

Response to review #1:

We appreciate Reviewer #1's interest in our project, and regret that they missed the main points of our paper, which we repeat below for clarity:

The main point of our paper is that chemically abraded reference ages should not be used for non-chemically abraded (untreated) reference zircons for SIMS U-Pb analysis. We reproduce chemically abraded CA-ID-TIMS reference ages for chemically abraded SIMS targets, and untreated ID-TIMS reference ages for untreated SIMS targets. This is the most important and urgent part of this paper, as many SIMS analysts currently use CA ages for natural material. While the differences for most reference zircons are too small to distinguish in an individual session, this will produce systematic errors over time for any reference zircon where these ages appreciably differ.

Our secondary point, and the reason the study was conducted, is to see if the chemical abrasion process interferes with measurements of other chemical and isotopic systems, either by mobilizing elements, or by altering the matrix in a way that interferes with the analysis. Our conclusions are that this does not happen.

Our tertiary point is that if SIMS U-Pb analysts wish to achieve sub-percent accuracy and precision, chemical abrasion will be necessary, but not sufficient, for all but the least radiation-damaged zircons. Natural zircon is not a reliably closed system at sub-percent levels. Once the open system material is removed by chemical abrasion, highly accurate and precise ages are possible, but are not yet routine.

We admit that due to the large amount of data in this paper has resulted in us plotting only the barest minimum of figures to make our point; perhaps we have been too miserly in this regard. We can happily produce the additional figures that reviewers request, with the caveat that this would lengthen what is already a long paper.

Should the editor prefer the paper be split into multiple papers (e.g. one of TIMS and SIMS U-Pb, and one on the rest of the periodic table and Hf and O isotopic systems), we are open to such an approach, however we would like to point out that reviewer's #1's desire to see this work both published soon and completely rewritten is self-contradictory.

Reviewer #1 engages in excessive speculation regarding the zircons from Mt Painter Volcanics, a S-type dacite in the Lachlan Fold Belt, Australia. We admit an oversight in representing the literature on this topic, the inclusion of which may answer some of reviewer #1's questions.

While we were preparing this manuscript for submission we failed to notice the publication of Vogt et al. (2023), which describes the chemical abrasion and subsequent SIMS analysis of S-type granitic zircons from Central Europe.

Like the study of Kryza et al. (2012), Vogt et al. (2023) chemically abrade only their unknown sample, and do not look at the effect of CA on well-characterised zircons. However, Vogt et al. (2023) provide SIMS U-Pb,  $\delta^{18}\text{O}$ , and trace element results for both untreated and chemically abraded samples which answer many of Reviewer #1's speculations on this topic.

We feel that that the very different aims of our study and Vogt et al.'s (2023) should make explaining their complementarity fairly straightforward. However, if the editors feel that Vogt et al. (2023) renders our S-type work redundant, and that this work is confusing to readers like Reviewer #1, we suppose we could remove it for clarity and publish just the reference zircon data.

With regards to the scatter of this data, almost all SIMS U-Pb data are scattered beyond the statistical expectations of the grouped spots. This is why it was noteworthy that the chemically abraded reference zircon analytical session (170123) did not require any spot-to-spot error component. Thus, the scatter in the MPC data is no different than the vast majority of SIMS U-Pb data collected over the last 40 years. This is why we don't dwell on it, aside from noting that it shows that chemical abrasion is not a magic bullet that will eliminate spot-to-spot error every time in every sample of every session. The outliers were worth checking to make sure there was no reproducible change that might indicate diffusion (or some other elemental mobility) induced by the CA process, but outlier spots are not that unusual. To put them in perspective, with the exception of spot MPC.21.1 (which clipped the tip of a core), the rest of the spots all group with a spot-to-spot error of 1.02%. While slightly higher than our lab's long term average, this is quite good by historical standards (see figure 1 of Magee et al. 2024 for context).

Similarly, CA-ID-TIMS data also often fail to group. Of the 36 TIMS results examined in Magee et al. (2023), only 13 had a single group. While many of these (particularly the Permian volcanics) are obviously geological (as shown in Metcalfe et al. 2015 and Laurie et al. 2016), rejection of outliers is commonplace in the CA-ID-TIMS literature. Outlier treatment in SIMS is a large and contentious topic which we feel is beyond the scope of this paper; this is why we provide both outlier rejected and included data for those samples which do not group (OGU, QNU, and MPC) in tables 2 and 3.

Finally, if there is a SIMS surface effect created by chemical abrasion, it is of a smaller magnitude than the variation of U-Pb measurements caused by open system behaviour in untreated zircon. The difference between MPC and MPU is slightly less than half a percent. However, Magee et al. (2024) show that SIMS dates of untreated zircon can be as much as 2% different to their CA-ID-TIMS dates, due to open system behaviour in the zircon. So even if we are introducing a new source of bias, it is several times smaller than the open system problem which it solves.

We do not use  $^{207}\text{Pb}$  common Pb correction for the age of the Mount Painter Volcanics because publishing their age is not the point of the paper. The point of the paper is to examine the effect of chemical abrasion on zircon. Obviously multi-age domain zircons are never used as reference zircons, so Mount Painter Volcanics zircons were picked because they were, in our opinion, the most likely multi-age zircon we had in our collection that we thought would behave well (As we state in section 1.2). Comparing zircons requires similar data treatment for all samples. And  $^{207}\text{Pb}$  corrections cannot be performed on OG1 because around  $\sim 3450\text{Ma}$ , the common Pb correction line and the concordia curve are parallel, so tiny deviations from concordance produce huge changes in age (and reverse concordant data has no intercept at all!). Thus for consistency we do not use  $^{207}\text{Pb}$  corrections on any samples.

With regards to the many comments on S-type zircons, the applicability of the S-type label to rocks beyond the Silurian of the Lachlan Fold Belt is well beyond the scope of this paper. We feel that data on LFB  $\sim 430\text{Ma}$  S-type granites (e.g. those of Ickert 2012 or Bodorkos et al 2015 ) are the best approximation for what to expect from the Mount Painter Volcanics, a  $\sim 430\text{Ma}$  S-type pyroclastic dacite. Again, the purpose of this sample in this study is to determine if chemical abrasion does unexpected things to multi-stage zircons, and as we state in section 1.2, zircons from the Mount Painter Volcanics were chosen as the simplest example of a zircon with multiple growth domains.

Addressing Reviewer #1's specific points:

We are well aware of the McKenna et al. (2023) study, as I (CM, the corresponding author) provided a community comment during the review process. It is complementary to our work, as their imaging

of ever-smaller channels of dissolution down to the resolution of their instrument is consistent with our observation that the removal of low level Pb loss is pervasive at the scale of our SHRIMP spot.

We do not maintain the argument that chemical abrasion of moderately damaged (or highly dissolved) zircon produces worse SIMS results because neither the literature nor our data support that argument. Previous studies (Kryza et al. (2012), Watts et al. (2016), and Vogt et al. (2023) between them showed improvement in all 5 of their unknowns. Our data for OG1 and QGNG similarly show vast improvements. Whether the MPC data is better or worse depends on one's outlier treatment philosophy and data quality metrics; MPC is certainly lower in common Pb than MPU, and the outlier-rejected age is more precise. But with 4 studies showing that 7 zircons give better CA data, 4 reference zircons (91500, R33, Temorax2) are unchanged within uncertainty, and only one might be worse depending on how you treat the data, to catastrophize about the use of CA for SIMS is nonsensical.

Reviewer 1's speculation about secondary ion emission behaviour due to the change in texture from large scale dissolution is interesting, but tangential. Should the editors feel it is important to address, we have data from follow-up experiments which show that OG1 chemically abraded with a more aggressive recipe produces results that are at least as good, if not better than the results published here, and with a much higher proportion of dissolution channels. We did not include that work in this paper because this study is long enough already, and it is a minor secondary concern. Similarly, we have hundreds of sessions worth of OG1 secondary standard data we could present to show that the SIMS age reproduces the untreated (and not the CA) ID-TIMS  $^{206}\text{Pb}/^{238}\text{U}$  age, but we didn't feel the need to belabour this point (Which was previously made in Black et al. (2003), Stern et al. (2009), and Magee et al. (2023)).

Reviewer 1's suggestion that the difference in igneous age cores between MPU and MPC is due to selection bias is backwards: inherited cores are easier to see in MPC, so it is less likely that we would hit igneous ones by accident while targeting detrital cores from the protolith, as was our stated goal (line: 256-257). All analyzed spots are annotated in the supplementary figure 2 if the reviewer is curious about spot placement. There is no need to speculate.

Reviewer 1's point about the radiation damage history of our reference zircons (sample selection) is valid. Sadly, not all reference zircons are completely characterized with regards to the time at which they cooled through their self-annealing temperature limit (lines 97-100). Our main aim was to find materials with a variety of ages which had both chemically abraded and untreated reference values. While damage estimates for OG1 and Temora-2 were published in Magee et al. (2017), we are not aware of any determinations for QGNG. Regional constraints suggest it cooled through  $\sim 300\text{C}$  (K-spar argon closure from Foster and Ehlers 1998) around 1600 Ma, and through apatite fission track closure ( $\sim 100\text{C}$ ) around 250Ma (Kohn et al. 2002). These provide uncertainty of more than a billion years for when QGNG dropped below the self-annealing temperature. None-the-less, most of central Australia has similarly nebulous cooling constraints as QGNG, so we think it is a useful reference material to investigate, as much of our work is done in the central Australian Paleoproterozoic.

Reviewer 1's comment about internal and external error is misplaced. Internal error is to be used when comparing results run from within the same analytical session, which is most of what we do in this paper. External error is also reported and used in relation to comparing our SIMS results to reference values. This is why we took the time in the introduction to carefully explain this difference.

Reviewer 1's comments on calibration slopes is confusingly written. Any of the eight reference zircon samples can be used as a primary standard to produce a calibration line. We mostly present data

using T2U or T2C as the primary reference materials for convenience. For each reference zircon, the calibration slope of the untreated and CA samples are within uncertainty of each other, as is shown in figure 1. 91500 (both untreated and CA) has a different slope than the other three. As we specifically state in the paper (line 400) the only permutation of references and unknowns which isn't within analytical uncertainty is the OGC-91U pair, and even these are only mismatched by less than half a percent.

Replies to the specific line-by-line comments:

L149. These CA-ID\_TIMS OG11 results were delivered to GA too late for inclusion in Stern et al. 2009, and instead presented in poster form by Bodorkos et al. 2010. We are presenting them in full here in supplement table 1.

L502. Uncertainties including tracer are always larger than without the tracer, because tracer uncertainties are non-zero. However, there is an error in table 1, in that the values tracer+random 95% on the ratio did not recalculate when our new tracer uncertainty was applied. We thank the reviewers for making us check this.

L582-3. The Mount Painter U-Pb data are shown in table 3 and supplementary table 5.

L613 Relative Rim-Core  $^{206}\text{Pb}$  concentration must be known to predict the direction of any potential diffusion.

L617: We are saying that chemical abrasion alone will not eliminate spot-to-spot error. There are still session-specific instrumental effects, not all of which we have solved, which can and do occur.

L627 It isn't a guess to say that 18 is larger than 6, or that 12 is smaller than 22.

L657 Rims on low  $\delta^{18}\text{O}$  cores having their  $\delta^{18}\text{O}$  values dragged down relative to the rim average (or vice versa for high  $\delta^{18}\text{O}$  cores) would be evidence of core-rim oxygen exchange. We do not see this, but it would be hard to see due to the heterogeneity in rim  $\delta^{18}\text{O}$  in this sample.

L658 Ti-is zircon thermometry is calculated as described in L470 and shown in table 6A (reference zircons and 6B (Mount Painter Volcanics).

L662. Attempts to get CA-ID-TIMS ages from Australian S-type igneous rocks often find cores are pervasive enough to make defining an igneous age group impossible, so this is an actual problem in need of a solution.

L668 We are happy to move this paragraph.

L677 We agree that OG1 is the only sample to lose water, which is why we state this in the text. Note that there is an error in the supplementary data table- we have copied the raw data twice instead of also presenting corrected data (which is what we plotted). We apologize for this .

We are happy to check the format of the supplementary tables.

SIMS measurements with low backgrounds and low abundance often have a distribution around zero, tailing in either direction. Replacing all the negative instances will bias the mean, so it is poor practice.

We agree that oxygen mean data should have full uncertainty propagation, but as those are systematic effects they should not be added to the individual spot data in the supplements, but to the aggregate value (in the main text, not the supplement).

From the reviewer's "Scientifically incorrect" section...

L89: The reviewer is wrong. Zero age Pb loss does not change the  $^{207}\text{Pb}/^{206}\text{Pb}$  ratio. The Pb is lost, and because it is modern loss, there is no subsequent ingrowth.

L91, 98: This is disingenuous. Obviously insufficient ion counts will limit precision of any mass spectrometric analysis, SIMS or otherwise. For Pb isotopic ratios, the limit of precision in SIMS is basically set by the precision on the Pb isotopic reference zircon, which is generally between 1 and 2 permil (2 sigma). See Stern et al. (2009) for a more thorough discussion. For U-Pb, there are a host of other issues, such as the calibration, Pb loss, etc., which limit precision more than counting stats for all but the youngest or lowest U zircons. See references in the introduction section for more details, or the introduction of Magee et al. (2023), and references therein, for a more thorough discussion of the calibration and related uncertainties.

L512: The sub-micron positioning of the peizo stage during autoanalysis, particularly in relation to working distance (X in the SHRIMP / American Chemical Society coordinate system), reduces the variability in steering of the secondary ion beam off of the sample (particularly the QT1Y voltage change spot-to-spot).

L513: A 95% confidence interval of 3 Ma out of 1000 Ma (or 1063 Ma ) is roughly three permil, not three percent.

L535: Wee happy to clarify that we refer to the Magee et al. (2016) supplement, not this one.

L554: As we state in L553, the difference in calibration is a function of the difference in slope, and the difference in UO/U.

Fig 7E (actually all Fig 7). As previously mentioned, I copied raw data into the supplement twice, and not the corrected data we plotted. I apologize for this error.

Fig 12. The session plotted is 210046. The original session was 170124. We are happy to edit the caption to make this clear.

References:

As in the manuscript, plus:

Foster, D. A., and Ehlers, K.: (1998),  $^{40}\text{Ar}$ - $^{39}\text{Ar}$  thermochronology of the southern Gawler Craton, Australia: Implications for Mesoproterozoic and Neoproterozoic tectonics of East Gondwana and Rodinia, *J. Geophys. Res.*, 103(B5), 10177–10193, doi:10.1029/98JB00151.

Kohn, B. P., Gleadow, A. J. W., Brown, R. W., Gallagher, K., O'Sullivan, P. B. & Foster, D. A.: (2002) Shaping the Australian crust over the last 300 million years: Insights from fission track thermotectonic imaging and denudation studies of key terranes, *Australian Journal of Earth Sciences*, 49:4, 697-717, DOI: 10.1046/j.1440-0952.2002.00942.x

Vogt, M., Schwarz, W.H., Schmitt, A.K., Schmitt, J., Trieloff, M., Harrison, T.M. and Bell, E.A., 2023. Graphitic inclusions in zircon from early Phanerozoic S-type granite: Implications for the preservation of Hadean biosignatures. *Geochimica et Cosmochimica Acta*, 349, pp.23-40.

Signed,

Charles Magee, on behalf of the team.