This paper further explores the design space of cost-efficient ice flow models that can capture both plug and shear flow. The more faithful but more expensive model has been known for a long time -- the first-order or Blatter-Pattyn equations, derived through a perturbative simplification of the full Stokes equations. Several authors have proposed further simplifications of the Blatter-Pattyn equations using a variety of techniques, for example the multi-layer shallow shelf approximation of Jouvet et al. (2015), which works by a finite difference semi-discretization in the vertical dimension. This paper proceeds instead by using a Galerkin-type semi-discretization in the vertical dimension. This idea has several precedents in the literature; the only citation that I’d add is Langdon and Raymond (1978), which is the earliest reference I can find of anyone trying this idea.

This paper does a good job illustrating the differences between the model they implement and closely related models that have appeared elsewhere in the literature, e.g. Brinkerhoff and Johnson (2015) and the work of myself and coauthors this year. Assuming that this model is part of ISSM or will be merged, this paper is also valuable as documentation of the underlying mathematics for future users. Having implemented a model in a similar spirit, I found the vertical averaging they used to be a convincing approach for keeping the computational cost down while preserving as much physical accuracy as possible. I also found their observation that the model has a bias for faster flow in ice sheet interiors to be especially important and motivating of future work that might correct this problem.

I have a few minor nitpicks. I try to think of any paper like this not just as a description of how the model was implemented in a particular software package but as a guide to the poor souls who might have to reimplement these ideas in whatever programming language people are using 20 or 30 years from now. The appendix shows a number of painstaking computations, and presumably getting all of these details correct is necessary to correctly implement the mathematical model that the authors have specified, which could be quite error-prone. Obviously I’m biased here but to me this speaks to the utility of software frameworks like FEniCS, dune-fem, Devito, and Firedrake, which automate away much of the symbolic derivation and generation of integration formulae. This becomes especially apparent when considering what to do in the face of a non-isothermal ice column. Any of the tools above (provided that they can work with tensor product elements) will generate the correct vertical quadrature formula to include temperature...
effects. The end of section 5 suggests that much of this symbolic manipulation would need to be redone to account for variable temperature. (If I've misread that then it should be reworded to make it clear that it doesn't.) In any case this more my opinion so take it with a grain of salt.

I wasn't sure where the footnote at the end of section 4.3 came from regarding an iterative solver with time complexity O(n^2) came from. A well-tuned multigrid or multilevel solver for a 2D elliptic equation converges to an accuracy comparable to the discretization error much faster than this -- O(n log n) or even O(n) if you include a global coarse level. See *Domain Decomposition Methods* by Toselli and Widlund. Sparse direct solvers are asymptotically faster than O(n^2) for 2D elliptic finite element problems too.

My final concern is that the higher-order model has a minimization principle which may not be preserved with the vertical averaging approximation that the authors use. Having a convex minimization principle means that you get more guarantees about convergence rate when you use Newton's method with a globalization strategy such as line searches or trust regions. The authors instead used Picard's method, which I've found often requires some manual and problem-specific tuning to get an acceptable convergence rate. This strikes me as more of a tradeoff than some fatal flaw and the gains in speed of the MOLHO model might outweigh any advantages of having a convex minimization principle.

In any case I recommend for publication almost as-is and look forward to seeing future work on paleo simulations with higher order models!