

マテリアルズ・インフォマティクスとは何か -物質材料科学とデータ駆動科学-

東京大学・大学院新領域創成科学研究科

岡田真人

自己紹介

- 大阪市立大学理学部**物理**学科 (1981 - 1985)
 - アモルファスシリコンの成長と構造解析
- 大阪大学大学院理学研究科(金森研) (1985 – 1987)
 - 希土類元素の光励起スペクトルの理論
- **三菱電機** (1987 - 1989)
 - 化合物半導体(半導体レーザー)のエピタキシャル結晶成長
- 大阪大学大学院基礎工学研究科生物工学 (1989 - 1996)
 - **ニューラルネットワーク(人工知能)**
- JST ERATO 川人学習動態脳プロジェクト (1996 - 2001)
 - 計算論的神経科学
- 理化学研究所 脳科学総合研究センター 甘利チーム (2001 - 2004/06)
 - 情報統計力学
 - ベイズ推論, **機械学習, データ駆動型科学**
- 東京大学・大学院新領域創成科学研究科 複雑理工学専攻 (2004/07 –)
 - 物質科学再開(強相関, 表面, 地球惑星科学)
- JST ERATO 岡ノ谷情動情報PJ GL (2008/10 – 2014/3)
- NIMS 情報統合型 物質・材料開発イニシアティブ
物理モデリンググループ GL (2015/8 -

内容

- マテリアルズインフォマティクス(MI)の背景
- MIとバイオインフォマティクス(BI)の相違点
- 構造材料を例とするMI具体例
- 文科省新学術領域 疎性モデリングの紹介
- データ駆動科学の三つのレベルとMI²Iの構成
- 物質材料開発へのデータ駆動的アプローチ

アメリカビッグデータプロジェクト始動

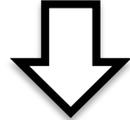
朝日新聞 2012年 5月26日

2012年3月29日 オバマ発表. 予算2億ドル
「ビッグデータ研究開発イニシアティブ」始動

- ・膨大な量のデータ管理や分析を必要とする最先端中核技術の発展を促すこと
- ・その技術を科学や工学分野における発見、国家安全保障の強化、教育に役立てること
- ・ビッグデータ技術分野の人材育成を達成すること

データ科学: Jim Gray (1944-2012)

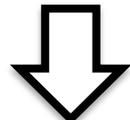
第1の時代: 経験科学
(数千年前～ アリストテレス)



第2の時代: 理論科学
(数百年前～ ライプニッツ)



第3の時代: 計算科学
(数十年前～ フォン・ノイマン)



第4の時代: データ科学

内容

- マテリアルズインフォマティクス(MI)の背景
- MIとバイオインフォマティクス(BI)の相違点
- 構造材料を例とするMI具体例
- 文科省新学術領域 疎性モデリングの紹介
- データ駆動科学の三つのレベルとMI²Iの構成
- 物質材料開発へのデータ駆動的アプローチ

BIとMIの相違点

- バイオインフォマティクス(BI)の典型例は遺伝子情報処理
- BI: DNAチップからの情報抽出
- DNAチップを21世紀に手にする医学者, 生物学者
- 病気の情報は入っているはずだが, DNAチップのパターンと, 病気の関係を因果的に追うことは不可能.
- 機械学習 / 高次元データ解析の導入

BIとMIの相違点

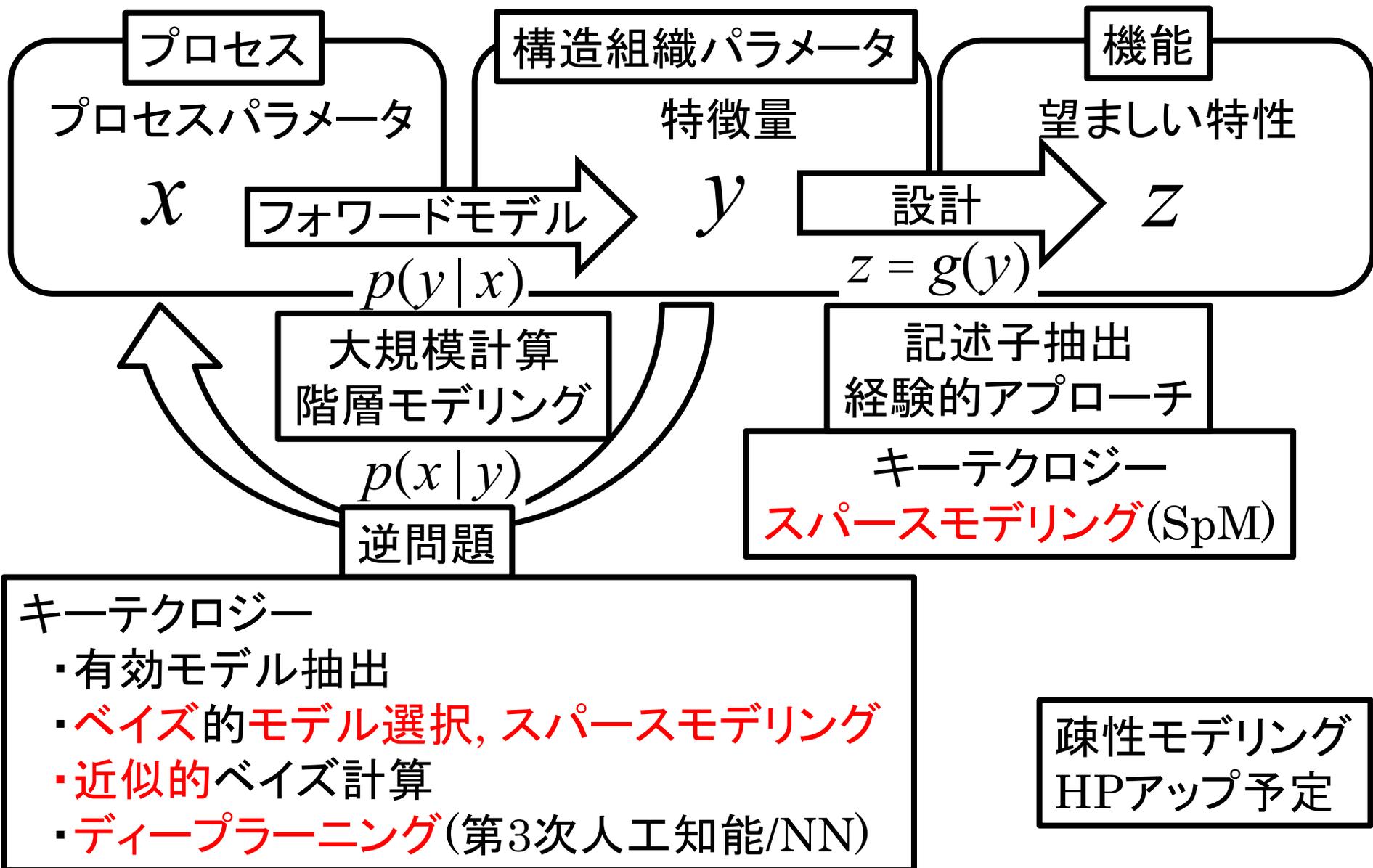
- 物質材料科学
 - 18世紀の産業革命：鉄は国家なり
- 紙と鉛筆と計算尺でデータ解析
- 第一原理にによるフォワードモデル
 - 第3の時代： 計算科学

- 先見的知識(物質材料科学の知見)が豊富
- データ解析の結果と先験的知識の整合性が常に問われる.

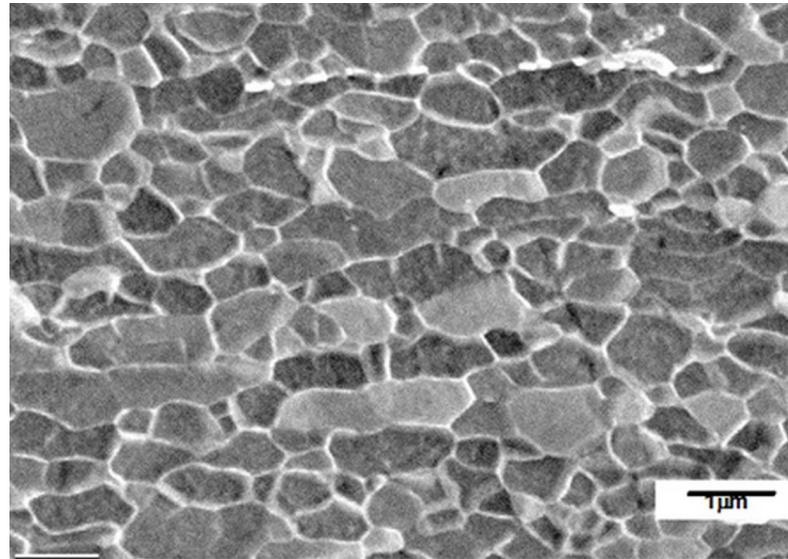
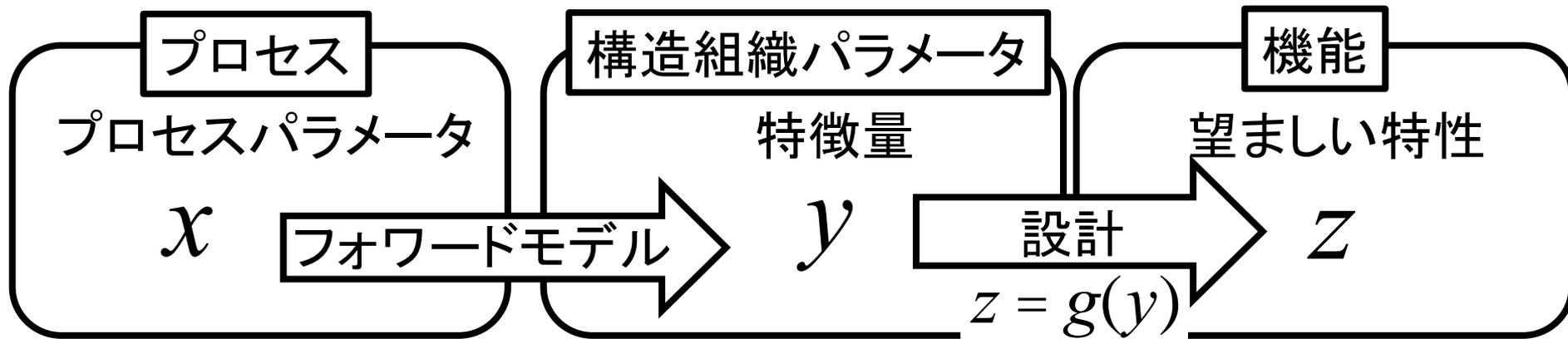
内容

- マテリアルズインフォマティクス(MI)の背景
- MIとバイオインフォマティクス(BI)の相違点
- 構造材料を例とするMI具体例
- 文科省新学術領域 疎性モデリングの紹介
- データ駆動科学の三つのレベルとMI²Iの構成
- 物質材料開発へのデータ駆動的アプローチ

データ駆動型物質材料科学の三つのステップ



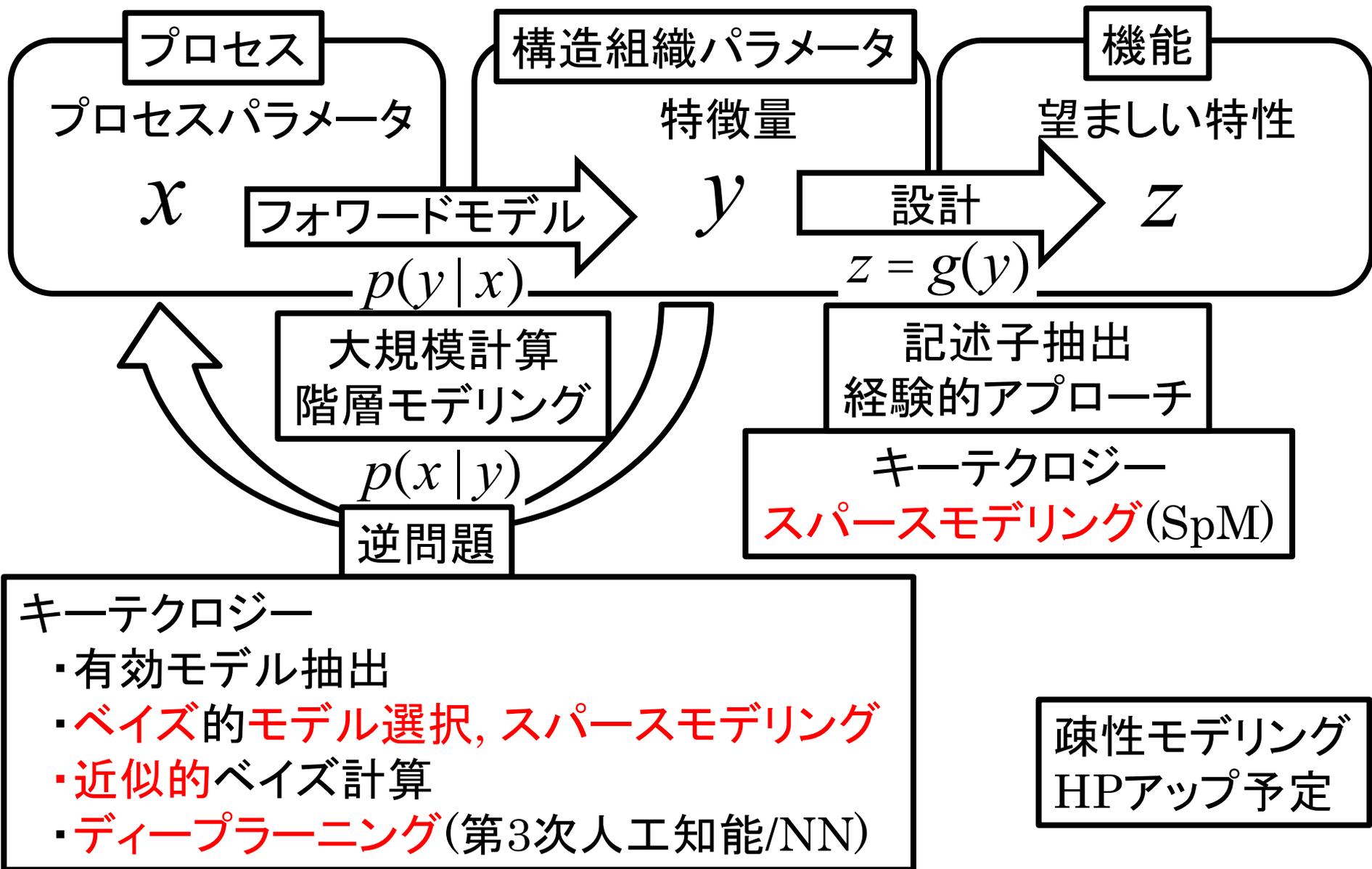
データ駆動型物質材料科学の三つのステップ



プレス資料中の図3: 超鉄鋼ワイヤーのマイクロ組織 (平均粒径0.5ミクロン)

NIMSのHPより掲載

データ駆動型物質材料科学の三つのステップ



内容

- マテリアルズインフォマティクス(MI)の背景
- MIとバイオインフォマティクス(BI)の相違点
- 構造材料を例とするMI具体例
- 文科省新学術領域 疎性モデリングの紹介
- データ駆動科学の三つのレベルとMI²Iの構成
- 物質材料開発へのデータ駆動的アプローチ

H25～29年度 新学術領域 「**スパースモデリング**の深化と高次元**データ駆動科学**の創成」

領域代表

東京大学・大学院新領域創成科学研究科 複雑理工学専攻

岡田真人

スパースモデリングに関するマスコミ報道

- 【テレビ】NHK サイエンスZERO 2015年8月23日放送,
 - 情報科学の名探偵！魔法の数式 **スパースモデリング**
- 【テレビ】NHK徳島ニュース2015年9月14日放送,
 - **ビッグデータ**解析で津波予測
- 【新聞】朝日新聞, 2015年1月19日
 - **人工知能**でカンニングを発見 京大などがプログラム開発
- 【新聞】日本経済新聞, 2015年5月3日
 - 宇宙や津波, 数学で迫る—少ないデータで**「本質」**解析—
- 【新聞】日刊工業新聞, 2015年8月31日
 - **ビッグデータ**絞り込み高速・高精度に—東大、**圧縮センシング**の解析因子の評価技術開発
- 【Web】「日経産業新聞, 財経新聞」, 2014年11月28日
 - 元素含有量で津波堆積物を識別

研究領域の目的及び概要

研究体制のコア形成

目的: 高次元データ駆動科学の創成

大量の高次元データから仮説(モデル)を系統的に導く方法論を「生物」、「地学」分野に確立し、それを実践するための研究体制のコアを我が国に形成する。

3つの戦略

1. スパースモデリング(SpM)に重点投資

今後5年で飛躍的発展が確実視される枠組み

2. 分野の壁を取り去り、知識伝播を飛躍的に加速

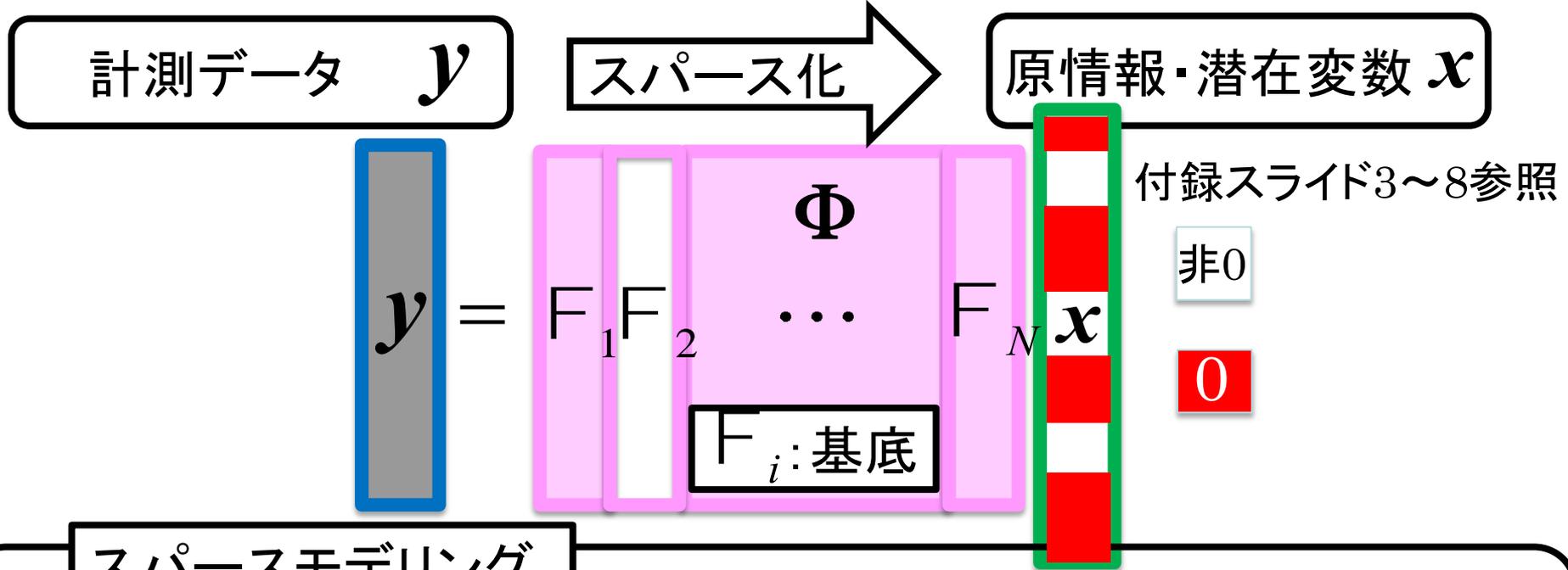
分野をまたぐモデルの構造的類似性を明確化

3. 実験家と理論家との有機的協働

仮説の提案／検証ループを効率的に稼働させる体制

スパースモデリング(SpM)

スパース原理による極限計測・潜在構造抽出



スパースモデリング

潜在変数が**スパース**(0が多い)状況で, 方程式を解く

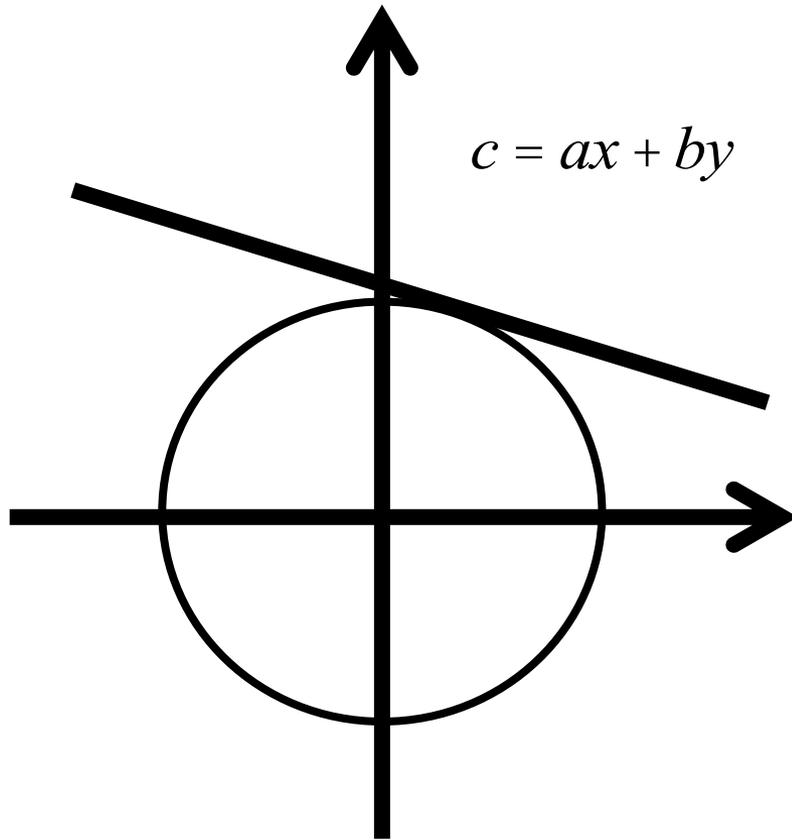
$$E(x) = \underbrace{\left\| y - \sum_i F_i x_i \right\|_2}_{\text{データの再構成}} + \underbrace{\lambda \sum_i |x_i|}_{\text{スパースな変数}} \quad O(2^N) \rightarrow O(N^3)$$

ある条件下で, L1とL0が一致する数理的証明[Candes-Tao, 2005]

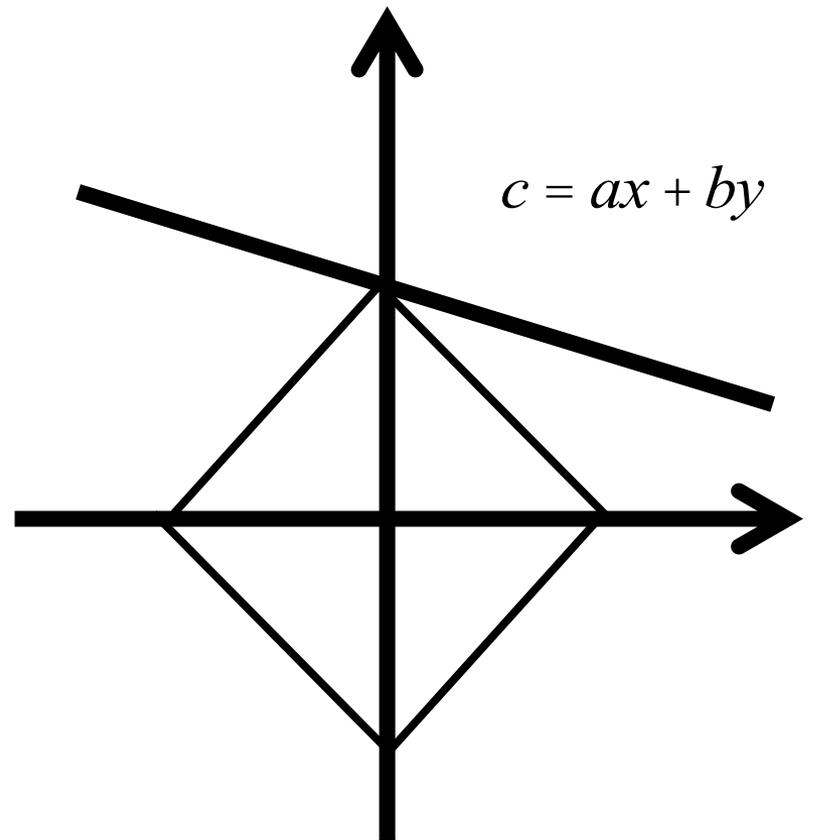
1次関数とスパースモデリング

スパース原理(先入観の積極的活用)

$$E(x, y) = (ax + by - c)^2 + \lambda(x^2 + y^2) \quad E(x, y) = (ax + by - c)^2 + \lambda(|x| + |y|)$$



解はスパースではない



解はスパース

幅広い生物・地学分野の喫緊のテーマ 各分野のフラッグシップを選定

A01-1: 医学班(富樫・京大)

新たな診断・治療の実現

A01-2: 生命科学班(木川・理研)

タンパク科学の質的変化

A01-3: 脳科学班(谷藤・理研)

モノを見分ける脳のしくみ

A02-1: 地球科学班(駒井・東北大)

津波防災対策への提言

A02-2: 惑星科学班(宮本・東大)

次世代探査戦略の創出

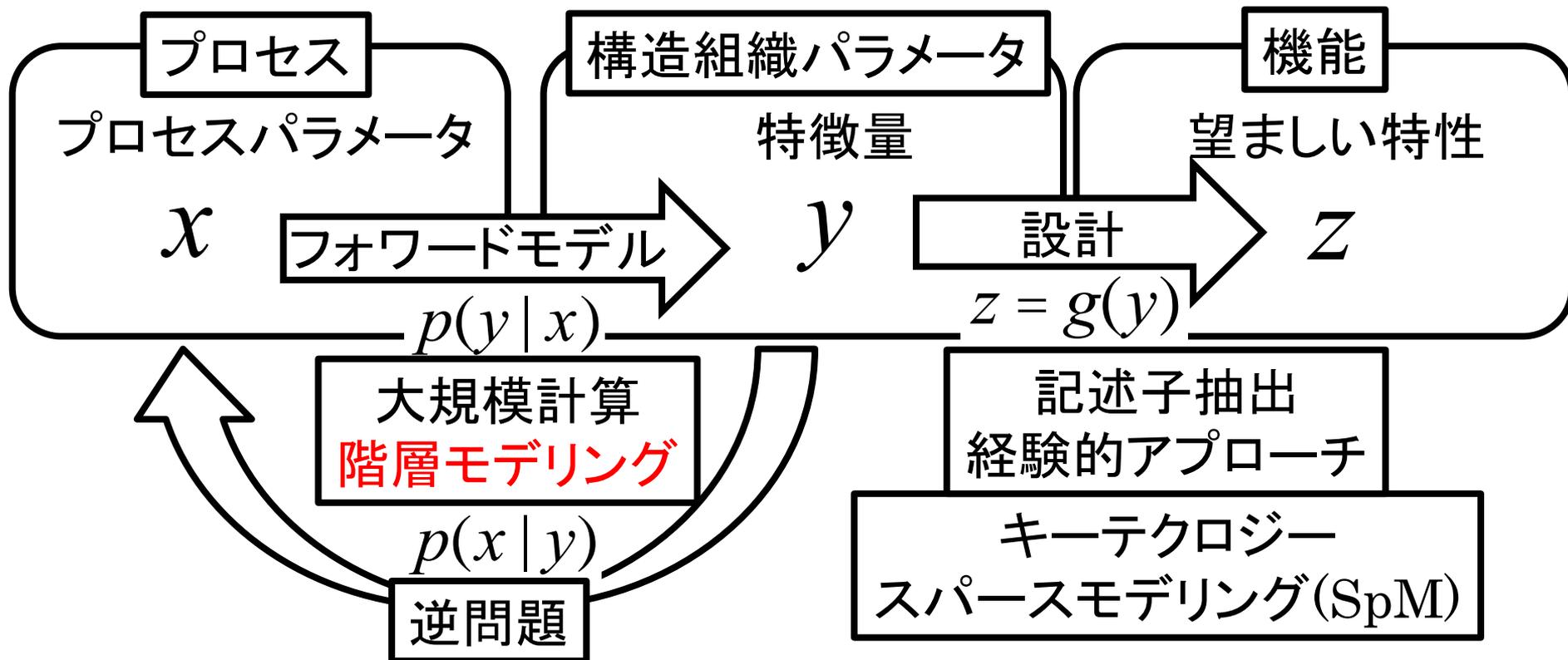
A02-3: 天文学班(本間・国立天文台)

ブラックホールの直接撮像

スパースモデリングの有用性が确实視できる題材を選定
⇒ これらを起爆剤に公募研究・周辺分野に成果を波及

スパースモデリングの 物質材料科学への応用

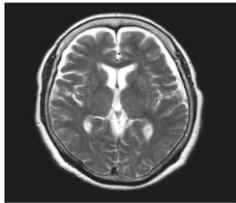
- 合金のクラスターモデルへの適用
- 第一原理計算からの非調和フォノンの有効モデル抽出



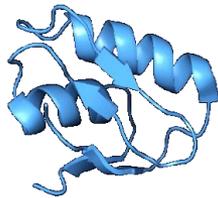
スパースモデリングの深化と 高次元データ駆動科学の創成

実験・計測グループA01,A02

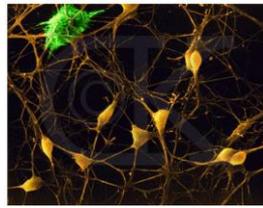
医学



生命科学



脳科学



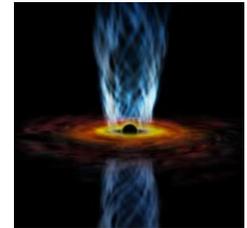
地球科学



惑星科学



天文学



モデリンググループB01

計測モデリング ↔ スパースモデリング ↔ 物理モデリング

情報科学グループC01

非線形

セミパラベイズ

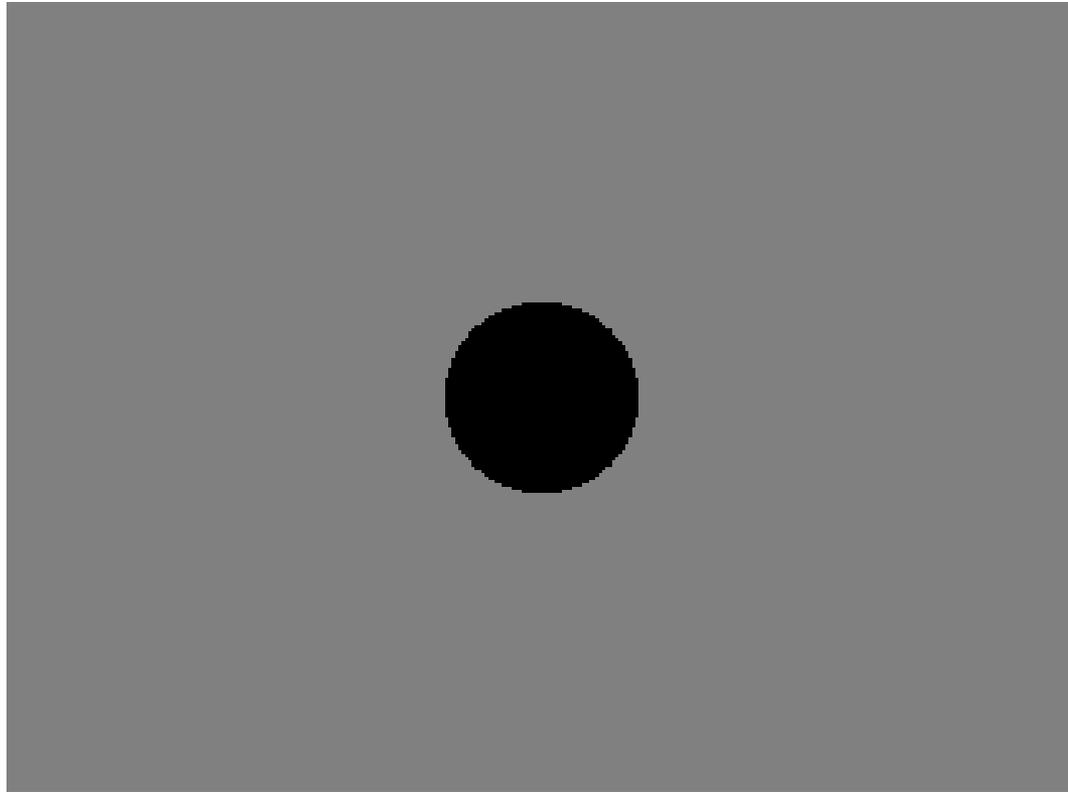
大自由度系

可視化

内容

- マテリアルズインフォマティクス(MI)の背景
- MIとバイオインフォマティクス(BI)の相違点
- 構造材料を例とするMI具体例
- 文科省新学術領域 疎性モデリングの紹介
- データ駆動科学の三つのレベルとMI²Iの構成
- 物質材料開発へのデータ駆動的アプローチ

データ解析：知りたいことが不足しているデータからどう知るか



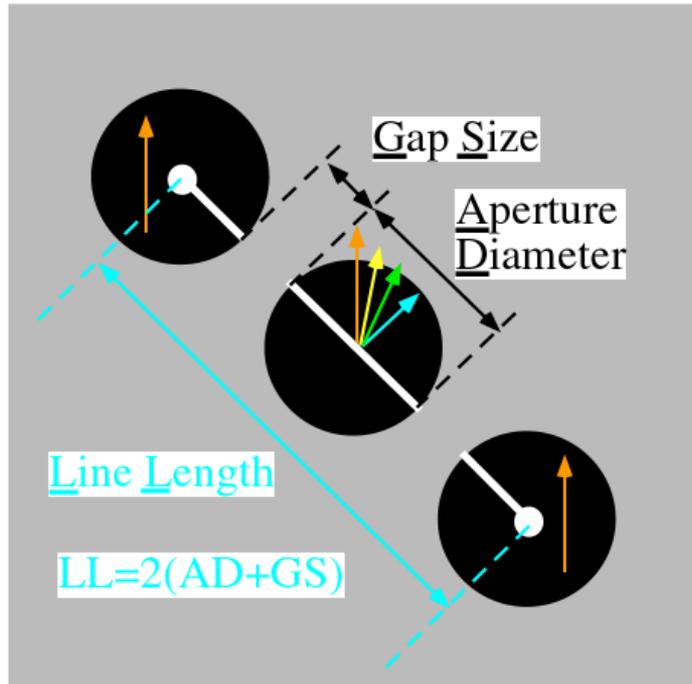
(Ben-Av and Shifrar, 1992)

(Okada, Nishina and Kawato, 2003)

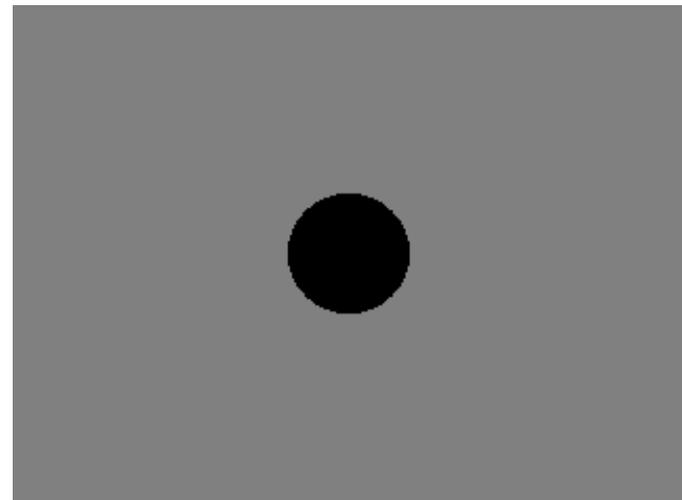
視覚計算の不良設定性(小窓問題)



視覚計算の不良設定性(小窓問題)



三つの小窓から見える三本の線分は、まるで一本の千分の一の一部であるように、同一直線上に乗り移動する。



(Ben-Av and Shifrar, 1992)

(Okada, Nishina and Kawato, 2003)

David Marrの3つのレベル

David Marrは複雑な情報処理装置を理解するには以下の3つのレベルが必要であると説いた

計算理論

情報処理(データ解析)の目標, 方略, 適切さ

表現とアルゴリズム

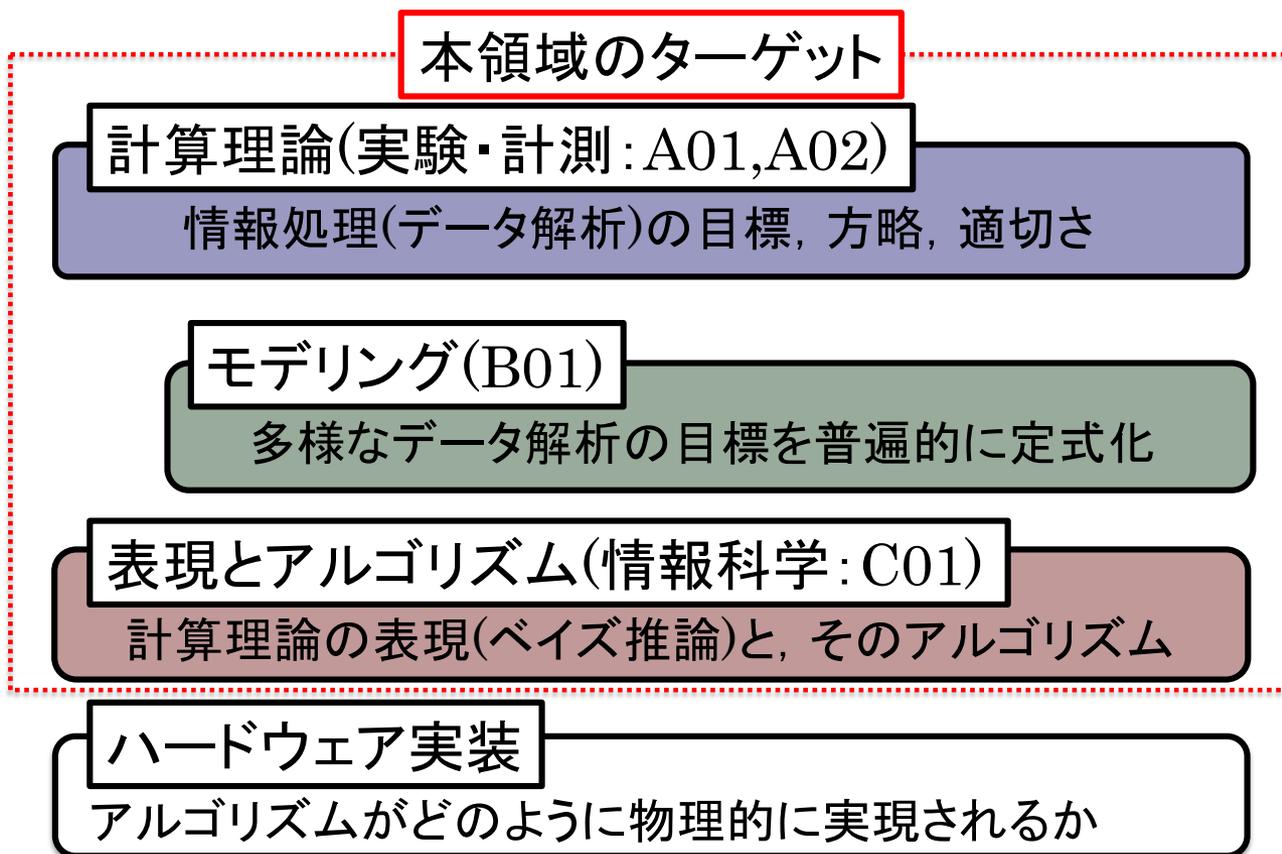
計算理論の表現(ベイズ推論)と, そのアルゴリズム

ハードウェア実装

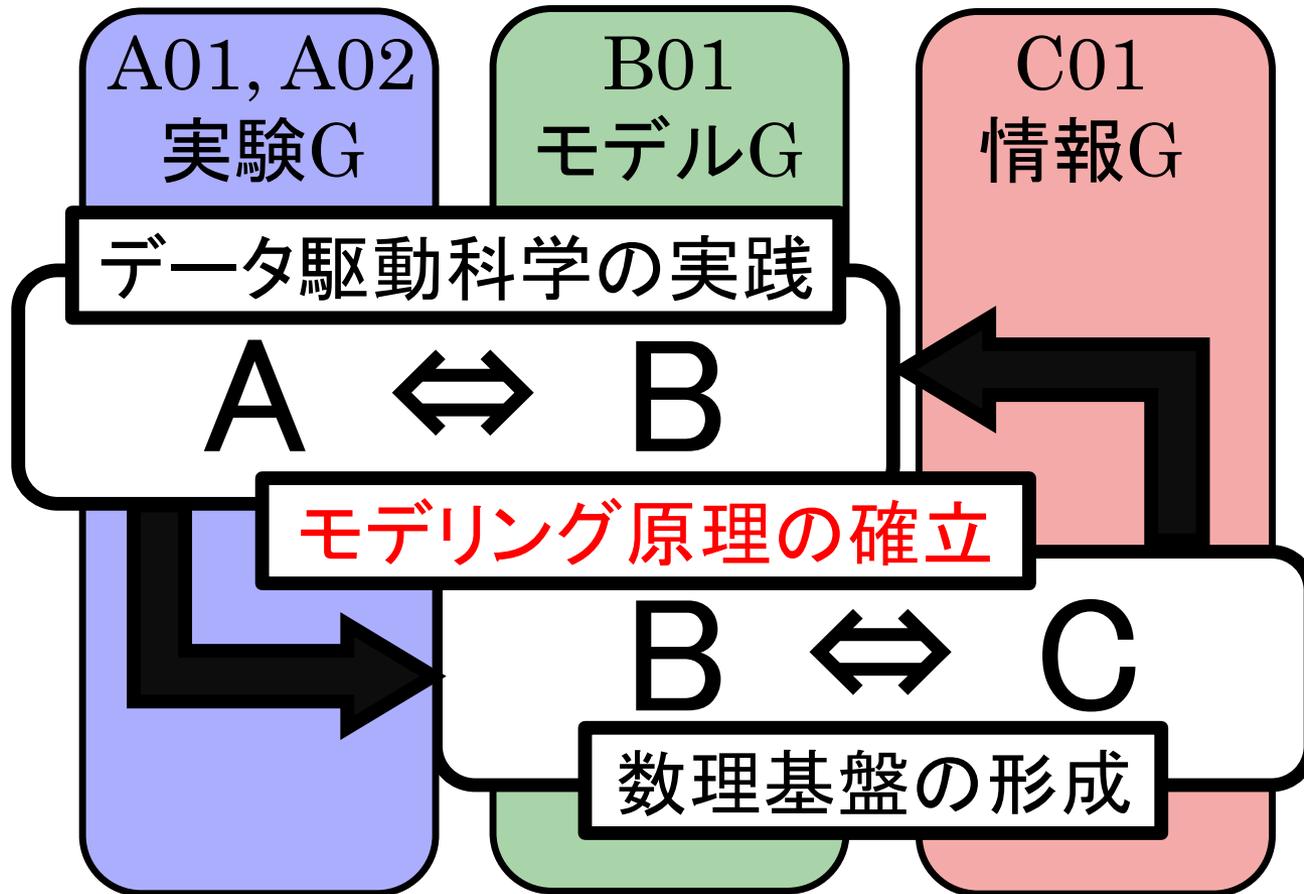
アルゴリズムがどのように物理的に実現されるか

高次元データ駆動科学の学理の原点の創成

- David Marrが指摘した三つのレベルを参考に、“データ駆動科学の三つのレベル”を提唱し、データ駆動科学の学理の原点に位置付けた。

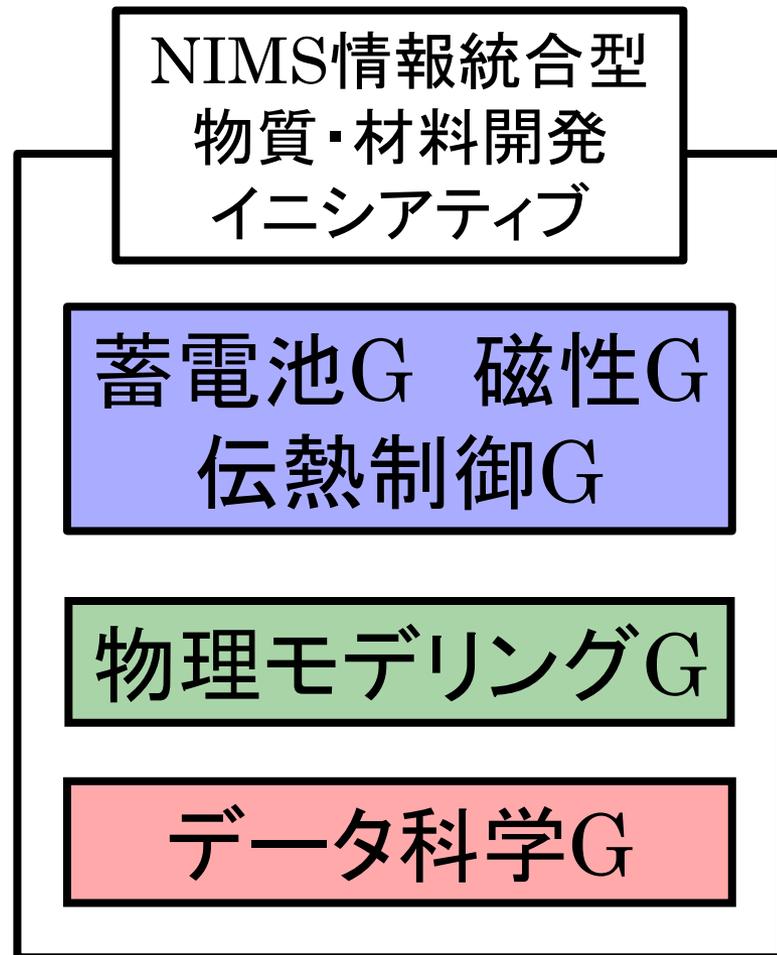


モデルGを要とした領域の融合



データ駆動科学の三つのレベル

NIMS情報統合型 物質・材料開発イニシアティブ



内容

- マテリアルズインフォマティクス(MI)の背景
- MIとバイオインフォマティクス(BI)の相違点
- 構造材料を例とするMI具体例
- 文科省新学術領域 疎性モデリングの紹介
- データ駆動科学の三つのレベルとMI²Iの構成
- 物質材料開発へのデータ駆動的アプローチ

人工知能との関係

- 第3次人工知能ブーム
 - IBM ワトソン
- 第3次ニューラルネットワークブーム
 - ディープラーニング
- 知識駆動型人工知能
 - 記号処理, テキストマイニング
- データ駆動型人工知能
 - 機械学習(カーネル法, ベイズ推論, ディープラーニング)
- MIについては, データ駆動型先行で進み, 後ほど知識駆動型を導入し, 統合する方向

データ駆動型物質材料科学の三つのステップ

