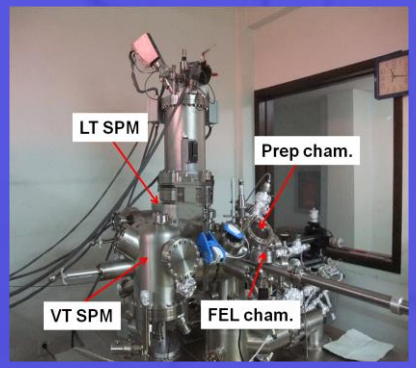
Materials informatics: the 4th paradigm

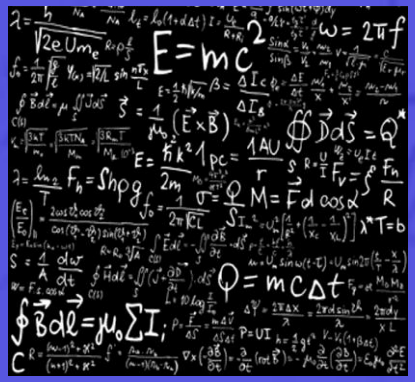
Hong Guo

RQMP, Center for the Physics of Materials, and Department of Physics,

McGill University, Montreal, Canada

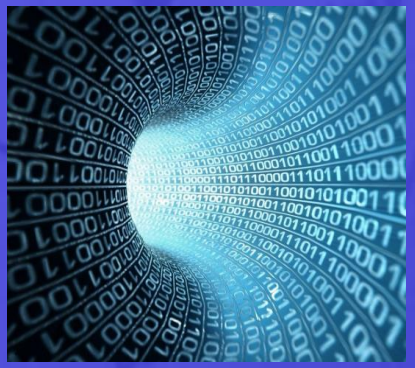


Experiment



Theory

Modeling



Material informatics

Acknowledgements

The iMAT platform:

Dr. Eric Zhu, Nanoacademic Technologies Inc., Montreal, machine learning
Dr. Zhongli Liu, McGill, code/platform design/development, superconductivity
Dr. Peter Kang, McGill, code/platform development, energetic materials, spectroscopy

Applications:

Dan Abarbanel, PhD student, McGill, X-ray spectroscopy, AIMD
Dustin Lebiadowski, undergraduate student, McGill, 2D ferroelectric materials
Yi-Li Lin, undergraduate student, McGill, material descriptors
Prof. Mark Sutton, McGill, X-ray spectroscopy
Dr. Xianghua Kong, McGill, photocatalysts

Thanks to: NSERC, FQRNT

Outline

- **Introduction: materials informatics & machine learning**
- Search for solid lithium-ion conductors
- Search for high Tc superconductors
- Search for energetic materials
- Summary

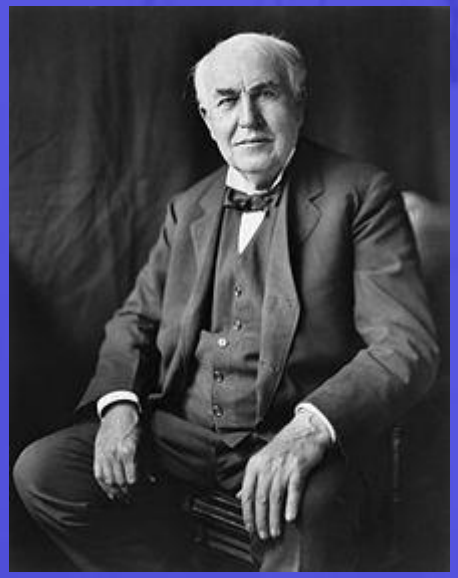
Introduction to materials informatics (MI)

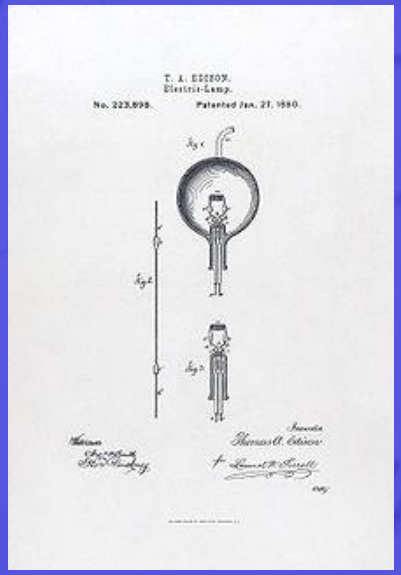
Materials informatics (MI)： data-driven approach for materials screening and design.

Input: some data and physical intuition (databases, experimental or computation).

Tools: machine learning, informatics toolkits … **iMAT**

Advantage 1: MI has larger searching space



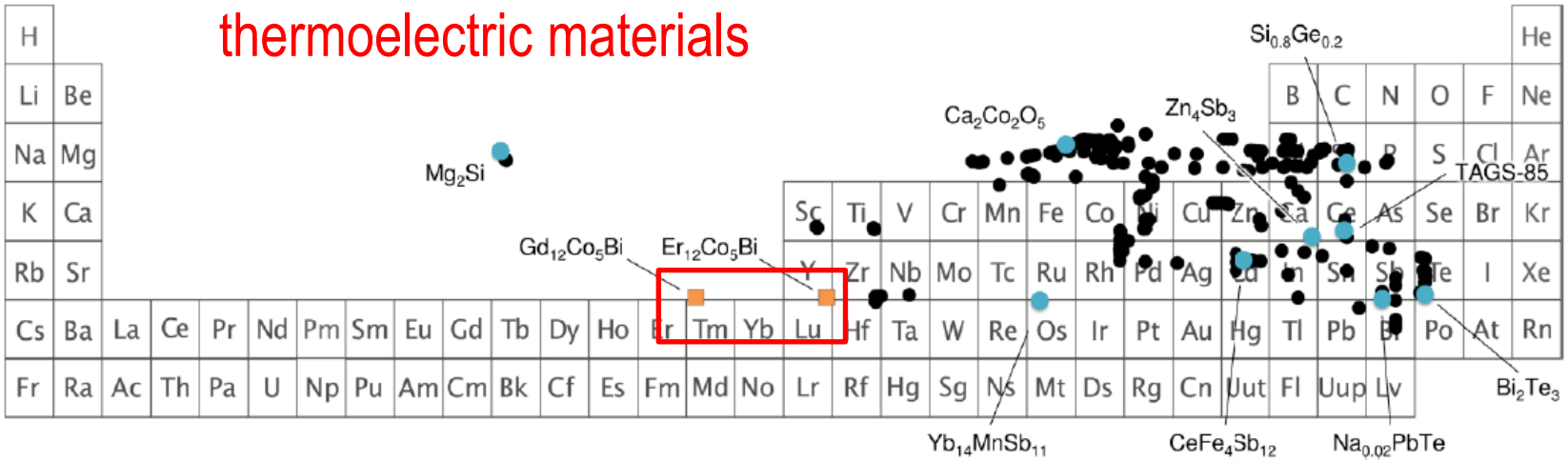


Human:: Thomas Edison performed ~3000 experiments to find an acceptable material for the filament of electric bulb (10^3)

MI: ~200,000 experimentally existing materials in the Inorganic Crystal Structure Database (ICSD, 10^5)

Introduction (cont.)

Advantage 2: MI can search unknown territory



Human: tend to search around known materials

MI: search all places - no bias toward any material class

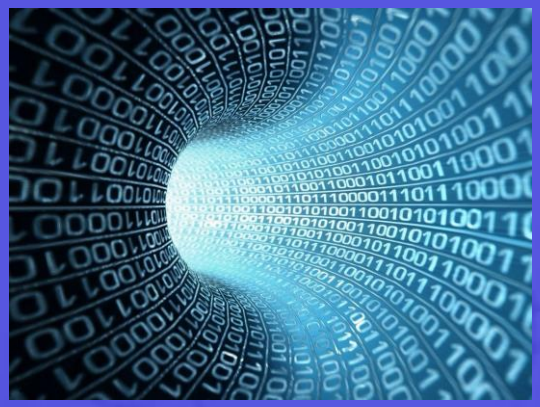
Advantage 3: MI presents huge opportunities

Human have discovered or synthesized ~400,000 compounds and over 140 million molecules and sequences, but what do we know about them?

- Elastic constants: ~500 compounds
- Dielectric constant: ~400
- Heat conductivity: ~200
- Superconductor transition temperature: ~30,000
- Topological insulators: ~50

Conclusion: we have the materials but don't know properties for most of them

Where to find materials data?



experimental data

computational data

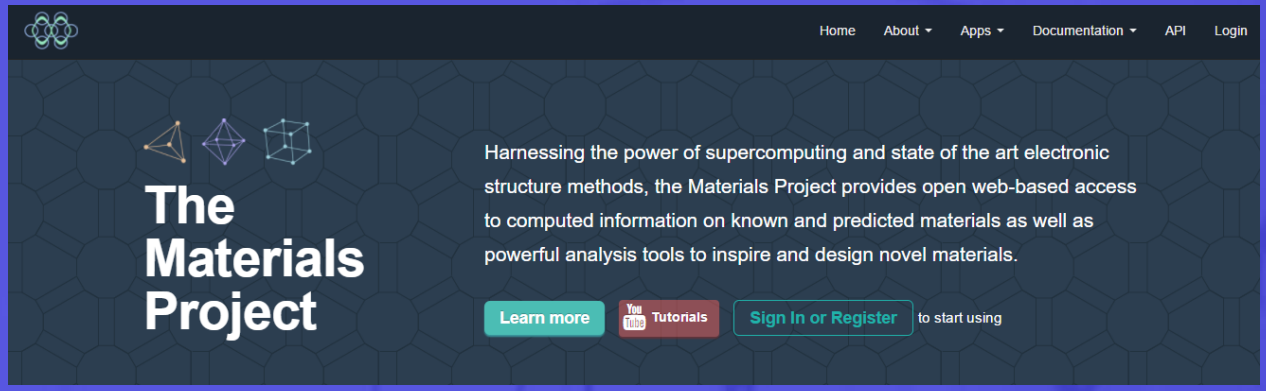
database

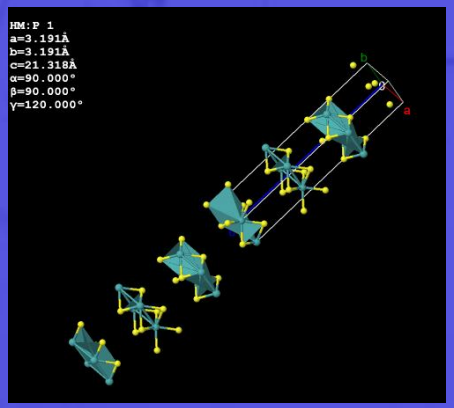
Table I. A list of some notable materials-data resources.

Name	URL	Category	Free/Non-Free
3D Materials Atlas	cosmicweb.mse.iastate.edu/wiki/display/home/Materials+Atlas+Home	3D Characterization	Free
AFLOWLIB	aflowlib.org	Computational	Free
AIST Research Information Databases	www.aist.go.jp/aist_e/list/database/riodb	General Materials Data	Free
American Mineralogist Crystal Structure Database	rruff.geo.arizona.edu/AMS/amcsd.phP	Minerals	Free
ASM Alloy Center Database	mio.asminternational.org/ac	Alloys	Non-Free
ASM Phase Diagrams	www1.asminternational.org/AsmEnterprise/APD	Thermodynamics	Non-Free
CALPHAD databases (e.g., Thermocalc SGTE)	www.thermocalc.com/products-services/databases/thermodynamic	Thermodynamics	Non-Free
Cambridge Crystallographic Data Centre	www.ccdc.cam.ac.uk/pages/Home.aspx	Crystallography	Non-Free
CatApp	suncat.stanford.edu/catapp	Catalysts	Free
Chemspider	www.chemspider.com	Chemical data	Free
CINDAS High-Performance Alloys Database	cindasdata.com/products/hpad	Alloys	Non-Free
Citrination	citrination.com	General Materials Data	Free
Computational Materials Repository	cmr.fysik.dtu.dk	Computational	Free
CRC Handbook	www.hbcpnetbase.com	General Materials Data	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.newglass.jp/interglad_n/gaiyo/info_e.html	Glass	Non-Free
Knovel	app.knovel.com/web/browse.v	General Materials Data	Non-Free
Matbase	www.matbase.com	General Materials Data	Free
MatDat	www.matdat.com	General Materials Data	Non-Free
Materials Project	www.materialsproject.org	Computational	Free
MatNavi (NIMS)	mits.nims.go.jp/index_en.html	General Materials Data	Free
MatWeb	www.matweb.com	General Materials Data	Free
Mindat	www.mindat.org	Minerals	Free
NanoHUB	nanohub.org	Nanomaterials	Free
Nanomaterials Registry	www.nanomaterialregistry.org	Nanomaterials	Free
NIST Materials Data Repository (DSpace)	materialsdata.nist.gov/dspace/xmlui	General Materials Data	Free
NIST Interatomic Potentials Repository	www.ctcms.nist.gov/potentials	Computational	Free
NIST Standard Reference Data	www.nist.gov/srd/dblistpcdatabases.cfm	General Materials Data	Non-Free
NIST Standard Reference Data	www.nist.gov/srd/onlinelist.cfm	General Materials Data	Free
NoMaD	nomad-repository.eu/cms	Computational	Free
Open Knowledge Database of Interatomic Models (Open KIM)	openkim.org	Computational	Free
Open Quantum Materials Database	oqmd.org	Computational	Free

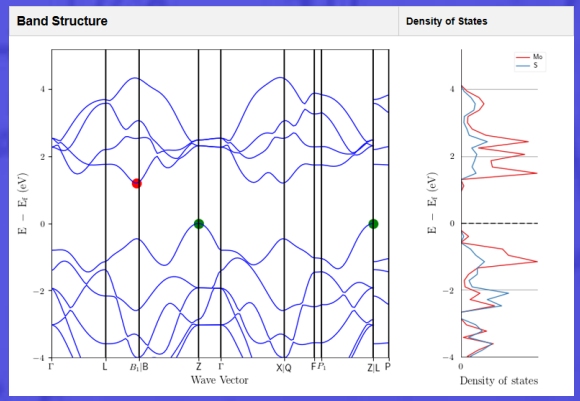
Informatics: 3 simplest examples

USA: materials genome





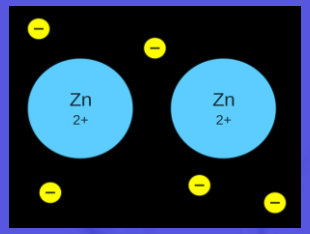
structure: 67483

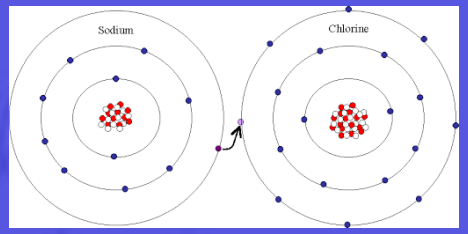


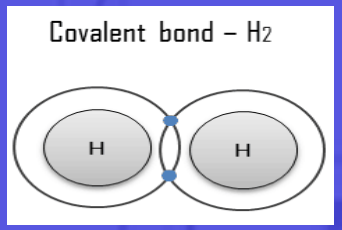
electronic structure: 47087

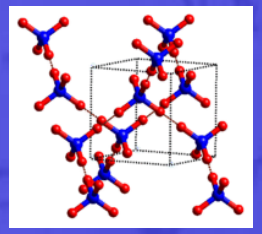
Ex.1 Can we determine crystal types by simple informatics ?

Metallic crystal, ionic crystal, covalent crystal, molecular crystal

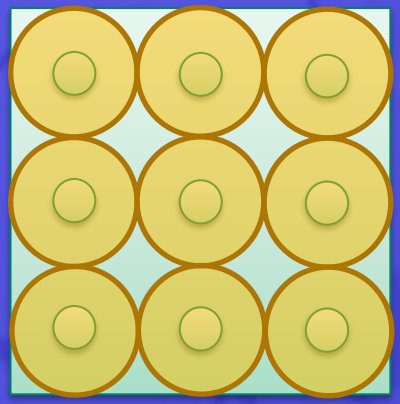




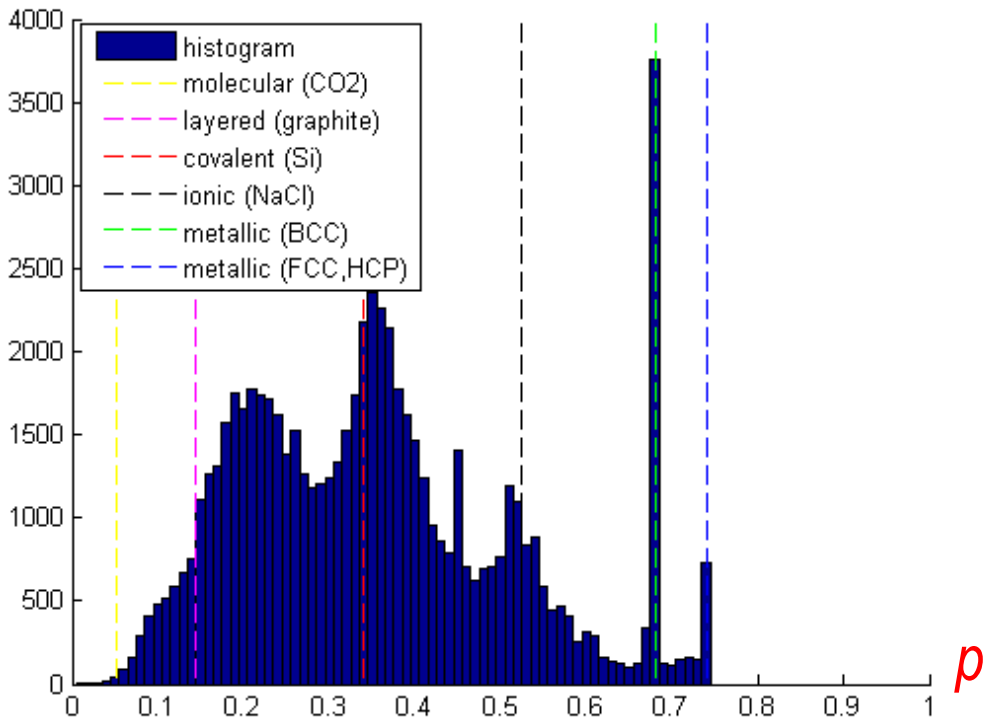




Packing factor $p = \frac{\sum V_{sphere}}{V_{unitcell}}$



Result: histogram of packing factor (for 67483 crystals in Materials Project)

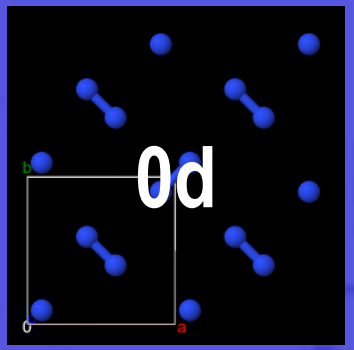


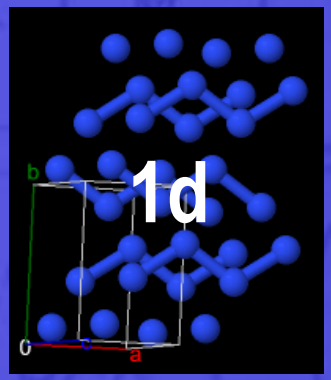
atomic structure	crystal type	packing factor
CO2	molecular	0.052
graphite	layered	0.143
MoS2	layered	0.354
Si	covalent	0.340
NaCl	ionic	0.523
BCC	metallic	0.740
FCC/HCP	metallic	0.740

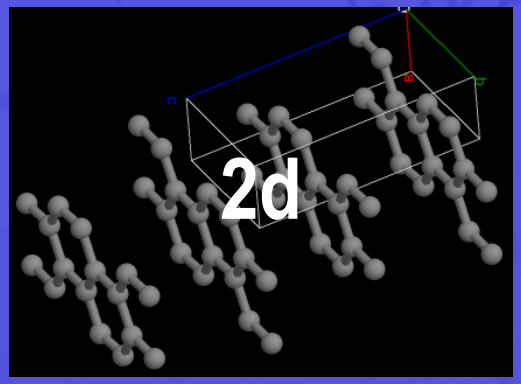
Conclusion by informatics - packing factor can *roughly* distinguish crystal types

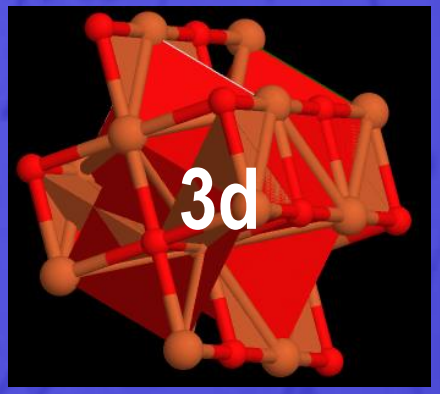
Ex.2: can we determine dimensionality by informatics ?

0d, 1d, 2d, 3d and layered structures?



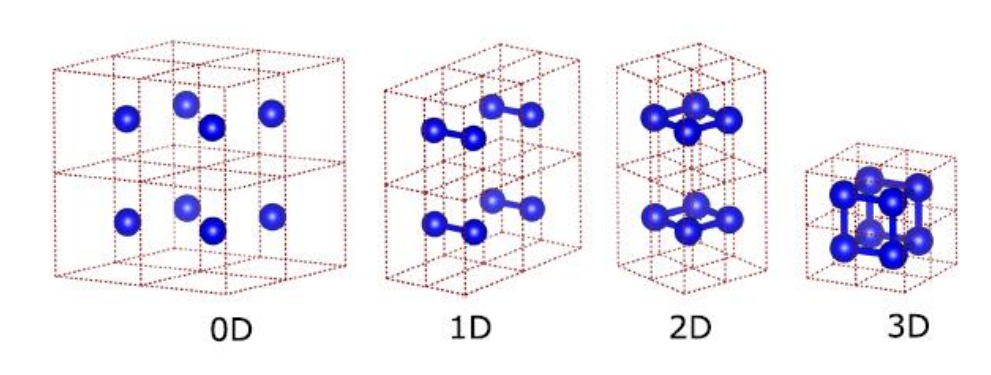






Algorithm for determining dimensionality of a crystal

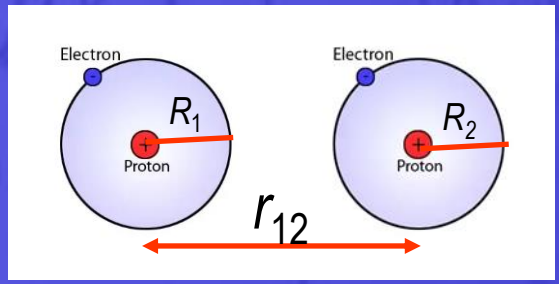
Count the connected atom number in a 2 x 2 x 2 supercell



$$2^d = \frac{N_{2\times2\times2}}{N_{1\times1\times1}}$$

How do we know if two atoms are "connected" ?

$$r_{12} < R_1 + R_2 + \delta R$$



R: covalent radius extrapolated from 426,000 crystal structures (Cambridge structure database).

The Cambridge Crystallographic Data Centre

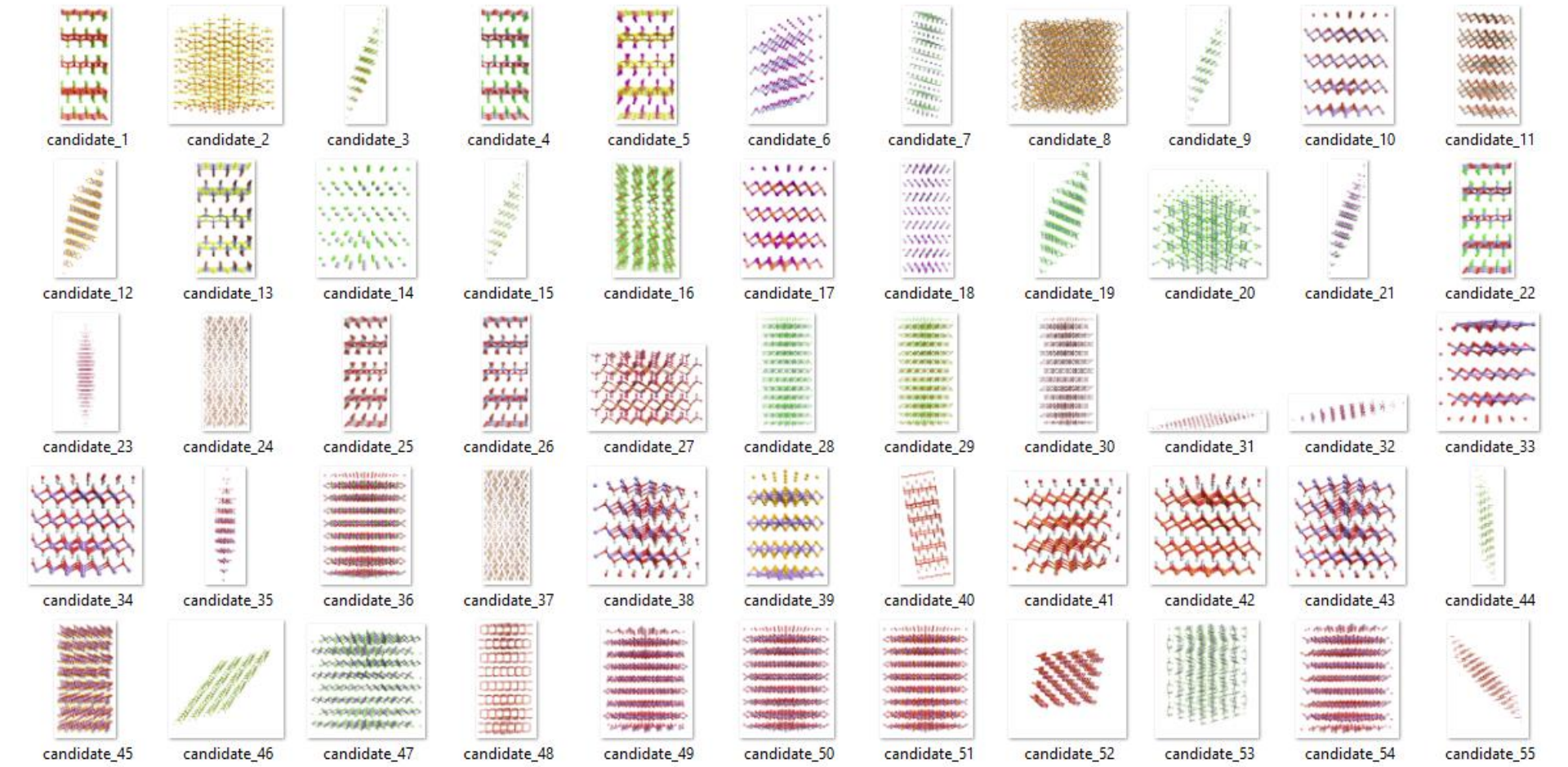
Results: dimensionality of materials (67483 crystals in Materials Project)

$$r_{12} < R_1 + R_2 + \delta R$$

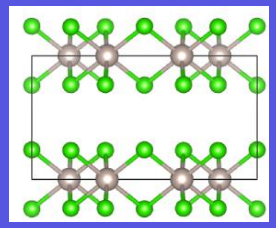
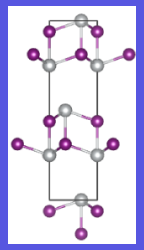
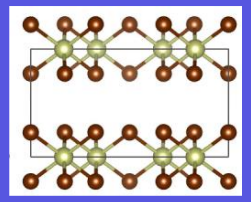
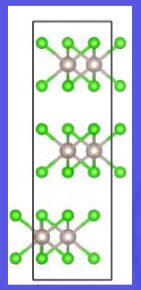
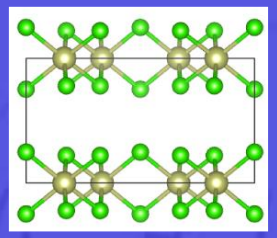
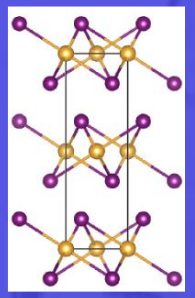
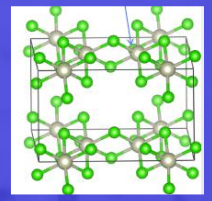
dimensionality	$\delta R = 0.5$	$\delta R = 1.0$	$\delta R = 1.5$
0d	3,506	1,839	910
1d	1,257	867	577
2d	3,531	2,510	1,802
3d	59,188	62,266	64,194

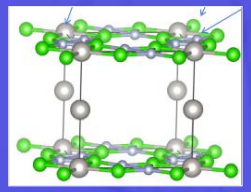
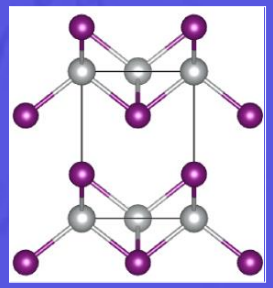
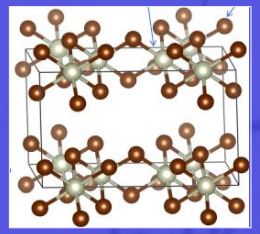
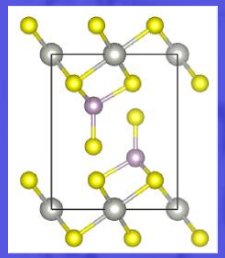
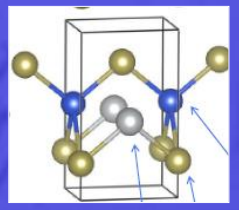
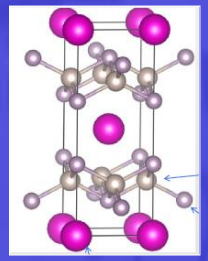
Conclusion: by simple informatics, we roughly find all the possible materials from 0d to 3d, from a database

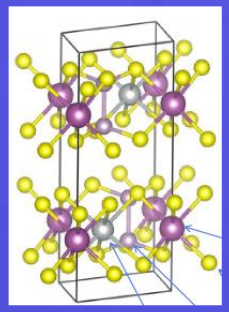
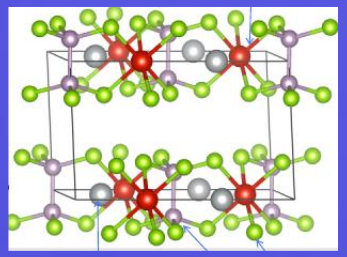
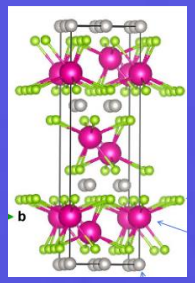
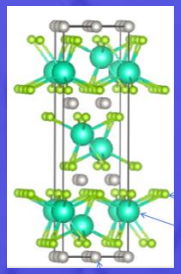
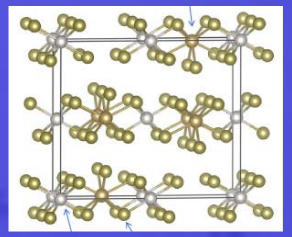
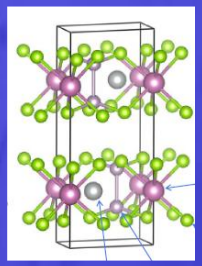
Results: possible magnetic materials with layered structure



Results: noble metal with layered structures

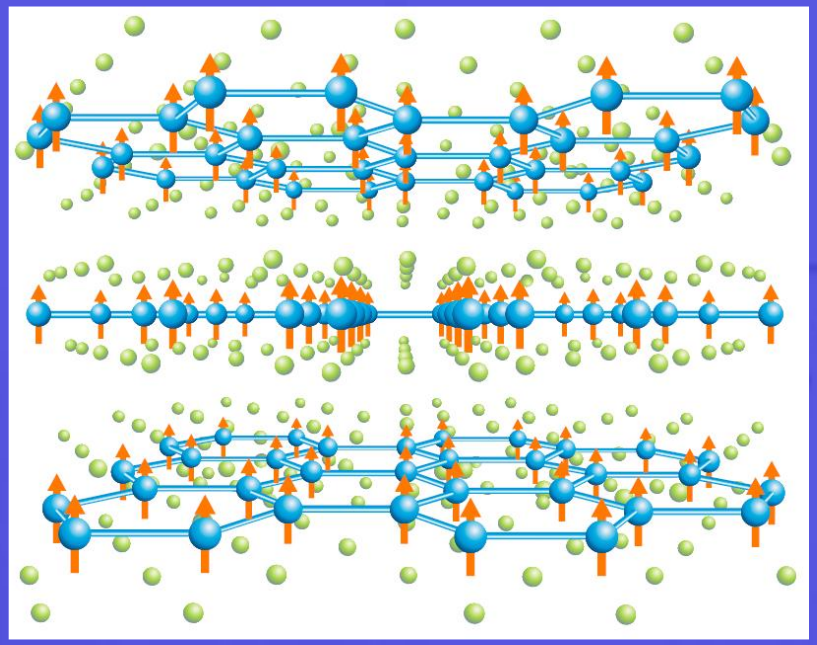







Ex.3 Search 2d ferromagnets by materials informatics

Thanks to: Dr. Eric Zhu, Dr. Xianghua Kong, Dr. Trevor David Rhone

Ferromagnetism was found in 2d materials experimentally



LETTER

doi:10.1038/nature22060

Discovery of intrinsic ferromagnetism in two-dimensional van der Waals crystals

Cheng Gong[1]*, Lin Li[2]*, Zhenglu Li[3,4]*, Huiwen Ji[5], Alex Stern[2], Yang Xia[1], Ting Cao[3,4], Wei Bao[1], Chenzhe Wang[1], Yuan Wang[1,4], Z. Q. Qiu[3], R. J. Cava[5], Steven G. Louie[3,4], Jing Xia[2] & Xiang Zhang[1,4]

LETTER

doi:10.1038/nature22391

Layer-dependent ferromagnetism in a van der Waals crystal down to the monolayer limit

Bevin Huang[1]*, Genevieve Clark[2]*, Efrén Navarro-Moratalla[3]*, Dahlia R. Klein[3], Ran Cheng[4], Kyle L. Seyler[1], Ding Zhong[1], Emma Schmidgall[1], Michael A. McGuire[5], David H. Cobden[1], Wang Yao[6], Di Xiao[4], Pablo Jarillo-Herrero[3] & Xiaodong Xu[1,2]

UC-Berkeley group	$Cr_2Ge_2Te_6$
UW and MIT group	CrI_3

Fundamental: why 2d ferromagnetism is interesting?

Fundamental physics: can ferromagnetism exist in 2d materials?

2d Heisenberg	$H = -J\sum_{\langle ij\rangle}(S_i^x S_j^x + S_i^y S_j^y + S_i^z S_j^z) - B\sum_i S_i^z$	No	thermal fluctuations
2d Ising	$H = -J\sum_{\langle ij\rangle} S_i^z S_j^z - B\sum_i S_i^z$	Yes	magnetic anisotropy

Mermin-Wagner theorem: Continuous symmetries cannot be spontaneously broken at finite temperature in systems with sufficiently short-range interactions in dimensions d ≤ 2.

Device application:: ultra-compact spintronics, magneto-optoelectronics

Simple and physics based descriptors:

Descriptor: parameter that describes some physical property

Descriptors:

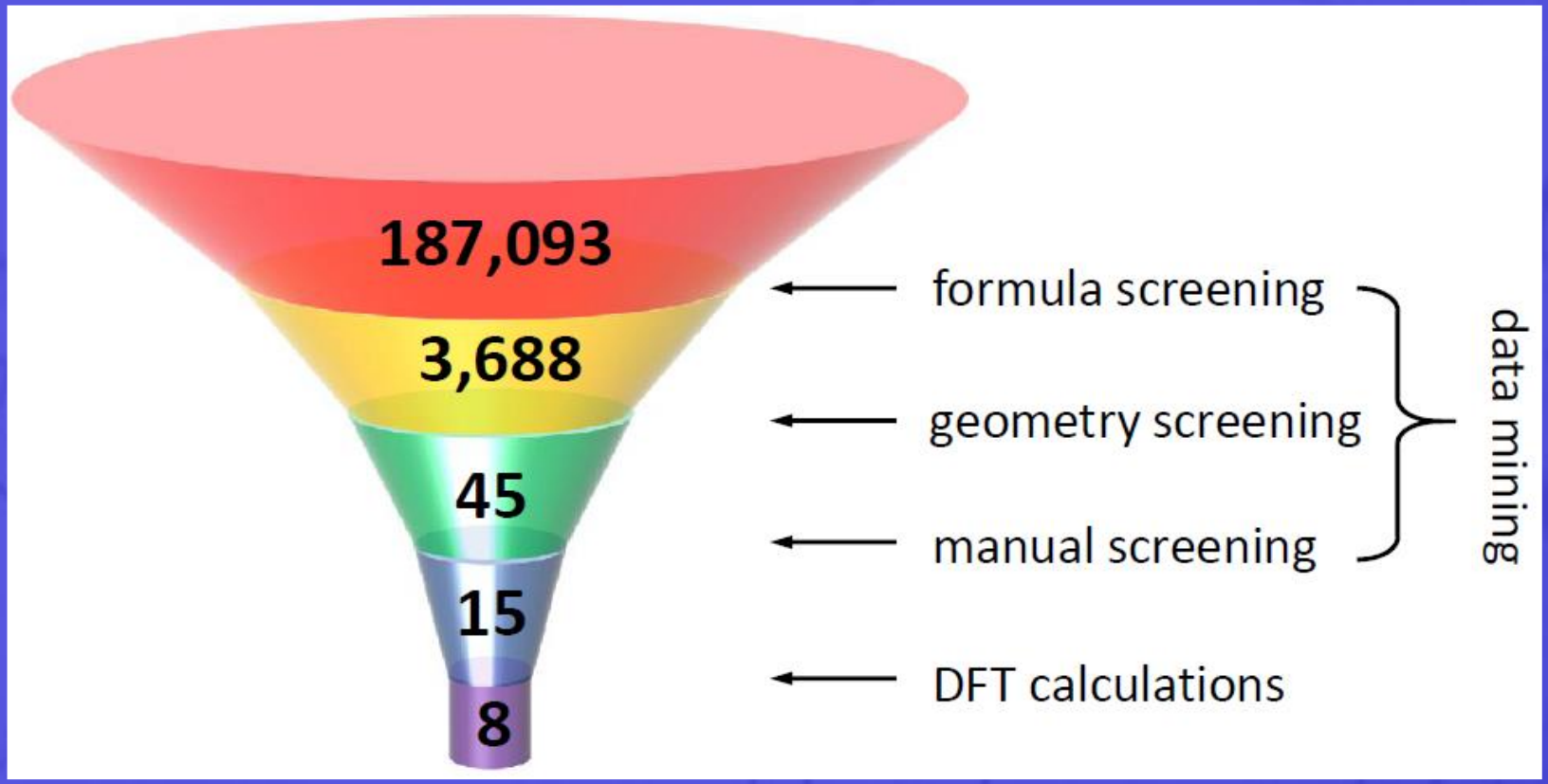
1) Material exited (use ICSD)
2) dimension = 2
3) packing factor < 0.354
4) magnetic moment per atom > 0.625 μ_B
5) Must have M-Z-M network:

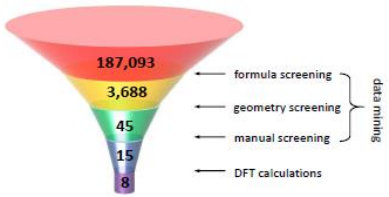
   M = transition metal atom,

   Z=heavy atom (Z > 49). This ensures some super-exchange.

materials	mp-id	quantity
diamond	mp-66	δE = 0.136
graphite	mp-568286	p = 0.143
MoS_2	mp-1434	p = 0.354
Fe	mp-13	μ = 2.332
Co	mp-54	μ = 3.319
Ni	mp-23	μ = 0.625

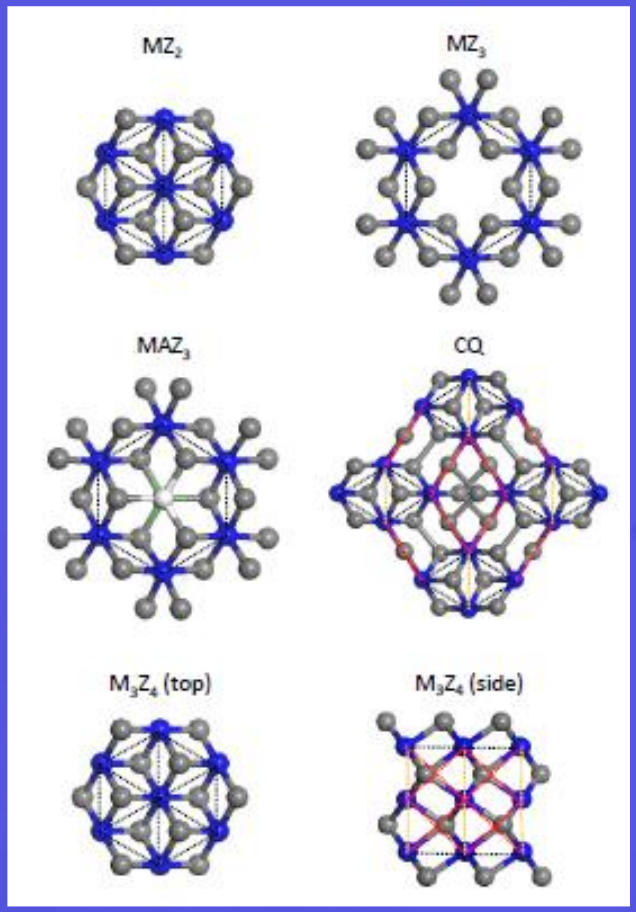
Mining ICSD (Inorganic Crystal Structure Database):





We calculate the 15 material candidates

$$H = -\sum_{\langle ij \rangle} J_{ij}\sigma_i\sigma_j$$



#	ICSD	formula	structure	ordering	J(meV)	T_c(K)
1	52368	CoI_2	-	AFM	-	-
2	52369	FeI_2	-	AFM	-	-
3	33673	MnI_2	MZ_2	FM	0.3	15
4	22108	NiI_2	MZ_2	FM	1.5	63
5	246907	VI_2	-	AFM	-	-
6	603582	VTe_2	MZ_2	FM	3.0	128
7	626809	$CrSiTe_3$	MAZ_3	FM	12.2	214
8	252343	$CrGeTe_3$	MAZ_3	FM	17.8	314
9	251655	CrI_3	MZ_3	FM	9.1	161
10	180602	FeTe	-	AFM	-	-
11	66958	MnCaSn	-	AFM	-	-
12	4073	CrI_2	-	AFM	-	-
13	76730	NiTe	-	AFM	-	-
14	35266	$CrTe_3$	CQ	FM	5.0 16.5 1.5	71
15	626873	Cr_3Te_4	sMZ_2	FM	16.6 18.8 12.0	2057

Using the 2 experimental data, get a heuristic scaling factor of T_c: 0.2 ~ 0.4.

C. R. Acad. Sc. Paris, t. 263 (29 août 1966). Série B — 621

PHYSIQUE DES SOLIDES. — *Étude par diffraction neutronique du composé ferromagnétique* $TiCr_2Te_4$. Note (*) de Mlle **Bernadette Andron**, MM. **Gilbert Bérodias**, **Maurice Chèvreton** et **Paul Mollard**, transmise par M. Louis Néel.

For **bulk** Cr_3Te_4, an old measurement reported ferromagnetic transition at T_c = 320K.

Our Ising simulation produced ~400K after applying the heuristic factor.

PHYSICAL REVIEW MATERIALS **2**, 081001(R) (2018)

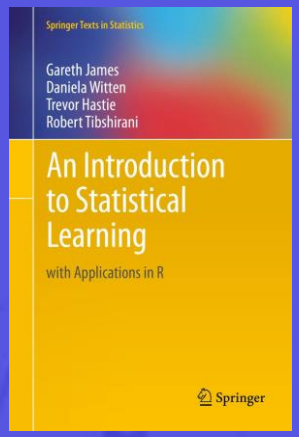
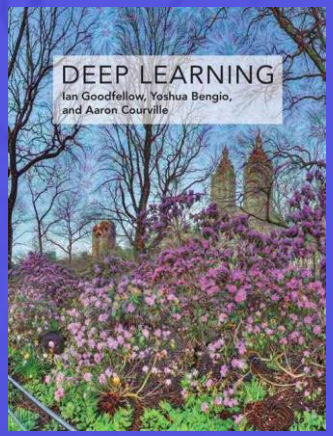
Rapid Communications | Editors' Suggestion

Systematic search for two-dimensional ferromagnetic materials

Yu Zhu,[1] Xianghua Kong,[2] Trevor David Rhone,[3] and Hong Guo[1,2,4]

2019: Prof. C.G. Zeng of USTC synthesized Cr_2Te_3: T_c=220K. Work on Cr_3Te_4 on going.

Advanced data analysis tools – machine learning methods

Statistical Learning		scikit learn	10^2 to 10^3 data points physics interpretation
Deep Learning		TensorFlow	> 10^4 data points large model compacity

Old days: a very old work on computer optimization:

Physica Scripta. Vol. T38, 40–44, 1991.

A Fast Algorithm for Simulated Annealing

Hong Guo, Martin Zuckermann, R. Harris and Martin Grant

Centre for the Physics of Materials, Department of Physics, McGill University, 3600 University Street, Montreal, Quebec, Canada H3A 2T8

Received September 24, 1990; accepted October 26, 1990

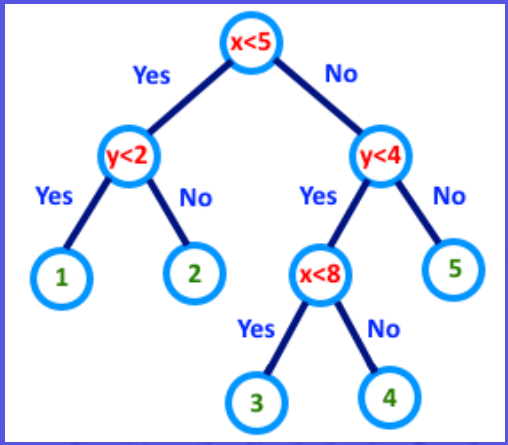
2.3. *Restoration of corrupted binary images*

In addition to the two examples discussed above, there are many other practical problems which can be studied with this algorithm. For example, a large class of computer vision and image interpretation problems can be described and discussed within the framework of optimization theory

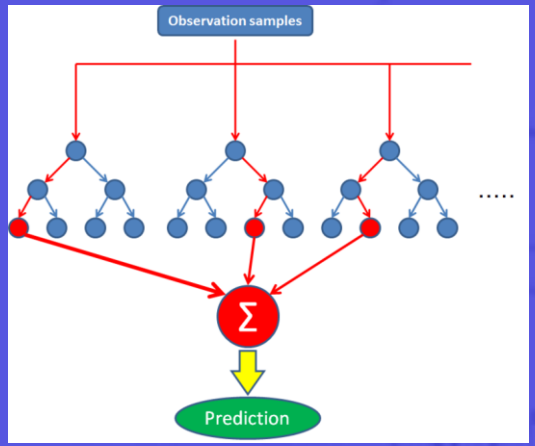
For an image (a Chinese character) of 50 000 pixels corrupted with noise strength $p = 0.25$ (25% of the pixels are corrupted), only 3% noise is left after 120 trials per pixel (6 annealing stages with 20 trials each).

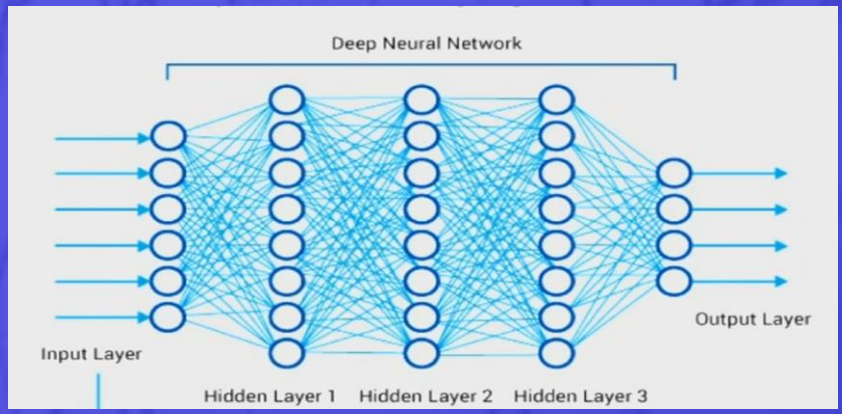
Today: many machine learning algorithms

Decision Tree



Random Forest





For materials informatics - supervised learning

Outline

- Introduction: materials informatics & machine learning
- **Search for solid lithium-ion conductors**
- Search for high Tc superconductors
- Search for energetic materials
- Summary

Search for solid Lithium-ion conductors (from Materials Project)

Sandek et al. Ener. & Environ. Sci. (2017)

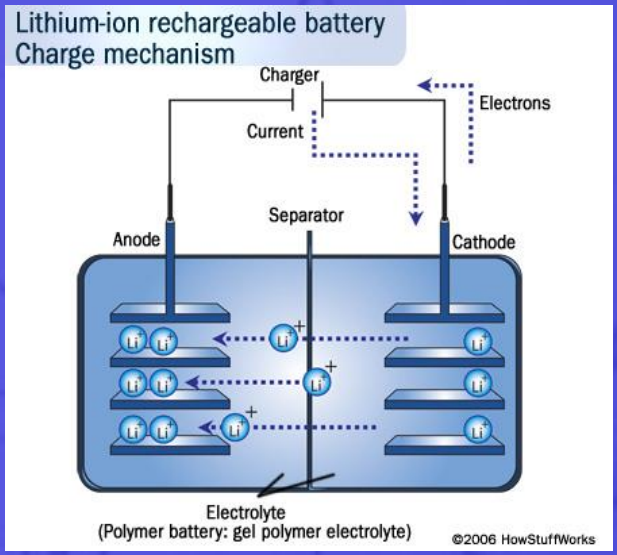
66,840 crystals candidates

contain Li atoms

12,831 candidates

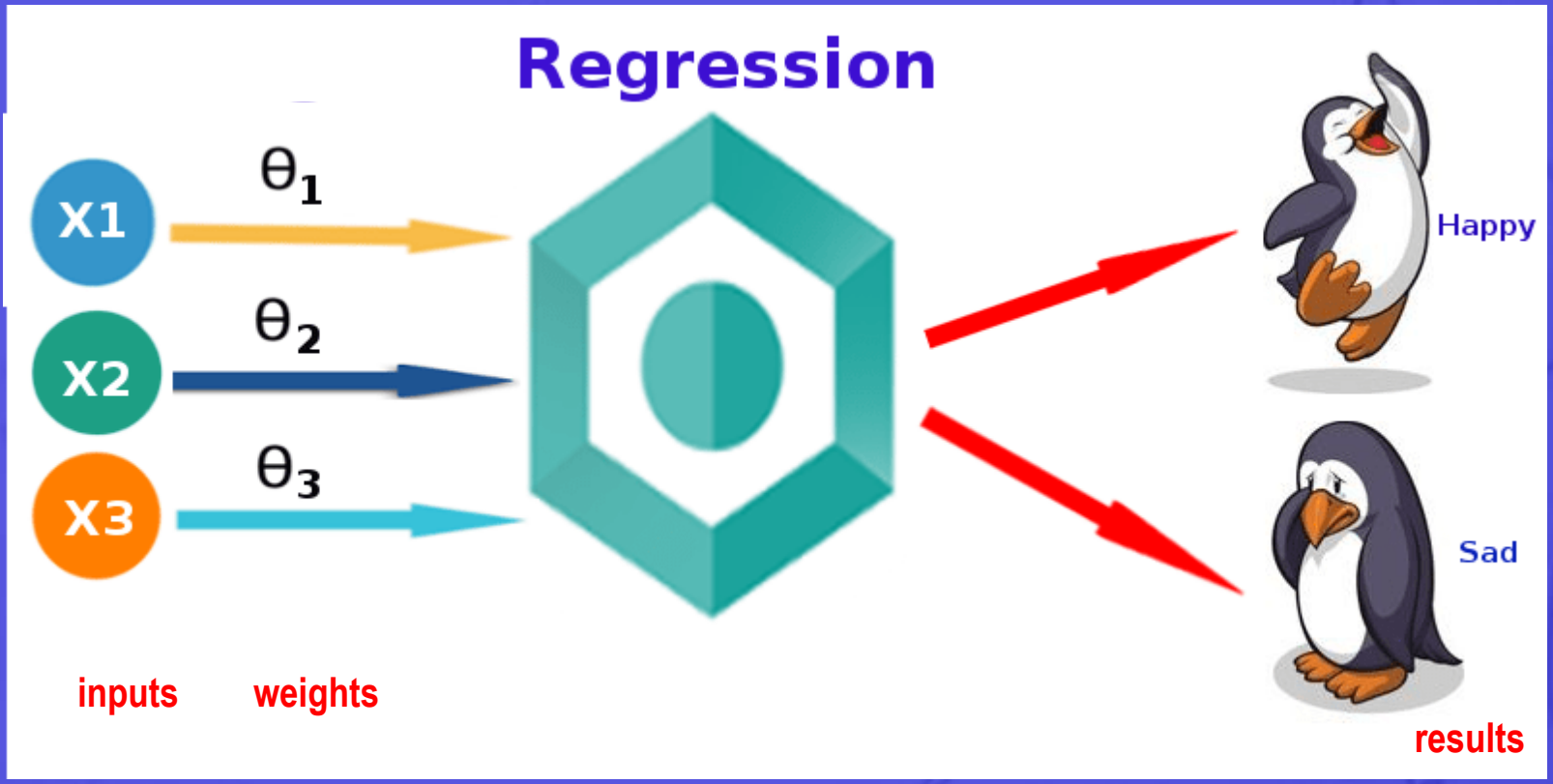
band gap, anti-oxidization, structural stability, no transition metal, cost, earth abundance

317 candidates



Next? Machine Learning to further narrow down the list

Machine learning: build a classifier for “good” Li-conductor



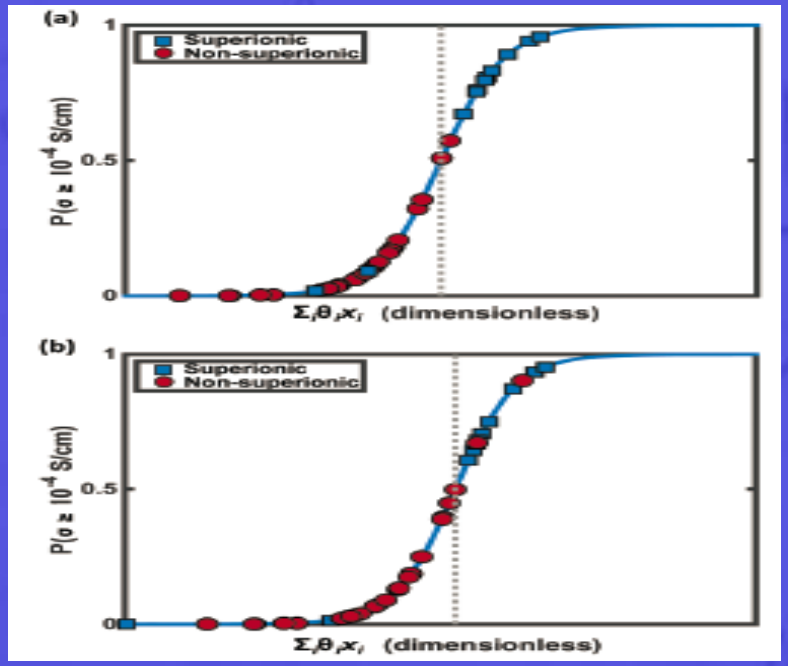
Simplest regression:

$$p(\{\theta_i\},\{x_i\}) = \frac{1}{1+e^{-(\theta_o+\theta_1 x_1+\cdots+\theta_n x_n)}}$$

Statistical Learning (SL): experimental property data

Training set: Li ionic conductivity of 40 measured materials

Ionic conductivity of training set differs by ten orders. Both good and bad ionic conductors are valuable for training.



Sandek et al. Ener. & Environ. Sci. (2017)

Composition	RT bulk ionic conductivity (S cm^{-1})	Ionic conductivity ref.	Structure ref.
$LiLa(TiO_3)_2$	1×10^{-3}	17	29
$Li_{9.81}Sn_{0.81}P_{2.19}S_{12}$	5.5×10^{-3}	30	30
$Li_{10}Ge(PS_6)_2$	1.4×10^{-2}	7	31
$Li_{10.35}Si_{1.35}P_{1.65}S_{12}$	6.5×10^{-3}	30	30
$Li_{14}ZnGe_4O_{16}(2)$	1.0×10^{-6}	32	33 and 34
$Li_2Ca(NH)_2$	6.4×10^{-6}	35	36
$Li_2Ge_7O_{15}$	5.0×10^{-6}	37	38
Li_2NH	2.5×10^{-4}	35	39
Li_2S	1.0×10^{-13}	40	41
$Li_{13.6}Si_{2.8}S_{1.2}O_{16}$	6.0×10^{-7}	42	43
$Li_{14}Ge_2V_2O_{16}$	7.0×10^{-5}	44	45
$Li_{15}Ge_3V_2O_4$	6.03×10^{-6}	46	47
$Li_{14.8}Ge_{3.4}W_{0.6}O_4$	4.0×10^{-5}	46	47
$Li_3Fe_2P_3O_{12}$	1.0×10^{-7}	48	49
Li_3N	5.75×10^{-4}	50	51
Li_3P	1.0×10^{-3}	8	52
γ-Li_3PS_4	3.0×10^{-7}	53	54
$Li_3Sc_2P_3O_{12}$	1.0×10^{-10}	55	56
β_{II}-Li_3VO_4	4.4×10^{-8}	57	58
$Li_4B_7O_{12}Cl$	1.0×10^{-7}	59	59
$Li_4BN_3H_{10}$	2.0×10^{-4}	60	61
γ-Li_4GeO_4	3.1×10^{-12}	37	62
Li_4SiO_4	2.4×10^{-10}	37	63
$Li_5La_3Bi_2O_{12}$	2.0×10^{-5}	64	64
$Li_5La_3Nb_2O_{12}$	8.0×10^{-6}	65	66
$Li_5La_3Ta_2O_{12}$	1.5×10^{-6}	65	66
Li_5NI_2	1.5×10^{-7}	67	68
$Li_6BaLa_2Ta_2O_{12}$	4.0×10^{-5}	69	70
Li_6FeCl_8	1.0×10^{-4}	71	72
Li_6NBr_3	1.5×10^{-7}	67	73
$Li_6SrLa_2Ta_2O_{12}$	7.0×10^{-6}	69	70
$Li_7La_3Zr_2O_{12}$	3.5×10^{-4}	74	75
$Li_7P_3S_{11}$	4.1×10^{-3}	9	76
$LiAlH_4$	2.0×10^{-9}	77	78
$LiAlSiO_4$	1.4×10^{-5}	79	80
$LiBH_4$	2.0×10^{-8}	60	81
LiI	1.0×10^{-6}	42	82
$LiNH_2$	4.0×10^{-10}	35	83
α'-$LiZr_2P_3O_{12}$	5.0×10^{-8}	84	84

Search for Lithium-ion conductors (from Materials Project)

Sandek et al. Ener. & Environ. Sci. (2017)

66,840 candidates

containing Li atoms

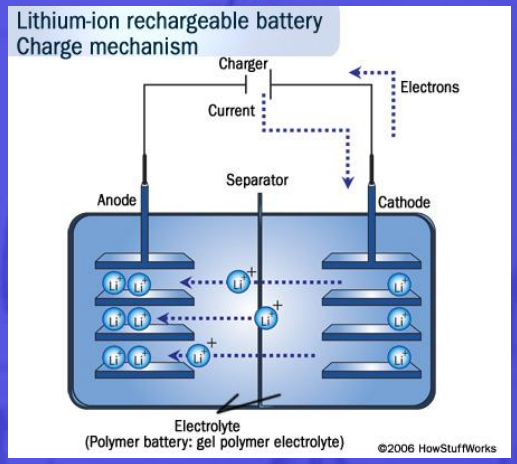
12,831 candidates

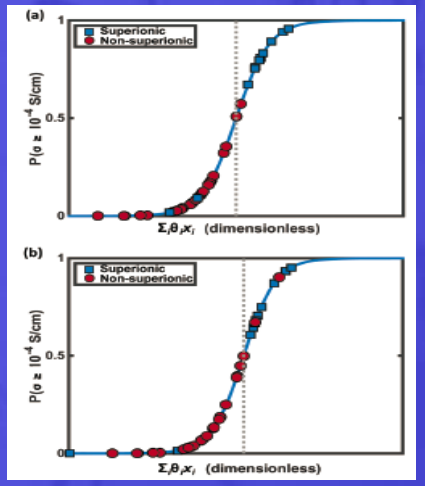
band gap, anti-oxidization, structural stability, no transition metal, cost, earth abundance

317 candidates

Li conduction: machine learning

21 candidates





Energy Env. Sci. 10, 306 (2017).

Results: compare to Sendek, all are the same or similar

MPID	Chemical formula
mp-554076	$BaLiBS_3$
mp-532413	$Li_5B_7S_{13}$
mp-569782[a]	$Sr_2LiCBr_3N_2$
mp-558219	$SrLi(BS_2)_3$
mp-15797	$LiErSe_2$
mp-29410	$Li_2B_2S_5$
mp-676361	Li_3ErCl_6
mp-643069[a]	Li_2HIO
mp-19896	Li_2GePbS_4
mp-7744[a]	$LiSO_3F$
mp-22905[b]	$LiCl$
mp-34477	$LiSmS_2$
mp-676109	Li_3InCl_6
mp-559238	$CsLi_2BS_3$
mp-866665[a]	$LiMgB_3(H_9N)_2$
mp-8751	$RbLiS$
mp-15789	$LiDyS_2$
mp-15790	$LiHoS_2$
mp-15791	$LiErS_2$
mp-561095[a]	$LiHo_3Ge_2(O_4F)_2$
mp-8430	$KLiS$

Sendek

[1]	mp-22905	Li(1)Cl(1)
[2]	mp-34477	Sm(1)Li(1)S(2)
[3]	mp-532413	Li(5)B(7)S(13)
[4]	mp-676361	Er(1)Li(3)Cl(6)
[5]	mp-8751	Rb(1)Li(1)S(1)
[6]	mp-29410	Li(2)B(2)S(5)
[7]	mp-15797	Er(1)Li(1)Se(2)
[8]	mp-15790	Ho(1)Li(1)S(2)
[9]	mp-8430	Li(1)K(1)S(1)
[10]	mp-676109	Li(3)In(1)Cl(6)
[11]	mp-15789	Li(1)S(2)Dy(1)
[12]	mp-15791	Er(1)Li(1)S(2)
[13]	mp-643069	O(1)H(1)I(1)Li(2)
[14]	mp-558219	Sr(1)Li(1)B(3)S(6)
[15]	mp-559238	Li(2)Cs(1)B(1)S(3)
[16]	mp-554076	Li(1)B(1)S(3)Ba(1)
[17]	mp-19896	Pb(1)Li(2)S(4)Ge(1)
[18]	mp-7744	O(3)Li(1)S(1)F(1)
[19]	mp-569782	Br(3)Sr(2)N(2)Li(1)C(1)
[20]	mp-866665	Mg(1)N(2)H(18)B(3)Li(1)
[21]	mp-561095	Ho(3)Li(1)F(2)Ge(2)O(8)

ours

What's next ?

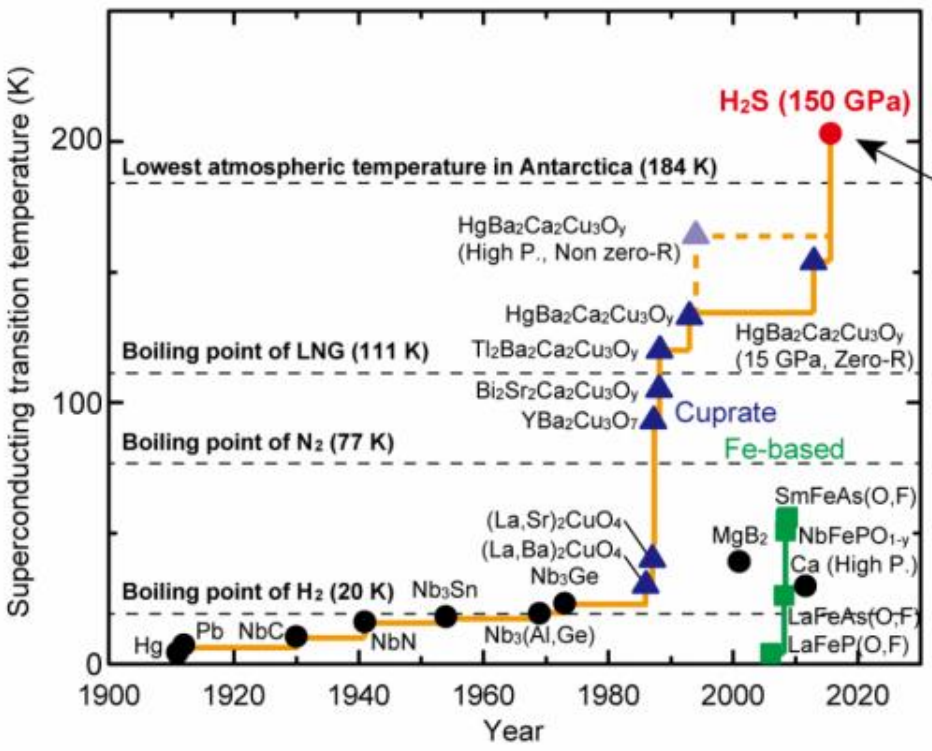
- For the newly found materials， calculate Li ionic conductivity.
- For those with high conductivity, seek experimental verification.

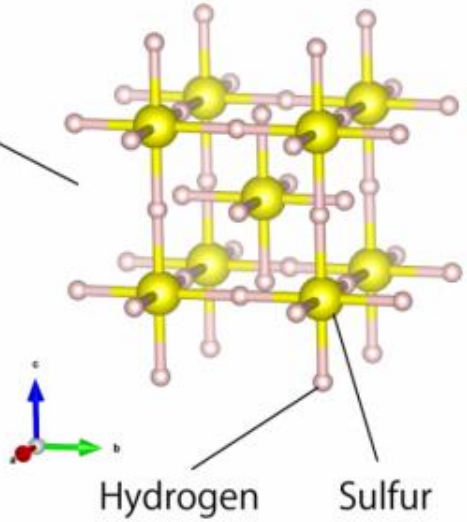
We collaborate with manufacturing firms

Outline

- Introduction: materials informatics & machine learning
- Search for solid lithium-ion conductors
- **Search for high Tc superconductors**
- Search for energetic materials
- Summary

History of superconductors





The superconductor T_c database: SuperCon

Japanese Superconducting Material Database (SuperCon)

31,364 records as of Aug. 2018

Removing data with near zero T_C

Removing apparently erroneous data

15,832 superconducting materials with $T_C > 0$

For degenerate records, keep the one with the highest T_C

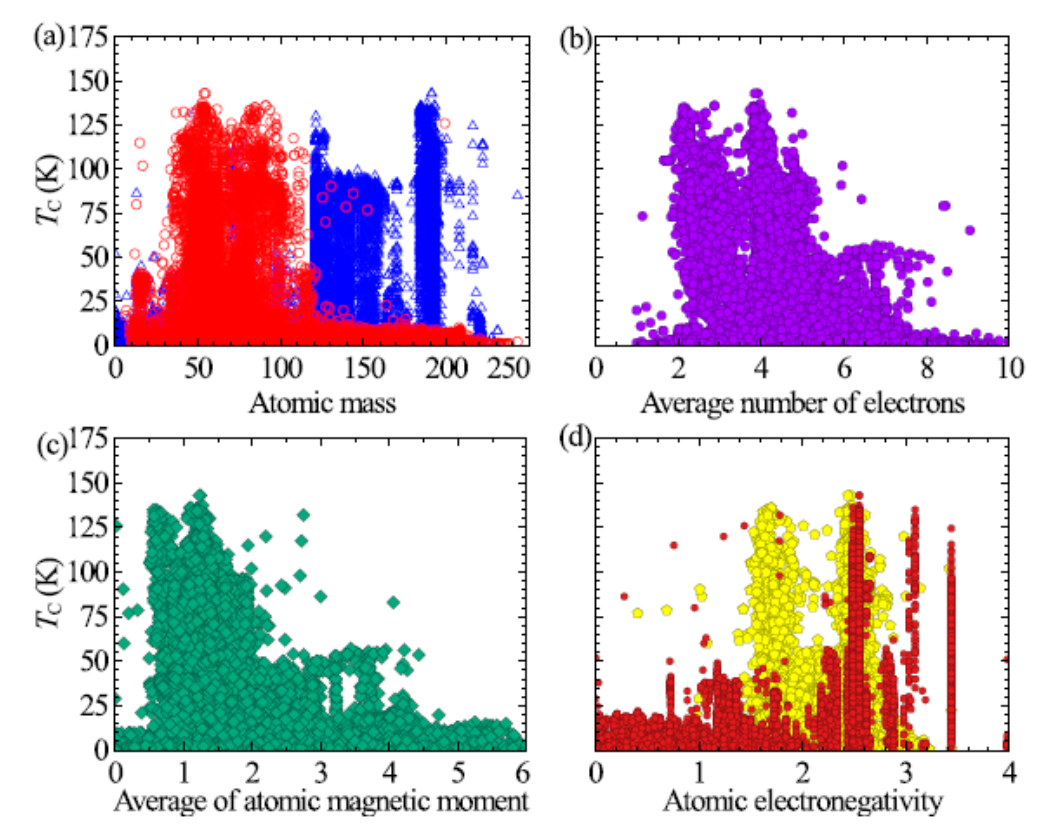
Keep all maerials having similar chemical formula

T_C database for machine learning

Training set: 70% data
Test set: 30% data

Descriptors from SuperCon (informatics)

- Average atomic mass and largest difference in atomic mass
- Average number of valence electrons
- Atomic magnetic moment
- Atomic electronegativity
- Largest difference in atomic number
- Average atomic number
- Average valence radius
- Largest difference of valence radius
- ……



Search high Tc materials by machine learning

Materials informatics:

~200,000 ICSD Entries

Find all the layered structures

Remove replicate structures

~2600 layered materials

Machine learning:

SuperCon: 15,832 superconducting materials with Tc > 0

Training with random forest, decision tree, SVM, and neural network

Multi-step training

Superconductor model

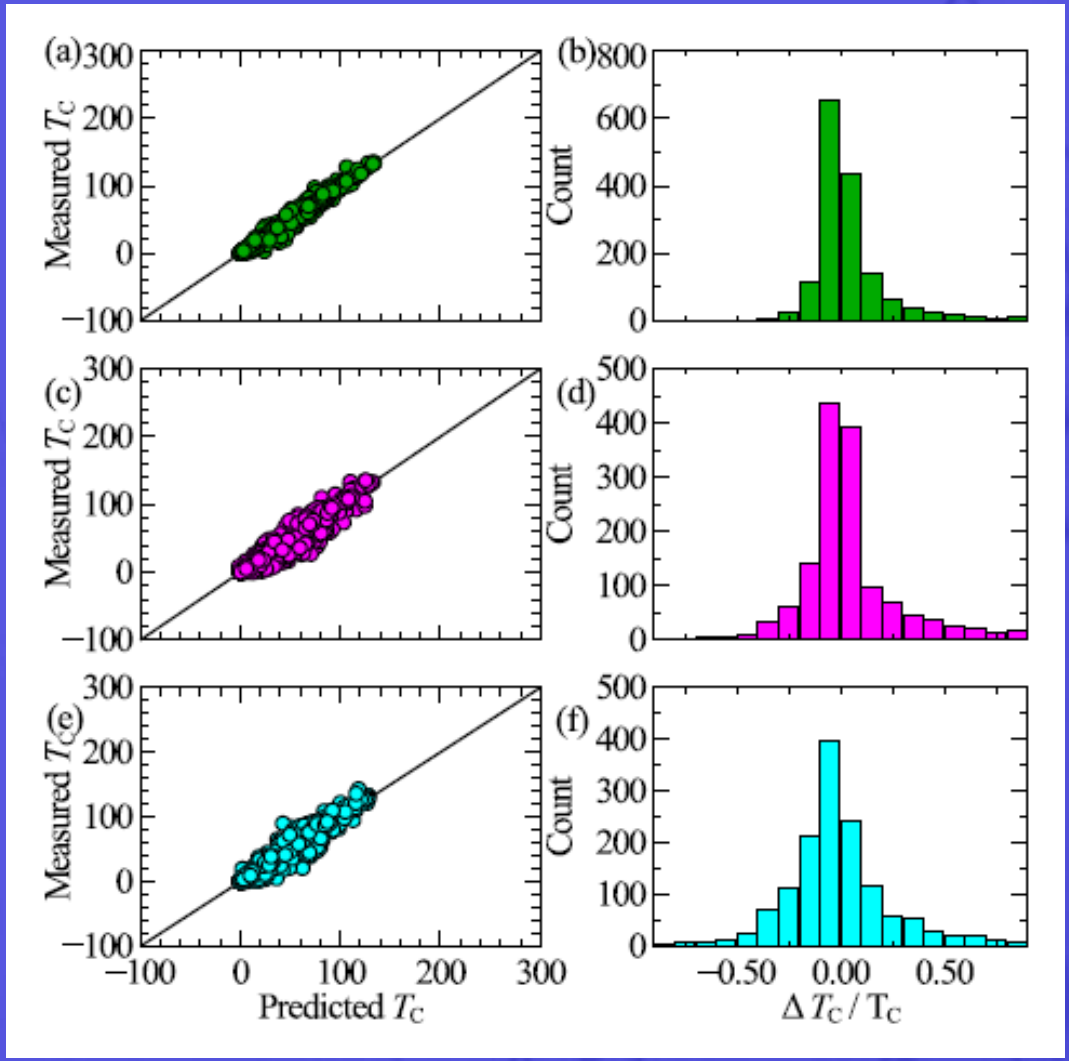
Found 38 layered materials with T_C from 30 K to > 110K

Testing the regression:

Random forest regression

Support vector machine

Artificial neural network



Results: 3 ML and experiments

Of the 38 predicted materials, 13 were already known to be superconductors:

Chemical formula	ICSD entry	SGN	RFR (K)	SVR (K)	DLR (K)	Measured (K)	Heuristic (K)
$HgBa_2Ca_2Cu_3O_8$	075730	123	134.2	125.3	119.0	135.7	133.6
$HgBa_2CaCu_2O_6$	075727	123	124.7	123.1	103.7	134.0	111.3
$HgBa_2CuO_4$	075720	123	95.5	92.4	87.6	98.0	80.0
$Tl_2Ba_2Ca_3Cu_4O_{12}$	068585	139	119.0	115.1	109.8	120.0	140.2
$Tl_2Ba_2CuO_6$	067334	139	94.1	68.0	59.4	117.0	79.5
$Tl_2Ba_2CaCu_2O_8$	065554	139	107.6	110.2	98.7	114.1	114.5
$TlBa_2CaCu_2O_7$	067128	123	94.2	113.0	105.0	93.0	109.5
$Bi_2Sr_2CaCu_2O_8$	068188	139	94.4	84.2	78.0	88.5	142.9
$EuSr_2Pb_2Cu_3O_8$	071481	65	60.8	70.9	70.2	71.0	71.0
La_2CuO_4	041643	139	38.2	33.9	38.8	41.2	85.4
$Ca_2KF_2Fe_4As_4$	239952	139	31.7	31.2	32.7	33.4	48.3
$CaRbFe_4As_4$	252344	123	34.2	34.5	35.4	35.0	55.6
KFe_2Se_2	186573	139	37.5	31.4	33.0	40.0	36.0

25 new materials are predicted to be high T_c superconductors

Chemical formula	ICSD entry	SGN	RFR (K)	SVR (K)	DLR (K)	Average (K)	Heuristic (K)
$Tl_3Ba_4Ca_4Cu_6O_{19}$	094965	139	113.6	114.3	108.0	112.0	132.8
$Tl_5Ba_6Ca_6Cu_9O_{29}$	094966	123	113.3	113.8	106.9	111.3	137.2
$Tl_7Ba_8Ca_8Cu_{12}O_{39}$	094967	139	113.5	113.6	106.3	111.1	138.3
$TlYBa_2Cu_2O_7$	074163	123	48.4	65.6	61.7	58.6	88.8
Ba_2CuF_6	021055	69	49.5	47.9	49.5	49.0	77.2
$Ba_2Cl_2Cu_3O_4$	081196	139	70.3	60.8	37.1	56.1	80.6
$Ba_2Cu_3Br_2O_4$	036128	139	55.3	67.7	45.4	56.1	55.6
BaCuFSe	075585	129	64.1	46.2	35.0	48.4	
TeBaCuF	245624	129	54.5	30.4	33.6	39.5	
$Sr_2ZnS_2Cu_2O_2$	084735	139	56.8	35.4	34.8	42.3	
GdSeCuO	080358	129	37.3	33.4	34.2	35.0	46.9
$Sr_2Cu_2O_3$	150912	64	60.1	37.1	84.0	60.4	87.8
$Sr_3Cu_2O_5Fe_2Se_2$	154203	139	32.1	50.0	30.5	37.5	
$Sr_2F_2OFe_2Se_2$	249690	139	34.8	46.3	34.5	38.5	35.5
Ca_2CuO_3FeSe	169993	129	53.7	48.7	42.1	48.2	
NdOFeAs	164676	129	37.2	48.1	49.3	44.9	42.9
NdOFeAs	236650	67	37.2	48.1	49.3	44.9	34.1
GdOFeAs	422003	129	39.5	48.0	52.2	46.6	44.6
GdOFeAs	425015	67	39.5	48.0	52.2	46.6	31.7
$Ba_3In_2O_6$	065258	139	36.9	42.4	58.3	45.9	78.0
$Ba_3In_2Cl_2O_5$	069636	139	39.0	38.2	39.0	38.7	83.7
Ba_2InClO_3	081877	129	39.7	38.2	39.4	39.1	81.8
$Ba_3In_2Br_2O_5$	071603	139	39.4	40.3	54.7	44.8	82.6
Ba_2InBrO_3	081878	129	40.3	40.3	55.0	45.2	82.0
Ba_2N	067510	166	42.6	31.3	40.4	38.1	

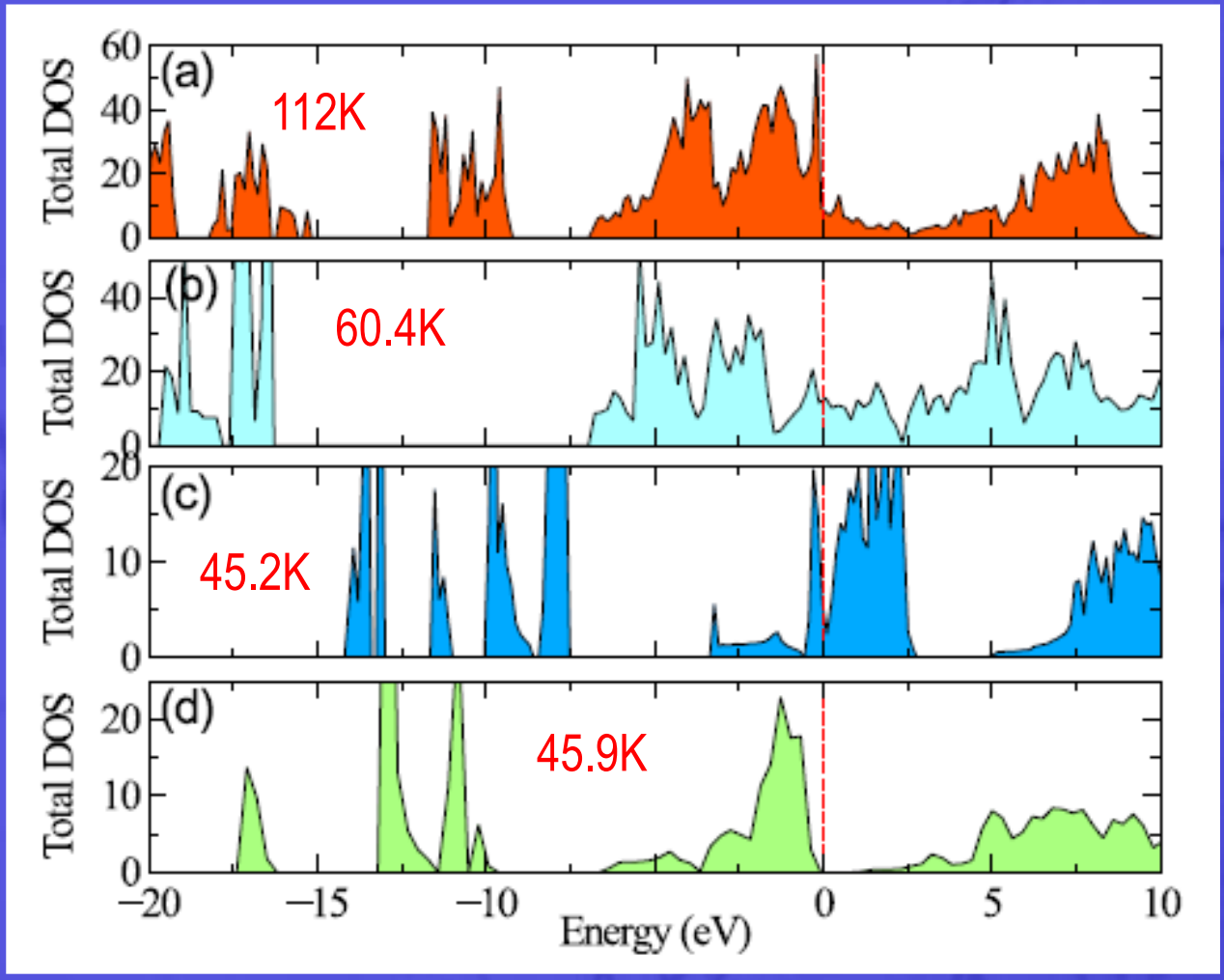
Calculated density of states of some predicted materials:

$Tl_3Ba_4Ca_4Cu_6O_{19}$

$Sr_2Cu_2O_3$

Ba_2InBrO_3

$Ba_3In_2O_6$



Role and location of the descriptors:

Material informatics for layered high-T_C superconductors

Cite as: APL Mater. 8, 061104 (2020); https://doi.org/10.1063/5.0004641
Submitted: 13 February 2020 . Accepted: 04 May 2020 . Published Online: 04 June 2020

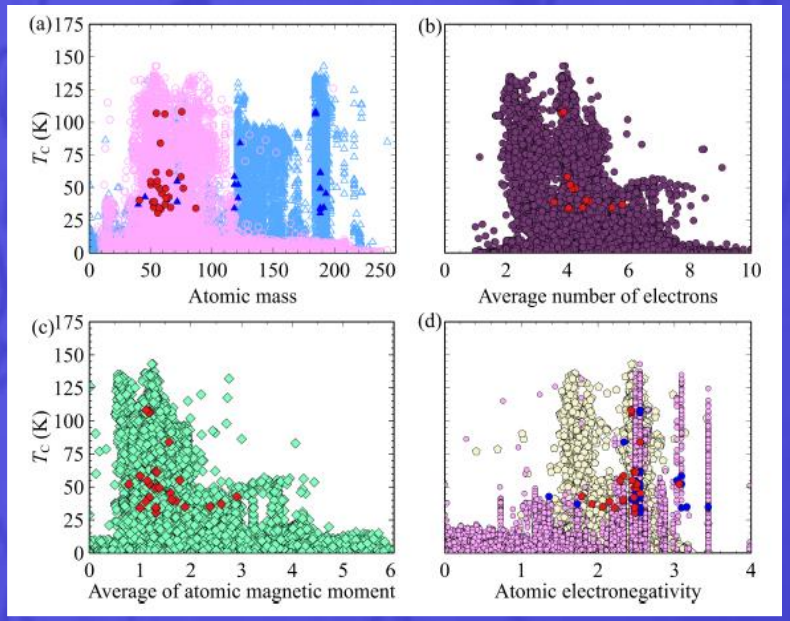
Zhong-Li Liu, Peng Kang, Yu Zhu, Lei Liu, and Hong Guo

COLLECTIONS

This paper was selected as Featured

1st place winner of the 2020 APL Materials Excellence in Research Award.
https://aip.scitation.org/apm/info/excellence

OPES	$\bar{M}$(%)	$\bar{N}$ (%)	$\bar{\mu}$ (%)	$\bar{\eta}$ (%)	ΔM (%)	$\delta\eta$ (%)
Yes	6.5	3.5	12.5	2.0	3.6	30.7
No	11.3	6.0	16.6	3.2	6.8	56.1



Outline

- Introduction: materials informatics & machine learning
- Search for solid lithium-ion conductors
- Search for high Tc superconductors
- **Search for energetic materials**
- Summary

Energetic materials:

Combustion

- Burning with flame
- Slow process
- Fully oxidized

Deflagration

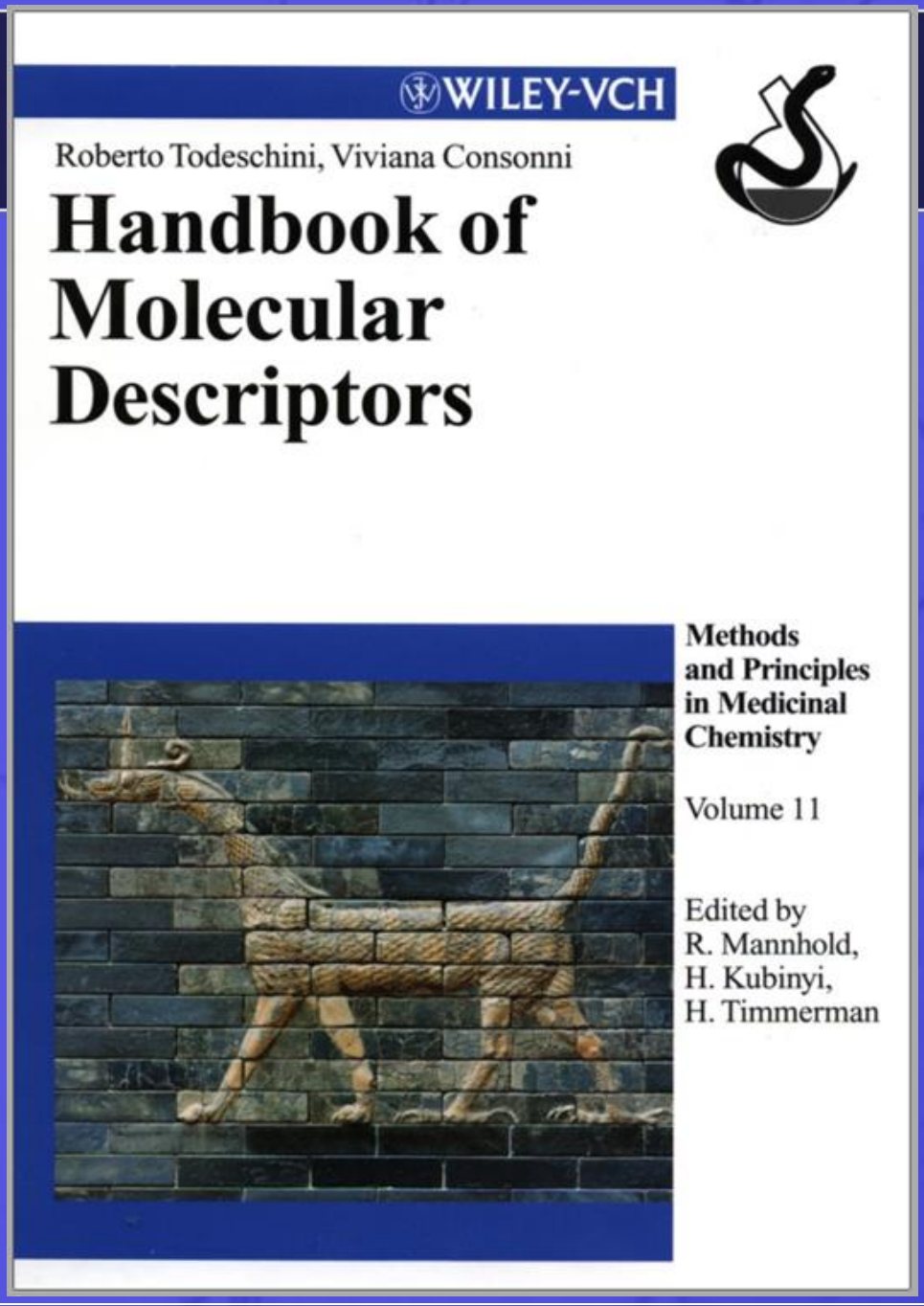
- Subsonic combustion
- Energy propagate through heat transfer
- External oxygen involved

Detonation

- Supersonic combustion
- Energy propagate though shock wave
- External oxygen not involved

Finding energetic materials can be a dangerous process. Let's try to find them by materials informatics and machine learning, verify by thermal chemistry.

What are the proper descriptors for energetic materials?



Some thermal chemistry of decomposition of $C_aH_bN_cO_d$

Example: Trinitrotoluene (TNT)

Oxygen balance (OB):

$$C_7H_5N_3O_6 \rightarrow xCO_2 + yH_2O + zN_2$$
$$C_7H_5N_3O_6 \rightarrow 7CO_2 + 2.5H_2O + 1.5N_2 - 10.5O$$
$$OB = (-10.5 \times 16)/227 \times 100\% = -74\%$$

Decomposition reaction:

Step	Conditions (K-W rules)
1	Carbon atoms are converted to CO
2	If any oxygen remains, then hydrogen is oxidized into H_2O
3	If any oxygen still remains, then CO is oxidized into CO_2
4	All the nitrogen atoms are converted to nitrogen gas N_2
	Conditions (Modified K-W rules)
1	Hydrogen atoms are converted to H_2O
2	If any oxygen remains, then carbon is converted to CO
3	If any oxygen still remains, then CO is oxidized into CO_2
4	All the nitrogen atoms are converted to nitrogen gas N_2

Some thermal chemistry of decomposition of $C_aH_bN_cO_d$ (cont.)

Heat of explosion & formation:

$$\Delta H_e = (\sum_{\text{products}} \Delta H_f - \sum_{\text{reactants}} \Delta H_f)/M$$

Volume of gas products:

$$C_3H_6N_6O_6 \rightarrow \underbrace{3CO + 3H_2O + 3N_2}_{V_g = 9 \text{ moles}}$$

Explosive power:

$$P_{e(\text{TNT})} = \frac{\Delta H_e \times V_g}{\Delta H_e(\text{TNT}) \times \text{V}_\text{g}(\text{TNT})}$$

The training set: 41 experimental data

PubChem CID	Chemical formula	OB %	ΔH_e
2537	$C_{10}H_{16}O$	-283.76	-1436.25
3026	$C_{16}H_{22}O_4$	-224.17	-1557.95
4510	$C_3H_5N_3O_9$	3.52	5117.78
6518	$C_5H_8N_4O_{12}$	-10.13	5725.39
6828	$C_{17}H_{20}N_2O$	-256.72	5.62
7434	$C_6H_3N_3O_6$	-56.34	3174.81
8376	$C_7H_5N_3O_6$	-74.01	2732.48
8490	$C_3H_6N_6O_6$	-21.62	5041.58
8561	$C_{18}H_{26}O_4$	-234.98	-1550.49
10271	$C_6H_4N_4O_6$	-56.14	2685.4
10800	$C_7H_6N_2O_5$	-96.97	1521.06
11817	$C_7H_5N_3O_7$	-62.55	2623.24
11917	$C_{15}H_{16}N_2O$	-246.67	-102.01
15353	$C_5H_9N_3O_{10}$	-26.57	4310.33
17596	$C_4H_8N_8O_8$	-21.62	5159.29
19431	$C_6H_3N_5O_8$	-32.23	4194.96
20150	$C_4H_8N_4O_8$	-26.67	4342.4
22933	$C_3H_6N_2O_6$	-28.92	4306.96
25711	$C_{14}H_{14}N_2O$	-240.71	13.69
26872	$C_4H_6N_4O_6$	-38.83	3113.79
28929	$C_{16}H_{15}N_2O$	-246.67	236.1
33603	$C_{13}H_{12}N_2O$	-233.96	6.32
34385	$C_{10}H_4N_4O_8$	-72.73	3107.86
43960	$C_6H_8N_6O_{12}$	-17.98	4525.58
61198	$C_4H_8N_2O_7$	-40.82	3929.39
61818	$C_6H_8N_6O_{18}$	7.08	5880.72
66261	$C_6H_{11}N_3O_9$	-50.56	3292.29
69050	$C_{15}H_{15}NO_2$	-235.68	-35.34
69442	$C_8H_7N_3O_6$	-89.63	2439.11
75802	$C_{13}H_{12}NO_2$	-218.99	-903.18
76423	$C_5H_9N_3O_9$	-34.51	4064.68
78227	$C_{16}H_{18}N_2O$	-251.97	114.5
83215	$C_{10}H_{16}N_6O_{19}$	-27.48	4725.03
143858	$C_8H_6N_4O_{10}$	-45.28	3316.94
143859	$C_8H_7N_3O_8$	-67.4	2306.11
267173	$C_8H_7N_3O_7$	-77.82	2334.06
522216	$C_4H_7N_3O_9$	-16.6	5381.1
3772977	$C_6H_6N_4O_4$	-88.89	2638.8
4166622	$C_9H_{12}N_4O_{13}$	-45.83	3229.11
5462985	$C_{14}H_6N_6O_{12}$	-67.56	3106.99
9889323	$C_6H_6N_{12}O_{12}$	-10.96	5910

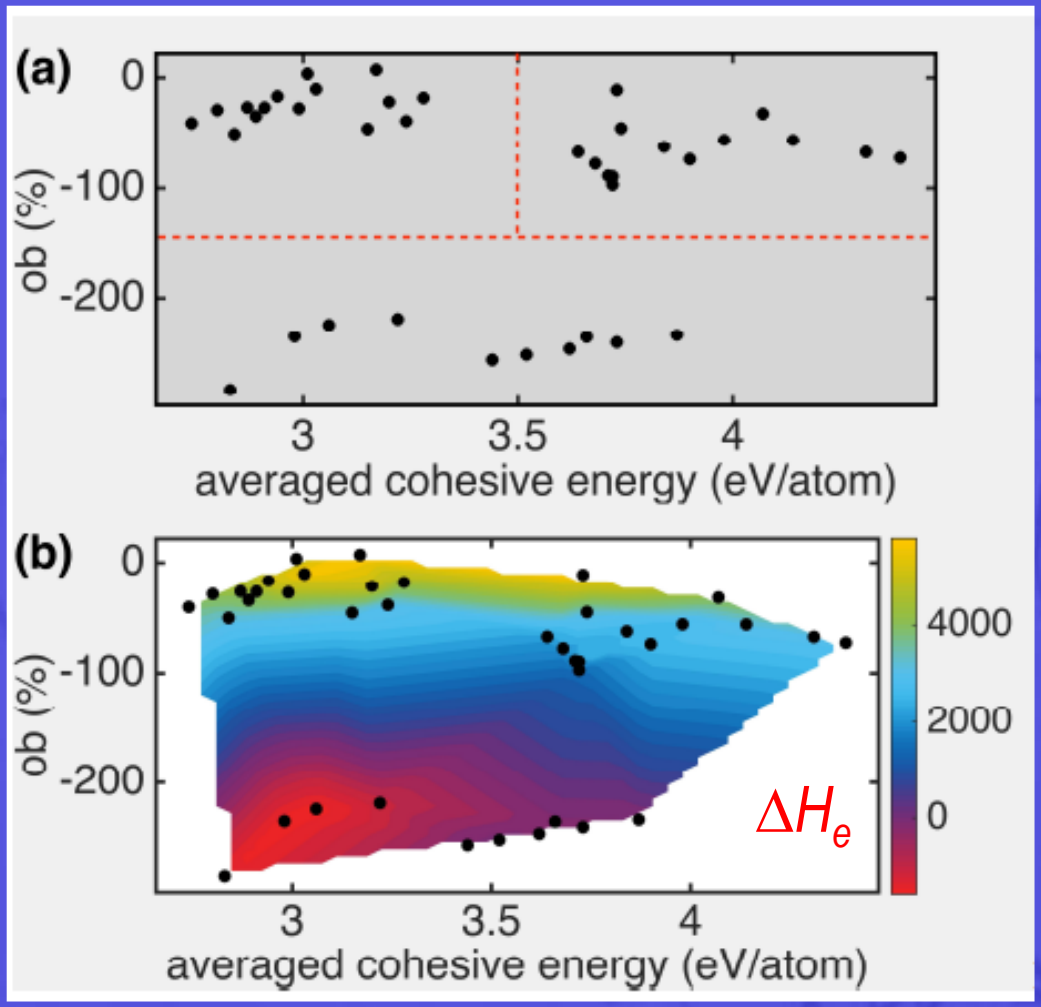
Descriptors:

- Oxygen balance;
- Cohesive energy averaged over all constituent elements'
- …

Target: ΔH_e - *heat of explosion*

Cohen, N., and S. W. Benson. "Estimation of heats of formation of organic compounds by additivity methods." Chemical Reviews 93.7 (1993): 2419-2438.

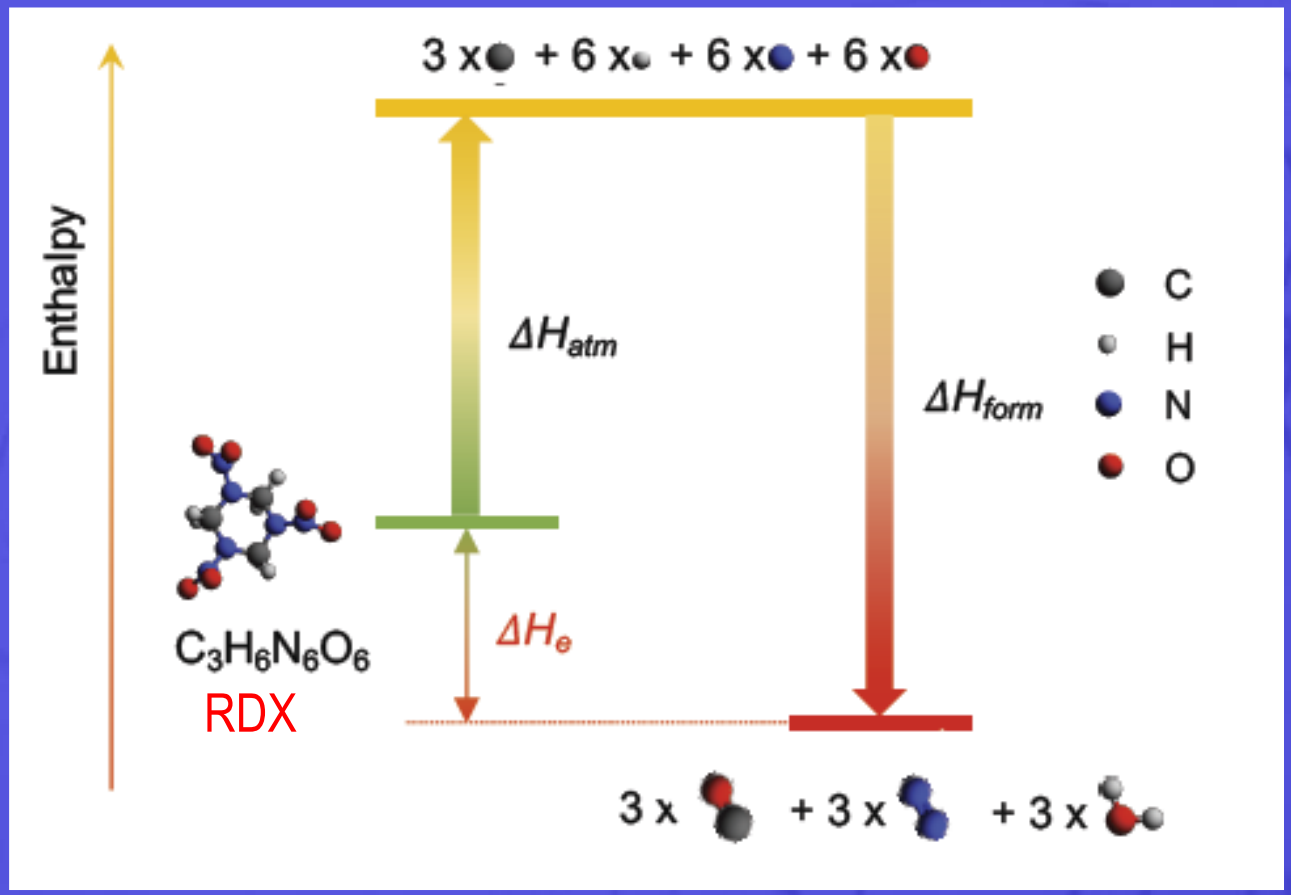
Training data distribution in descriptor space



Training data are nicely partitioned in 3 (or more) regions naturally.

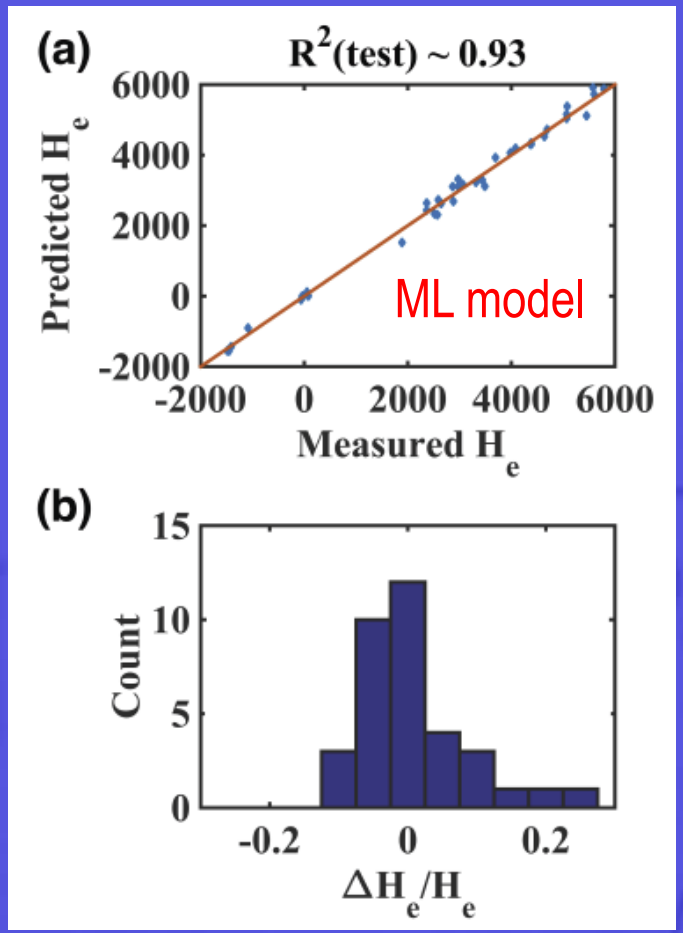
Tree based ML should be good

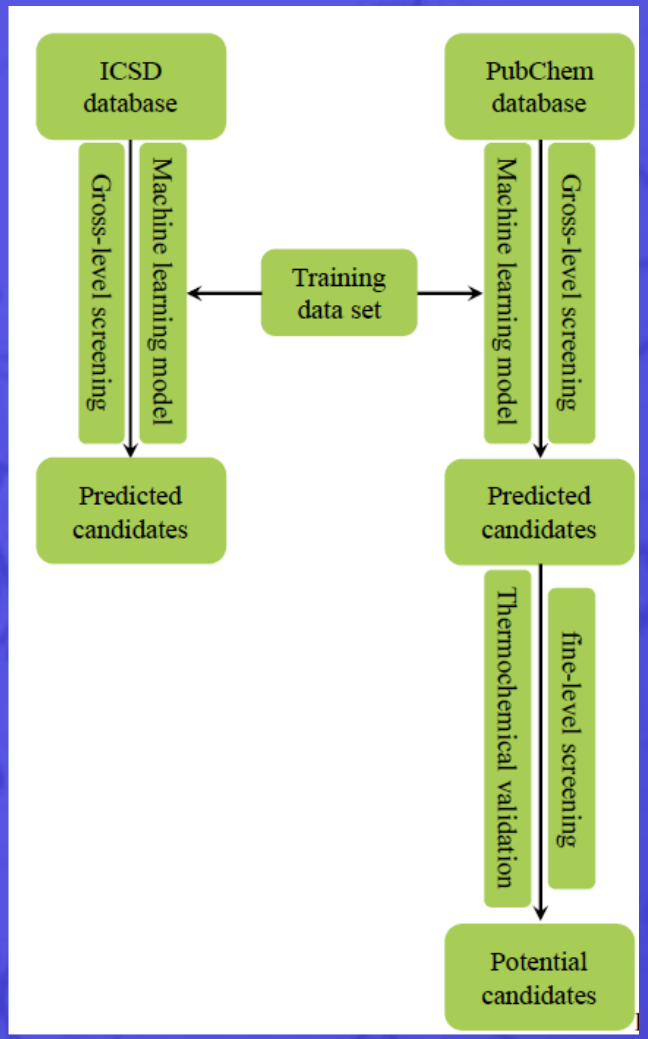
Why are OB and average cohesive energy good descriptors ?



Average cohesive energy Oxygen balance

ICSD (185000 compounds) & PubChem (~140M molecules)





Screening ICSD: 116 CHNO molecular crystals are found

Machine learning predicts 15 of the 116 to have ΔH_e > *4500 kJ/kg*, and some of them are well-known explosives:

TABLE III: The predicted and reported ΔH_e of explosive candidates screened out from ICSD database. The unit of ΔH_e is kJ/kg. [a] reported data from Ref.[61-68]

ICSD id	Chemical formula	Compound name	Predicted ΔH_e	Reported ΔH_e
161203	$C_3H_6N_6O_6$	RDX	5065.99	5075.5[a]
172533	$C_2H_4N_4O_4$	FOX-7	5065.99	3881.13[a]
417771	$C_2H_7N_7O_6$	AGNF	5510.75	5617[a]
417768	$CH_5N_5O_6$	HNF	5393.63	5443 ~ 5451[a]
417769	$C_2H_8N_6O_7$	GNFH	5391.43	5250[a]
417772	$C_2H_8N_8O_6$	DAGNF	5357.92	5853[a]
417770	$C_2H_9N_9O_6$	TAGNF	4982.49	6274[a]
281338	$CH_3N_5O_4$	1,2-Dinitroguanidine	5600.27	
170681	CH_2N_6O	Nitroguanyl azide	5576.16	
084595	$C_2H_9N_{11}O_8$	$(BIGH_2)(DN)_2$	5449.78	
034697	$CH_5N_3O_4$	Urea nitrate	5371.68	
250532	$CH_6N_4O_4$	HDM	5362.20	
281711	$CH_6N_6O_4$	Guanidinium Dinitramide	5352.97	
281712	$CH_6N_6O_5$	Hydroxyguanidinium Dinitramide	5342.99	
082738	$C_2H_9N_7O_6$	$(BIGH_2)(NO_3)_2$	4669.22	

So, our machine learning model of energetic materials works

Screening PubChem: ~140 million molecules

ML model predicts **2732** CHNO candidates to have ΔH_e > *4500 kJ/kg !*

Next: do thermal chemistry validation (since there are no experimental data).

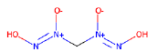
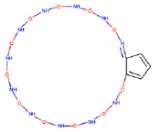
Results: equivalent power of explosion to TNT

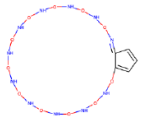
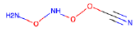
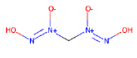
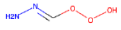
If $$P_{e(\mathrm{TNT})} = \frac{\Delta H_e \times V_g}{\Delta H_e(\mathrm{TNT}) \times \mathrm{V_g(TNT)}}$$ > 1.5, **162** candidates remain;

If requiring 1.8 power of explosion of TNT, **29** candidates remain.

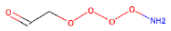
29 energetic materials not known before:

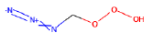
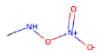
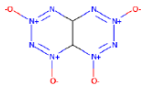
TABLE IV: The screened out 29 candidates from PubChem database with $P_{e(TNT)} > 1.8$ are listed. The units of ΔH_f, ΔH_e and V_g are kJ/mol, kJ/kg and dm^3/kg, respectively.

PubChem CID	Chemical formula	2D molecular depiction	OB(%)	ΔH_f	ΔH_e	ΔH_e(predicted)	V_g	$P_{e(TNT)}$
135474996	$CH_4N_4O_4$		0.00	51.41	7151.01	5337.90	823.53	1.86
53906144	CH_3NO_3		-10.39	-95.66	7169.75	5352.97	872.73	1.98
21621129	$CH_3N_5O_2$		-20.51	354.82	6232.06	4723.47	957.26	1.89
100257582	$C_5H_{13}N_{11}O_{11}$		-21.84	1016.61	7821.56	4968.48	944.91	2.34
90058261	$CH_{11}N_7O_8$		3.21	346.54	8799.19	5342.99	899.60	2.50

PubChem CID	Chemical formula	2D molecular depiction	OB	ΔH_f	ΔH_e	ΔH_e(predicted)	V_g	$P_{e(TNT)}$
135981456	$CH_4N_4O_4$		0.00	51.41	7151.01	5337.9	823.53	1.86
90471911	$C_5H_{13}N_{11}O_{11}$		-21.84	1016.61	7821.56	4968.48	944.91	2.34
58638113	$CH_3N_3O_3$		-7.62	103.77	7157.18	5461.45	853.33	1.93
135538843	$CH_4N_4O_4$		0.00	51.41	7151.01	5337.90	823.53	1.86
58170408	$CH_4N_2O_3$		-17.39	-84.16	6021.57	4669.22	973.91	1.85
117768295	$C_4H_{11}NO_{11}$		-16.06	-524.62	7200.42	4969.98	899.60	2.05

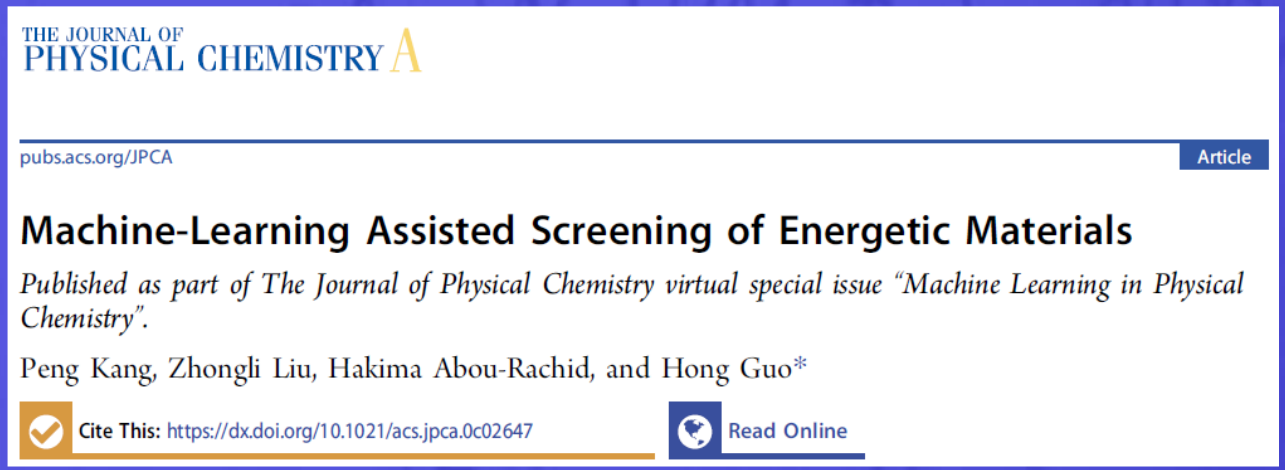
29 energetic materials not known before (cont.):

PubChem CID	Chemical formula	2D molecular depiction	OB	ΔH_f	ΔH_e	ΔH_e(predicted)	V_g	$P_{e(TNT}$
129689476	$CH_5N_3O_3$		-22.43	21.89	6168.55	4641.80	1046.73	2.04
117768110	$C_2H_5NO_5$		-19.51	-217.42	6541.84	4655.67	910.57	1.88
129691886	$CH_3N_5O_2$		-20.51	344.72	6145.74	4723.47	957.26	1.86
21494569	CH_3NO_3		-10.39	-49.01	7775.65	5352.97	872.73	2.14
129634872	$CH_2N_2O_2$		-21.62	121.75	6703.89	5065.99	908.11	1.92
123320779	$CH_4N_2O_4$		0.00	-116.23	7452.79	5342.99	829.63	1.95

PubChem CID	Chemical formula	2D molecular depiction	OB	ΔH_f	ΔH_e	ΔH_e(predicted)	V_g	$P_{e(TNT}$
117787017	$CH_3N_3O_3$		-7.62	83.05	6959.85	5461.45	853.33	1.88
57500267	$CH_4N_2O_3$		-17.39	-49.32	6400.24	4669.22	973.91	1.97
134861857	$C_2H_2N_8O_4$		-7.92	720.43	7367.71	5564.05	776.24	1.81
71319450	$CH_2N_2O_3$		0.00	47.65	7832.93	5708.30	746.67	1.85
18402732	$CH_2N_2O_2$		-21.62	117.90	6651.76	5065.99	908.11	1.91
89774793	$CH_4N_6O_2$		-24.24	513.94	6729.31	4780.01	1018.18	2.17

Next ?

- First principles calculation to further validate these 29 candidates.
- Calculate other useful properties (sensitivity…).
- Speaking to experimentalists.



THE JOURNAL OF PHYSICAL CHEMISTRY A

pubs.acs.org/JPCA Article

Machine-Learning Assisted Screening of Energetic Materials

Published as part of The Journal of Physical Chemistry virtual special issue "Machine Learning in Physical Chemistry".

Peng Kang, Zhongli Liu, Hakima Abou-Rachid, and Hong Guo*

Cite This: https://dx.doi.org/10.1021/acs.jpca.0c02647 Read Online

Summary:

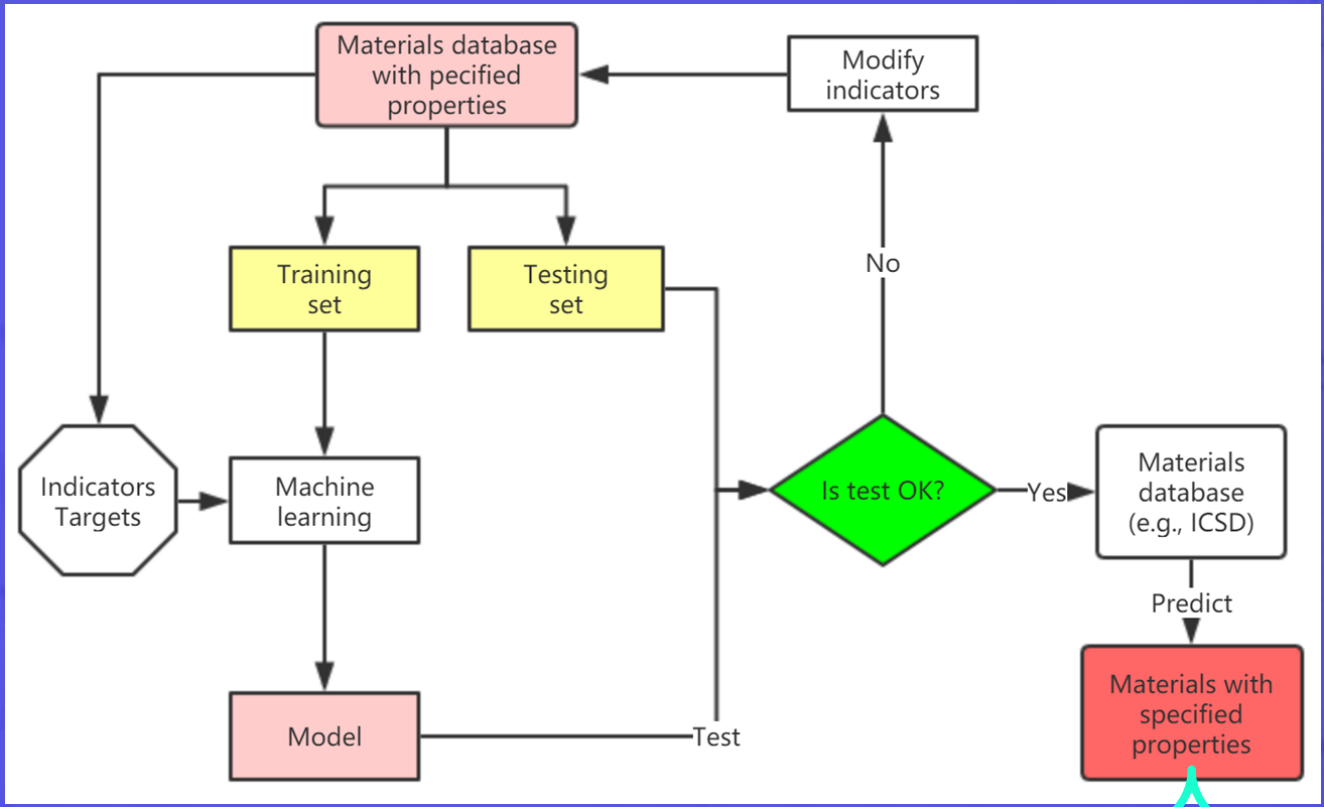
- Materials informatics is the 4th paradigm of scientific exploration. It is NOT for finding very precise numbers, but for efficiently classifying materials.

- Three important ingredients:
  - Data: those who have data are the winners
  - Methods to analyze data: so far standard and public domain
  - Physical indicators: critical !

- **iMAT:** pre-screening + machine learning + simulation

Simulation + experimental verification

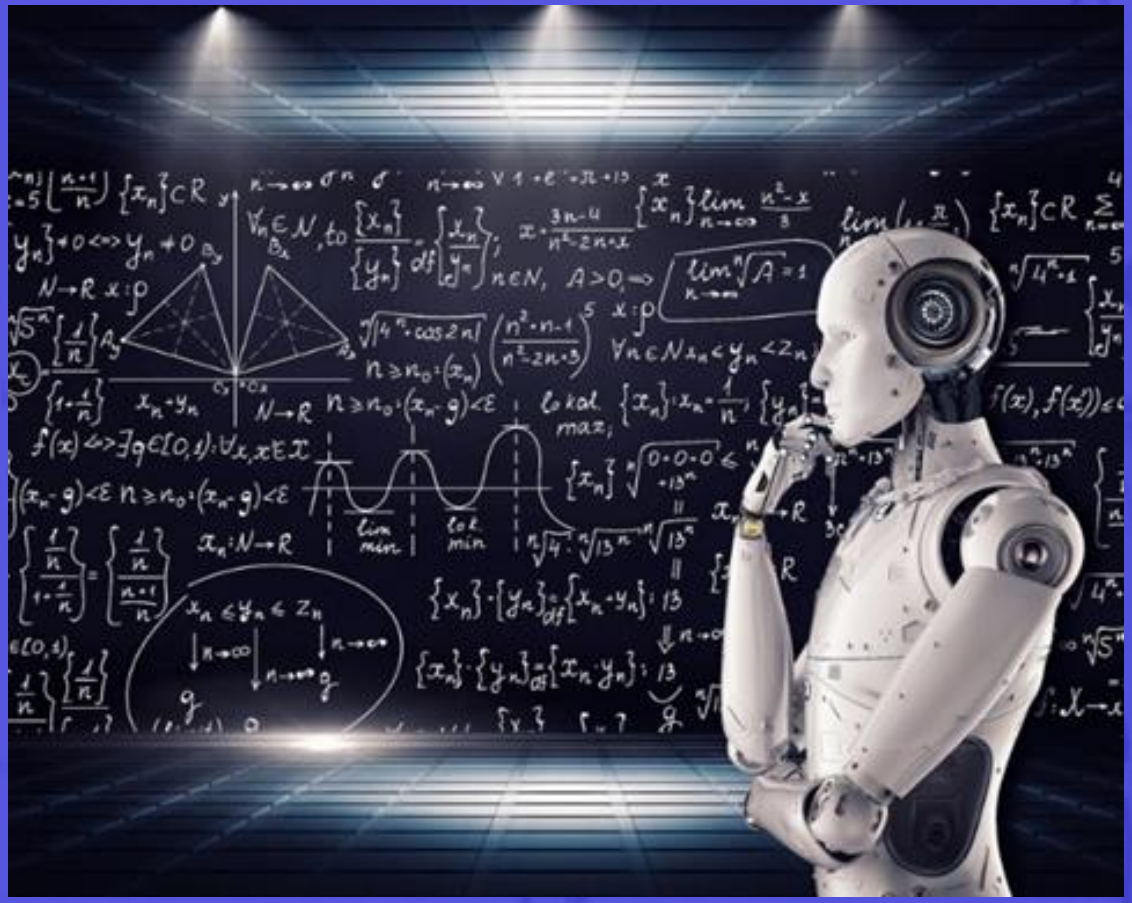
Technical summary: iMAT



simulation

experiment

What will be the 5th paradigm ?



Thank you !