Response to Reviewers' Comments
Algorithm Selection and Auto-Tuning in AutoPas

Manuel Lerchner

I sincerely thank the reviewers for their valuable feedback and constructive suggestions, which helped shape the final version of the paper.

In this response letter, I address each reviewer's comments and describe the changes made to the paper accordingly. Since the reviewers were anonymous, I will refer to them in an arbitrary order.

Reviewer 1

Comment 1: [**Title Change**] The title does not directly convey the main contribution of the paper.
Response: Since the main focus of the paper is on the drawbacks of existing auto-tuning approaches, with a special focus on the idea of "Early Stopping", I have updated the title to "*Auto-Tuning with Early Stopping in AutoPas*".

Comment 2: [**Wording**] AutoPas is a library, and can't be directly compared to full MD engines.
Response: I changed the introduction to introduce AutoPas differently. Namely, as a logical improvement over "single-implementation" MD engines, rather than a direct competitor, this avoids the mentioned issue.

Comment 3: [**Wording**] ls1-mardyn is not independent of AutoPas.
Response: I explicitly added that I just consider the version of ls1-mardyn without AutoPas.

Comment 4: [**Extra Info**] Linked cells can also only be rebuilt every n steps when using a skin-factor.
Response: I decided to **not** include this information for clarity and as it is not directly a key feature of standard linked cells.

Comment 5: [**Wording**] Simulation speed is unclear.
Response: As proposed, I changed the wording to runtime.

Comment 6: [**Duplicate Info**] The infeasibility of exhaustively testing all configurations is mentioned multiple times.
Response: As proposed, I changed the order of the sections to avoid this. Now I first introduce all tuning strategies and then use the problems of sometimes suggesting bad / too many configurations as a motivation for early stopping.

Comment 7: [**Clarification**] The exact part of the Figure readers should look at is not clear.
Response: As proposed, I added stated the section I refer to in the text.

Comment 8: [**Question**] Does the data of md-flexible runs contain MPI runs?
Response: Yes, the data of md-flexible contains some MPI runs (which also don't benefit from returning). I made this more clear in the text.

Comment 9: [**Reordering**] The last comparison section between the MD engines is out of place.
Response: I agree and moved the comparison to a "Related Work" section at the start of the paper. As proposed: This change also allows for a nice transition to the need for auto-tuning libraries such as AutoPas.

Reviewer 2

Comment 1: [**Possibly Missing Citation**] It is not clear where the idea of "stopping further samples" comes from.
Response: As I came up with this idea independently, I did **not** include a citation. I'm also unaware of any prior work that directly suggests this idea.

Comment 2: [**Reordering**] Reordering of the comparison section.
Response: Also mentioned by reviewer 1. See the response above.

Comment 3: [**Question**] Why are the bar charts separated in "tuning" and "simulation" time?
Response: The original text already included this information indirectly in the analysis of both limiting cases, however, I agree that it requires some prior knowledge to connect the dots. To make this clearer, I added two short sentences to the text explicitly explaining the effect of $allowedSlowdownFactor$ on the duration of tuning phases.

Reviewer 3

Comment 1: [**Clarity**] The introduction focuses too much on the comparison of MD engines.
Response: This has already addressed in parts by the previous reviewers. I rephrased the introduction to not focus on the comparison of MD engines as much.

Comment 2: [**Missing Reference**] The GitHub link to AutoPas is never mentioned in the text.
Response: I added a reference to the GitHub repository in the AutoPas section.

Comment 3: [**Reordering**] The comparison section is out of place.
Response: Also mentioned by reviewers 1 and 2. See the response above.

Comment 4: [**Wording**] "combine to form a configuration" sentence structure seems off to me
I changed it to the passive voice: "... are combined to form a configuration".

Comment 5: [**More details**] Is the load estimator used in single-ranked runs or in MPI runs? + How is the coloring affected by different cell size factors?
Response: I think those questions are very implementation-specific and feel too detailed for the scope of the paper. I decided to **not** include this extra information.

Comment 6: [**More details**] How does Bayesian Search work?
Response: I think this is again too detailed for the scope of the paper. For curious readers, I already provided the reference to the original paper, so I did not change anything here.

Comment 7: [**Wording**] Change ”Benefits of Auto-Tuning” to ”Benefits of AutoPas”.
Response: I feel like the mentioned points are the benefits of auto-tuning in general, and AutoPas is just the example used in this paper. Also the text explicitly mentions when some aspects are specific to AutoPas. Therefore, I decided to **not** change this.

Comment 8: [**Clarity**] What is md-flexible?
Response: I added a short explanation of md-flexible to the text.

Comment 9: [**Clarity**] Why do most scenarios not benefit from re-tuning
Response: The text already contains a rough explanation of this. Namely, that many scenarios tend to be stable over time. I'm unaware of any rigorous proof of this / which formal conditions need to be met for this to be true. Therefore, I am unable to provide a more detailed explanation here. Maybe this could be a topic for future work.

Comment 10: [**Clarity**] What is the ”Exploding Liquid” scenario?
Response: I added a short YouTube link for a video of the scenario as well as a link to the .yaml file for the scenario.

Comment 11: [**Question**] General Questions about the scenario.
Response: The Exploding Liquid scenario was just meant as a showcase for the idea of Early Stopping, but I feel the general benefits would be quite similar for other scenarios. As stated in the text, more (scenario-specific) benchmarks are needed to make a general statement about the benefits of Early Stopping. I think most of the questions about the scenario are answered by the added .yaml file, so I decided to **not** change anything here.

Comment 12: [**Clarity**] Early Stopping should not be able to increase the runtime.
Response: As early stopping essentially reduces the number of tests for a configuration, the resulting configurations tend to be ”noisier”. This can lead to worse runtimes when timing fluctuations cause a worse configuration to win. This is however already mentioned in the text, so I decided to leave this as is.

Comment 13: [**Error**] Ls1-mardyn can converts to SOA on the fly. **Response:** I added this information to the text to make this more clear.

Summary of Changes

The biggest change was the reordering of the comparison section, which was mentioned by all reviewers. I also changed the title to better reflect the main contribution of the paper. The content of the paper was mostly left unchanged, with some minor additions to clarify some points aswell as slight reordering of some sections.

I hope that the changes made address the reviewers' concerns and improve the quality of the paper. I am looking forward to hearing your feedback on the revised version.

Sincerely,
Manuel Lerchner