計算科学における情報圧縮

Information Compression in Computational Science

2018.9.27

#1:現代物理学における巨大なデータ

Huge data in modern physics

理学系研究科　物理学専攻　大久保　毅

Department of Physics, **Tsuyoshi Okubo**

Outline

- Background
  - Background of the lectures
  - Computational Science Alliance, The University of Tokyo
  - Computational science and data science
  - Tentative lecture schedule
  - Evaluation
- Huge data in physics
  - Why we need information compression?
  - Examples of information compression

Background of lecturer

大久保　毅（OKUBO Tsuyoshi)

Project Lecturer,
Department of Physics,
Sci. Bldg. #1 940

Research:

Statistical Physics, Condensed matter physics, Magnetism,
(Computational Physics)

- Random packing of disks
- Mean-filed analysis of hierarchical society
- Ordering of (classical) frustrated spin system
    - Z_2-vortex, skyrmion, multiple-Q states, …
- Deconfined quantum criticality
- **Tensor network methods**
- ….

From Yamaji-sensei's slide

Background of Lecturers

山地 洋平 YAMAJI, Youhei (Project Associate Professor, Eng. Bldg. #6, 422)

Research:
Theoretical condensed matter physics
Computational method of many-body quantum systems

Developer of open source codes for supercomputers

Quantum lattice model solver HΦ

http://ma.cms-initiative.jp/ja/index/ja/listapps/hphi

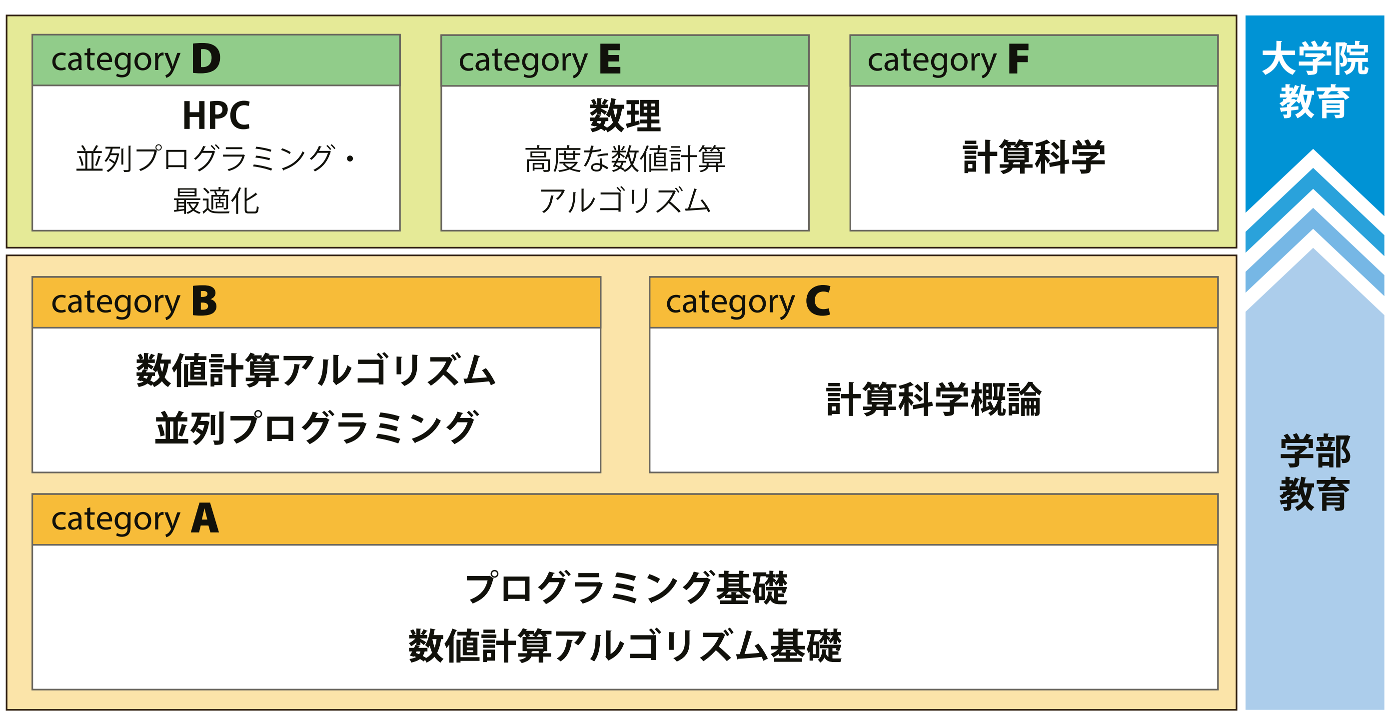
Computational Science Alliance, The University of Tokyo

http://www.compsci-alliance.jp

Train experts for computational and computer sciences.

計算科学アライアンス認定講義

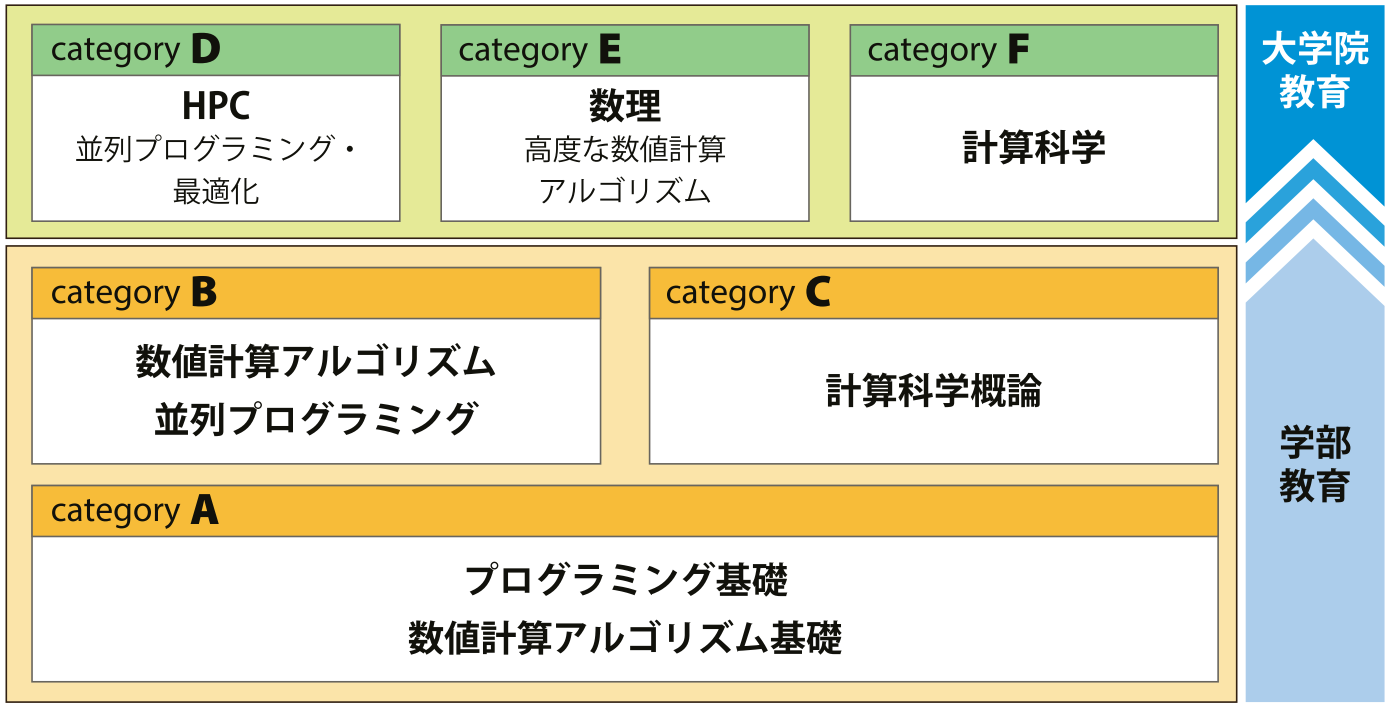
- 平成２９年度から実習にも力点をおいた新しい講義を立ち上げ
- 計算科学・計算機科学に関する80以上の学部・大学院講義とあわせ、「計算科学アライアンス認定講義」として体系化
- 認定講義を内容に応じて6つのカテゴリに分類
- 所定の単位を取得した学生には「修了認定証」を発行
- この講義はカテゴリE



- 学部
  - カテゴリA,B,Cからそれぞれ1.5単位以上
- 大学院
  - カテゴリD,E,Fのうち2つのカテゴリを選択しそれぞれから2単位以上

計算科学アライアンス認定講義：大学院

- カテゴリD - 最先端のスーパーコンピュータを駆使するのに必要とされる技術。種々の並列アルゴリズム、MPI並列やOpenMP並列などの並列プログラミング、メモリアクセス最適化などのチューニング
  - 例：**物質科学のための計算数理I**
- カテゴリE - 最先端の数値計算アルゴリズムとその数理的基礎付け。差分法・有限要素法・有限体積法、特異値分解、最適化問題などの手法とその応用
  - 例：**計算科学における情報圧縮**
- カテゴリF - 各分野におけるシミュレーション手法とその研究成果。電子状態計算、分子動力学、量子多体計算、数値流体力学、構造計算、ゲノム解析など
  - 例：**多体問題の計算科学、物質科学のための計算数理II**



海外派遣

- 目的：国際的にトップクラスの計算科学・計算機科学の人材を育成する。
- 応募資格：東京大学に在籍する学部生・大学院生で、計算科学および計算機科学に関連する研究を行っているもの（登録者を優先）。
- 期間及び内容：1〜2週間程度。 海外の国際会議における研究発表、サマースクール等への参加、大学・研究機関における国際共同研究など。
- 報告書：海外派遣終了後2週間以内にA4紙1枚以上（写真貼付可）を提出。

- 随時募集。
- 募集要項と応募申込書は計算科学アライアンスのWebページに掲載済み。
  http://www.compsci-alliance.jp/overseas-support/

- 昨年度は10件実施。

海外派遣

所属	学年	渡航先	日程
工学系研究科	博士１年	29th European Summer School in Logic, Language, and Information, Toulouse, France	2017.7.15 - 7.31
情報理工学系研究科	博士１年	SC17, Denver, USA	2017.11.11 - 11.18
理学系研究科	博士１年	2nd WCRP summer school on Climate Model Development, Cachoeira Paulista, Brazil	2018.1.20 - 2.3
理学系研究科	修士１年	University of Milan, Milan, Italy	2018.1.29 - 2.12
理学系研究科	修士１年	University of Milan, Milan, Italy	2018.1.29 - 2.12
理学系研究科	修士１年	University of Milan, Milan, Italy	2018.1.29 - 2.12
理学系研究科	博士１年	Marbug, Germany - Zurich, Switzerland - LA, USA	2018.2.18 - 3.11
工学系研究科	修士２年	APS March Meeting, Los Angeles, USA	2018.3.4 - 3.11
工学系研究科	博士２年	APS March Meeting, Los Angeles, USA	2018.3.4 - 3.11
理学系研究科	博士１年	APS March Meeting, Los Angeles, USA	2018.3.4 - 3.11

Lecture schedule

Okubo	第１回：	現代物理学における巨大なデータ
	第２回：	現代物理学と情報圧縮
	第３回：	情報圧縮の数理１　(線形代数の復習)
	第４回：	情報圧縮の数理２　(特異値分解と低ランク近似)
Yamaji	第５回：	情報圧縮の数理３　(スパース・モデリングの基礎)
	第６回：	情報圧縮の数理４　(クリロフ部分空間法の基礎)
	第７回：	物質科学における情報圧縮
	第８回：	データ解析の高速化：スパース・モデリングの物質科学への応用
	第９回：	データ空間の圧縮：クリロフ部分空間法の物質科学への応用
Okubo	第１０回：	高度なデータ圧縮：情報のエンタングルメントと行列積表現
	第１１回：	行列積表現の固有値問題への応用
	第１２回：	テンソルネットワーク表現への発展
	第１３回：	テンソルネットワーク繰り込みによる情報圧縮

Important infomations

Evaluation: Based on 3 reports:

Exercise include algorithms and computer simulations.

(We provably provide python codes.)

Notice!

The grade will be evaluated based on
the sum of scores of three reports.
(So, if you will miss one of them, it will be a big disadvantage.)

Slides:

The lecture slides will be uploaded to

- ITC-LMS (Information Technology Center Learning Management System)
- https://github.com/compsci-alliance/information-compression

Huge data in physics

Computer science and data science

1. Experimental Science
2. Theoretical Science
3. Computational Science
4. Data Science
5. AI?

"Information compression in computational science"
is related to 3rd and 4th sciences.
(It might be slightly related to AI.)

Huge data in physics

Many-body problems in physics

- Celestial movement （天体運動）
- Gases, Liquids
- Molecules, Polymers (eg. Proteins), ...
- Electrons in molecules and solids
- Elemental particles (Quantum Chromo Dynamics)
（量子色力学）

In these problems, "systems" contain huge degrees of freedoms:

$6N$-dimensional phase space for classical mechanics

$O(e^N)$-dimensional Hilbert space for quantum system

Complex particle system

Eg. Poliovirus capsid in electrolyte solution

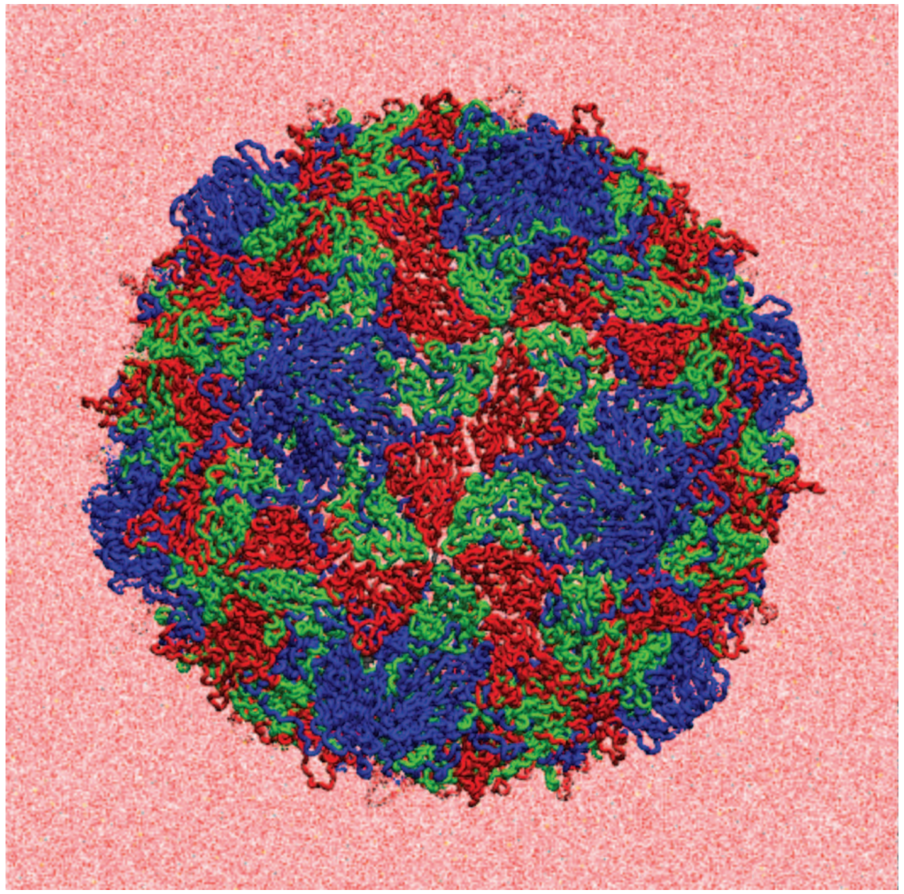
（ポリオウイルス　カプシド）　　　（電解質溶液）

Y. Ando et al, J. Chem. Phys. **141**, 165101(2014).

Long-range coulomb interaction

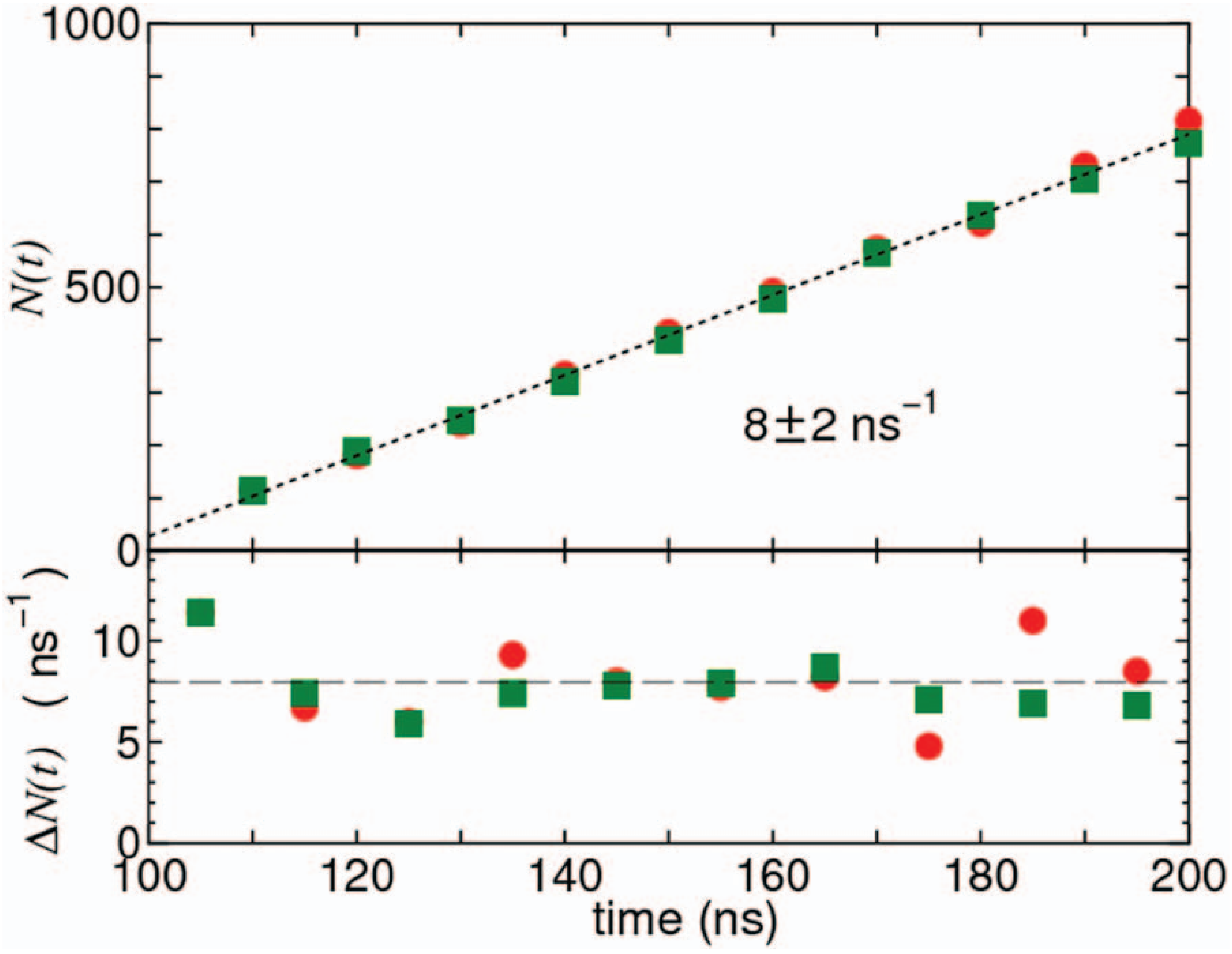
（クーロン相互作用）

Poliovirus capsid



6.5 million atoms

Dynamics of water molecules



Localized electrons as quantum spin systems

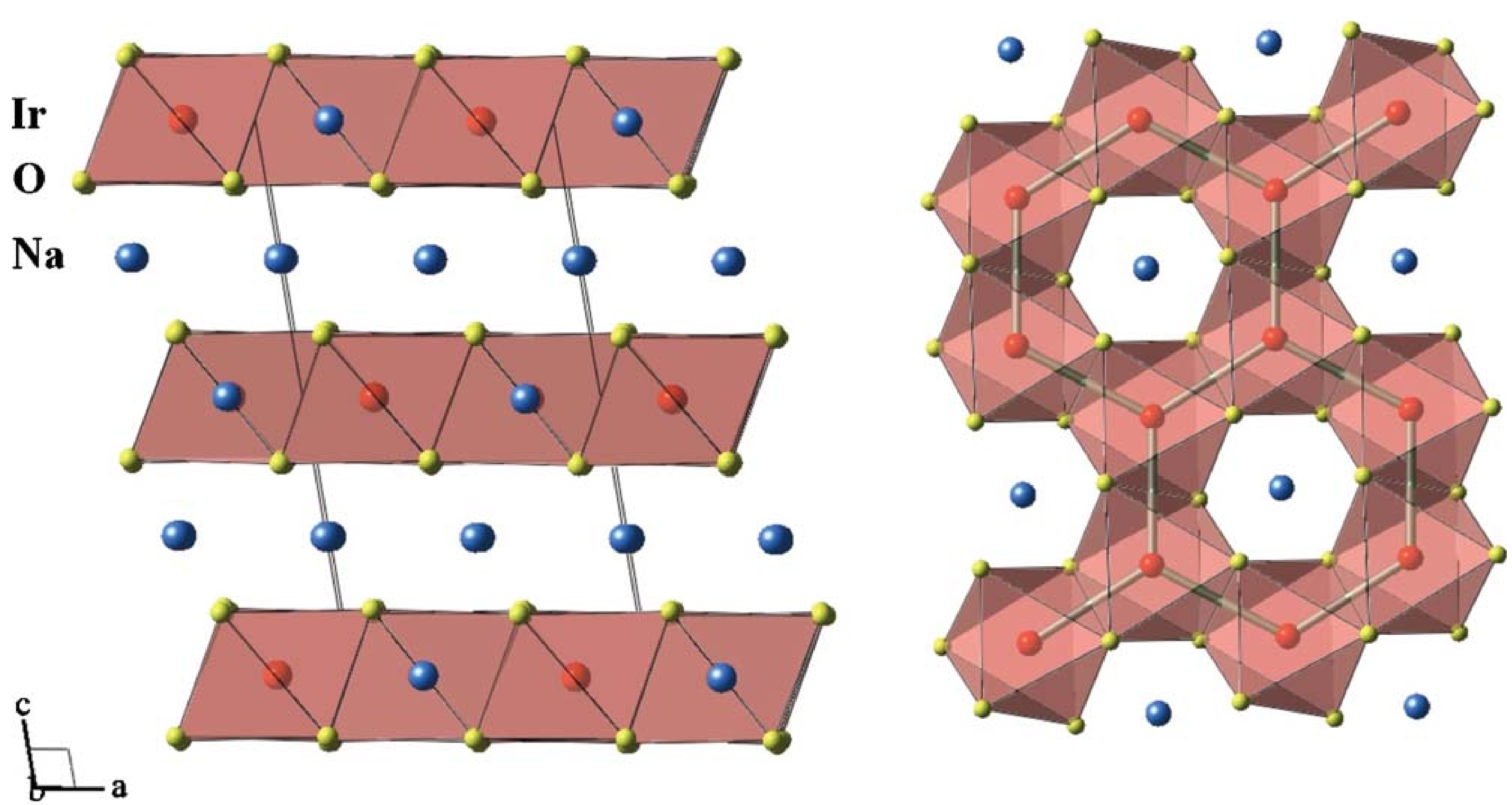
Eg. Antiferromagnetic Mott insulator Na_2IrO_3

（反強磁性）　　　（モット絶縁体）

Y. Singh and P. Gegenwart, Physical Review B **82**, 064412 (2010)

$$\mathcal{H} = \sum_{i,j} J_{ij} S_i S_j$$

S_i : spin operator



(a)　　　(b)

Why we need information compression?

1. We can not understand huge information directly

We try to characterize "systems" thorough a few parameters

Examples:

Thermodynamics:

Systems are characterized by thermodynamic quantities,

Internal energy, Entropy, Pressure, Volume, Particle number,...

Critical phenomena:

Critical systems are characterized by a few critical exponents.
(臨界指数)

Related topics in 13th lecture: "Renormalization（繰り込み）"

Why we need information compression?

1. We can not understand huge information directly

We try to characterize "systems" thorough a few parameters

William of Ockham

(from wikipedia)

Modeling:

We want to determine "essential" parameters
to explain observed phenomena.

Occam's razor:
In order to understand the essence,
we should not assume too much things.

Related topics in 5th and 8th lectures: "Sparse Modeling"

Example of sparse modeling (Compressed sensing)

Reconstruction of MRI image
from smaller samplings

M. Lustig et al, Magnetic Resonance in Medicine **58**, 1182 (2007).

Experiment: Random sampling in k-space

Image reconstruction by assuming
sparsity in wavelet transform domain

$$\begin{aligned} &minimize \quad \|\Psi m\|_1 \\ &s.t. \quad \|\mathcal{F}_u m - y\|_2 < \epsilon \end{aligned}$$

m :image

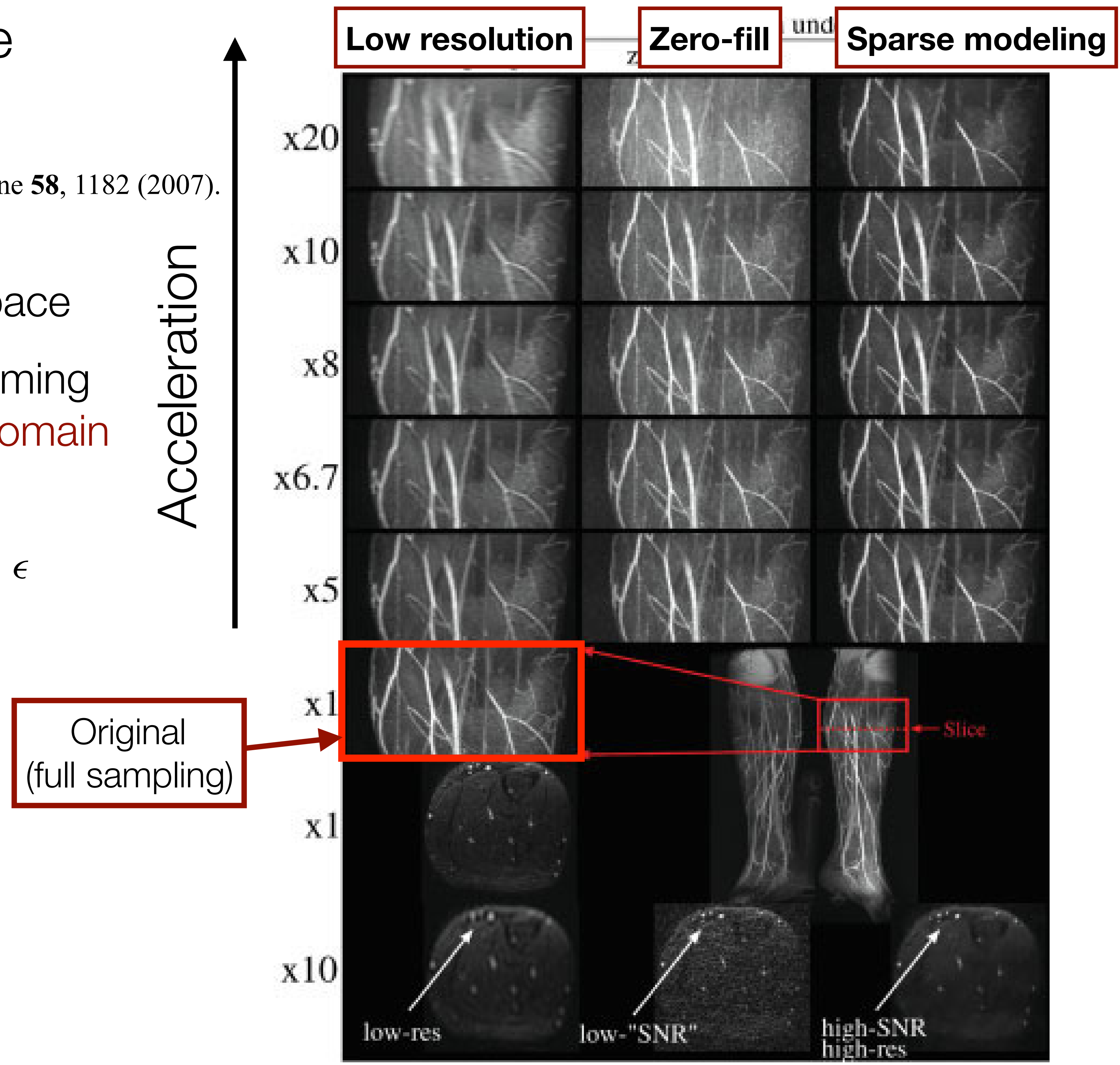
Ψ :wavelet transform

$\mathcal{F}_u$:Fourier transform

y :Experimental data in k-space

L1 norm: $||x||_1 = \sum_i |x_i|$

L2 norm: $||x||_2 = \sqrt{\sum_i |x_i|^2}$



Why we need information compression?

2. We can not treat entire data in the present computers

Available memories in the present computers

		Double precision real number = 8 Bite
Personal computers:	~10 GB	$\sim 10^9$
Super computers:	~100 GB / node	$\sim 10^{10}$
K@RIKEN, **Oakforest-PACS**@UTokyo and Tsukuba Univ, **Sekirei**@ISSP, UTokyo ...	~1 PB (whole system)	$\sim 10^{14}$

Notice: In quantum system, the size of Hilbert space is $O(e^N)$

Why we need information compression?

2. We can not treat entire data in the present computers

→ Try to reduce the "effective" dimension of (Hilbert) space

By taking proper basis set,
we can represent a quantum state efficiently.

- Krylov subspace (6th and 9th lectures)
- Matrix product state
- Tensor network states (10th, 11th 12th lectures)
- ...

Examples of information compression 1

Krylov subspace

linear subspace generated by a square matrix (M) and a vector (v) as

$$\mathcal{K}_n(M,\vec{v}) = \mathrm{span}\left\{\vec{v}, M\vec{v}, M^2\vec{v}, \ldots, M^{n-1}\vec{v}\right\}$$

For quantum many body problems:

$M = \mathcal{H}$:Hamiltonian

$\vec{v} = |\phi\rangle$:wavevector

Solve the eigenvalue problem within
a restricted space (Krylov subspace)

Lanczos method, Arnoldi method

* In these method, we do not necessarily need explicit matrix.
It is enough to know the result of matrix vector multiplication.

Examples of information compression 2

Compression of an image

$\boxed{\text{image} = \text{matrix}}$

Original

$\chi = 768$

of "singular values"

Compressed

$\chi = 10$

$\chi = 100$

Examples of information compression 2

Compression of a color image

$\boxed{\text{image} = \text{tensor}}$

Original

$$\chi = 768$$

About 20% compressed

SVD

HOSVD

Singular value decomposition （特異値分解）

Singular value decomposition (SVD):

$U, V^\dagger$: (half) unitary

Λ : Diagonal

For a K × L matrix M,

$$M = U\Lambda V^\dagger$$

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_N \end{pmatrix}$$

$$M_{i,j} = \sum_m U_{i,m}\lambda_m V^\dagger_{m,j}$$

Singular values: $\lambda_m \geq 0$

Singular vectors:
$$\sum_i U_{i,m}U^\dagger_{m,j} = \delta_i,j$$
$$\sum_i V_{i,m}V^\dagger_{m,j} = \delta_i,j$$

By taking only several larger singular values, we can approximate M as a lower rank matrix.

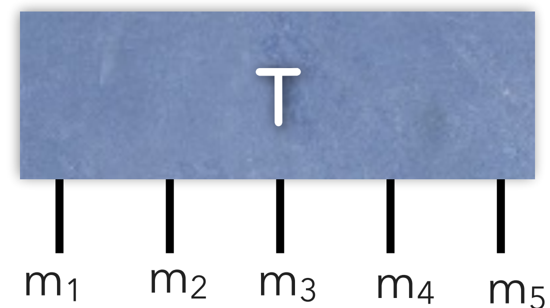
Examples of information compression 3

Wave function: $|\Psi\rangle = \sum_{\{m_i = \uparrow\downarrow\}} T_{m_1,m_2,\cdots,m_N} |m_1, m_2, \cdots, m_N\rangle$

(波動関数)

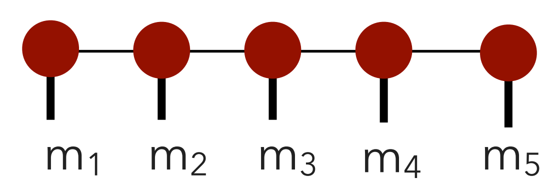
or $\{m_i = 0, 1\}$

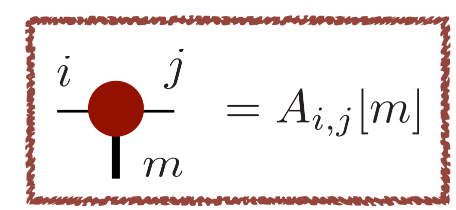
$T_{m_1,m_2,\cdots,m_N}$ N-rank tensor (or Vector)



# of Elements＝2^N

Approximation as a product of "matrices"





Matrix Product States (行列積状態)

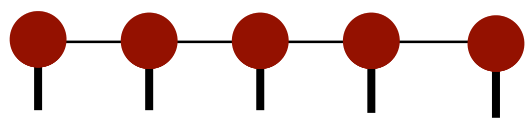
(Tensor train decomposition)

$$T_{m_1,m_2,\cdots,m_N} \simeq A_1[m_1]A_2[m_2]\cdots A_N[m_N]$$

$A[m]$: Matrix for state m

Examples of tensor decompositions

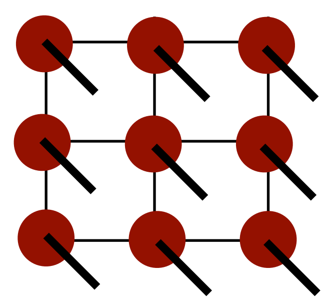
MPS:



Good for 1d gapped systems

(1d correlation in data)

PEPS, TPS:

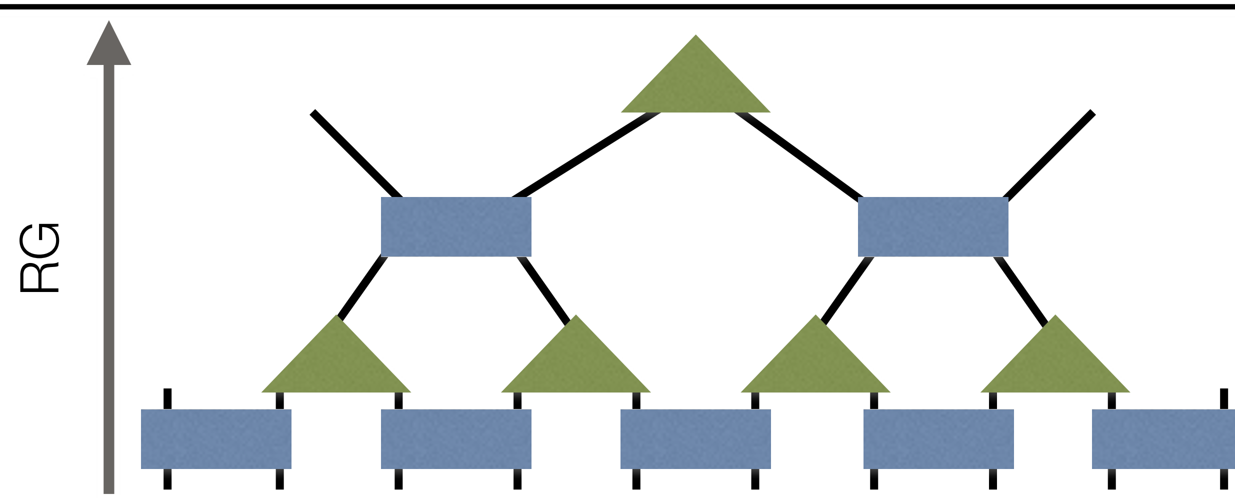


For higher dimensional systems

Extension of MPS

(higher correlation in data)

MERA:

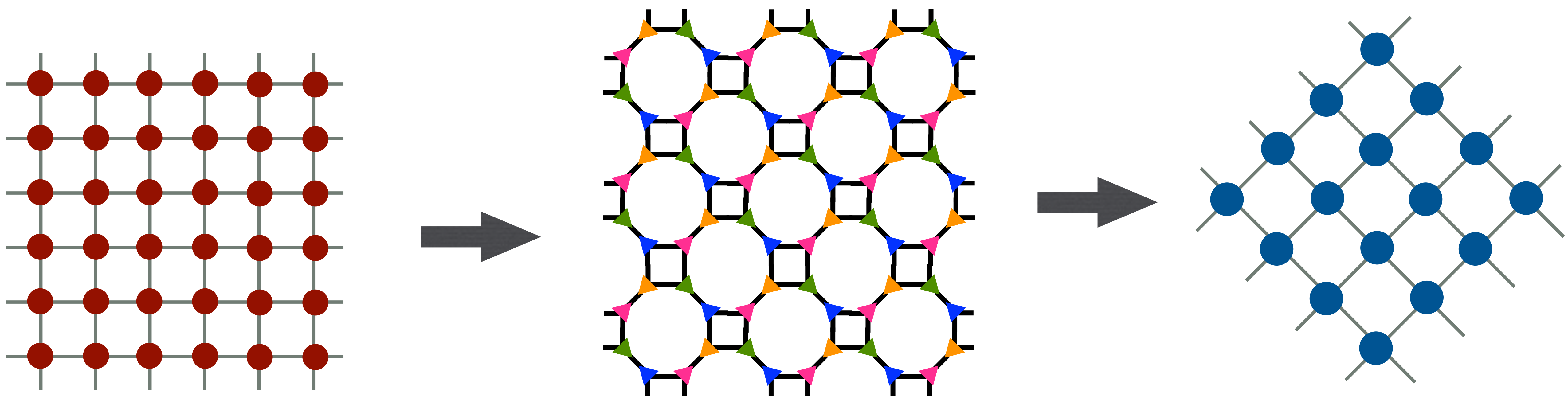


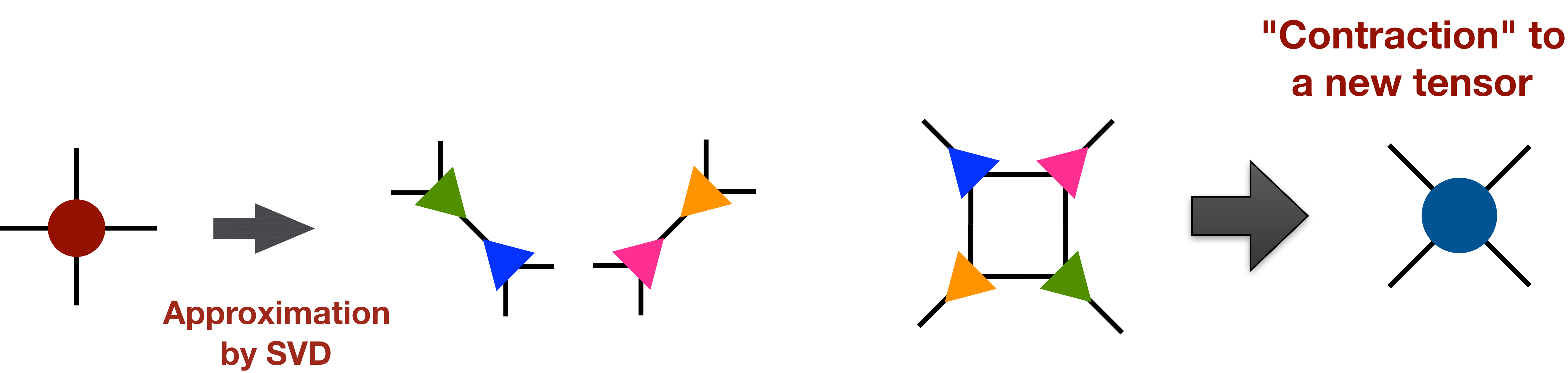
Scale invariant systems

(スケール不変)

Real space renormalization（実空間繰り込み）

Corse graining of a tensor network





Next week

第１回：　現代物理学における巨大なデータ

第２回：　現代物理学と情報圧縮

(Information compression in modern physics)

第３回：　情報圧縮の数理１　(線形代数の復習)

第４回：　情報圧縮の数理２　(特異値分解と低ランク近似)

第５回：　情報圧縮の数理３　(スパース・モデリングの基礎)

第６回：　情報圧縮の数理４　(クリロフ部分空間法の基礎)

第７回：　物質科学における情報圧縮

第８回：　データ解析の高速化：スパース・モデリングの物質科学への応用

第９回：　データ空間の圧縮：クリロフ部分空間法の物質科学への応用

第１０回：　高度なデータ圧縮：情報のエンタングルメントと行列積表現

第１１回：　行列積表現の固有値問題への応用

第１２回：　テンソルネットワーク表現への発展

第１３回：　テンソルネットワーク繰り込みによる情報圧縮