- *Authors*: Chloé Mercier, Thierry Vieville
- *Title*: Algorithmic ersatz for VSA: Macroscopic simulation of Vector Symbolic Architecture
- *Submission Type*: 'Regular Paper'
- *Tracking number*: 838-1839
- *Assigned editor*: Alessandro Oltramari <alessandro.oltramari@us.bosch.com>

Editors-in-chief feedback

Thank you very much for considering this draft. We fully understand the need for a major revision and detail how we address these reviewer comments here, using italics.

... based on the reviewer's comments, your paper requires major revisions.

Review #1 feedback by Alessandro Oltramari*<alessandro.oltramari_960>*

Recommendation: Minor revision

Detail Comments: The authors have made significant strides in addressing the initial concerns, notably by expanding their empirical validation to include large-scale experiments and by clarifying the scope and limits of their symbolic representation choices. Their responses demonstrate both engagement with the critique and a commendable effort to strengthen the theoretical and practical grounding of their approach. This seems to be the case not only within the scope of the specific review but also when addressing comments from the other reviewers. The revised manuscript is notably improved and offers an interesting perspective in connecting neural computation with symbolic processing. Some minor revisions are required:

Thanks for this caring and positive evaluation of this new version, and thanks a lot for your help regarding final corrections.

- Table formatting is not uniform across tables. Also, I'd strongly advise against vertically positioning Table 4;

Indeed, sorry, now corrected, thanks to both remarks.

- Missing commas, such as: "Interestingly enough, deriving a macroscopic ersatz of complex instead of real VSA specification seems straightforward." –> Suggestion: "Interestingly enough, deriving a macroscopic ersatz of complex instead of real VSA specification seems straightforward."

Thanks done, including rereading the whole draft ... [@Chloé .. help :)].

- "explicitized" is a bit archaic and seldom used –> Suggestion: replace with "made explict" or "elaborated" where suitable

Thanks for the advice, implemented.

- Unclear sentences: e.g., "This result is not directly useful for our purpose, but is in favor of using the VTB algebra, including at the mesoscopic level, despite a less performing computation time." Suggestion: "This result is not directly useful for our purpose but supports the use of VTB algebra, including at the mesoscopic level, despite its lower computational performance." I

Thanks for the advice, implemented.

- In general, I'd recommend an extensive proof-checking of the article before final submission.

`TBD [@Chloé .. help :)].`

Review #2 feedback by Anonymous

Recommendation: Minor revision

Thanks for the complementary work that helps make the present work acceptable.

Detail Comments: In symbol encoding (p. 3), we have verified for d=100...1000d that unary vectors are generated with a relative precision on the magnitude below 0.3%, while orthogonality is verified with a relative precision below 0.4%, and the noise standard deviation prediction relative precision is below 0.3%. However, I didn't find the detailed verification for these claims, which are foundational to the hypothesis testing described later in Appendix A and Footnote 5. For example, the explanation of how the difference between the magnitude of the generated vectors and the expected magnitude is very small—less than 0.3%—is not clearly outlined.

You are entirely right: following your important comment, we have reworked this part of the subsection and added details of the numerical simulation in Appendix A, in addition to the existing (but neither described nor correctly documented, sorry) related open-source code.

Following that, the author discusses partial knowledge encoding, a method for semantically encoding uncertainty knowledge, and how it can be implemented in common data structures. The author has addressed previous comments regarding the enumeration in the DS structure. However, further clarification is needed for the bundling and binding rules mentioned.

Yes, thanks for inviting us to improve further

Specifically, for bundling, the concepts of similarity and relatedness need further explanation. I believe cosine similarity is implied here, but it should be stated explicitly. Adding a single word could significantly reduce the ambiguity and jargon in this section.

Thanks again, we have rewritten these lines, to follow this advice.

Regarding binding, I have additional questions. The text states (p. 5): '...enjoys the property that the corresponding unbinding operator Bs1 allows retrieving s2.'

Firstly, I recommend avoiding the use of the word 'enjoys' in academic writing, as it can sound too informal and is not appropriate for a serious journal.

We understand, done.

Secondly, this phrasing raises questions for the reader: How and why can the operator 'retrieve' s2? The internal mechanism behind this operation, whether mathematical or computational, has not been explained and cannot be sufficiently clarified with the phrase 'can simply enjoy.' More detail is needed to explain how the unbinding operator works clearly. Then, in symbol indexing and specification.

Indeed, we have added an informal explanation regarding the internal mathematical mechanism behind this operation, before the technical development in Appendix C.

The equation 0=x-x'=(tau-tau')uk+nu(sigma+sigma'), up to the first order, with |tau-tau'|<sigma+sigma' < is referenced as being developed in Appendix A. However, it's unclear how this relates to the hypothesis testing of whether two vectors are orthogonal or not. Since tau represents the belief value and sigma is the noise deviation, the connection between the equation and the hypothesis test needs further clarification from my view.

Yes, we have rewritten this section to make the rationale for the design choice more explicit.

On page 15, the newly added benchmark. 'Let's consider two rather large VSA experiments' but it only has one experiment, which is the KJB dataset.

This was poorly formulated. Sorry, we realized two experiments (word search and word prediction), but with a single dataset. This is now corrected, and we added subsection titles for better readability of this experimental part.

The phrase 'Thanks to the ...' should be rephrased to something simpler, like 'A being modeled by B.'

Now rewritten, thanks.

A bigger issue is the lack of a descriptive explanation for Table 4. I had to try to understand it from the table caption alone. Specifically, what do you mean by the second block being the macroscopic prediction for the mesoscopic? Why is the macroscopic prediction of the bias and standard deviation not provided for the KJB data sample? I do not think this comparison yields meaningful results unless more explanation is provided since you are comparing the meso to the KJB sample to the macro to meso.

Oh yes, so a bad formulation, sorry: the simulation is based on the same data set, indeed, but we failed to make this explicit, now done. We have rewritten both the caption and the text that refer to this table. We have also canceled the display of the τ because it was not informative and was generating confusion.

We also have made explicit that we were not using the KJB data sample but the same *"synthetic´´ data set at both scales for the comparison.*

The practical experiment is still very preliminary with a better presentation though.

Yes indeed. We had emphasized this section, following your previous review, and made our best to be clear with respect to the reader.

Overall, I appreciate the authors' efforts in improving the work. They have added some empirical testing on the macro implementation.

Thanks.

However, the manuscript still suffers from a lack of clarification on the key technical and mathematical parts.

We confirm and have make th ebest re-reading we could along the whole manuscript.

Review #3 feedback by Kaushik Roy <*kaushikr@email.sc.edu*>

Recommendation: Major revision

Comments Note: I conducted extensive background research on the topics covered in this paper, and I am not an expert in these areas.

Detail Summary: The manuscript presents an "algorithmic ersatz" simulator for Vector-Symbolic Architectures (VSAs). The authors recast core VSA operations, symbol encoding, bundling, and their non-commutative Vector-Derived-Transformation-Binding (VTB), in closed-form equations that can be executed symbolically. They extend the framework with a scalar-based belief weighting scheme, propose a relation-map data structure, and release reference C++ code.

Thanks, for this summary.

Detail Comments: Although the algebraic derivations are correct and potentially helpful for large-scale neurosymbolic applications, empirical validation and exposition fall short of the expectations of a top-tier journal.

Strengths:

- Well-founded algebra: Formal proofs for VTB orthogonality, inverse, and noise characteristics are sound and extend prior work.

- Conceptual clarity: The graded-truth (tau-scaling) mechanism elegantly folds uncertainty into the same vector space. - Open-source implementation: A public GitLab repository promotes transparency and reproducibility.

- Potential impact for the VSA community: A symbolic-level simulator could accelerate experimentation without the need for costly mesoscopic simulations.

Thanks, for making explicit these elements.

Weaknesses:

- Limited experiments: Benchmarks are restricted to micro-latency tests with 256-dimensional vectors

`@Chloe : quoi que il veut dire par micro-latency tests à ton avis ?`

Your feedback is precious, because it means we have missed explaining properly that the macroscopic approach is independent from the dimension: it is the same calculation at any dimension, and only the result prediction precision is function of the dimension, computed in by a closed-form formula. This is now clarified in the paper. It is true that one experiment involves 256-dimensional vectors, so the macroscopic approach is also 256-dimensional, in this case, for comparison, whereas benchmarking has involved comparing mesoscopic versus

marcroscpic approaches at several dimensions up to the limit of the used laptop coomputer.

- Limited experiments: Lack real-world cognitive tasks, statistical rigor, and comparisons to diverse baselines.

We agree with these weaknesses:
- Lack of real-world cognitive tasks, because this modest paper only introduces the idea of simulating any VSA mechanism based on biding and bundling at a macroscopic level, this being indepent on the application task.
- Lack of statistical rigor, because we are in a determistic context, and it is not expected for the number of operations to vary with such algorithms, unless operational hazards at low-level (e.g., memory swap, multi-core temporary deadlock), but this is out of the scope of the considered issues.
Lack comparisons to diverse baselines, this a definite limitation of the study, which focus only on VTB for binding, whereas other operators will definitely chance the balance differently at ethe computation performances study. In order to take your feedback into account, we have developed a basic argument regarding the fact that as soon as, for bundling, it is faster to turn to macroscopic simulation for bundling, it must also be the case for any binding.

- Presentation issues: Dense prose, heavy reliance on footnotes for essential proofs, several grammatical errors, and awkward figure placement impede readability.

`TBD [@Chloé .. help :): je peux être ton stagiaireet faire du taff mais il me faut ton regard ¸, ].`

- Overstated claims: Assertions of biological plausibility are not backed by empirical evidence at realistic vector dimensions.

We have completely rewitten this section, because we did had introduced a misunderstanding: we do neither have any claim nor contribution regarding biological plausibility, and have only refered to previous work on NEF, this is now corrected. We also feel important to clarify that NEF biologically plausible implementation corresponds more to an numerical anchoring than a grouding as reviewed by Taddeo and Floridi 2005.

- Overstated claims: Assertions of performance gains are not backed by empirical evidence at realistic vector dimensions.

We again feel very sorry: this feedback is induced by the fact we have not clearly stated what we were investigating here. Of course the higher the dimension, the better the advantage to avoid mesoscopic calculations, and realy on a macroscopic simulation. But, sincerely, you are right again, we can not stand on saying "we have checked until 10^4 *and this should be even better at* $10^d, d > 4$*".*

Thanks to your feedback, we have now pushed our calculations up to $d = 8$ *which is up to our best knowledge beyond both VSA computations and the size human neural maps (e.g., up to 50 millions in the human hippocampus, see reference in the text.).*

`@vthierry ET CA FAUT QUE JE LE FASSE MAIS C EST SANS SOUCI`

- Incomplete related-work coverage: Recent non-commutative binding schemes and graph-VSA literature receive only cursory mention.

This was a lack, now completed thanks to your feedback. in the discussion; our "bad" :) excuse is that the work is now quite "old" but yes it was worth doing this state of the art comparison with recent publications.

`@vthierry ET CA FAUT QUE JE LE FASSE MAIS C EST SANS SOUCI`

- Scalability analysis gaps: The retrieval complexity and false-positive rates for the proposed relation map are not quantified.

This is absolutely right. We have now a new subsection "Relational maps/Relational map performances" that discussed this aspect in details, building on existing results in the field.

Your remark also raised another key limitation of our study : we simulate here "perfect" mesoscopic implementation which are, in fact, subject to errors such as false positives [**?**]. Such errors are definitely considered in our framework.

This is now made explicit in the paper.

Conclusion: The paper presents promising algebraic contributions and a functional codebase; however, substantial revisions, particularly ... a more in-depth empirical section, clearer writing, and stronger engagement with the current literature, are necessary before it can meet the journal's high publication standards.

Thanks, if sufficiently improved, it will be thanks to your feedback.