

機器學習於材料資訊的應用

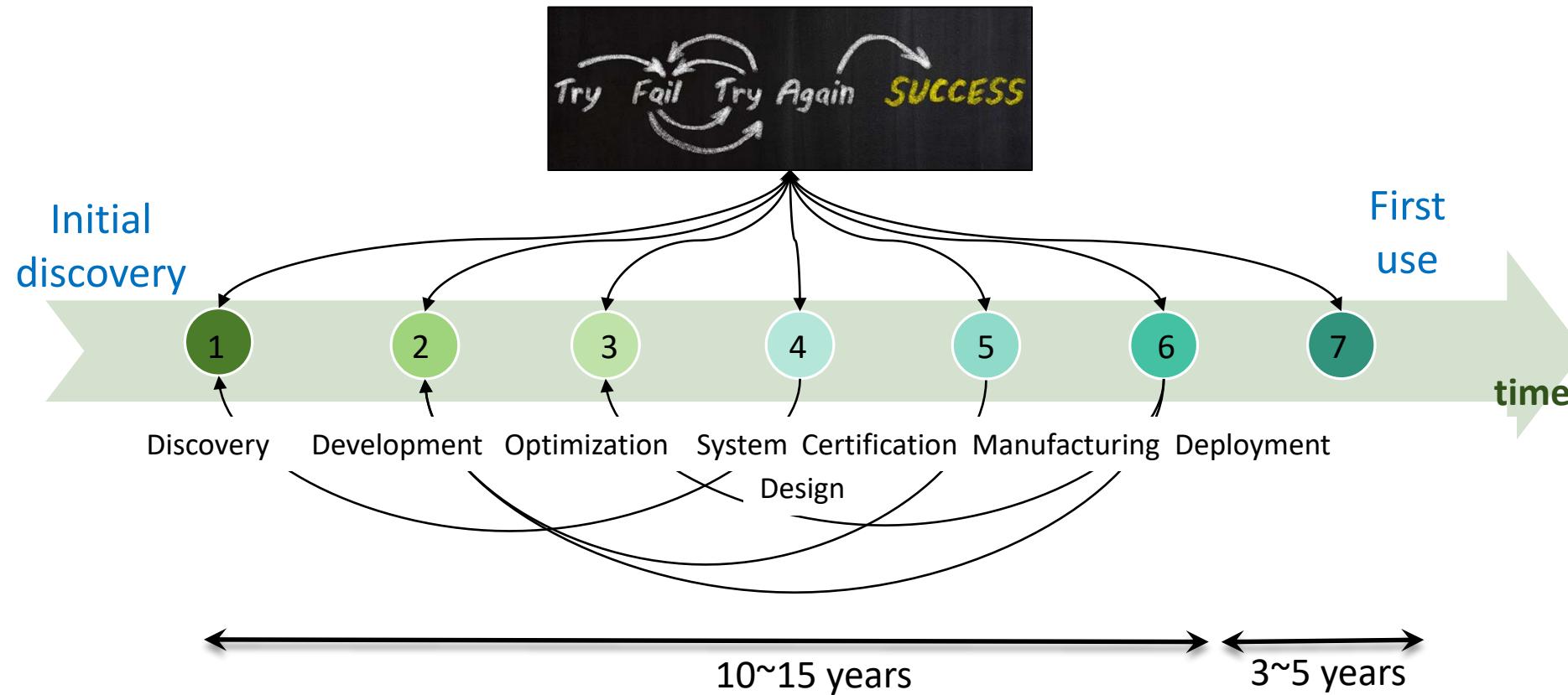
Machine Learning on Material Informatics

陳南佑(NAN-YOW CHEN)

楊安正(AN-CHENG YANG)

Streams of Materials Discovery and Development

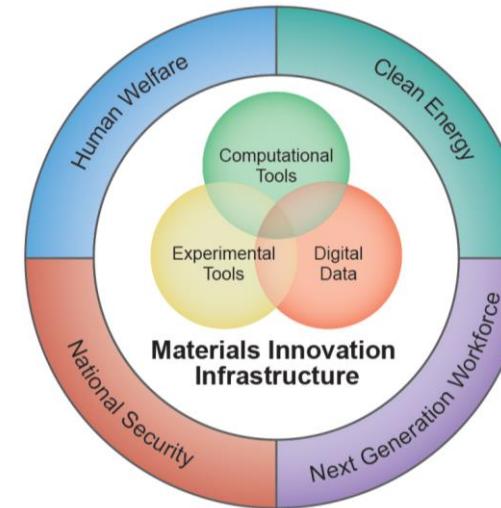
The process of finding new materials using traditional methods



- *Finding new materials is complex, expensive, and time-consuming*
- *Time to introduce new materials is much longer than that of design and prototyping.*

Advanced Manufacturing Partnership, AMP

材料基因計畫(Material Genome Initiative, MGI)加速了材料資訊(Material Informatics)的發展



To help businesses discover, develop, and deploy new materials twice as fast, we're launching what we call the **Materials Genome Initiative**.
-President Obama
Carnegie Mellon University, June 2011

About the Materials Genome Initiative

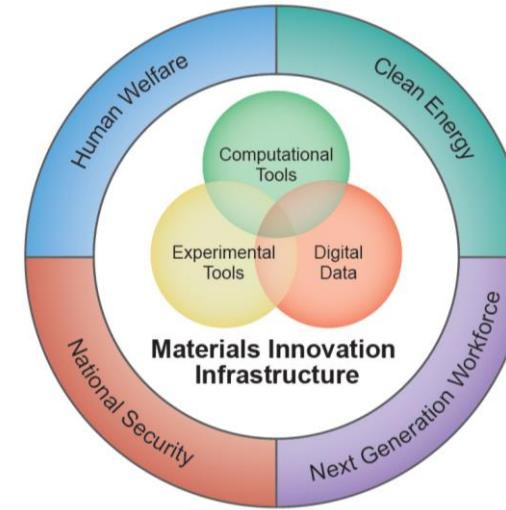
<https://www.mgi.gov/>

The **Materials Genome Initiative** is a multi-agency initiative designed to create a new era of policy, resources, and infrastructure that support U.S. institutions in the effort to discover, manufacture, and deploy advanced materials twice as fast, at a fraction of the cost.

Advanced materials are essential to economic security and human well being, with applications in industries aimed at addressing challenges in clean energy, national security, and human welfare, yet it can take 20 or more years to move a material after initial discovery to the market. Accelerating the pace of discovery and deployment of advanced material systems will therefore be crucial to achieving global competitiveness in the 21st century.

Advanced Manufacturing Partnership, AMP

材料基因計畫(Material Genome Initiative, MGI)加速了材料資訊(Material Informatics)的發展



Since the launch of MGI in 2011, the Federal government has invested over \$250 million in new R&D and innovation infrastructure to anchor the use of advanced materials in existing and emerging industrial sectors in the United States.



美國
「材料基因
組計畫」

歐盟
「歐洲冶金
復興計劃」
1億歐元

中國
成立“上海市
材料基因組
工程研究院”
材料基因工
程北京市重
點實驗室
3.2億人民幣

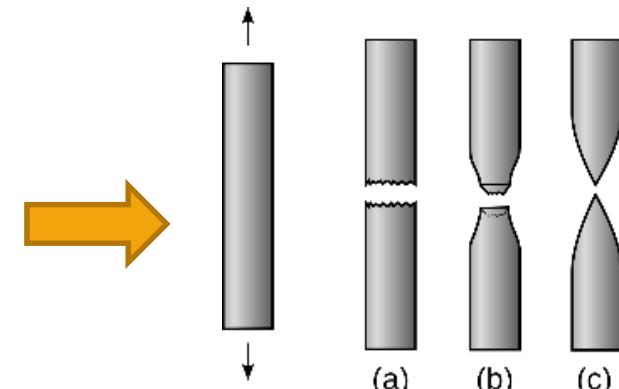
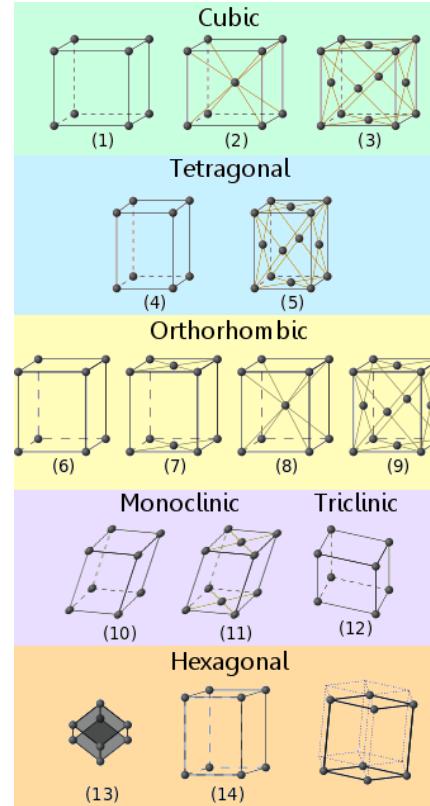
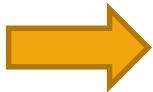
台灣
陳力俊院士
呼籲成立
MGI專案計
畫

中國
「材料基因
工程關鍵技
術與支撐平
台計畫」
3億人民幣

台灣
科技部啟動
「智慧仿生
材料與數位
設計平台」
計畫

基因材料

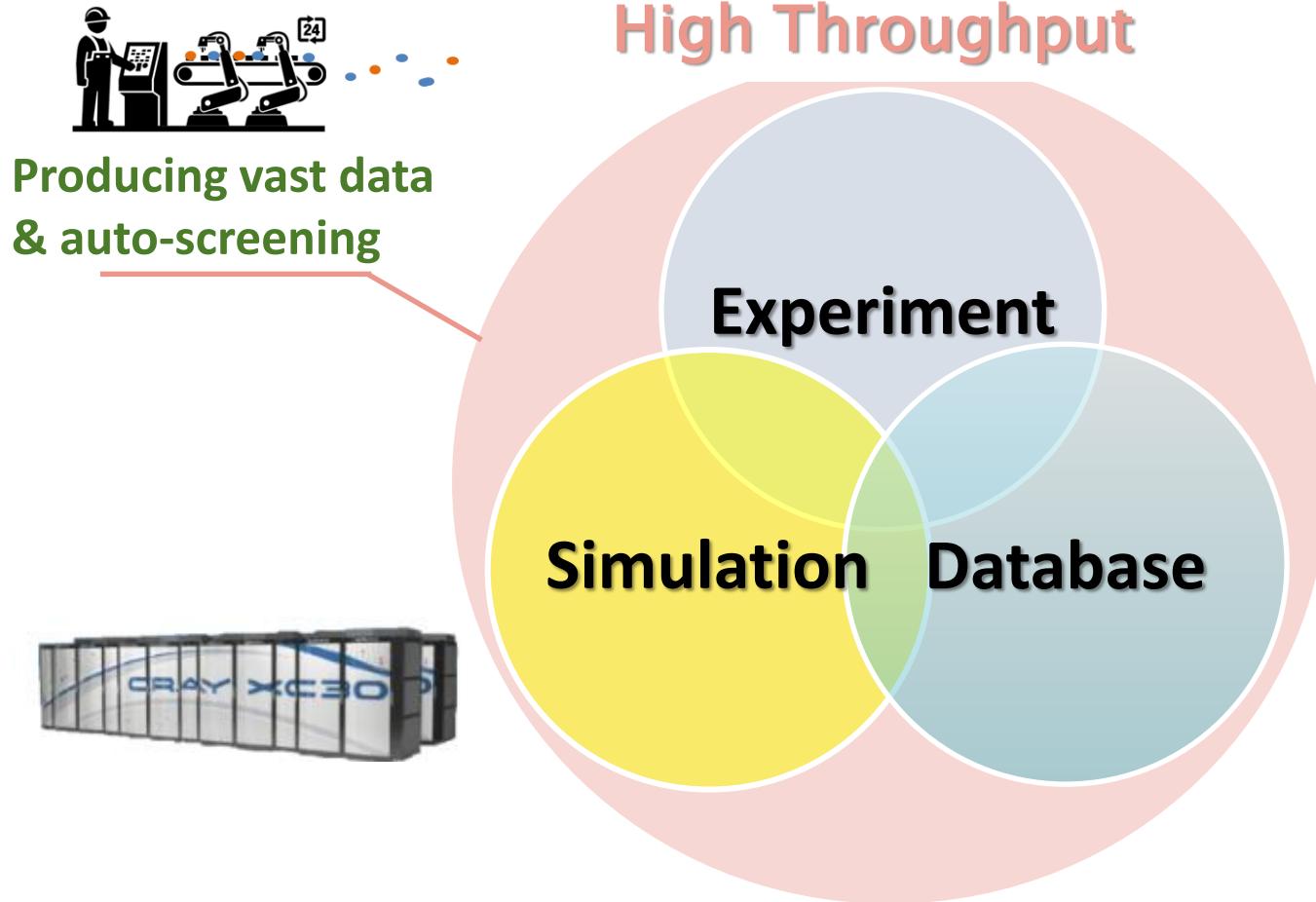
Human Genome V.S. Material Genome



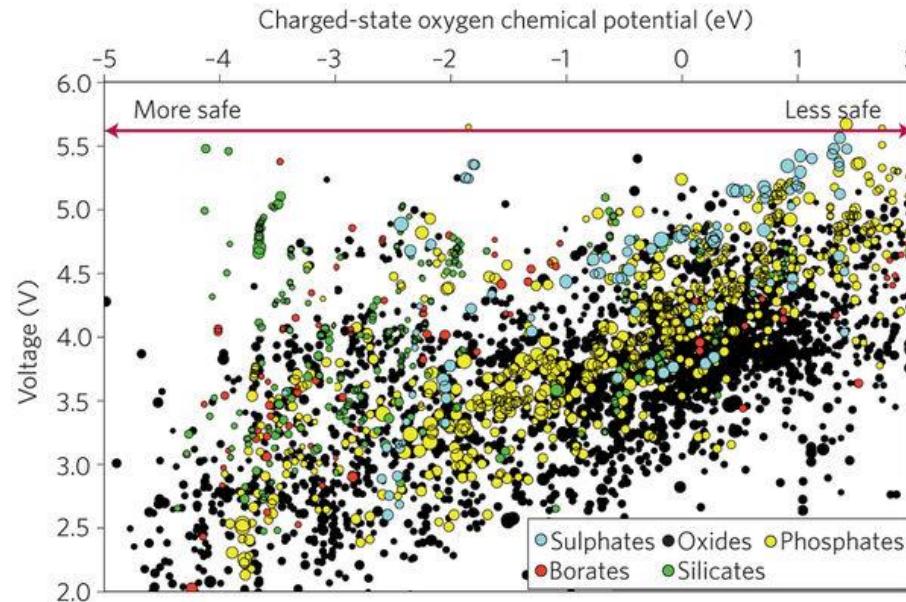
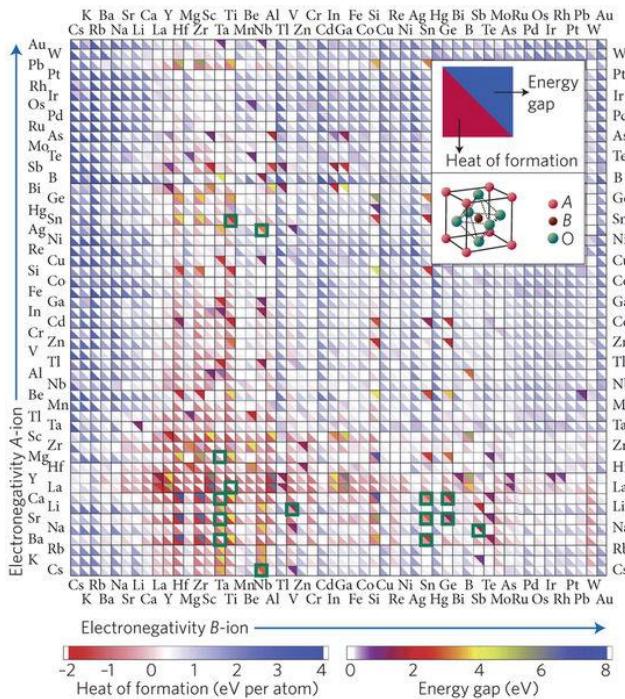
生物個體的外在表現叫作「表現型 (phenotype)」，由內在的「基因型 (genotype)」決定。改變表現型，首先要改變基因型。.

材料在巨觀尺度表現出來的材料特性是由微觀尺度下的原子組態所決定的。要改變材料特性，首先要改變微觀的原子組態。.

Materials Innovation Infrastructure

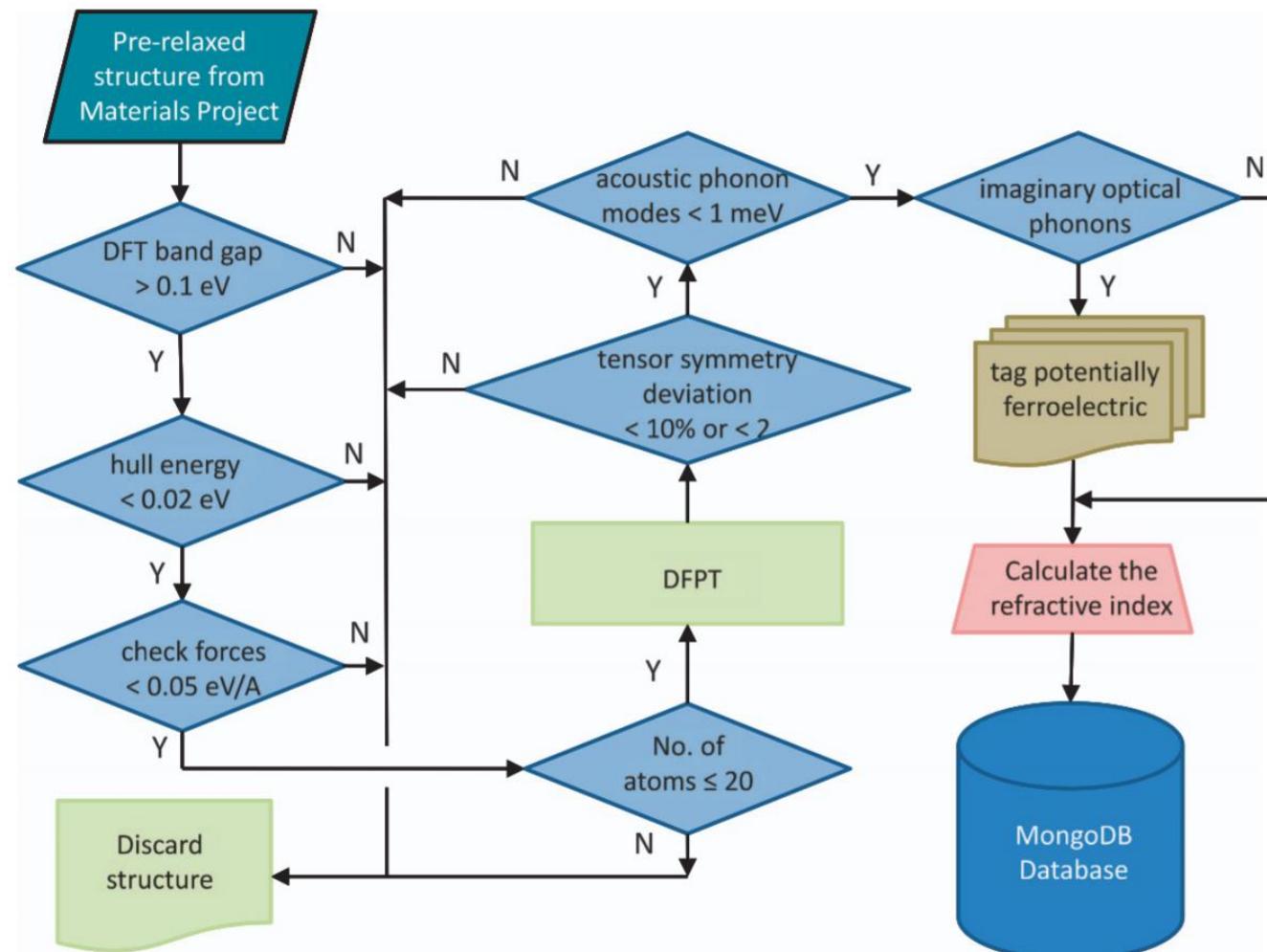


The High-Throughput Highway to Computational Materials Design



➤ *Nature Materials* **12** (2013) 191–201

Virtual Screening Process



1,056 compounds

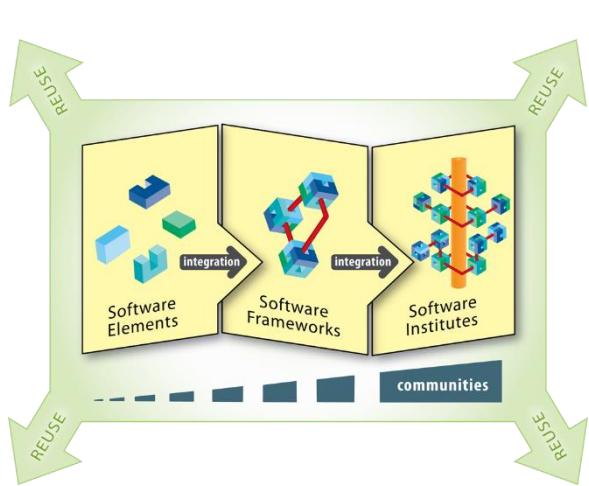
Data Descriptor: High-throughput screening of inorganic compounds for the discovery of novel dielectric and optical materials

Ioannis Petousis¹, David Mrdjenovich², Eric Ballouz³, Miao Liu⁴, Donald Winston⁴, Wei Chen^{5,6}, Tanja Graf⁶, Thomas D. Schladt⁶, Kristin A. Persson⁷ & Fritz B. Prinz^{1,3}

Figure 2. The workflow for calculating the dielectric tensor.

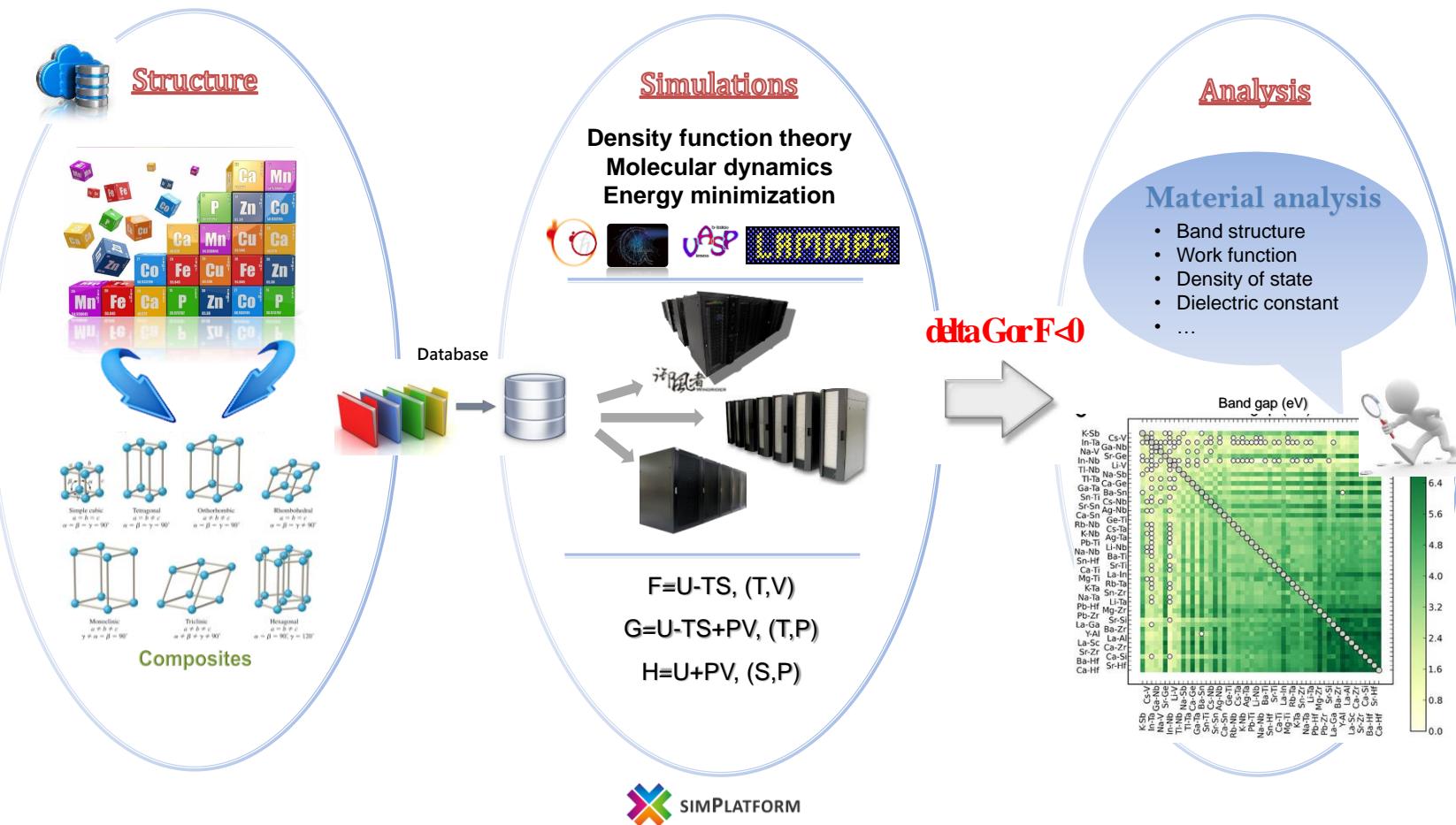
➤ *Scientific Data* 4 (2017) 160134

Demand for new materials

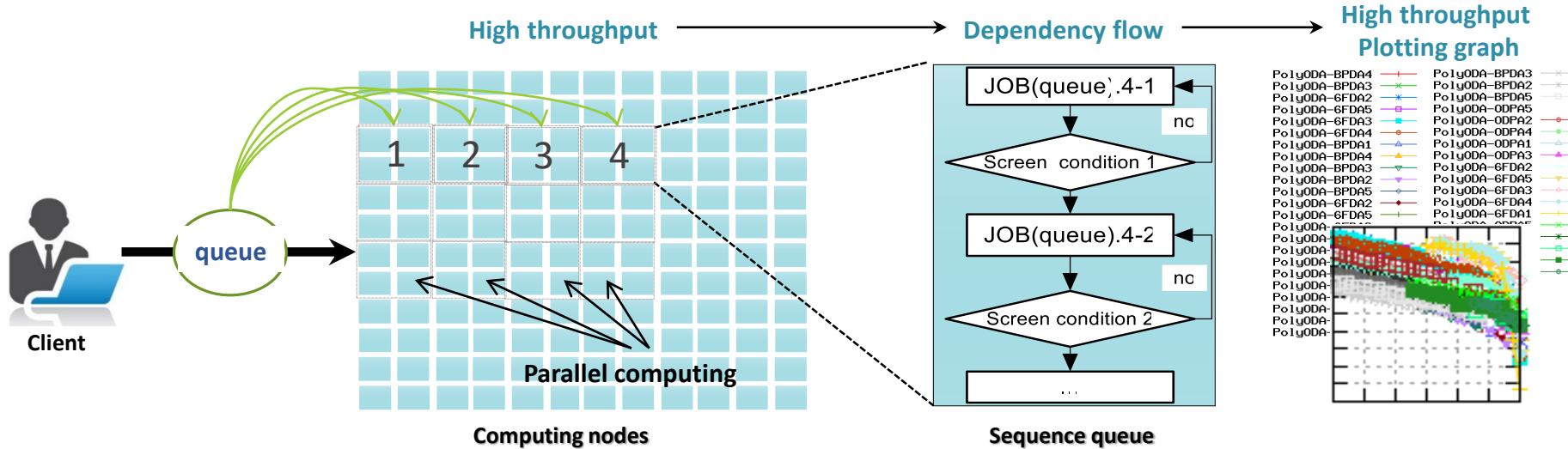


- 整合與開發不同尺度/領域之模擬與實驗計算工具
- 結合資料庫已知之數據與參數特性
- 發展可重複使用之高通量計算流程平台

Auto Simulation Flow

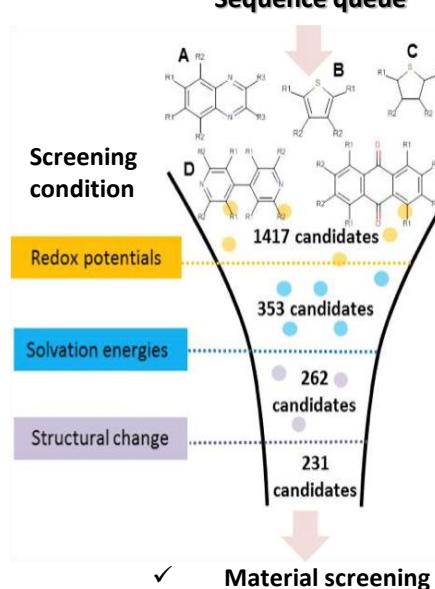


High Throughput Computing Workflow Framework



Advantages

- High throughput on parallel jobs
- Job dependency
- Condition screening
- Software transferable
- Flow extendable
- Allow hybrid program



Benefits

- Produce massive tasks/data
- Easy to customize personal flow
- Efficient (Removes launch speed bottleneck & script editing (x))
- Avoiding negligence

Materials Innovation Infrastructure

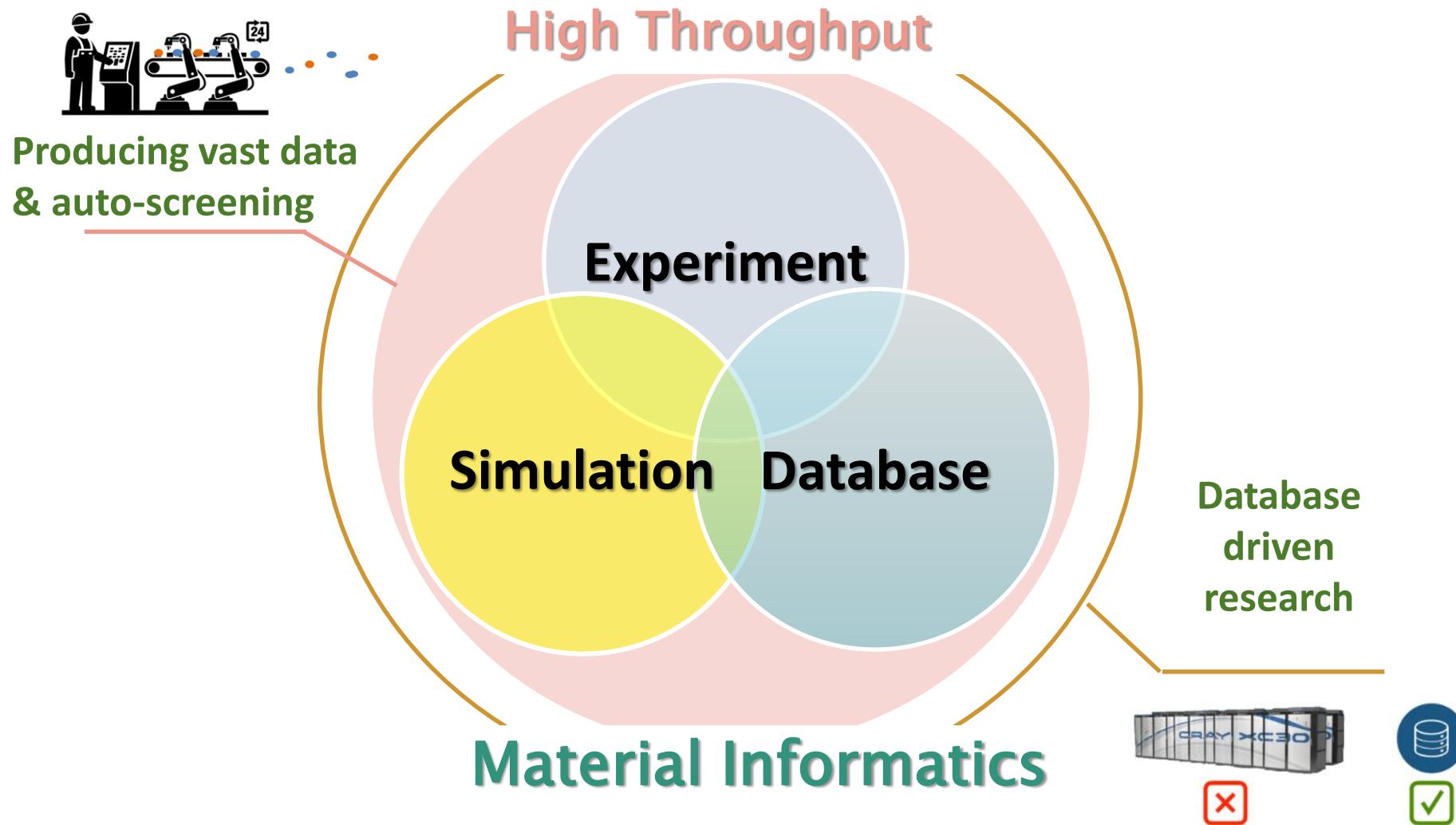


Table I. A list of some notable materials-data resources.

Name	URL	Category	Free/Non-Free
3D Materials Atlas	cosmicweb.mse.iastate.edu/wiki/display/home/Materials+Atlas+Home	3D Characterization	Free
AFLIB	aflowlib.org	Computational	Free
AIST Research Information Databases	www.aist.go.jp/aist_e/list/database/riodb	General Materials Data	Free

Table I. A list of some notable materials-data resources.

Name	URL	Category	Free/Non-Free
CrystMet	cds.dl.ac.uk/cgi-bin/news/disp?crystmet	Crystallography	Non-Free
Crystallography Open Database	http://www.crystallography.net	Crystallography	Free
DOE Hydrogen Storage Materials Database	www.hydrogenmaterialssearch.govtools.us	Hydrogen Storage	Free
Granta CES Selector	www.grantadesign.com/products/ces	General Materials Data	Non-Free
Handbook of Optical Constants of Solids, Palik	N/A	Hard-Copy Sources	Non-Free
Harvard Clean Energy Project	cepdb.molecularspace.org	Computational	Free
Inorganic Crystal Structure Database	cds.dl.ac.uk/cds/datasets/crys/icsd/llicsd.html	Crystallography	Non-Free
International Glass Database System	www.newg.org	General Materials Data	Non-Free

Table I. A list of some notable materials-data resources.

Name	URL	Category	Free/Non-Free
Pauling File	paulingfile.com	General Materials Data	Non-Free
Pearson's Handbook: Crystallographic Data	N/A	Hard-Copy Sources	Non-Free
Powder Diffraction File (PDF)	www.icdd.com/products/index.htm	Crystallography	Non-Free
PubChem	pubchem.ncbi.nlm.nih.gov	Chemical data	Free
Reaxys	www.elsevier.com/solutions/reaxys	Chemical data	Non-Free
Scifinder/ChemAbstracts	scifinder.cas.org	Chemical data	Non-Free
SciGlass	www.sciglass.info	Glass	Non-Free
SpringerMaterials	materials.springer.com	General Materials Data	Non-Free
Metallurgical Thermochemistry, Kubaschewski	N/A	Hard-Copy Sources	Non-Free
TEDesignLab	www.tedesignlab.org	Thermoelectrics	Free
Total Materia	www.totalmateria.com	General Materials Data	Non-Free
UCSB-MRL thermoelectric database	www.mrl.ucsb.edu:8080/datamine/thermoelectric.jsp	Thermoelectrics	Free

Note: AFLIB, Automatic-FLOW for Materials Discovery; AIST, National Institute of Advanced Industrial Science and Technology (Japan); ASM, American Society for Metals; CALPHAD, CALCULATION of PHase Diagrams; SGTE, Scientific Group ThermoData Europe; CINDAS, Center for Information and Numerical Data Analysis and Synthesis; CRC, Chemical Rubber Company; DOE, US Department of Energy; CES, Cambridge Engineering Selector; NIMS, National Institute for Materials Science; NIST, National Institute of Standards and Technology; KIM, Knowledge Database of Interatomic Models; UCSB MRL, University of California, Santa Barbara Materials Research Laboratory.

Database

The screenshot shows the Materials Project website. At the top, there's a navigation bar with links for Home, About, Apps, Documentation, API, and Login. Below the navigation is a large banner with the text "Harnessing the power of supercomputing and state of the art electronic structure methods, the Materials Project provides open web-based access to computed information on known and predicted materials as well as powerful analysis tools to inspire and design novel materials." It features three small molecular models (icosahedron, cube, and tetrahedron) and a "Learn more" button. Below the banner is a section titled "Explore Materials" with a plot of Fermi energy (E_Fermi) versus band index (k). To the right are "Database Statistics" showing counts for various material types: INORGANIC COMPOUNDS (69,640), BANDSTRUCTURES (53,648), MOLECULES (21,954), NANOPOROUS MATERIALS (530,243), ELASTIC TENSORS (6,073), PIEZOELECTRIC TENSORS (941), INTERCALATION ELECTRODES (3,628), and CONVERSION ELECTRODES (16,128). A central feature is a 3D ball-and-stick model of a crystal lattice.



User-friendly web interface
(but unfriendly for advanced
users requiring large quantities
of data!)

A screenshot of the Materials Explorer interface. It shows a sidebar with various search and filter options like "Explore Materials", "Explore Batteries", "Visualize Stability", "Invent Structures", and "Calculate". The main area displays a complex network graph of elements and their interactions. A callout box highlights several tools: Materials Explorer, Battery Explorer, Crystal Toolkit, Structure Predictor, Phase Diagram App, Pourbaix Diagram App, and Reaction Calculator.

A screenshot of the Electronic Structure interface. It shows plots of energy levels (E) versus wave vector (k) for different materials. Callout boxes highlight Energetic properties, Structure, Electronic Structure, XRD, and Elastic properties.

Development of the Materials Project was funded by the U.S. Department of Energy (DOE) Office of Energy Efficiency and Renewable Energy (EERE) and Advanced Scientific Computing Research (ASCR) programs, and through its Office of Energy Efficiency and Renewable Energy (EERE), via the Battery Materials Research (BMR, formerly BATT) program. A notable source of support within DOE-BES is the Joint Center for Energy Storage Research (JCESR).

The Materials Project is also supported by a Laboratory Directed Research and Development grant from LBNL. Disseminated science is supported by DOE (BES and BMR), the National Science Foundation (NSF), Gillette, Volkswagen, Umicore, and Bosch.

Partners include the Lawrence Berkeley National Laboratory, MIT, NERSC, University of California Berkeley, NUS, University of Wisconsin, Duke University, UK, and UCL. Key contributors listed include Kristin Persson, Anubhav Jain, Daniel Gunter, Maciej Haranczyk, Wei Chen, Dane Morgan, Gerbrand Ceder, Shreyas Cholia, Jack Deslippe, David Skinner, Alan Dozier, Raphael Fink, Stefan Adams, and Geoffroy Hautier.

Database



A screenshot of a computer desktop showing the AFLOW distributed materials property repository. The desktop has two monitors. The left monitor displays the AFLOW homepage, featuring a molecular network background, navigation links for HOME, CONSORTIUM, PUBLICATIONS, FORUM, and SEARCH, and a welcome message: "Welcome to the AFLOW distributed materials property repository: share with us your passion for innovation and technology." It also mentions a database of 1,729,101 material compounds with over 171,180,999 calculated properties. The right monitor shows a detailed view of the search interface and prediction results for the material Cr3 Si6.

The search interface includes a search bar for "Aflow Unique IDentifier or advanced search", a dropdown for "Type" set to "POSCAR" (highlighted in blue), and a "Quantum Espresso" tab. The search results for Cr3 Si6 show several entries with numerical values. A "RUN PREDICTION" button is visible at the bottom right of the search interface.

The prediction results for Cr3 Si6 are displayed on the right monitor. The material is identified as an "Insulator". Key properties listed include:

Band Gap	Energy
0.92 eV	-5.76 eV/cell
Bulk Modulus	Shear Modulus
178.54 GPa	140.12 GPa
Heat Capacity C_p (cell)	Heat Capacity C_p (atom)
23.36 J/g-cell	2.33 J/g/atom
Heat Capacity C_v (cell)	Heat Capacity C_v (atom)

Logos for Duke University, Department of Energy, NSF, Homeland Security, CRAY, and UNT are visible at the bottom of the screen.

Database

The image shows the Kazana 1.0 logo on the left, featuring a stylized 'K' composed of orange and yellow dots. To the right of the logo is the text "A Computational Materials Knowledgebase" in white. Below this, a subtitle reads "A platform to store structure and property data created by atomistic simulations, and tools to design materials by learning from the data." At the bottom of the slide, there is a navigation bar with links: "Polymer Genome", "Data Repository", "Graphical Visualization", "wjlee@nchc.narl.org.tw | Out".



Khazana

A platform to store structure and property data created by atomistic simulations, and tools to design materials by learning from the data.

Polymer Genome Data Repository Graphical Visualization wjlee@nchc.narl.org.tw |

Explore the Polymer Genome polyimide Experimental DFT Advanced Options

17 experimental results for 'polyimide' / (ML prediction)

Formula Unit	Repeat Unit SMILES	Polymer Name or Class	Band Gap (eV)	Dielectric Constant	Refractive Index	Electron Affinity (eV)	Ionization Energy (eV)	DFT Result
C ₈ H ₅ N ₂ O ₂ (ID: E0002)	CO-NH-CO-C ₆ H ₄ -C(=O)NC(=O)C(C=C1)=CC=C1	Polyimide	4 (2.92)	4.80 (4.98)	(1.93)	(3.36)	(6.86)	0344 0347 0348
C ₂₀ H ₁₂ N ₂ O ₅ (ID: E0004)	N(CO)C ₂ C ₆ H ₃ -CO-C ₆ H ₃ (CO)2N-CH ₂ -CH ₂ -CH ₂ -N(C(=O)C1=CC2(=O)(=O)C1=CC2C(=O)(=O)C=C3C(=O)(=O)C=C3C(=O)N4CCC4	Polyimide	3.79 (2.88)	4.01 (5.14)	(1.95)	(3.91)	(7.32)	N/A
C ₂₃ H ₁₈ N ₂ O ₅ (ID: E0005)	N(CO)C ₂ C ₆ H ₃ -CO-C ₆ H ₃ (CO)2N-CH ₂ -CH ₂ -CH ₂ -CH ₂ -N(C(=O)C1=CC2(=O)(=O)C1=CC2C(=O)(=O)C=C3C(=O)(=O)C=C3C(=O)N4CCCCC4	Polyimide	3.42 (3.46)	3.57 (4.66)	(1.86)	(3.45)	(6.97)	3200
C ₃₀ H ₃₂ N ₂ O ₉ (ID: E0009)	N(CO)C ₂ C ₆ H ₃ -CO-C ₆ H ₃ (CO)2N-HK511	Polyimide	3.48 (4.57)	7.80 (4.1)	(1.74)	(2.5)	(6.83)	3201

Material Contents

Classified and tagged materials information obtained by quantum mechanical calculations that numerically and visually demonstrate the property. Provide simulation methods in detail.

3350 Materials in 4 classes

Class	Percentage	Count
Crystalline polymers	27%	897
Polymers(experimental)	2%	68
Molecular crystals	8%	253
Inorganic materials	64%	2132

Element Contents

Search the targeting materials or make a prediction via composition based searching interface with the elements of your interests.

13284 Species in 3319 materials

Element	Percentage	Count
H	19%	2527
C	17%	2315
O	12%	1570
N	13%	1702
Sn	5%	649
Ge	4%	474
Pb	4%	477
F	4%	470
Cl	3%	384
I	3%	384
Br	3%	367
S	2%	367
Mn	2%	321
La	2%	321
Cd	1%	304
Zn	1%	304
Hf	1%	304
Other	6%	860

Property Contents

Uniformly curated material properties aided by computational materials science. User controllable and customizable list of field for property make the search result easier to read.

Property	Percentage
Band gap(HSE)	12%
Band gap	11%
Spontaneous polarization	0%
Bulk modulus	1%
Exp. band gap	1%
Breakdown Field	1%
Ionization energy	6%
Formation energy	6%
Density	5%
Ionization potential	2%
Electron affinity	2%
Phonon cutoff	1%
Spontaneous polarization	0%

Database

The NOMAD Laboratory, A European Centre of Excellence, maintains the largest Repository for input and output files of all important computational materials science codes. From its open-access data, it builds several Big-Data Services helping to advance materials science and engineering.

To learn more, click on the summary on the **NOMAD** homepage.

NOMAD Scope and Overview

Data is a crucial raw material

Surprisingly, extreme-scale aspects of science and engineering, one reason being that standard hardware and software calculations is wasted without scale computing challenge added.

Element **Formula/Material** **Properties** **AND** **OR** **NOT** **[]**

NOMAD Success Stories

Thermal-barrier coatings fuel-efficiency improvement over the last 30 years

NOMAD Archive

The NOMAD Archive provides the open access data from the repository in a code-independent, normalized representation.

The NOMAD Archive stores in a code-independent format calculations performed with all the most important and widely used electronic-structure and force-field codes.

A summary statistics of the Archive content, update to Aug 11, 2017

44,179,006	39,402,326	220,918	4,517,016
Total-Energy Calculations	Bulk Crystals	Surfaces	Molecules/Clusters

40,481,615

Different Geometries

- 9,274 Bio Archives for parsing
- Data extracted with parsing
- Data classified using 168 parameters
- Number of parsed quantities

Below, we show a breakdown of the Archive are displayed.

The plot is interactive. You can do the following:

- Type
- Total-Energy Calculations

Add chemical elements to your query.

NOMAD REPOSITORY

MATERIALS ENCYCLOPEDIA

HPC EXPERTISE & HARDWARE

BIG-DATA ANALYTICS

ADVANCED GRAPHICS

Database

NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

NIST Materials Genome Initiative
Gateway to Materials Genome Information

Material Measurement Laboratory

HOME JARVIS-DFT JARVIS-FF DOCUMENTATION OTHER RESOURCES CONTACT

JARVIS (Joint Automated Repository for Various Integrated Simulations) is a repository designed to automate materials discovery using classical force-field, density functional theory, machine learning calculations and experiments.

The Force-field section of JARVIS (JARVIS-FF) consists of thousands of automated LAMMPS based force-field calculations on DFT geometries. Some of the properties included in JARVIS-FF are energetics, elastic constants, surface energies, defect formations energies and phonon frequencies of materials.

The Density functional theory section of JARVIS (JARVIS-DFT) consists of thousands of VASP based calculations for 3D-bulk, single layer (2D), nanowire (1D) and molecular (0D) systems. Most of the calculations are carried out with optB88vDW functional. JARVIS-DFT includes materials data such as: energetics, diffraction pattern, radial distribution function, band-structure and gamma-point phonons.

The Machine-learning section of JARVIS, energetics, heat of formation, GGA/ME

[Presentation slides for brief overview](#)

Web-search

JARVIS

Properties (scaled)

JARVIS-ID	Formula	Space-group	Functional	Calculation type	optB88vDW-bandgap	MBJ-bandgap	HSE bandgap	b_v (GPa)	g_v (GPa)	Formation energy p
JVASP-8118	SiC	P6_3mc	optB88-vDW	Bulk	2.62	3.43	na	213.533	190.24	-0.239
JVASP-107	SiC	P6_3mc	optB88-vDW	Bulk	2.49	3.23	na	213.344	190.627	-0.244
JVASP-22643	SiC	P3m1	optB88-vDW	Bulk	2.07	2.85	na	213.3	193.02	-0.244
JVASP-10867	SiC	R3m	optB88-vDW	Bulk	2.25	na	na	213.3	190.96	-0.244
JVASP-182	SiC	P6_3mc	optB88-vDW	Bulk	2.3	3.08	na	213.278	190.727	-0.245
JVASP-8158	SiC	F-43m	optB88-vDW	Bulk	1.62	2.31	na	212.767	197.62	-0.245
JVASP-22633	SiC	P6_3mc	optB88-vDW	Bulk	2.06	2.82	na	213.2	190.76	-0.245
JVASP-22644	SiC	P3m1	optB88-vDW	Bulk	2.27	3.02	na	213.289	190.653	-0.245

<http://www.hitwebcounter.com/htmltutorial.php>

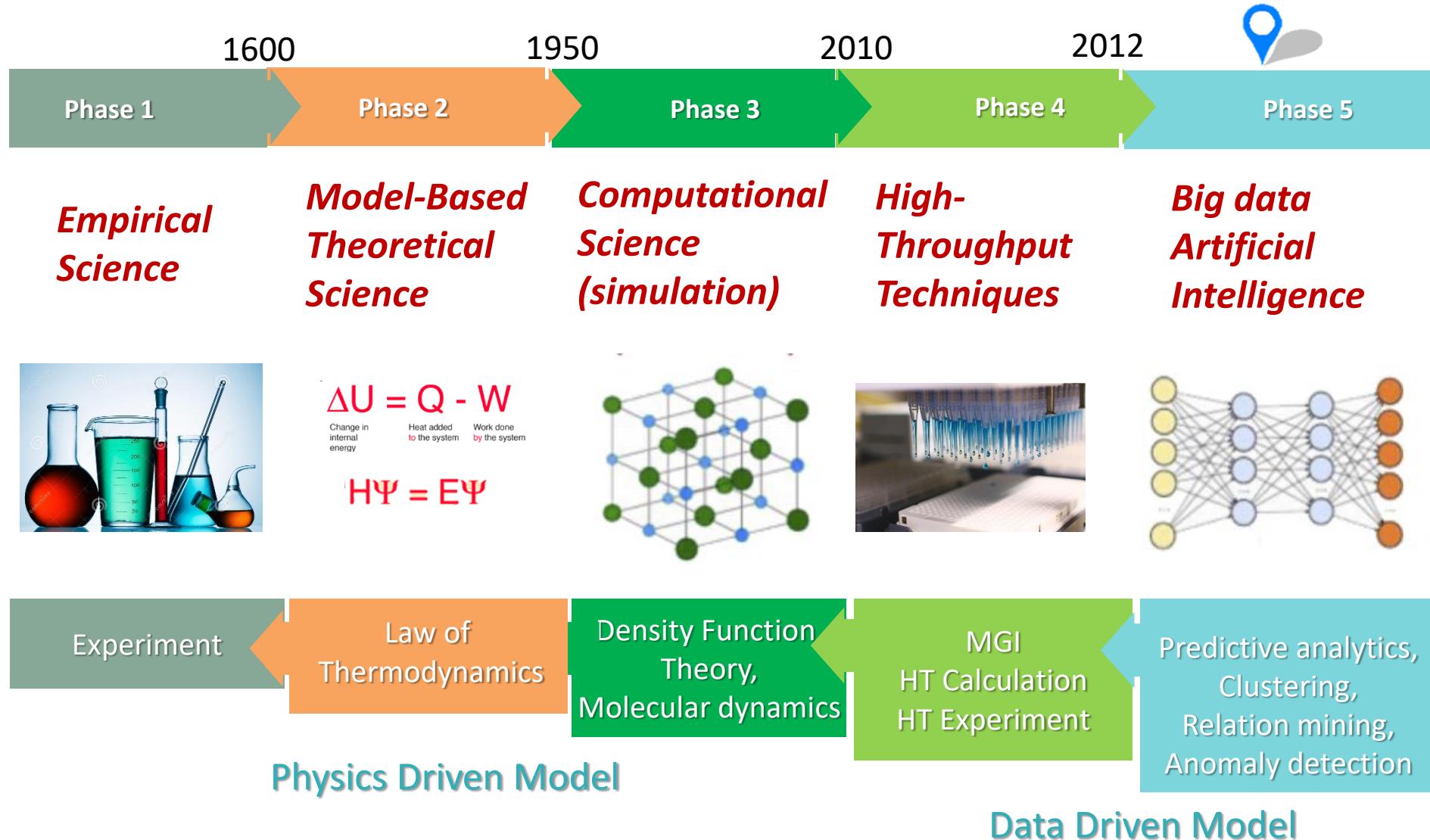
Please send comments and suggestions to kamal.choudhary@nist.gov.

- Formation energy/atom
- Relaxed energy/atom
- Bandgap
- Primitive cell lattice parameters
- Cell lattice
- X-ray diffraction (XRD) pattern
- Radial distribution function (RDF)
- Electrostatic potential
- Optical properties Semi-local
- Optical properties METAGGA-MBJ
- Elastic tensor and derived phonon properties
- Thermoelectric properties
- Magnetic moment

JARVIS Force-fields

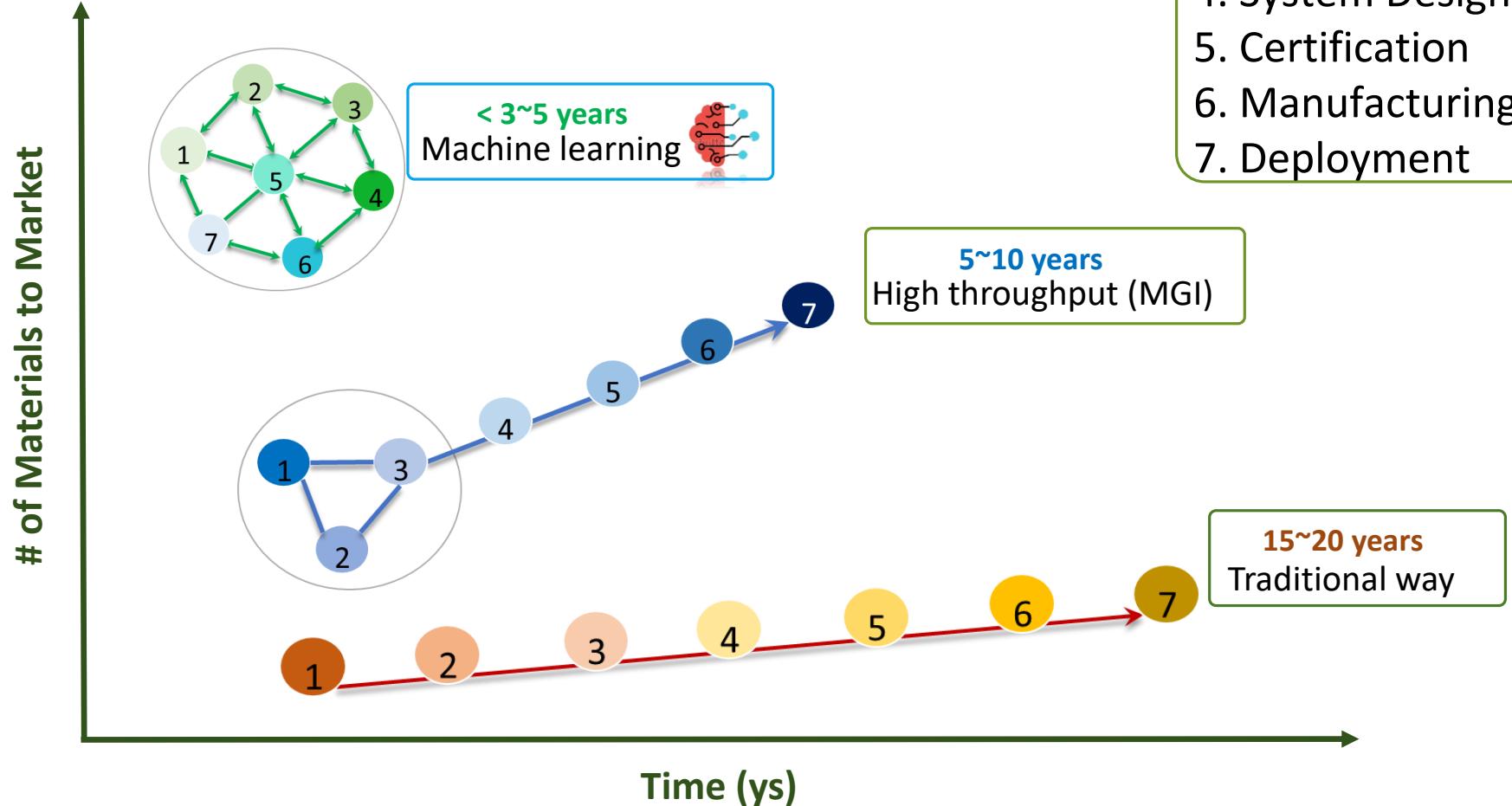
JARVIS-ID	MPID	Formula	Energy/atom (eV)	c_{11} (GPa)	c_{22} (GPa)	c_{33} (GPa)	c_{12} (GPa)	c_{13} (GPa)	c_{23} (GPa)	c_{44} (GPa)	c_{55} (GPa)	c_{66} (GPa)	b_v (GPa)	g_v (GPa)	Formation energy_hull (eV)	Force-field
Calo-1794	imp-11713	SiC5	-6.16	523.6	523.6	56.3	53.1	53.1	191.8	191.8	213.9	224.2	213.6	0.0	SiC_1989 tersoff	
Calo-1856	imp-11713	SiC5	-6.43	506.9	506.9	534.4	121.8	94.2	176.2	176.2	192.5	241.0	191.5	0.0	SiC_1994 tersoff	
Calo-1918	imp-11713	SiC5	-6.34	485.9	485.9	543.9	122.3	64.4	64.4	159.7	159.7	181.8	224.2	184.5	0.0	SiC_Ehrt-Albe tersoff
Calo-2763	imp-11713	SiC5	-6.16	523.5	523.5	56.2	95.8	53.1	191.7	191.7	213.8	224.1	213.5	0.0	SiC tersoff	
Calo-2890	imp-11713	SiC5	-6.43	467.5	467.5	497.9	98.1	87.7	87.7	167.2	167.2	184.7	211.1	183.8	0.0	SiC mean
Calo-3252	imp-11713	SiC5	-6.21	483.7	483.7	510.2	118.1	91.5	91.5	167.1	167.1	182.8	231.1	181.8	0.0	SiC_1990 tersoff
Calo-1793	imp-11714	SiC4	-6.16	523.2	523.2	56.3	96.2	53.1	191.8	191.8	213.5	224.2	213.3	0.0	SiC_1989 tersoff	
Calo-1855	imp-11714	SiC4	-6.43	506.6	506.6	534.4	122.1	94.2	94.2	176.2	176.2	192.2	241.0	191.4	0.0	SiC_1994 tersoff
Calo-1917	imp-11714	SiC4	-6.34	485.5	485.5	543.9	122.8	84.4	84.4	159.7	159.7	181.4	224.2	184.4	0.0	SiC_Ehrt-Albe tersoff
Calo-2762	imp-11714	SiC4	-6.16	523.1	523.1	56.2	96.2	53.1	191.7	191.7	213.5	224.1	213.4	0.0	SiC tersoff	
Calo-2889	imp-11714	SiC4	-6.43	467.2	467.2	497.9	98.4	87.7	87.7	167.2	167.2	184.4	211.1	183.7	0.0	SiC mean
Calo-2351	imp-11714	SiC4	-6.21	483.4	483.4	510.2	118.4	91.5	91.5	167.1	167.1	182.5	231.1	181.7	0.0	SiC_1990 tersoff
Calo-1795	imp-56705	CS98	-6.16	523.1	523.1	56.3	96.3	53.1	191.8	191.8	213.4	224.2	213.4	0.0	SiC_1989 tersoff	
Calo-1857	imp-56705	CS98	-6.43	506.8	506.8	534.4	121.9	94.2	94.2	176.2	176.2	192.5	241.0	191.5	0.0	SiC_1994 tersoff
Calo-1919	imp-56705	CS98	-6.34	484.6	484.6	542.9	123.6	84.4	84.4	159.7	159.7	180.5	224.2	184.0	0.0	SiC_Ehrt-Albe tersoff
Calo-2764	imp-56705	CS98	-6.16	523.0	523.0	56.2	96.3	53.1	191.7	191.7	213.3	224.1	213.3	0.0	SiC tersoff	
Calo-2891	imp-56705	CS98	-6.43	467.2	467.2	497.9	98.4	87.7	87.7	167.2	167.2	184.4	211.1	183.0	0.0	SiC mean
Calo-2553	imp-56705	CS98	-6.21	483.6	483.6	510.2	118.6	91.5	91.5	167.1	167.1	182.7	231.1	181.8	0.0	SiC_1990 tersoff
Calo-1796	imp-56705	CS98	-6.16	523.1	523.1	56.3	96.3	53.1	191.8	191.8	213.4	224.2	213.4	0.0	SiC_1989 tersoff	

Paradigm Shift in Science



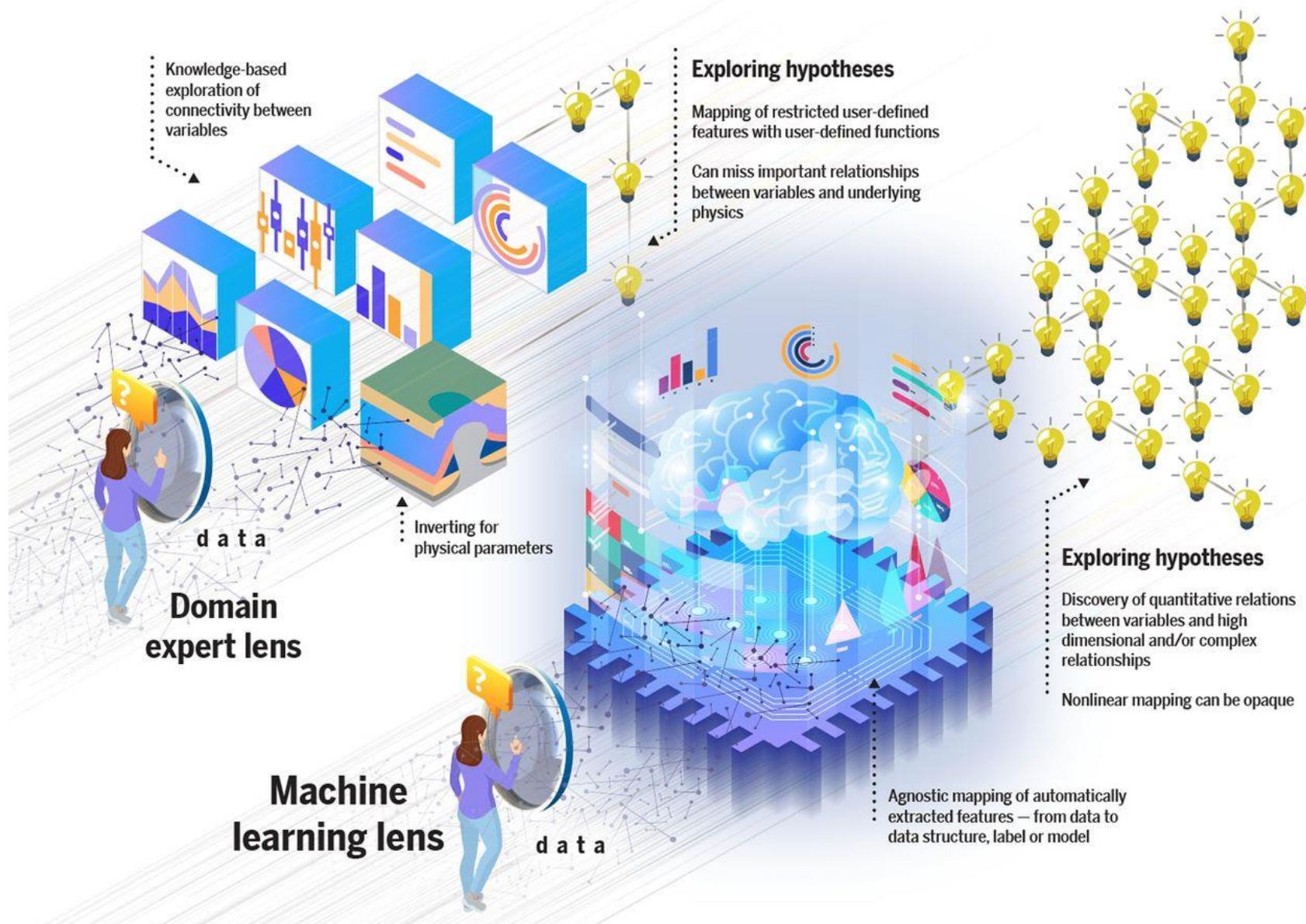
AI加速新材料的開發

QuesTek's
Ferrium® M54® steel

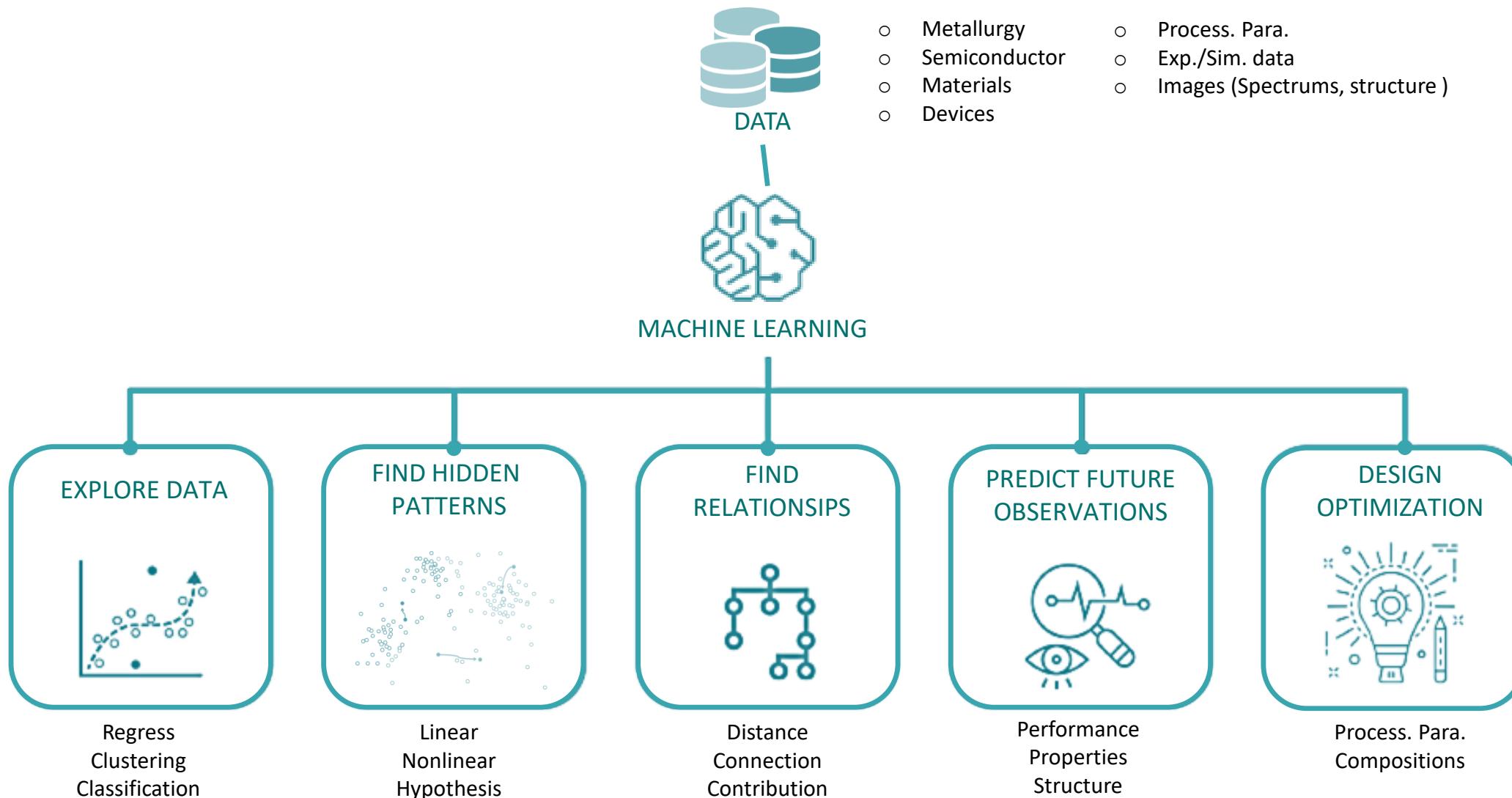


1. Discovery
2. Development
3. Optimization
4. System Design
5. Certification
6. Manufacturing
7. Deployment

Tradition method vs. Machine Learning

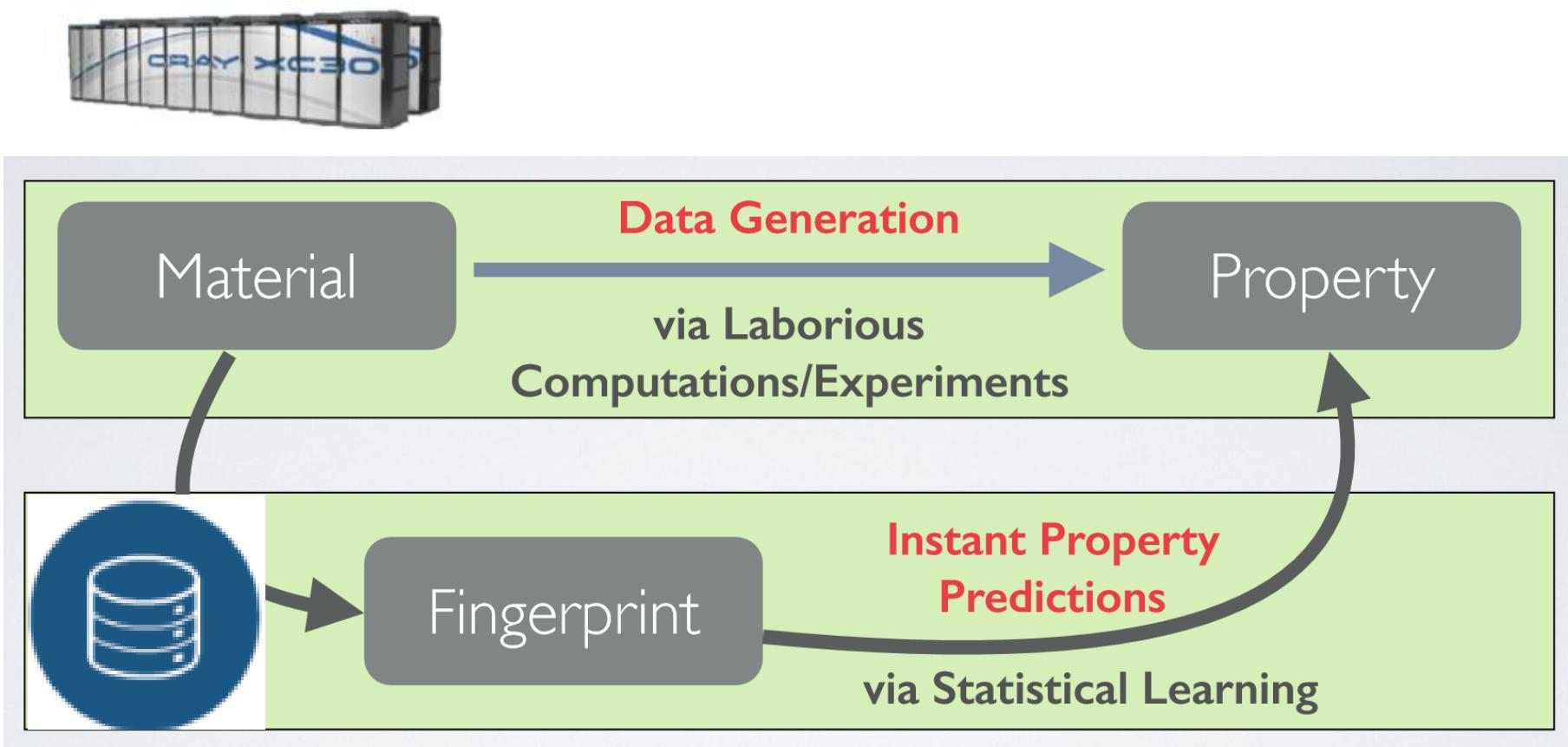


What can Machine Learning Do



Machine Learning in Material Science

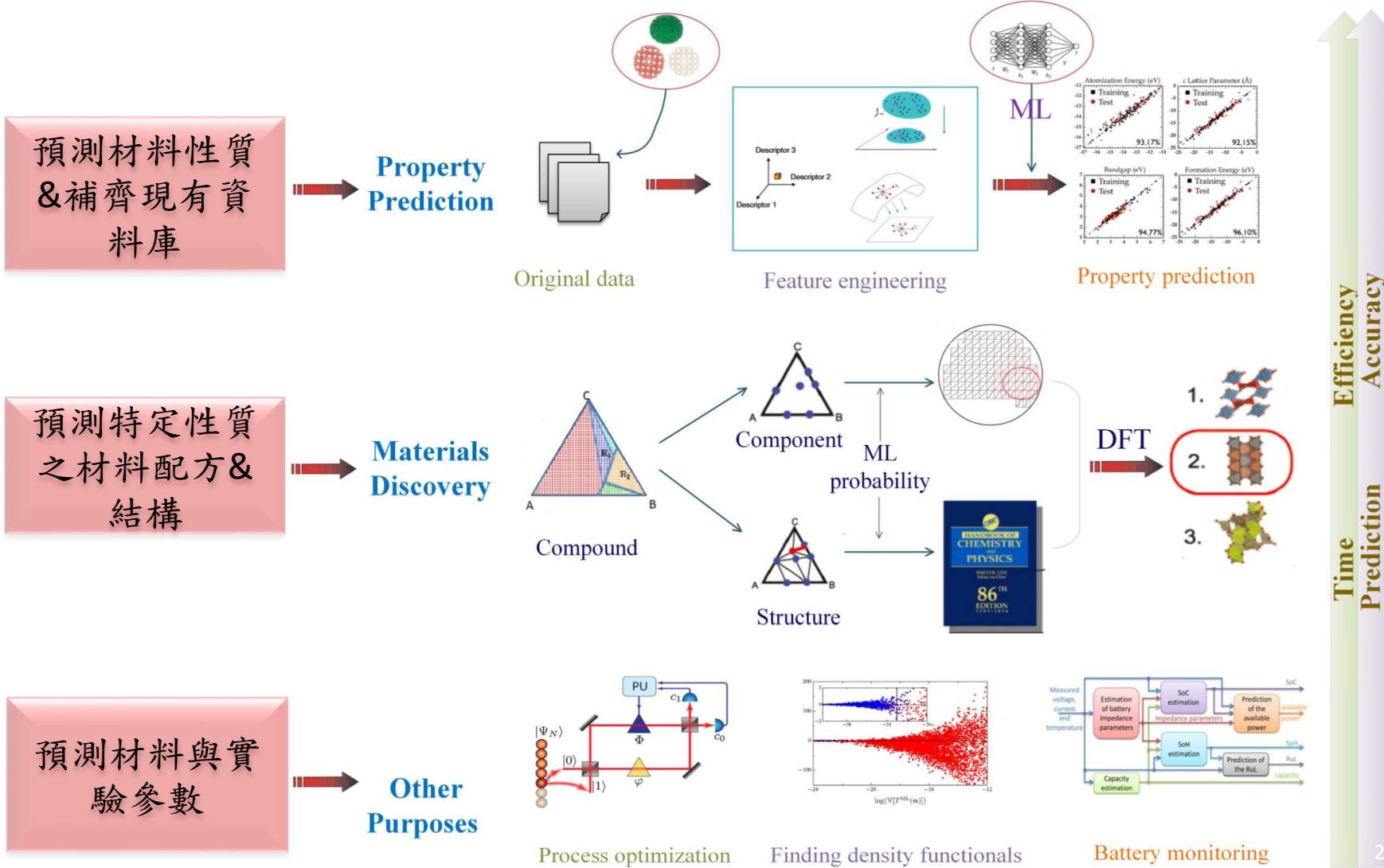
Database Driven Research



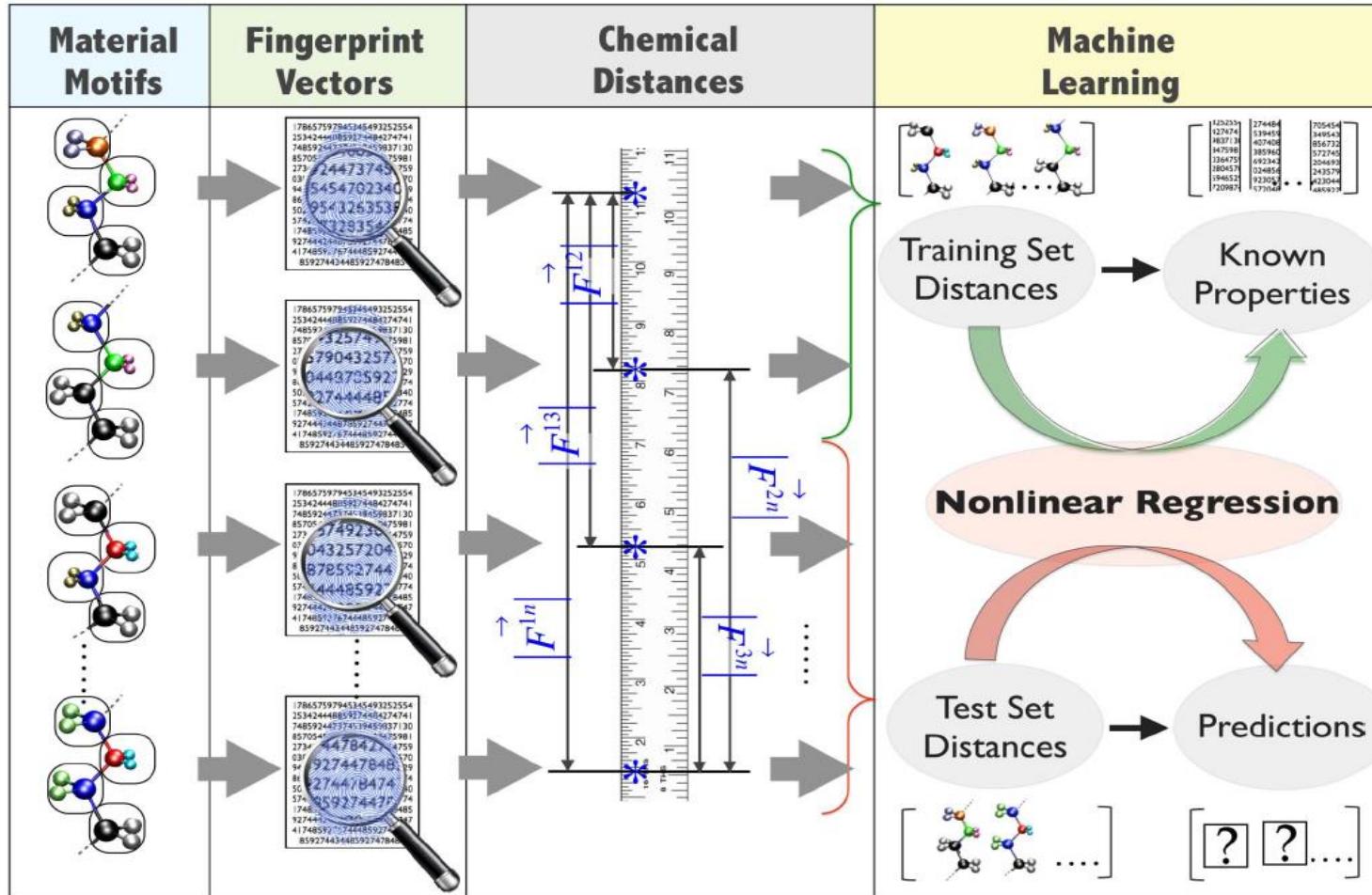
"Machine learning in materials science: Recent progress and emerging applications", Mueller, Kusne & Ramprasad, Reviews of Computational Chemistry (2016)

"Atomistic calculations and materials informatics: A review", Ward & Wolverton, Current Opinion in Solid State and Materials Science (2016)

Machine Learning in Material Science



Machine Learning on the Prediction of Physical Property of Polymer



fingerprints :

$$\vec{F} = f_1 i + f_2 j + f_3 k \dots$$

chemical distance ($\rightarrow x$)

$$|\vec{F}^a - \vec{F}^b| = \vec{F}^{ab}$$

Kernel function P ($\rightarrow y$):

$$P = \sum_{a=1}^N \alpha_a K_{ab}(\vec{F}^{ab})$$

Learning/Training → Minimize the cost function :

$$\sum_{a=1}^N (P_{\text{Est}}^a - P_{\text{DFT}}^a)^2 + \lambda \sum_{a=1}^N \alpha_a^2$$

Pearson's correlation function

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}.$$

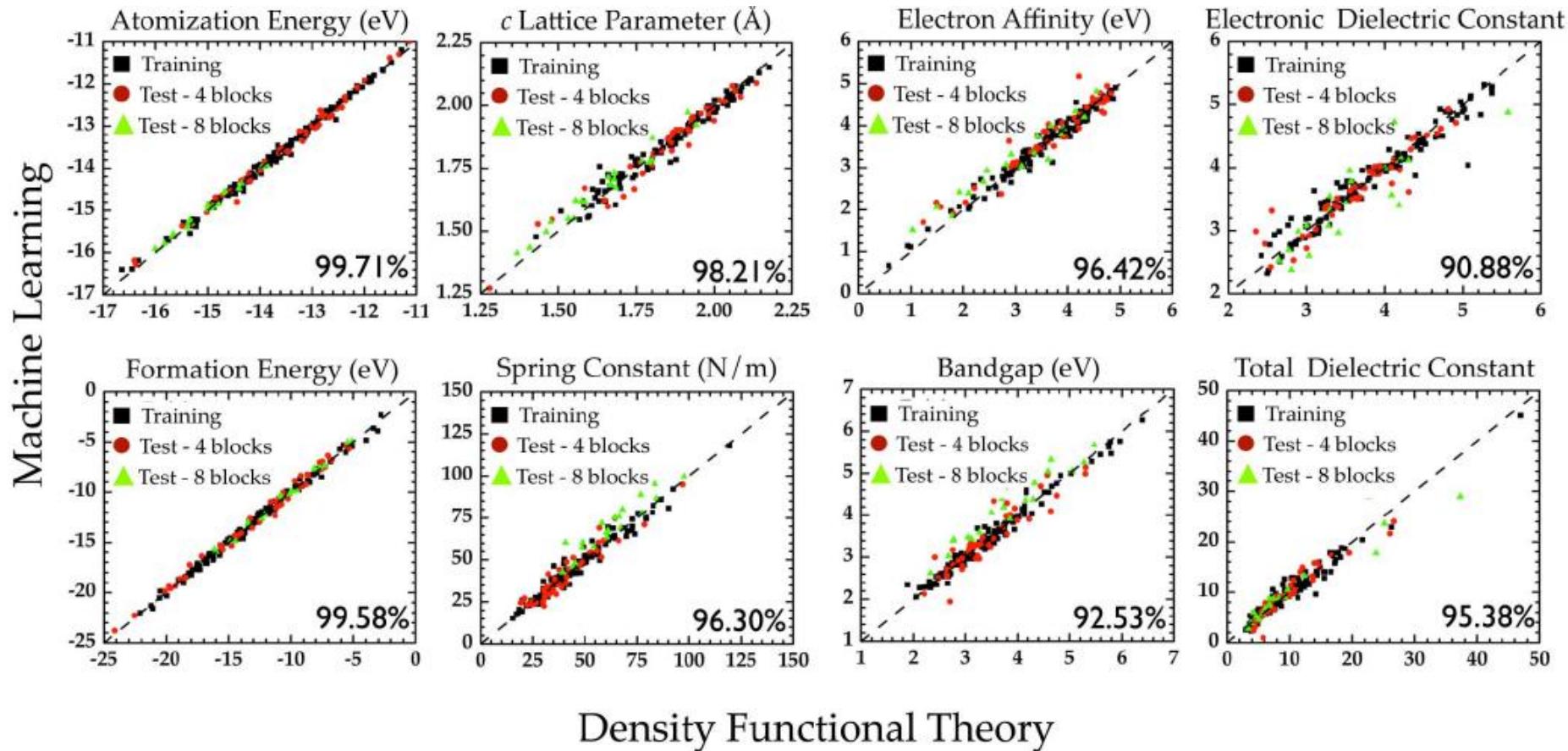
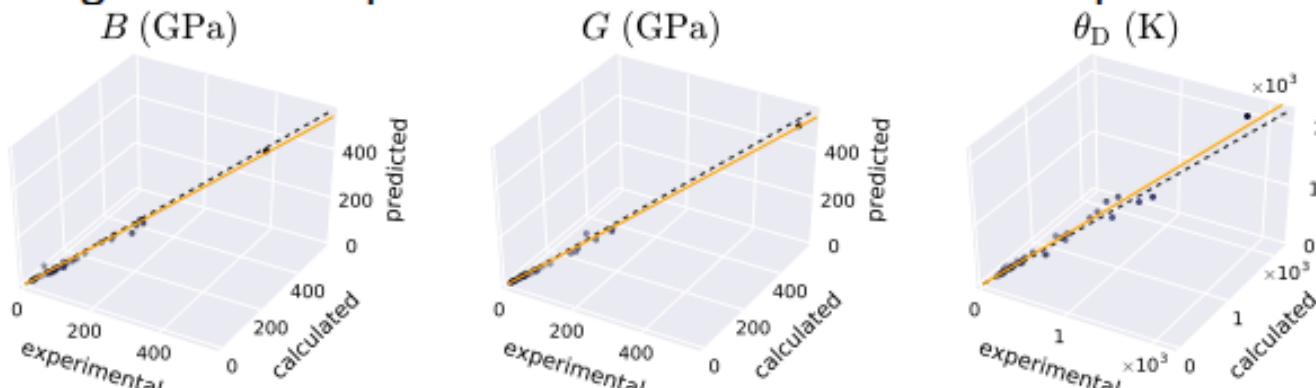


Figure 2 | Learning performance of chemo-structural fingerprint vectors. Parity plots comparing property values computed using DFT against predictions made using learning algorithms trained using chemo-structural fingerprint vectors. Pearson's correlation index is indicated in each of the panels to quantify the agreement between the two schemes.

Calc. vs. Exp. vs. ML Prediction

- Good agreement of predictions with both DFT and experiment



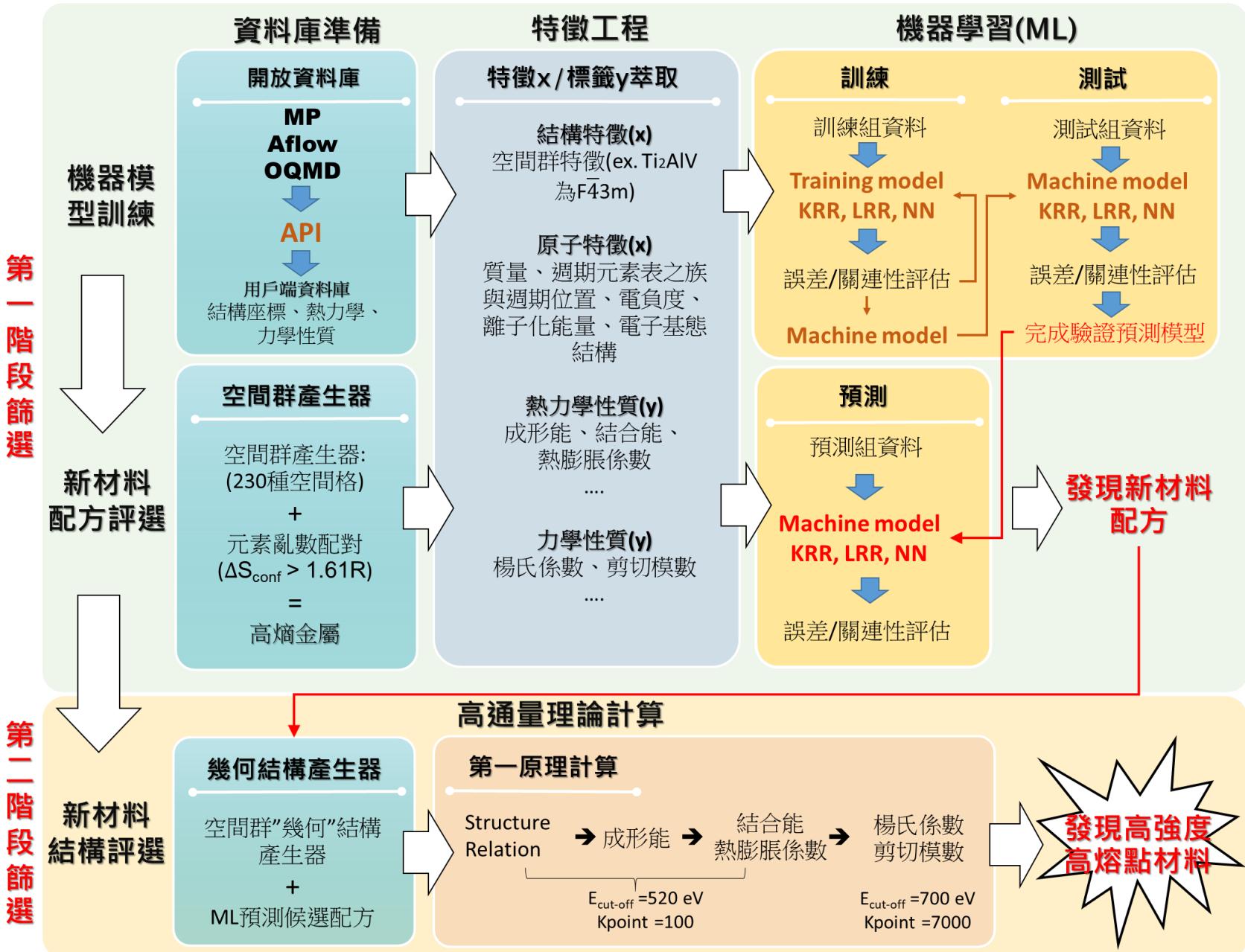
$$B, G : r^2 = 0.99; \theta_D : r^2 = 0.97$$

Figure 7 | Comparison of the AEL—AGL calculations and ML predictions with experimental values for three thermomechanical properties. (a) bulk modulus (B), (b) shear modulus (G), and (c) Debye temperature (θ_D).

Table 3 | Statistical summary of the AEL-AGL calculations and ML predictions versus experimental values for three thermomechanical properties.

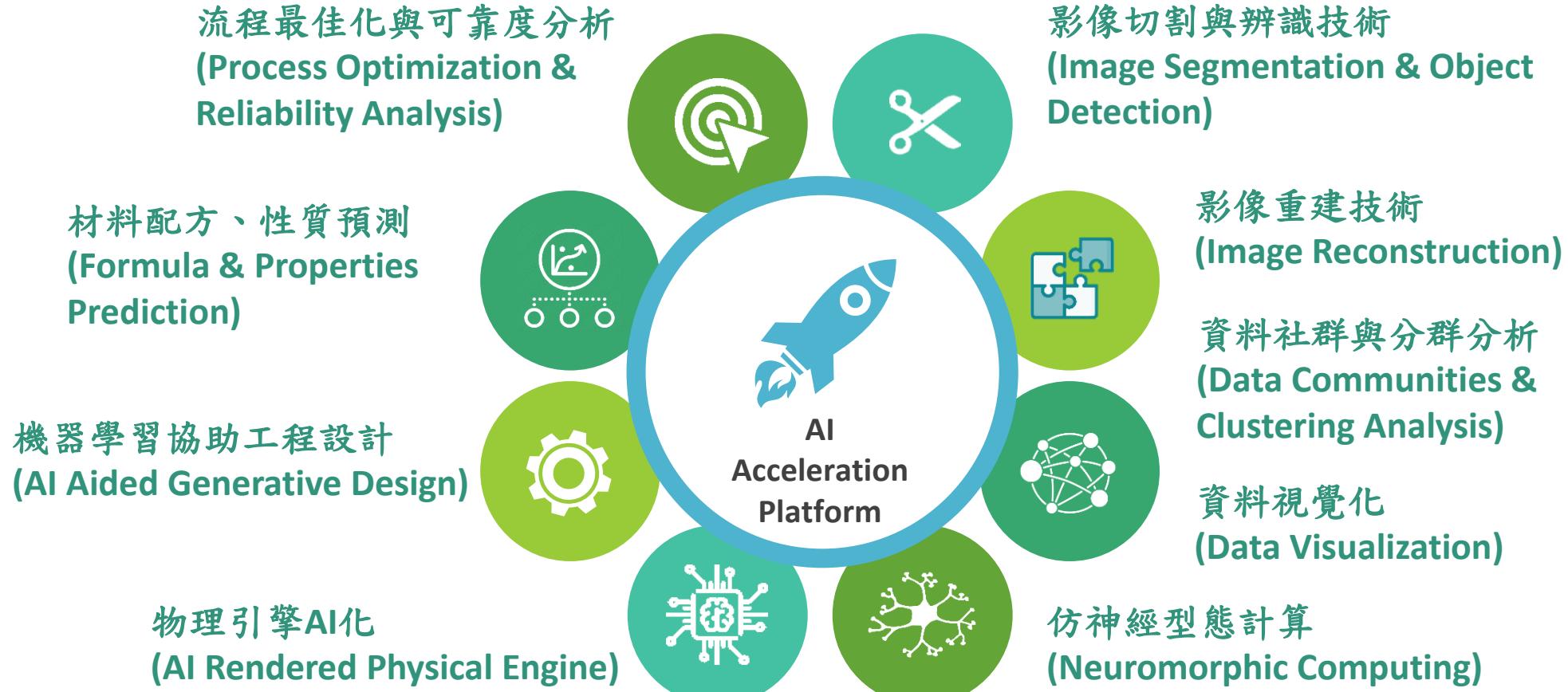
Property	RMSE		MAE		r^2	
	exp. versus calc.	exp. versus pred.	exp. versus calc.	exp. versus pred.	exp. versus calc.	exp. versus pred.
B	8.90 GPa	10.77 GPa	6.36 GPa	8.12 GPa	0.99	0.99
G	7.29 GPa	9.15 GPa	4.76 GPa	6.09 GPa	0.99	0.99
θ_D	76.13 K	65.38 K	49.63 K	42.92 K	0.97	0.97

The summary corresponds with Fig. 7.



人工智能智慧加速科學應用

半導體製程 / 冶金科學與工程 / 有機無機材料 / ...



Artificial Intelligence for Research & Design

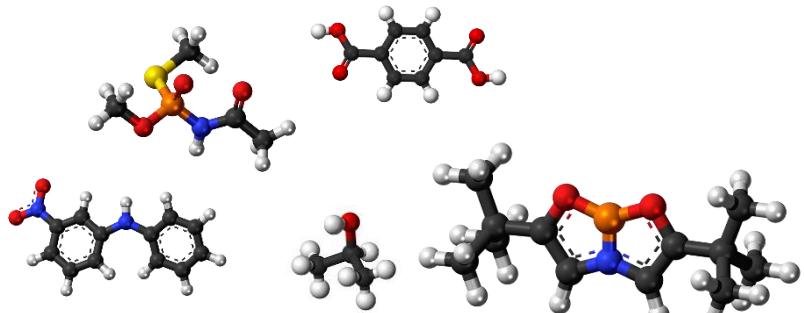


材料配方研發策略

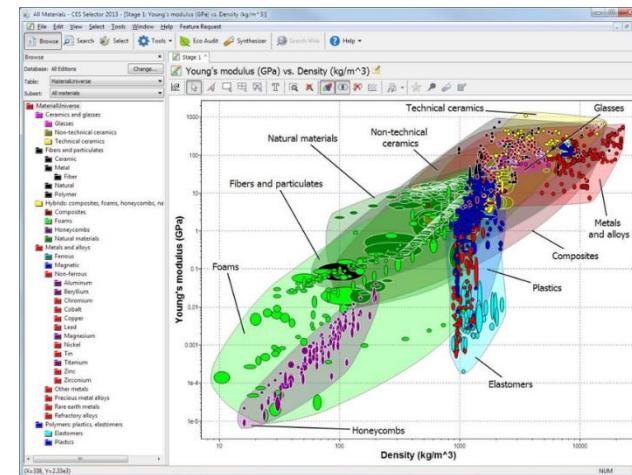
產品規格

40HRB < Y2 < 50HRB	Y1. Density
	Y2. Hardness
	Y3 > 500N/mm
	Y4 > 650N/mm
	Y5 > 35%
Y6 < ...	Y6. Cost

化合物組合與比例



資料庫



實驗&模擬



材料配方研發策略

配方比例最佳化

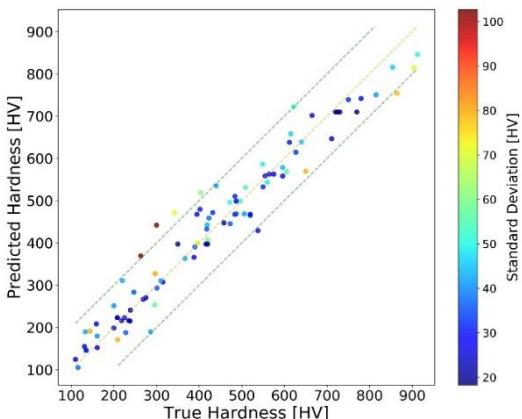
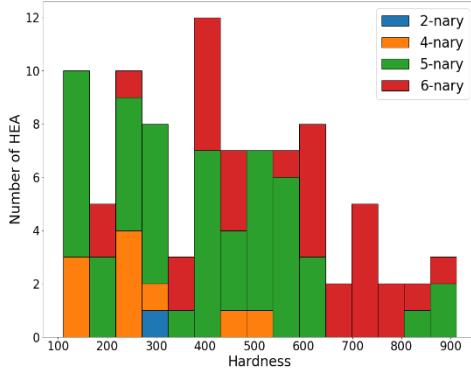
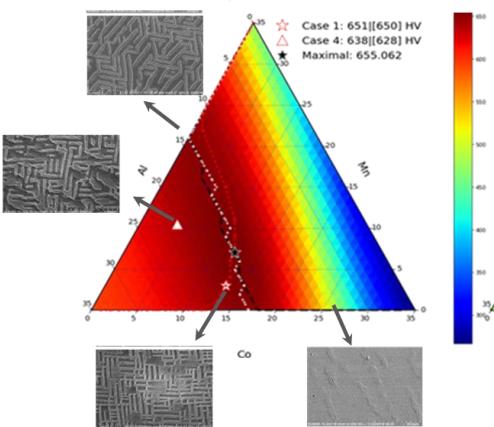
需求

- 複合材料: CoCrFeMoNiAl
- 製程: 電弧融煉
- 規格: 高硬度

資料與來源

- 360筆資料 → 90筆
- 科學文獻

Property Distribution



問題描述：材料配方排列組合種類難以計數、實驗成本非常昂貴 (ex. 單晶柱樣本單一製程>5萬元以上/組，單一製程製作時程至少4~6hr)

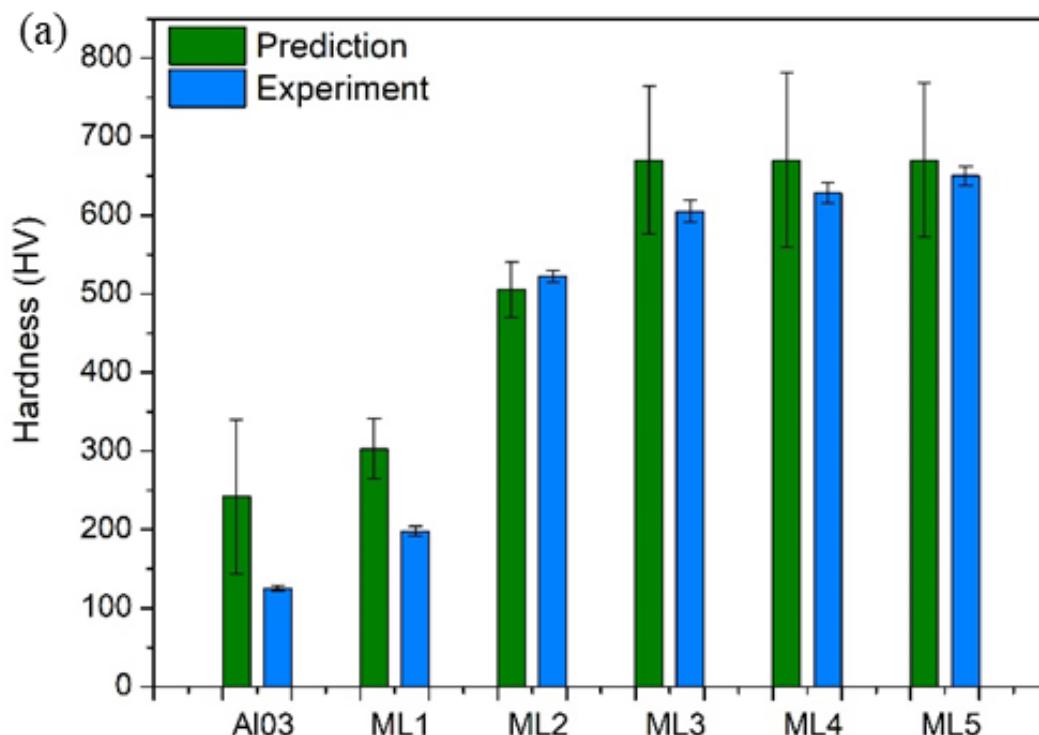
成果：開發材料性質與成分預測之模型，找到材料成分對目標物理量之關聯性，從難以計數的成份比例配對中尋找適合應用端的材料配方，大量減少高成本的實驗試誤。高熵合金的實際案例，已成功找到特定成份的最高硬度配方比 (超越目前發表文獻的最高值)。實驗驗證誤差7%以下，可降低9成實驗成本，加速實驗時程2000倍。其應用端可拓展至半導體、高分子、陶瓷材料...等的發展，驅動實驗前進。



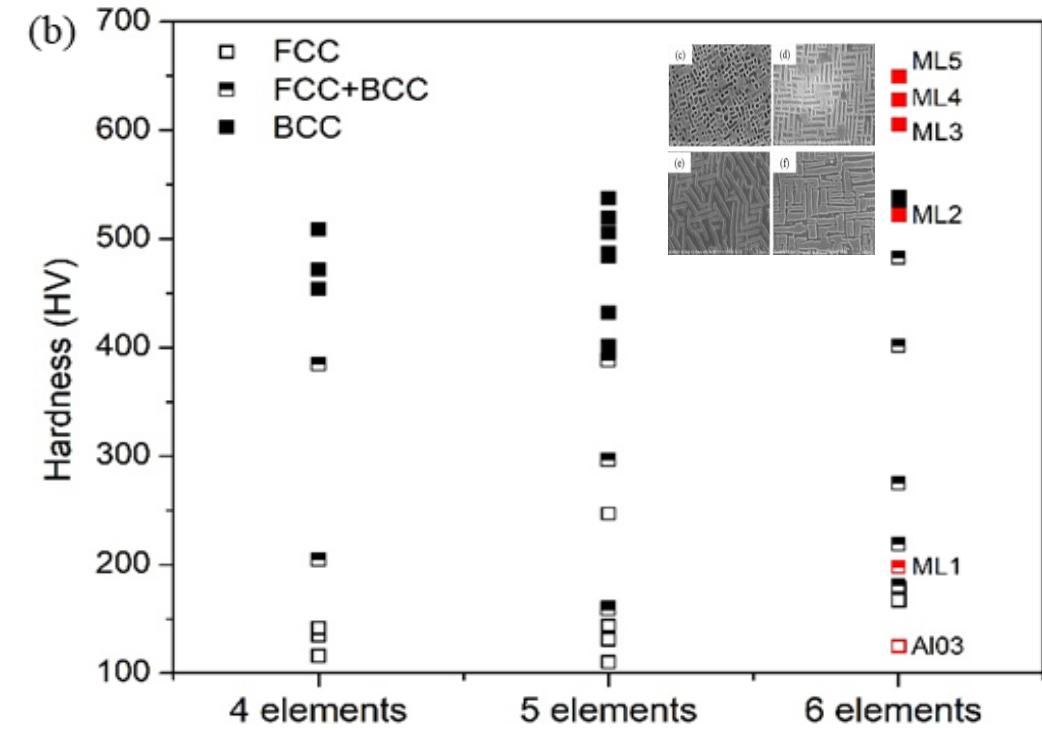
跨越知識屏障、發現新材料物理

材料配方研發策略

Predictions VS. Experiments



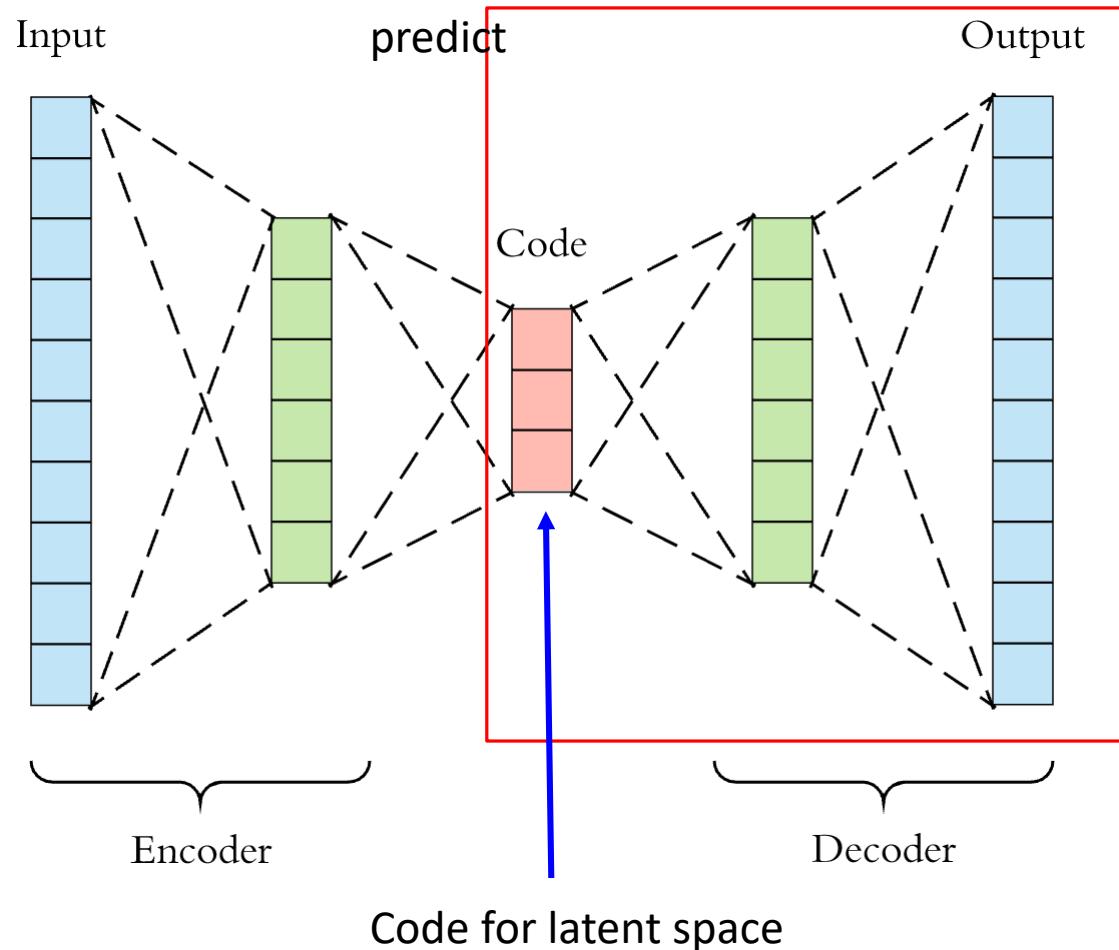
- The Hardness of prediction with uncertainty and experiments with standard error of Al03 and ML alloys;



- the reported hardness of VAM as-cast HEAs containing 4-6 elements chosen from Al, Co, Cr, Fe, Mn, and Ni

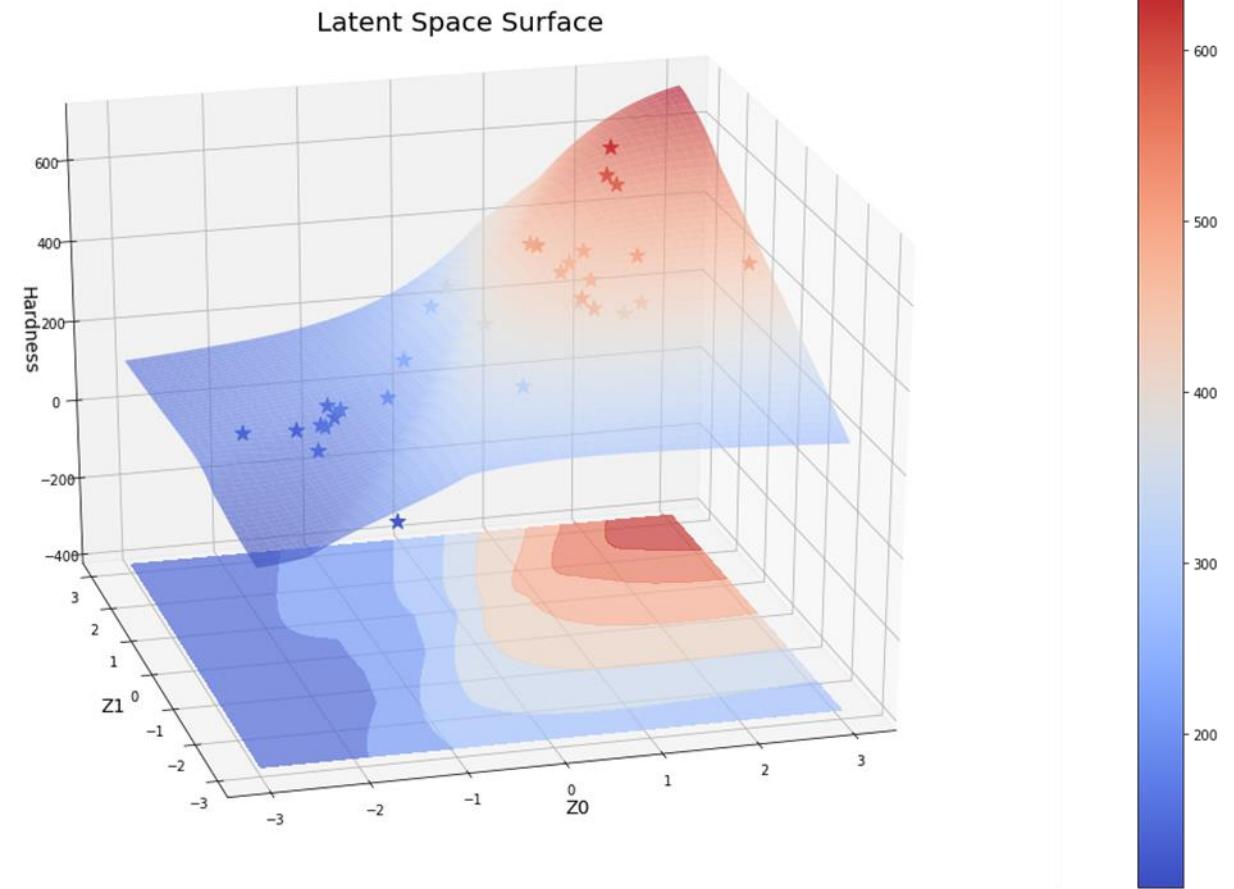
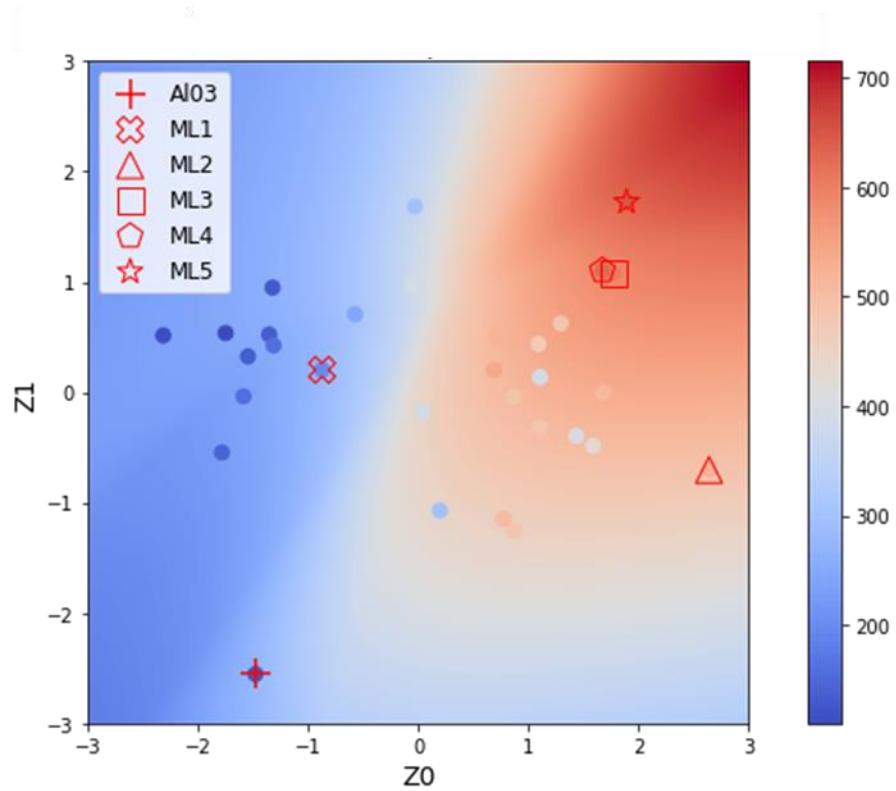
Autoencoder

□ Input and output will be same



材料配方研發策略

□ Hardness on Latent Space



Phase on Latent Space

□ Phase on Latent Space

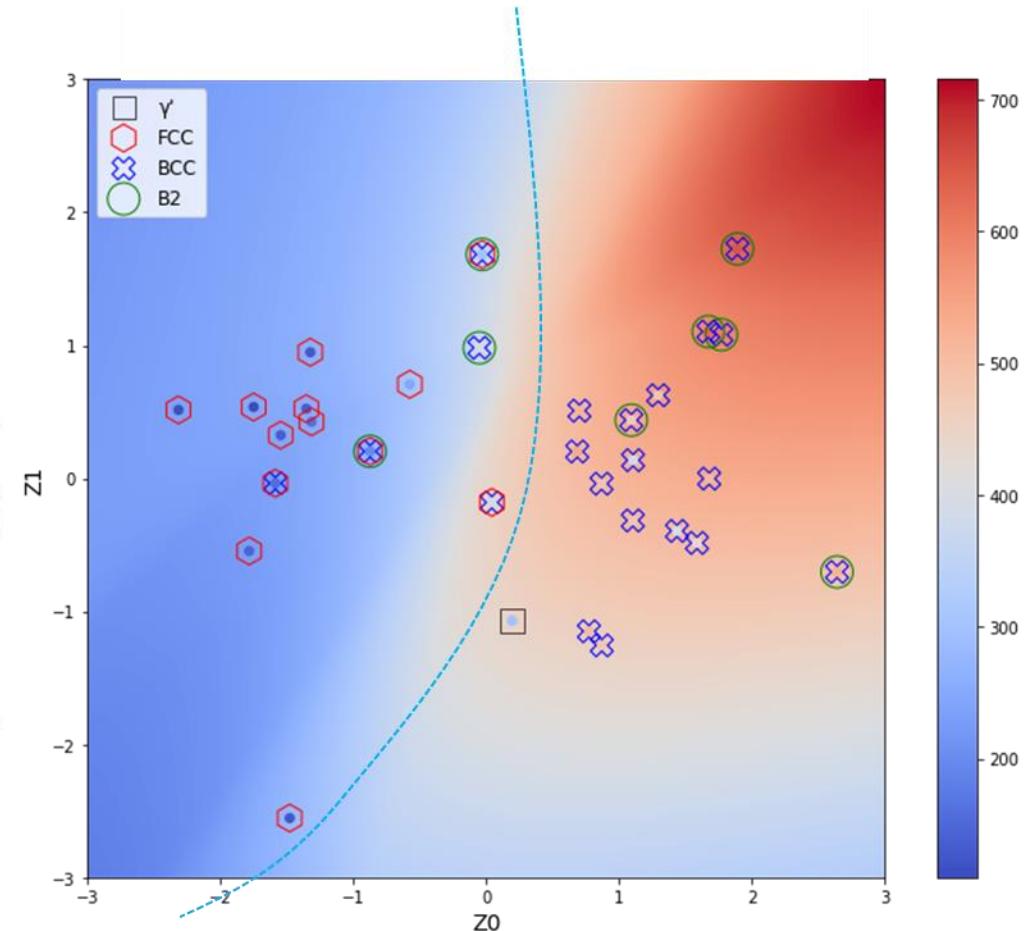
- model trained **without phase info**
- phase transition learned implicitly?

Table I. Nominal composition (at.%) of Al03 and ML alloys, and their hardness values

Alloy	Al	Co	Cr	Fe	Mn	Ni	Experimental hardness	Predicted hardness (HV)
Al03	5.66	18.87	18.87	18.87	18.87	18.87	125 ± 3	242 ± 98
ML1	11	18	22	22	5	22	198 ± 6	303 ± 38
ML2	30.5	16	18.5	16.5	5	13.5	522 ± 8	505 ± 35
ML3	30	6	35	6	18	5	605 ± 14	670 ± 94
ML4	25.5	9	35	10	15.5	5	628 ± 13	670 ± 111
ML5	24	18	35	10	7.5	5.5	650 ± 12	670 ± 98

JOM 71, 3433 (2019)

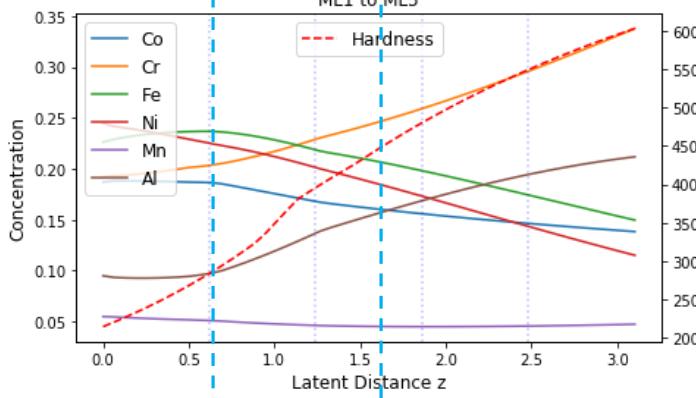
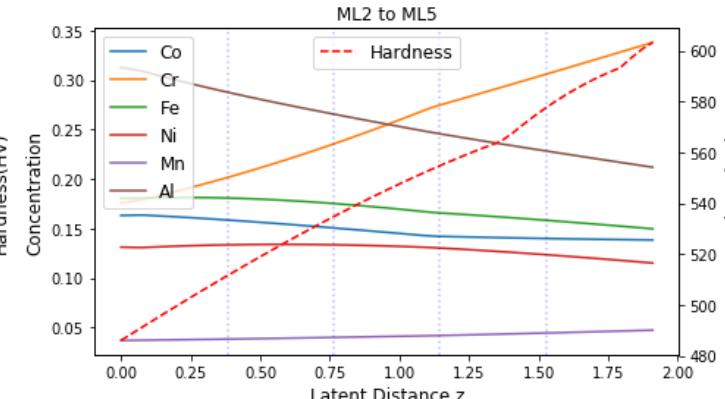
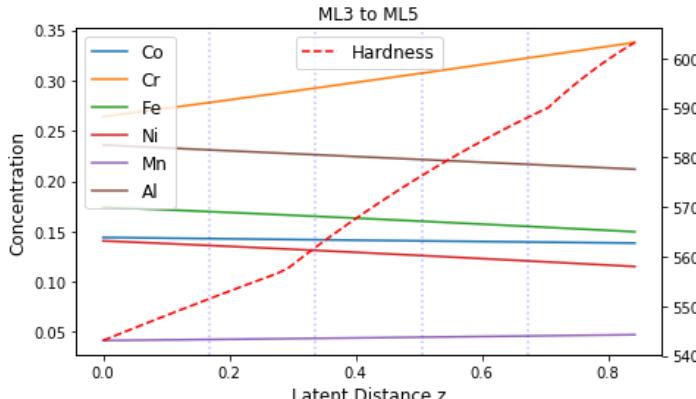
Boundary between FCC & BCC?



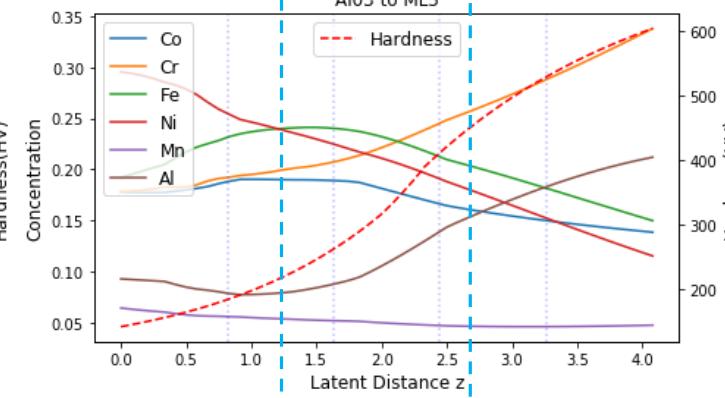
Phase on Latent Space

Phase on Latent Space

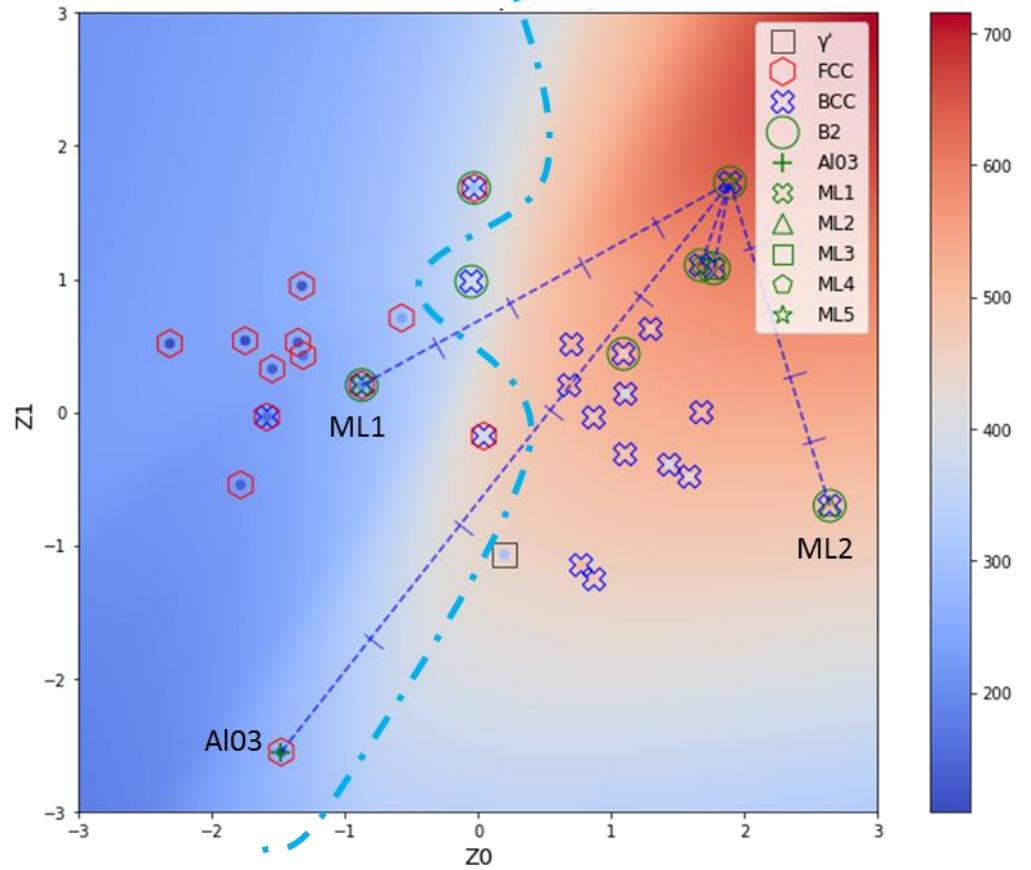
➤ recognize phase transition implicitly?



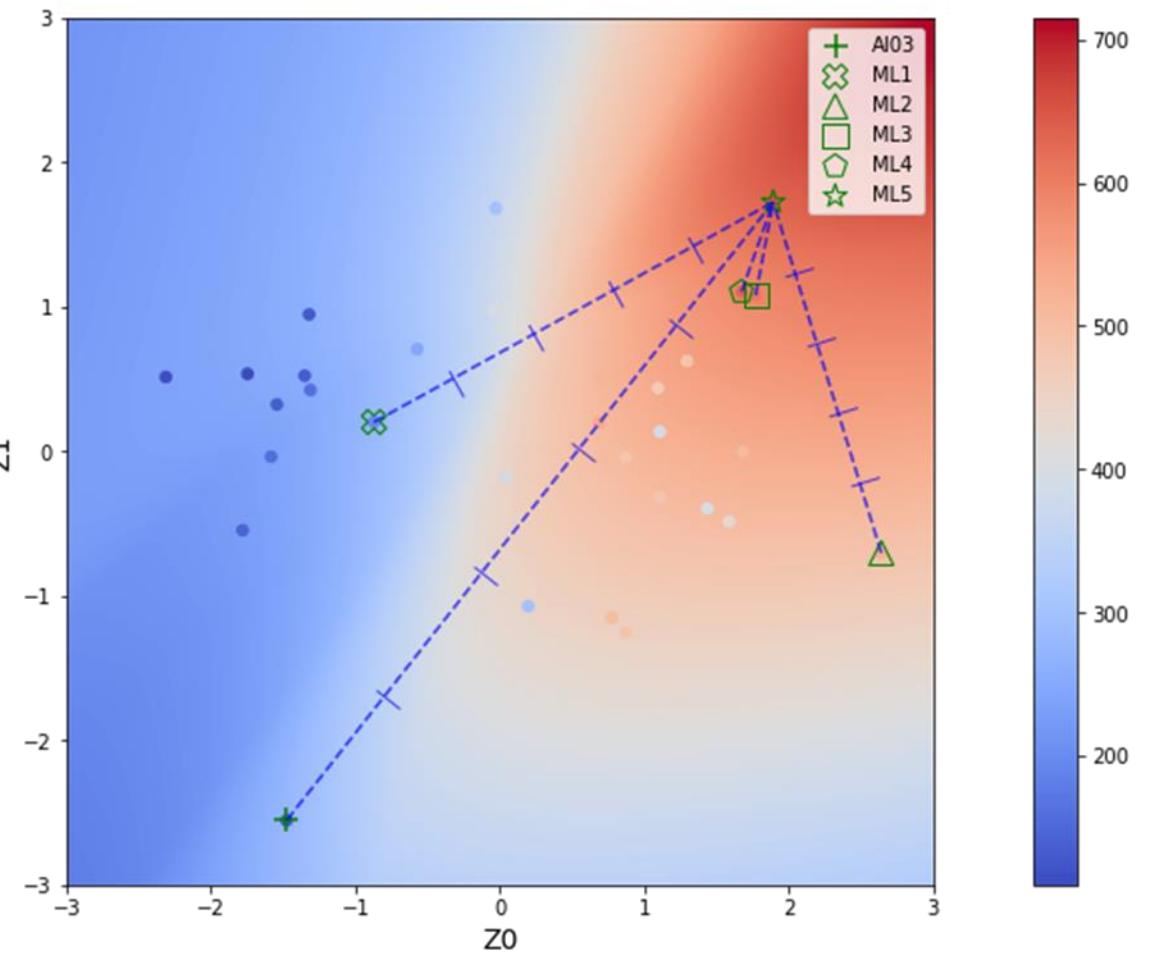
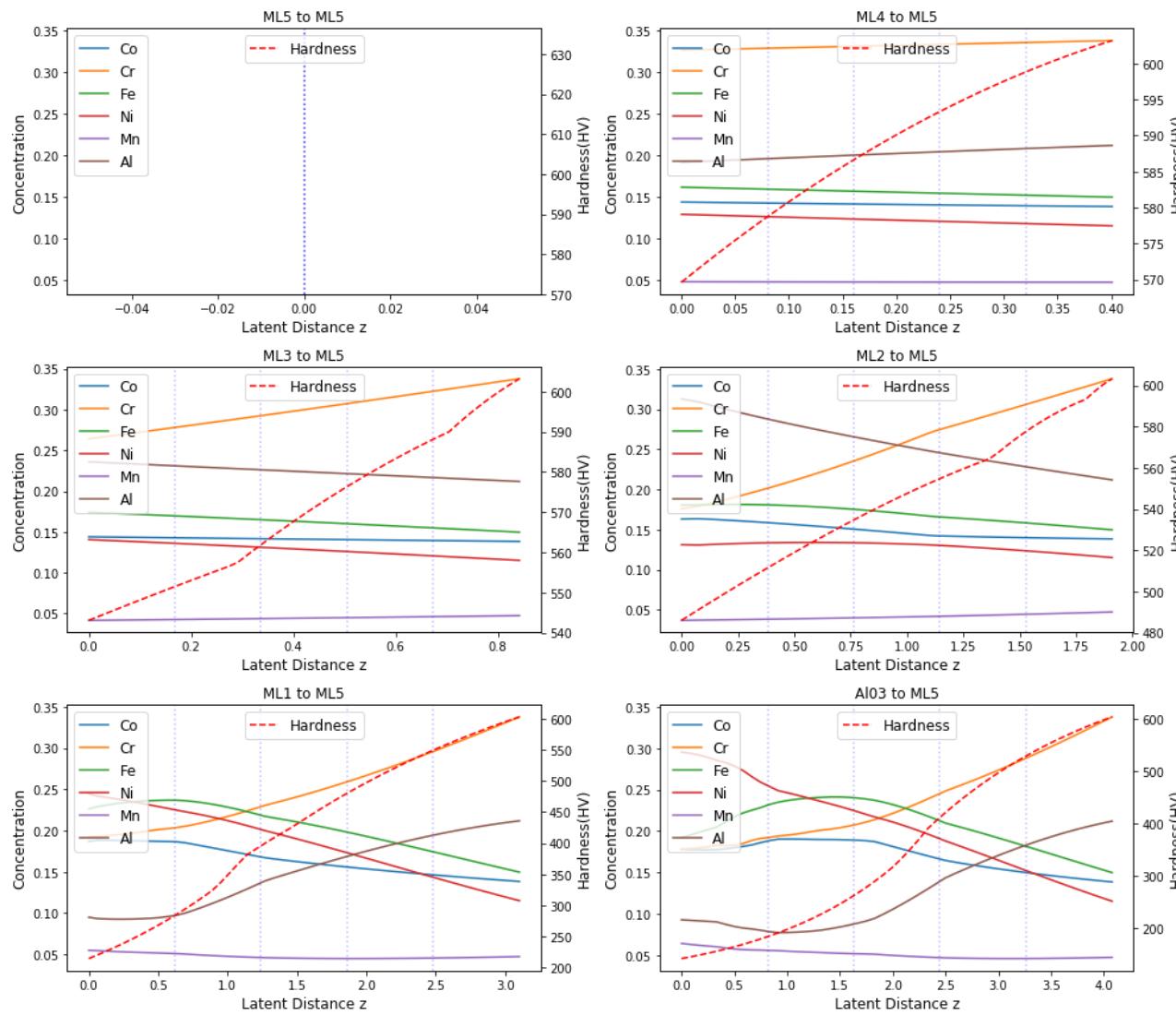
phase transition?



Implicit clusters of FCC & BCC?



Traces of Composition and Property

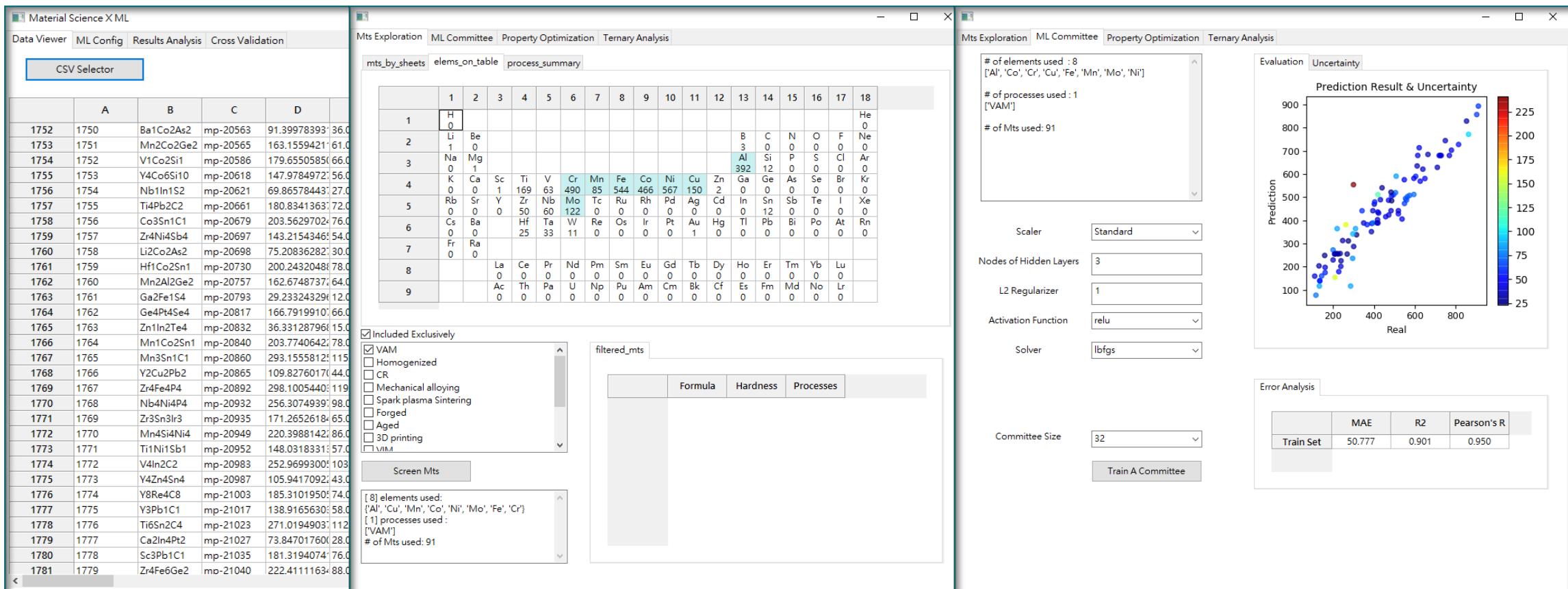


Material Design Platform

Input Data

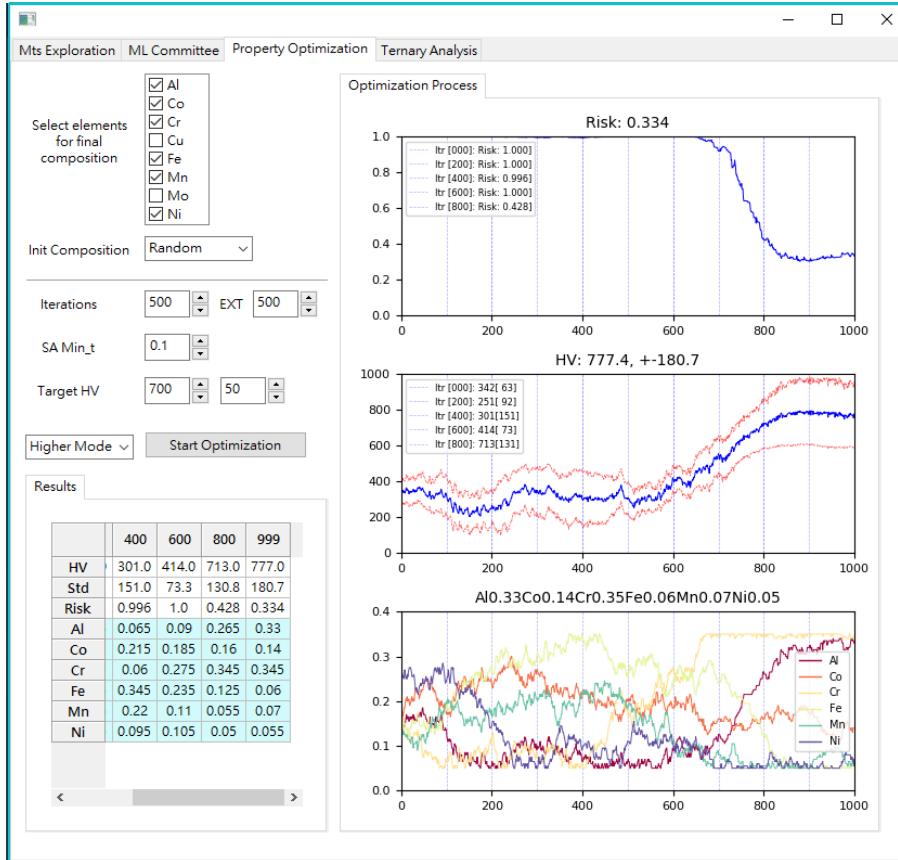
Data selection and Feature engineering

Training/Prediction

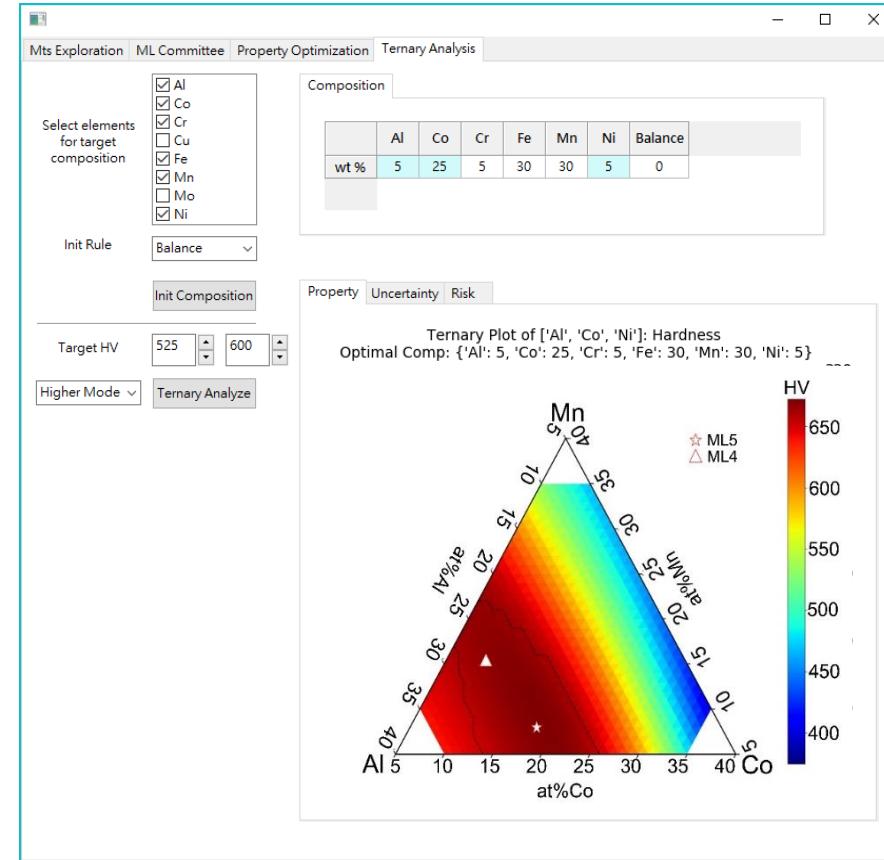


Material Design Platform

Optimization



Visualization

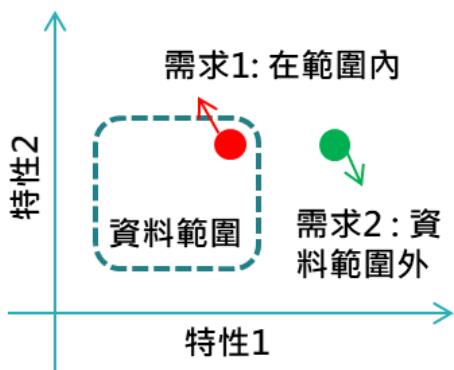


材料配方研發策略

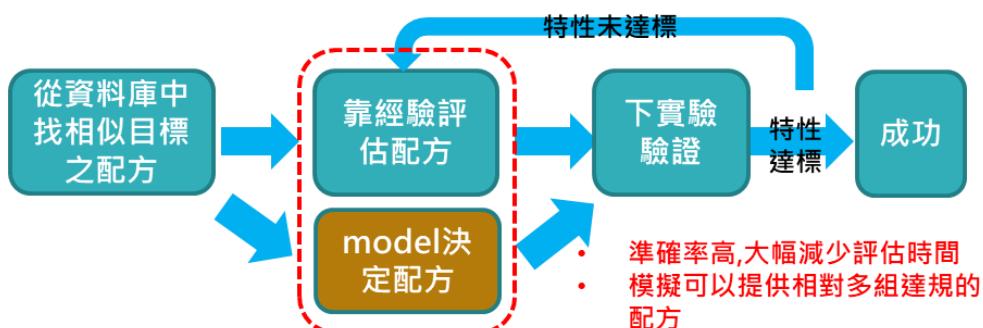
多重產品規格之配方最佳化

產品需求

- 複合材料：160種可能分子組成
- 規格：15種材料性質



	Requirement
Y1	>89
Y2	1.54 ~ 1.74
Y3	>4.9
Y6	>11
Y8	< 75.4
Y9	> 0.104
Z3	>10



問題描述：滿足多產品規格需求之複合材料配方發現是相當困難，更遑論從數以百計的候選配方中選擇。

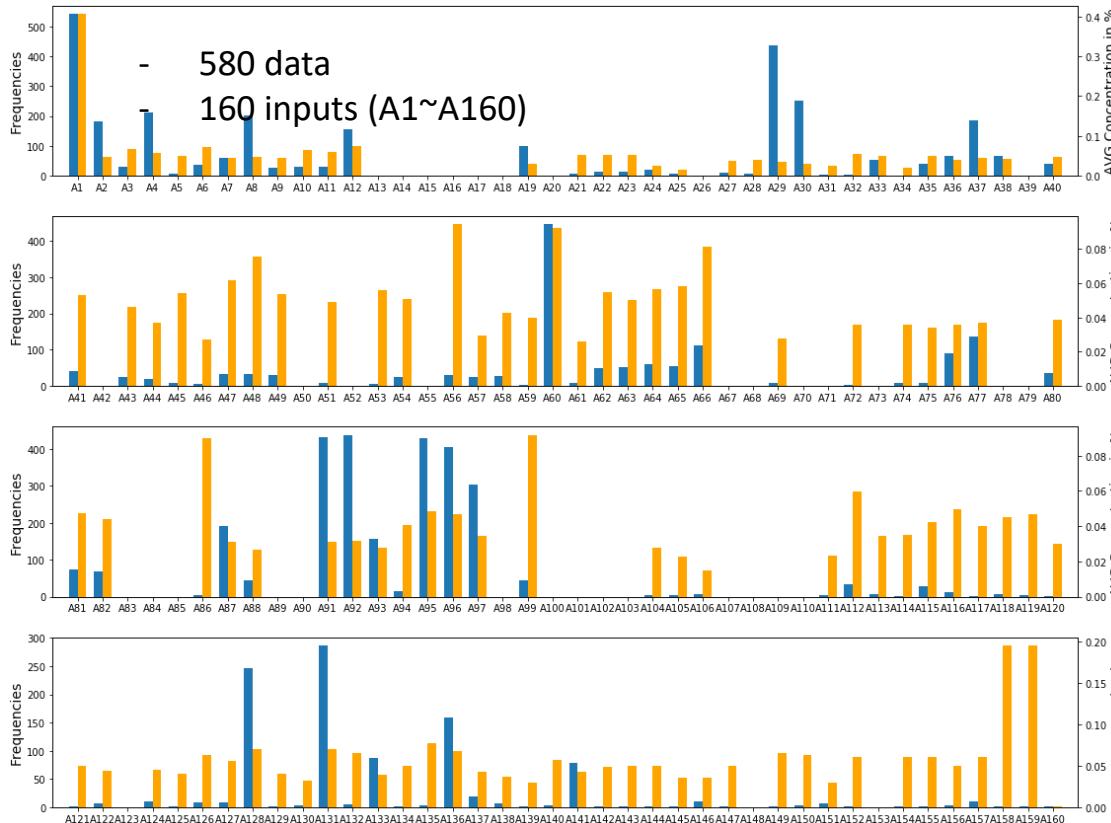
	到達目標值所需花費的配方數量	所需花費的時間
既有方法	需求1: 30組配方	至少一個半月
	需求2: 100組配方	至少三個月

成果：開發**多性質目標預測模型與材料配方最佳化之模型**，使客戶從過去只能追隨已知專利的配方改質，到自行研發配方。驗證準確率高於**9成**，大幅度加速實驗時程。

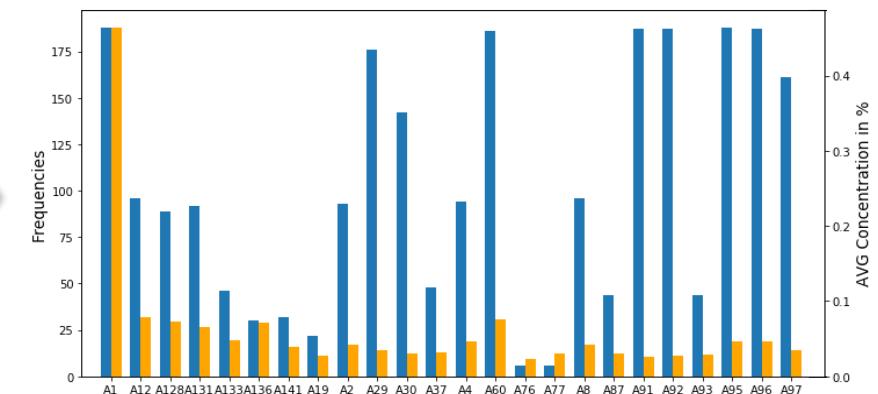


跨越知識屏障、發現新材料物理

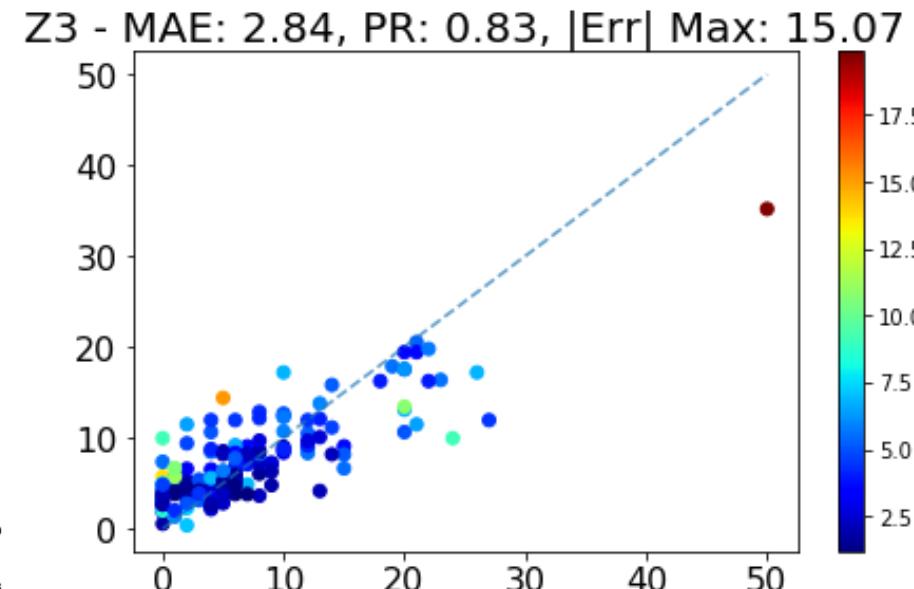
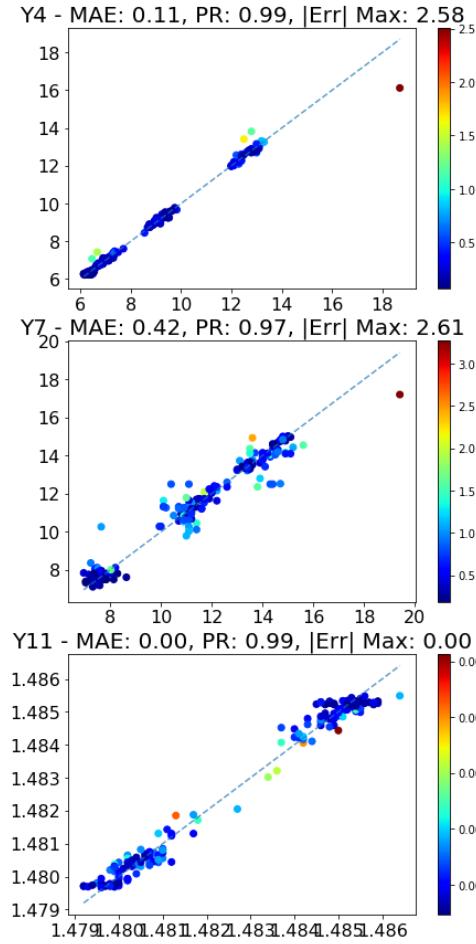
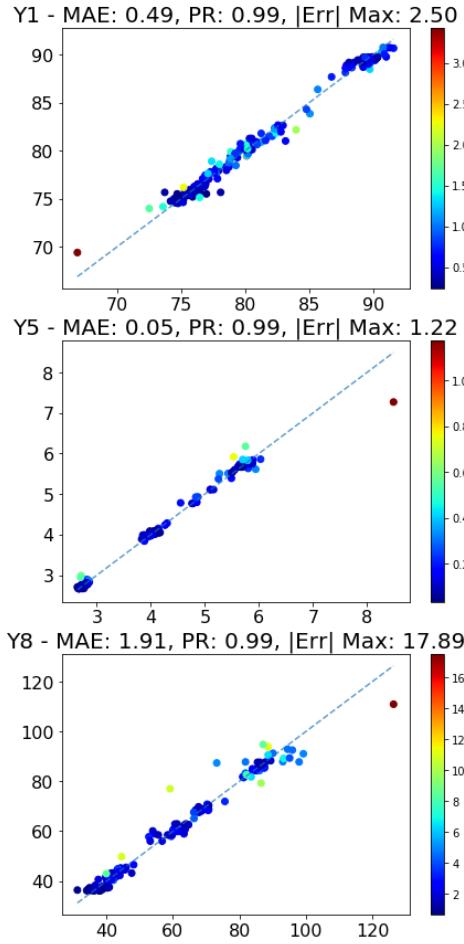
資料清除與採用



- 188 data
- 24 inputs

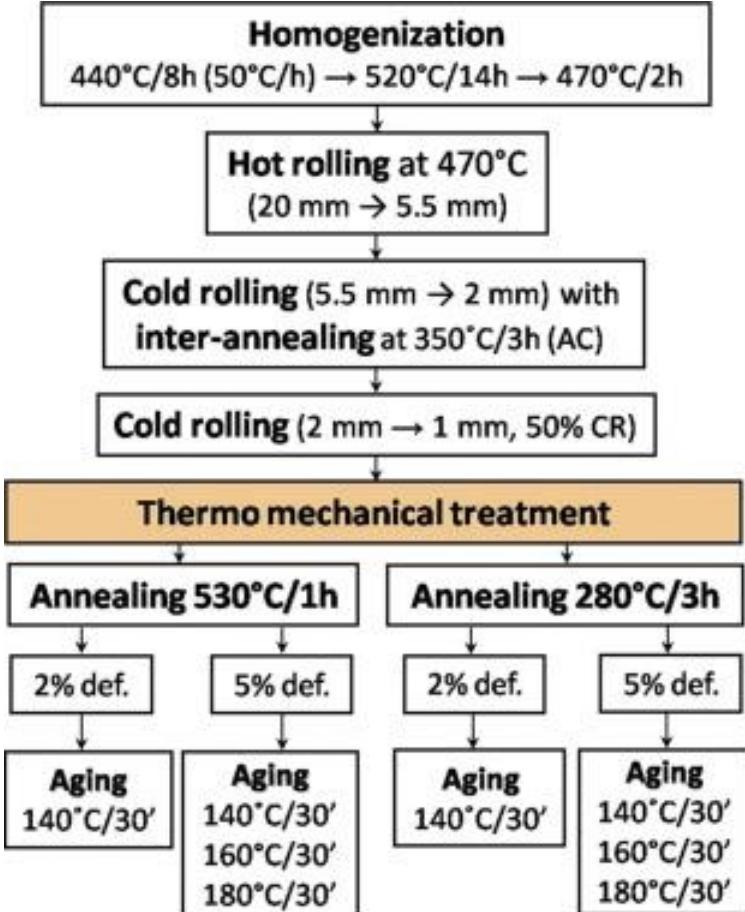


性質預測



製程流程設計

精準材料製程流程設計



問題描述：多樣性實驗流程或參數搭配對物理性質有決定性影響（例如冶金製程的均質化、冷壓、退火、...等），除了製作程序耗時且成本昂貴外，製程與參數間的影響難以評估。

成果：本案例藉由機器學習建立多參數空間模型，成功預測不同高熵合金配方在不同製程與環境參數下的材料物理量及影響，並可分析參數或製程工序間的關聯。本案例預測準確度達94%，可降低實驗成本9成，實驗時程加速約7000倍。大幅縮減實驗測試與洞悉流程深層的關聯性。



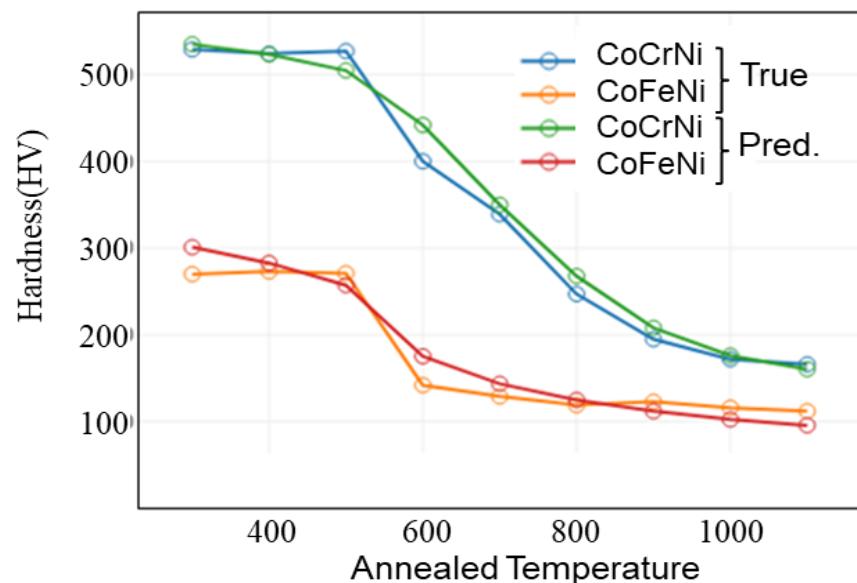
洞悉流程間關聯、了解製程-特性關係

製程流程設計

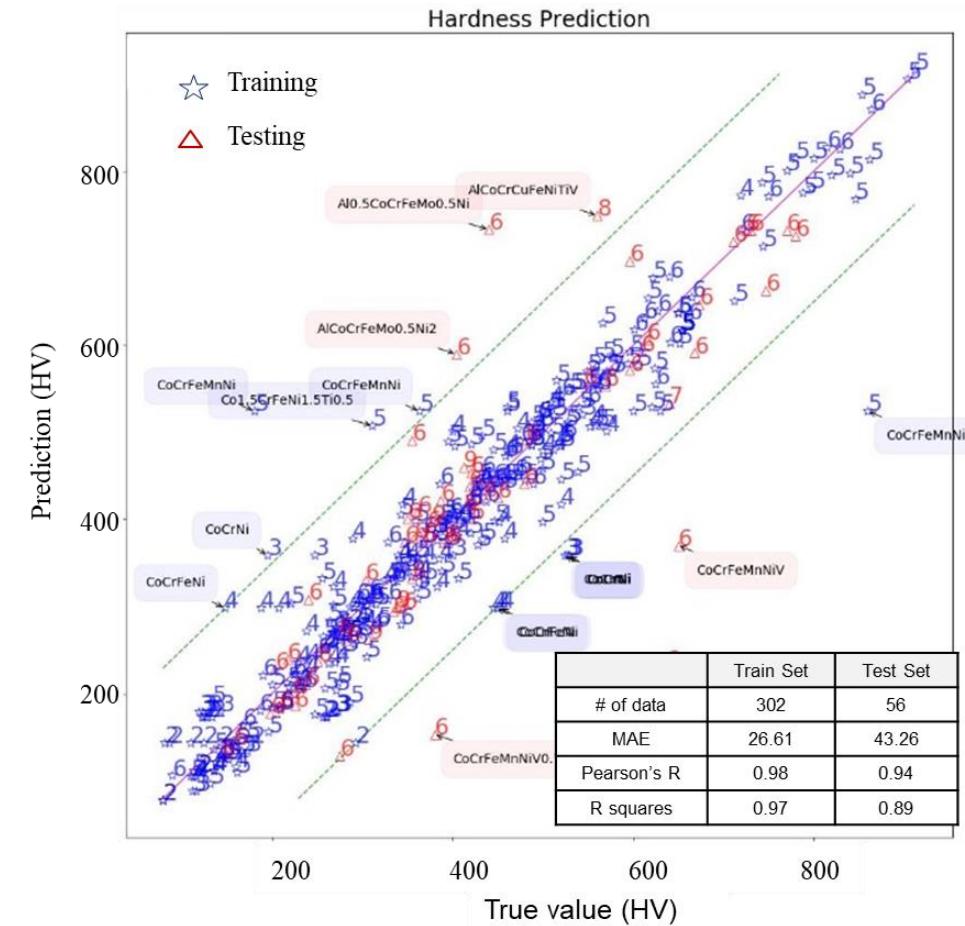
精準材料製程流程設計



- Hardness Prediction of alloy with multi manufacturing processes



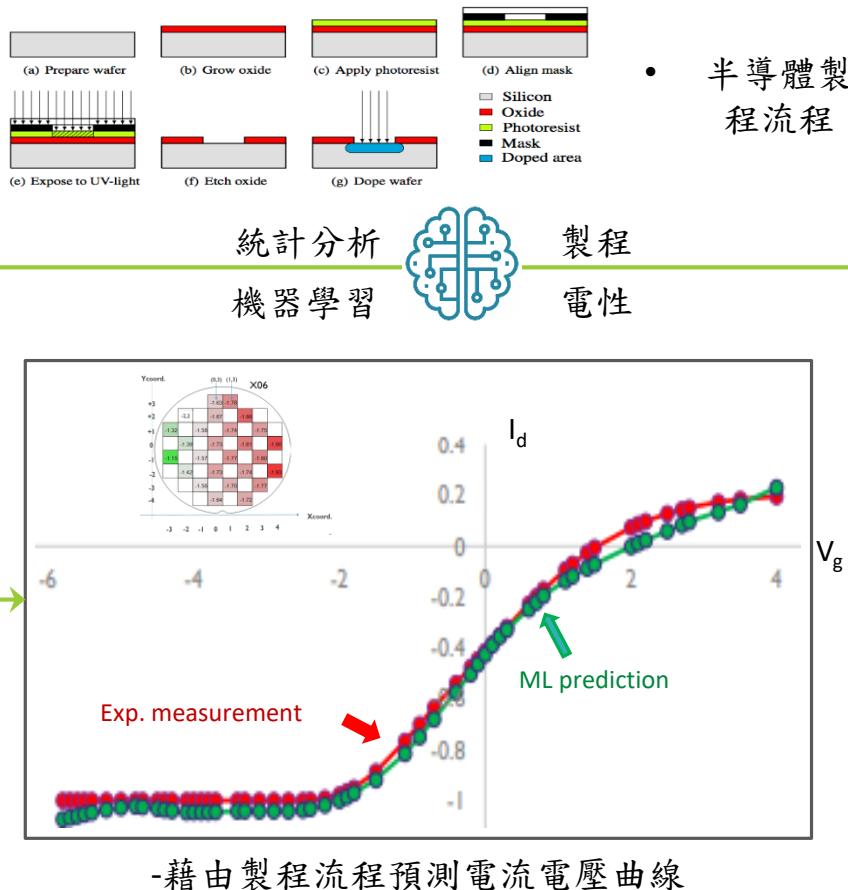
- Hardness as a function of annealing T



- Prediction values by ML vs. True value by experiment

製程流程設計

精準元件製程流程設計



問題描述：不同製程工序參數的綜合效應及晶圓上不同位置晶片的電性影響難以評估，更遑論製程參數對於晶片的可靠度分析影響。

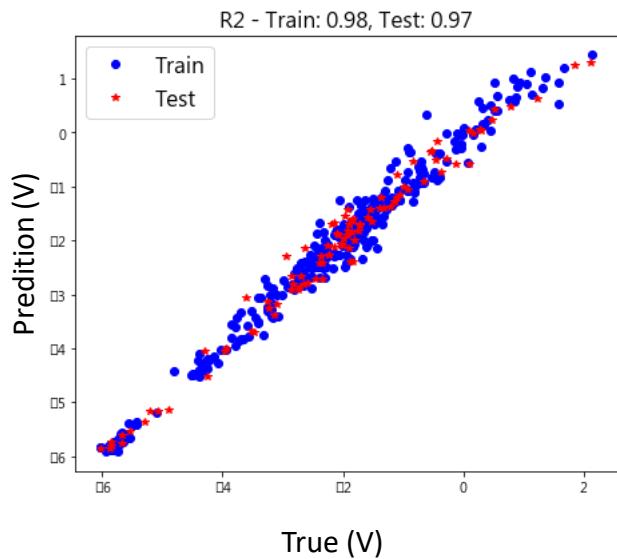
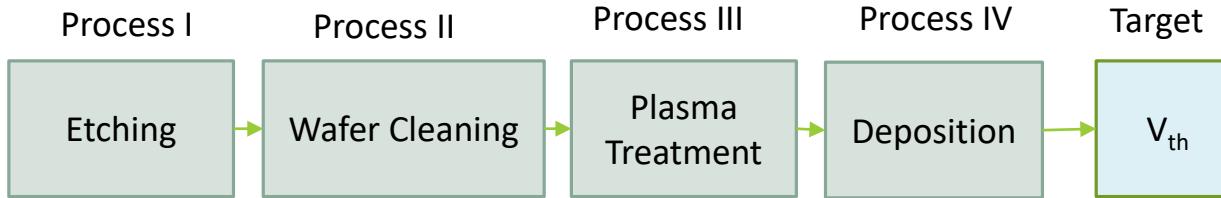
成果：本案例建立多重製程工序之元件性質預測模型，應用於預測不同參數條件下的臨界電壓、導通電阻、關閉電壓、甚至電流電壓特徵曲線。可進一步協助製程參數最佳化，輔助尋找製程、效能與可靠度間的關聯性，進而改良製程獲得最佳電性，並大幅降低實驗成本(ex. 歐洲工業級製程單次>300K EUR, NDL製程500K NTD)。



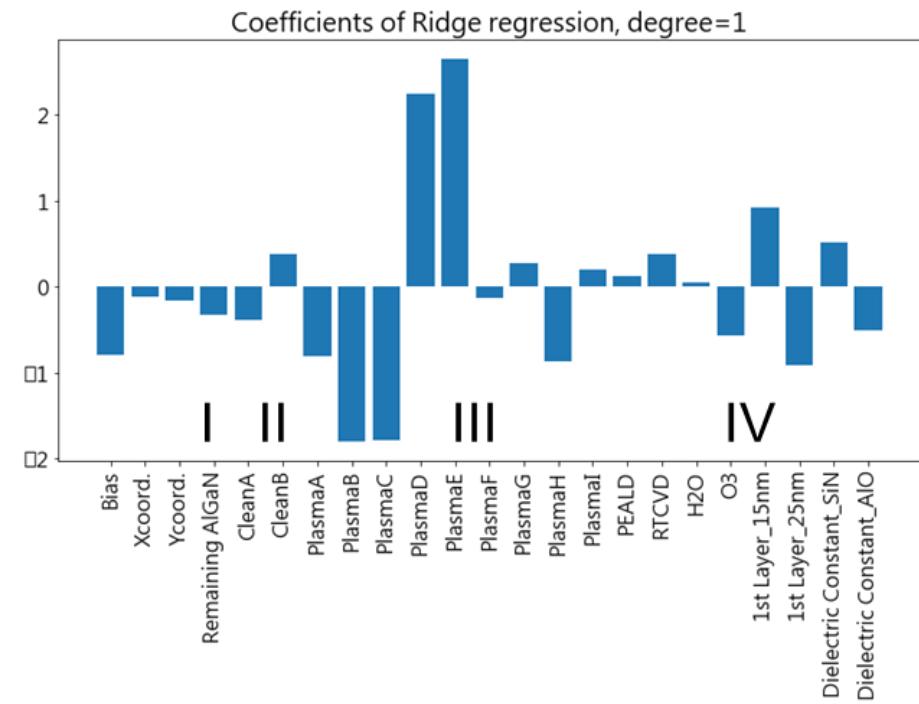
解析製程-性質-可靠度等系統性關係

製程流程設計

關聯性分析



Correlation analysis



❖ Attribution analysis

製程流程設計

逆向設計

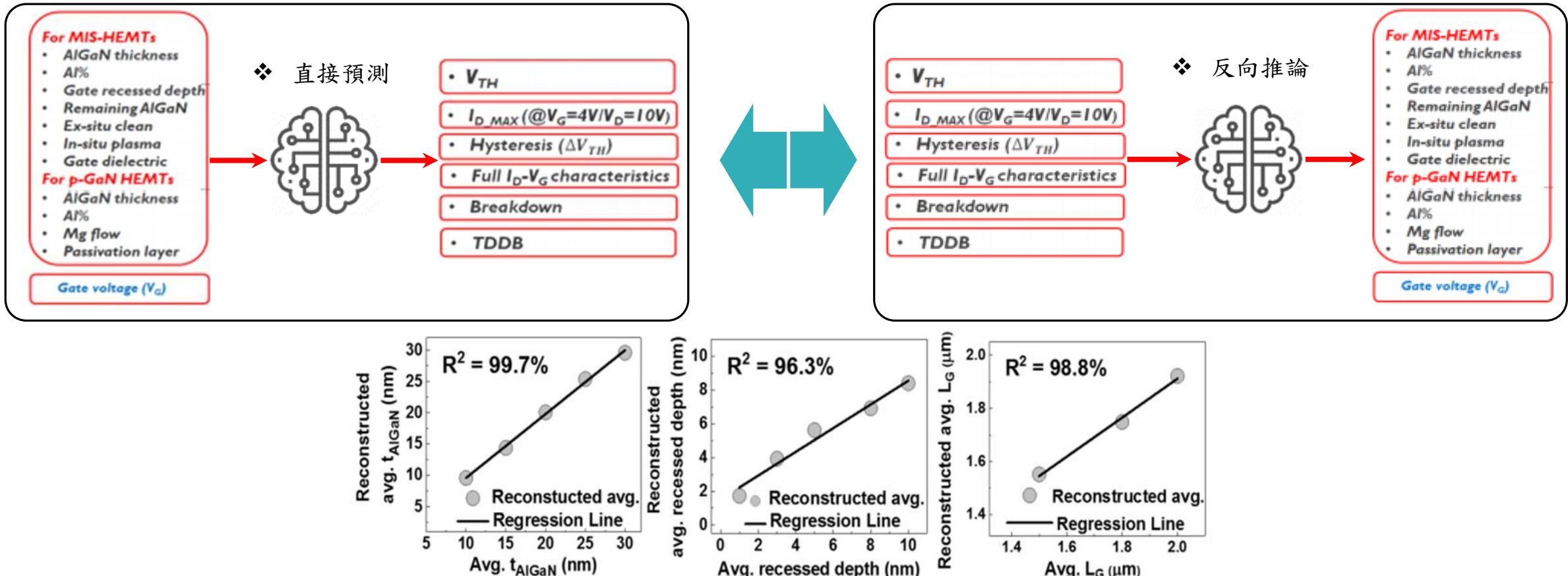
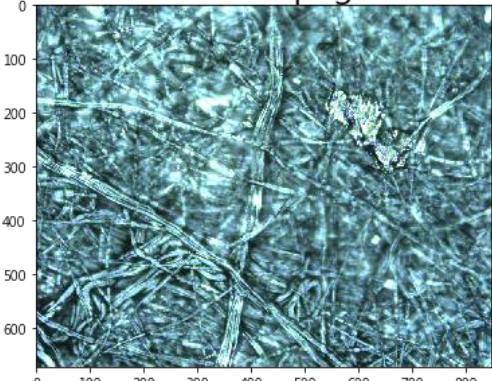


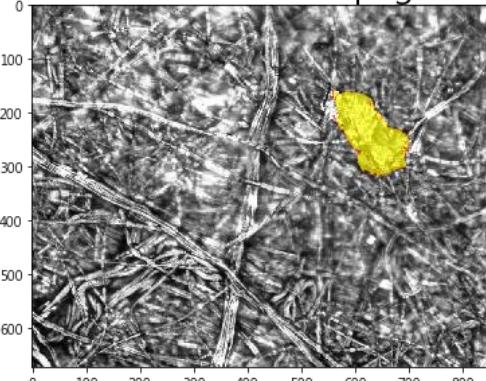
Fig. 12. Regression analyses for 3 reconstructed device parameters in MIS-HEMTs. Device parameters are reconstructed using the autoencoder for all of the input device parameters and the average (avg.) results are used to evaluate R^2 .

智慧影像處理與瑕疵檢測

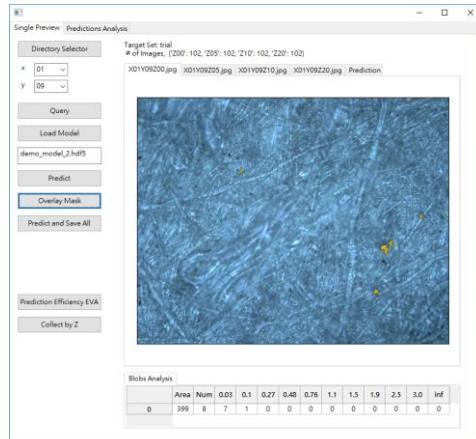
高階影像處理機器



(a)optical image



(b)DL segmentation result



雜質分析使用者介面

問題描述：二維材料表面或三維重建體影像之材料缺陷、雜質或邊界的辨識，除了沒有固定的判讀標準外，更無量化之技術。

成果：本技術已成功應用在**材料影像的雜質自動偵測、邊界輪廓辨識、及其幾何物理量的量化**。其中材料雜質偵測案例，成功建立判別標準，準確率超越工程師人眼判別能力，並節省廠商檢測設備購置至少270萬元與人力資源75%。其技術可應用於3D半導體之斷層攝影結構之缺陷檢測、SEM微結構分析與雜質檢測、金屬材料相結構之分析、...等有關產品品質與影像關聯性高之產業。

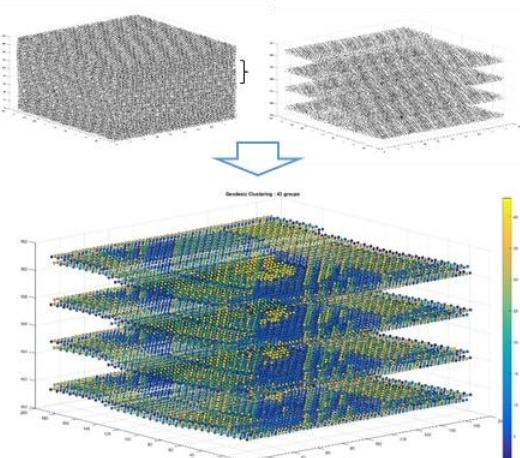


避免人眼判斷之偏見或不一致，提升產品的可靠度、穩定度、與產品效能。

深度學習協助材料影像重建技術

原子結構影像重建

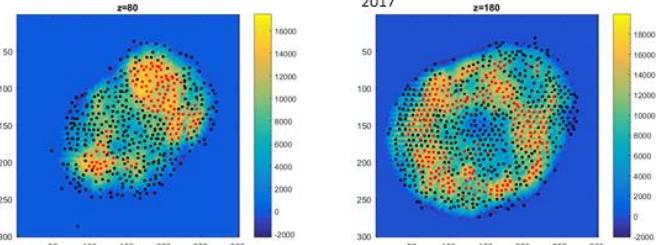
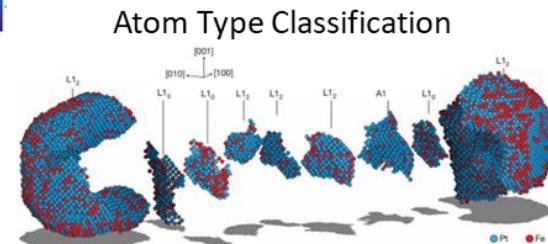
Grain Boundary Detection



- 晶粒邊界辨識、結構分群，了解區塊結構相似度。

相較於2D繞射結果分析，X光3D斷層掃描重建原子結構讓科學家可更直觀的觀察原子排列特徵，為近年來材料領域相當熱門之議題。透過機器學習技術，建立分析流程應用於材料之原子級二維、三維斷層攝影影像的原子座標定位、種類辨識、與晶粒邊界辨識等相關研究工作，達成原子結構分群，以進行材料缺陷型態、區塊結構相似度、與原子種類分佈等探討，避免人為誤判進而提升整體研究效率。將有助於凝態物理、化學、材料科學、生醫材料的研發與應用。

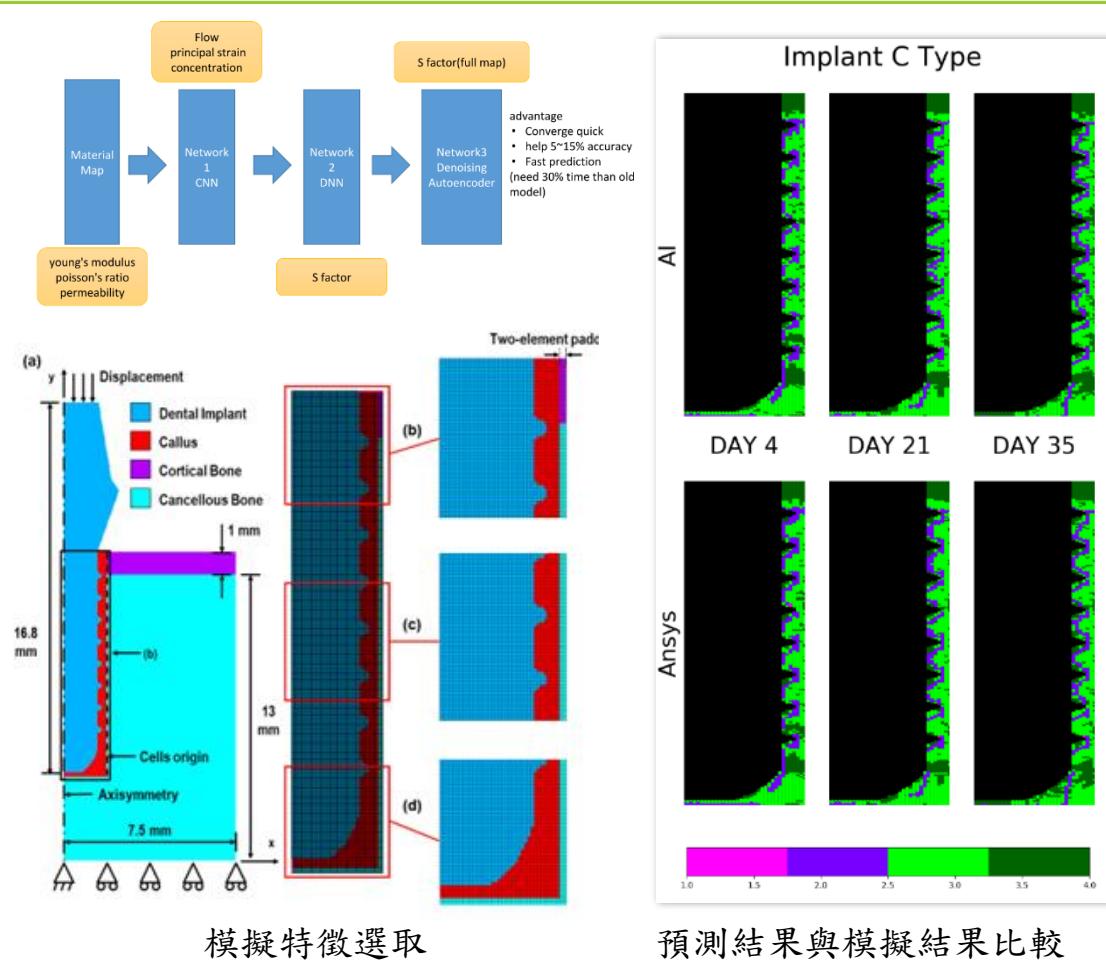
- 原子座標定位、種類辨識，進行原子種類分佈分析。



解析材料缺陷型態與相似度分析。

機器學習協助工程設計

機器學習加速工程結構設計



問題描述：理論計算雖成本較低，但實際模擬單一案例所需耗時至少4小時以上(>24核心計算力)，對於工程最佳化設計而言，總體計算需求將更為可觀。

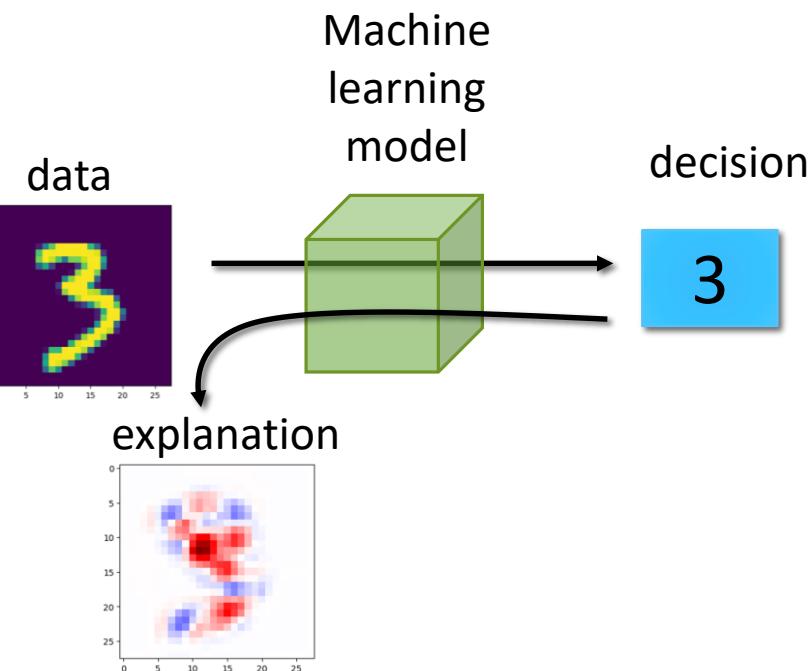
成果：使用多個神經網路取代有限元素求解器，可快速預測材料之力學與生物力學等性質。與理論計算結果相較，預測準確率可達90%以上，加速預測時間至少約36,000倍(ex. 24core)。



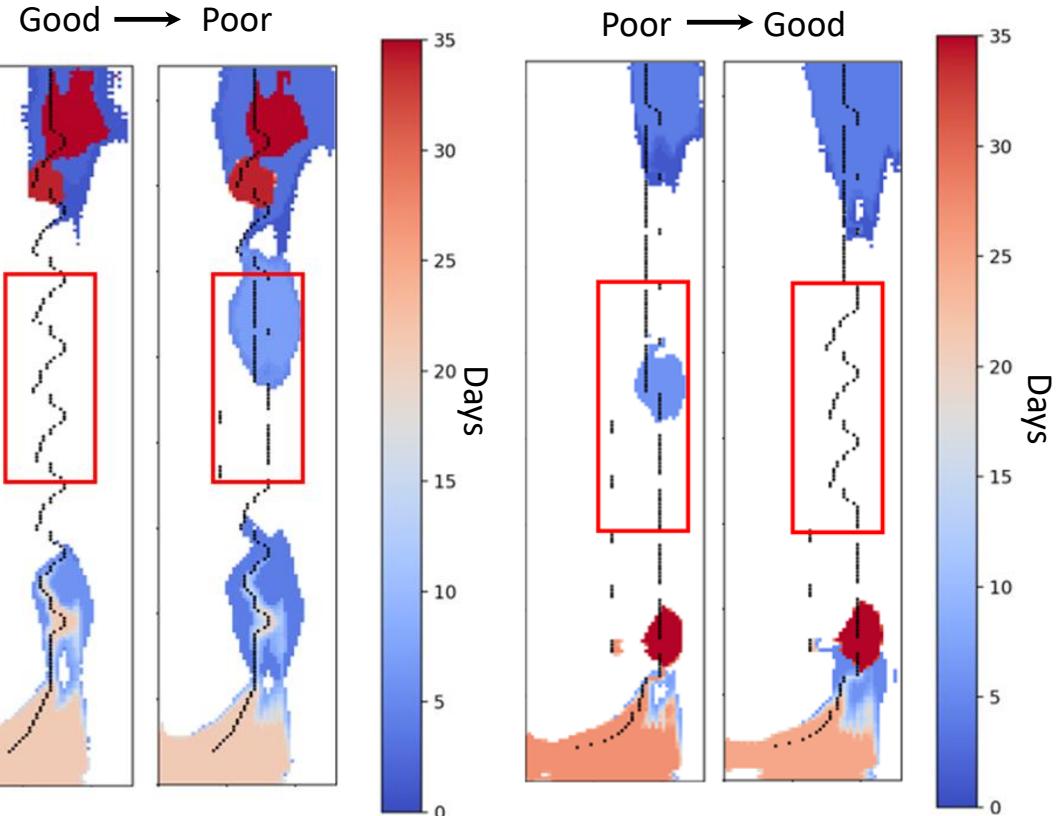
利用AI引導結構設計方針

機器學習協助工程設計

骨釘植體構型設計最佳化，歸因與敏感性分析



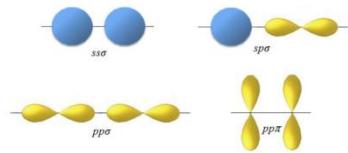
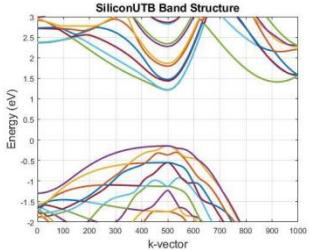
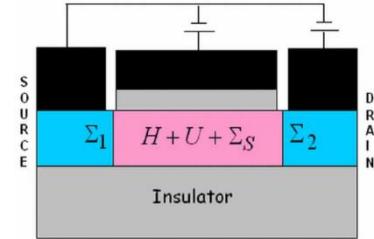
打開機器學習的黑盒子，重新了解問題的本質



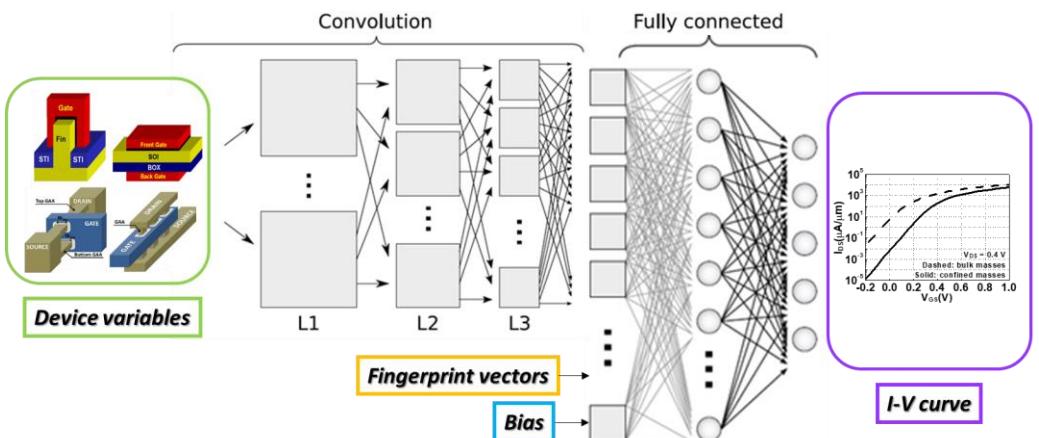
特徵的敏感度與歸因分析，得到過往量測或模擬不易獲得的關鍵因子。

機器學習協助工程設計

機器學習加速電子元件設計



Tight-binding model
The non-equilibrium Greens function & Poisson equation



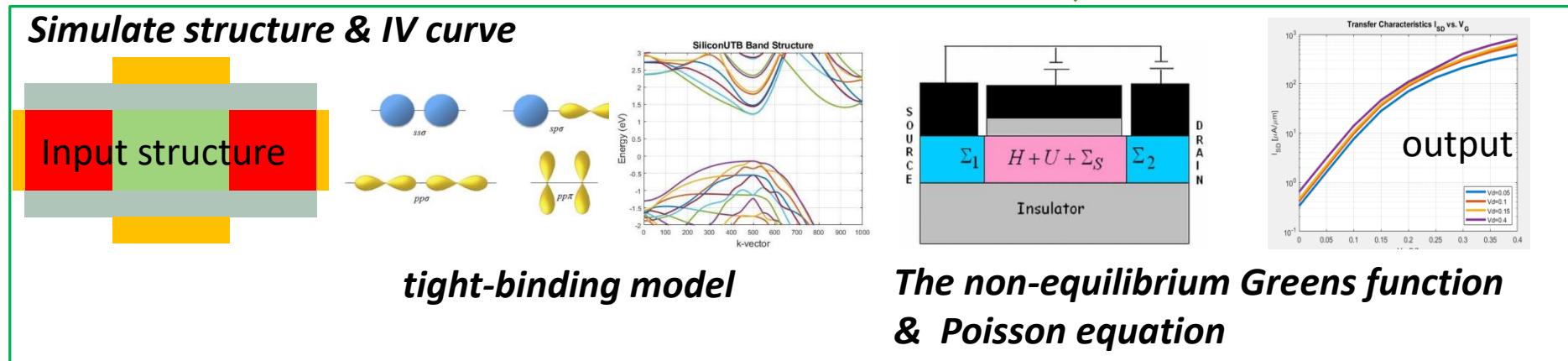
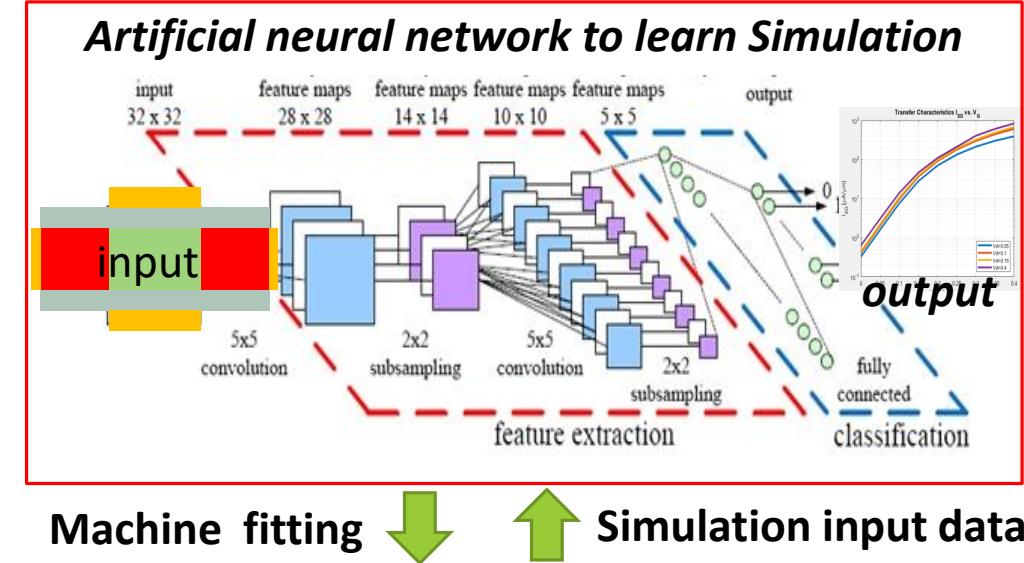
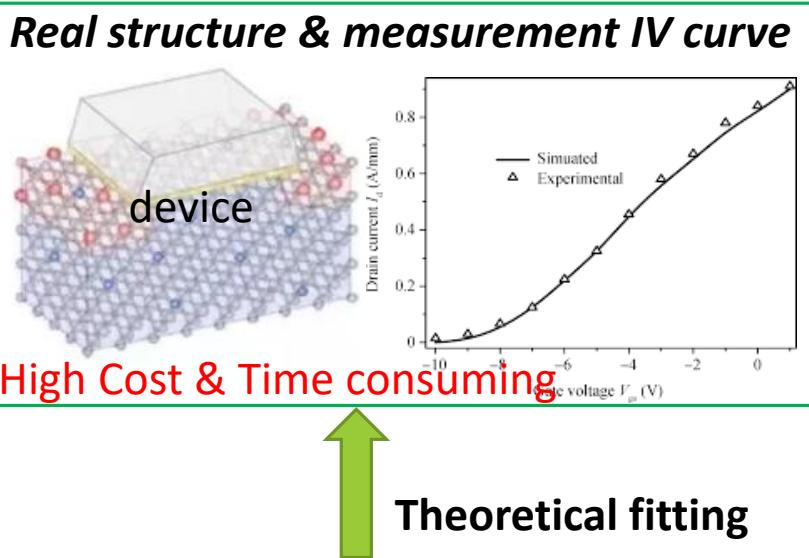
問題描述：半導體元件傳輸性質預測是基於緊束縛(tight-binding)與量子傳輸(NEGF)理論計算，因為求解涉及自洽過程，元件模擬計算的時間相當可觀。

成果：利用過往累積之計算資料，訓練神經網路模型取代傳統求解計算，在元件特徵曲線預測模型，預測準確率達95%以上，可取代現有精準的量子計算物理引擎，加速預測時間170倍。

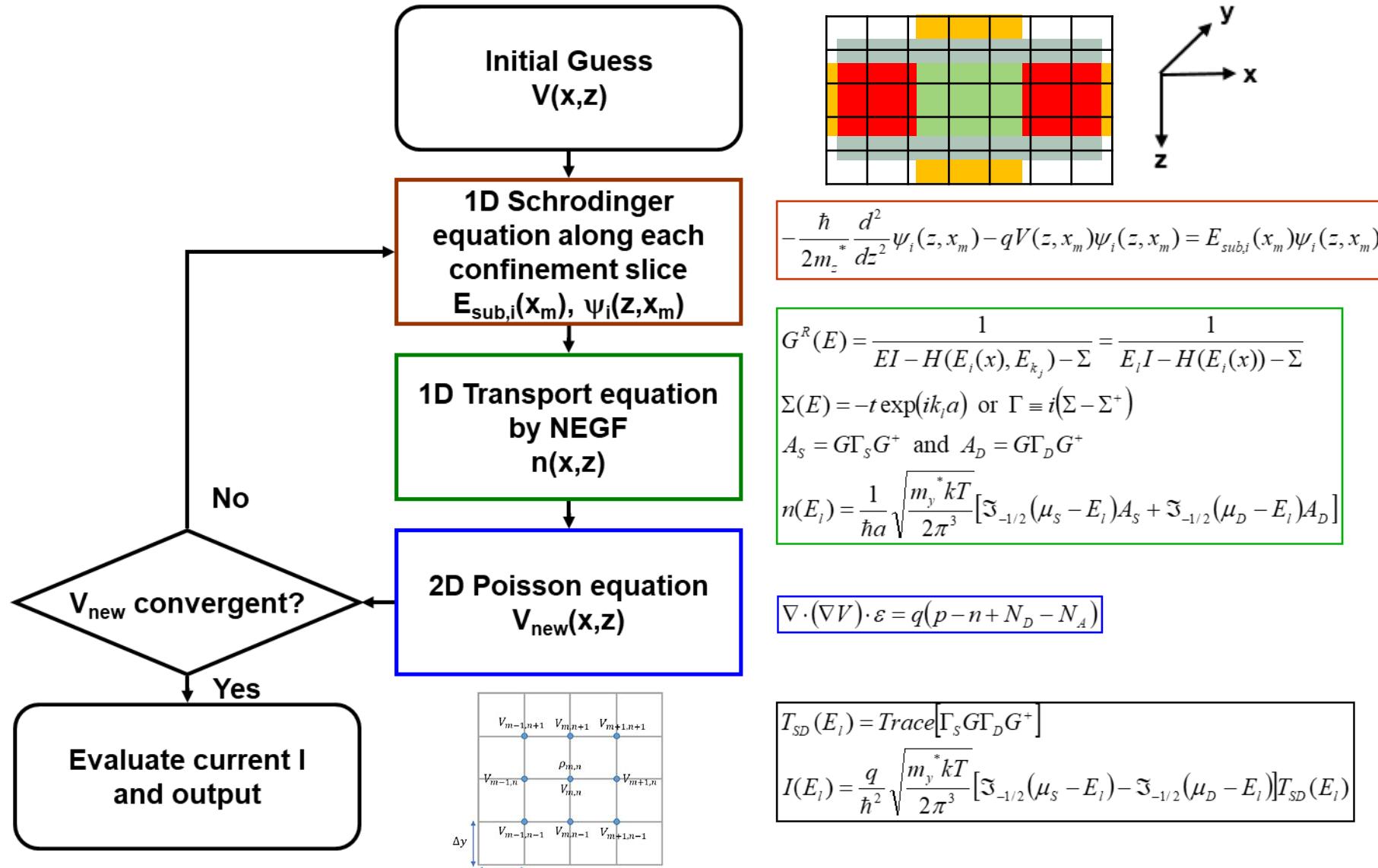


加速模擬計算

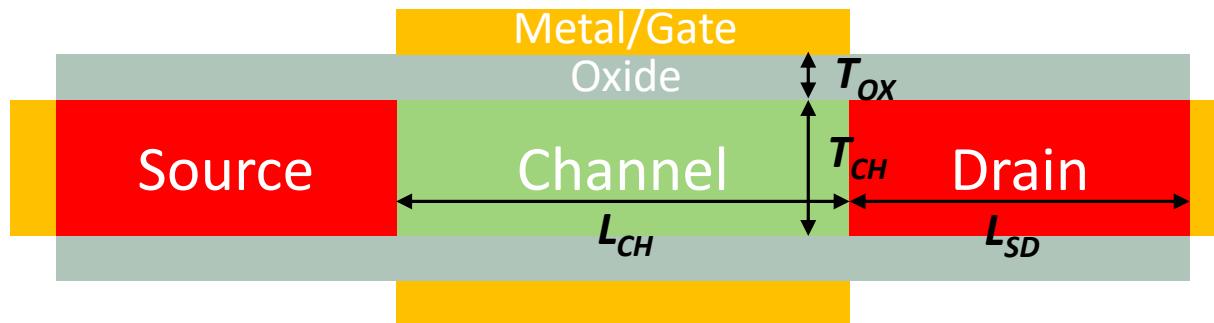
Machine Learning Assisted Non-Equilibrium Green's Function Simulations of Double-Gate nMOSFET



NEGF simulations workflow

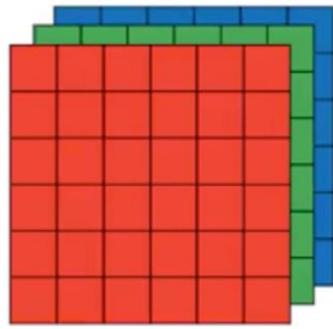


Parameters of simulation



- Gate work function: 4.22 eV (metal)
- Channel length (L_{CH}): 8~30 nm
(small: leakage current; long: scattering)
- Channel thickness (T_{CH}): 3~10 nm
- Channel length/thickness > 3 (short channel effect)
- Source/Drain length(L_{SD}): 5 nm (5~10 nm)
- Metal thickness: 0 nm (1 nm)
- Oxide thickness(T_{ox}): 0.8 nm (0.6 ~ 3 nm)

CNN input matrix



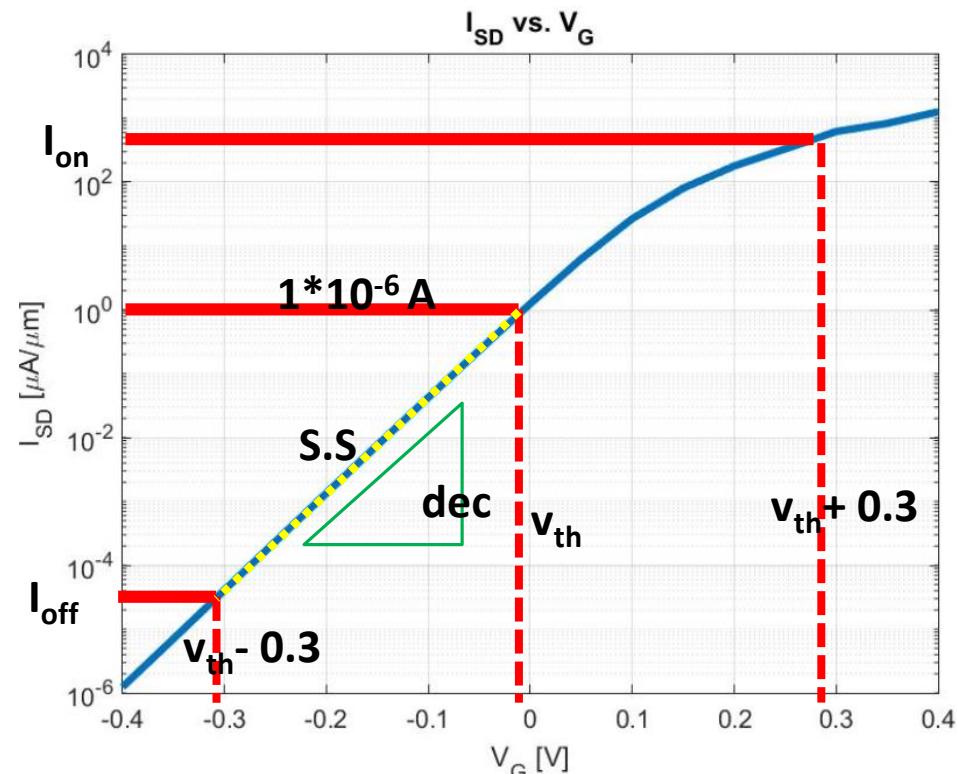
doping

dielectric
constant

work function / electron affinity

Output: Id-Vg analysis

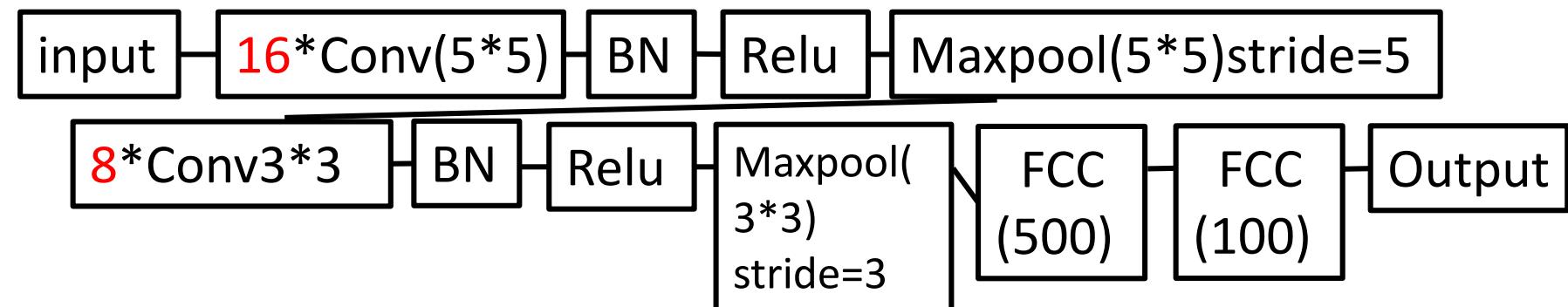
- Set $I_d = 1 \times 10^{-6}$ (A) as v_{th}
- Set $v_{th} + 0.3$ as I_{on}
- $v_{th} - 0.3$ as I_{off}
- Calculate I_{on}/I_{off}
- Calculate S.S between v_{th} to $v_{th} - 0.3$
- Total output: 5

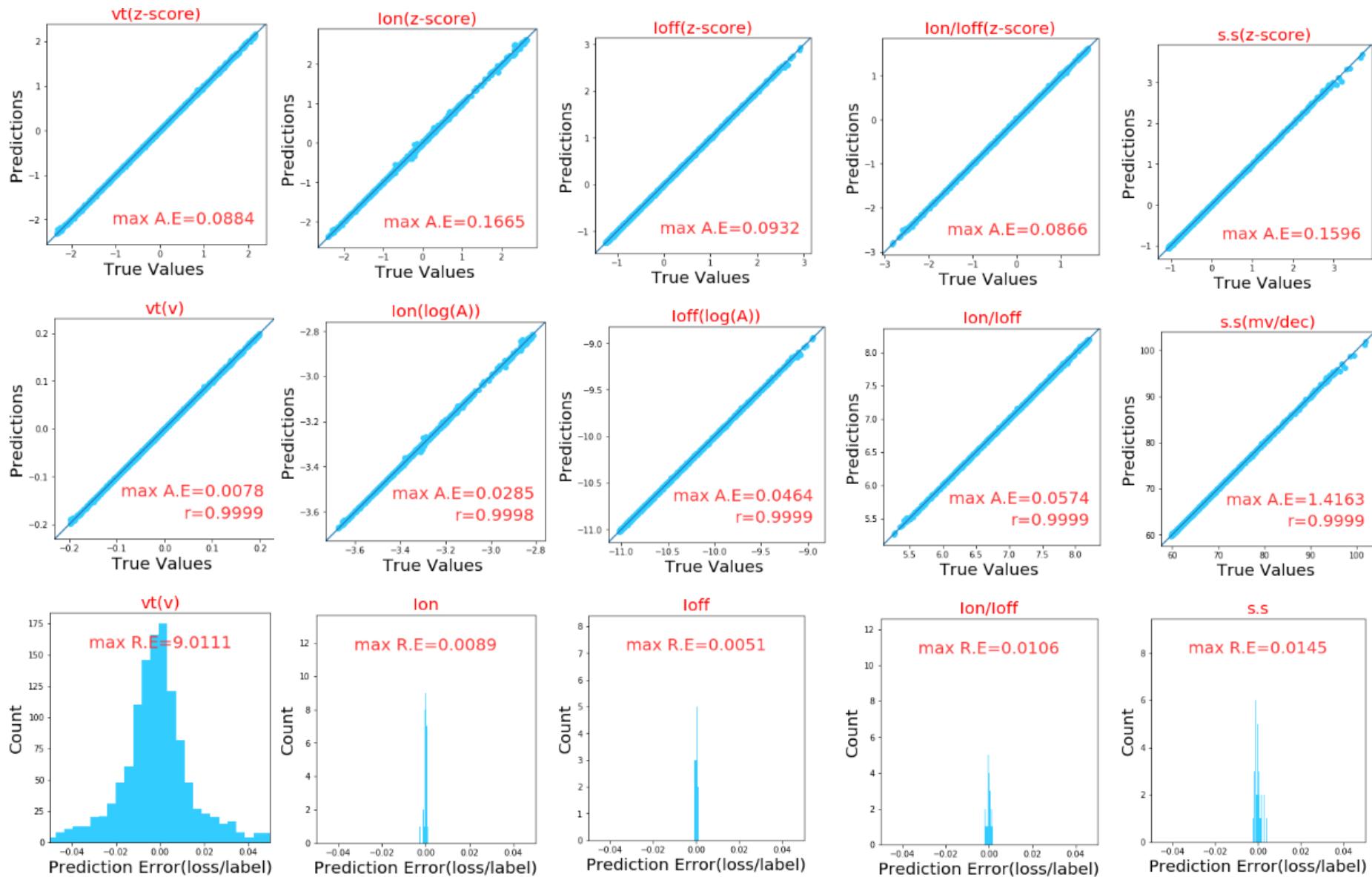


	A	B	C	D	E
1	vt(V)	Ion[log(I)]	Ioff[log(I)]	Ion/off[log(I)]	S.S[mv/dec]
2	-0.00678	-3.234268169	-10.490611	7.256342825	87.63822008

CNN結構設計

- Training data: 13288 * 0.8
- Validation data: 13288 * 0.1
- Testing data: 13288 * 0.1
- Optimizer: Adam (learning rate=0.003)
- Active function: Relu
- Batch Normalization: used
- Batch size: 1024
- Epoch: 5000
- Conv stride: 1
- Padding: 'SAME'





Train different models

CONV1 s	pool1	CONV2	pool2	Fcc	Batch size
8*(5*5)	5*5(stride=5)	16*(3*3)	3*3(stride=3)	1000-300-100-30	512
8*(5*5)	5*5(stride=5)	8*(3*3)	3*3(stride=3)	500-150-30	512
4*(5*5)	5*5(stride=5)	16*(3*3)	3*3(stride=3)	1000-300-100-30	512
4*(5*5)	5*5(stride=5)	8*(3*3)	3*3(stride=3)	500-150-30	1024

Results in different models

	vt	lon	loff	lratio	ss	evg.	min train loss	min test loss
ANN 30-20-20	0.99998	0.999837	0.999985	0.999983	0.999968	0.999950716	5.2*e-5	3.6*e-4
ANN 30-40-20	0.999988	0.999858	0.999992	0.999985	0.999982	0.999961166	8.2*e-5	3.6*e-4
ANN 40-30-20	0.999987	0.999893	0.99999	0.999985	0.999984	0.99996778	3.3*e-5	2.9*e-4
ANN 40-50-30	0.999998	0.999887	0.999998	0.999991	0.999995	0.9999738	3.3*e-5	2.9*e-4
ANN 50-40-30	0.999992	0.999863	0.999993	0.999983	0.999987	0.99996357	3.2*e-5	2.8*e-4
ANN 50-50-30	0.999992	0.999792	0.999993	0.99998	0.999988	0.99994894	2.2*e-5	3.1*e-4
ANN 40-50-30(1-2)	0.999983	0.999688	0.999986	0.999965	0.999969	0.9999183	2.7*e-6	1.6*e-5
ANN 50-40-30(1-2)	0.999985	0.999388	0.999999	0.999933	0.999975	0.9998561	1.2.e-6	1.0*e-5
CNN 8-16	0.999994	0.999602	0.999992	0.999962	0.999987	0.99990705	1.1*e-5	1.1*e-4
CNN 4-16	0.99994	0.999524	0.999973	0.999948	0.999963	0.99986934	1.0*e-5	1.2*e-4
CNN 8-8	0.999993	0.999926	0.999939	0.999991	0.999939	0.99995726	1.9*e-5	6.3*e-5
CNN 4-8	0.999938	0.999639	0.999954	0.999935	0.999947	0.99988244	3.1*e-5	1.1*e-4
CNN 4-8 noBN	0.999993	0.999947	0.999995	0.999993	0.999995	0.999984734	5.2*e-6	1.2*e-4

Summary

- Simulation tools can help scientists to understand the fundamental physics and mechanism.
- Material informatics combined with simulations as a virtual screen technique would help the scientists to find/create the next wonder materials.
- Machine learning can help scientists to predict the properties in material space without exploring all of it.
- Machine learning can **boost the new material findings and reveal the insight between physical properties of materials.**

資訊化的時代來臨

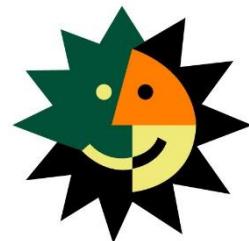
AI浪潮將會像過去的E化一樣，未來每一個環節都會用到AI



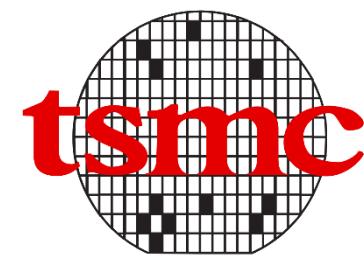
台塑企業
FORMOSA PLASTICS GROUP



中錠公司



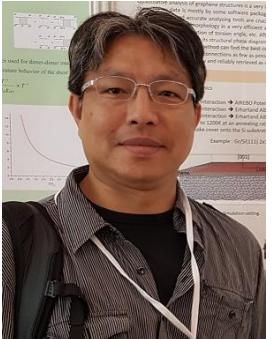
ASE GROUP
日月光集團



台灣積體電路製造股份有限公司
Taiwan Semiconductor Manufacturing Company, Ltd.

關於我們

材料小組(Material Research Team)



陳南佑 博士

清華大學物理博士

李致頤 博士

中山大學機械博士

楊安正 博士

陳冠朋

中央大學應數碩士

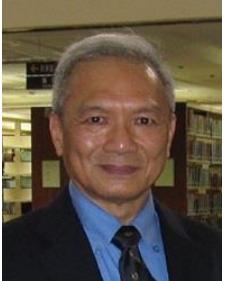
材料計算
半導體量子傳輸模擬
大尺度計算與GPU計算
機器學習與深度學習
影像資訊學、生物物理

微觀模擬
材料計算與材料資訊
平行計算
半導體量子傳輸模擬
機器學習與深度學習

微觀模擬
材料計算與材料資訊
平行計算
機器學習與深度學習
系統整合

統計分析
機器學習與深度學習

重要合作夥伴



交大材料
杜經寧院士



成大電機
高國興教授
哥倫布計畫主持人



清大材料
葉安洲教授
高熵材料整合型計畫
主持人



清大腦科中心
江安世院士



交大半導體學院
吳添立教授
愛因斯坦計畫主持人



食品工業發展研究所
盧懷恩博士
智慧仿生整合型計畫
主持人



中研院物理所
李定國院士



成大電機
盧達生教授
AI專案計畫主持人



交大材料
鄒年棣教授



國家實驗研究院
翁政義董事

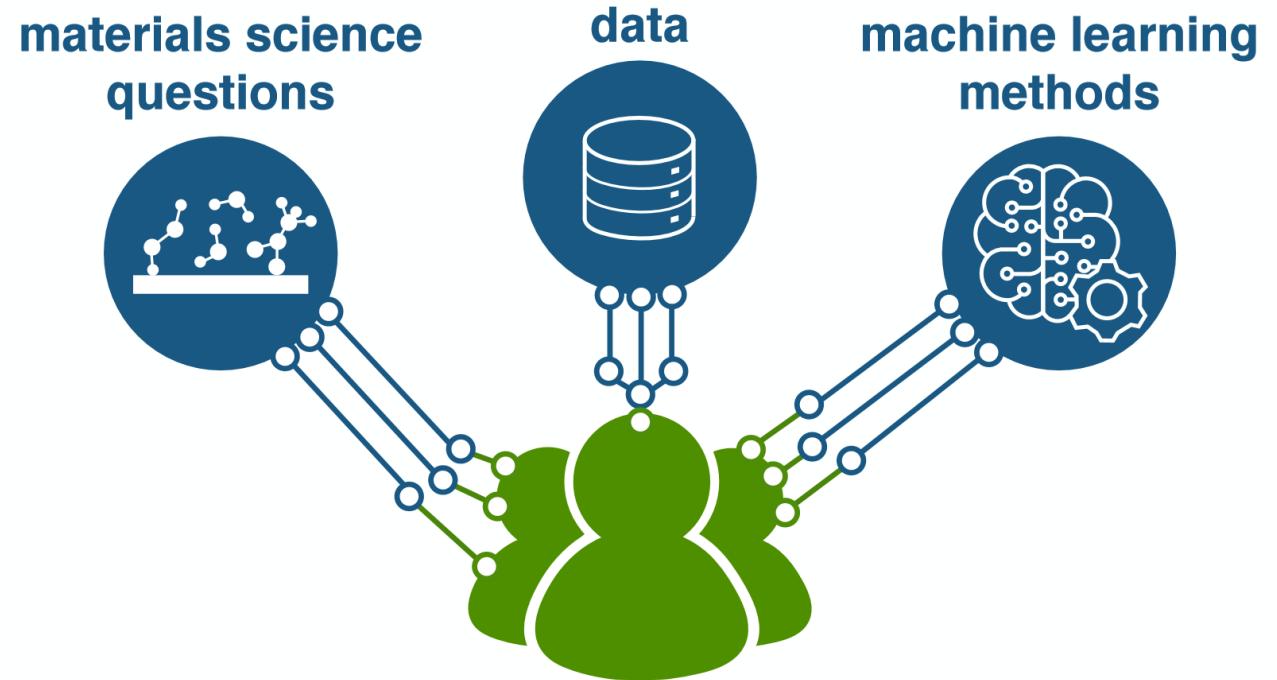


東海物理系
施奇廷教授



清大工科
陳健群教授

*Thanks for your
attention*

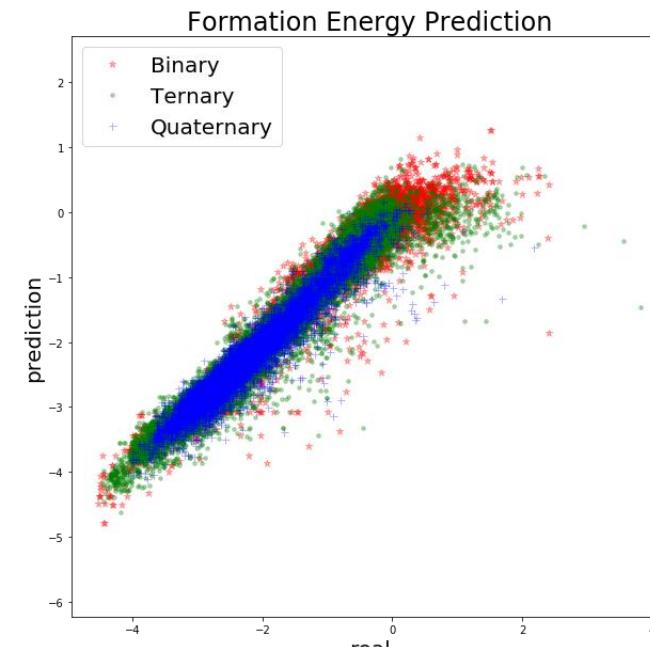


National Applied Research Laboratories

National Center for
High-performance Computing

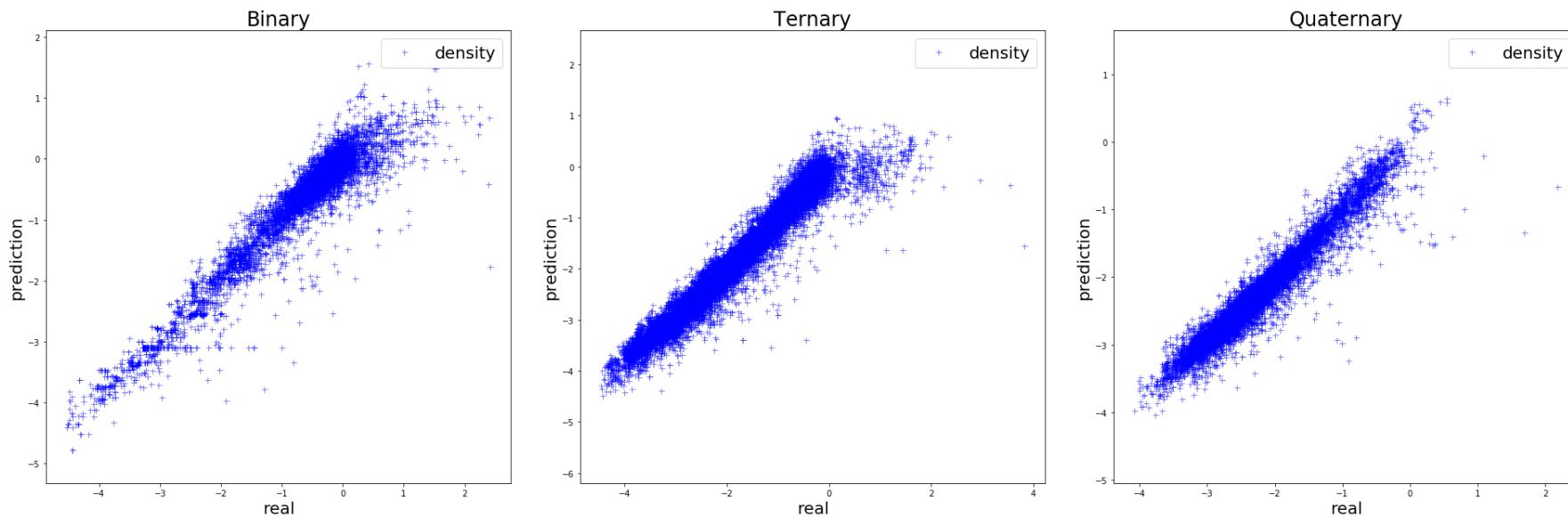
Formation Energy Prediction (w/o crystal Info)

ML Setting	
Data	<ul style="list-style-type: none"> 52959 materials from Material Project with complete [cohesive energy] 70% (37071) for training 30% (15888) for testing
Feature Engineering	<ul style="list-style-type: none"> Features: weighting sum of <ol style="list-style-type: none"> atomic fraction atomic mass electronegativity melting point 5). crystal structure Apply robust scaling
ML Model	<ul style="list-style-type: none"> Multi-Layer Perceptron (MLP) Hidden layer arch: 10 Activation functions: relu Solver: lbfgs
Evaluation	<ul style="list-style-type: none"> MAE (mean absolute error) Pearson's correlation 5-fold cross-validation /r-squared



	Metal/Alloy	Complete Information
Unary	223	220
Bin.	9221	8975
Ter.	29134	28374
Quarter.	16299	15390
Total	54877	52959

Formation Energy Prediction (w/o crystal Info)



Experimental Results

	Train Set	Test Set	Binary	Ternary	Quaternary
Pearson's R	0.9777	0.9773	0.9605	0.9760	0.9665
R-Square	0.9559	0.9551	0.9224	0.9524	0.9341
MAE	0.1513	0.1522	0.1916	0.1576	0.1153

5-fold Cross Validation

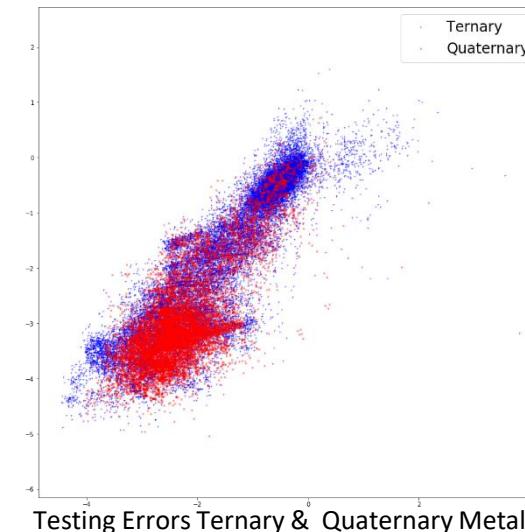
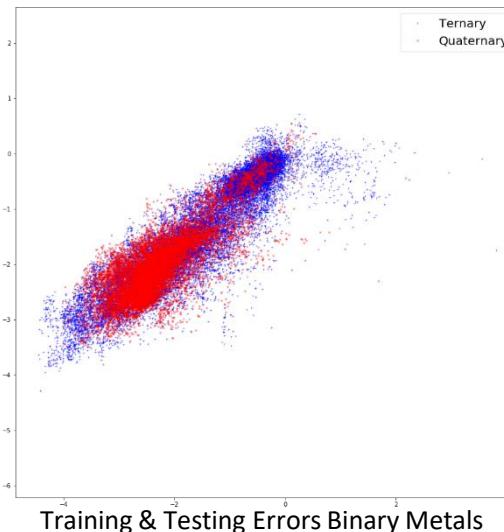
Fold	1	2	3	4	5
R-Square	0.9533	0.9514	0.9500	0.9457	0.9503

Formation Energy Prediction: Train with Binary only

Dataset

- filter for metals
- filter mts with incomplete
 - 1). cohesive energy 2). Atomic radius
- 8975 bin, 28351 ter, 15363 qua

Data Set	Materials
Train Set	7180 binary (80%)
Test Set 1	1795 binary (20%)
Test Set 2	28351 ternary, 15363 quaternary



Network Info

- arch : (20,)
- activation function : relu
- alpha = 20.0

Feature Info

- atomic ratio (row, group based)
- crystal system
- cohesive energy, atomic mass, atomic radius, youngs_modulus, electronegativity, melting point

Comparison of Training & Testing Errors

	Train Set (7180 binary)	Test Set 1 (1795 binary)	Test Set 2 (Ternary)	Test Set 2 (Quaternary)
Pearson's R	0.9671	0.9639	0.8947	0.7647
R-Square	0.9348	0.9288	0.6826	-0.6262
MSE	0.1727	0.1902	0.4361	0.7475

Train with Binary Only

Features Used	Stats	Train Set (7180 binary)	Test Set 1 (1795 binary)	Test Set 2 (Ternary)	Test Set 2 (Quaternary)
Atomic ratio + Crystal system	Pearson's R	0.9671	0.9639	0.8947	0.7647
	R-Square	0.9348	0.9288	0.6826	-0.6262
	MSE	0.1727	0.1902	0.4361	0.7475
Atomic ratio + Crystal system + [cohesive energy, atomic mass, atomic radius, young's modulus, electronegativity, melting point]	Pearson's R	0.9393	0.9405	0.9129	0.8063
	R-Square	0.8805	0.8824	0.8023	0.5053
	MSE	0.2475	0.2588	0.3677	0.3752

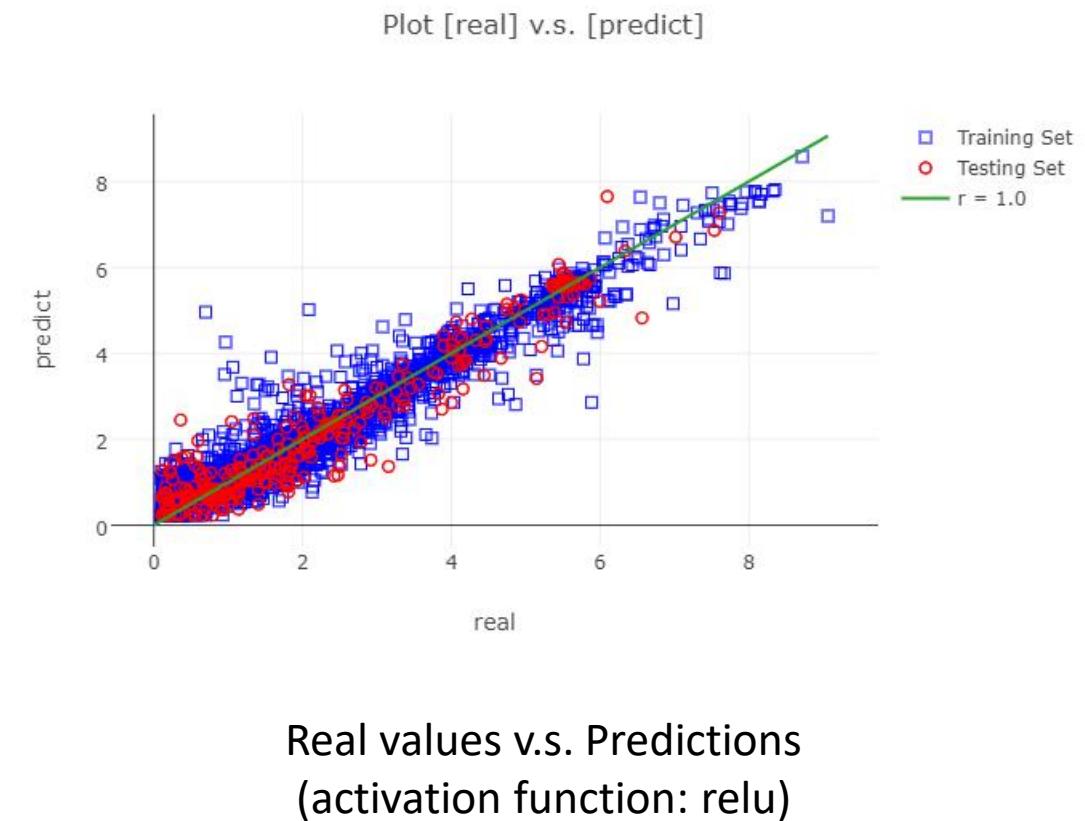
Train with 3k n-nary for each n-nary

Features Used	Stats	Train Set (3k n-nary)	Test Set 1 (7975 binary)	Test Set 2 (25351 Ternary)	Test Set 2 (12363 Quaternary)
Atomic ratio + Crystal system	Pearson's R	0.9580	0.9316	0.9487	0.9269
	R-Square	0.9168	0.8666	0.8992	0.8582
	MSE	0.2231	0.2641	0.2437	0.1777
Atomic ratio + Crystal system + [cohesive energy, atomic mass, atomic radius, young's modulus, electronegativity, melting point]	Pearson's R	0.9755	0.9582	0.9673	0.9530
	R-Square	0.9512	0.9177	0.9355	0.9080
	MSE	0.1688	0.2027	0.1929	0.1446

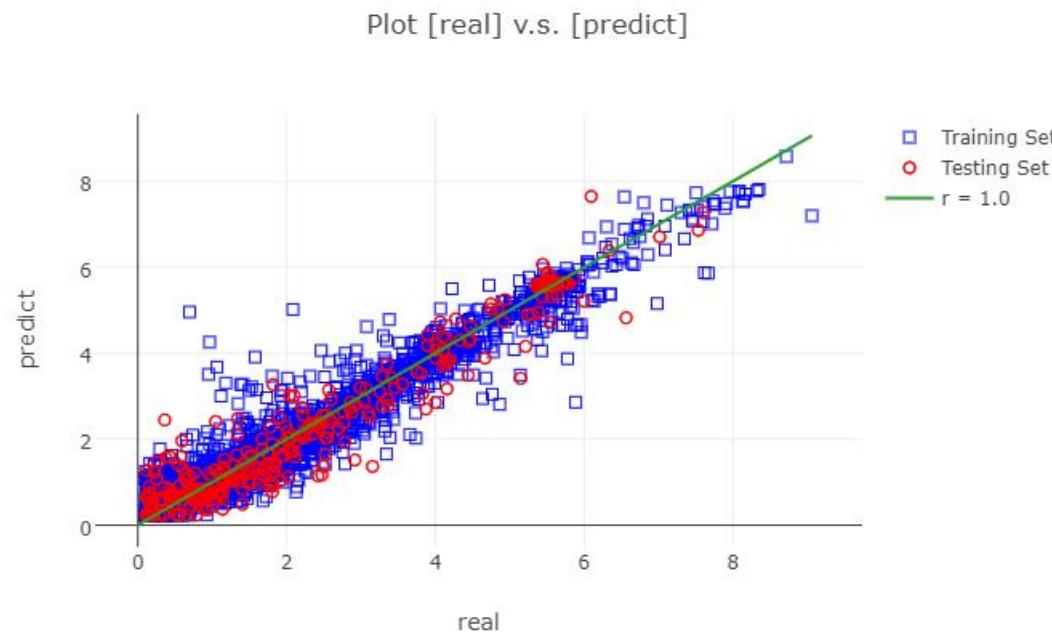
Band Gap Prediction: (band gap > 0.1)

ML Settings

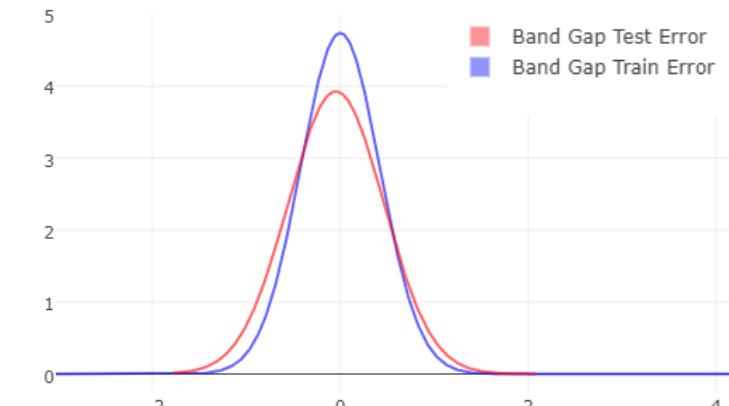
Data	<ul style="list-style-type: none">3124 binary composition from Material Project (band gap > 0.1)90% (2811) for training 10% (313) for testing
Feature Engineering	<ul style="list-style-type: none">Features: atomic fraction, density , energy per atom, formation energy per atomApply robust scaling
ML Model	<ul style="list-style-type: none">Multi-layer Perceptron (MLP)Hidden layer arch: 20, 10, 5Activation functions: reluSolver: lbfgs
Evaluation	<ul style="list-style-type: none">MAE (mean absolute error)Pearson's correlation5-fold cross-validation/r-squared



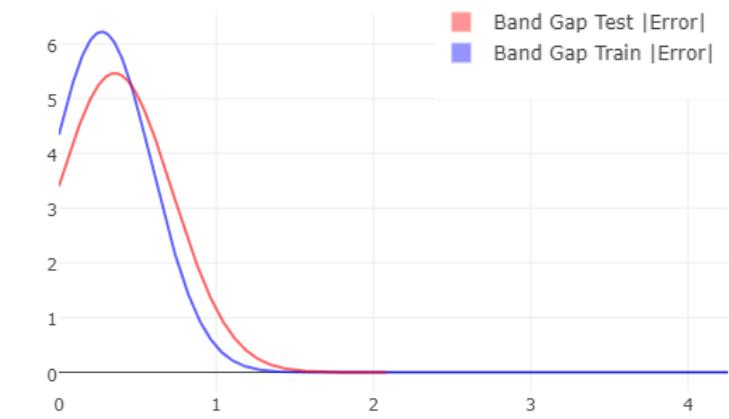
Band Gap Prediction: (band gap > 0.1)



	MAE	Pearson's Coefficient	Percent of Predictions within Relative Error of			
			5%	10%	20%	30%
Training Set	0.272	0.971	0.326	0.473	0.646	0.754
Testing Set	0.355	0.958	0.323	0.467	0.641	0.749



Error Distribution



Absolute Error Distribution

Band Gap Prediction: (band gap > 0.1)

Discussions

- **relu** made best generalization (according to cross-validation)
- **relu** leads best result with testing set
- **overfitting** with **tanh, logistic**

5-Fold Cross-Validation (r-squared)

	identity	relu	logistic	tanh
Fold-1	0.734502	0.801586	0.553183	0.547325
Fold-2	0.755639	0.835816	0.612451	0.657760
Fold-3	0.686981	0.783382	0.632355	0.632988
Fold-4	0.713144	0.827609	0.574432	0.559907
Fold-5	0.737791	0.783864	0.773284	0.581522
AVG	0.725611	0.806451	0.629140	0.595900

Experimental Results - Training

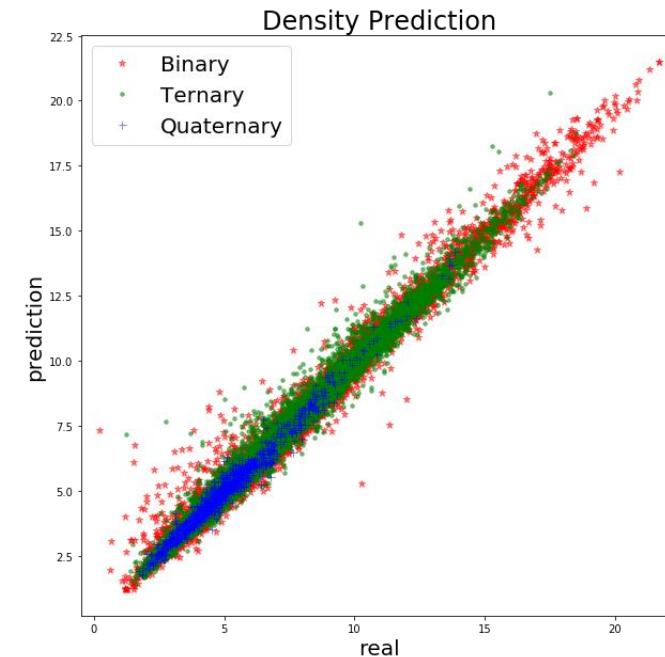
Activation Function	Mean Absolute Error (MAE)	Pearson's Correlation
identity	0.621059	0.869738
relu	0.212171	0.983325
logistic	0.146818	0.990581
tanh	0.122655	0.993558

Experimental Results - Testing

Activation Function	Mean Absolute Error (MAE)	Pearson's Correlation
identity	0.700814	0.824414
relu	0.455767	0.918947
logistic	0.695141	0.764507
tanh	0.590612	0.842838

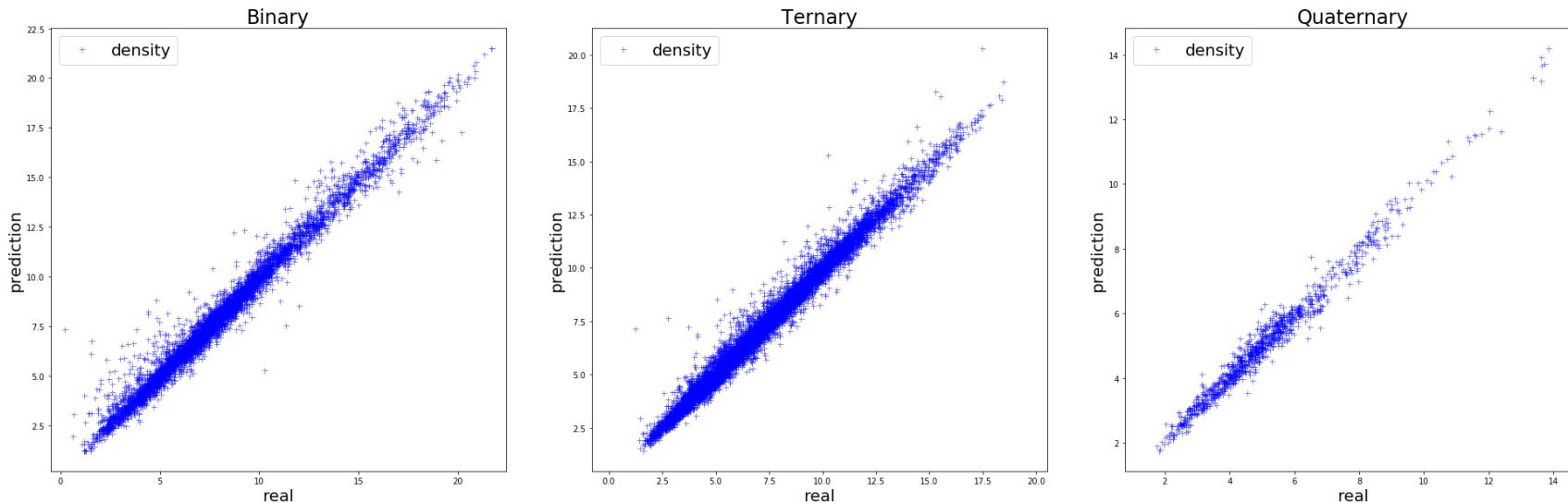
Alloy Density Prediction

ML Setting	
Data	<ul style="list-style-type: none"> 19173 materials from MP with complete <ol style="list-style-type: none"> cohesive energy density of solid 70% (13421) for training 30% (5752) for testing
Feature Engineering	<ul style="list-style-type: none"> Features: weighting sum of <ol style="list-style-type: none"> atomic fraction atomic mass density of solid electronegativity melting point 6)-crystal structure Apply robust scaling
ML Model	<ul style="list-style-type: none"> Multi-layer Perceptron (MLP) Hidden layer arch: 20, 10, 5 Activation functions: relu Solver: lbfgs
Evaluation	<ul style="list-style-type: none"> MAE (mean absolute error) Pearson's correlation 5-fold cross-validation/r-squared



	Metal/Alloy	Complete Information
Unary	223	210
Bin.	9221	6528
Ter.	29134	11377
Quarter.	16299	1058
Total	54877	19173

Alloy Density Prediction



Numerical Results

	Train Set	Test Set	Binary	Ternary	Quaternary
Pearson's R	0.9911	0.9884	0.9896	0.9905	0.9882
R-Square	0.9824	0.9769	0.9793	0.9810	0.9764
MAE	0.2760	0.3102	0.3277	0.2666	0.2167

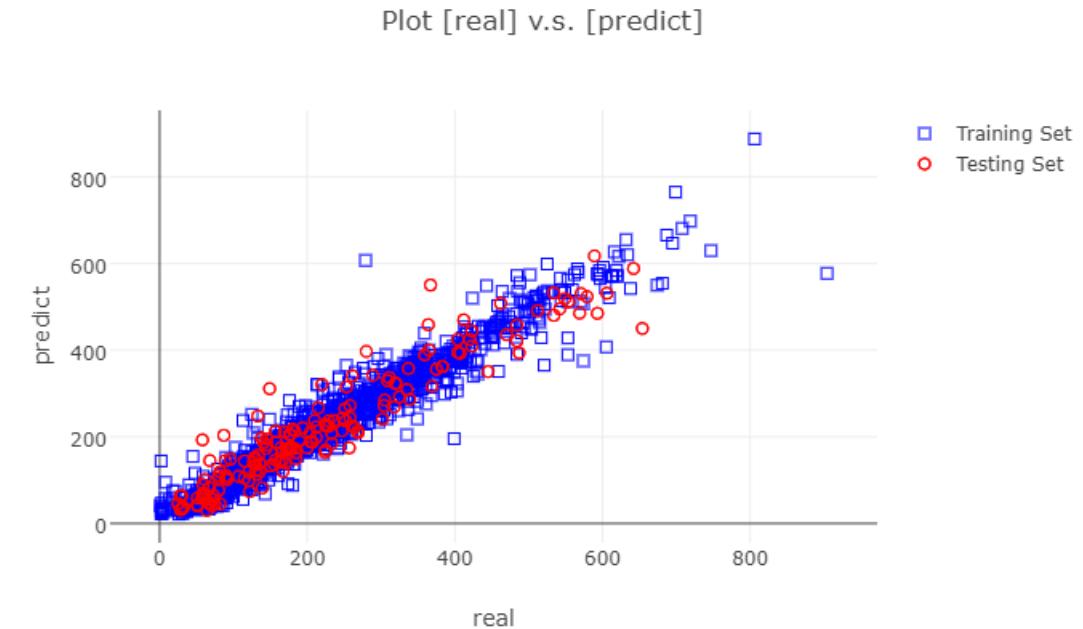
5-fold Cross Validation

Fold	1	2	3	4	5
R-Square	0.9804	0.9758	0.9773	0.9753	0.9803

Stiffness Tensor Prediction

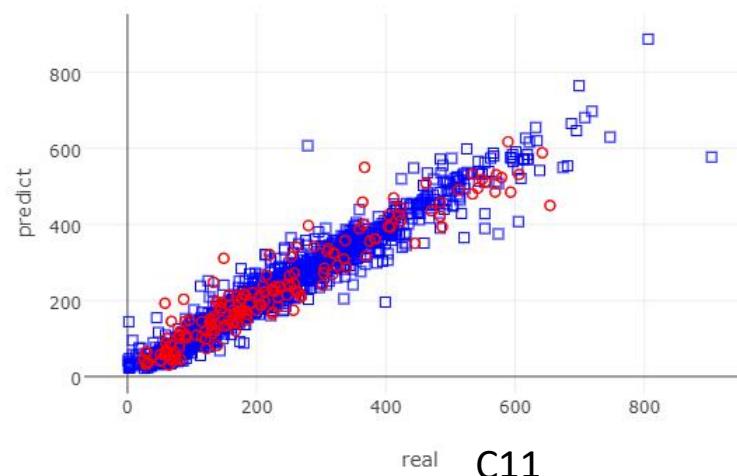
ML Settings

Data	<ul style="list-style-type: none">1675 binary composition from Material Project (with band gap < 0.5 & $0 <$ Poisson's ratio < 0.5)90% (1507) for training 10% (168) for testing
Feature Engineering	<ul style="list-style-type: none">Features: atomic fraction, density , energy per atom, formation energy per atomApply robust scaling
ML Model	<ul style="list-style-type: none">Multi-layer Perceptron (MLP)Hidden layer arch: 20, 10, 5Solver: lbfgsOutput: stiffness tensor C11, C22, C33
Evaluation	<ul style="list-style-type: none">MAE (mean absolute error)Pearson's correlation

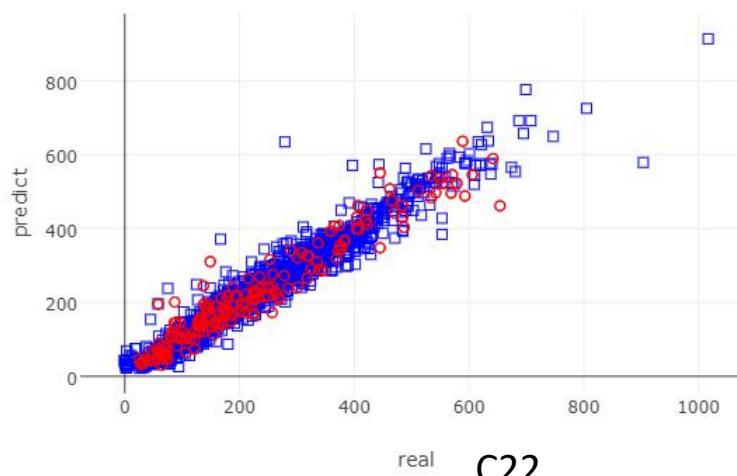


Real values v.s. Predictions:
Elastic Tensor – C11
(activation function: relu)

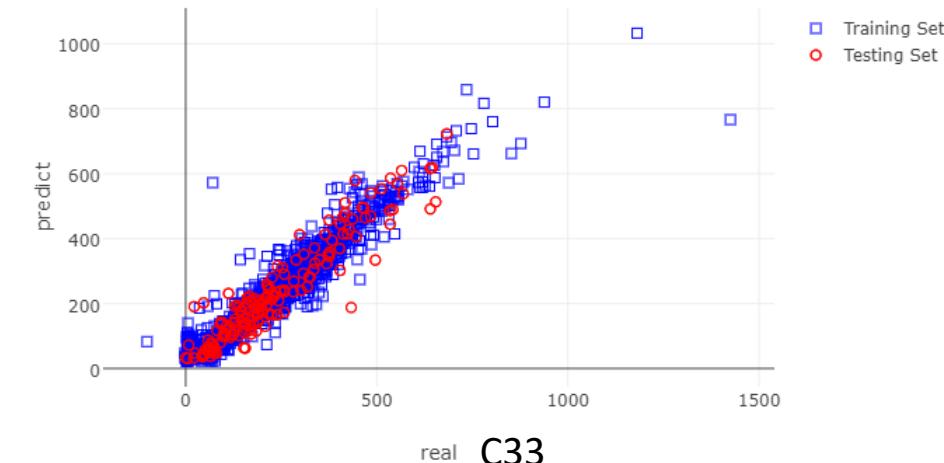
Stiffness Tensor Prediction



C11



C22



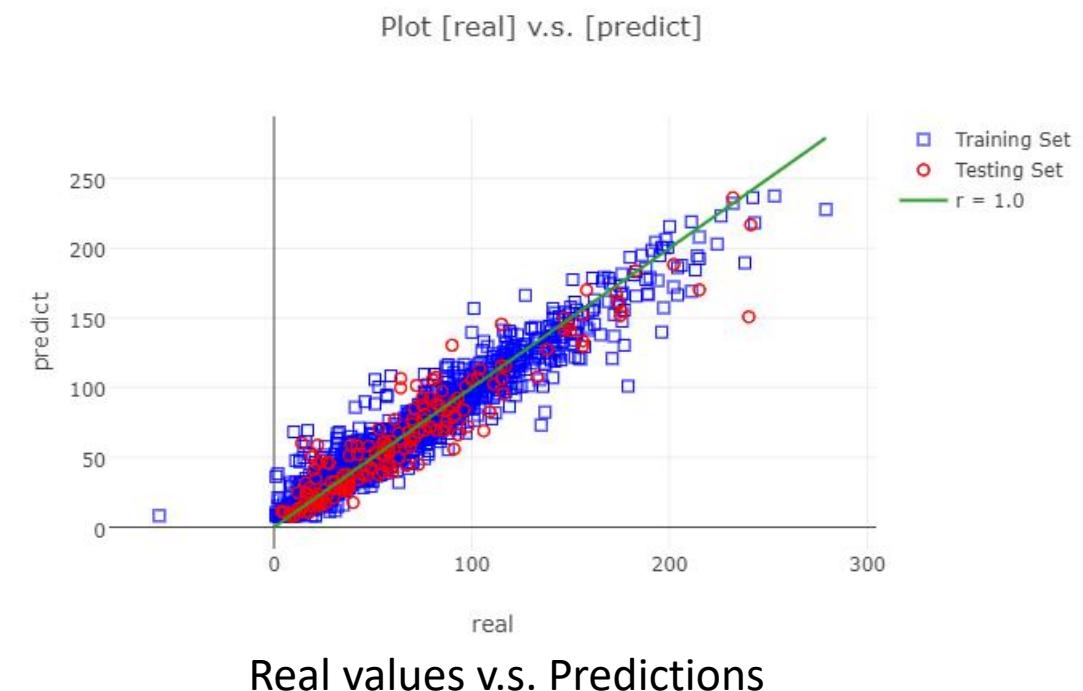
C33

	Stiffness Tensor	MAE	Pearson's Coefficient	Percent of Predictions within Relative Error of			
				5%	10%	20%	30%
Training Set	C11	21.349	0.969	0.365	0.589	0.822	0.914
	C22	20.884	0.972	0.361	0.595	0.837	0.927
	C33	26.760	0.955	0.311	0.545	0.777	0.879
Testing Set	C11	33.764	0.951	0.178	0.386	0.684	0.809
	C22	30.028	0.963	0.214	0.452	0.750	0.869
	C33	36.747	0.944	0.214	0.416	0.648	0.797

Shear Modulus Prediction (G_VRH)

ML Settings

Data	<ul style="list-style-type: none">1675 binary composition from Material Project (with band gap < 0.5 & 0 < Poisson's ratio < 0.5)90% (1507) for training 10% (168) for testing
Feature Engineering	<ul style="list-style-type: none">Features: atomic fraction, density , energy per atom, formation energy per atomApply robust scaling
ML Model	<ul style="list-style-type: none">Multi-layer Perceptron (MLP)Hidden layer arch: 20, 10, 5Activation functions: relu Solver: lbfgs
Evaluation	<ul style="list-style-type: none">MAE (mean absolute error)Pearson's correlation



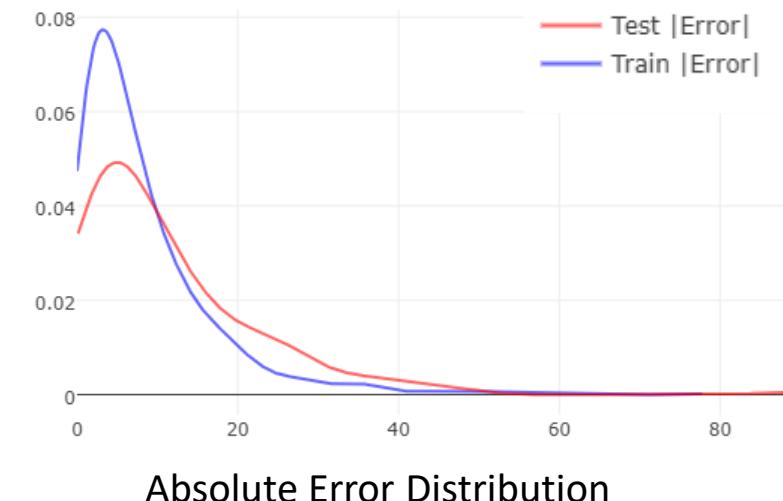
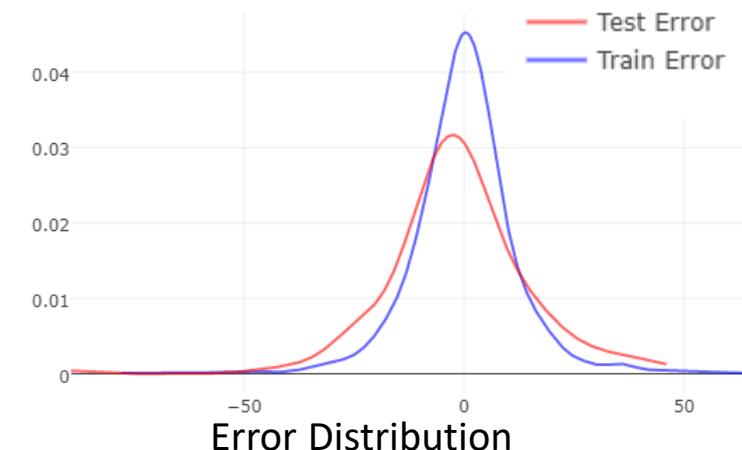
Shear Modulus Prediction (G_VRH)

Percent of Predictions within Relative Error of

	5%	10%	20%	30%
Training Set	0.238	0.448	0.714	0.831
Testing Set	0.232	0.442	0.704	0.823

MAE & Pearson's Correlation of Predictions

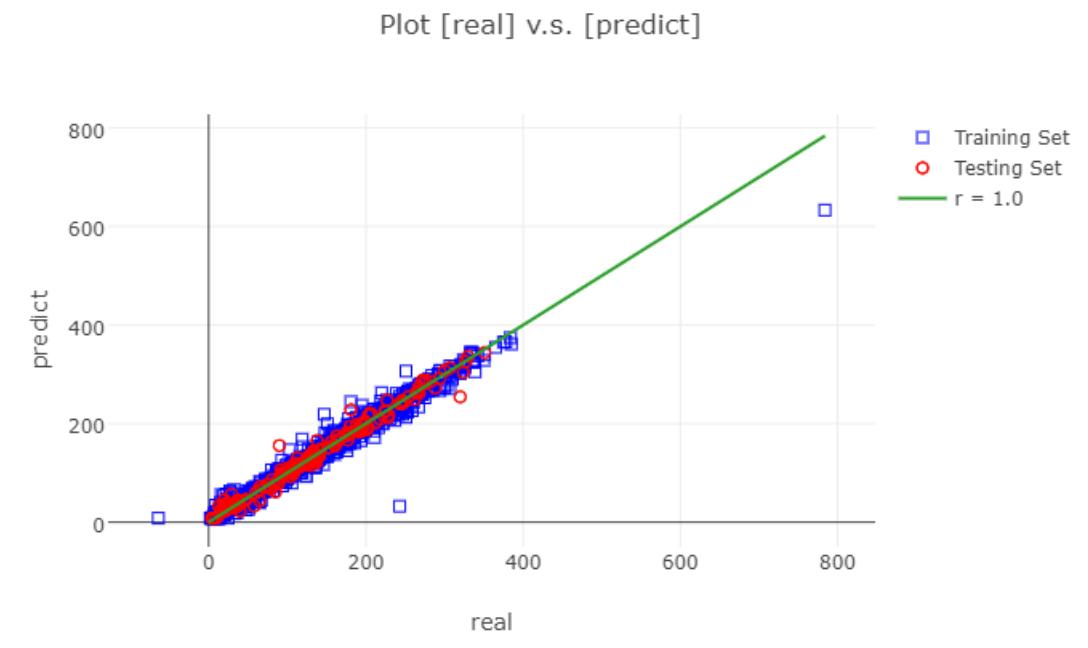
	MAE	Pearson's Correlation
Training Set	8.406	0.963
Testing Set	11.373	0.946



Bulk Modulus Prediction (K_VRH)

ML Settings

Data	<ul style="list-style-type: none">1675 binary composition from Material Project (with band gap < 0.5 & 0 < Poisson's ratio < 0.5)90% (1507) for training 10% (168) for testing
Feature Engineering	<ul style="list-style-type: none">Features: atomic fraction, density , energy per atom, formation energy per atomApply robust scaling
ML Model	<ul style="list-style-type: none">Multi-layer Perceptron (MLP)Hidden layer arch: 20, 10, 5Activation functions: relu Solver: lbfgs
Evaluation	<ul style="list-style-type: none">MAE (mean absolute error)Pearson's correlation

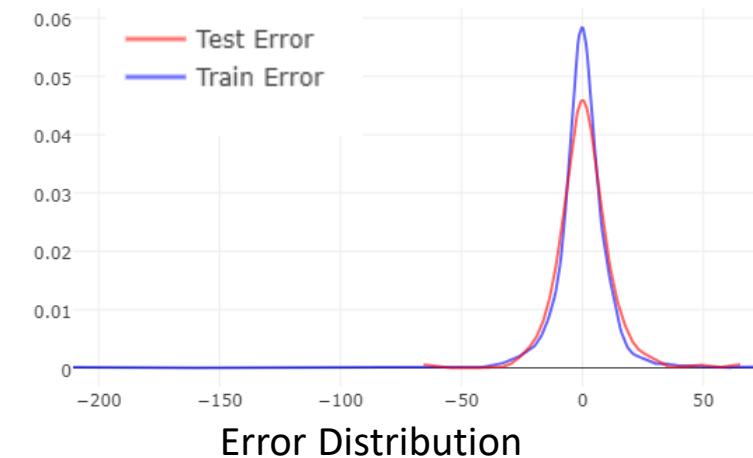


Real values v.s. Predictions

Bulk Modulus Prediction (K_VRH)

Percent of Predictions within Relative Error of

K_VRH	5%	10%	20%	30%
Training Set	0.621	0.826	0.9157	0.945
Testing Set	0.614	0.819	0.911	0.941



MAE & Pearson's Correlation of Predictions

	MAE	Pearson's Correlation
Training Set	6.835	0.988
Testing Set	7.567	0.989

