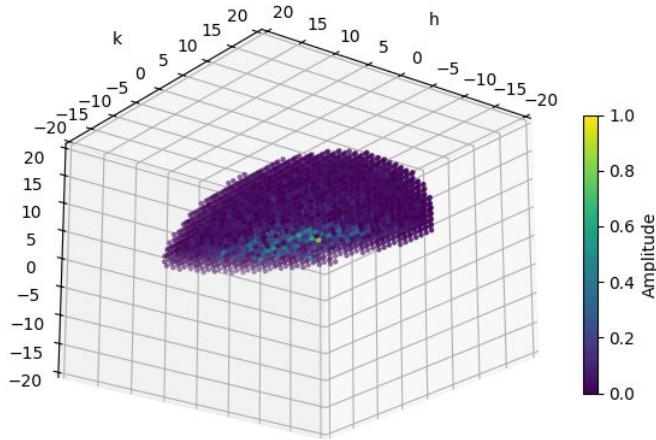


First status report (01.10.2024-22.10.2025) for study course:
[SDSKD041] Seminar in research field III (3 ECTS)
22.10.2025 (Wednesday), 16:30
Raina Boulevard 19, 20 Auditorium



Solving the phase problem in crystallography with deep learning

Ronalds Zakovskis, M.Sc. in Computer Science, rz18026

2nd year (out of 4) PhD Student – University of Latvia FST DCS
Research Assistant – IMCS, University of Latvia, RLSMST



UNIVERSITY OF
LATVIA

FACULTY OF
SCIENCE AND
TECHNOLOGY



**Institute of Mathematics and
Computer Science University of Latvia**

General details

Thesis name in English – "Solving the phase problem in crystallography with deep learning".

Thesis name in Latvian – "Fāzes problēmas kristalogrāfijā risināšana ar dziļo mašīnmācīšanos".

Current supervisor – Prof. Guntis Bārzdiņš, PhD in Computer Science.

Prospective supervisor – Assoc. Prof. Agris Šostaks, PhD in Computer Science.

About me

Education

Riga State Gymnasium No.1:



- [09/2012-06/2018]
Secondary Education Diploma.

University of Latvia:



- [09/2018-01/2021] Associate's Degree with honors;
- [09/2018-06/2023] Bachelor's degree;
- [09/2023-06/2024] Master's degree;
- **[09/2024-current] PhD studies.**

Work Experience

[Research] Institute of Mathematics and Computer Science, University of Latvia:

- [05/2020-09/2021]
Senior Software Engineer / Research Assistant.



[Industry] Visma Resolve:

- [01/2022-08/2023]
AI Developer.



[Research] Institute of Mathematics and Computer Science, University of Latvia:

- **[01/2024-currently]**
Senior Software Engineer / Research Assistant.



Courses completed

- First semester:
 - [SDSKD039] Seminar in research field I (3 ECTS) – passed;
 - [DatZD012] Methods of Research in Computer Science (6 ECTS) – **waiting for grade**;
 - [SDSKD036] Doctoral Thesis I (21 ECTS) – **waiting for “passed”**.
- Second semester:
 - [SDSKD040] Seminar in research field II (3 ECTS) – passed;
 - [ValoN012] Promotional examination in a foreign language (6 ECTS) – 9;
 - [DatZD025] Deep Learning (6 ECTS) – 10;
 - [SDSKD035] Doctoral Thesis II [DatzM] (24 ECTS) – **waiting for “passed”**;
 - [DatZD008] Preparing research publications (3 ECTS) – **waiting instructions** 🙄.
- Third semester:
 - [SDSKD041] Seminar in research field III (3 ECTS) – currently completing;
 - [SDSKD034] Doctoral Thesis III (24 ECTS) – currently completing.

Project foundation

I mainly work with Assoc. Prof. Agris Šostaks and a little bit PhD Artūrs Sprōģis.

Collaborating with chemists from University of Latvia Faculty of Medicine and Life Sciences.

Started off as a UL Fund project "Solving Phase Problem in Crystallography Using Deep Learning" from 01/2024 to 12/2024.

Continued as LZP FLPP project with the same name from 01/2025 to 12/2027.

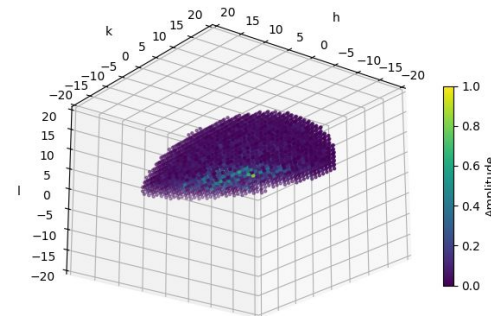
Fun facts

A significant number of Nobel Prizes in the natural sciences have been awarded for research where crystallography played a central or enabling role.

If we could solve the problem – faster drug and materials discovery, etc.

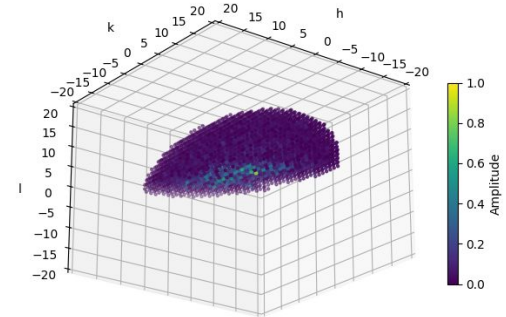
General details

The goal is to determine the structure of an unknown crystal. To do this, the crystal is irradiated with X-rays, and a detector records the resulting diffraction intensities (see image to the right). These measured amplitudes are imperfect, and they don't give us the phases. If we had phases, we could easily reconstruct the electron density, and thus the crystal structure. Since we don't, we must infer the missing phases to reveal the arrangement of atoms, in our case by using amplitudes.

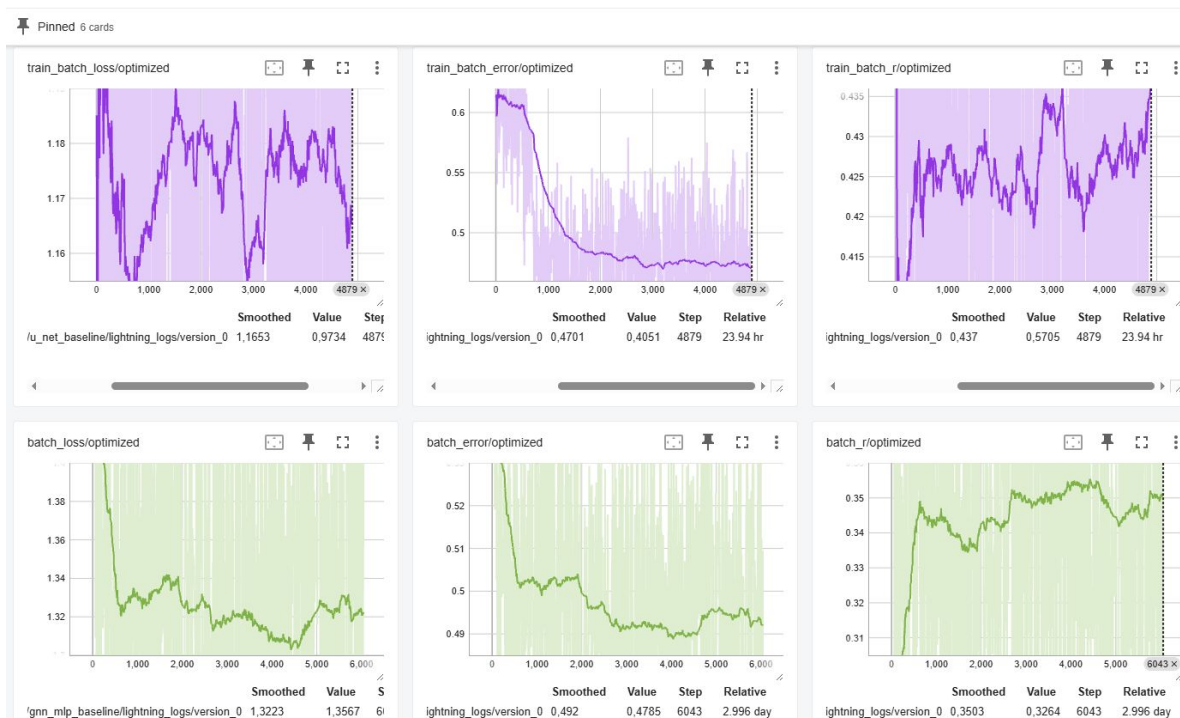


General details 2

- The problem can be looked in different difficulties (symmetries). We look at two variants: classification variant with P-1 data (each point needs a phase of 0 or π , with 8 parallel correct answers); regression variant with P1 data (angle anywhere between 0..360 degrees, infinite correct answers), but it is the general case.



Performance



Research limitations

- Lack of resources:
 - It takes weeks or months to fully evaluate the performance of a model with our UL IMCS GPU (at the moment our project has one UL IMCS server with a “GeForce RTX 2080 Ti” GPU).
- Lack of researchers:
 - In serious AI projects, for example, projects done by Google or other companies, there can be tens or even hundreds of employees on a single project.

For example – “The AlphaFold model was trained on over 10 million samples with 128 TPUs, took over 11 days (7 days initial training and 4 days finetuning) to converge.”

Results

- Decent results on the classification variant (currently continued by the chemist team);
- Pipeline ready for the regression variant, trying to make the model achieve good performance;
- Won “EuroHPC Development Access Call” computing call:
 - Received a node with 4500 hours (half a year) with 4 “AMD MI250x” GPUs, which results in around 50 times faster performance than our UL IMCS server.
 - I suggest applying to this or similar calls, if you need resources for your code.

Current architectures

- NN, CNN, ResNet, U-Net, GNN+MLP, ...

Future

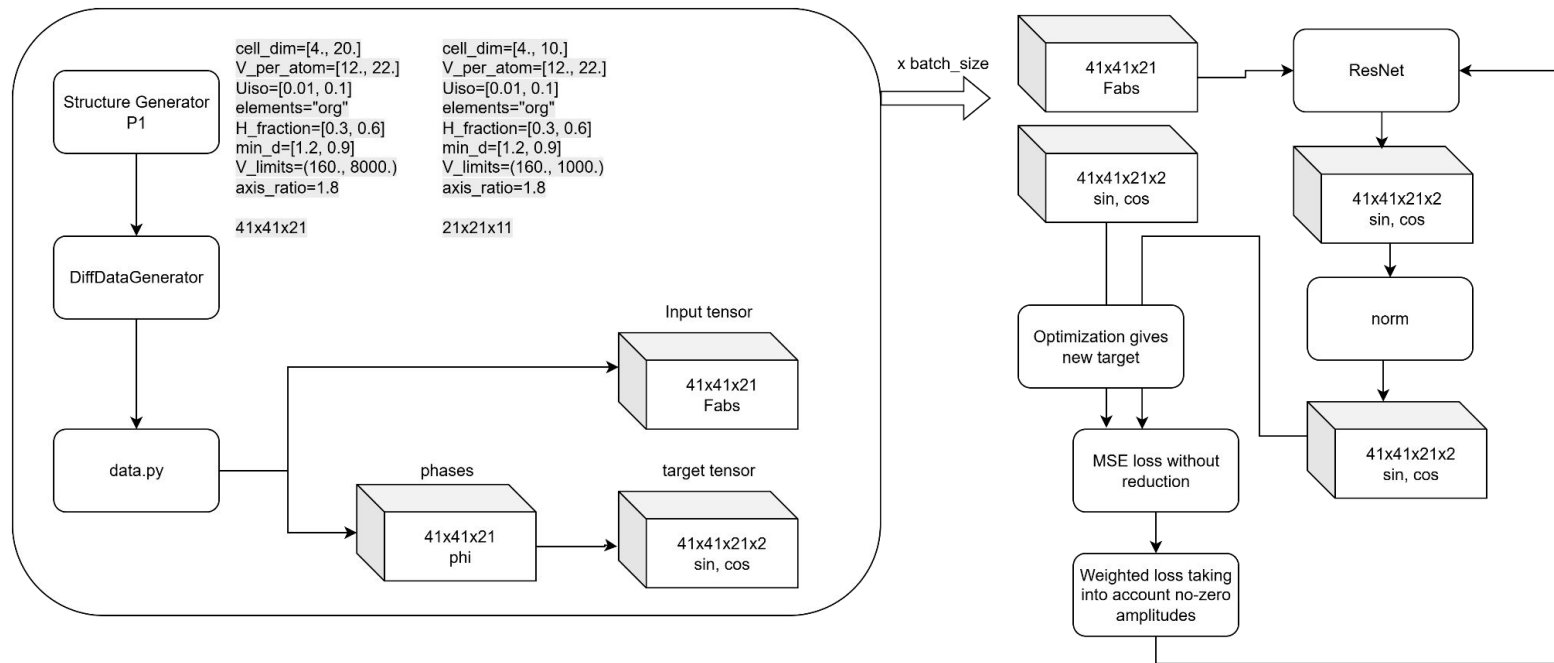
General:

- Trying to start collaborations with researchers from USA, Denmark and Czech Republic, possibly more.

Experiments:

- Modifications of current architectures;
- Diffusion U-Net;
- 1 B or 10 B model;
- ...

Current pipeline



Tech stack

- Python;
- PyTorch;
- PyTorch Geometric;
- TensorBoard;
- Git (GitHub);
- PyCharm.

Planned publicity

Nothing published yet, but planned:

- First paper – problem description and initial experiments;
- Second paper – architecture that solves the regression variant at least partially;
- Third and following papers:
 - Solving problem with even worse data;
 - Publishing data base and analysis;
 - Architecture that solved the classification variant with good performance.

References

- UL FST logo
https://eztf.lu.lv/fileadmin/user_upload/lu_portal/eztf.lu.lv/EZTF_logo/Fakultates_logo/Anglu_valoda/EZTF_eng_horiz.png;
- UL IMCS logo <https://www.facebook.com/100064780443958/photos/1266155592220444/>;
- Hex <https://upload.wikimedia.org/wikipedia/commons/a/a3/Hex-board-11x11-%282%29.svg>;
- RSG No.1 logo:
<https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcSSOU3IEf8LubyZfJoqs66FF5cXOtI22-h98w&s>;
- UL logo:
https://www.lu.lv/fileadmin/user_upload/LU.LV/www.lu.lv/Logo/Logo_jaunie/2024_LU_logotips_ENG_horiz.png.png;
- UL IMCS logo: <https://lumii.lv/static/images/IMCSULlogoLV.svg>;
- Visma logo https://commons.wikimedia.org/wiki/File:Visma_logo.png;
- Visma Resolve logo:
https://nabla.no/media/thumbnails/uploads/news_pictures/Logo_Resolve_color_3.png.770x300_q95_background-white_detail_upscale.jpg.

Thank you for your attention!
Questions? Suggestions?

