

Enhancing Materials Discovery with Valence Constrained Design in Generative Modeling

Corresponding Author: Professor Mingda Li

This file contains all editorial decision letters in order by version, followed by all author rebuttals in order by version.

Version 0:

Decision Letter:

**** Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your co-authors. ****

Dear Professor Li,

Your manuscript "Enhancing Materials Discovery with Valence Constrained Design in Generative Modeling" has now been seen by 4 referees, whose comments are appended below. You will see that while they find your work of interest, they have raised points that need to be addressed before we can make a decision on publication.

The referees' reports seem to be quite clear. Naturally, we will need you to address ***all*** of the points raised.

While we ask you to address all of the points raised, the following points need to be substantially worked on:

- 1) Please provide additional methodological detail to ensure clarity
- 2) Please provide additional comparisons, including quantitative comparisons in both performance and efficiency

Please use the following link to submit your revised manuscript and a point-by-point response to the referees' comments (which should be in a separate document to any cover letter):

Link Redacted

**** This url links to your confidential homepage and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this e-mail to co-authors, please delete this link to your homepage first. ****

To aid in the review process, we would appreciate it if you could also provide a copy of your manuscript files that indicates your revisions by making use of Track Changes or similar mark-up tools. Please also ensure that all correspondence is marked with your Nature Computational Science reference number in the subject line.

In addition, please make sure to upload a Word Document or LaTeX version of your text, to assist us in the editorial stage.

If you have any issues when updating your Code Ocean capsule during the revision process, please email the Code Ocean support team Cc'ing me.

To improve transparency in authorship, we request that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit [please visit](http://www.springernature.com/orcid) www.springernature.com/orcid.

We hope to receive your revised paper within three weeks. If you cannot send it within this time, please let us know.

We look forward to hearing from you soon.

Best regards,

Kaitlin McCardle, PhD
Senior Editor
Nature Computational Science

Reviewers comments:

Reviewer #1 (Remarks to the Author):

This manuscript addresses a critical limitation in generative AI for materials design: the tendency to produce chemically invalid or thermodynamically unstable structures. The proposed framework, CrysVCD, introduces a two-step generative process that first enforces valence balance at the composition level and then applies a diffusion model for structural generation. This approach is conceptually appealing because it integrates fundamental chemical constraints early in the pipeline, potentially reducing the inefficiency of generating and discarding non-physical candidates. The authors also incorporate property-guided generation and report improved stability metrics compared to prior models, which is an important advancement for practical materials discovery.

Despite these strengths, several concerns limit the impact and applicability of the method. First, the strict enforcement of charge neutrality at the composition level may inadvertently exclude realistic materials that achieve stability through point defects such as vacancies or interstitials. These defect-mediated mechanisms are common in functional materials and are essential for practical applications, yet the current approach screens them out. Second, the treatment of oxidation states appears simplistic. While the model uses a valence-labeled vocabulary, it is unclear whether it adapts oxidation states based on local coordination environments or simply maps common oxidation states from a predefined table. This raises questions about its ability to handle multivalent elements and complex chemistries where oxidation states are context-dependent.

The stability screening strategy also warrants scrutiny. The use of an energy-above-hull cutoff of 0.1 eV/atom is relatively permissive, but for metallic systems, their stability margins are much tighter. As a result, the reported improvement in E_{hull} distribution, while notable, may not translate into practically synthesizable materials. Furthermore, the reliance on surrogate models for property-guided generation introduces additional uncertainty, as these models may deviate from experimental reality. The authors acknowledge dataset limitations, such as the exclusion of lanthanides, actinides, and complex frameworks, which further constrains the generalizability of the approach.

In summary, this work presents an innovative and promising direction for improving chemical validity in generative materials design. However, the current implementation raises important questions about physical realism, stability thresholds, and practical utility. Addressing these concerns would significantly strengthen the manuscript and its contribution to the field.

Reviewer #2 (Remarks to the Author):

Please check with the attached file.

Reviewer #3 (Remarks to the Author):

"Co-reviewed with Reviewer #4"

The manuscript entitled "Enhancing Materials Discovery with Valence Constrained Design in Generative Modeling" by Cheng et al. provides an exciting methodology to map materials with specific properties. The novelty of this approach lies in incorporating physical and chemical boundaries, such as oxidation states and chemical stoichiometry, to both speed up exploration and generate meaningful structures. The manuscript is well organised and written, though I recommend that the authors reduce jargon to broaden the audience and enhance impact.

I would like to encourage the incorporation of details in the main paper regarding the pre-trained MLIPs. What is the effect of the environment on the pre-trained potentials? Is the accuracy secured? The last part of the paragraph before the results is rather obscure. Can it be rephrased?

On page 2, the dot in parentheses indicates a scalar product; it should be made clear. Is p_T the prior noise distribution at the final temperature?

There are minimal details how the crystals are generated. Are there dictionaries with predefined distances and angles, for instance? Similarly, how is the periodic behaviour implemented?

Is there any experimental value to validate the CrysVCD results? It is critical to validate predictions that lie outside the learning set, i.e. projections that extrapolate from learnt trends commonly imply large deviations from measurements. How is the loss function of the different learning processes?

The discussion, and especially the critical reflection at the end of the section, is very much welcome.

Reviewer #4 (Remarks to the Author):

Co-reviewed with reviewer #3

Title: Enhancing Materials Discovery with Valence Constrained Design in Generative Modelling

Article Type: Research

The manuscript entitled "Enhancing Materials Discovery with Valence Constrained Design in Generative Modelling" by M. Cheng et al. introduces an innovative methodology that associates large language models (LLMs) with existing diffusion algorithms to accelerate material screening by imposing physically motivated constraints on the chemical space. Boundaries are introduced by specifying known properties, e.g. stoichiometry and oxidation states of the elements present in the material.

The introduction of the physically-motivated LLM in a crystal generator enables the dissociation of an equation of 3 variables, namely the material's formula, its lattice structure, and the position of atoms, into two simpler equations. The first one, the object of the present paper, generates the material's formula, subject to appropriate boundaries, and the second searches for positions and a lattice.

While the plan followed by M. Cheng et al. is clear, some sections of this paper could be improved for better clarity.

Particularly, I believe Eq. (1) could be written in a more reader-friendly fashion by introducing the Markov chain of crystals as $M_T \rightarrow M_{(T-1)} \rightarrow \dots \rightarrow M_1 \rightarrow M_0$ (M_T being the diffused structure, and M_0 the target), and write Eq. (1) as follows,

$$\begin{aligned} & \int p(A_0, X_0, L_0) \\ &= p_T(A_T, X_T, L_T) \prod_{t=1}^T \\ & \int dM_{(T-t+1)}; \\ & p_{\theta} \left(A_{(T-t)}, X_{(T-t)}, L_{(T-t)} \right. \\ & \left. \middle| A_{(T-t+1)}, X_{(T-t+1)}, L_{(T-t+1)} \right) \\ & \left. \right) \end{aligned}$$

Which uses a single product operator, which instinctively captures the 'optimisation' path M_T from to M_0 , rather than 2 product operators. The same treatment could be applied to Eq. (2) and render it easier to read.

The SI provides details on the dataset used to train the networks in this work, as well as specific information on the LLM used for formula calculations. If possible, could the SI include brief details on the MLIP used for energy predictions (e.g., neural network or Gaussian approximation) and a short description of the descriptors used (e.g., symmetry functions, smooth overlap of atomic positions, etc.).

I also wonder whether experimental observations of the predicted GeC crystal have been made, whether it has been synthesised, and whether its properties have been characterised and agree with CrysVCD and DFT? In addition to being an extension of the available model DiffCSP, will CrysVCD be publicly available online in the future?

The work proposed in this paper is original and highly valuable to the field of data-guided material discovery. Associating LLMs with diffusion models is an innovative approach that accelerates material scanning by orders of magnitude compared to non-physically guided models. I would recommend this paper for publication after addressing the comments given in the previous sections of this report.

Reviewer #4 (Remarks on code availability):

Co-reviewed with reviewer #3

Assuming the code is to be released along with the paper, the choice of Python language and specifically the PyTorch library ensures clarity and easy access/understanding of the code for a majority of members of the scientific community.

Version 1:

Decision Letter:

** Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your co-authors. **

Dear Professor Li,

Your manuscript "Enhancing Materials Discovery with Valence Constrained Design in Generative Modeling" has now been seen by 4 referees, whose comments are appended below. You will see that while they find your work of interest, they have raised points that need to be addressed before we can make a decision on publication.

The referees' reports seem to be quite clear. Naturally, we will need you to address *all* of the points raised.

While we ask you to address all of the points raised, the following points need to be substantially worked on:

1) Please address the remaining requests from Reviewer #2, including the points about code/workflow accessibility

Please use the following link to submit your revised manuscript and a point-by-point response to the referees' comments (which should be in a separate document to any cover letter):

Link Redacted

** This url links to your confidential homepage and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this e-mail to co-authors, please delete this link to your homepage first. **

To aid in the review process, we would appreciate it if you could also provide a copy of your manuscript files that indicates your revisions by making use of Track Changes or similar mark-up tools. Please also ensure that all correspondence is marked with your Nature Computational Science reference number in the subject line.

In addition, please make sure to upload a Word Document or LaTeX version of your text, to assist us in the editorial stage.

To improve transparency in authorship, we request that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit <http://www.springernature.com/orcid>.

We hope to receive your revised paper within three weeks. If you cannot send it within this time, please let us know.

We look forward to hearing from you soon.

Best regards,

Kaitlin McCardle, PhD
Senior Editor
Nature Computational Science

Reviewers comments:

Reviewer #1 (Remarks to the Author):

The authors have addressed my comments well and clarified the scope of the study. I recommend the manuscript be accepted.

Reviewer #2 (Remarks to the Author):

See attached

Reviewer #3 (Remarks to the Author):

The authors have satisfactorily addressed all of the concerns raised in the previous round of review, and I appreciate the care taken in revising the manuscript accordingly. The responses are clear, thorough, and demonstrate a genuine engagement with the feedback provided.

In its current form, the paper presents a well-considered and methodologically sound contribution to the field of computational chemistry. In particular, the proposed approach to valence-constrained design within generative modelling frameworks appears both robust and well thought out. The authors have made a convincing case for its utility in enhancing the reliability and chemical validity of generated structures, which is a meaningful step forward in materials discovery workflows.

Overall, I find the revised manuscript to be clear, rigorous, and suitable for publication.

Reviewer #4 (Remarks to the Author):

Revised article report: NATCOMPUTSCI-25-2419A
Reviewer #4 (Co-reviewed with reviewer #3)

Review regarding my initial feedback:

The authors have thoroughly addressed most of the points raised in my initial report by (i) reformulating Eqs. (1) to (3) to a more reader-friendly format, (ii) including details regarding experimental investigation of materials similar to h-GeC reported

in the manuscript, and (iii) introducing the MLIP employed in more detail.

Review regarding the other reviewer's feedback:

I believe the updated discussion the authors included in the main document addresses the reviewer's #1 concern regarding the model's prediction of oxidation states. A performance of 97.6% is very satisfying for an ML model. It is also appreciated that the authors explicitly report limits of their model (i.e. exotic oxidation states and crystal structures not present in the MP-20 dataset).

The details provided regarding the prediction of different materials' oxidation states against existing tools (pymatgen `oxi_state_guess()`) are welcome in the answer letter as well as in the main body of the paper, as it sheds light on the ambiguity highlighted by reviewers #1 and #2 in the initial review, and highlights the superior performance of the proposed algorithm.

The authors address the reviewer's #2 comment regarding CrysVCD access to an "oxidation state" database (i.e., generating knowledge from known, established scientific principles, such as known oxidation states). To the best of my knowledge, this approach has been employed so far for atomistic force fields (Physics-informed neural networks, PINNs), not in the present form of a language-like model. The novelty of this paper lies in the association of this type of model and strategy for materials screening, filtering, and discovery.

Overall assessment:

My first impression of the paper was quite positive, though some corrections were required before it was fit for publication. Upon reviewing other reviewers' feedback, I acknowledged that the authors needed to make further corrections to make this publication fit for Nat. Comp. Sci.

In my opinion, the authors' letter and the multiple updates to the main body of the paper and the supporting information shed light on the various points initially addressed by the reviewer. Particular points addressed cover:

- i) Treatment of oxidation states via a large-language-like model (LLM) that outperforms existing algorithms, with examples for which the existing approaches fail,
- ii) The different limitations of the proposed model are explicitly stated and discussed in the main body of the article (e.g. defective systems, the specific classes of material not present in the database employed by the authors),
- iii) More clarity on the performance of the presented model, which lies on the LLM, efficiently screening most unphysical materials before the denoising step,
- iv) Enhanced result reproducibility, by including a more detailed DFT section and all generated candidate structures online,
- v) Clarifying the jargon employed and the formulae presented in the main body of the article,
- vi) Including in the main body of the text a discussion regarding experimental advances in parallel with computational research and progress, and whether experimental research has been performed or initiated on the materials identified by CrysVCD.

In conclusion, I believe the current version of the submitted article has been significantly strengthened by the authors' multiple updates and clarifications. The innovations brought by the authors result in an algorithm that is faster and more reliable than previous versions it is compared to. The innovation lies in the inclusion of a physically informed large-language-like model for filtering structure. While some existing drawbacks are highlighted (need for fine-tuning for optimal performance, certain classes of materials missing from the dataset, and inability to work with non-pristine materials), I believe these issues can be addressed by downstream users (e.g., through fine-tuning of available datasets) and do not mitigate the algorithmic innovations brought in by the authors.

The current version of the paper appears fit for publication to me.

Version 2:

Decision Letter:

Our ref: NATCOMPUTSCI-25-2419B

7th May 2026

Dear Dr. Li,

Thank you for submitting your revised manuscript "Enhancing Materials Discovery with Valence Constrained Design in Generative Modeling" (NATCOMPUTSCI-25-2419B). It has now been seen by the original referees and their comments are below. The reviewers find that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Computational Science, pending minor revisions to satisfy the referees' final requests and to comply with our editorial and formatting guidelines.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements in the next few weeks. Please do not upload the final materials and make any revisions until you receive this additional information from us.

TRANSPARENT PEER REVIEW

Nature Computational Science offers a transparent peer review option for original research manuscripts. We encourage increased transparency in peer review by publishing the reviewer comments, author rebuttal letters and editorial decision

letters if the authors agree. Such peer review material is made available as a supplementary peer review file. **Please remember to choose, using the manuscript system, whether or not you want to participate in transparent peer review.**

Please note: we allow redactions to authors' rebuttal and reviewer comments in the interest of confidentiality. If you are concerned about the release of confidential data, please let us know specifically what information you would like to have removed. Please note that we cannot incorporate redactions for any other reasons. Reviewer names will be published in the peer review files if the reviewer signed the comments to authors, or if reviewers explicitly agree to release their name. For more information, please refer to our [FAQ page](https://www.nature.com/documents/nr-transparent-peer-review.pdf).

Thank you again for your interest in Nature Computational Science. Please do not hesitate to contact me if you have any questions.

Sincerely,

Kaitlin McCardle, PhD
Senior Editor
Nature Computational Science

ORCID

IMPORTANT: Non-corresponding authors do not have to link their ORCIDs but are encouraged to do so. Please note that it will not be possible to add/modify ORCIDs at proof. Thus, please let your co-authors know that if they wish to have their ORCID added to the paper they must follow the procedure described in the following link prior to acceptance: <https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research>

Author names using non-Roman characters

Nature Portfolio journals can support presentation of author names using non-Roman characters in the HTML version of the article. If you wish to, please include author names in parentheses after the Roman-character spelling; [see example online here](https://www.nature.com/articles/s44222-024-00258-2). Currently supported scripts are: Arabic, Chinese, Cyrillic, Devanagari, Greek, Hebrew, Hangul, Japanese and Persian. You will be asked to verify the rendering is correct at proof stage.

Reviewer #2 (Remarks to the Author):

We thank the authors for their point-by-point responses. The concerns have been adequately addressed, the manuscript has been significantly improved, and I recommend publication.

Version 3:

Decision Letter:

Dear Professor Li,

We are pleased to inform you that your Article "Enhancing Materials Discovery with Valence Constrained Design in Generative Modeling" has now been accepted for publication in Nature Computational Science.

Once your manuscript is typeset, you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required.

Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates.

If your research is supported by a funder that requires immediate open access (e.g. according to [Plan S principles](https://www.springernature.com/gp/open-science/plan-s-compliance) or the [NIH public access policy](https://www.springernature.com/gp/open-science/us-federal-agency-compliance)) then you should select the gold OA route, and we will direct you to the compliant route where possible. Because authors warrant under our subscription licensing terms that they haven't committed to licensing any version of their article under a license inconsistent with the terms of our agreement – including the applicable embargo period – publication under the subscription model isn't suitable for authors whose funders require no embargo.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

Acceptance of your manuscript is conditional on all authors' agreement with our publication policies (see <https://www.nature.com/natcomputsci/for-authors>). In particular your manuscript must not be published elsewhere and there must be no announcement of the work to any media outlet until the publication date (the day on which it is uploaded onto our web site).

Before your manuscript is typeset, we will edit the text to ensure it is intelligible to our wide readership and conforms to house style. We look particularly carefully at the titles of all papers to ensure that they are relatively brief and understandable.

Once your manuscript is typeset, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately.

If you have queries at any point during the production process then please contact the production team at rjsproduction@springernature.com.

You may wish to make your media relations office aware of your accepted publication, in case they consider it appropriate to organize some internal or external publicity. Once your paper has been scheduled you will receive an email confirming the publication details. This is normally 3-4 working days in advance of publication. If you need additional notice of the date and time of publication, please let the production team know when you receive the proof of your article to ensure there is sufficient time to coordinate. Further information on our embargo policies can be found here:

<https://www.nature.com/authors/policies/embargo.html>

An online order form for reprints of your paper is available at <https://www.nature.com/reprints/author-reprints.html>. All co-authors, authors' institutions and authors' funding agencies can order reprints using the form appropriate to their geographical region.

We welcome the submission of potential cover material (including a short caption of around 40 words) related to your manuscript; suggestions should be sent to Nature Computational Science as electronic files (the image should be 300 dpi at 210 x 297 mm in either TIFF or JPEG format). We also welcome suggestions for the Hero Image, which appears at the top of our [home page](http://www.nature.com/natcomputsci); these should be 72 dpi at 1400 x 400 pixels in JPEG format. Please note that such pictures should be selected more for their aesthetic appeal than for their scientific content, and that colour images work better than black and white or grayscale images. Please do not try to design a cover with the Nature Computational Science logo etc., and please do not submit composites of images related to your work. I am sure you will understand that we cannot make any promise as to whether any of your suggestions might be selected for the cover of the journal.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

We look forward to publishing your paper.

Best regards,

Kaitlin McCardle, PhD
Senior Editor
Nature Computational Science

P.S. Click on the following link if you would like to recommend Nature Computational Science to your librarian: <https://www.springernature.com/gp/librarians/recommend-to-your-library>

** Visit the Springer Nature Editorial and Publishing website at <http://editorial-jobs.springernature.com> for more information about our career opportunities. If you have any questions please click [here](mailto:editorial.publishing.jobs@springernature.com). **

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Response letter on revision made to manuscript entitled “Enhancing Materials Discovery with Valence Constrained Design in Generative Modeling”

Reviewer #1 (Remarks to the Author):

This manuscript addresses a critical limitation in generative AI for materials design: the tendency to produce chemically invalid or thermodynamically unstable structures. The proposed framework, CrysVCD, introduces a two-step generative process that first enforces valence balance at the composition level and then applies a diffusion model for structural generation. This approach is conceptually appealing because it integrates fundamental chemical constraints early in the pipeline, potentially reducing the inefficiency of generating and discarding non-physical candidates. The authors also incorporate property-guided generation and report improved stability metrics compared to prior models, which is an important advancement for practical materials discovery. Despite these strengths, several concerns limit the impact and applicability of the method.

We sincerely thank the reviewer for the positive assessment of our manuscript and for the pertinent comments highlighting several important limitations of the current work. The issues raised, such as defect effects, stability metrics, and coverage of broad chemical space, are indeed common and challenging problems in AI-driven materials discovery. In this study, our primary focus is to address one such fundamental challenge: enforcing charge neutrality, a basic physical constraint for stable materials.

We acknowledge the broader limitations noted by the reviewer and have expanded the manuscript to explicitly discuss them. In addition, we have made substantial revisions to strengthen the analysis of stability metrics, clarify the treatment of mixed-valence compositions, as well as the verification of surrogate model predictions against DFT calculations. A detailed point-by-point response is provided below.

First, the strict enforcement of charge neutrality at the composition level may inadvertently exclude realistic materials that achieve stability through point defects such as vacancies or interstitials. These defect-mediated mechanisms are common in functional materials and are essential for practical applications, yet the current approach screens them out.

We thank the reviewer for highlighting the role of point defects in functional materials. We acknowledge that the scope of the present work is the discovery of ideal crystalline materials, and our treatment of charge neutrality is simply aligned with this current scope.

However, this limitation does not arise solely from our charge-neutrality constraint. Explicit representation of point defects at the structure level requires support of fractional atomic

occupancies. Most mainstream crystal structure generative models, including DiffCSP and MatterGen, use individual atoms as the elementary generation units and do not support fractional occupancies. There is also a lack of defect-containing structures in the relevant training datasets, as their scope is similarly restricted to ideal crystals. Our contribution in this work is aimed at a general framework to improve the chemical validity and stability of outputs with a model-agnostic design for the underlying structural generative models. Therefore, its applicability to defect modeling is inherently tied to such degrees of freedom of the underlying structure generators.

Still, compared to generative models without any constraints on chemical composition, we believe that our framework provides a chemically grounded starting point that could serve as a foundation for generating more realistic and stable defect-containing structures, because materials with point defects typically derive from perturbations of ideal valence-balanced compositions rather than complete violations of the valence rules.

In the revised manuscript, we have added discussions and perspectives on this point in the **Discussion** section of the main text.

Second, the treatment of oxidation states appears simplistic. While the model uses a valence-labeled vocabulary, it is unclear whether it adapts oxidation states based on local coordination environments or simply maps common oxidation states from a predefined table. This raises questions about its ability to handle multivalent elements and complex chemistries where oxidation states are context-dependent.

We thank the reviewer for raising this important point regarding the treatment of oxidation states. In our framework, oxidation-state assignment is a foundational component, and we have given it careful and explicit consideration.

Oxidation states in our model are not inferred implicitly by the structural context, but are derived through a rule-based valence interpretation algorithm. We start from a manually curated list of commonly observed oxidation states for each element. For materials where oxidation states are inherently ambiguous, such as alloys and hydrides, we introduce a complementary element list that allows compounds composed exclusively of these elements to be represented in a zero-valence form, which is consistent with chemical intuition. To ensure that these predefined rules are sufficiently expressive rather than overly restrictive, as we have mentioned in the manuscript, we evaluated the algorithm's coverage on our MP-20 dataset, and it successfully explains 97.6% of all compositions (as is shown in the **Discussion** section of the main text).

To map a given chemical composition to valid (element, oxidation state) combinations, we developed a tree-based search algorithm that enumerates charge-neutral solutions from our predefined list. This approach works on the atom level of the stoichiometric unit instead of the element level, which naturally supports multivalent elements and mixed-valence systems. For example, simple oxidation state guess methods (as implemented in pymatgen,

smact and other packages) fail in the stoichiometry of Fe(II) and Fe(III) in Fe_3O_4 and Prussian Blue ($\text{Fe}_4[\text{Fe}(\text{CN})_6]_3$). These packages fail to assign discrete Fe(II), Fe(III), C(II), and N(-III) oxidation states by chemistry. In contrast, our algorithm correctly resolves them. Our algorithm achieves more than 3 fold improvement of the valence assignment failure rate (from 8.5% to 2.4%) on our relevant chemical space in the MP-20 dataset compared with the element-level treatment.

In the revised manuscript, we have added the Prussian Blue example, the improvement of the valence assignment failure rate and more details on our treatment on chemical valence in the **Result (Page 4)** and **Method/”Chemical formula generation”** section.

The stability screening strategy also warrants scrutiny. The use of an energy-above-hull cutoff of 0.1 eV/atom is relatively permissive, but for metallic systems, their stability margins are much tighter. As a result, the reported improvement in E_{hull} distribution, while notable, may not translate into practically synthesizable materials.

We thank the reviewer for this thoughtful comment regarding the stability screening criterion. We agree that different material classes warrant different stability thresholds to be considered practically synthesizable.

In response, we have revised the manuscript to include additional analyses using more stringent and class-aware stability criteria. Specifically, we now:

1. **Report stability rates as a function of the energy-above-hull threshold**, rather than relying solely on the 0.1 eV/atom cutoff. As shown in **Figure R1**, CrysVCD consistently achieves higher stability rates than baseline methods across a range of thresholds from 0.05 to 0.15 eV/atom, including more stringent values relevant for metallic systems.
2. **Separate stability statistics for ionic and alloy systems**. As shown in **Figure R2**, CrysVCD achieves stability rates of 0.60 and 0.76 for ionic systems under 0.05 and 0.10 eV/atom thresholds, respectively, and 0.39 and 0.48 for alloy systems. This demonstrates that the improvement is not limited to ionic compounds and persists for metallic systems with tighter stability margins.
3. **Evaluate dynamical stability through phonon calculations**. As shown in **Figure R3**, the phonon stability rate improves from 0.14 (CDVAE) and 0.33 (MatterGen) to 0.49 for the unconditional CrysVCD model, and further to 0.68 after fine-tuning. This indicates that the generated structures are not only lower in energy but also more likely to be dynamically stable.

Across these additional analyses, CrysVCD consistently produces a larger fraction of low- E_{hull} and phonon-stable structures than the baseline models. These results indicate that the observed improvement is not limited to a single permissive cutoff, but reflects a

systematic shift toward greater thermodynamic and dynamical stability across different material classes.

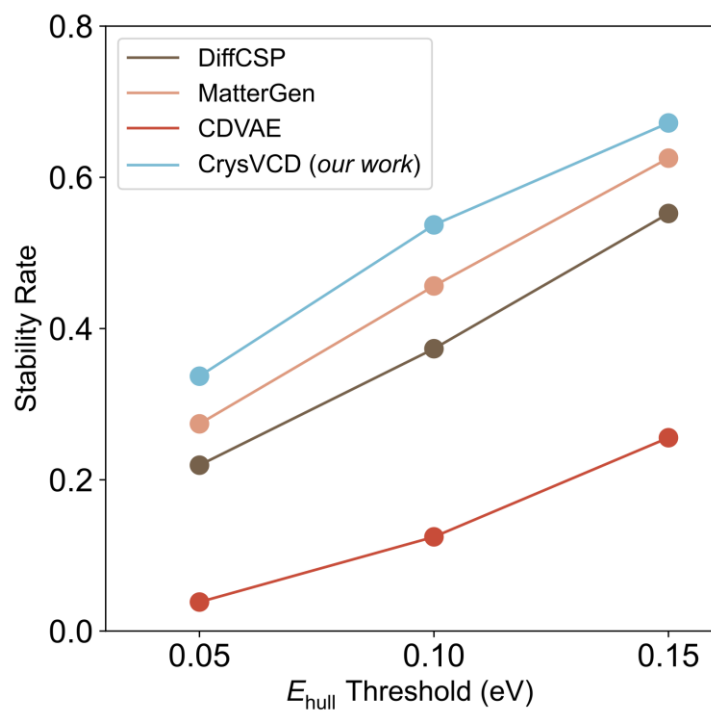


Figure R1. Stability rate comparison as a function of energy-above-hull threshold between CrysVCD and different state-of-the-art generative models, including DiffCSP, MatterGen, and CDVAE.

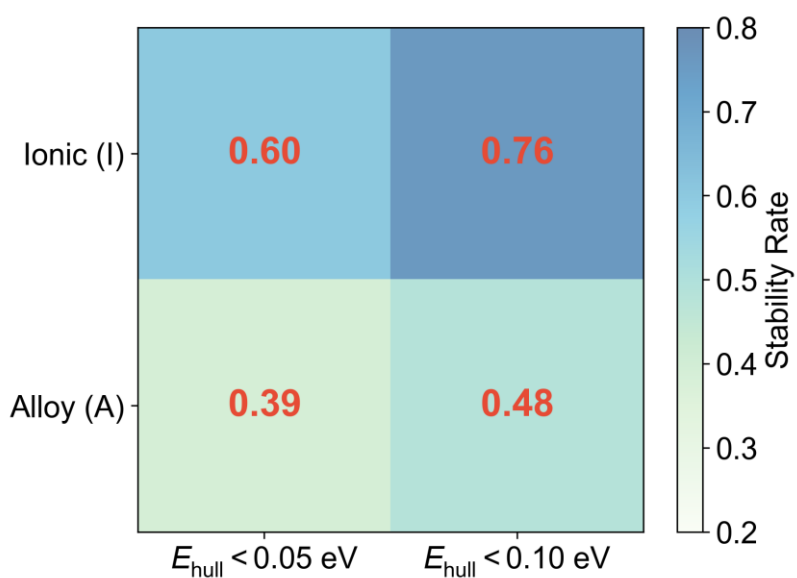


Figure R2. Stability rate of generated structures for ionic compounds (I) and alloys (A) under different energy-above-hull thresholds. CrysVCD achieves stability rates of 0.60 and 0.76 for ionic systems under 0.05 and 0.10 eV/atom thresholds, respectively, and 0.39 and 0.48 for alloy systems.

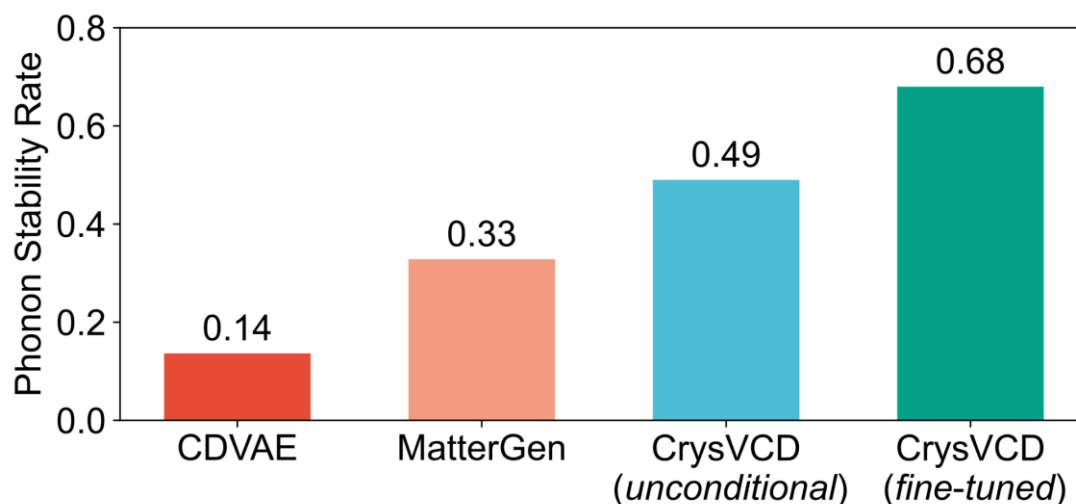


Figure R3. Comparison of phonon stability rates for structures generated by different models. CDVAE and MatterGen achieve phonon stability rates of 0.14 and 0.33, respectively. The unconditional CrysVCD model improves this value to 0.49, and the fine-tuned CrysVCD further increases it to 0.68.

In the revised manuscript, we have included these benchmarking results in **Supplementary Information Section 3.1** and referenced them in the main text (**Page 6**). Specifically, we state: “We further conducted systematic benchmarking of computational speed and generated-structure stability against MatterGen and CDVAE, which are implemented using architectures distinct from the DiffCSP backbone (Supplementary Information Section 3.1). Under identical training conditions on the MP-20 dataset, our model outperforms or matches these methods in both efficiency and quality.”

Furthermore, the reliance on surrogate models for property-guided generation introduces additional uncertainty, as these models may deviate from experimental reality. The authors acknowledge dataset limitations, such as the exclusion of lanthanides, actinides, and complex frameworks, which further constrains the generalizability of the approach.

We thank the reviewer for raising this important point regarding the reliability of surrogate models and the limitation of the chemical space. We fully acknowledge that relying on surrogate property predictors can introduce deviation from experimental reality.

To mitigate this risk, we took care to benchmark the surrogate models against higher-fidelity DFT calculations before their use in the generative loop, and we restricted property-guided generation to domains where the surrogate demonstrated reliable performance. Nonetheless, we recognize that DFT itself can exhibit systematic errors from experimental measurements for certain properties, and that experimental datasets for key properties remain sparse. These limitations are not unique to our work but reflect broader challenges in computational materials science. Within this context, our results also try to demonstrate the potential of conditional generative models in low-data regimes in the case of high thermal conductivity materials. The model can leverage limited and imperfect supervision to identify meaningful trends and propose good candidates. The detailed protocol on surrogate model screening as well as DFT verification are shown in **Methods** section in the revised manuscript.

Additionally, we agree that the exclusion of lanthanides, actinides, and complex structural families currently constrains the scope of applicability. We have expanded the revised manuscript **Discussion** section to address these limitations and outline future directions, including the incorporation of more accurate and generalizable surrogate models, the use of advanced electronic-structure methods, the gap between computational thermodynamic stability and experimental synthesizability, and the integration of experimental datasets as they become available. We believe that continued methodological advances and data accumulation will continue enhancing the impact and generalizability of the proposed framework.

In summary, this work presents an innovative and promising direction for improving chemical validity in generative materials design. However, the current implementation raises important questions about physical realism, stability thresholds, and practical utility. Addressing these concerns would significantly strengthen the manuscript and its contribution to the field.

We thank the reviewer for this constructive and insightful summary and for recognizing the novelty and potential impact of our work. In response, we have revised the manuscript to address these points in a systematic manner, including clarifying the scope of the current implementation, strengthening the discussion of the underlying physical and chemical assumptions, refining the stability analysis with more context-aware criteria, and expanding the discussion of limitations and future directions. We hope that these revisions improve the clarity and rigor of the manuscript and address the reviewer's concerns, making our manuscript more suitable for publication in *Nature Computational Science*.

Reviewer #2 (Remarks to the Author):

This paper presents CrysVCD, a two-stage, modular deep generative framework that incorporates chemical knowledge, such as oxidation state balance, into both composition and structure generation, enabling the design of stable, chemically valid, and functionally superior crystalline materials. The study demonstrates the method's potential in generating low formation energy crystals with enhanced thermodynamic and dynamic stability, and successfully identifies high thermal conductivity material GeC and high dielectric constant material CoSnO₂, with first-principles (DFT) validation further supporting the predicted performance.

However, several critical issues identified during review limit the reliability of the conclusions and the generality of the method: (1) the complexity of ensuring oxidation state balance in the first stage of the generative model appears to be overestimated, as it can be efficiently verified using existing tools (e.g., pymatgen); (2) the efficiency comparison is methodologically problematic, as it considers only the speed of the two internal stages rather than the entire workflow; (3) the baseline DiffCSP setup and fairness are unclear, with no explicit information on the source of chemical formulas or fine-tuning strategy; and (4) the high-performance material generation experiments in Figure 4 lack detailed descriptions of the screening process and sample sizes, limiting reproducibility.

In summary, although CrysVCD demonstrates methodological innovation and shows potential in materials discovery, the current manuscript does not yet meet the stringent criteria of Nature Computational Science for high-level, broad scientific impact and interdisciplinary frontier breakthroughs. Nature Computational Science consistently prioritizes publishing pioneering research that can significantly advance the entire field of computational science. Its review process emphasizes not only methodological novelty but also the universality and scientific insight of the work. If the key issues mentioned above can be adequately addressed, this work would be more suitable for submission to specialized journals such as Physical Review Materials, Journal of Chemical Physics, Physical Review B or Applied Physics Letters.

We sincerely appreciate the reviewer's careful reading and the detailed and candid comments. On the other hand, we believe that several of the major concerns may reflect areas where our original presentation lacked clarity to fully highlight the novelty. We are grateful for the opportunity to improve the manuscript, and we have revised the text to explain these points more clearly and thoroughly.

CrysVCD is not intended as a faster diffusion model to fully replace the state-of-the-art achievements, but as a general, modular plugin (like VCD can be inserted in a DVD player) that injects basic chemical constraints into generative pipelines. Our work highlights the importance of explicitly incorporating domain knowledge into model design, rather than relying solely on data-driven learning. We believe this work has immediate impact and broad relevance for the computational science and materials science community, particularly for researchers developing AI-driven generative models who may be less aware

of the foundational chemical constraints and physical principles. In the revised main text, we have fully addressed the reviewer's concerns and restated the positioning and perspectives of our work in the Discussion section. Additionally, please find the point-to-point response to the reviewer's comments below.

Comments (1) In the first stage of the generative model, the authors employ a Transformer-based elemental language model to ensure the oxidation state balance of chemical formulas. While this design demonstrates a certain degree of novelty, the manuscript appears to overestimate the complexity of the problem. In practice, the oxidation state balance of a chemical formula is inherently a deterministic combinatorial problem that can be efficiently addressed using existing, well-established methods. For instance, the `oxi_state_guesses()` function in the `pymatgen` library enables robust verification of complex formulas, including those with mixed valence, using only a few lines of code (see the example provided at the end of this comment).

We thank the reviewer for pointing out the existing toolkits in the community regarding the oxidation state issue. Oxidation state balance is not a trivial post-hoc check in our framework. First, our Stage-I model is more powerful than oxidation balance verification in existing tools (e.g., `pymatgen.oxi_state_guesses`). A concrete example is Fe_3O_4 . `oxi_state_guesses()` assigns Fe an *average* oxidation state of +2.66, as shown in the code block below, which was run exactly as provided by the reviewer.

```
from pymatgen.core import Composition

def check_oxidation_state(formula):
    comp = Composition(formula)
    ox_guesses = comp.oxi_state_guesses()
    if not ox_guesses:
        print(f"Cannot guess oxidation states for {formula}")
        return
    for ox_state in ox_guesses:
        total_charge = sum(ox_state[el] * comp[el] for el in ox_state)
        status = "Balanced" if total_charge == 0 else "Unbalanced"
        print(f'Oxidation state guess: {ox_state}, Total charge: {total_charge}- {status}')

check_oxidation_state("Fe3O4")
# Output for the code:
# Oxidation state guess: {'Fe': 2.6666666666666665, 'O': -2.0}, Total charge: 0.0- Balanced
check_oxidation_state("Fe7C18N18") # Fe4[Fe(CN)6]3
```

```

# Output for the code:
# Oxidation state guess: {'Fe': 3.0, 'C': 1.8333333333333333, 'N': -
3.0}, Total charge: 0.0- Balanced
# Oxidation state guess: {'Fe': 2.857142857142857, 'C':
1.8888888888888888, 'N': -3.0}, Total charge: 0.0- Balanced
# Oxidation state guess: {'Fe': 2.7142857142857144, 'C':
1.9444444444444444, 'N': -3.0}, Total charge: 0.0- Balanced
# Oxidation state guess: {'Fe': 2.5714285714285716, 'C': 2.0, 'N': -
3.0}, Total charge: 0.0- Balanced
# Oxidation state guess: {'Fe': 2.0, 'C': 2.2222222222222223, 'N': -
3.0}, Total charge: 0.0- Balanced
# Oxidation state guess: {'Fe': 2.2857142857142856, 'C':
2.111111111111111, 'N': -3.0}, Total charge: 0.0- Balanced
# Oxidation state guess: {'Fe': 2.4285714285714284, 'C':
2.0555555555555554, 'N': -3.0}, Total charge: 0.0- Balanced
# ...(876 possible oxidation states in total)

```

The function can indeed check charge neutrality of a chemical formula. In fact, this can be readily done by various packages, including pymatgen, SMOG, and our in-house code. However, the output of this pymatgen function is chemically ambiguous and unusable for generative modeling. In contrast, CrysVCD explicitly represents Fe^{2+} and Fe^{3+} as distinct, quantized tokens, learns their co-occurrence statistics, and enforces charge neutrality *by construction* during generation, not verification. Another extreme example is Prussian blue $\text{Fe}_4[\text{Fe}(\text{CN})_6]_3$. Here, there are Fe(II), Fe(III), C(II), and N(-III) oxidation states by chemistry, but the pymatgen function gives 876 possible oxidation states in total. These examples demonstrate the limited capacity of the pymatgen oxidation state guessing function in the oxidation state assignment.

In the revision of the main text, we have added the example of the Prussian blue in **Page 4**, and emphasized how much the oxidation state resolution can be improved by our treatment of mixed-valence species across the MP-20 dataset (from 8.4% to 2.6% mismatch, more than 3-fold reduction of error rate).

More importantly, because our goal is inverse design of materials, not filtering after unconditional generation, we must *reinforce* chemically valid compositions in the model itself. Therefore, in our model, oxidation states are also tokenized with electronic configuration embeddings, which is coupled to the atom type as an additional degree of freedom. Then mixed-valence systems, alloys vs ionic compounds, and multi-valence coexistence can be learned by the model from the data and checked within the generative process. Post-hoc checking tools (including pymatgen) only serve as verification pipelines and cannot replace this knowledge of in the model. This distinction is central to the

philosophy of CrysVCD and has been described in the **Methods** and **Fig. 2** of the manuscript.

(2) In the first paragraph of page 4, the authors state that “CrysVCD brings efficiency and flexibility to crystal generation through a two-step architecture, with the autoregressive elemental language model requiring only 4–5 inference steps, far faster than the diffusion process (typically 1000 denoising steps).” However, this comparison is methodologically problematic, as the overall workflow efficiency depends on both stages. A more appropriate baseline would compare the first-stage composition generation with a traditional “generate-and-filter” approach (e.g., using established pymatgen scripts). Emphasizing only the speed of the first stage does not reflect the true computational advantage, since the diffusion model in the second stage is indispensable and computationally expensive, and the actual runtime and resource usage are determined by both stages combined.

We thank the reviewer’s suggestion on the speed comparison. Instead of end-to-end wall time, our efficiency claim regards the difference between

1. Generating more structures with a stand-alone diffusion model and filtering out the chemically invalid structures from the generated candidates afterwards;
2. Generating structures directly with the chemically valid compositions, where the compositions come from a fast elemental language model whose inference time is negligible compared to the diffusion model.

We have added a direct comparison of the diffusion process in **Figure R4** (Revised **Figure S13**), but do not attempt to claim that CrysVCD makes the computation cost of each diffusion process cheap. Our claim is more specific: In existing pipelines, chemical validity is learned more implicitly and enforced via generate-and-filter strategies. This results in large numbers of chemically implausible samples and significant wasted diffusion and screening costs. On the contrast, our model’s stage I is a negligible overhead (0.11 second/structure compared to 1.26 seconds/structure) that eliminates a large fraction of chemically invalid compositions before any diffusion or even DFT-level screening is attempted (see CrysVCD compared to DiffCSP, the backbone of CrysVCD for diffusion crystal generation in **Figure R4**).

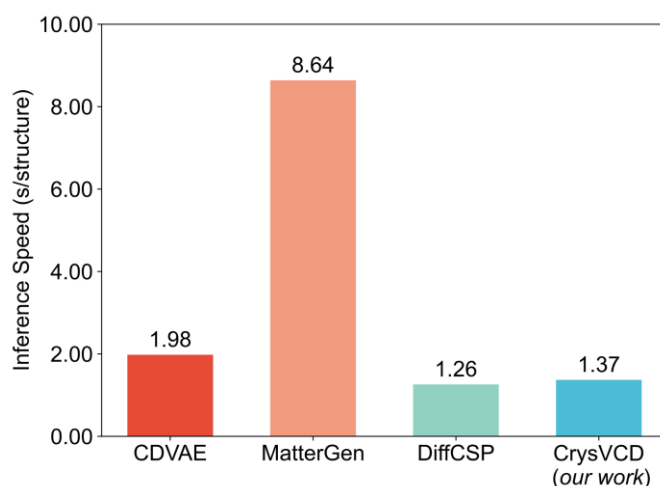


Figure R4: Inference speed comparison between CDVAE, MatterGen, DiffCSP, and CrysVCD. CDVAE, MatterGen, DiffCSP, and CrysVCD require approximately 1.98, 8.64, 1.26, and 1.37 seconds per generated structure, respectively.

We changed the main text on **Page 4** to clarify this point, and added **Figure S13 (Figure R4)** in the Supplementary Information section.

(3) In the comparison presented in Figure 3(a), the source of the chemical formulas used by DiffCSP to generate low-Ehull crystals is not clearly stated. The authors should clarify whether these formulas were taken from an external dataset or generated ab initio by DiffCSP's own composition module. If the latter, it is important to specify whether DiffCSP was fine-tuned on the MP-20-valence dataset or on low-Ehull samples. Without such adjustments, CrysVCD benefits from chemically informed constraints (e.g., valence balance) in its composition stage, whereas the baseline does not receive comparable chemical priors, leading to a methodologically uneven comparison.

We thank the reviewer for the concern of the interpretation between the CrysVCD and DiffCSP. In our context, because both CrysVCD and DiffCSP are trained on the same MP-20 chemical space, the difference we want to study is that: CrysVCD explicitly encodes a fundamental chemical constraint, while DiffCSP, and most existing generative models, do not. This is not an unfair advantage; it is essentially the scientific point of the paper. Our position is that basic chemical laws should not be expected to be fully rediscovered from data, especially in low-data or extrapolative regimes. CrysVCD demonstrates that injecting minimal, well-established chemical structures at the model level leads to more reliable generation. This highlights a room of consideration and improvement of current generative approaches rather than biasing the comparison.

In the revised main text, we modified the caption of **Fig. 3** to clarify that DiffCSP here was not finetuned on the low Ehull dataset. This figure was to show that, by the explicit chemical

constraint, CrysVCD had been increasing the thermodynamic stability of generated candidates, and could further improve via fine tuning on the stability metrics of materials.

(4) The selection of baseline models in Figure 3(a) appears insufficient. Recent studies have demonstrated strong performance in generating low-Ehull stable materials using several advanced approaches, including fine-tuned models such as MatterGen, Con-CDVAE, Cond-CDVAE, InvDesFlowAL, as well as reinforcement-learning-based methods like CrystalFormer-RL. To more convincingly demonstrate the advantages of CrysVCD in generating thermodynamically stable materials, the authors are encouraged to include comparisons with one or two representative state-of-the-art methods and to discuss the observed performance differences and their underlying causes. This would substantially strengthen the credibility of the conclusions.

We fully agree with the reviewer that including additional state-of-the-art baselines (e.g., MatterGen and CDVAE variants) would significantly strengthen the results and impact of the paper.

In the revised manuscript, we incorporated additional representative baseline methods, including MatterGen, CDVAE and, DiffCSP models, trained and evaluated on the same MP-20 dataset for a fair comparison. The updated results are shown in **Figure R1-R5**. From these additional comparisons, several consistent trends are observed:

1. **Improved thermodynamic stability across thresholds.** As shown in **Figure R1**, CrysVCD achieves higher stability rates than baseline methods across a range of energy-above-hull thresholds from 0.05 to 0.15 eV/atom. This indicates that the improvement is not limited to a single permissive cutoff, but persists under stricter stability criteria.

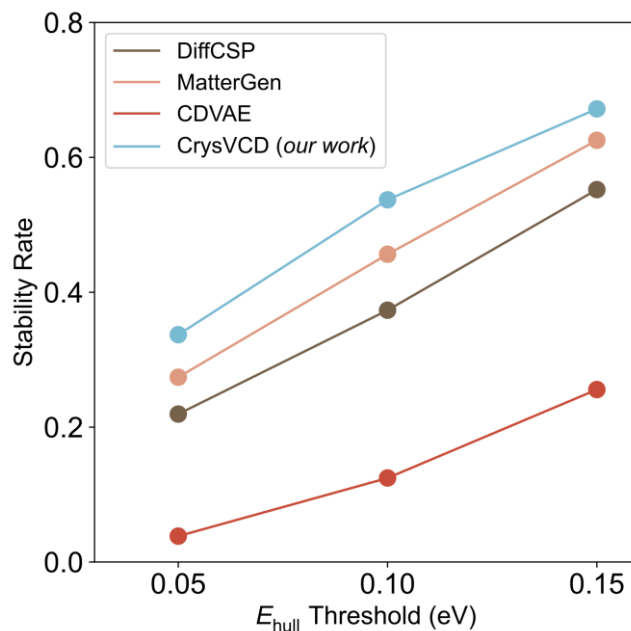


Figure R1. Stability rate comparison as a function of energy-above-hull threshold between CrysVCD and different state-of-the-art generative models, including DiffCSP, MatterGen, and CDVAE.

2. **Consistent gains for both ionic and metallic systems.** Figure R2 separates the results for ionic compounds and alloys. CrysVCD achieves stability rates of 0.60 and 0.76 for ionic systems under 0.05 and 0.10 eV/atom thresholds, and 0.39 and 0.48 for alloys, respectively. This demonstrates that the method improves stability across different material classes, including metallic systems with tighter stability margins.

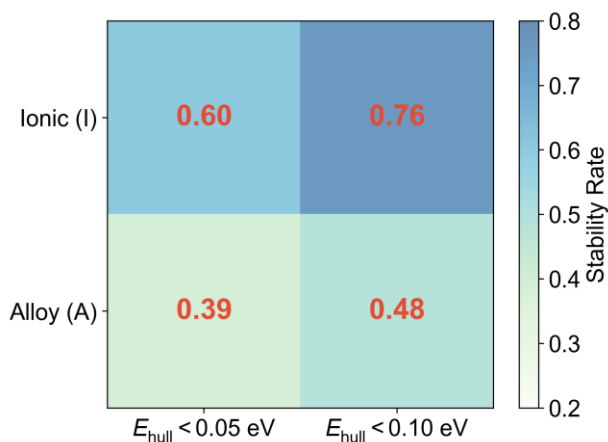


Figure R2. Stability rate of generated structures for ionic compounds (I) and alloys (A) under different energy-above-hull thresholds. CrysVCD achieves stability rates of 0.60 and 0.76 for ionic systems under 0.05 and 0.10 eV/atom thresholds, respectively, and 0.39 and 0.48 for alloy systems.

3. **Improved dynamical stability.** As shown in **Figure R3**, the phonon stability rate increases from 0.14 (CDVAE) and 0.33 (MatterGen) to 0.49 for the unconditional CrysVCD model, and further to 0.68 after fine-tuning. This indicates that the generated structures are not only lower in energy but also more likely to be dynamically stable.

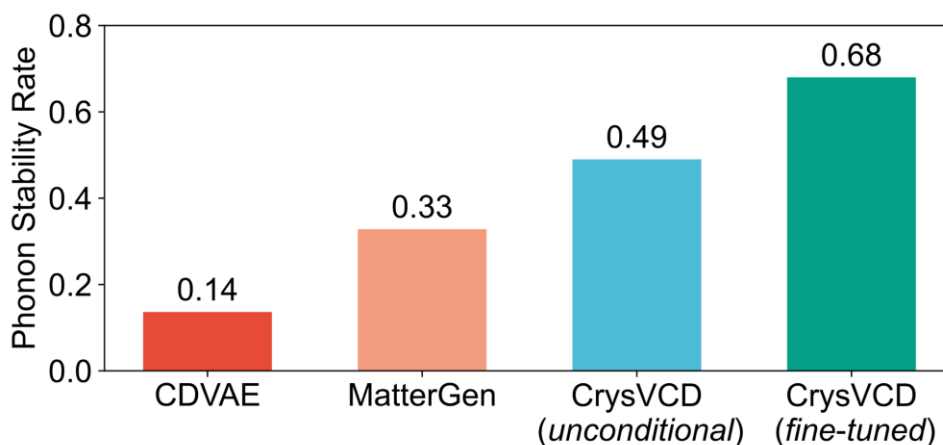


Figure R3. Comparison of phonon stability rates for structures generated by different models. CDVAE and MatterGen achieve phonon stability rates of 0.14 and 0.33, respectively. The unconditional CrysVCD model improves this value to 0.49, and the fine-tuned CrysVCD further increases it to 0.68.

4. **Comparable inference efficiency.** As shown in **Figure R4**, CrysVCD achieves an inference speed of approximately 1.37 seconds per structure, comparable to DiffCSP (1.26 s) and CDVAE (1.98 s), and significantly faster than MatterGen (8.64 s), while achieving improved stability metrics.

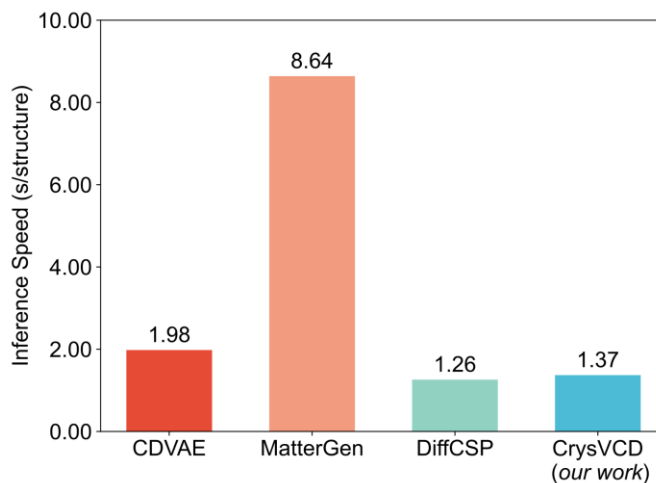


Figure R4: Inference speed comparison between CDVAE, MatterGen, DiffCSP, and CrysVCD. CDVAE, MatterGen, DiffCSP, and CrysVCD require approximately 1.98, 8.64, 1.26, and 1.37 seconds per generated structure, respectively.

5. **Favorable energy-above-hull distribution.** Figure R5 shows that CrysVCD produces a distribution of E_{hull} values that is systematically shifted toward lower energies compared with the baselines, with a higher density of structures near the thermodynamic stability region.

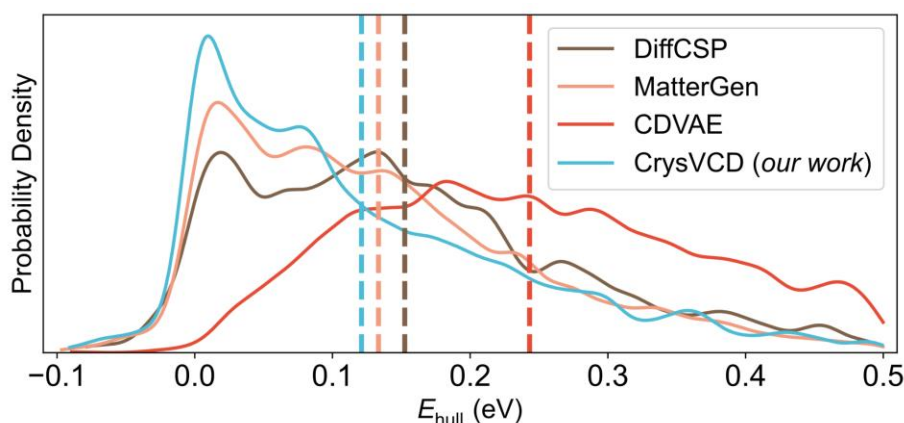


Figure R5. Distribution of energy-above-hull for structures generated by different models trained on the MP-20 dataset, including DiffCSP, MatterGen, CDVAE, and CrysVCD. Dashed lines correspond to the average E_{hull} for each model.

All these results above demonstrate the advantage of incorporating valence constraints in CrysVCD relative to other leading diffusion-based models. In the revised manuscript, we have included these benchmarking results in **Supplementary Information Section 3.1** and referenced them in the main text (**Page 6**). Specifically, we state: “We further conducted systematic benchmarking of computational speed and generated-structure stability against MatterGen and CDVAE, which are implemented using architectures distinct from the DiffCSP backbone (**Supplementary Information Section 3.1**). Under identical training conditions on the MP-20 dataset, our model outperforms or matches these methods in both efficiency and quality.”

(5) The authors demonstrate the potential of CrysVCD in generating high-performance functional materials, successfully identifying representative candidates such as GeC (high thermal conductivity) and CoSnO₂ (high dielectric constant). However, the description of the high-performance material generation experiments in Figure 4 lacks critical details, which limits reproducibility and the strength of the conclusions. Specifically, the authors should

clarify how many candidate materials were generated by CrysVCD for the thermal conductivity and dielectric constant tasks, how many samples were retained at each stage of the screening process (e.g., surrogate model filtering, thresholding, manual inspection), and how many candidates underwent first-principles (DFT) validation. Additionally, the number of materials underlying the thermal conductivity distribution in Figure 4(b) should be specified, along with whether the values were obtained via DFT, the mattersim surrogate model, or other approximate methods. Similar details should be provided for the dielectric constant task. To ensure reproducibility, it is recommended that all generated candidates and their corresponding predicted or computed properties be included in the appendix.

We thank the reviewer for this constructive comment regarding the reproducibility and transparency of the high-performance material generation experiments shown in Figure 4.

In response, we have substantially expanded the Methods “MLIP and DFT calculations” section to explicitly describe the candidate generation and screening pipeline for both the thermal conductivity and dielectric constant tasks. This now includes: the number of materials calculated/labelled by DFT for fine-tuning the CrysVCD, as well as the total number of candidate materials generated by CrysVCD and screened by the surrogate models. The materials GeC and CoSnS₂ highlighted in the main text are further verified by DFT calculations, besides the screening protocol.

To further ensure full reproducibility, we have open-sourced all generated candidate structures together with their corresponding predicted properties in a public data repository. This repository is referenced in the revised manuscript (same as the code repository at GitHub) in the “Data and code availability” statement. We have also uploaded a pre-trained unconditional CrysVCD model, making it easier for users to fine-tune based on their own downstream applications. We now believe these additions significantly strengthen the reproducibility and clarity of the high-performance material generation results, and we thank the reviewer for highlighting this important point.

In summary, we hope that our clarifications have addressed any potential misunderstandings and further highlight the novelty and impact of our approach. We also believe that the revisions improve the clarity and presentation of the manuscript, making it more accessible to a broader readership. We are confident that the revised manuscript is now suitable for publication in *Nature Computational Science*.

Reviewer #3 (Remarks to the Author):

"Co-reviewed with Reviewer #4"

The manuscript entitled “Enhancing Materials Discovery with Valence Constrained Design in Generative Modeling” by Cheng et al. provides an exciting methodology to map materials with specific properties. The novelty of this approach lies in incorporating physical and chemical boundaries, such as oxidation states and chemical stoichiometry, to both speed up exploration and generate meaningful structures. The manuscript is well organised and written, though I recommend that the authors reduce jargon to broaden the audience and enhance impact.

We sincerely thank the reviewer for acknowledging the novelty of our work and for the positive assessment of the manuscript. We fully agree that the clarity of our manuscript, especially in the technical presentation, is essential to ensure accessibility for a broad readership, and we have revised the manuscript accordingly. Below, we provide a point-by-point response to the reviewer’s comments.

I would like to encourage the incorporation of details in the main paper regarding the pre-trained MLIPs. What is the effect of the environment on the pre-trained potentials? Is the accuracy secured? The last part of the paragraph before the results is rather obscure. Can it be rephrased?

We appreciate the reviewer for pointing out that our original description regarding MLIPs lacked clarity and could be unclear to a broader readership. In the revised main text, we have expanded the last part of the **Introduction** section to better explain our motivation for using MLIPs and their significance in our workflow. We have also added a more detailed description in the **Methods/”MLIP and DFT calculations”** section, including the basic concept of MLIPs, the architecture of the selected model, and the rationale for choosing MatterSim for this study.

We also clarify that all structural relaxations, energy evaluations, and phonon calculations are carried out within the Atomic Simulation Environment (ASE) on the potential energy surfaces predicted by MatterSim. ASE is a widely adopted community platform that provides standard implementations of common simulation algorithms. With appropriately chosen convergence criteria, the resulting calculations are stable, reproducible, and reliable for high-throughput screening purposes. We hope these additions improve the clarity and transparency of our computational workflow.

On page 2, the dot in parentheses indicates a scalar product; it should be made clear. Is p_T the prior noise distribution at the final temperature?

We apologize for the confusion caused by the notation. The dot in parentheses $p(\cdot)$ was intended only as a shorthand indicating the arguments of the probability distribution, rather than a scalar product. To avoid any ambiguity, we have removed this notation entirely in the revised manuscript and now write the probability distributions using p only.

Regarding p_T , it denotes the prior noise distribution at the final time step T . This has now been explicitly clarified in the text immediately following **Eq. (1)** in the revised manuscript.

There are minimal details how the crystals are generated. Are there dictionaries with predefined distances and angles, for instance? Similarly, how is the periodic behaviour implemented?

We appreciate the reviewer for encouraging us to add those details. We have explained the representation of periodic crystals in the diffusion model, and the periodic treatment with Fourier features on fractional coordinates. Please refer to the **Supplementary Information Section 2.2** for more details.

Is there any experimental value to validate the CrysVCD results? It is critical to validate predictions that lie outside the learning set, i.e. projections that extrapolate from learnt trends commonly imply large deviations from measurements. How is the loss function of the different learning processes?

We thank the reviewer for these important questions regarding experimental validation, extrapolation, and the learning objectives of CrysVCD.

At present, direct experimental thermal conductivity data are unavailable for several newly proposed candidates, including hexagonal GeC. This reflects a broader limitation of the field rather than a shortcoming of the model, as experimental synthesis and thermal transport measurements remain scarce even for many well-studied compounds. Prior first-principles studies have identified several II–VI beryllium-based semiconductors (BeO, BeS, BeSe, BeTe) as well as cubic GeC as high-thermal-conductivity materials, with cubic GeC predicted to reach $\sim 270 \text{ W m}^{-1} \text{ K}^{-1}$ [Phys. Rev. B 88, 144306 (2013)]. In our work, CrysVCD identifies a hexagonal GeC phase with similarly high thermal conductivity ($183 \text{ W m}^{-1} \text{ K}^{-1}$), suggesting that GeC is a robust and promising thermal conductor across competing polymorphs. Notably, while amorphous Ge-C alloys have been synthesized, single-crystal GeC of any phase has not yet been realized experimentally. We view this as a materials-discovery opportunity, where CrysVCD proposes physically consistent targets that can motivate future experimental exploration.

Importantly, the reported functional predictions are not based solely on machine-learning inference. The highlighted material candidates, including GeC, are explicitly validated using independent first-principles calculations, including full structural relaxation, phonon dispersion, and anharmonic phonon thermal conductivity computed via Density Functional Theory (DFT) and phono3py (**Fig. 4c**). DFT has a well-established record of reliability for lattice dynamics and phonon-limited thermal transport, particularly for covalent

semiconductors, providing strong computational validation of these predictions. While extrapolation beyond the training distribution is inherently challenging, it is not a major concern here because CrysVCD enforces hard chemical constraints (charge neutrality and oxidation states), incorporates stability guidance (Ehull and phonon stability), and requires first-principles verification before any functional claim is made. Thus, extrapolation is physics-anchored rather than unconstrained.

Regarding the learning objectives, CrysVCD uses distinct loss functions at different stages:

- (i) the elemental language model is trained with a standard autoregressive cross-entropy loss over valence-labeled tokens and atom counts;
- (ii) the geometric diffusion model is trained using a denoising score-matching loss typical of diffusion models;
- (iii) Property conditioning is implemented via classifier-free guidance, without introducing an explicit regression loss on the generative model itself.

In the revised manuscript, we have made the following modification:

- We have added relevant discussion on the GeC material, including its connection to literature study, and lack of experimental observation on the compound, on **Page 7** in the main text.
- We have added (more quantitative) description on the training loss functions for different stages in the CrysVCD training in the **Methods** section “Property-guided conditional crystal generation”.

The discussion, and especially the critical reflection at the end of the section, is very much welcome.

We fully agree that some additional discussion and reflection could be helpful. In the revised manuscript we have expanded the manuscript **Discussion** section to address some limitations and outline future directions, including:

- The limited chemical space, excluding lanthanides, actinides, structures with defects and complex structures like metal-organic frameworks, etc;
- Incorporation of more accurate and generalizable surrogate models, and the use of advanced electronic-structure methods;
- The gap between computational thermodynamic stability and experimental synthesizability, and the integration of experimental datasets as they become available.

Overall, we believe that continued methodological advances and data accumulation will continue enhancing the impact and generalizability of the proposed framework.

In summary, we thank the reviewer for the positive assessment of our manuscript and for recognizing the novelty, impact, and broader message of our approach. We have revised the manuscript accordingly to address the comments, particularly to add more details on the methodology we used in this paper. We hope that these revisions clarify the overall workflow of the manuscript, improve its accessibility to a broader readership, and strengthen its suitability for publication in *Nature Computational Science*.

Reviewer #4 (Remarks to the Author):

Co-reviewed with reviewer #3

Title: Enhancing Materials Discovery with Valence Constrained Design in Generative Modelling

Article Type: Research

The manuscript entitled “Enhancing Materials Discovery with Valence Constrained Design in Generative Modelling” by M. Cheng et al. introduces an innovative methodology that associates large language models (LLMs) with existing diffusion algorithms to accelerate material screening by imposing physically motivated constraints on the chemical space. Boundaries are introduced by specifying known properties, e.g. stoichiometry and oxidation states of the elements present in the material.

The introduction of the physically-motivated LLM in a crystal generator enables the dissociation of an equation of 3 variables, namely the material’s formula, its lattice structure, and the position of atoms, into two simpler equations. The first one, the object of the present paper, generates the material’s formula, subject to appropriate boundaries, and the second searches for positions and a lattice.

We sincerely thank the reviewer for the positive assessment of the manuscript. We are also grateful to the reviewer for the detailed suggestions that help us improve the clarity and presentation of our paper. Please see the point-to-point response below to the reviewer’s comments in more detail.

While the plan followed by M. Cheng et al. is clear, some sections of this paper could be improved for better clarity. Particularly, I believe Eq. (1) could be written in a more reader-friendly fashion by introducing the Markov chain of crystals as $M_T \rightarrow M_{T-1} \rightarrow \dots \rightarrow M_1 \rightarrow M_0$ (M_T being the diffused structure, and M_0 the target), and write Eq. (1) as follows,

$$\begin{aligned} & \prod_{t=1}^T p(A_t, X_t, L_t) \\ &= p_T(A_T, X_T, L_T) \prod_{t=1}^{T-1} p_t(A_t, X_t, L_t | A_{t+1}, X_{t+1}, L_{t+1}) \end{aligned}$$

$$\int dM_{T-t+1};$$

$$p_{\theta}(\mathbf{A}_{T-t}, \mathbf{X}_{T-t}, \mathbf{L}_{T-t} | \mathbf{A}_{T-t+1}, \mathbf{X}_{T-t+1}, \mathbf{L}_{T-t+1})$$

Which uses a single product operator, which instinctively captures the ‘optimisation’ path M_T from to M_0 , rather than 2 product operators. The same treatment could be applied to Eq. (2) and render it easier to read.

We totally agree with the reviewer that the formula can be improved for clarity. In the revised manuscript, we have updated **Eq. 1** and **Eq. 2/3** according to the reviewer’s suggestion. The updated formulas only contain one product operator and resemble the optimization process from M_T to M_0 .

The SI provides details on the dataset used to train the networks in this work, as well as specific information on the LLM used for formula calculations. If possible, could the SI include brief details on the MLIP used for energy predictions (e.g., neural network or Gaussian approximation) and a short description of the descriptors used (e.g., symmetry functions, smooth overlap of atomic positions, etc.).

We appreciate the reviewer encouraging us to add more architectural details of the MLIP. In the **Methods** section, we introduce the graph neural network choice of the MatterSim MLIP model and its presentation of the crystal geometry. Please refer to the main text for more details.

I also wonder whether experimental observations of the predicted GeC crystal have been made, whether it has been synthesised, and whether its properties have been characterised and agree with CrysVCD and DFT? In addition to being an extension of the available model DiffCSP, will CrysVCD be publicly available online in the future?

We thank the reviewer for this important question regarding the present status of research on the GeC crystal. At present, direct experimental thermal conductivity data are unavailable for hexagonal GeC discovered by CrysVCD. This reflects a broader limitation of the field rather than a shortcoming of the model, as experimental synthesis and thermal transport measurements remain scarce even for many well-studied compounds. On the computational side, however, prior first-principles studies have identified several II–VI

beryllium-based semiconductors (BeO, BeS, BeSe, BeTe) as well as cubic GeC as high-thermal-conductivity materials, with cubic GeC predicted to reach $\sim 270 \text{ W m}^{-1} \text{ K}^{-1}$ [Phys. Rev. B 88, 144306 (2013)]. In our work, CrysVCD identifies a hexagonal GeC phase with similarly high thermal conductivity ($183 \text{ W m}^{-1} \text{ K}^{-1}$), suggesting that GeC is a robust and promising thermal conductor across competing polymorphs. Notably, while amorphous Ge-C alloys have been synthesized, single-crystal GeC of any phase has not yet been realized experimentally. We view this as a materials-discovery opportunity, where CrysVCD proposes physically consistent targets that can motivate future experimental exploration. In the revised manuscript, we have clarified the current status of GeC-related research on **Page 7** of the main text accordingly.

Regarding the availability of CrysVCD, the current implementation, together with the required training and generated data, has already been publicly open-sourced on GitHub, as stated in the “Code and Data availability” section. During the revision process, we have also uploaded the pre-trained weights of the unconditional CrysVCD model, making it more convenient for people to fine-tune according to their own downstream tasks. Moreover, although CrysVCD builds upon the DiffCSP framework, it is designed as a modular and model-agnostic extension applicable to a broad class of generative models for materials. We plan to actively support plug-in integration of additional generative models. A current challenge is the heterogeneity of backbones and data representations across existing crystal generative models, in contrast to more standardized communities (e.g., pure large language models). Nevertheless, we expect CrysVCD to evolve into a unifying and extensible platform for benchmarking and deploying diverse crystal generative approaches, and we are committed to maintaining and expanding the public repository.

The work proposed in this paper is original and highly valuable to the field of data-guided material discovery. Associating LLMs with diffusion models is an innovative approach that accelerates material scanning by orders of magnitude compared to non-physically guided models. I would recommend this paper for publication after addressing the comments given in the previous sections of this report.

We are grateful to the reviewer for the positive assessment of our manuscript and for recognizing the novelty and value of our approach to the field. We have revised the manuscript accordingly to address the reviewer’s comments, particularly to add more details on the methodology we used in this paper. We hope that these revisions can clarify the overall workflow of the manuscript and make our manuscript suitable for publication in *Nature Computational Science* for a broad readership.

Reviewer #4 (Remarks on code availability):

Co-reviewed with reviewer #3

Assuming the code is to be released along with the paper, the choice of Python language and specifically the PyTorch library ensures clarity and easy access/understanding of the code for a majority of members of the scientific community.

We thank the reviewer for this positive comment. We agree that implementing CrysVCD in Python using PyTorch facilitates clarity and accessibility, and we have released the code alongside the manuscript to ensure broad usability within the scientific community.

Second-round response letter on revision made to manuscript entitled “Enhancing Materials Discovery with Valence Constrained Design in Generative Modeling”

Reviewer #1 (Remarks to the Author):

The authors have addressed my comments well and clarified the scope of the study. I recommend the manuscript be accepted.

We sincerely thank the reviewer for recommending publishing our work on *Nature Computational Science*.

Reviewer #2 (Remarks to the Author):

Overall, the revised manuscript has been significantly improved and addresses several of my previous concerns, particularly through more comprehensive benchmarking against state-of-the-art methods and clearer validation of the advantages of CrysVCD. The conceptual innovation of incorporating valence states as intrinsic constraints is both meaningful and well-motivated. However, several aspects still require clarification or refinement to further strengthen the work, including more precise descriptions of methodological details (e.g., phonon stability criteria), improved transparency and reproducibility (e.g., workflow release and reporting of sample sizes), and more cautious wording in certain claims. Addressing these points would further enhance the rigor, clarity, and practical impact of the manuscript.

We sincerely thank the reviewer for the positive assessment of our revised manuscript and for the previous comments, which helped us greatly improve the quality of the manuscript. In the second round of the revised manuscript, we have addressed the comments of the reviewer, including a more precise description and improved reproducibility. Please find the point-to-point response below. We are confident that our revised manuscript is now suitable for publication in *Nature Computational Science*.

1. The core innovation of CrysVCD lies in transforming valence states from a “post hoc validation” into an “intrinsic constraint.” Traditional tools such as pymatgen can rapidly verify charge balance, but their outputs are often ambiguous and multi-valued (e.g., the average valence of +2.666 in Fe₃O₄, or 876 possible solutions for Prussian blue). These tools can only tell you “possibly correct” after generation is complete, but cannot guide the model on how to generate structures. In contrast, CrysVCD converts valence states from “a result to be checked” into “a prerequisite for generation,” and this paradigm shift itself constitutes a methodological innovation.

We thank the reviewer for acknowledging the methodological innovation of our CrysVCD model regarding the charge balance constraints. Indeed, we are excited to share these advances with the community through *Nature Computational Science*, and we hope they provide useful and meaningful insights for future research.

2. Although the valence-constrained design in CrysVCD is meaningful for ensuring chemical validity, it would be beneficial for the authors to further clarify its scope and limitations. Specifically, charge neutrality does not necessarily correspond to a unique or physically correct valence assignment, as multiple valence configurations may satisfy the same stoichiometry. For materials outside the training distribution, the validity of a specific valence configuration can only be indirectly supported through downstream evaluations (e.g., low energy, phonon stability, or favorable properties). Therefore, the strength of CrysVCD lies not in identifying a uniquely correct valence state, but in providing chemically self-consistent constraints that guide the generation toward stable structures. Clarifying this point would further strengthen the conceptual positioning of the method.

We thank the reviewer for this thoughtful suggestion. In the revised manuscript, we have further clarified the scope of our method in the Discussion section, Page 9, which reads “... *For those successful valence decompositions, we emphasize that they provide a chemically consistent interpretation of composition, rather than a unique assignment of oxidation states. Multiple valence configurations may satisfy the same stoichiometry and represent distinct redox tendencies. Therefore, in our framework, valence serves primarily as a constraint to guide the generation toward chemically plausible compositions, while the definitive valence assignment must ultimately be assessed through downstream evaluations, such as electronic structure analysis and crystal structure validation.*”

3. The authors emphasize the speed advantage of Stage I by stating that it is “orders of magnitude faster than the diffusion process.” This comparison may be misleading, as it could give the impression that the overall generation pipeline is significantly accelerated. In practice, the diffusion process in Stage II remains the computational bottleneck (approximately 1.26 s per structure), and the efficiency of Stage I does not change this fact. It is therefore recommended to adopt a more precise formulation, for example: “Stage I imposes chemical constraints at negligible cost, thereby avoiding expensive diffusion calculations on invalid compositions and improving the overall discovery efficiency.” This wording more accurately reflects the role of Stage I while avoiding potential misinterpretation.

We thank the reviewer for pointing out this potential source of confusion. To further clarify the relation between Stage I and Stage II, as well as the advantage of CrysVCD, in the revised manuscript we have followed the reviewer’s suggestion and revised the related discussion in P4 to “*Compared with Stage II, the autoregressive elemental language model in Stage I requires only a few inference steps (typically 4 or 5), orders of magnitude faster than the diffusion process (typically 1,000 denoising steps), thus imposing chemical constraints at*

negligible cost (more details in Supplementary Information 3). These constraints avoid expensive diffusion calculations during the CSP step on chemically invalid compositions to be dropped by post-generation filters. Therefore, our method improves the overall discovery efficiency on valid materials.”

4. The authors have supplemented the revised manuscript with systematic comparisons against advanced methods such as MatterGen and CDVAE, and comprehensively demonstrated the consistent advantages of CrysVCD across multiple dimensions, including different energy-above-hull thresholds (0.05-0.15 eV/atom), material categories (ionic compounds vs. alloys), dynamical stability, and inference efficiency. These additions effectively address my concerns regarding baseline model selection and significantly strengthen the conclusions. However, several key details still require clarification to improve interpretability and reproducibility. Specifically, the authors should report: (i) the number of candidate structures generated by each model in the low Ehull regime (ii) the criteria and threshold used to define phonon stability (and whether they are consistent with those used in the thermal conductivity screening) and (iii) the total number of materials evaluated for phonon calculations and thermal conductivity. Providing these details would ensure a fair comparison and clarify the statistical significance of the screening process.

We thank the reviewer for this important comment regarding the sampling size and screening criteria. To clarify these choices, in the revised manuscript, we have directly highlighted additional sentences for:

- The number (4000) of candidate structures generated by each model for evaluating Ehull (Figure 3(b) caption);*
- The threshold (0.01 THz imaginary frequency) used to define phonon stability (Methods section, Page 12);*
- The total number of materials evaluated for phonon stability screening (4000, Figure 3(c) caption) and thermal conductivity (1800, Figure 4(e) caption).*

5. The authors are encouraged to open-source the MatterSim + phono3py workflow for thermal conductivity calculations to improve reproducibility. While the current pipeline—using MatterSim to provide second- and third-order force constants and phono3py for thermal conductivity evaluation—is well-motivated and practically useful, the released repository only includes the generative model and lacks implementation details for this key downstream task. Given that force constants predicted by MatterSim (as an MLIP) may have limited accuracy, the resulting thermal conductivity is more suitable for relative ranking rather than absolute prediction. To facilitate reproducibility and broader applicability, the authors are encouraged to release the full workflow, including scripts for extracting force constants from MatterSim outputs, interfaces to phono3py, example cases (e.g., GeC), and recommended settings (e.g., supercell size and k-point sampling). This would significantly enhance the practical impact and usability of the work.

We thank the reviewer for encouraging us to make the entire workflow accessible to the whole community. We furnished our README file that links to our updated scripts in our GitHub repository, [mlip_screen_ehull.py](#) and [mlip_screen_kappa.py](#).

6. The statement regarding the imaginary frequency threshold on page 12 of the revised manuscript (“If an imaginary frequency larger than 0.01 THz appears in the computed phonon band structure, the structure is labeled unstable.”) is inaccurate. It should be revised to: “A structure is considered dynamically unstable if significant imaginary phonon modes (frequencies below -0.01 THz) are present, excluding numerical noise near zero.”

We thank the reviewer for this advice. In the revised manuscript, we have revised the statement regarding imaginary frequency as “*A structure is considered dynamically unstable if significant imaginary phonon modes (imaginary frequencies over $\backslash SI{0.01}{THz}$) are present, excluding numerical noise near zero.*”

Reviewer #3 (Remarks to the Author):

The authors have satisfactorily addressed all of the concerns raised in the previous round of review, and I appreciate the care taken in revising the manuscript accordingly. The responses are clear, thorough, and demonstrate a genuine engagement with the feedback provided.

In its current form, the paper presents a well-considered and methodologically sound contribution to the field of computational chemistry. In particular, the proposed approach to valence-constrained design within generative modelling frameworks appears both robust and well thought out. The authors have made a convincing case for its utility in enhancing the reliability and chemical validity of generated structures, which is a meaningful step forward in materials discovery workflows.

Overall, I find the revised manuscript to be clear, rigorous, and suitable for publication.

We are deeply grateful to the reviewer for acknowledging the impact of our work and recommending publishing our paper on *Nature Computational Science*.

Reviewer #4 (Remarks to the Author):

Review regarding my initial feedback:

The authors have thoroughly addressed most of the points raised in my initial report by (i) reformulating Eqs. (1) to (3) to a more reader-friendly format, (ii) including details regarding experimental investigation of materials similar to h-GeC reported in the manuscript, and (iii) introducing the MLIP employed in more detail.

Review regarding the other reviewer's feedback:

I believe the updated discussion the authors included in the main document addresses the reviewer's #1 concern regarding the model's prediction of oxidation states. A performance of 97.6% is very satisfying for an ML model. It is also appreciated that the authors explicitly report limits of their model (i.e. exotic oxidation states and crystal structures not present in the MP-20 dataset).

The details provided regarding the prediction of different materials' oxidation states against existing tools (`pymatgen oxi_state_guess()`) are welcome in the answer letter as well as in the main body of the paper, as it sheds light on the ambiguity highlighted by reviewers #1 and #2 in the initial review, and highlights the superior performance of the proposed algorithm.

The authors address the reviewer's #2 comment regarding CrysVCD access to an "oxidation state" database (i.e., generating knowledge from known, established scientific principles, such as known oxidation states). To the best of my knowledge, this approach has been employed so far for atomistic force fields (Physics-informed neural networks, PINNs), not in the present form of a language-like model. The novelty of this paper lies in the association of this type of model and strategy for materials screening, filtering, and discovery.

Overall assessment:

My first impression of the paper was quite positive, though some corrections were required before it was fit for publication. Upon reviewing other reviewers' feedback, I acknowledged that the authors needed to make further corrections to make this publication fit for Nat. Comp. Sci.

In my opinion, the authors' letter and the multiple updates to the main body of the paper and the supporting information shed light on the various points initially addressed by the reviewer. Particular points addressed cover:

- i) Treatment of oxidation states via a large-language-like model (LLM) that outperforms existing algorithms, with examples for which the existing approaches fail,
- ii) The different limitations of the proposed model are explicitly stated and discussed in the main body of the article (e.g. defective systems, the specific classes of material not present in the database employed by the authors),
- iii) More clarity on the performance of the presented model, which lies on the LLM, efficiently screening most unphysical materials before the denoising step,
- iv) Enhanced result reproducibility, by including a more detailed DFT section and all generated candidate structures online,
- v) Clarifying the jargon employed and the formulae presented in the main body of the article,
- vi) Including in the main body of the text a discussion regarding experimental advances in parallel with computational research and progress, and whether experimental research has been performed or initiated on the materials identified by CrysVCD.

In conclusion, I believe the current version of the submitted article has been significantly strengthened by the authors' multiple updates and clarifications. The innovations brought by

the authors result in an algorithm that is faster and more reliable than previous versions it is compared to. The innovation lies in the inclusion of a physically informed large-language-like model for filtering structure. While some existing drawbacks are highlighted (need for fine-tuning for optimal performance, certain classes of materials missing from the dataset, and inability to work with non-pristine materials), I believe these issues can be addressed by downstream users (e.g., through fine-tuning of available datasets) and do not mitigate the algorithmic innovations brought in by the authors.

The current version of the paper appears fit for publication to me.

We appreciate the reviewer's detailed and thoughtful comments on our previous round of revision. We also sincerely thank the reviewer for the positive assessment on CrysVCD and the recommendation for publishing our paper on *Nature Computational Science*.

Third-round response letter on revision made to manuscript entitled “Enhancing Materials Discovery with Valence Constrained Design in Generative Modeling”

Reviewer #2 (Remarks to the Author):

We thank the authors for their point-by-point responses. The concerns have been adequately addressed, the manuscript has been significantly improved, and I recommend publication.

We sincerely thank the reviewer for recommending publishing our work on *Nature Computational Science*.