Optimizing Detector-Grade Yield in HPGe Crystal Growth for Rare-Event Searches Using Machine Learning

The University of South Dakota

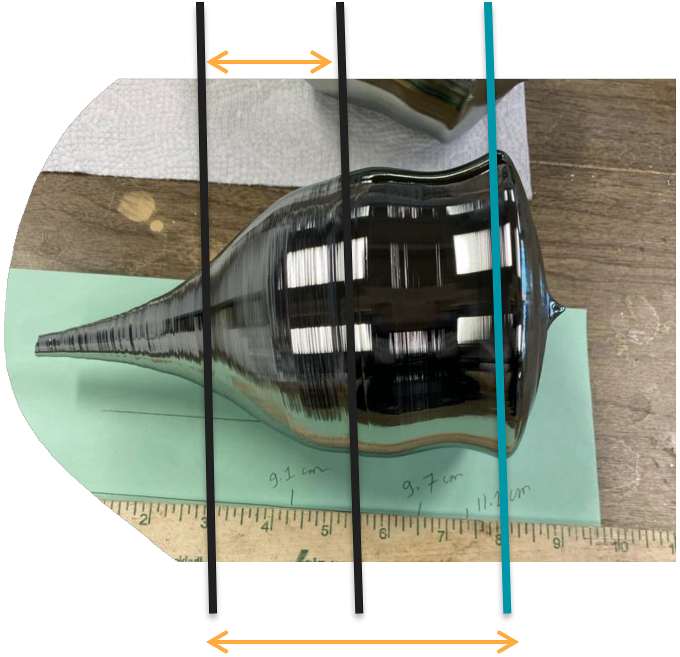
Athul Prem, Sanjay Bhattarai, Narayan Budhathoki, Kunming Dong, Austin Warren, Sunil Chhetri, Dongming Mei

This work was supported in part by NSF OISE 1743790, NSF PHYS 2310027, DOE DE-SC0024519, DE-SC0004768, and research center supported by the State of South Dakota.

Current region of
detector grade region
(20%-30%)

Motivations: Increase the detector grade portion in Ge Crystals

- ➢ Large size detectors reduces background rate per kilogram significantly due to less cable and electronic requirement
- ➢ Extremely difficult to get large portion of crystal to be detector grade due to difference in segregation of various impurities
- ➢ Detailed study of impurity segregation required.
- ➢ Plan to use machine learning tools to analyze different variables and build a predictive model for growing large portion of a crystal to be a detector grade



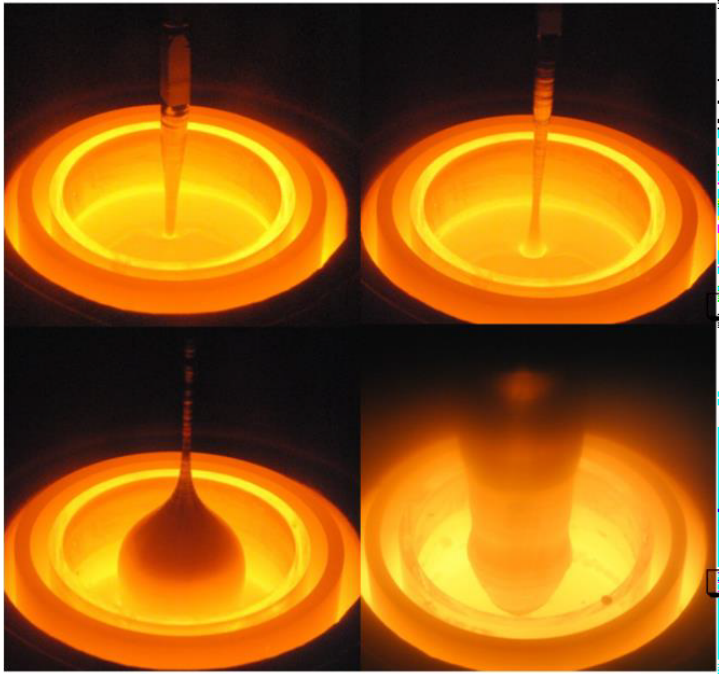
Goal to get detector grade (60 % of total crystal mass)

Crystal growth process

- The materials used for input are either freshly zone-refined ingots or portions of crystals previously identified as non-detector-grade.

- Growth is initiated by immersing a (100) Ge seed crystal in the melt and withdrawing it gradually, ensuring the melt temperature remains just above the melting threshold.

- To control CZ crystal growth, parameters such as pull speed, seed rotation, and melt temperature (set by applied power) are systematically varied.



Direct Detection: Importance of HPGe in Dark Matter Detection

In direct detection, we do **not observe the dark matter particle itself**, but the **nuclear recoil energy** it imparts when it scatters off a target nucleus.

The recoil energy E_R imparted to the nucleus is given by:

$$E_R = \frac{2\mu^2 v^2}{m_N}(1 - \cos\theta)$$

$$\text{where } \mu = \frac{m_\chi m_N}{m_\chi + m_N} \text{ is the reduced mass}$$

When it comes to the germanium detectors the recoil energy and the resolution is precise compared to any other detection method:

Recoil energy for $m_\chi = 100\ \mathrm{GeV}/c^2$ on Ge:

$$E_R^{\mathrm{max}} \approx 25.8\ \mathrm{keV}$$

Modern HPGe detectors can detect down to:

Threshold $E_{\mathrm{th}} \sim 50 - 100\ \mathrm{eV}$

High energy resolution will be <0.2% at MeV scale. They also have **Dual signal readout**: phonon + ionization → accurate recoil discrimination

Challenges in Current HPGe detectors

The main challenges in the HPGe crystals are the scalability.

Current detectors are only in range of 1-10kg, whereas when it come to xenon we can scale it to ton scale.

when the HPGe detectors scale is low their possibility of detection also reduces significantly.

Dark matter interacts extremely weakly ($\sigma_{\chi N} \sim 10^{-46}$ cm^2 or smaller)

Expected event rate: $\lesssim 1$ event per ton-year

Smaller detectors (e.g., 1-10 kg) see < 1 event in years

Ton-scale mass is required to build up enough exposure ($M \times T$)

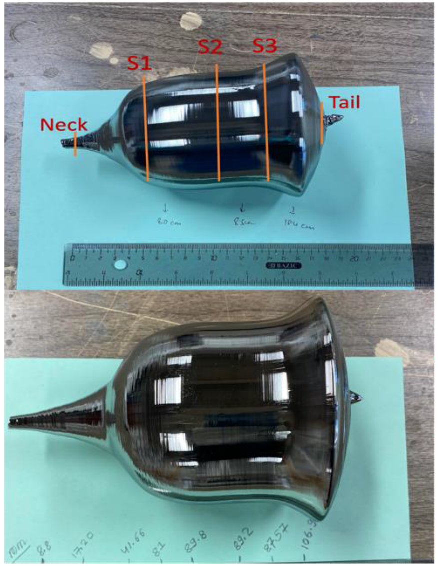
Only then can we explore the full parameter space down to the neutrino floor

Ge crystal purification and crystal growth is a rigorous task, getting ~$10^{10}/cm^3$ impurity range in every crystal growth is extremely hard.

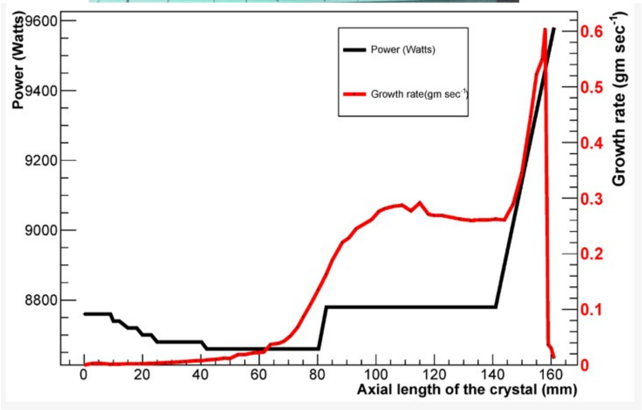
At USD we have 10 years of data from past crystal growth that can be used for Machine learning to increase detector grade portion

Input parameters

position	Net Impurity ($/cm^3$)	Resistivity (Ω cm)	Mobility (cm^2/Vs)
Neck	1.01×10^{11}	5.40×10^{2}	3.24×10^{4}
S1	6.31×10^{10}	3.74×10^{3}	3.63×10^{4}
S2	1.55×10^{10}	4.92×10^{3}	3.48×10^{4}
S3	-1.13×10^{11}	3.80×10^{2}	2.14×10^{4}
Tail	-1.04×10^{13}	6.32×10^{1}	1.41×10^{4}



- The average net impurity of previous crystal or zone refined ingots acquired using Hall effect measurement , mass of input materials, applied power in the process, growth rate of the crystal (time dependent).
- Rotation rate, Pull rate of crystal are kept same for uniformity for all growths



Data Preparation

- Already have a large data set from the crystals grown over 10 years of time.

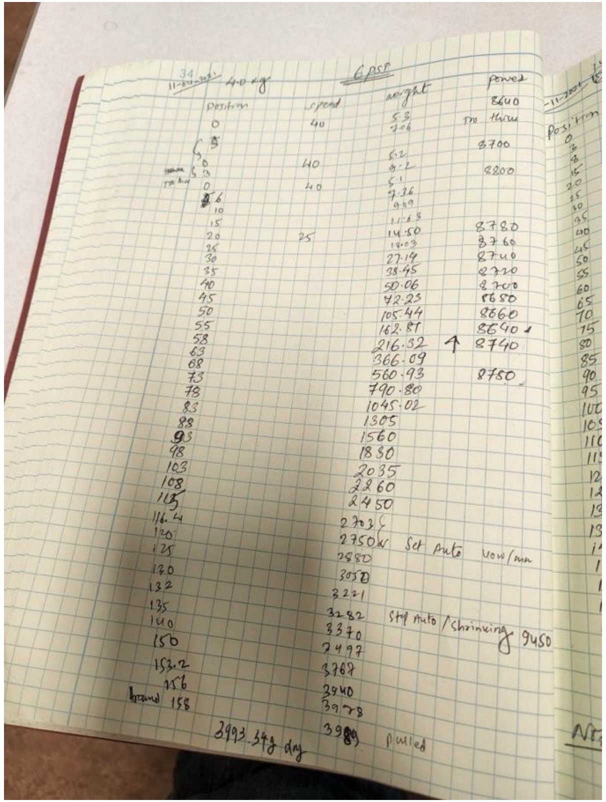
- Data taken include quality of input material, pull rate, rotation rate, applied power, growth rate, real time mass of the crystal and real time photo of the growth process.

- Used data sets from 20 crystal growths for first study, then expanded to 50 crystal growth data.

  Input Parameters: Time, Growth Rate, Power, Impurity of previous crystal etc.

Output

- The percentage of detector grade region.



Example data set

- Using pull rate, position of the crystal and real time mass of the grown crystal, time and growth rate can be determined.

- From the hall effect measurement of the input materials and their masses, number of net impurity atoms is estimated

Time (sec)	Power(W)	Growth Rate (gm/sec)	No. of net impurity atoms added	Number of net impurity of previous crystal	Output net impurity(/cm^3)	Detector grade region(%)
0	8720	0	3.53383E+13	3.42969E+14	4.74341E+11	38
900	8700	0.091111111	3.53383E+13	3.42969E+14	4.74377E+11	38
1350	8700	0.054074074	3.53383E+13	3.42969E+14	4.73089E+11	38
2880	8680	0.015277778	3.53383E+13	3.42969E+14	4.72314E+11	38
3600	8660	0.011944444	3.53383E+13	3.42969E+14	4.71557E+11	38
4320	8640	0.018287037	3.53383E+13	3.42969E+14	4.70171E+11	38
5040	8620	0.02797619	3.53383E+13	3.42969E+14	4.67707E+11	38
5760	8600	0.031944444	3.53383E+13	3.42969E+14	4.64508E+11	38
6480	8580	0.047222222	3.53383E+13	3.42969E+14	4.59235E+11	38
7200	8560	0.076527778	3.53383E+13	3.42969E+14	4.4988E+11	38
8928	8560	0.329189068	3.53383E+13	3.42969E+14	4.02918E+11	38
10080	8560	0.416666667	3.53383E+13	3.42969E+14	3.43706E+11	38
10800	8560	0.407407407	3.53383E+13	3.42969E+14	2.90464E+11	38
12960	8660	2.345679012	3.53383E+13	3.42969E+14	8.61E+10	38
13464	8700	0.935828877	3.53383E+13	3.42969E+14	5.04E+10	38
15120	8700	3.121693122	3.53383E+13	3.42969E+14	-5.03E+10	38
15984	8700	1.551551552	3.53383E+13	3.42969E+14	-5.87E+10	38
16560	8700	0.899758454	3.53383E+13	3.42969E+14	-6.48E+10	38
18000	8700	2	3.53383E+13	3.42969E+14	-1.4565E+11	38
18720	8700	0.822649573	3.53383E+13	3.42969E+14	-2.4313E+11	38
19440	8700	0.601851852	3.53383E+13	3.42969E+14	-3.67953E+11	38
20160	8700	0.694444444	3.53383E+13	3.42969E+14	-6.2403E+11	38
21600	8700	0.925925926	3.53383E+13	3.42969E+14	-6.63659E+11	38
23040	8700	0.737847222	3.53383E+13	3.42969E+14	-7.24308E+11	38
24048	8700	0.461576846	3.53383E+13	3.42969E+14	-7.71031E+11	38
26582.4	9404	1.241422896	3.53383E+13	3.42969E+14	-9.60471E+11	38
29232	9500	2.035440613	3.53383E+13	3.42969E+14	-1.85107E+12	38

Example data set for one crystal growth

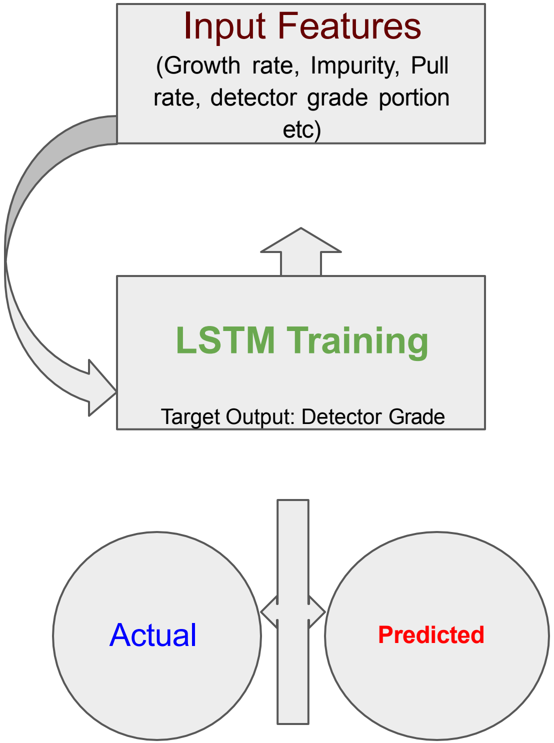
Long Short-Term Memory(LSTM) in Crystal Growth Optimization

- LSTM (Long Short-Term Memory) is a type of recurrent neural network (RNN) designed to handle sequential and time-series data by learning long-range dependencies.

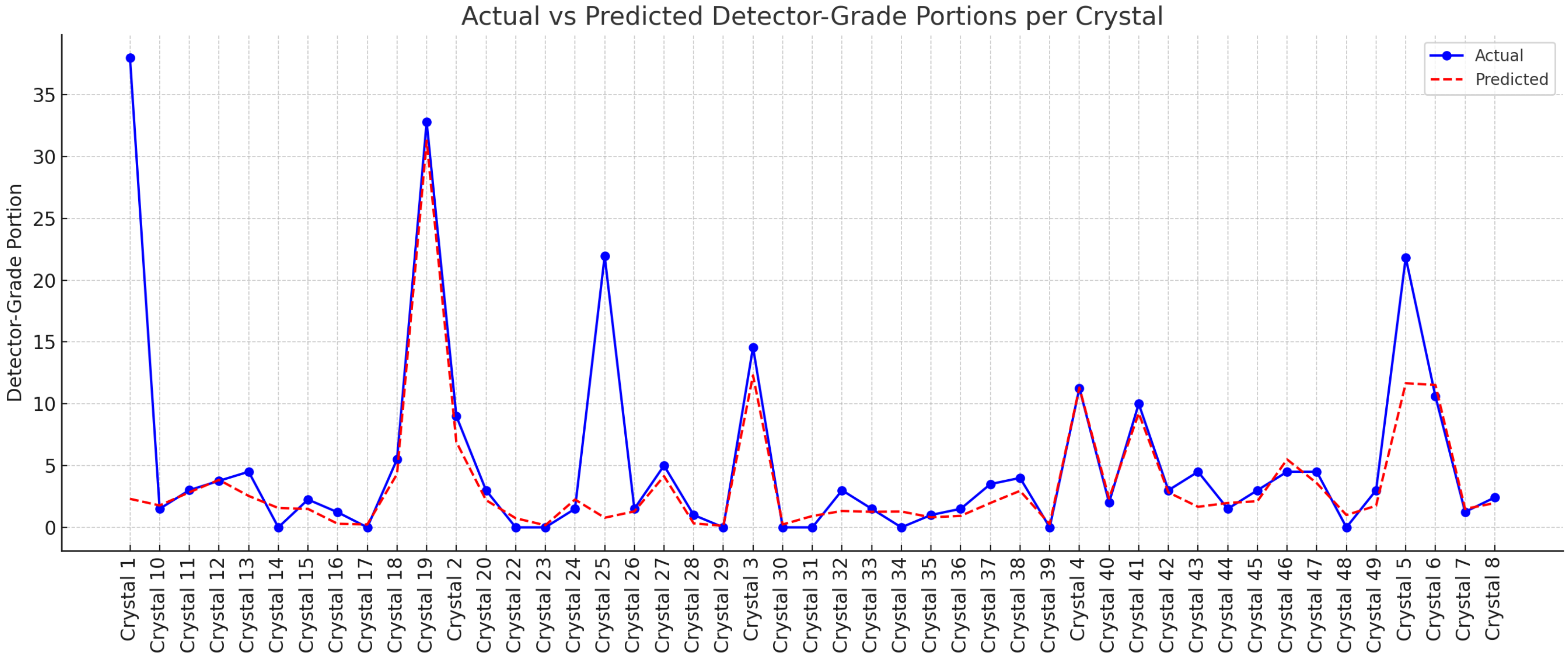
- LSTMs use gated memory cells to retain or forget information selectively, making them well-suited for dynamic process prediction.

- Crystal growth is a time-dependent process, LSTMs capture temporal relationships in experimental data, making them ideal for predicting detector-grade yield based on past growth conditions.

- Unlike traditional regression models, LSTMs learn complex, nonlinear correlations between parameters without requiring predefined equations.

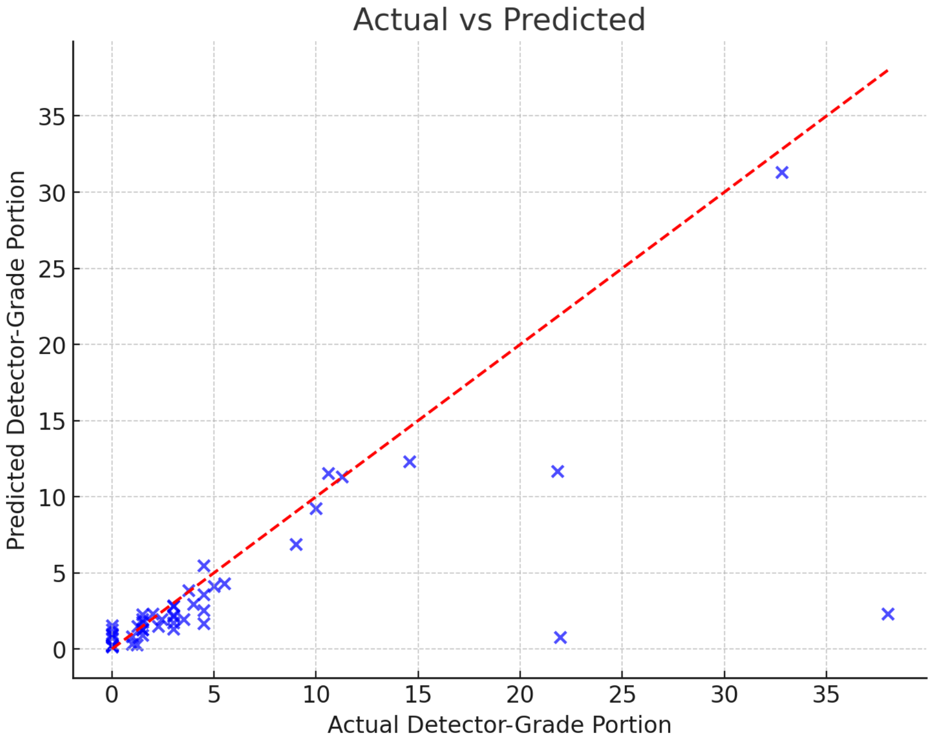


Data Driven LSTM: Actual vs Predicted Detector Grade



- At USD, Ge crystals are grown frequently and numerous HPGe crystal growth data have been collected.
- Past 50 HPGe crystal growth data have been used to train the LSTM model.
- Model has generated successful prediction attaining an average of **90% accuracy** in most trials.
- The LSTM model can also be used to predict the detector grade portion of newer crystal based on input physical parameters, like impurity, growth rate etc.
- Higher the number of data, higher the accuracy of the LSTM model.

Accuracy of the Model



- Most data points correspond to detector-grade portions below 20%, with very few samples above 20%.
- The model predicts well for the majority (below 20%).
- Limited data for higher values, affecting accuracy in that range above 20%

LSTM Feature Importance Analysis (SHAP analysis)

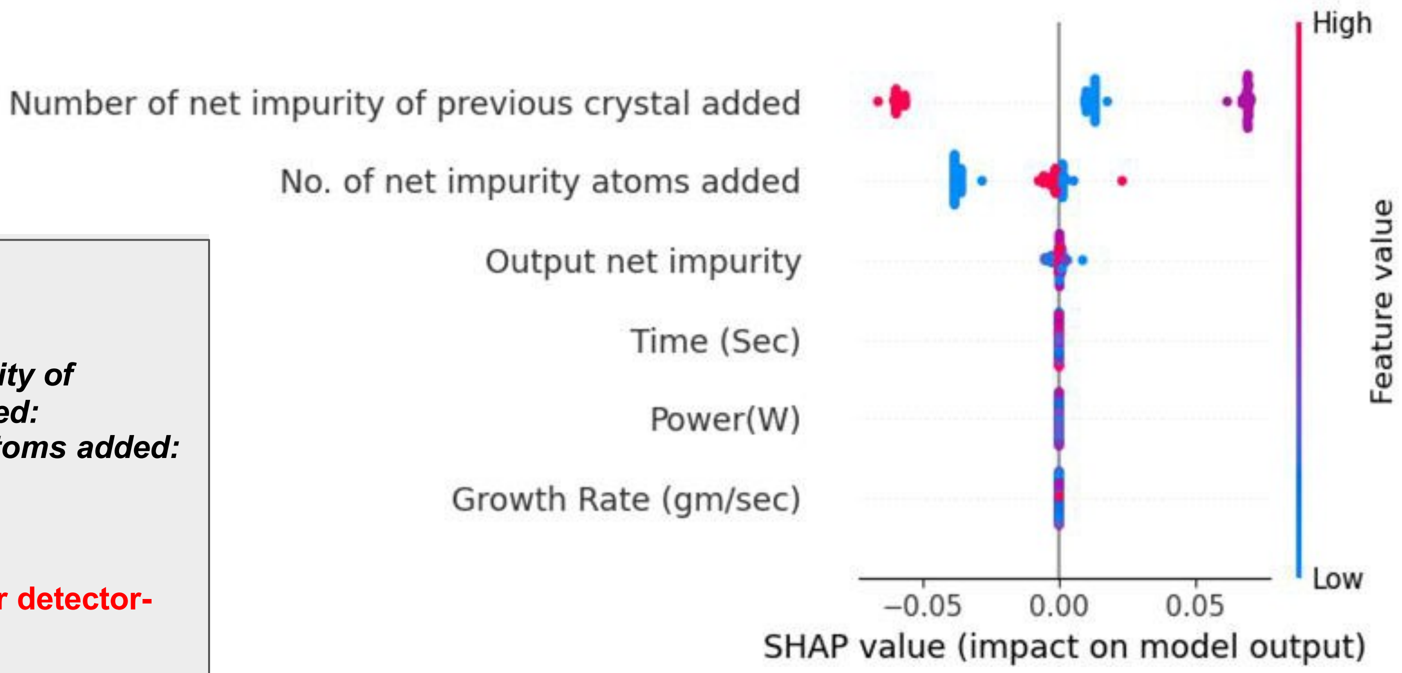
Important Features:

1. ***Number of net impurity of previous crystal added:***
2. ***No. of net impurity atoms added:***

Results:

→ **Higher impurity → Lower detector-grade**

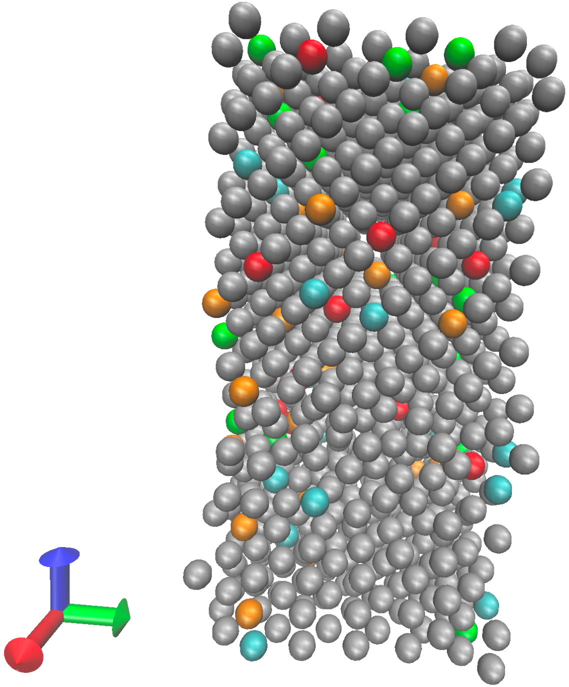
→ **Lower impurity → Higher detector-grade**



→ SHAP analysis indicates that Time, Power, and Growth Rate(different from experience) have minimal impact on the detector-grade portion.

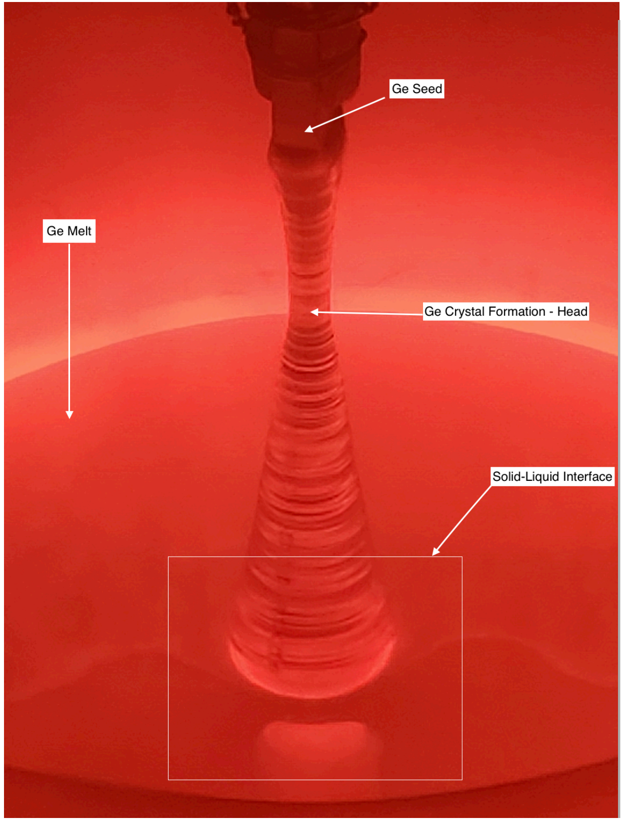
Ongoing Research

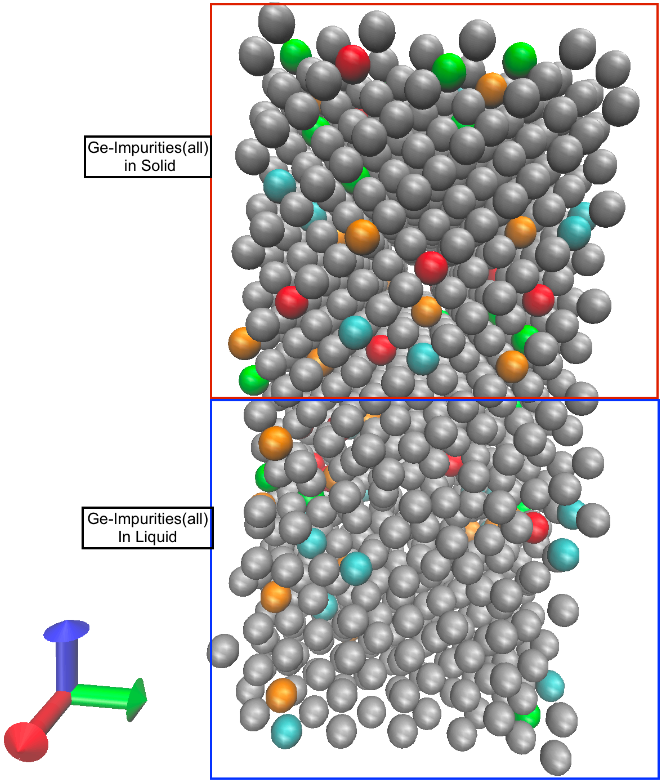
- From ML model we understood impurities impact the detector grade portion
- Investigating impurity segregation during crystal growth is key to understanding impurity transport behavior under varying thermal profiles
- Molecular dynamics enables detailed analysis of the solid–liquid interface in crystal growth by modeling both pure Ge-Ge interaction and Ge–impurity systems
- With no available SW potentials for Ge–impurity systems, ML-generated potentials can be employed to model the melt region in simulations.



Solid-Liquid interface of Ge

Impurity segregation at Solid-Liquid Interface during Crystal Growth





Future Work:

- Expanding the Dataset for Improved Model Accuracy, Include more diverse process conditions to make the model robust against variations. Integrate historical growth data from past experiments to enhance predictive capabilities, optimized search for higher detector grade feeding optimized input parameter.

- Enhancing Detector-Grade Yield through Zone Refining Integration, Expand the dataset structure by incorporating zone refining process data alongside crystal growth parameters. Model impurity movement across refining stages to optimize the number of passes required for higher purity.

- Real-Time Optimization for HPGe Growth, Implement a real-time machine learning feedback loop to adjust parameters dynamically during growth. Develop an automated process control system integrating LSTM predictions with experimental setups.

Questions or Suggestions?