General Assessment

The paper by Spange et al. compares different methods to regularize ductile localization associated with thermal runaway in the ductile regime. This is a challenging and important issue for the community, since the length scales of shear zones produced by this process in the Earth's mantle may be on the order of nanometers to millimeters and evolve over seconds, while models typically operate at scales of hundreds of kilometers and millions of years. Regularization is therefore essential.

The paper is well written and the results are presented clearly. The main weakness, in my opinion, is the lack of separation between the physical aspects and the numerical methodology in Sections 2 and 3. I believe this could be addressed by a modest reorganization of the text.

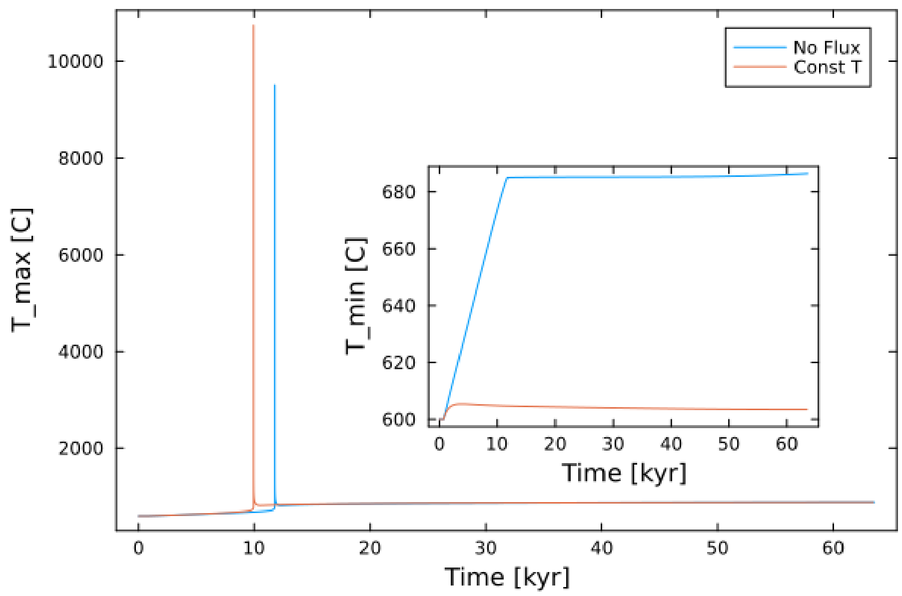
Broader Comments

- The paper is primarily methodological, but a stronger discussion of the underlying physics would be very valuable. In particular, it would help to reflect on how the approximations in the heat equation, mass conservation, boundary conditions, and the assumption of a single shear band simplify or complicate the problem compared to more Earth-like conditions.

  Reply: We have added a new section (6) to discuss the simplifications and design choices regarding inertia, gravity, thermal expansion, adiabatic heating, and grain size evolution. The section also discusses the case of multiple perturbations that could act as nucleation sites for thermal runaway.

  Regarding thermal boundary conditions, they have no impact during the runaway as the heating rate in the shear zone is orders of magnitude faster than thermal diffusion. In fact, once runaway starts, heating close to the boundaries essentially ceases as all dissipative work occurs in the center. Using constant temperature boundaries would increase runaway intensity as the host rock would be cooled during the LTP-phase, increasing the viscosity contrast between host rock and shear zone. We are attaching the results of a 1D model below.

  Regarding the periodic boundary conditions, we can create a similar result with pure shear boundary conditions. In this case, the shear zone forms at an angle of 45°.



Maximum and minimum (inset) temperature for no-flux boundary condition (blue, used in our study) and constant temperature boundary condition (orange) in a 1D model. Temperature is saved at cell centers, so even for constant T boundaries, there is a little bit of increase in T_min.

- It would also be interesting to discuss whether the length scales or gradients introduced by the regularization might have a physical meaning. Ideally, the problem should be governed by material parameters rather than numerical parameters. While the proposed regularizations alleviate mesh dependence, they introduce sensitivities to new parameters that are equally numerical. Introducing physically motivated mechanisms that could act as natural regularization would strengthen the study.

  Reply: We agree that physically motivated regularizations or, ideally, actual physical mechanisms would be more satisfying. However, such processes can introduce new uncertainties if they are not constrained by sufficient experiments. Furthermore, adding physical processes can add more non-linearities to the system and potentially make the numerical solution even less stable. We introduced a new subsection (4.3.5) to discuss the relation of both regularization approaches to physical processes, the value of adding more physical processes, and the potential problems related to these additions. Furthermore, we added a third regularization method with the latent heat of melting based on the suggestions of reviewer 3.

Suggestions for Reorganization

1. Introduce the governing equations (Sec. 2.2.1), rheology (Sec. 3.2), and density (Sec. 3.3) **before** describing the model setup (Sec. 2.1) and the 1D simplifications (Sec. 2.2.2).

   Reply: We adopted this order for section 2.

2. In Section 3, begin with the spatial discretization, then proceed through the pseudo-transient scheme, nonlinear viscosity iterations, and finally regularization.

   Reply: We adopted this order for section 3.

3. Please add the thermal boundary conditions to the description of the model setup.

   Reply: We have added the thermal boundary conditions to the beginning of the model setup section.

Notation and Tables

- Since you use λ as a regularization length scale, I suggest using κ for heat diffusivity (instead of λ). Consequently, use *k* for heat conductivity in Eq. 17 and in the text (L.122), and update the last line of Table 1. Currently, you list conductivity but label it as diffusivity.

Reply: We agree that using λ and λ_{reg} with different units is inconsistent. We have therefore replaced λ (thermal conductivity) by k.

- In Table 1, group the elastic bulk modulus *K* with the shear modulus *G*. It does not need the subscript "b". You could compute *K* using Poisson's ratio of 0.25 and then remove Poisson's ratio from the table, along with Eq. 28. For an isotropic elastic material, only *K* and *G* are needed; Poisson's ratio can simply be mentioned as an explanatory note.

Reply: We added bulk modulus to Table 1, next to shear modulus, and removed Poisson ratio. We now note that bulk modulus is computed from G and Poisson ratio at the bottom of the table and removed Eq. 28. We decided to keep the subscript "b" to avoid potential confusion with the new symbol for thermal conductivity (k).

Technical and Line-by-Line Comments

- **L.23:** While plate-scale models may overestimate shear zone width due to coarse resolution, could you add a reference or explanation for grain-scale models potentially underestimating them? If Braeck et al. are correct, shear zones could be as thin as nanometers—smaller than grain-scale models can capture.

  Reply: This section was poorly worded. We split the sentence in two and extended the second part to explain that a grain scale model might be too small cover the relevant geological context to accurately compute the size of the shear zone.

- **Eq. 27:** Please provide a brief explanation of how it follows from the mass conservation equation. I suggest moving this equation (with the explanation) next to Eqs. 2 and 4, without devoting a full subsection to density.

  Reply: Agreed. We moved this equation to the governing equations section and added that it follows from the combination of equations 2 and 4 and the integration of the changes of pressure and density. We then removed the density subsection.

- **Eq. 3:** Explicitly note that $\tau_{ij} e_{ij}^{vi}$ is the shear heating term, and that all dissipated work is assumed to convert into heat (i.e., no grain size reduction).

  Reply: We added this clarification below equation 3.

- Consider placing remarks on neglected terms close to the relevant equations:
  - inertial and gravity terms after Eq. 1,
  - adiabatic and radiogenic heating after Eq. 3,
  - thermal expansion after Eq. 4.

  Reply: We would like to keep the equations in one block. To improve clarity, we now mention which equations the neglected terms belong to.

- Use e instead of ε ˙ \dot{\varepsilon} for the deviatoric strain rate, consistent with your notation for deviatoric stress (τ vs. σ). Also, define τ explicitly, just as you do for the strain rate tensor.

Reply: We would like to stick with the use of \dot{\varepsilon}. We do not mention volumetric strain rate or full strain rate anywhere in the text and clearly state how we define \dot{\varepsilon} in equations 5 and 6 (now 6 and 7) and the lines in between. We also think the use of e is not ideal as it can be mistaken for the Euler number. One other common notation is \dot{\varepsilon}' but this is not ideal when specifications like viscous or dislocation are added in the exponent as we do in equation 18 (now 8). The accepted papers of Popov et al. and Li et al. in this issue also use \dot{\varepsilon} for deviatoric strain rate.

- **L.84:** effective shear viscosity (clarify wording).

  Reply: We added this clarification.

- **L.89:** Gravity does not need to be listed here, as it does not affect Eqs. 7–8 if initial stress conditions already include gravity and they should.

  Reply: Gravity is not taken into account for the initial stress conditions. The initial stress is zero in the entire domain. We list it because if we were taking gravity into account, divergence of velocity would not necessarily be equal to zero as material would compact due to gravitational forces.

- **L.97:** Since you previously called them "deviatoric," maintain that wording consistently.

  Reply: In order to not clutter the text in the following sections, and since we never discuss volumetric stress or strain rate, we instead added a statement at the end of the governing equations section that all following references to stress and strain rate refer to the deviatoric components.

- **Eq. 17:** Replace λ with κ

  Reply: All instances of λ were replaced by k.

- **L.123:** Use $k=\kappa/(\rho Cp)$.

  Reply: We now use $\kappa = k/(\rho Cp)$ in line with our change from λ to k.

- **L.145–150:** Please separate physics from numerics; for instance, move stabilization viscosity details to a new subsection of Section 3.

  Reply: We split this part of the rheology section from the physical part and it is now the section "3.3 Viscosity update"

- **L.203:** Clarify why τ is used instead of change in strain rate? If this comes from the cited reference, a short explanation in the text would help.

  Reply: This decision does not come from the cited references. We chose temperature and stress change as thermal runaway is driven by the conversion of elastic energy (i.e. stress) to thermal energy (i.e. temperature). We included this reasoning to the manuscript.

- **L.260–265:** A short note on how elasticity may improve conditioning would be useful.

  Reply: The comment appears to imply that we do not use elasticity, but we do. Elasticity acts as a buffer when stress conditions change rapidly. In areas of low viscosity, it limits the magnitude of deformation. We added a sentence about this to the section in question.

- **L.355/L.375:** You mention using the same parameters in 1D and 2D, but then describe differences in how the anomaly is defined. Please clarify.

  Reply: The parameters are the same. However, there is a small difference in how the weakening parameter ω was applied. This is also described in the Model setup section. The difference in applying ω was necessary due to the geometry of the 1D model. In 1D, weakening the LTP flow law means that the entire 1D model cannot reach 1.8 GPa, due to the shear stress being constant throughout the model. In 2D, only the small inclusion cannot reach this stress while the rest of the domain can. We have restructured section 5 which should help clarify this issue.

- **L.370:** A comment on the effect of periodic boundary conditions would be helpful. In principle, once the shear band spans the entire domain, 1D and 2D solutions should converge and not diverge as 1D is an infinite shear band.... I don't really understand.

Reply: Once the rupture spans the entire 2D domain across periodic boundaries, it indeed represents an infinite shear band like the 1D model. However, there are still differences between the two settings. In 1D, the entire shear zone is represented by a weakened rheological flow law whereas in 2D, only a very small portion of the shear zone (the initial anomaly) has a perturbed flow law.

What is probably more important is the heterogeneous stress field in 2D (Fig. 8, left column). The rupture tips increase the stress ahead of themselves. In the 1D model, runaway releases 1.8 GPa of stress. In 2D, the average stress across the model domain is the same, but in the stress lobes ahead of rupture tips, the stress exceeds 2 GPa. When the two tips unite, they have more stress (i.e. energy) to release than in the 1D example and consequently reach larger temperature and slip velocity. We added this explanation to the section in question.

Once the stress has been fully released in the 2D model, we would expect it to be very similar to the 1D case again.

- **L.382:** When the timestep drops to seconds, and given the large stress drops you show, is it still valid to neglect inertial terms? Could you add a small comment on that?

  Reply: It is correct that inertia may start to play a role at peak runaway conditions and reduce accelerations.

  For 2D, we can make a simple estimation of the inertia term. Figure 8e shows the peak velocities during the runaway. For an upper limit estimate, we assume that a grid node accelerates from stationary to peak velocities (5 mm/s) in one time step (15 s). With a density of about 3300 kg/m^3, this yields about 1 kg/(m^2 s^2). As a comparison, the stress gradient term reaches up to 10^7 kg/(m^2 s^2). So, inertia should still be negligible.

  In our 1D models, we cannot make this comparison as the stress gradient is always zero following the momentum conservation. For cases with very low eta_reg or lambda_reg, the inertia term would also be larger than 10^6 kg/(m^2 s^2). Here, it should have a significant impact. We added these considerations in the new section 6.1.

Recommendation

I recommend **minor revisions**, focused mainly on clarifying the separation between physical and numerical aspects, improving the discussion of the physical meaning of regularization, and addressing the minor notational and organizational issues listed above.

Laetitia Le Pourhiet

Reply: We thank the reviewer for their encouraging words and constructive criticism. We have adopted the proposed separation between physics and numerics. Furthermore, we have added a discussion of the physical effects of the regularization techniques alongside other small corrections.