"Materials research by Information Integration" Initiative
情報統合型物質・材料開発イニシアティブ

物質・材料記述基盤グループ

GL 吉田 亮

劉 暢 小山 幸典


Liu, C.[3]


Koyama, Y.[3]


Yoshida, R.[1,2,3]

[1] The Institute of Statistical Mathematics
[2] SOKENDAI
[3] National Institute for Materials Science

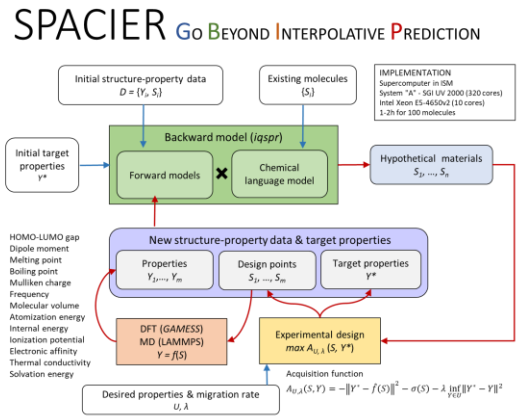
XenonPy

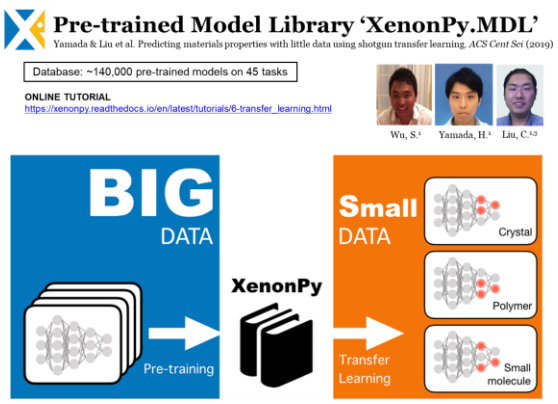
"Materials research by Information Integration" Initiative
情報統合型物質・材料開発イニシアティブ

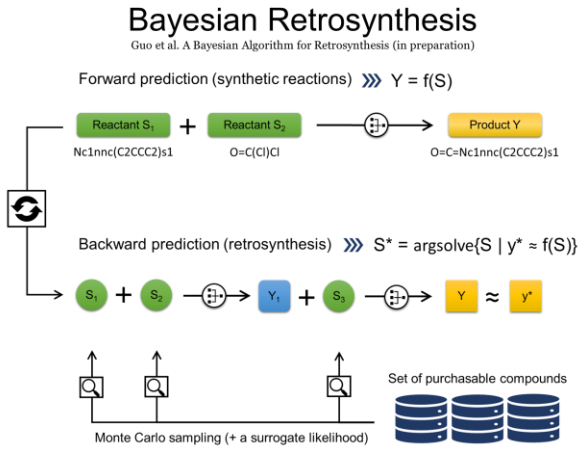
物質・材料記述基盤グループ

Liu, C.[1]

- Descriptor library
- Database of 140,000 pre-trained models (XenonPy.MDL)
- Transfer learning
- Bayesian molecular design algorithm (iQSPR-X, SPACIER) and so on







- Yamada & Liu et al. Predicting materials properties with little data using shotgun transfer learning. *ACS Cent Sci*. 5:1717-1730 (2019).
- Wu et al. Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm. *npj Comput Mater* 5:66 (2019).
- Wu et al. iQSPR in XenonPy: a Bayesian inverse molecular design algorithm. *Mol Inform* (2020). doi:10.1002/minf.201900107
- Ju et al. Exploring ultrahigh lattice thermal conductivity crystals via feature-based transfer learning. *ChemRxiv* (2019). doi:10.26434/chemrxiv.9850301.v1
- Ikebata et al. Bayesian molecular design with a chemical language model. *J Compt Aided Mol Des*. 31:379-391 (2017)
- Kusaba et al. Recreation of the periodic table with an Unsupervised Machine Learning Algorithm. arXiv (2020). arXiv:1912.10708
- Guo et al. A Bayesian algorithm for retrosynthesis (submitted)

Materials Descriptor Platform group (2017.4 -)

Research Organization of Information and Systems
The Institute of Statistical Mathematics

Materials Descriptor Platform group


Koyama, Y.[3]


Yoshida, R.[1,2,3]


Liu, C.[1]

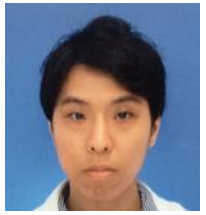
ISM Data Science Center for Creative Design & Manufacturing


Liu, C.[1]


Wu, S.[1]


Noguchi, Y.[1]


Hayashi, Y.[1]


Yamada, H.[1]


Fukumizu, K.[1]


Guo, Z.[1]


Kusaba, M.[2]


Minami, S.[2]


Zhang, Q.[2]

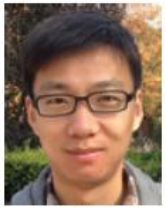

Iwayama, M.[2]


Lambard, G. (NIMS)

Shiomi group @mi2i-thermal

Shiomi, J.[6], Ju, S[6], Hongo, K.[4]


Shiomi, J. (U Tokyo)


Ju, S. (U Tokyo)

Morikawa group @mi2i-thermal

Morikawa, J.[5], Kakimoto, M.[5], Kondo, Y.[3], Kuwajima, I.[3], Xu, Y.[3], Hongo, K.[4]

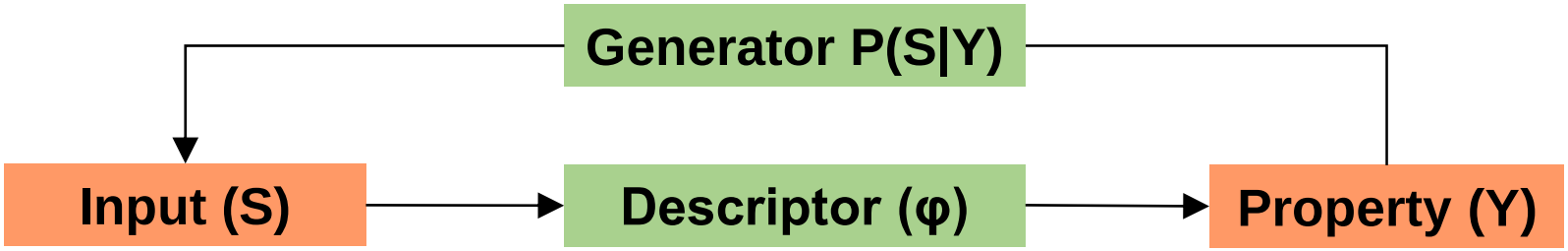
PolyInfo


Morikawa, J. (Tokyo Tech)

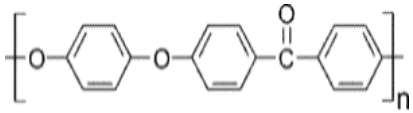

Wu, S.[1]

A General Workflow of MI

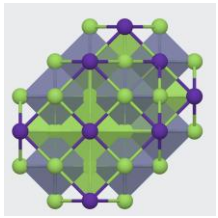
Representation, learning and generation for various materials



Chemical structure



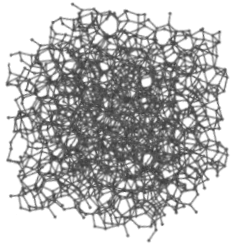
Crystal



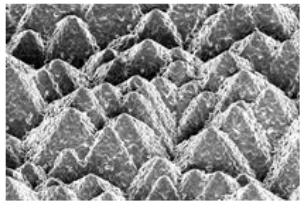
Composition

$A_aB_bC_c$

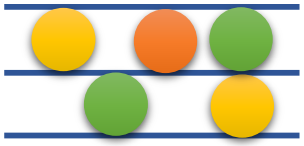
Amorphous



Microstructure



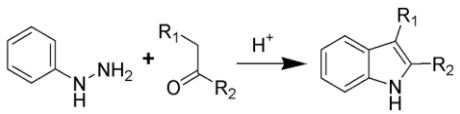
Composite / hybrid



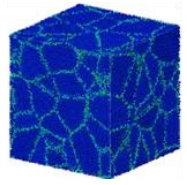
Process parameter

Θ, T

Chemical reaction



Nanostructure



ML-assisted discovery of new polymers

Wu, Yamada, Yoshida (ISM)
Morikawa, Kakimoto (Tokyo Tech)
Xu, Kuwajima, Kondo, Lambard (NIMS)
Hongo (JAIST)
Shiomi (Univ of Tokyo)
Yang, Schick (Univ of Rostock)


Morikawa, J
(Tokyo Tech)


Wu S
(ISM)

機械学習の「記憶」を活用し、高分子の熱伝導性の大幅な向上に成功
～少ないデータでも高精度な予測が可能に　高分子での材料インフォマティクス加速に期待～

(press release: https://www.ism.ac.jp/ura/press/ISM2019-07.html)

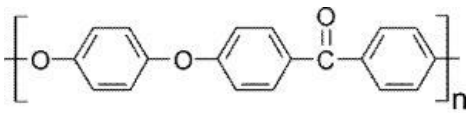
Wu et al. Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm. *npj Comput Mater* 5:66 (2019).zzzz

WORLD LARGEST POLYMER DATABASE　　OPEN SOURCE PLATFORM OF MI

PolyInfo × XenonPy

Target: higher thermal conductivity polymers

Structure (S)　　**Descriptor (φ)**　　**Properties (Y)**



0100010000011010011

λ Thermal conductivity
T_g Glass transition temp.
T_m Melting points
C_p Heat capacity
and so on

S = chemical structure of constitutional repeat unit

"Materials research by Information Integration" Initiative
情報統合型物質・材料開発イニシアティブ

ML-assisted discovery of new polymers

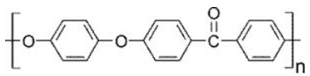
Wu, Yamada, Yoshida (ISM)
Morikawa, Kakimoto (Tokyo Tech)
Xu, Kuwajima, Kondo, Lambard (NIMS)
Hongo (JAIST)
Shiomi (Univ of Tokyo)
Yang, Schick (Univ of Rostock)


Morikawa, J
(Tokyo Tech)


Wu S
(ISM)

機械学習で物性予測モデルを導く

ポリマー (S) → 記述子 (φ) → 物性 (Y)



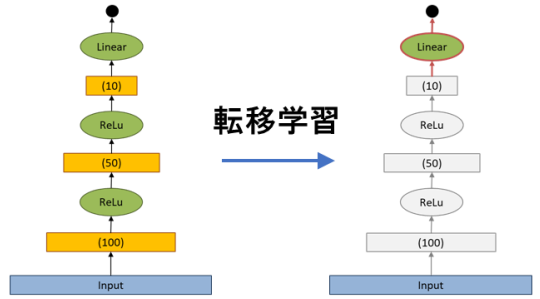
01000100000011010011

λ 熱伝導率
T_g ガラス転移温度
T_m 融点
C_p 比熱

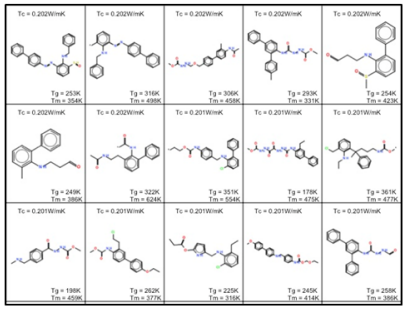
所望の物性を持つポリマーの構造を生成する

関連物性の訓練済みモデルライブラリ

28個のデータで熱伝導率の予測モデルを導く

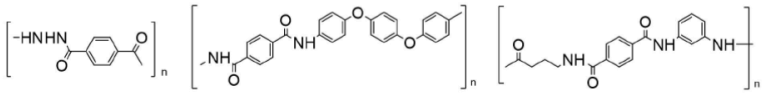


仮想ライブラリの生成

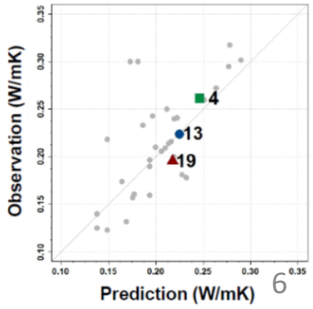


三種類の新規ポリマーを合成，超高速熱分析による熱物性の検証

熱伝導率：従来比最大80%増
高耐熱性
有機溶媒への溶解性
フィルム加工の容易性



転移学習による熱伝導率の予測



"Materials research by Information Integration" Initiative
情報統合型物質・材料開発イニシアティブ

Ultrahigh lattice thermal conductivity crystals

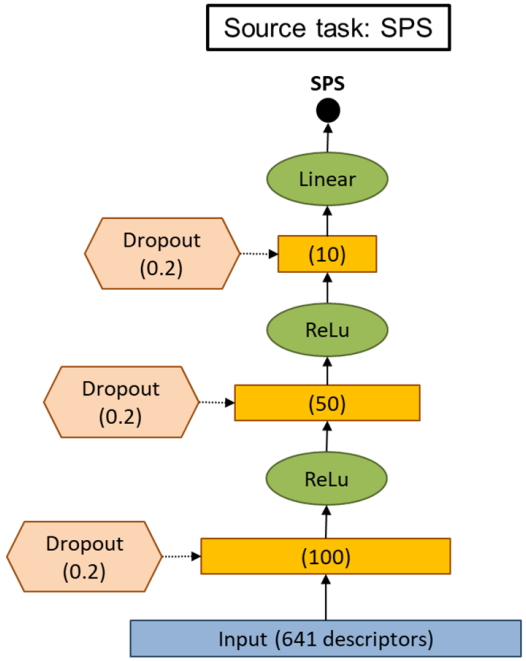
Liu, Yoshida (ISM)
Hongo (JAIST)
Ju, Shiomi (Univ of Tokyo)
Tadano (NIMS)

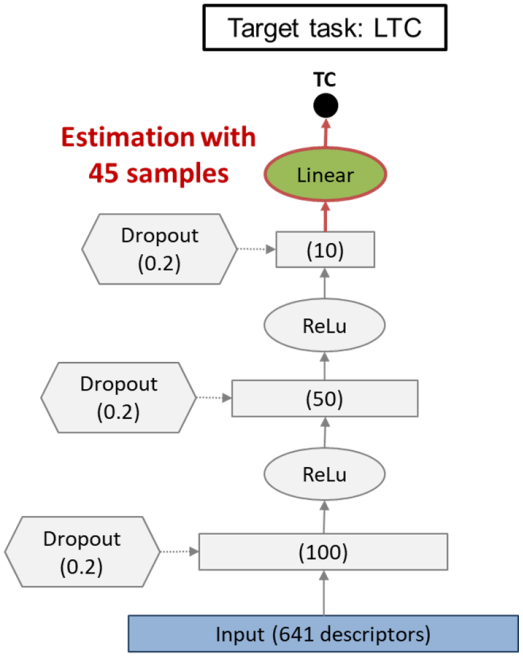


Shiomi, J. (U Tokyo)
Ju, S. (U Tokyo)
Liu, C.[1]

Number of samples on SPS and thermal conductivity

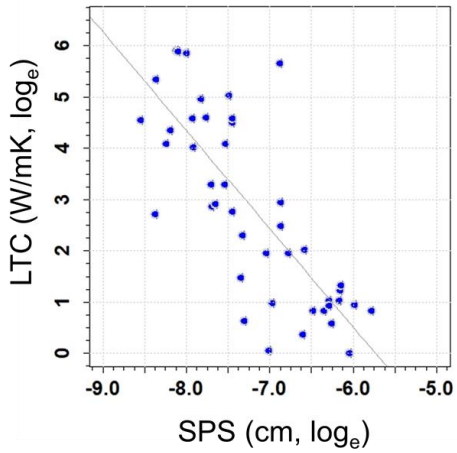
Property	Number of samples
Lattice Thermal conductivity (target)	45 materials
SPS (intermediate)	320 materials





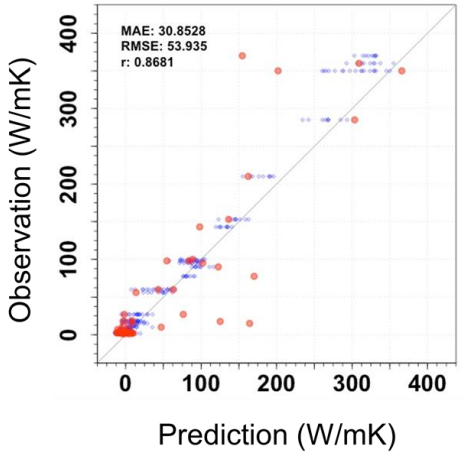
a

Distribution of LTC and SPS



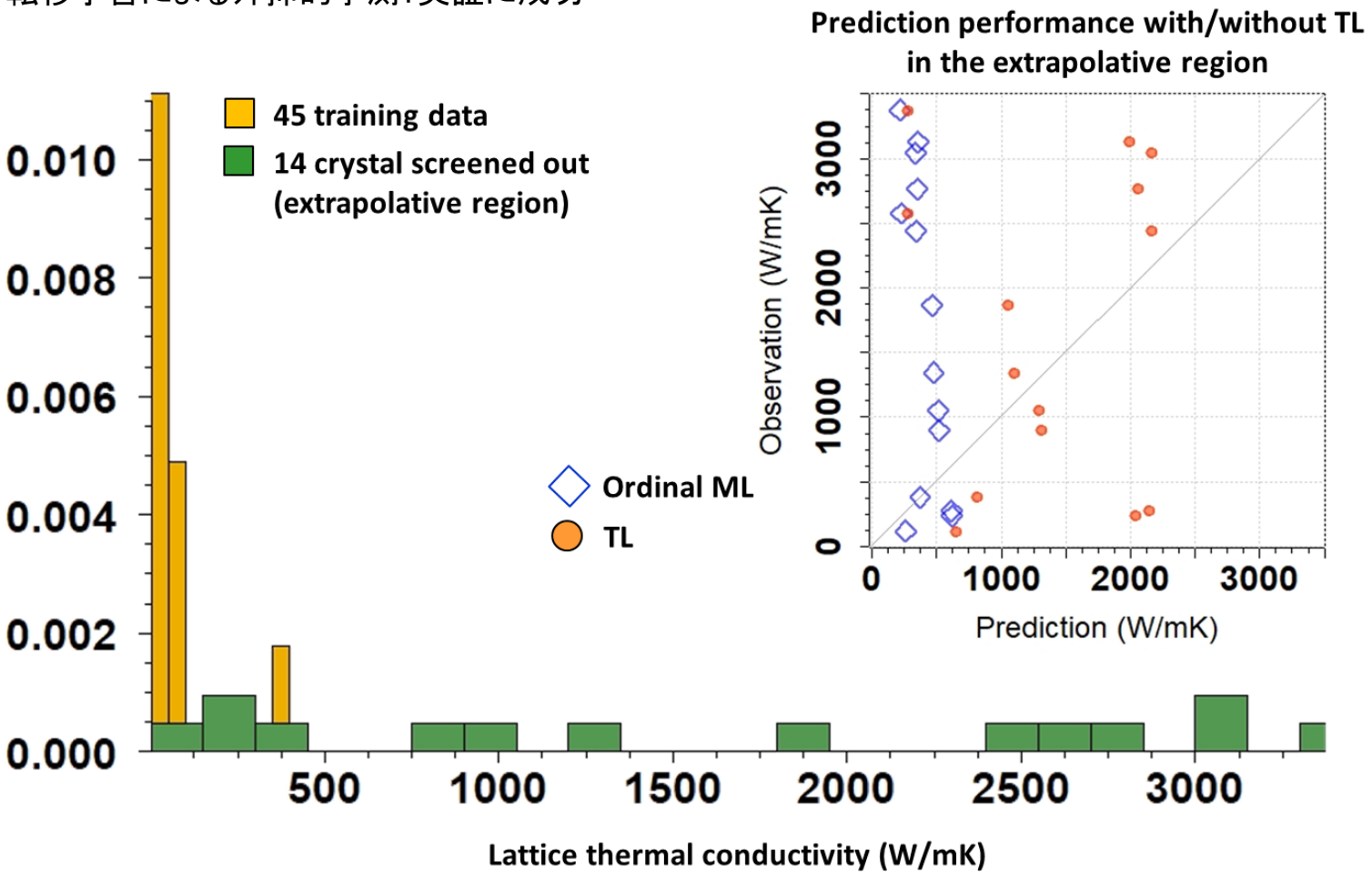
b

Prediction of LTC for the transferred model



TL & Extrapolative Prediction

- 訓練済みモデルライブラリXenonPy.MDLと転移学習を活用したハイスループットスクリーニングにより超高熱伝導性無機化合物を発見
- 転移学習による外挿的予測：実証に成功



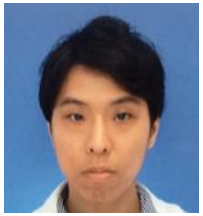
- Ju et al. Exploring ultrahigh lattice thermal conductivity crystals via feature-based transfer learning. *ChemRxiv* (2019)
- Yamada & Liu et al. Predicting materials properties with little data using shotgun transfer learning. *ACS Cent Sci* (2019)

Pre-trained Model Library ‘XenonPy.MDL’

Yamada & Liu et al. Predicting materials properties with little data using shotgun transfer learning. *ACS Cent Sci* (2019)

Database: ~140,000 pre-trained models on 45 tasks

ONLINE TUTORIAL
https://xenonpy.readthedocs.io/en/latest/tutorials/6-transfer_learning.html


Wu, S.[1] Yamada, H.[1] Liu, C.[1,3]

- Small data
- Devise performance
- Computational vs experimental properties
- Monomer vs polymers, organic vs inorganic materials

Monomers

Dipole moment
Isotropic polarizability
HOMO
LUMO
Gap, difference between LUMO and HOMO
Electronic spatial extent
Zero point vibrational energy
Internal energy at 0 K
Internal energy at 298.15 K
Enthalpy at 298.15 K
Free energy at 298.15 K
Heat capacity
Hydration free energy
Melting temperature
Boiling temperature

Polymers

Melting point
Glass transition temperature
Density
Viscosity
Dielectric constant
Ionic dielectric constant
Electronic dielectric constant
Refraction index
Atomization energy
Ionization energy
Electron affinity
Cohesive energy
Hildebrand solubility parameter
Molar heat capacity
Molar volume
Bandgap
Thermal conductivity
Heat capacity

Inorganic

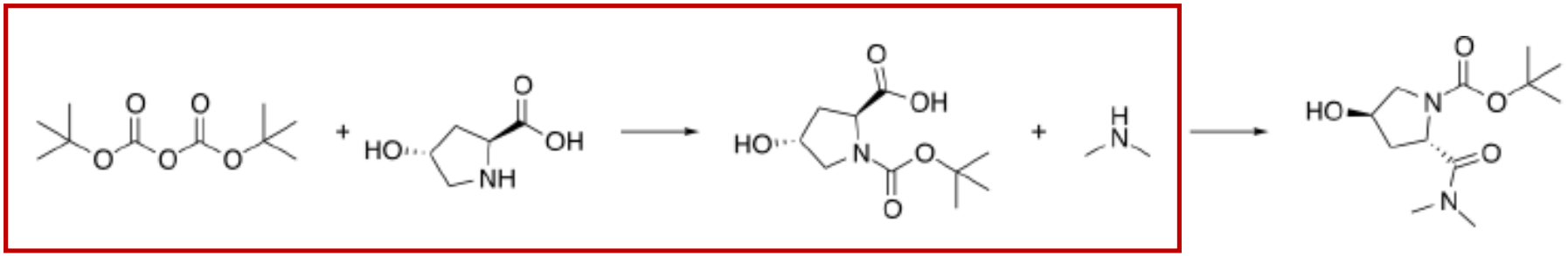
Bandgap
Density
Volume
Fermi energy
Formation energy
Final energy/atom
Total magnetization
Dielectric constant
Refractive index
Space group
Point group
Scattering phase space
Thermal conductivity

Machine Learning for Retrosynthesis

Highly expertized tasks in organic chemistry based on tacit knowledge

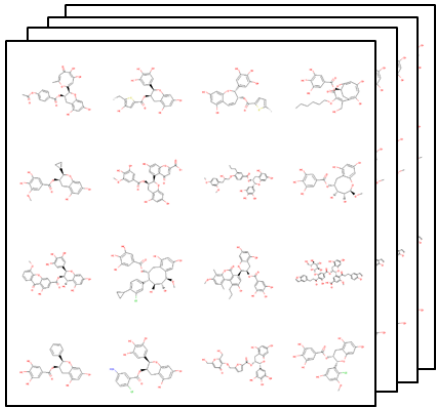
Unknown combinations (10^{18}-10^{21})

Design target



List of commercial compounds (10^{6}-10^{7})

- Initial compounds
- Reactants
- Reagents



Forward Prediction: Synthetic Reaction

Machine Translation & Prediction of Synthetic Reactions

Input

Reactant 1	Reactant 2	Reagent
Nc1nnc(C2CCC2)s1	. O=C(Cl)Cl	> A_CCOC(C)=O

Output

Product
O=C=Nc1nnc(C2CCC2)s1

[SMILES] Seq2Seq (https://github.com/google/seq2seq)
[SMILES] Transformer (https://github.com/pschwllr/MolecularTransformer)
[Graph] Rexgen (https://github.com/connorcoley/rexgen_direct)

Prediction accuracy of DNN trained on NextMove and Reaxys

	Top 1	Top 5
Transformer (Guo et al. in preparation)	**90.9**	**95.5**
Seq2Seq + Reaxys (Schwaller et al. 2017)	80.3	87.5
Template-based (Coley et al. 2017)	72.1	90.7

Guo, Z.
(ISM)

Bayesian Retrosynthesis

Guo et al. A Bayesian algorithm for retrosynthesis (to appear)

Guo, Z.
(ISM)

Forward model $p(Y|L_1, R_2, A_2)p(L_1|S_0, R_1, A_1)$

$$S_0 + R_1 \overset{A_1}{\rightarrow} H_1 + R_2 \overset{A_2}{\rightarrow} Y$$

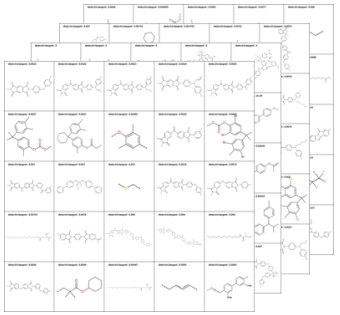
(given design target)

Backward calculation
(Sequential Monte Carlo)

$$p(H, R, A, S|Y)$$

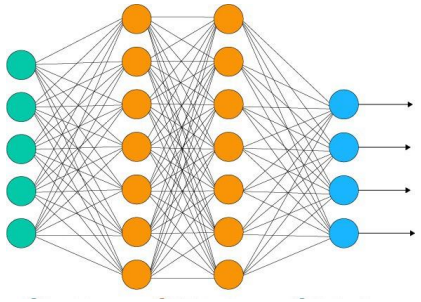
The unknown to be selected from DB (NextMove)

$\boldsymbol{S_0}$: initial compound
$\boldsymbol{R_1}, \boldsymbol{R_1}$: reactants
$\boldsymbol{A_1}, \boldsymbol{A_1}$: catalytic agent



The unknown to be predicted by DNN

$\boldsymbol{H_1}$: intermediate product



Historical View

Rule-based synthetic prediction (manually or computationally compiled)

- Corely et al (1969) "Organic Chemical Simulation of Synthesis"
- LHASA, SYNCHEM, WODCA …
- Low prediction performance (taking into account only local molecular structures)
- Small reaction space

50 years

Open dataset: Lowe's USPTO dataset (10^6 synthetic reactions) for US patented compounds

- **[Text mining]** Lowe. Extraction of chemical structures and reactions from the literature (Doctoral thesis) (2012)
- **[Open data]** Lowe et al. Chemical reactions from US patents (1976-Sep2016). figshare. Dataset. (2017)

2 years

After 2017: Deep Neural Networks – on the eve of groundbreaking days?

- **[seq2seq: Stanford]** Liu et al. ACS Cent Sci (2017)
- **[seq2seq: IBM]** Schwaller et al. Chem Sci, 28 (2018)
- **[Molecular Transformer: IBM]** Schwaller et al. arXiv:1811.02633 (2018)
- **[Graph CNN: MIT]** Jin et al. NIPS (2017)
- **[Molecular Chef]** Bradshaw et al. NIPS (2019)

…..

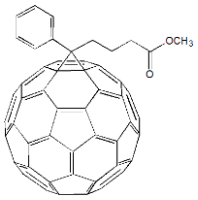
Rediscovery of OPVs: date back to 2009

Use of a time machine to date back to 2009 and the reproduction of the historical developments

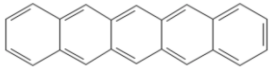
Acceptor (given)

Donner = design target

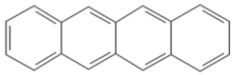
PCBM



pentacene (2004)

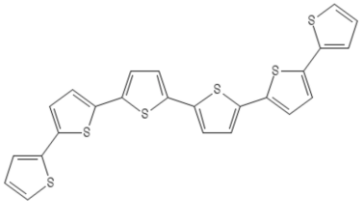


tetracene (2005)

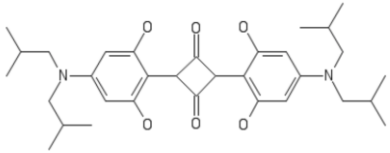


(some known compounds in 2009)

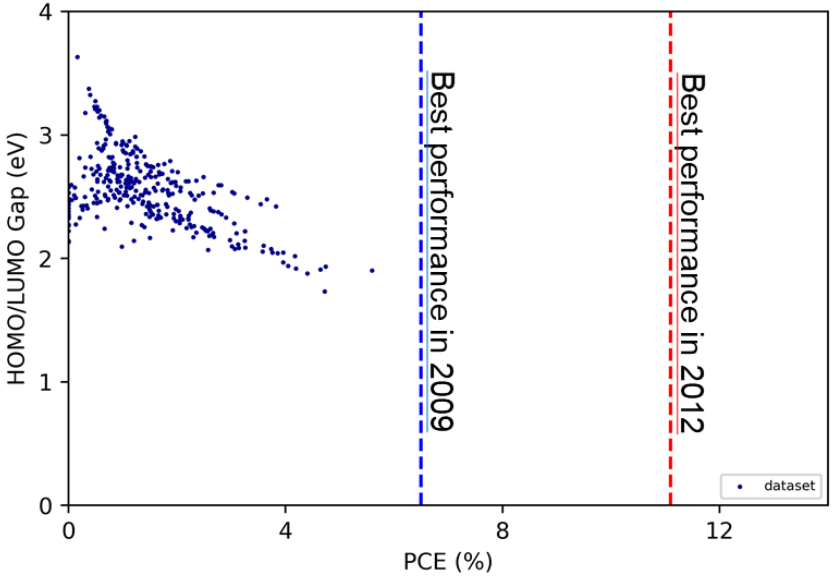
α-sexithiophene (2008)



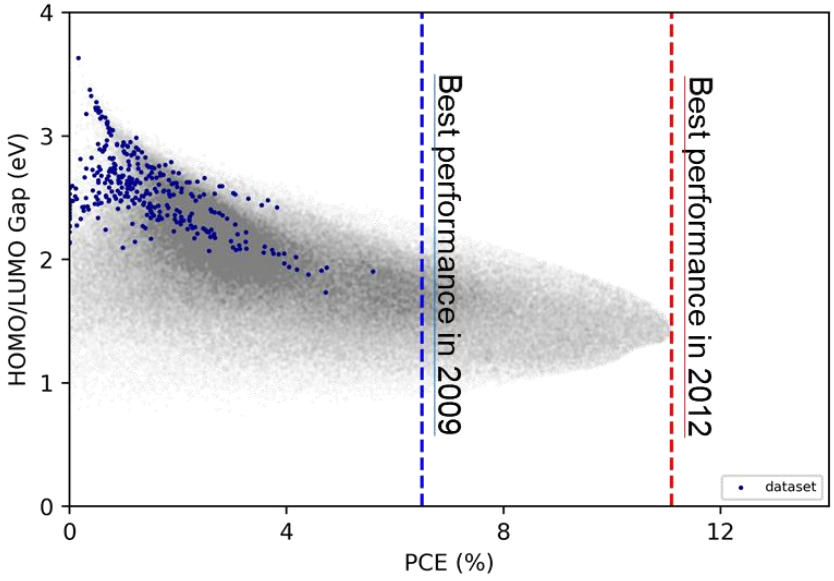
2,4-Bis[4-(*N*,*N*-diisobutylamino)-2,6-dihydroxyphenyl] squaraine (2009)



Distribution in 2009

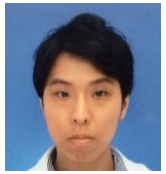


Recreation of the landscape in 2012

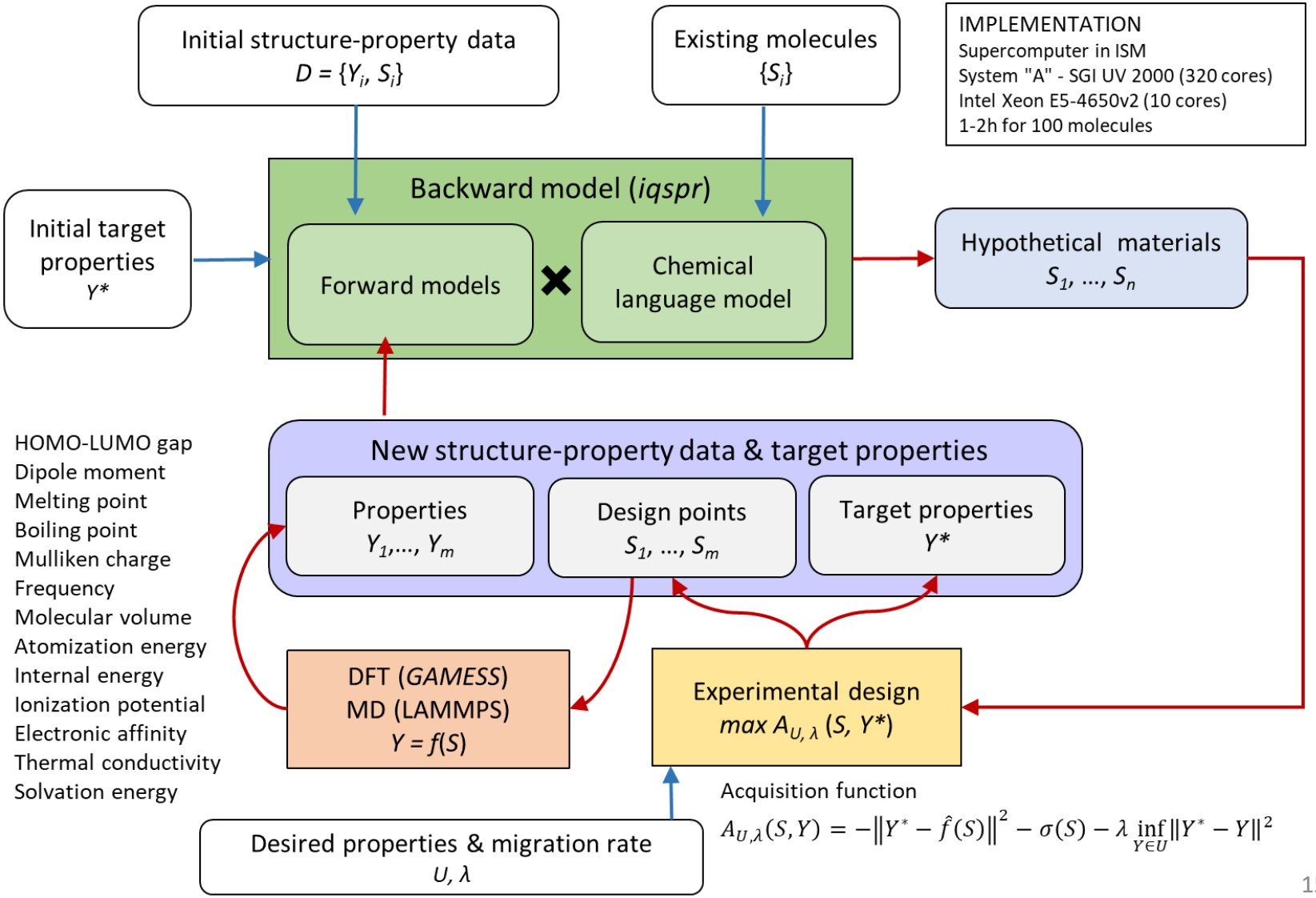


SPACIER Go Beyond Interpolative Prediction


Noguchi, Y


Yamada, H
ISM→Toyaku

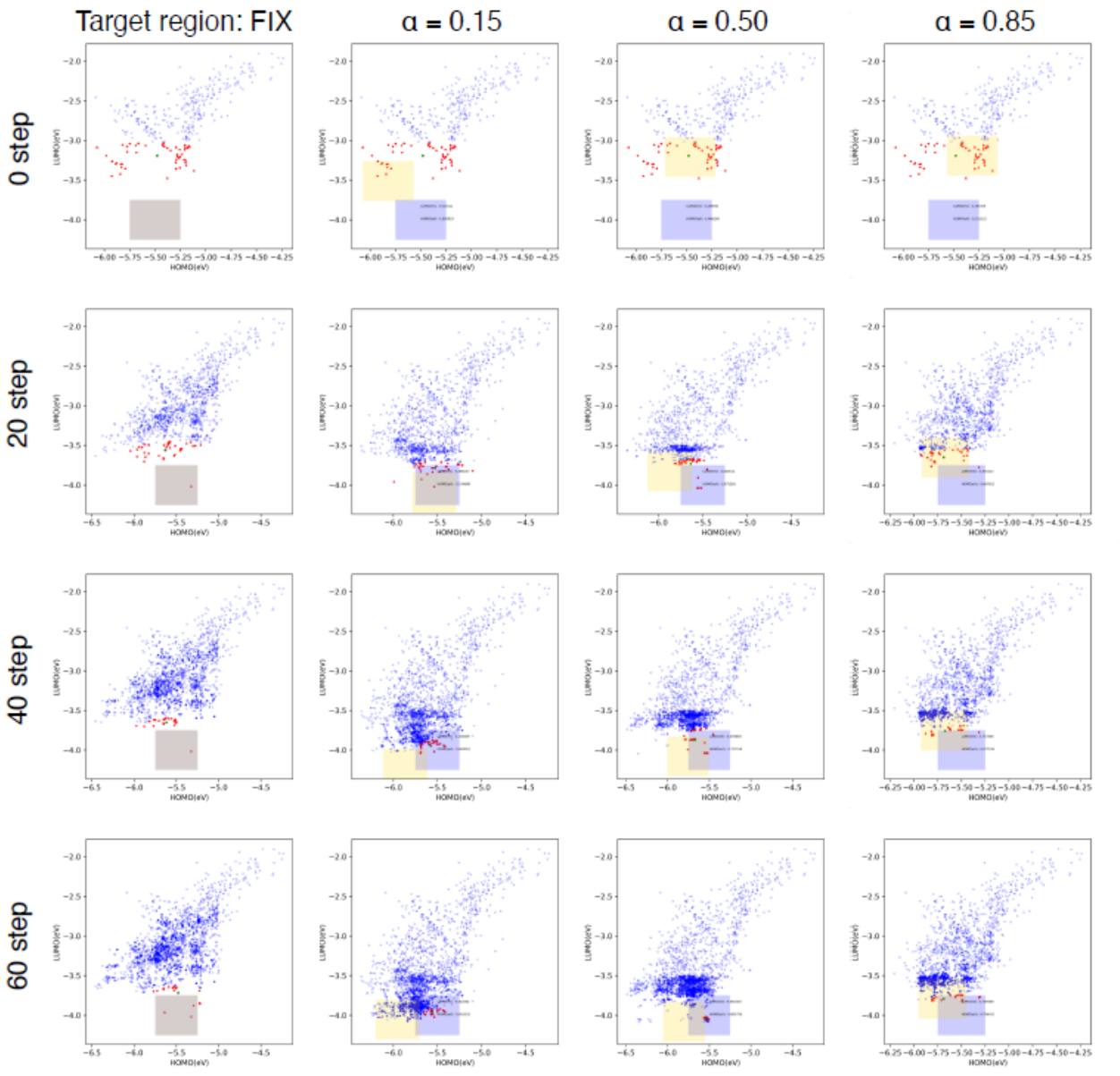

Lambard, G
ISM→NIMS



Scharber's formula for PCBM (Adv Mater 18:789 (2006))

PCE & HOMO-LUMO gap ➤ HOMO & LUMO (DFT-computable)

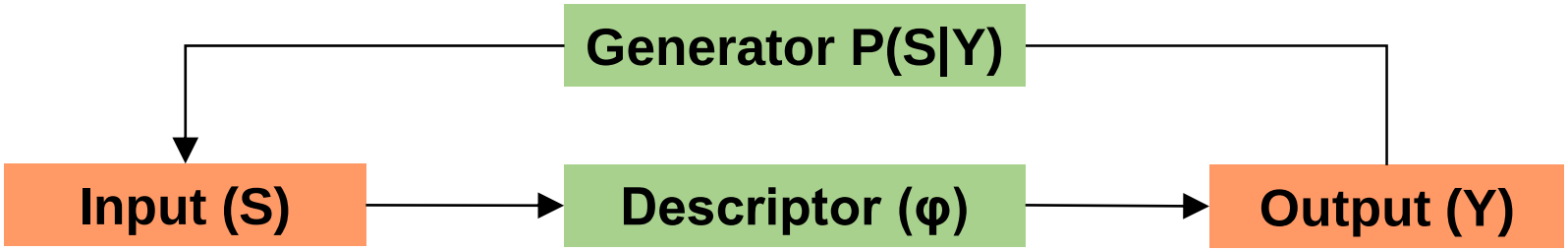
Noguchi, Y



Y-axis: LUMO
X-axis: HOMO
Blue: generated molecules
Red: molecules selected for DFT
Blue zone: final target
Yellow zone temporary target

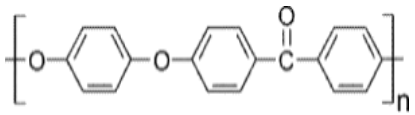
Towards New Horizons

Representation, learning and generation for various materials

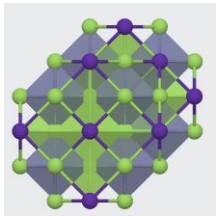


The 1st phase

Chemical structure



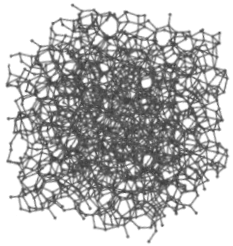
Crystal



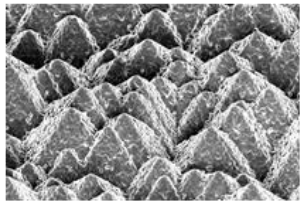
Composition

$A_aB_bC_c$

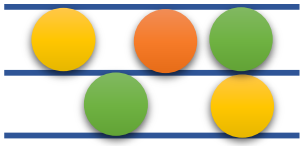
Amorphous



Microstructure



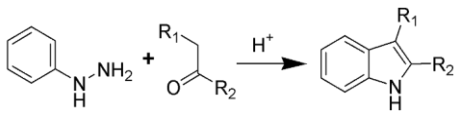
Composite / hybrid



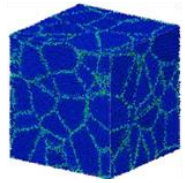
Process parameter

Θ, T

Chemical reaction



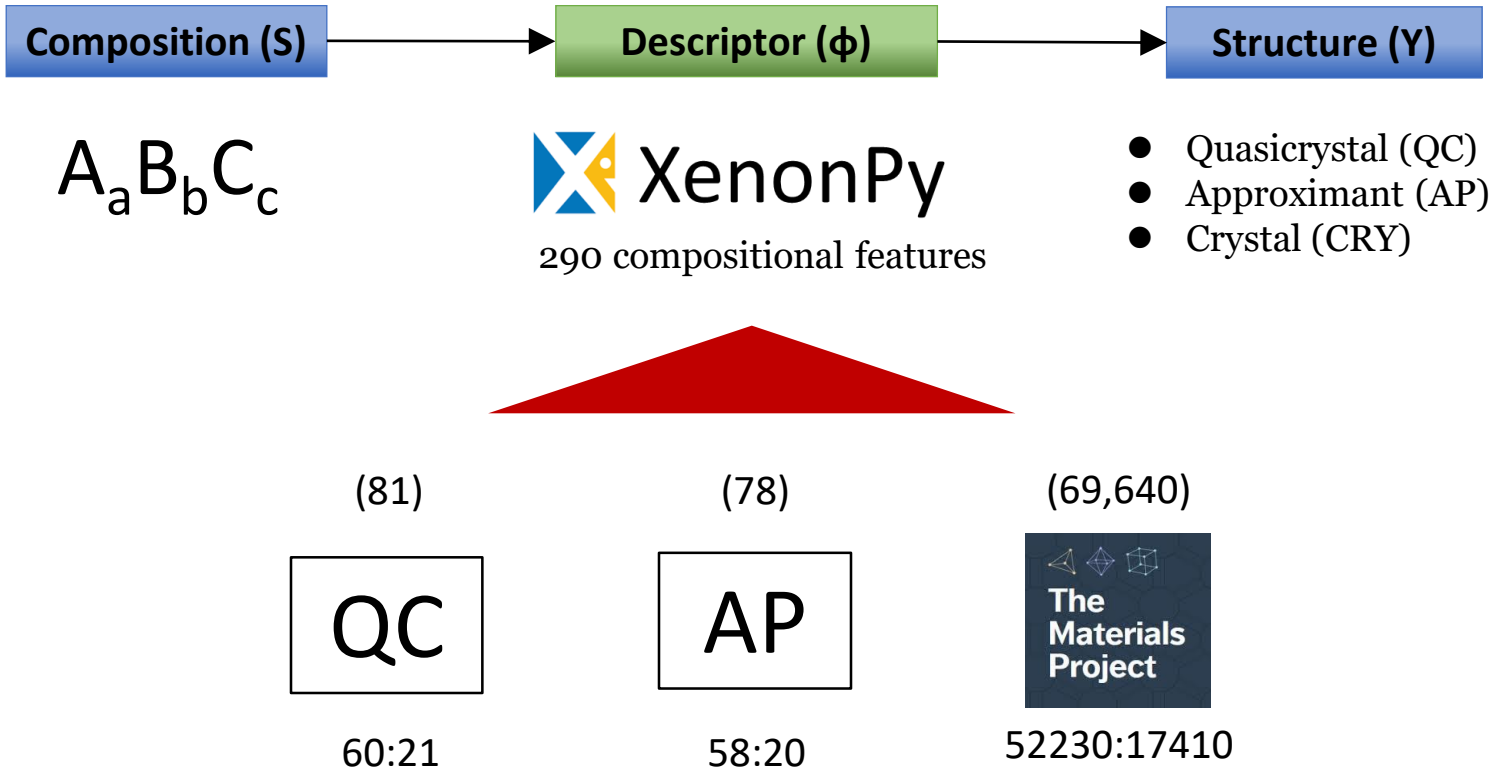
Nanostructure



Discovery of New Quasicrystals

新学術領域「ハイパーマテリアル:捕空間が創る新物質科学」(領域代表:田村隆治(東京理科大))
2019年8月~ 計画研究"ハイパーマテリアルズインフォマティクス"(代表:吉田)


Yoshida, R. Katsura, Y.
(U Tokyo)



Morikawa, J.
(Tokyo Tech)

Designing Process & Composition

Surface, morphology, and so on

Iwayama, M.[2]

Process
Composition

→

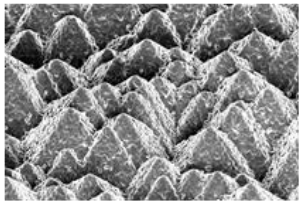
Microstructure
(e.g. SEM image)

→

Property
Function

$A_a B_b C_c D_d$

T, P, Θ



Thermoelectric property
Optical property
Energetic property
Mechanical property

[CGAN/VAE] Banko et al. Mastering processing-microstructure complexity through the prediction of thin film structure zone diagrams by generative machine learning models. arXiv (2019) arXiv:1910.09468

[CNN (VGG-19)+TL] Li et al. A transfer learning approach for microstructure reconstruction and structure-property predictions. Sci Rep. 8:13461 (2018)

[CRBM] Chang et al. Microstructure representation and reconstruction of heterogeneous materials via deep belief network for computational material design. J Mech Des. 139(7): 071404 (2017)

[GAN+BO] Yang et al. Microstructural materials design via deep adversarial learning methodology. J Mech Des 140; 10.1115/1.4041371 (2018).

[GAN] Li et al. A deep adversarial learning methodology for designing microstructural material systems. ASME 2018 (2018)

Microstructure Prediction

Banko et al. Mastering processing-microstructure complexity through the prediction of thin film structure zone diagrams by generative machine learning models. arXiv (2019) arXiv:1910.09468

Parameters

Composition: $Cr_{1-x} Al_x O_y N$

Process:

- Deposition temperature
- Deposition pressure
- Average ion energy
- Degree of ionization

Prediction & Design →

- Conditional GAN
- VAE

Microstructure

Data:

123 SEM images * 128 patches (128px * 128px)

高分子(熱物性)MIの学術創出

JST-CREST 熱制御 「高分子の熱物性マテリアルズインフォマティクス」
代表：森川淳子：2019.10-
機械学習・計算科学グループ：吉田*・林・塩見・川内・Wu ・野口・Liu


森川　淳子
(東工大)


吉田　亮
(統数研)


戸木田　雅利(東工大)


ミゼイキス
ビガンタス
(静岡大)


早川晃鏡
(東工大)

計測

- マクロ
- ミクロ
- ナノ


Morikawa, J.

高分子熱物性データベース

データ

シミュレーション

- 自動計算システム
- 分子動力学計算
- 第一原理計算


Shiomi, J.

材料創製・合成

候補提案

機械学習


Yoshida, R.


Hayashi, Y.

候補提案

- モノマー・高分子
- 複合材料
- 加工プロセス

- 実験計画法(ベイズ最適化)
- 仮想スクリーニング
- 分子・プロセス逆設計

XenonPy
https://xenonpy.readthedocs.io/en/latest/index.html

統計数理研究所 ものづくりデータ科学研究センター（2017年7月発足）

マテリアルズインフォマティクスの実践・実証

- 産学連携：共同研究・学術指導15社（2019年度） プロジェクト参画者～80名
- JSTイノベーションハブ構築支援事業「情報統合型物質・材料開発イニシアティブ」（物質・材料研究機構）
- JST-CREST 熱制御「高分子の熱物性マテリアルズインフォマティクス」（森川淳子：2019.10-）
- 新学術領域「ハイパーマテリアル：補空間が創る新物質科学」（2019.8-）
- 科研費基盤A「機械学習の先進技術による革新的機能性物質の発掘」（2019.4-）
- ISM-MCCフロンティア材料設計拠点（2019.10-）

物質・材料研究機構統合型材料開発・情報基盤部門との
MOU締結（2017.11.7-）

ISM-MCCフロンティア材料設計拠点（2019.10-）