolution of these zones cannot account for the change in the relationship between the two Raman bands; dissolution would remove material with broader peaks but would not result in a change of slope. We have removed Section 4.3 "Implications for Radiation Damage Annealing Models" which discusses hydrothermal annealing of radiation damage in revisions. There are a few studies that report structural recovering during hydrothermal treatment (Rizvanova et al. 2000; Geisler et al. 2001b, 2002, 2003) that we cite, and the color change evident in grains after chemical abrasion suggests that chemical abrasion anneals color centers. However, we recognize that this discussion is needs additional data to support it and detracts from the main purpose of this paper.

Line 405: "... many most ..." remove one of those words

The suggested correction has been made.

The section 4.2.2 Inclusions and zircon trace element analyses is not discussing the impact of dissolving compositional zones within zircons that are observed in this study.

We thank the reviewer for this suggestion, and we will add text regarding this point (line 648-649).

The section "4.4 Imaging radiation damage zoning: Implications for (U-Th)/He thermochronology" seems disconnected from the main point of the paper and is a little distracting.

This section is admittedly disconnected, but we feel it important to include. This paper is the first contribution to demonstrate that μCT can be used to image radiation damage zoning in zircon non-destructively and in 3D. There are big implications for this beyond U-Pb.