Formal Review (Paul Bons, 9 Feb 2021):

I already commented briefly in my unsolicited comment on the interesting and provocative manuscript of Behn et al. that proposes a novel (at least in glaciology, I believe) way to address the question of grain size in glaciers and ice sheets and its relationship with the rheology and stress exponent for power-law creep of ice. I was fortunate that by the time I was asked for a review, one thorough review was already published. I concur with the anonymous reviewer and need not repeat her/his comments.

I hope the manuscript by Behn et al. will be published in TC as it gives the community valuable food for thought. However, I would suggest to first address a few issues: 1) Does the paradox on which the paper is based really exist? 2) Grain-growth parameters may be over-simplified. 3) The merits of alternative explanations for the grain size - stress relationship could be discussed more.

These issues are discussed in more detail below.

Thank you for your thorough reading of our manuscript. Below we respond in detail to the issues that you have raised.

In the section starting at line 40, a crucial aspect is missing. Glen and some other authors made it very clear that their stress exponent was determined for the minimum strain rate/maximum stress and not for steady state. Comparing the low n (≈3) at very low strain (about 1-3%!) with high-strain steady-state flow may be like comparing apples and oranges.

The manuscript is based on the "paradox" mentioned in line 59. Simply put it is postulated that experiments indicate a stress exponent n of either ca. 1.8 (low stress) or ca. 4 (high stress), while natural flow is closer to n=3, the value generally (and uncritically!) used in flow modelling. The question is whether this paradox really exists. In lines 28- 34 it is argued that natural flow is consistent with n≈3. Although several studies indeed come to this conclusion, others do not. For example, Bons et al. (2018) deduced n≈4 for a large area of the Greenland Ice Sheet (excluding the divides, ice-sheet margins and ice streams), while Pettit & Waddington (2003) find n≈1 at divides. Glen (1955) himself wrote "... it is noteworthy that practically observable long-time creep rates, as in a glacier, would probably depend on a higher power of the stress than the 3.2 found here", although he did not actually determine this in natural ice. Cuffey and Kavanaugh (2011) write "we conclude that the effective n must be between 2.6 and 5.1 (99% confidence). The best match occurs with n ≈ 3.5". However, in the conclusions they also write "... supports the nearly universal practice of treating ice as an n = 3 nonlinear fluid in analyses of glacier flow". This may be symptomatic: despite evidence or indications to the contrary, some authors appear to (want to) stick to n=3, even if the data are inconclusive or allow alternatives. Another example is fig. 14 in Budd & Jacka (1989). They plot surface velocity/height against driving stress and find a best fit with a slope between n=3 and n=4. However, assuming n=3, they interpret the range in data in terms of temperature differences. Close (re-) assessment of the literature shows that there is quite abundant evidence for n unequal to 3 for natural ice flow, even though the literature unfortunately does not always fairly acknowledge this. I suggest the authors: (1) qualify their basic starting assumption that natural ice follows n≈3 (2) and include *in their following analysis what the consequences would be if n for natural flow is not 3, but perhaps indeed 4 as some claim to have measured in nature. Would this, for example, mean no contribution of GBS? Would the wattmeter work and give reasonable results?*

This is a fair criticism. As Dr. Bons correctly points out, observational studies find a range of values for the effective stress exponent in deforming ice sheets and glaciers. Our model provides a framework in which to interpret such variations depending on the relative contributions of dislocation creep and grain boundary sliding—which are in turn modulated by the grain size. This point is illustrated in Figure 4b, where we show how the effective stress exponent can vary as a function of strain rate. The goal of our study is really to emphasize the point expressed in Dr. Bon's initial comment (22-Nov-2020) that "*Glen's law with n=3 is usually assumed uncritically and other parameters, such as basal friction coefficients are derived instead*." This is an important point, because if variations in grain size and/or other aspects of the flow regime result in an effective stress exponent that deviates from the Glen law value, then any other parameters (e.g., basal friction) derived from models that assume n=3 will be incorrect.

In my unsolicited comment I already briefly addressed the grain-growth "constant" K and the grain-growth exponent p. The authors use p≈6, based on natural grain sizes in drill core and experiments with bubbly ice. There are a number of issues that I would ask the authors to consider.
(1) The exponent p reflects the scaling of the governing process(es). If grain growth is driven by unrestricted reduction of grain-boundary curvature and grain-boundary velocity is linearly proportional to the driving force (curvature), p should be 2. Restricted grain-boundary movement due to pinning or drag leads to a slow-down of growth, which gives a growth curve that may be fitted with a power law, but which is not a power law. The exponent p is "effective" or "apparent", but has little physical meaning and cannot be regarded as a universal material property. Growth then just does not follow a power law. If bubbles hinder growth, the effective p will depend on bubble size and distribution, relative to grain size (Arena et al., 1997; Roessiger et al. 2014). The main factor is probably the fraction of boundaries that is hindered in their movement by bubbles. If that fraction is small at the equilibrium grain size, the exponent p is expected to be close to 2, as most boundaries simply "don't know that they are in bubbly ice". In a grain- growth experiment that runs for long enough, one inevitably comes in the range where a significant number of boundaries interact with bubbles, which slows down the growth. The effective mobility of grain boundaries goes down, which raises the apparent p. This apparent p may not be relevant to the wattmeter if grain sizes are below this interaction range. It should be noted that in the numerical simulations of Roessiger et al. (2014) p is always 2, just because of the scaling of the numerical simulations and governing equations. However, the growth curves would give a wide variety of p>2 values, if one would erroneously assume a power law.

We agree with the argument here—namely, that the power law exponent reflects an "apparent" exponent, or perhaps is better thought of as an empirical exponent. The key here is that a power law is not necessarily "erroneous" – as it is an empirical fit. Where problems can arise is when the empirical relationship breaks down when extrapolated outside of the range of conditions where it was quantified. Determining how to deal with this was actually one of the most difficult aspects of this study. In the end, we were struck by the correspondence of how well the extrapolation of the power-law fit predicted the grain size in the shallow parts of the ice sheets (where dynamic recrystallization is not active) – which supports applying the empirical p value, but indeed does not prove its applicability.

As pointed out by Dr. Bons and *Arena et al*. (1997) the role of pores can also be thought of as changing the *K* value in the grain growth law. If *K* varies with the microstructure (bubble size / bubble topology) and this scales with grain size, then *K* will be proportional to some function of grain size *f(d)*. In our formulation, the empirically fit p-value is mathematically similar to a *K* term with a power-law relationship to grain size.

Thus, while we acknowledge the importance of more sophisticated models that specifically account for how bubble mobility and bubble size impact grain growth, with the caveats described above, we note that an effective grain growth exponent of order 6–7 can fit both the laboratory and ice core data (Figure 2), as well as the grain growth in the shallow part of the GISP2 ice core (Figure 7).

A revised manuscript will provide a more detailed description of the assumptions that have gone into our parameterization of the empirical grain growth exponent, as well as a discussion of potential future theoretical developments in this area.

Arena, L., Nasello, O. B. & Levi, L., 1997, Effect of Bubbles on Grain Growth in Ice. *J Phys Chem B*, v. 101, 6109–6112.

(2) K is also not a universal constant, because it depends on the microstructure. This was actually one outcome of my very first paper: Bons & Urai (1992; I was so proud that I sent reprints to my whole family!). Static grain growth typically leads to a particular microstructure (grain shape and size distribution): a foam texture as in a soap froth. Changing the microstructure means changing K. Growth experiments are probably often hampered by this effect: it takes some growth to establish the steady-state growth rate. Measurements of K and p should only start after this is reached. Roessiger et al. (2014) therefore suggest a grain size increase of at least about 4-5 times. The resulting K is for static grain growth and does not apply to a dynamic grain-size equilibrium under consideration in the manuscript, where the microstructure is expected to be quite different. The distribution of bubbles may also be different during deformation compared to static experiments (Steinbach et al. 2016). It is not clear if a different, but constant K applies, or that K is a function of stress and/or strain rate.
The bottom line is that one should not consider a single, constant p and K. It is very well possible that p=2, but K varies depending on a variety of factors. How would this affect the analysis?

Following on the points made above, we can describe these issues as part of the assessment of uncertainty in how our assumption of the power law form of the grain growth law impacts our interpretations. As a starting point, it is instructive to compare our model predictions with data for grain growth under bubble-free conditions. In doing so, we are essentially assuming that the "drag-drop" condition leads to no hinderance of grain boundary mobility. Extrapolation of the bubble-free grain growth data to natural conditions predicts extremely large grain sizes. For example, using the grain growth parameters from *Azuma et al*. (2012) Exp. 15 our application of the wattmeter predicts grain sizes > 100 mm at all depths in the GISP2 core. These predictions are of course significantly larger than the observed grain sizes, emphasizing that natural settings likely still include limited mobility owing to pinning, which is captured in our use of an empirical grain growth exponent.

Azuma, N., Miyakoshi, T., Yokoyama, S. and Takata, M., 2012, Impeding effect of air bubbles on normal grain growth of ice, J. Struc. Geol., 42(C), 184–193, doi:10.1016/j.jsg.2012.05.005.

Line 83: " However, the piezometer does not account for the physical processes that control ice grain size - namely the competition between grain growth and grain-size re- duction via recrystallization (e.g., Alley, 1992)." I suggest qualifying this rather sweeping sentence. There is a huge body of literature in materials science, metallurgy, geology, etc. on the physical processes that determine the piezometer. These models cannot be dismissed as "simple", nor do all say that grain size is the inverse of stress. The authors cite Jacka and Jun (1994). The authors of the paper are T.H. Jacka and Li Jun. The header of the original printed paper reads: "Jacka and Li: Steady-state crystal size of deforming ice". I therefore assume that the surname is "Li", not "Jun" and the Chinese convention of surname first was used. They do not find that grain size is inversely proportional to stress, but by an exponent of about -1.5. I do appreciate that the Jacka & Li piezometer is plotted in fig 3. It plots pretty much exactly on the boundary between the two deformation mechanisms as is acknowledged in the manuscript. So far, the data of Jacka & Li appear the only experimental grain size - stress data published in the literature and they would at first sight strongly support the de Bresser model. The slope of the piezometer is actually quite in line with that found for several other minerals, as pointed out by Jacka & Li and de Bresser et al (2001). Considering that natural flow of ice appears to be faster than experiments predict (compare the n=4 rates in Bons et al. (2018) with those used in the manuscript), the difference between gran size predicted by experiments and natural ice may be due to the infamous and "ad-hoc" enhancement factor. Line 284 is of interest: "Overall the piezometer [of Jacka & Li, 1994] results in smaller strain-rates throughout most of the column and a significantly higher effective stress exponent (neff ~ 3.9), similar to the experimental value for dislocation creep." This n≈4 is exactly what is proposed by some authors for natural flow, which would fit very well with the piezometer. I suggest not to be too dismissive of the de Bresser model and the data of Jacka & Li (nor assume that natural flow has n=3).

Both of the points are fair, and can be addressed with a more thorough discussion of the limitations and assumptions involved with the application of piezometers. The advantage of the wattmeter is that it provides scaling relationships for both steady state and transient grain sizes. The basic components of the wattmeter (as a competition between grain growth and grain size reduction) are actually based on the same logic as the piezometer model of *Jacka & Li* (1994), with the added insight that the grain size reduction processes involve the increase in internal energy within the crystals that drive grain size reduction – and how these depend on the product of stress times strain rate.

Regarding the field boundary hypothesis of de *Bresser et al*. (2001) – the basic idea to that model is that the driving force for grain boundary reduction becomes negligible when diffusion creep dominates. However, this is not applicable for the field boundary between DisGBS and dislocation creep, where easy slip on the basal plane of ice will produce intracrystalline deformation, similar to observations in olivine (e.g., *Hansen et al*., 2012). These observations suggest that the similarity with the de Bresser field boundary may simply be a "compelling" coincidence – but certainly one worth noting in the revised the manuscript.

de Bresser, J., Ter Heege, J. and Spiers, C.: Grain size reduction by dynamic recrystallization: can it result in major rheological weakening?, Int. J. Earth Sci., 90, 28–45, doi:10.1007/s005310000149, 2001.
Hansen, L. N., Zimmerman, M. E. and Kohlstedt, D. L.: The influence of microstructure on deformation of olivine in the grain-boundary sliding regime, J. Geophys. Res., 117(B9), doi:10.1029/2012JB009305, 2012.
Jacka, T. H. and Li, J.: The steady-state crystal size of deforming ice, Ann. Glaciol., 20, 13–18, doi:10.3189/1994AoG20-1-13-18, 1994.

Line 86: Typo in Roessiger Line 376: typo in Kipfstuhl

Thanks – will be corrected