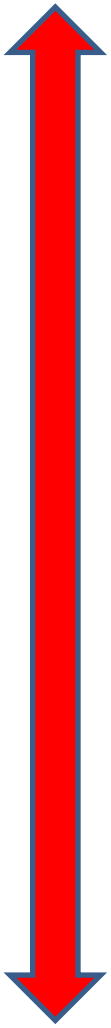
Discovery of Functional Molecules and Materials by Artificial Intelligence Techniques

津田　宏治
東京大学

Discovery of new functional molecules and materials is of national importance

First Principles Calculations

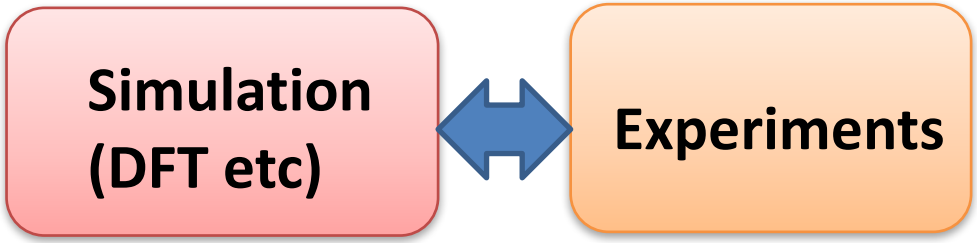
Accurate, Slow



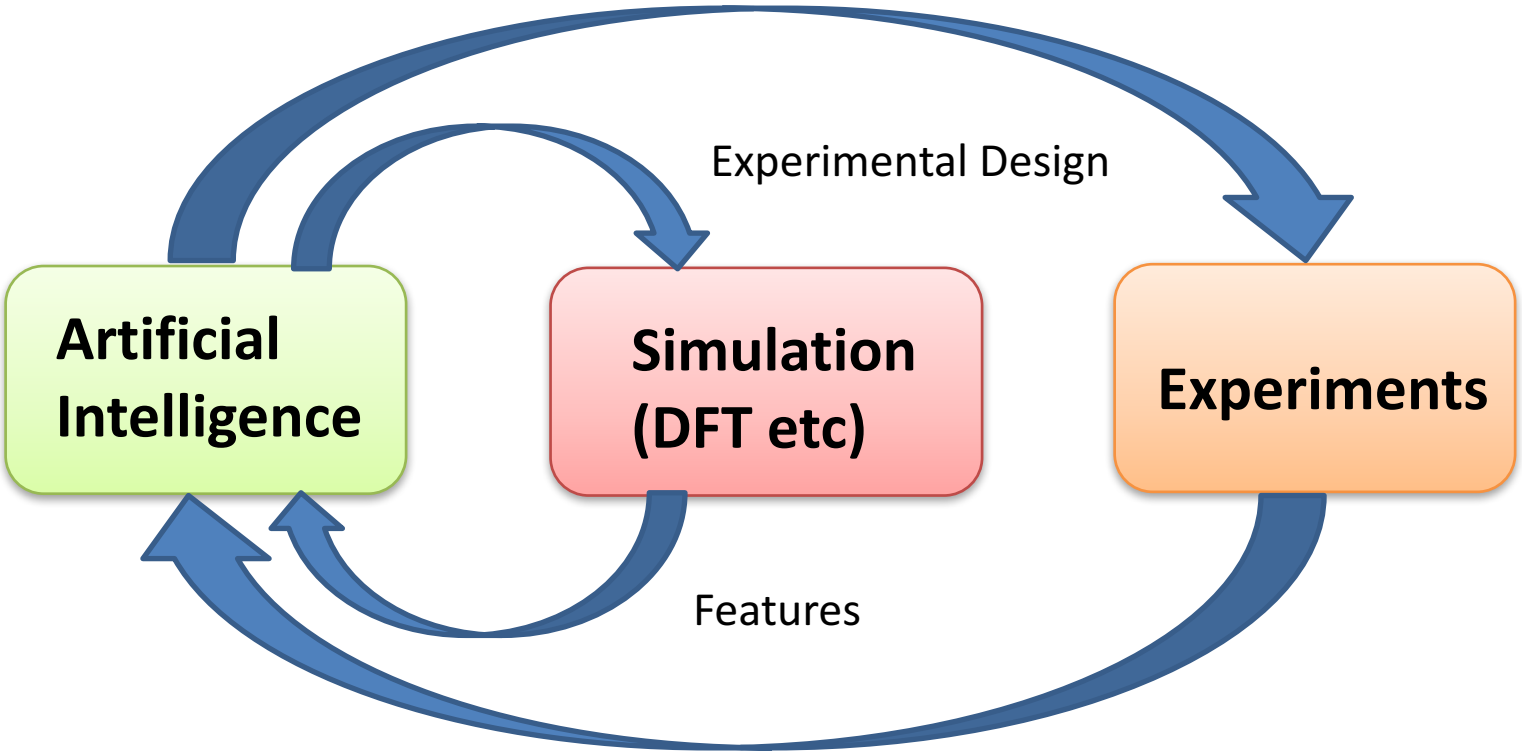
- Full configuration interaction
- Wave function based
- Density functional theory
- Semi-empirical
- Empirical potentials

Inaccurate, Fast

Old Picture



New Picture


Experimental Design
Artificial Intelligence
Simulation (DFT etc)
Experiments
Features

3 Techniques and 4 Studies

- Virtual Screening
  - Discovering low-LTC compounds
- Bayesian Optimization
  - Optimization of grain boundary
  - Design of Si-Ge nanostructures
- Monte Carlo Tree Search
  - Designing optimal RNA sequences

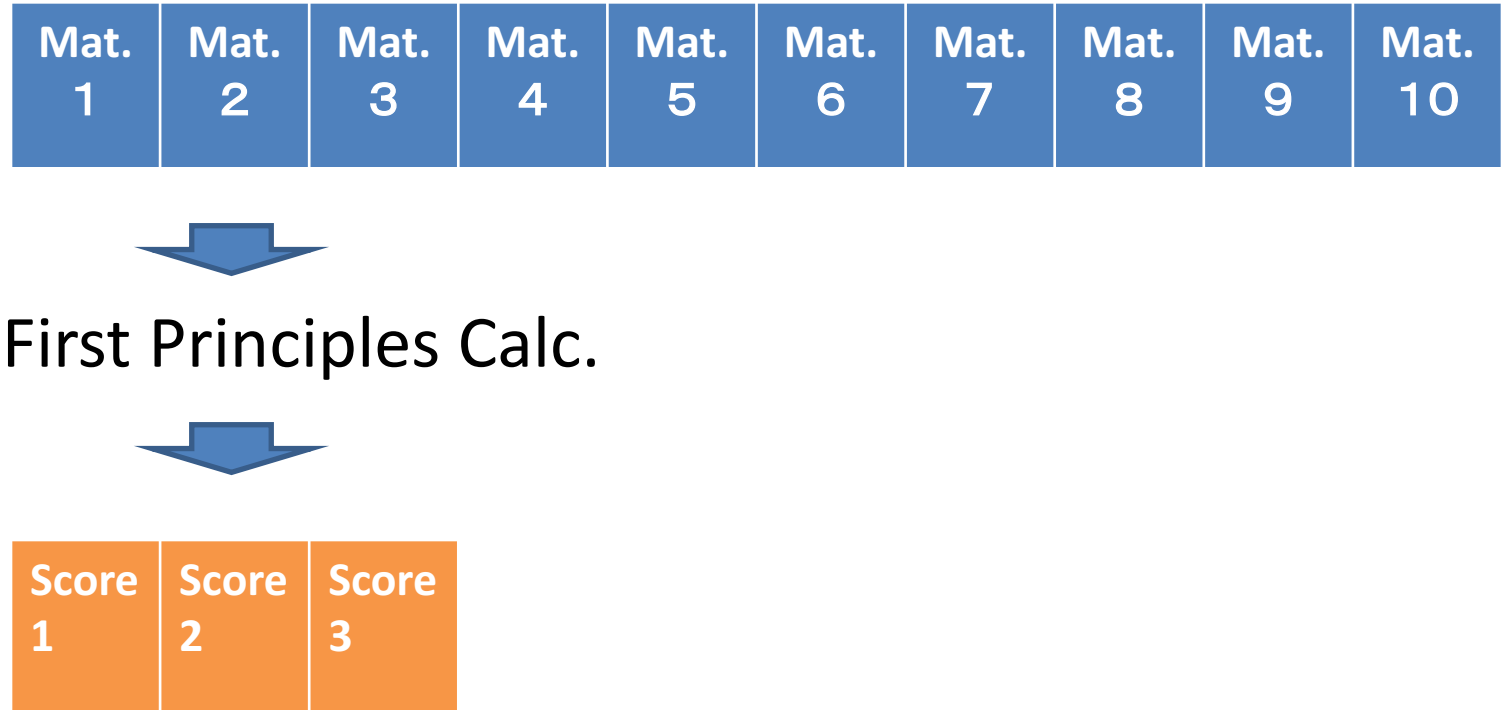
Screening by first principles calculations alone

Mat. 1	Mat. 2	Mat. 3	Mat. 4	Mat. 5	Mat. 6	Mat. 7	Mat. 8	Mat. 9	Mat. 10

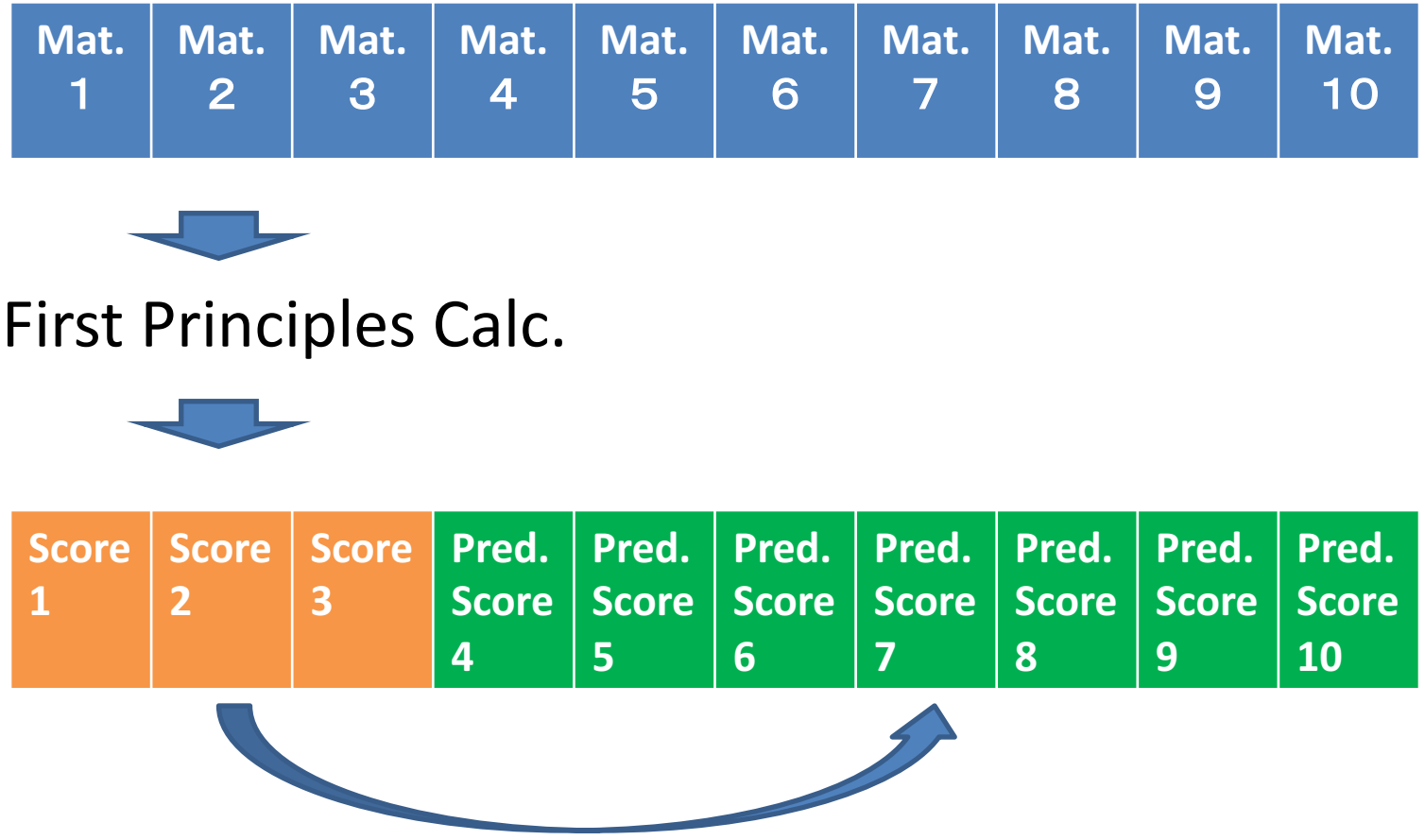
First Principles Calc.

Score 1	Score 2	Score 3	Score 4	Score 5	Score 6	Score 7	Score 8	Score 9	Score 10

Virtual Screening



Virtual Screening



マテリアルズ・インフォマテ ×

Koji

www.kyoto-u.ac.jp/ja/research/research_results/2015/151113_1.html

アプリ　Gmail - 受信トレイ -　Pocket：マイリスト　ログイン - サイボウズ　GitLab Tsudalab　JBpress(日本ビジネス　ホーム - 東大ポータル　ナノ構造情報のフロン　»　その他のブックマーク

MENU

京都大学
KYOTO UNIVERSITY

研究成果

マテリアルズ・インフォマティクス手法により超低熱伝導物質を高効率に多数発見 －材料科学と情報科学の融合研究に革新的成果－

2015年11月13日

世古敦人 工学研究科准教授、林博之 同特定助教、田中功 同教授、東後篤史 学際融合教育研究推進センター構造材料元素戦略研究拠点ユニット（ESISM）特定准教授らは、津田宏治 東京大学新領域創成科学研究科教授、Laurent Chaput フランス・ロレーヌ大学准教授との共同研究により、材料科学と情報科学が融合したマテリアルズ・インフォマティクス手法に基づいて、超低熱伝導物質を高効率に多数発見するという革新的な成果を上げました。従来知ら

Thermoelectric materials

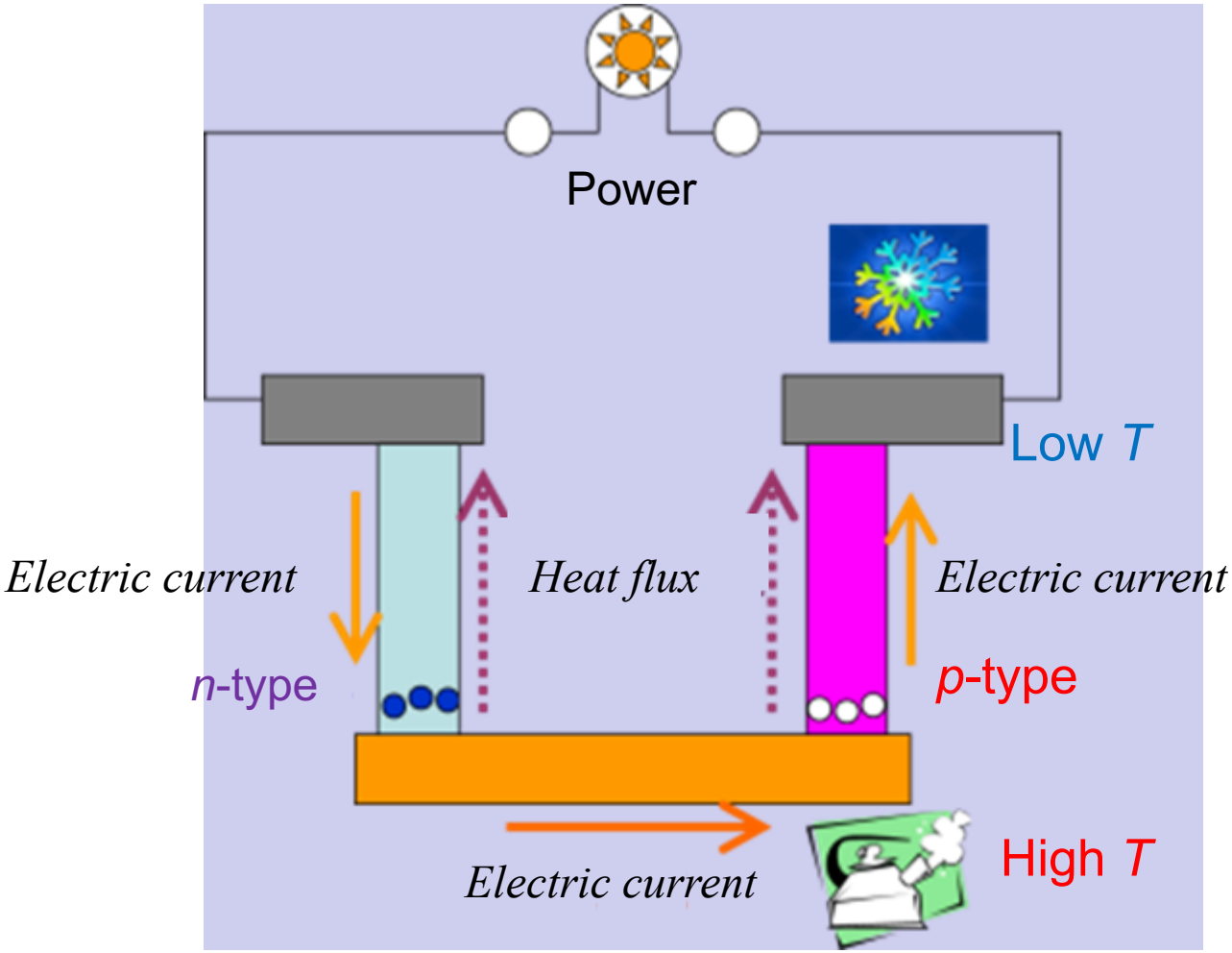


Figure of Merit $(S^2 \sigma/\kappa)T$

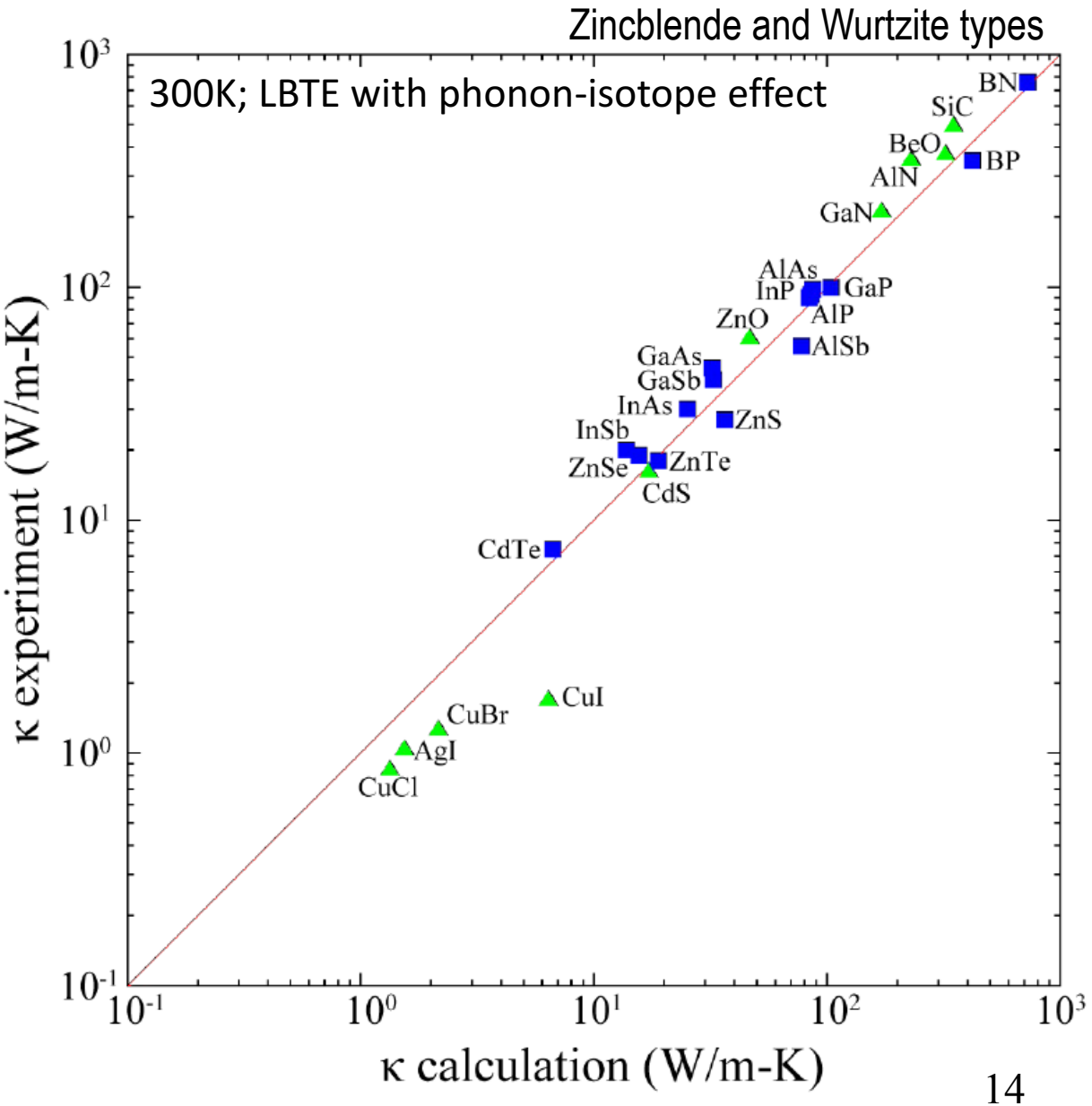
S: Seebeck coefficient
σ: electrical conductivity
κ: thermal conductivity

Discovering Compounds of Low Thermal Conductivity: Motivation

- Isao Tanaka's Lab developed a system capable of calculating lattice thermal conductivity (LTC)
  - First-principles anharmonic lattice dynamics calculations
  - Solving Boltzmann transport equation with the single-mode relaxation-time approximation

- Too slow for screening in a large database
  - One LTC can take one WEEK with hundreds of cores

Lattice thermal conductivity : calc. vs. exp.



14

Discovering Compounds of Low Thermal Conductivity from Database

(Seko et al., PRL 2015)

- Compute LTC of 101 prototypical compounds
  - Rocksalt, Zincblende, Wurtzite-type
  - Best LTC: 0.9 W/m·K
- Predict LTC of 54779 compounds in Materials Project Database
  - Additional LTC calculations for best 8 compounds
  - Five had impressive LTC of < 0.2 W/m·K (@300K)

TABLE I. First principles LTCs and Z-scores for highly ranked compounds by the virtual screening. Band gaps by DFT-PBE are taken from MPD library[29, 33].

Ranking	Z-score	Formula	Space group	LTC (W/mK)	Band gap (eV)
1	1.90	$PbRbI_3$	$Pnma$	0.10	2.46
2	1.76	PbIBr	$Pnma$	0.13	2.56
3	1.56	$PbRb_4Br_6$	$R\bar{3}c$	0.08	3.90
4	1.56	PbICl	$Pnma$	0.18	2.72
5	1.56	PbClBr	$Pnma$	0.09	3.44
7	1.44	PbI_2	$R\bar{3}m$	0.29	2.42
8	1.43	PbI_2	$P6_3mc$	0.29	2.45
121	0.39	K_2CdPb	$Ama2$	0.45	0.18
144	0.29	$Cs_2[PdCl_4]I_2$	$I4/mmm$	0.31	0.88

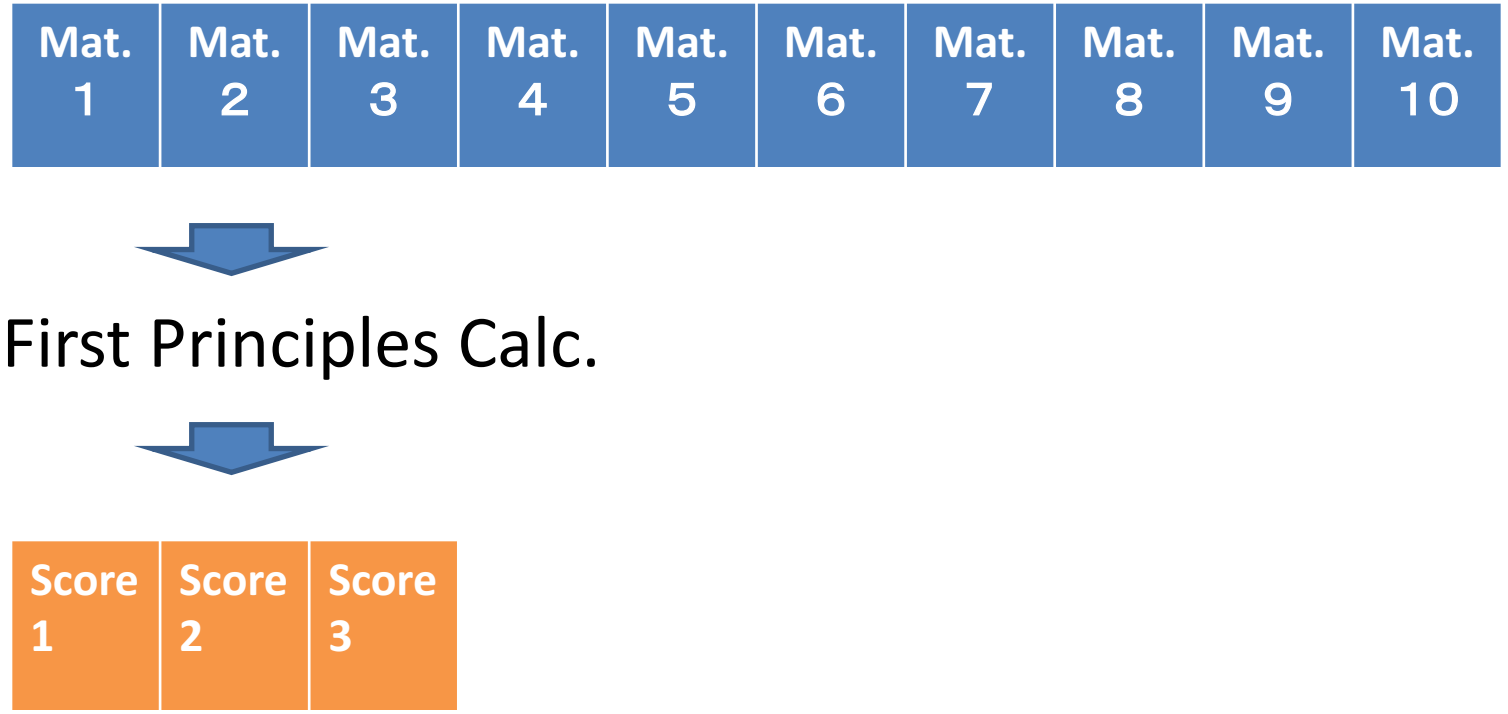
Bayesian Optimization

(Jones et al., 1998)

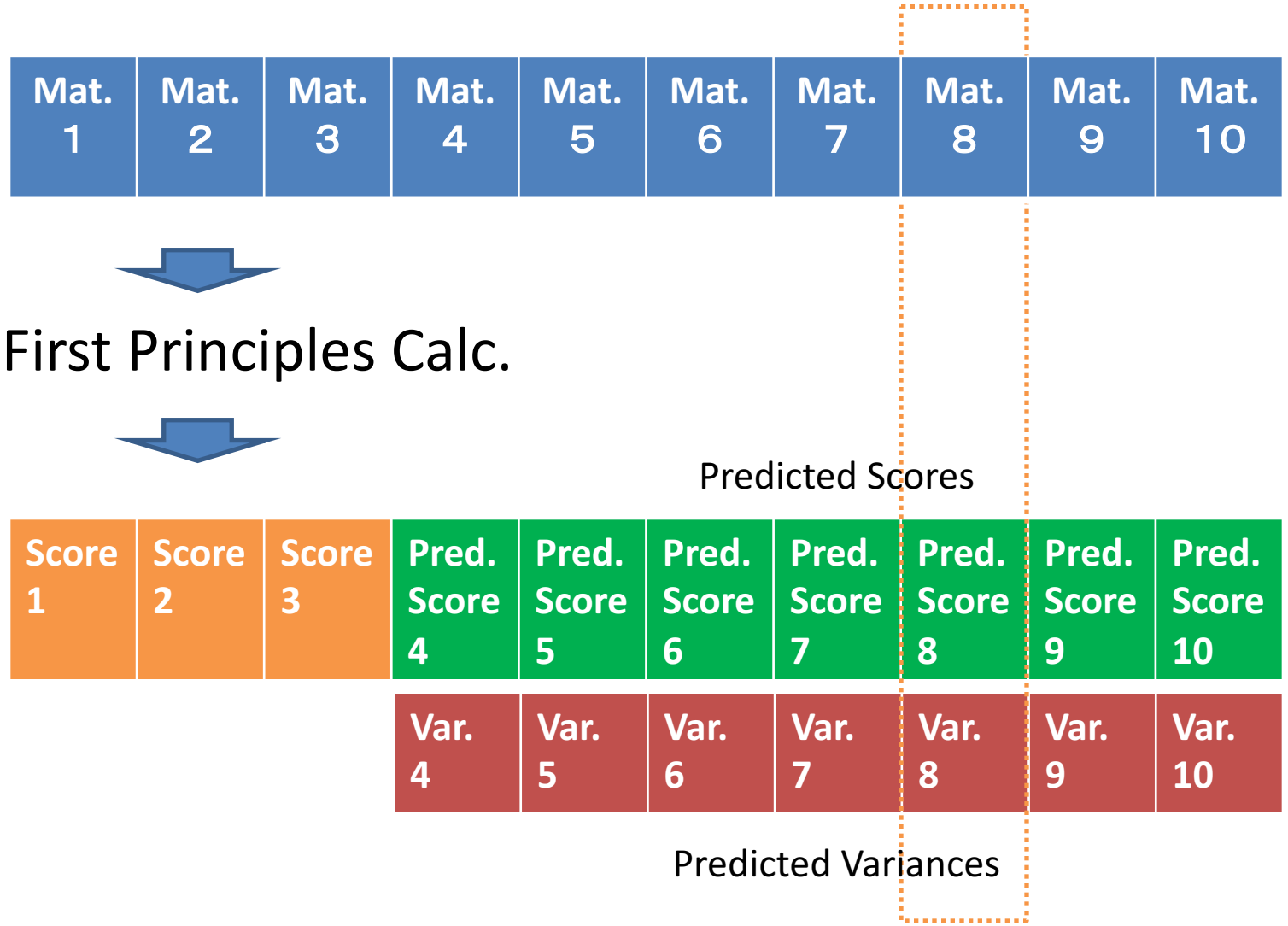
- Find best data points with minimum number of observations

- Choose next point to observe to discover the best ones as early as possible

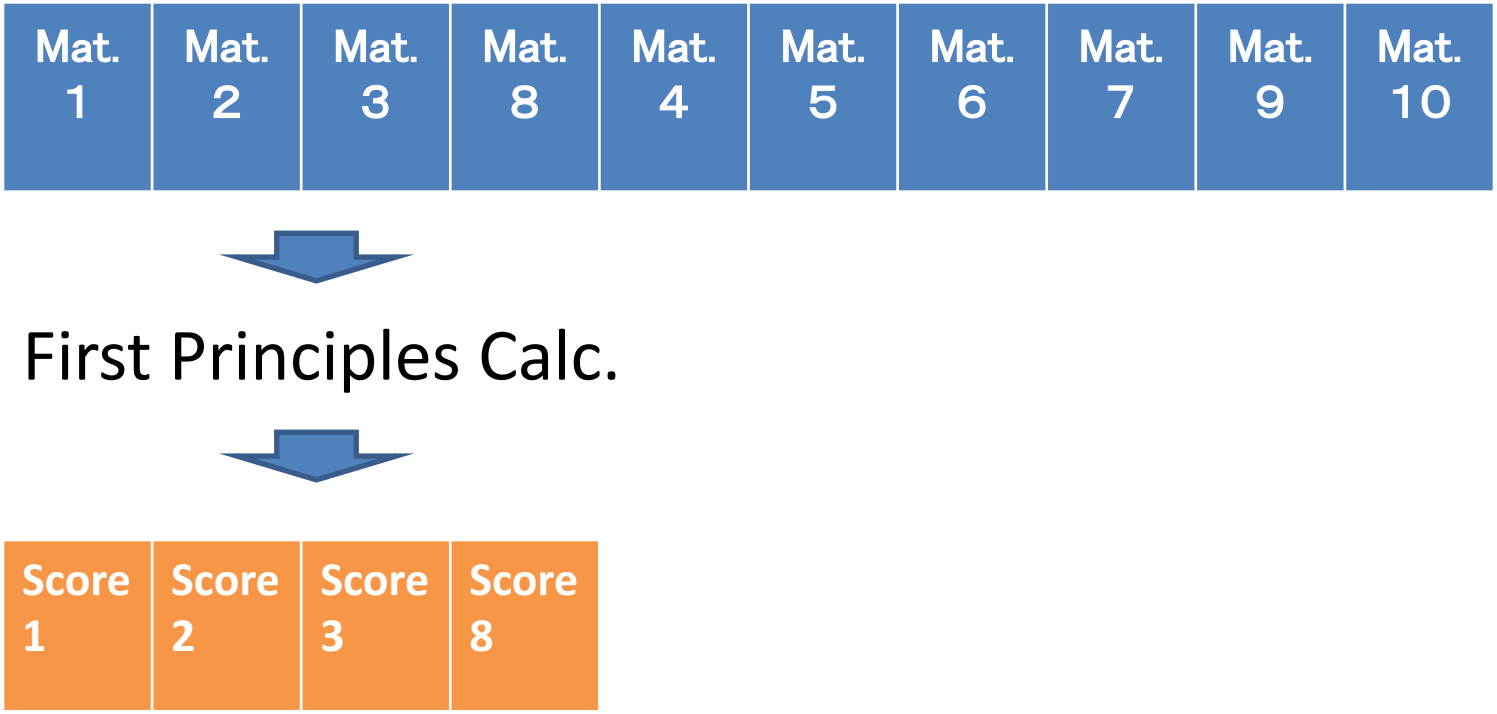
Bayesian Optimization (1)



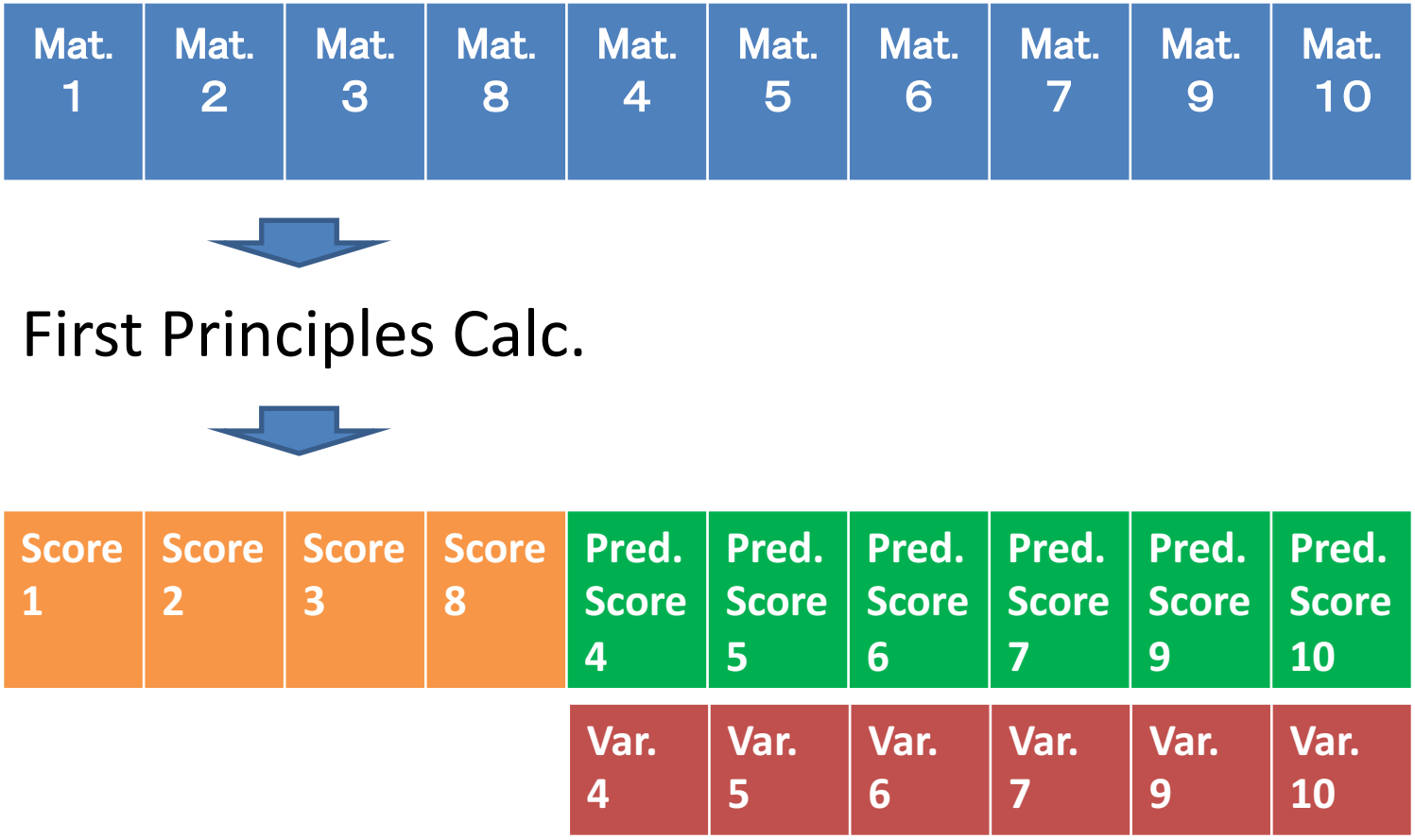
Bayesian Optimization (2)



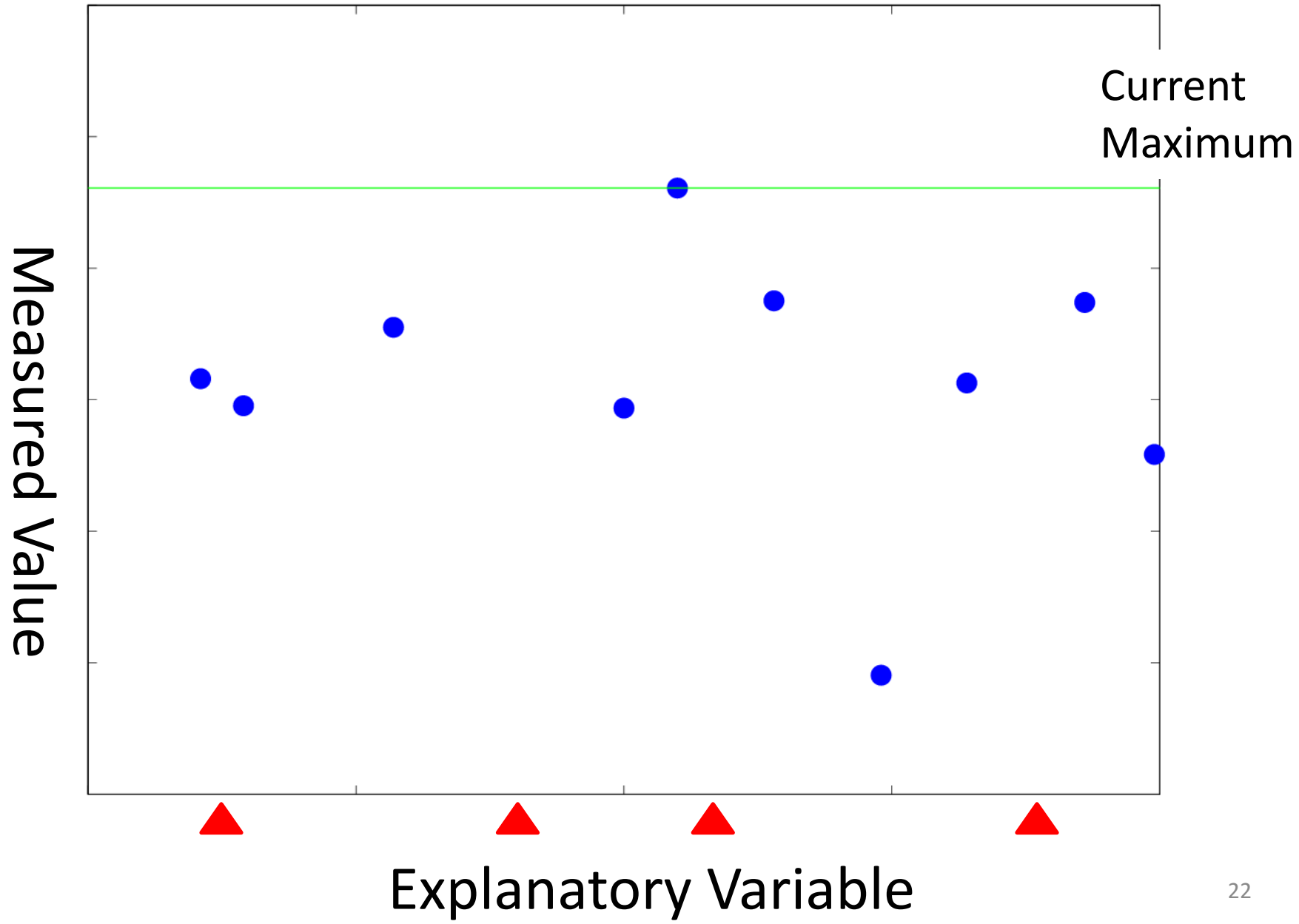
Bayesian Optimization (3)



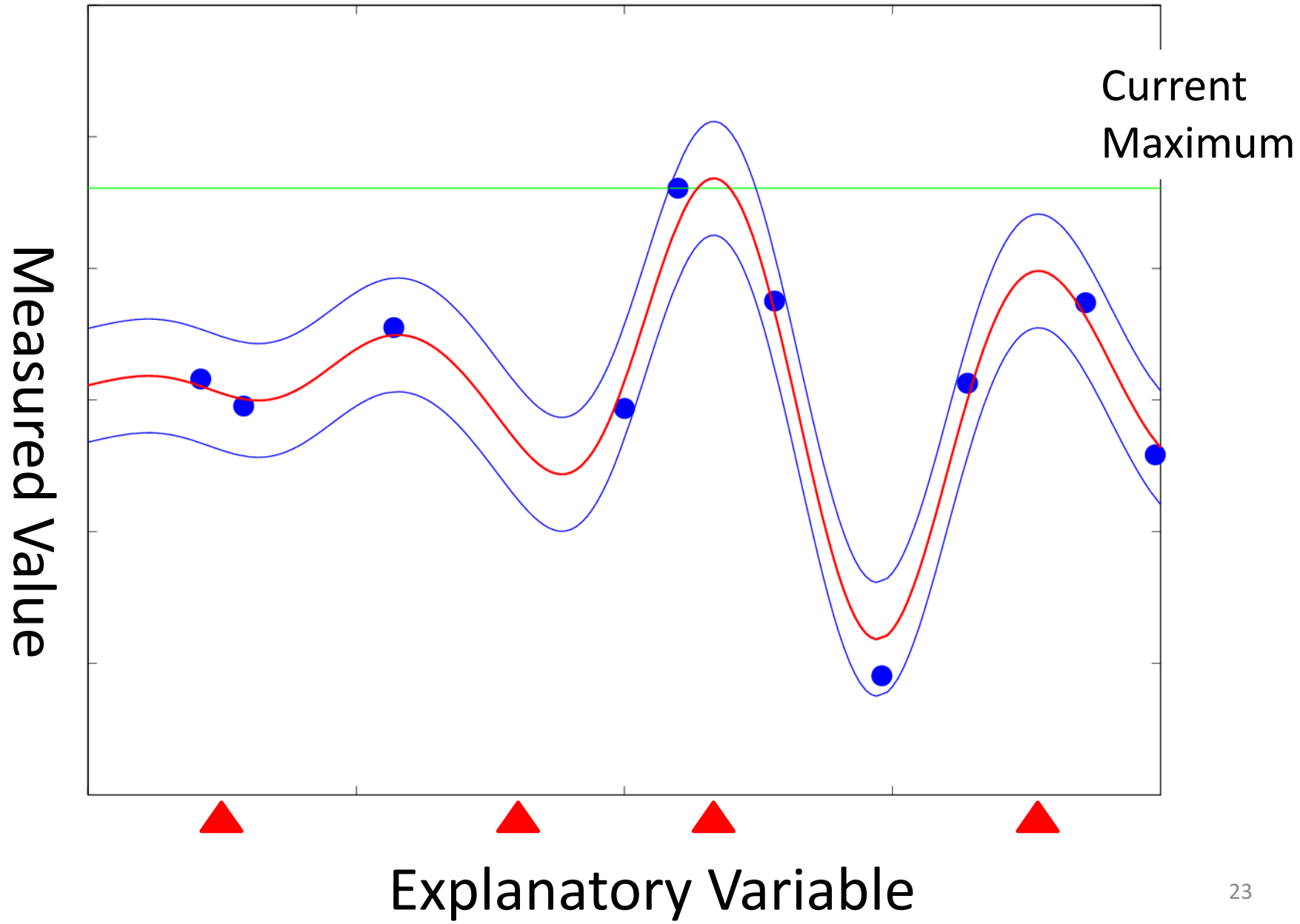
Bayesian Optimization (4)



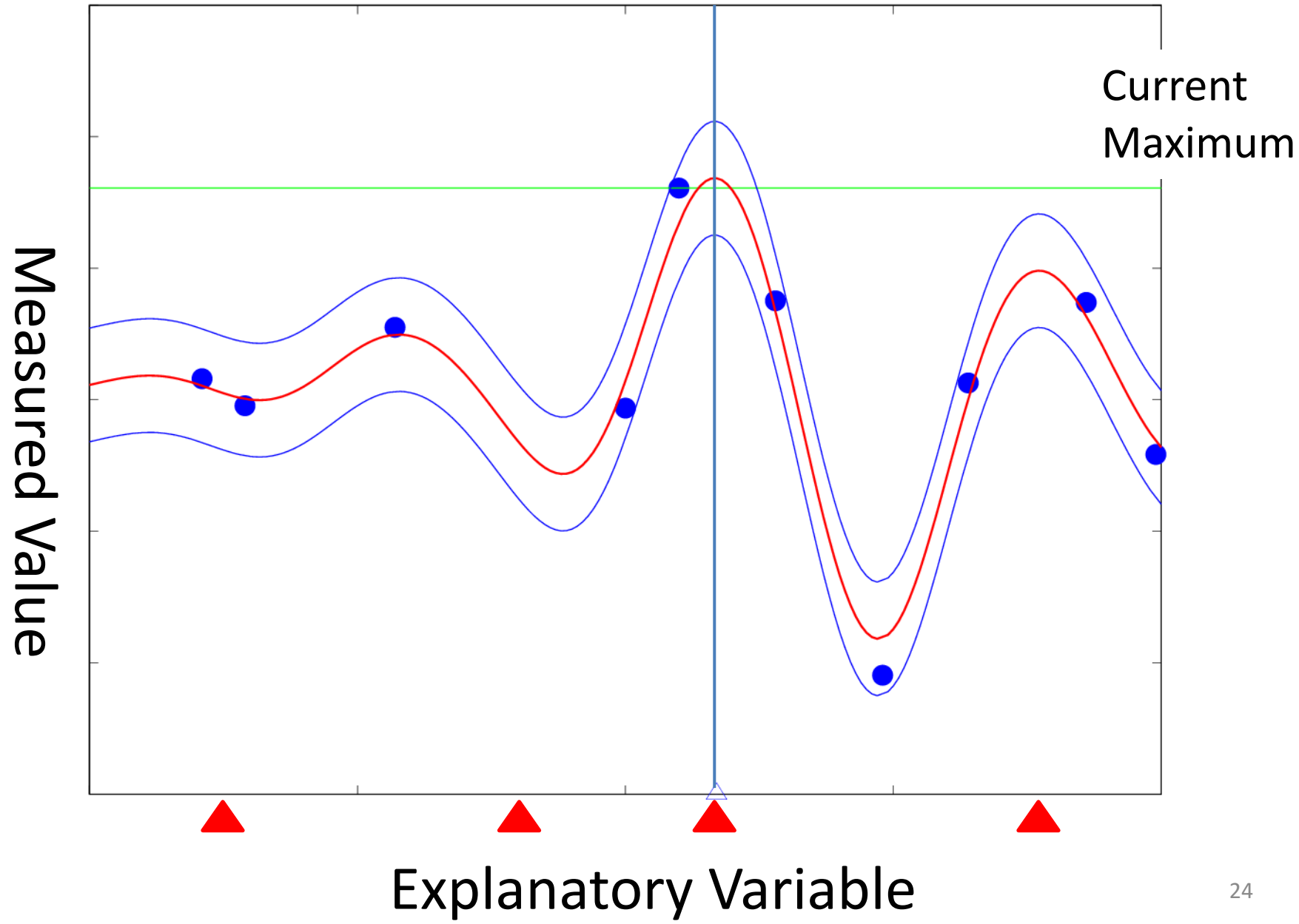
Where to observe next?



Gaussian Process



Maximum probability of improvement



Finding single and binary component solids with high melting temperature

(Seko et al., PRB 2014)

- Apply Bayesian optimization (BO) to discover the one with highest melting temperature from 226 materials

- Initially, 5% of materials are chosen randomly and their melting temperatures are observed
- After that, next point to observe is determined by BO

Descriptors

DFT Calculated Features

Cohesive Energy

Bulk modulus

Volume

Nearest-neighbor pair distance

Elemental Features

Composition

Atomic number

Atomic mass

Number of valence electrons

Group

Period

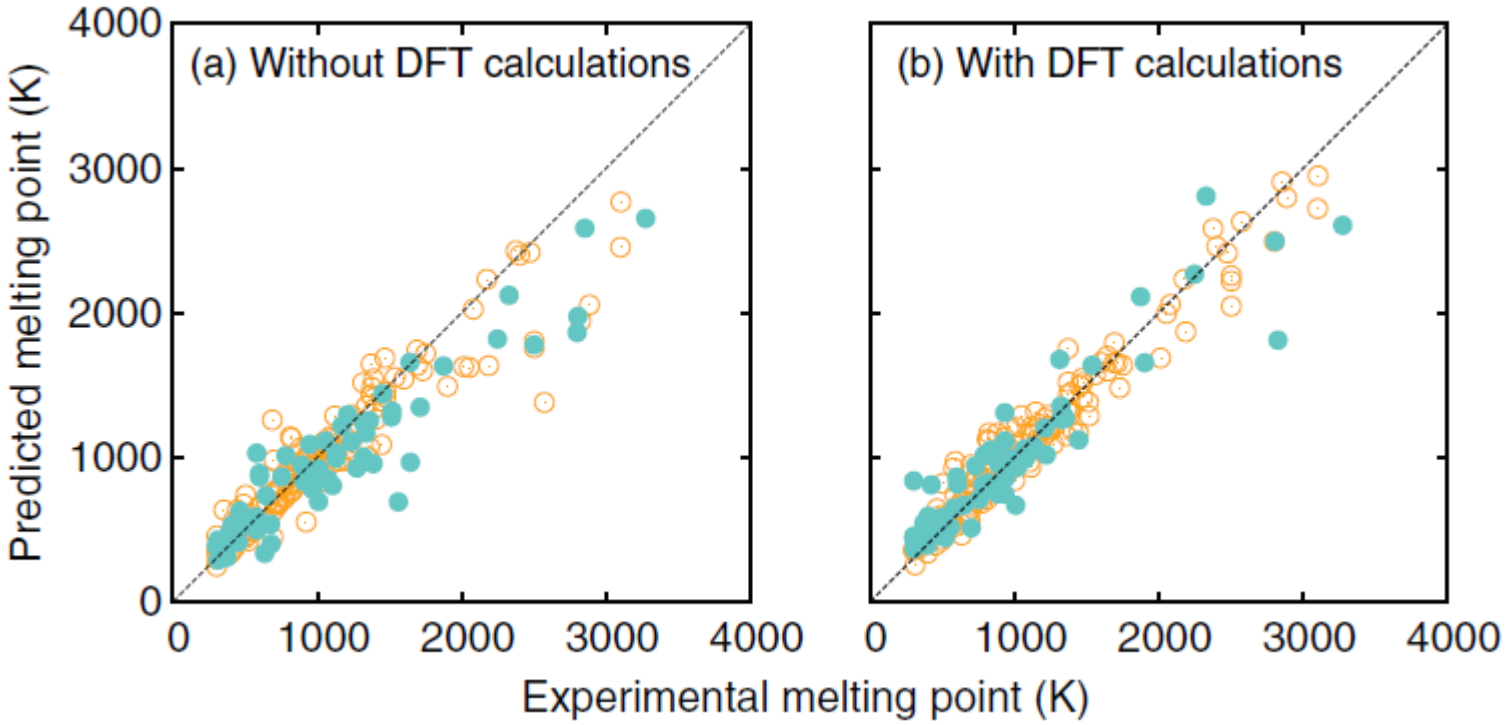
Van del Waals radius

Covalent radius

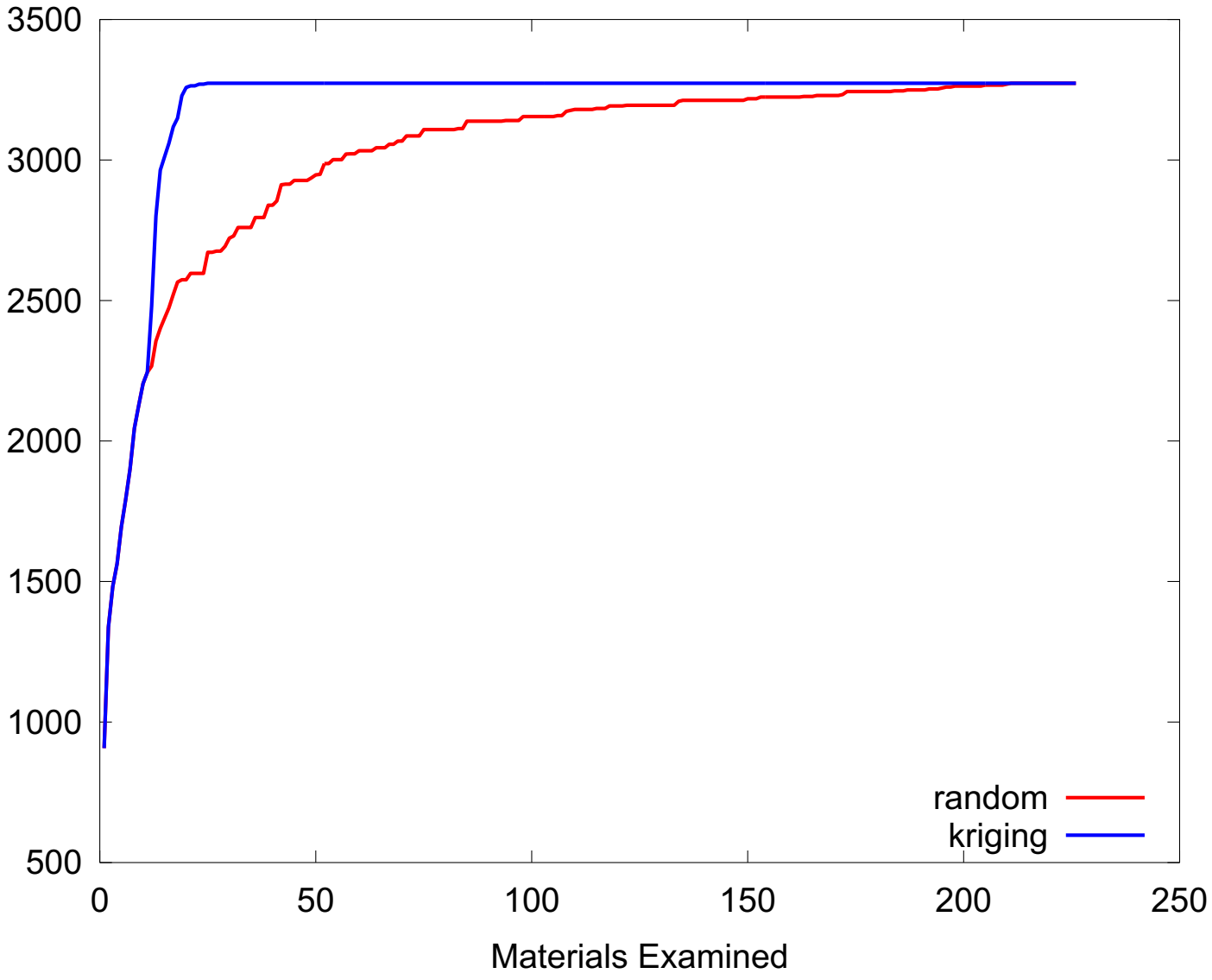
Electronegativity

First-ionization energy

Prediction Accuracy



Open circle: Training examples
Closed circle: Test examples


3500
3000
2500
2000
1500
1000
500
0
50
100
150
200
250
Materials Examined
random
kriging

Number of measurements required to find the solid of highest melting temperature

Designed	Random
16.1	133.4

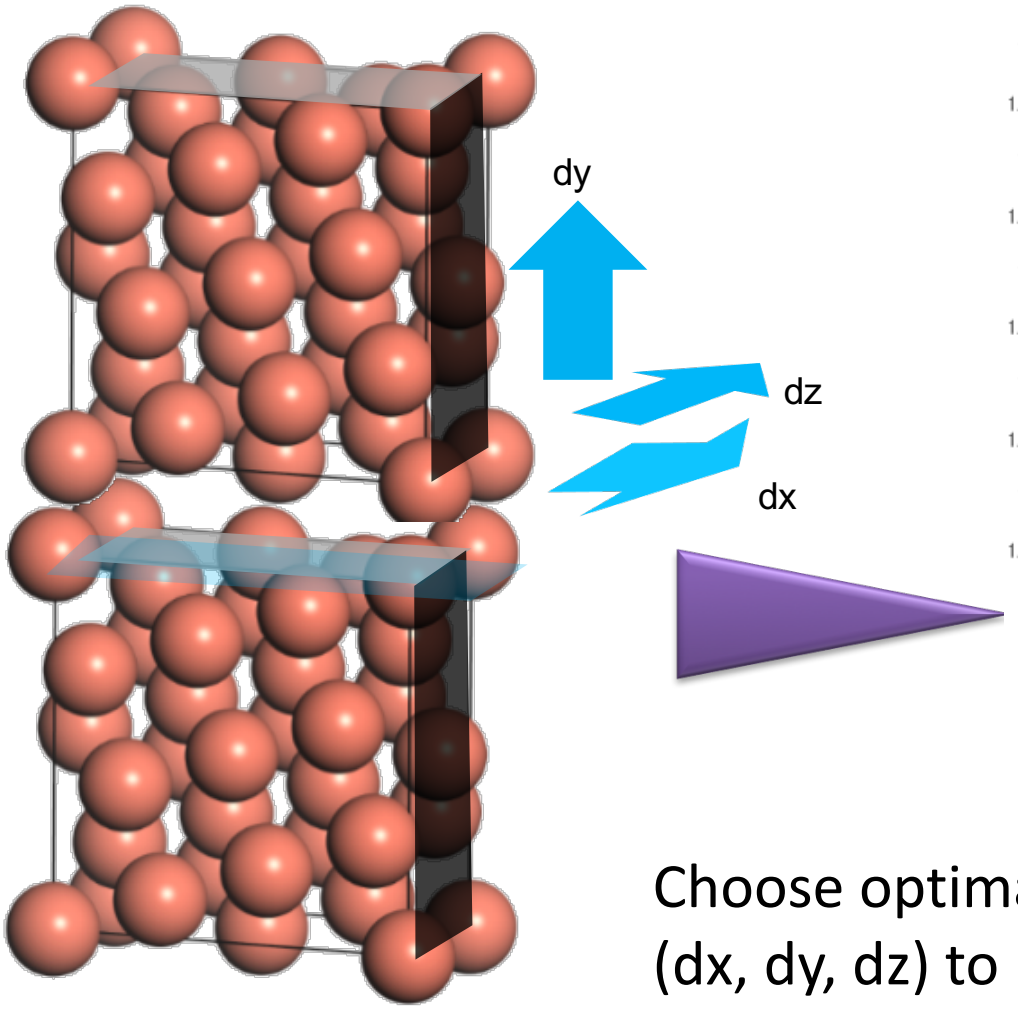
- AlBr3 - As4S4 - GeSe - Se - BaSe - SnO2 - Sb2S3 - Sb2Te3 - Pb - SnF2 - GeBr2 - SnSe2 - BaO - BaS - SrSe - SiC - BeO - **[AlN]** - Be3N2 - Al2O3 - Si3N4 - Al4C3 - MgO - CaO - CaC2 - LiH - Cs - Be - BaH2 - Bi2O4 - K - BeF2 - Tl - RbN3 - LiF - PbTe - CsI - Li - P2O5 - Tl2O3 - BaF2 - Bi - Ba - CaS - SrO - CaSi - PbO - CaF2 - Rb - MgH2 - Si - BaSi2 - IBr - Bi2O3 - SrS - NaF - Ga2O3 - Al - TlI - CsO2 - KCl - In - I2 - BiF3 - SrF2 - LiCl - InN - CsBr - ICl - SrH2 - Pb3O4 - Na - Na2O2 - In2O3 - RbI - S - PbF2 - Bi2Te3 - Sn - CaH2 - KF - InSb - Ca - BiI3 - CsCl - K2O2 - MgF2 - Ge - PbS - SrSi2 - TeO2 - TlSe - Sr - BaI2 - AlP - Li2O - RbO2 - CsF - P4S3 - BiF5 - Mg - GeO2 - NaCl - CaSi2 - BaCl2 - Te - PbSe - TeF4 - PbI2 - TlF - KI - P - MgS - SnTe - NaO2 - GaAs - RbCl - Tl2O - SiS2 - KO2 - InAs - BaBr2 - P2S3 - Sb - KBr - TeI4 - Li3N - TeO3 - RbBr - SiI4 - LiBr - GaSb - TlCl - SeO3 - GaP - RbF - SnI4 - Cs2O - As2O3 - SrCl2 - Mg2Si - TlBr - AlAs - LiI - P4S7 - Bi2S3 - Mg2Sn - CaCl2 - AlI3 - As2O5 - SnSe - Ca3N2 - Li2S - NaBr - InI3 - BeCl2 - Sb2O3 - NaI - Mg2Ge - InI - BiBr3 - GeS - BeI2 - SeBr4 - Tl2S - InP - GaTe - P2S5 - SbF3 - K2S - BiCl3 - SrBr2 - InF3 - GeTe - SbI3 - AlSb - In2Te3 - GeF2 - Mg3Sb2 - SrI2 - PbCl2 - GaS - PI3 - Na2S - SnS - Al2S3 - GaI3 - Rb2S - GaSe - MgCl2 - TeCl4 - Rb2Se - PbBr2 - GeI4 - K2Se - CaI2 - BeBr2 - P2I4 - Sb2Se3 - CaBr2 - As2Te3 - In2Se3 - AlCl3 - InS - GeBr4 - As2S3 - Ga2Se3 - SnBr4 - InCl - As2Se3 - AsBr3 - AsI3 - GaBr3 - Al2Te3 - In2S3 - SbBr3 - MgI2 - InBr3 - GeS2 - MgBr2 - Ga2S3 - GaCl3 - SbCl3 - SnBr2 - GaCl2 - SnCl2 -

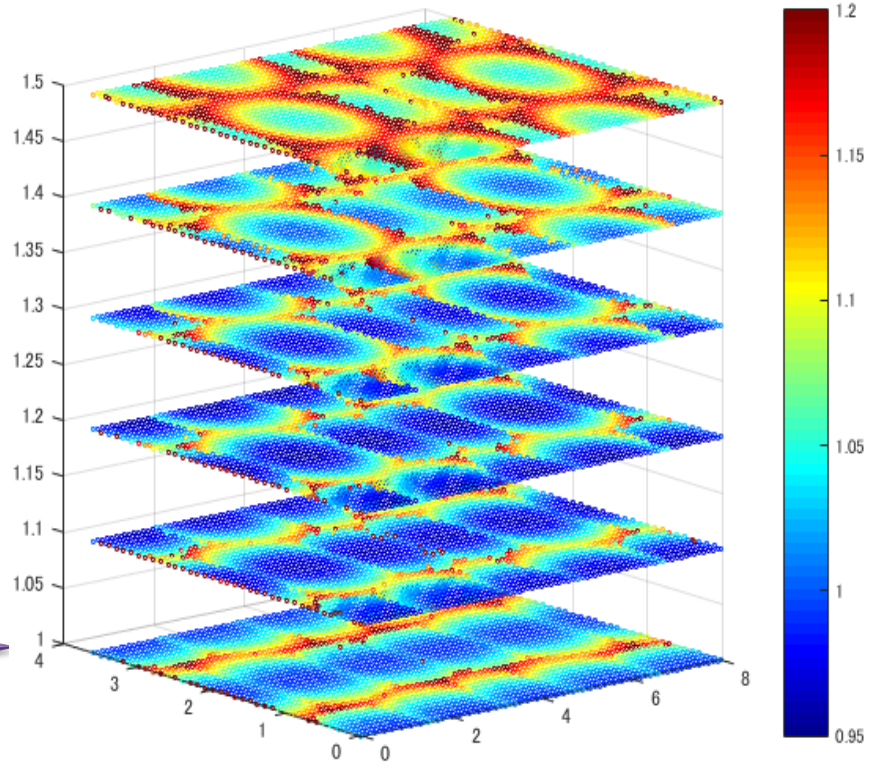
新材料開発にビッグデータ　東大、解析を１００倍高速化

山崎啓介　2016年3月20日08時40分

電池の電極や触媒などの新材料を開発する研究に ビッグデータ を取り入れ、原子レベルの構造を解析する計算を従来の１００倍以上高速化することに、東京大のチームが成功した。この手法は幅広い分野の材料に応用できる可能性があり、開発期間の短縮につながると期待される。

Grain boundary structure determination



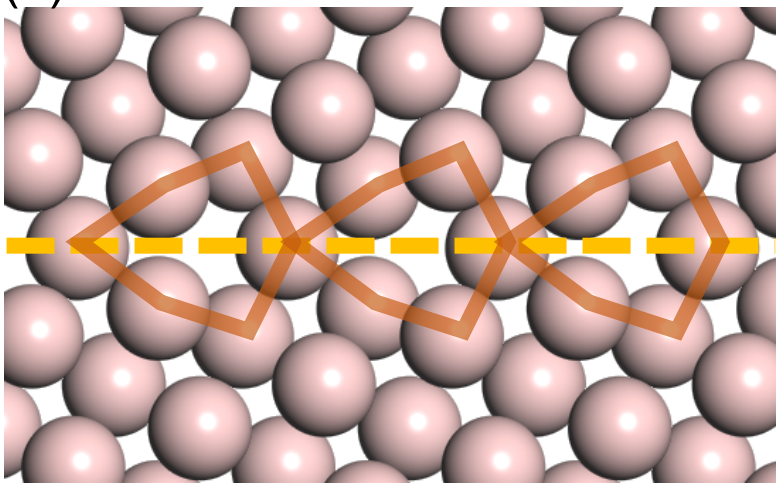


Choose optimal translation parameters (dx, dy, dz) to minimize the grain boundary energy

Acceleration of Discovery

Cu [001] (210) Σ5 grain boundary

(a)

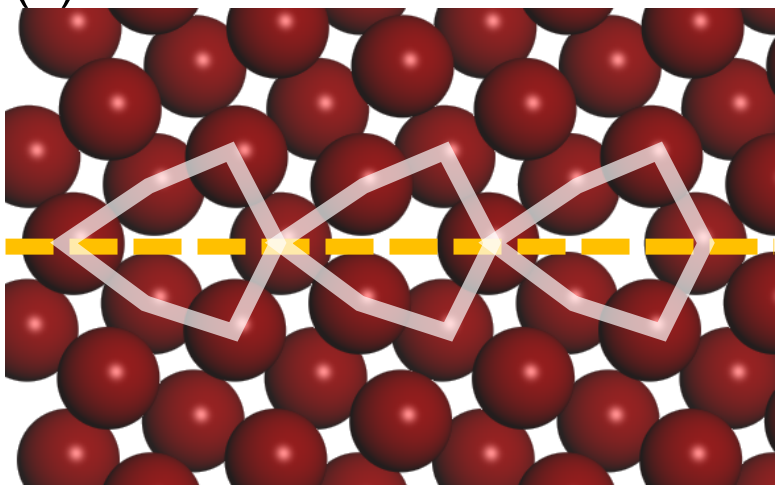


Exhaustive calculations

GB energy=0.96J/m^2

Number of energy calculations
＝16,983

(b)



Bayesian optimization

GB energy=0.96J/m^2

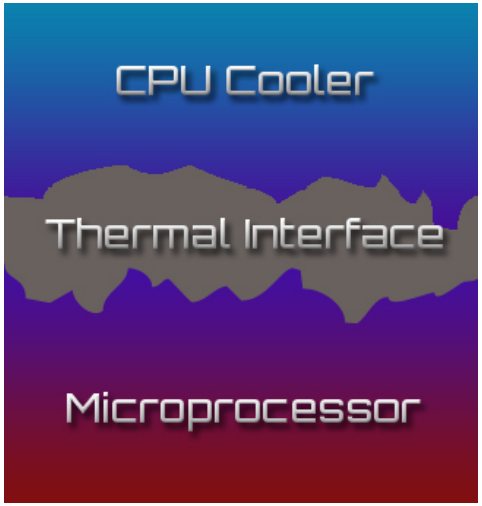
Number of energy calculations
＝69

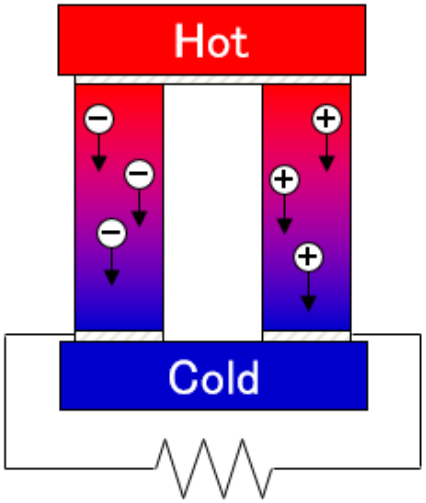
S. Kiyohara et al., Jpn. J. Appl. Phys., 2016.

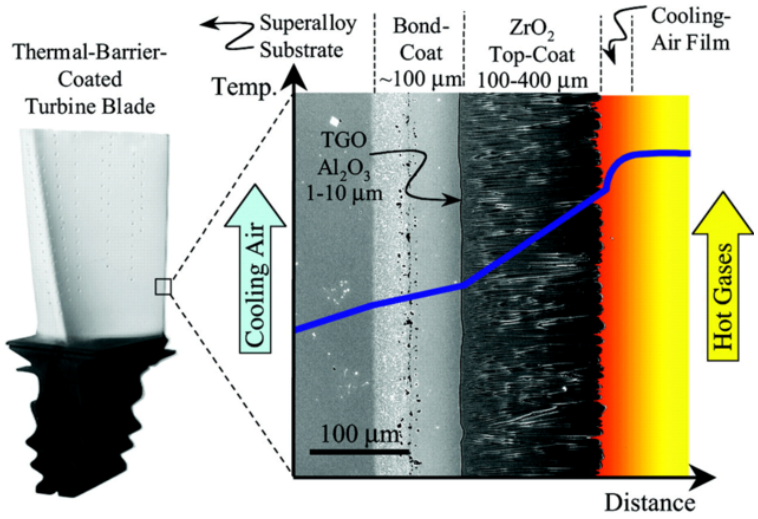
Design nanostructures for phonon transport via material informatics

Interface structure design has wide application in thermal devices.

High Conductance ⟷ **Low Conductance**







Interface materials | **Thermoelectric** | **Thermal barrier coating**

- ❑ various parameters
- ❑ parameters effect coupled
- ❑ transport vs local atomic configurations
- ❑ interference/resonance effects

⇒

- ➢ low developing efficiency
- ➢ high experimental cost
- ➢ long calculation time

Alloy Structure Optimization

Question: How to organize 16 alloy atoms (Si: 8, Ge: 8) to obtain the largest and smallest interfacial thermal conductance?

Descriptors: $C_{16}^{8} = 12{,}870$

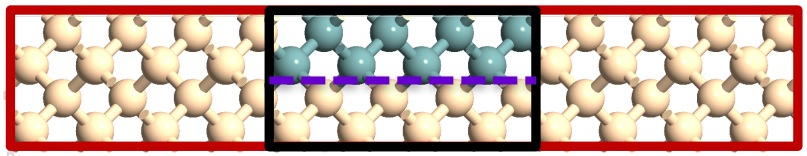
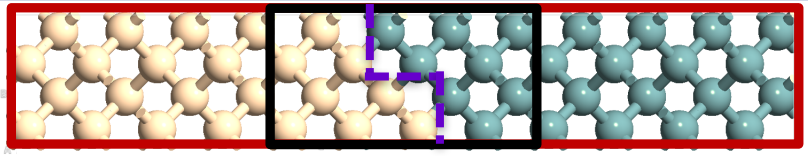
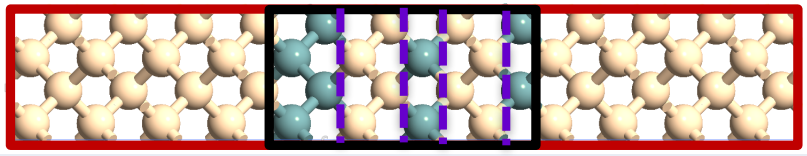
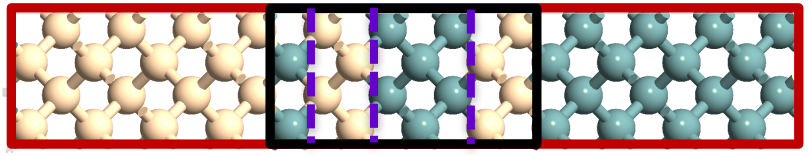
Case	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
2	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0
3	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

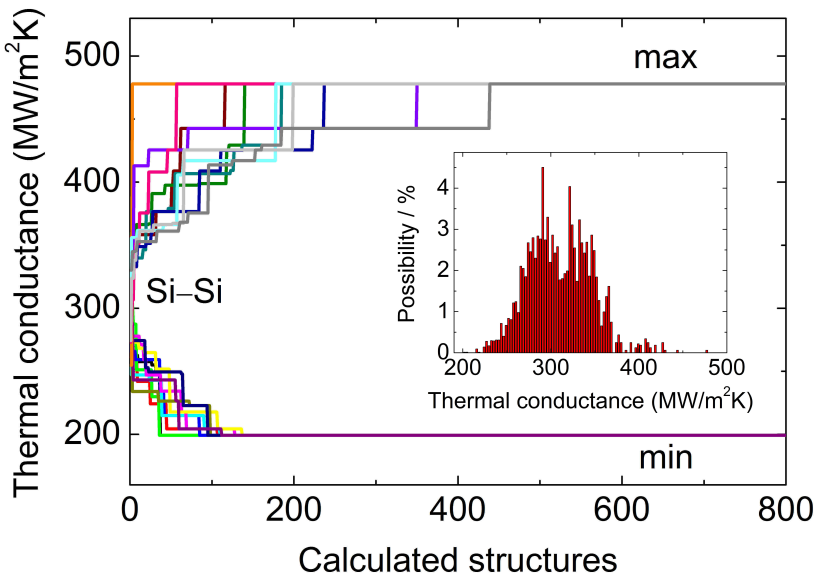
Calculator: Atomistic Green's Function (AGF): Phonon transmission

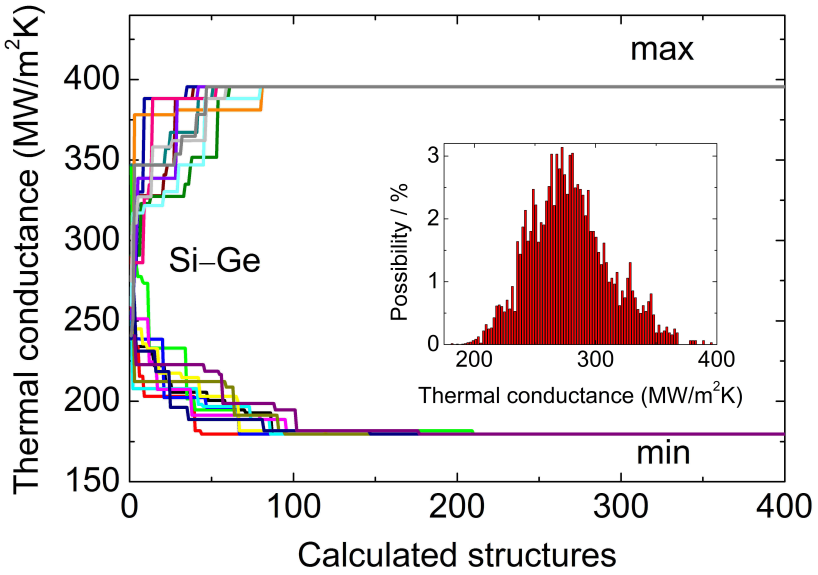
Evaluator: Interfacial Thermal Conductance (ITC)

Optimization method: Bayesian Optimization

Alloy Structure Optimization

ITC	Si-Si	Si-Ge
Max		
Min		

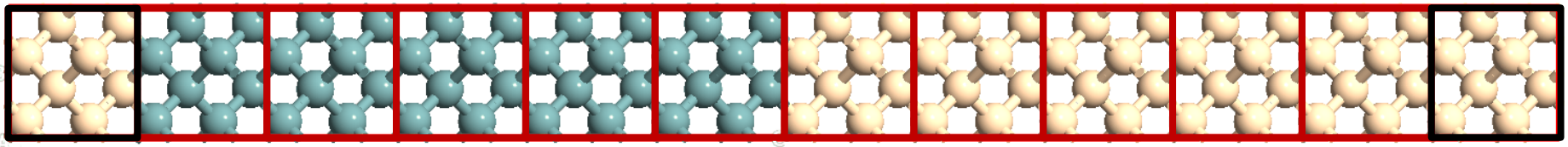




Optimal structures were obtained by calculating only 3.4% of all candidates.

Superlattices Structure Optimization

Topic: Arrange 10-layer superlattices structure (5 layers Si + 5 layers of Ge) between Si and Si to obtain minimal thermal conductance (1 layer thickness = 5.43 A)

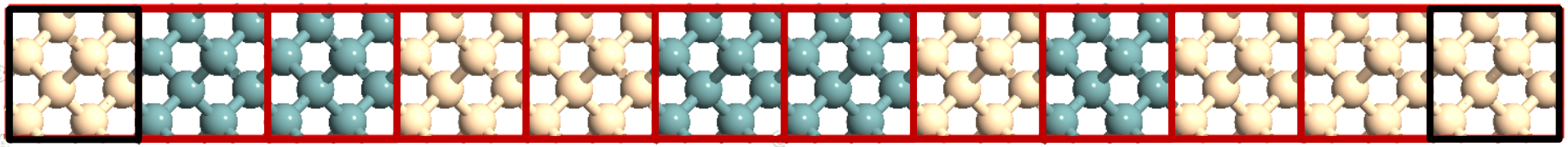


Lead Si Ge Si/Ge superlattice region Lead

Descriptors: 0 1 $C_{10}^{5} = 252$

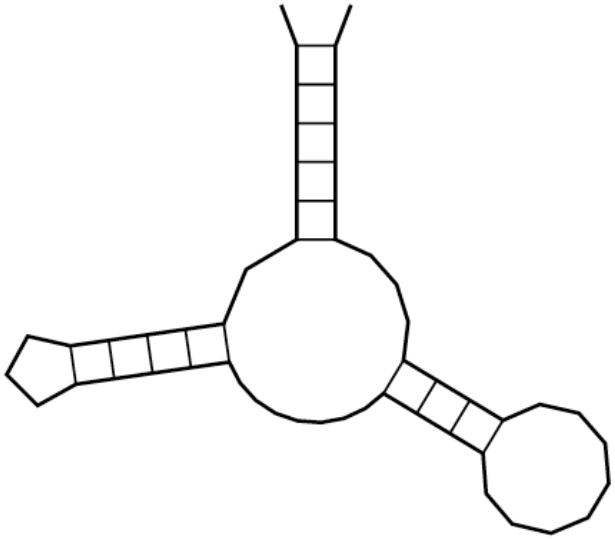
Case	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	0	0	0	0	0
2	1	1	1	1	0	1	0	0	0	0
3	1	1	1	1	0	0	1	0	0	0
…	…	…	…	…	…	…	…	…	…	…

Best Structure: (1101010001)

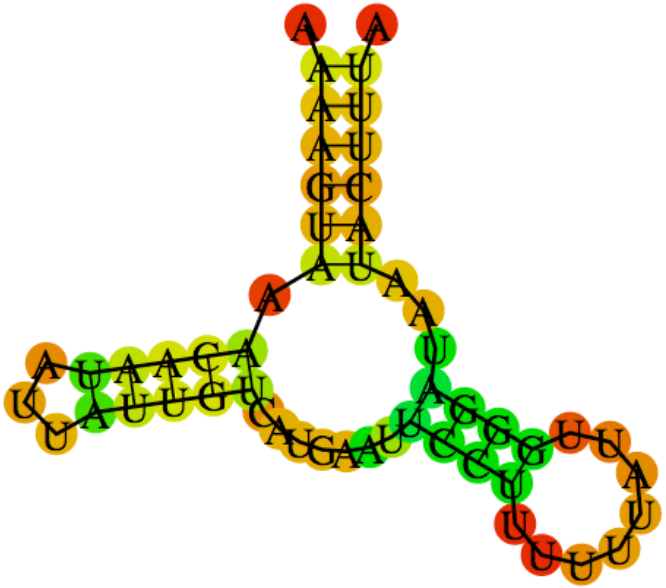


RNA Inverse Folding

- Design RNA whose structure matches target
- 4^n candidates: Too many for Bayesian Opt



Target Structure



AAAAGUAAACAAUAUUAUUGUCAUGAAUUCC
UUUUUUAUUGGGAUAAUACUUUA

Monte Carlo Tree Search

ARTICLE

doi:10.1038/nature16961

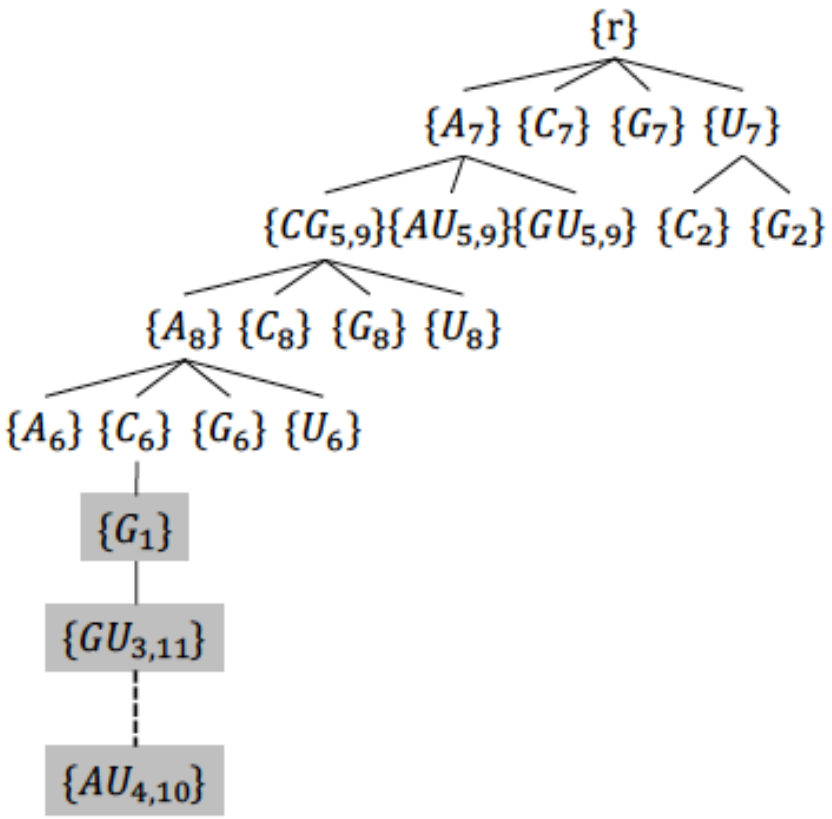
Mastering the game of Go with deep neural networks and tree search

David Silver[1]*, Aja Huang[1]*, Chris J. Maddison[1], Arthur Guez[1], Laurent Sifre[1], George van den Driessche[1], Julian Schrittwieser[1], Ioannis Antonoglou[1], Veda Panneershelvam[1], Marc Lanctot[1], Sander Dieleman[1], Dominik Grewe[1], John Nham[2], Nal Kalchbrenner[1], Ilya Sutskever[2], Timothy Lillicrap[1], Madeleine Leach[1], Koray Kavukcuoglu[1], Thore Graepel[1] & Demis Hassabis[1]

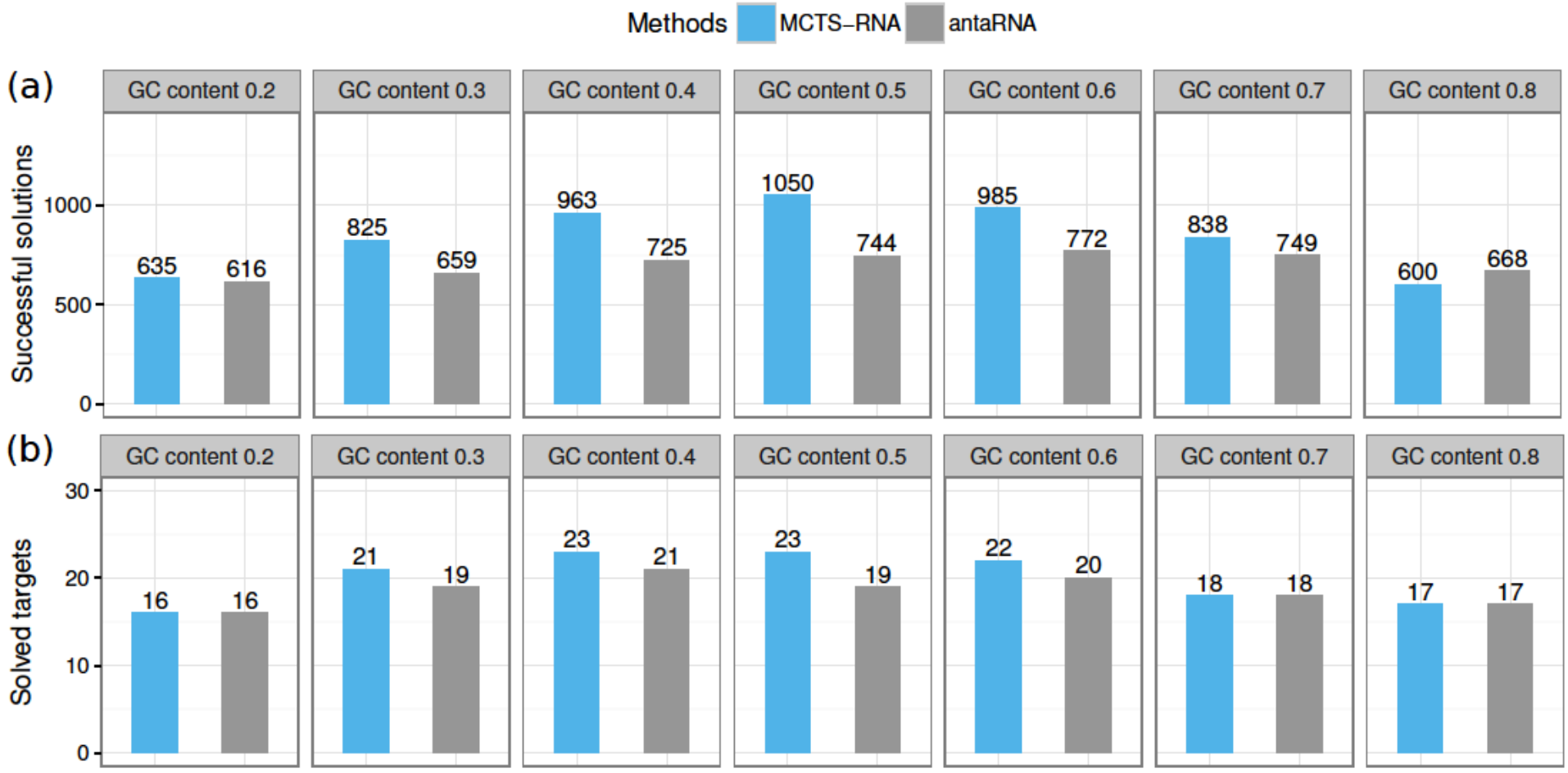
The game of Go has long been viewed as the most challenging of classic games for artificial intelligence owing to its enormous search space and the difficulty of evaluating board positions and moves. Here we introduce a new approach to computer Go that uses 'value networks' to evaluate board positions and 'policy networks' to select moves. These deep neural networks are trained by a novel combination of supervised learning from human expert games, and reinforcement learning from games of self-play. Without any lookahead search, the neural networks play Go at the level of state-of-the-art Monte Carlo tree search programs that simulate thousands of random games of self-play. We also introduce a new search algorithm that combines Monte Carlo simulation with value and policy networks. Using this search algorithm, our program AlphaGo achieved a 99.8% winning rate against other Go programs, and defeated the human European Go champion by 5 games to 0. This is the first time that a computer program has defeated a human professional player in the full-sized game of Go, a feat previously thought to be at least a decade away.

Monte Carlo Tree Search

- Candidates at leafs of search tree
- Reward at leafs: Similarity to target struc
- Score at intermediate node by Playout
  - Random traverses to leafs
  - UCB score: Average reward + Penalty



Comparison with antaRNA



Conclusion

- Artificial intelligence techniques combined with first principles calculation have enormous power