

ERATO湊プロジェクトの概要と 日本科学未来館を通した研究成果の発信

湊 真一

JST ERATO 研究総括

(北海道大学 大学院 情報科学研究科 教授)



自己紹介

■ 湊 真一 (みなと しんいち)

北海道大学 大学院 情報科学研究科 教授

JST ERATO湊離散構造処理系プロジェクト 研究総括 札幌

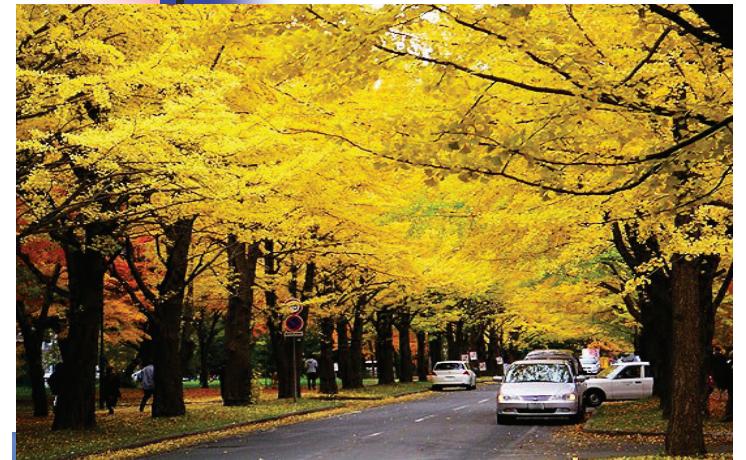


- 石川県出身 金沢大学附属高校卒
- 1990年 京都大学 大学院 修士課程 修了
(1995年 博士(工学))
- 1990年～2004年 NTT研究所(厚木/横須賀) 研究員
- 1997年 米国スタンフォード大学 客員研究員(1年間)
- 2004年～ 北海道大学 情報科学研究科 助教授
- 2010年～ 同 教授

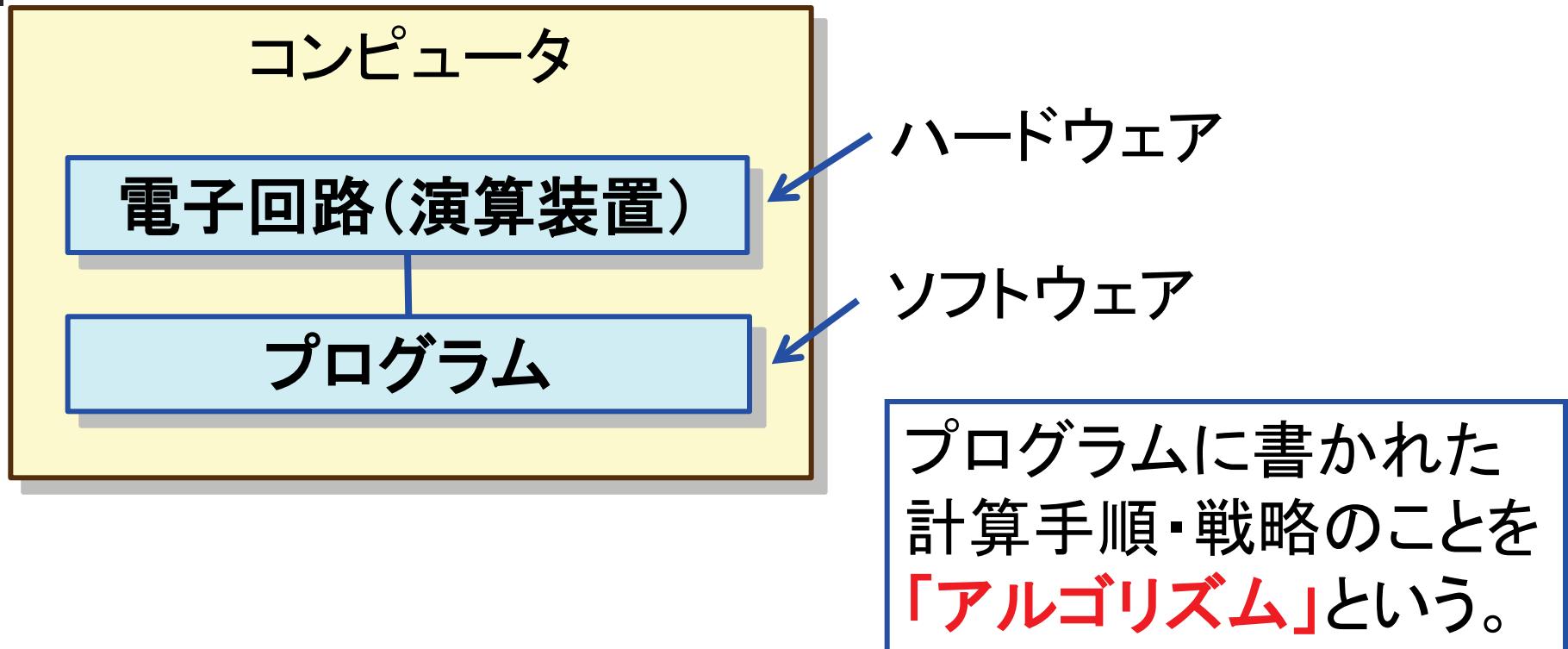
(北大に着任して11年目)



北海道大学@札幌



「アルゴリズム」を研究するERATOプロジェクト



- コンピュータで様々な計算を行う際に、
アルゴリズムの工夫によって、
何十倍、何百倍も計算時間が違ってくることがある。

計算時間を短縮するには



■ スーパーコンピュータを使う？

- スパコン「京」は
64万個以上の演算装置を並べて
毎秒1京回(兆の1万倍)計算できる。

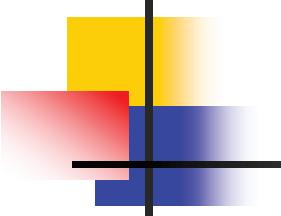


■ ただし誰もが使えるものではない

- 1000億円規模の開発費・年間数十億円の運用費
→ 国家的/世界的に重要な問題を解くために使われる。

■ アルゴリズム技術

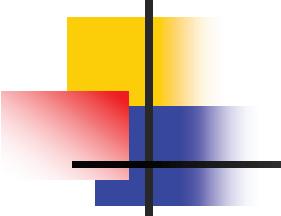
- 与えられた問題を解くために、できるだけ少ない基本演算回数で済まそうとする工夫
- 同じコンピュータでも、何百倍も速くなることがある。
- または、もっと安価で低消費電力なコンピュータで済む。



計算時間がかかる問題とは

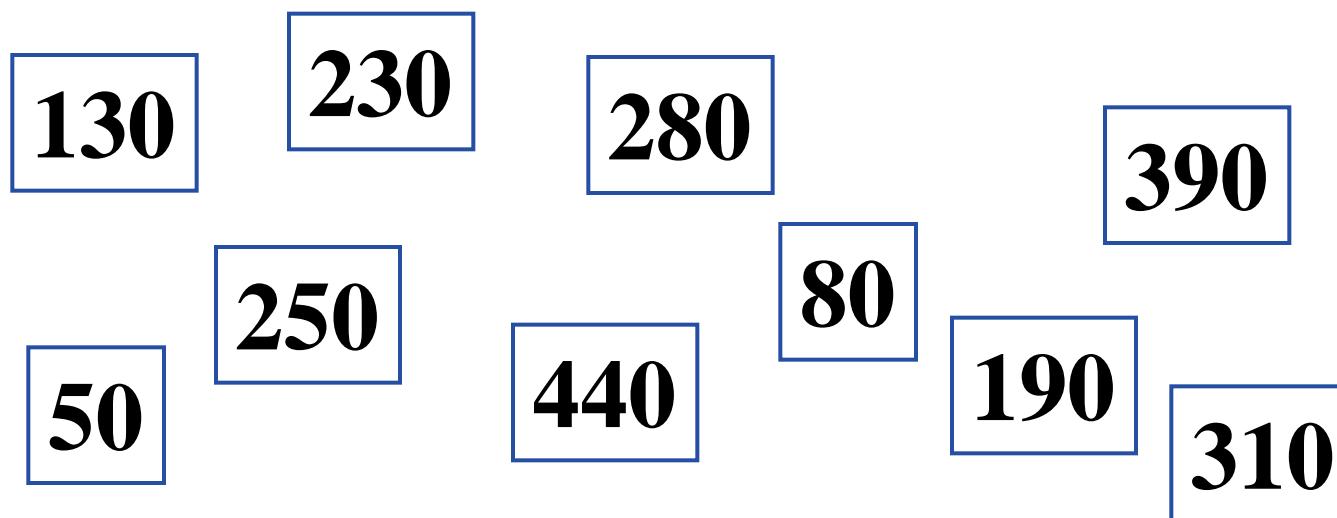
- 超巨大なデータを扱う計算
 - 全国のコンビニの購買データは、年に100万件以上
 - ヒトゲノム(人間の遺伝情報)は、3億文字以上の長さ
 - インターネットのwebページの総数は、(推定)1兆ページ以上
- データはそれほど巨大でないが、膨大な数の場合分けが必要な計算
 - 囲碁・将棋の勝ち負けパターンの分類
 - 数独のヒントの配置
 - 鉄道や道路の経路探索

本プロジェクトでは
主にこちらを考える



身近な問題

- クーポン券が10枚あります。
1000円の商品を買いたいのですが、おつりは出ません。
ぴったりの組合せはあるでしょうか？
- そもそも組合せ方は全部で何通り？

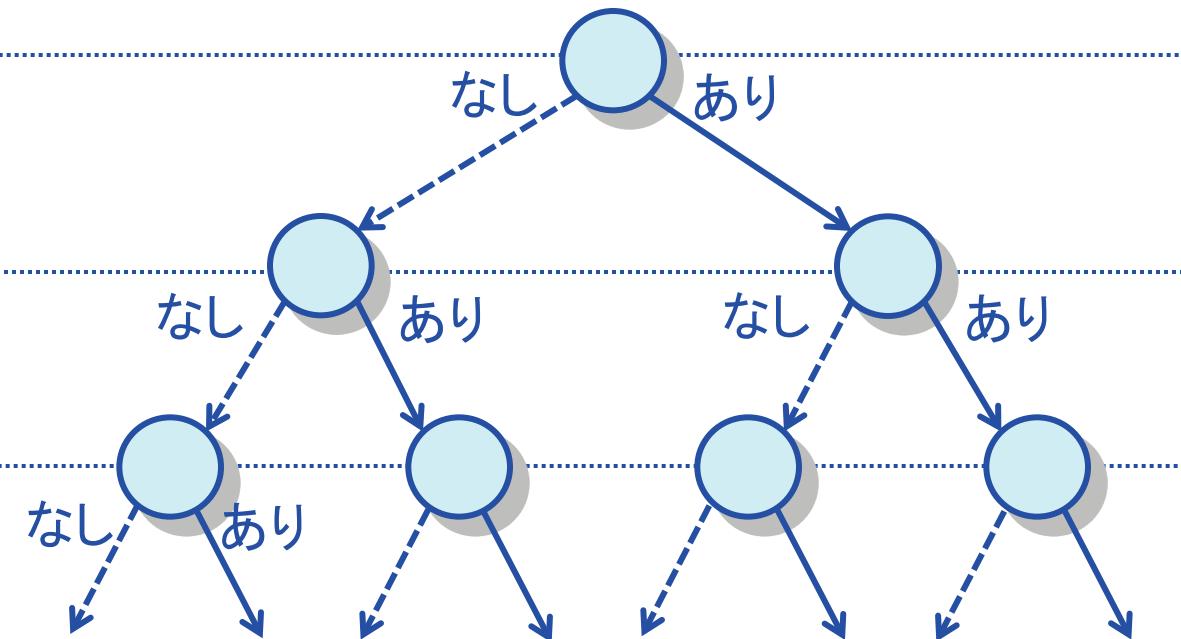


二分木の考え方

クーポン1

クーポン2

クーポン3



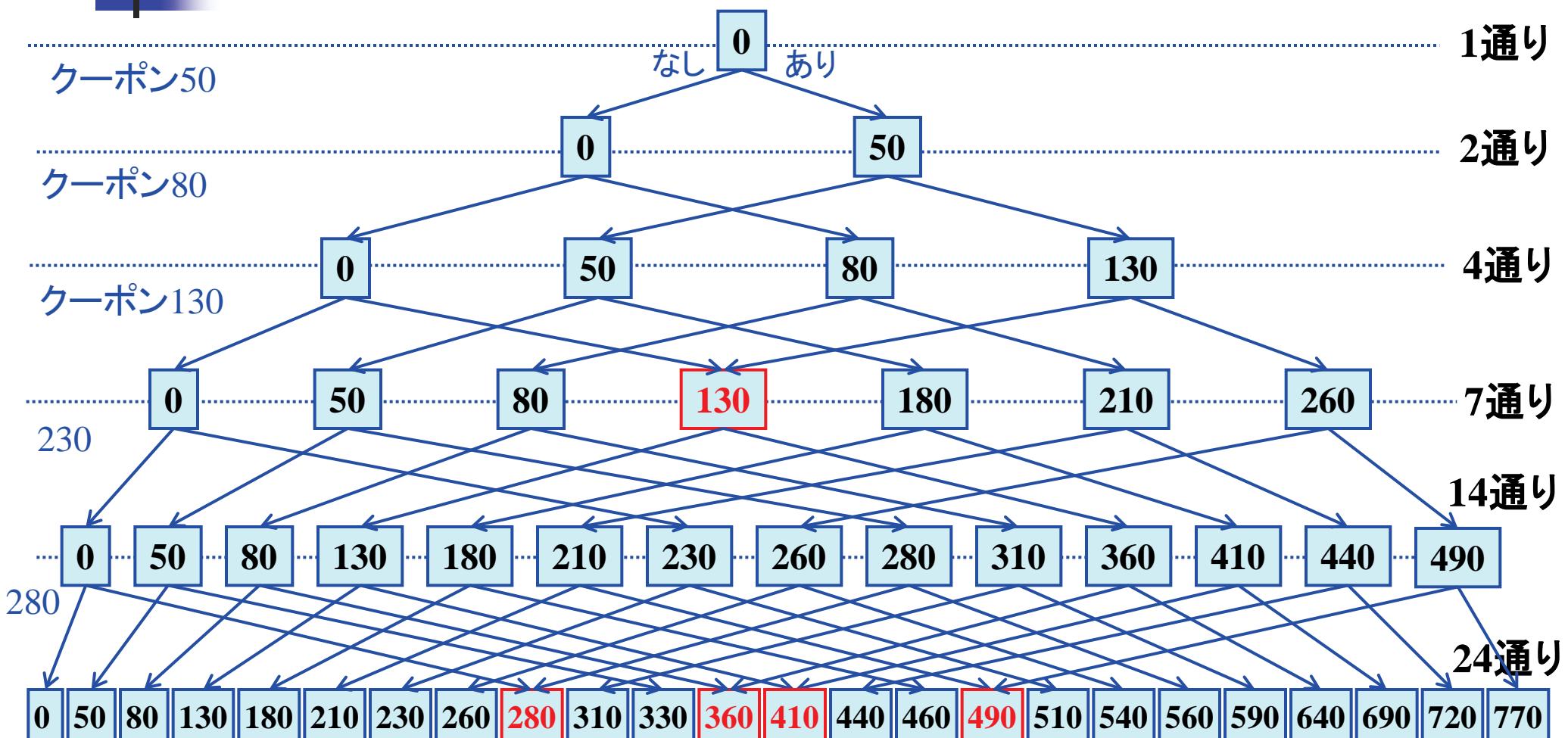
- クーポンが1つ増えるたびに組合せの総数は2倍になる。
 - クーポンの枚数を n とすると、組合せの総数は 2^n 通り
 - $n=10$ のときは全部で1024通り
(ちょうど1000円になるのは、その中の10通りだけ)

2のn乗の脅威（→指数爆発）

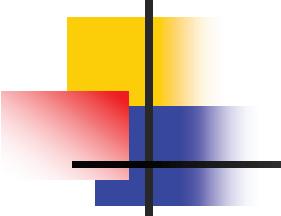
n	2^n
1	2
2	4
3	8
4	16
5	32
6	64
7	128
8	256
9	512
10	1,024

n	2^n
10	1,024
20	1,048,576
30	1,073,741,824
40	1,099,511,627,776
50	1,125,899,906,842,624
60	1,152,921,504,606,846,976
70	1,180,591,620,717,411,303,424
80	1,208,925,819,614,629,174,706,176
90	1,237,940,039,285,380,274,899,124,224
100	1,267,650,600,228,229,401,496,703,205,376

アルゴリズムの工夫



最後まで行っても198通りしか出て来ない (→ 1024通り調べなくても済む！)



アルゴリズム技術の重要性

- 今回の例では、1024通りが198通りに減るので、アルゴリズムの工夫で約5倍の高速化
→ 問題が大規模になるほど高速化の倍率が大きくなる。
(クーポンの枚数が20枚になると、効果が数十倍、数百倍になる例も)
- スーパーコンピュータ開発のような、コンピュータそのものの高速化は、問題の規模に関係なく一定の倍数だけ速くなる。
- アルゴリズムの改良は、問題の規模が大きくなるほど高速化の倍率が高くなることが多い。→劇的な効果
 - 特別な装置を使わないので費用もからない。
 - ただし、どんな問題でも高速化できるとは限らない。
問題ごとに最適なアルゴリズムの研究開発が必要。
→ アルゴリズムの工夫が効かない問題ではスパコンを使うしかない

ERATO湊離散構造処理系プロジェクト

■ 様々な工学的応用を持つ基盤技術として
「離散構造処理系」に着目し、研究開発を行う

システム設計自動化

データマイニングと知識発見

大規模システム故障解析

機械学習と自動分類

制約充足問題

生命情報科学

web情報解析

工学的応用
→ 社会への
影響大

離散構造処理系

本研究の
対象領域

集合論理

記号論理

帰納的証明

組合せ論

グラフ理論

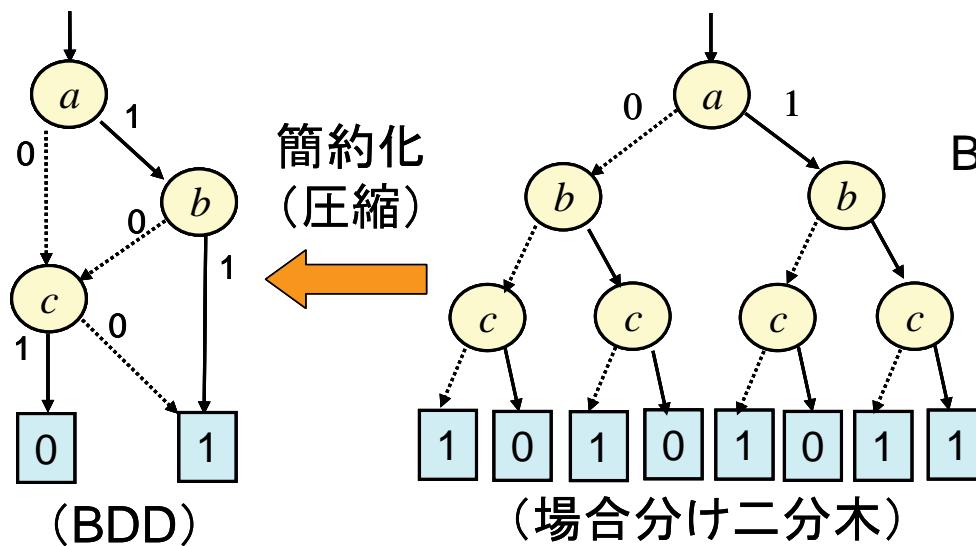
確率論

性能向上
(10倍～
100倍以上)

数学的
概念構造

BDD(二分決定グラフ)

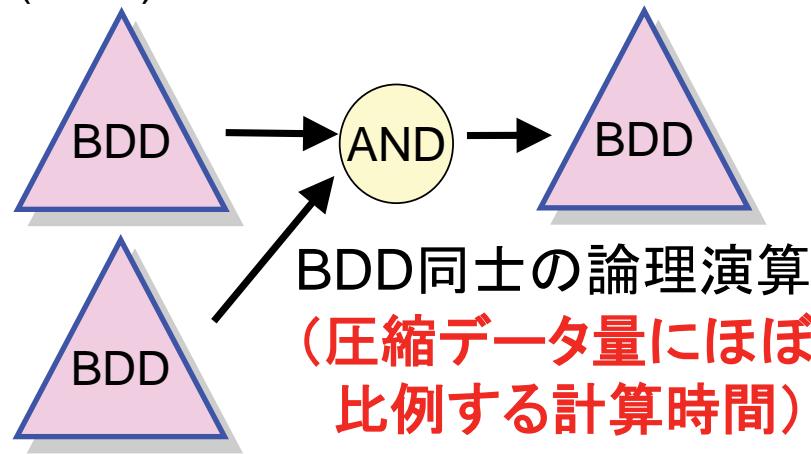
離散構造の最も基本的なモデルである「論理関数」の処理技法



- ・場合分け二分木グラフを簡約化(データ圧縮)
- ・多くの実用的な論理データをコンパクトかつ一意に表現。(数十～数百倍以上の圧縮率が得られる例も)



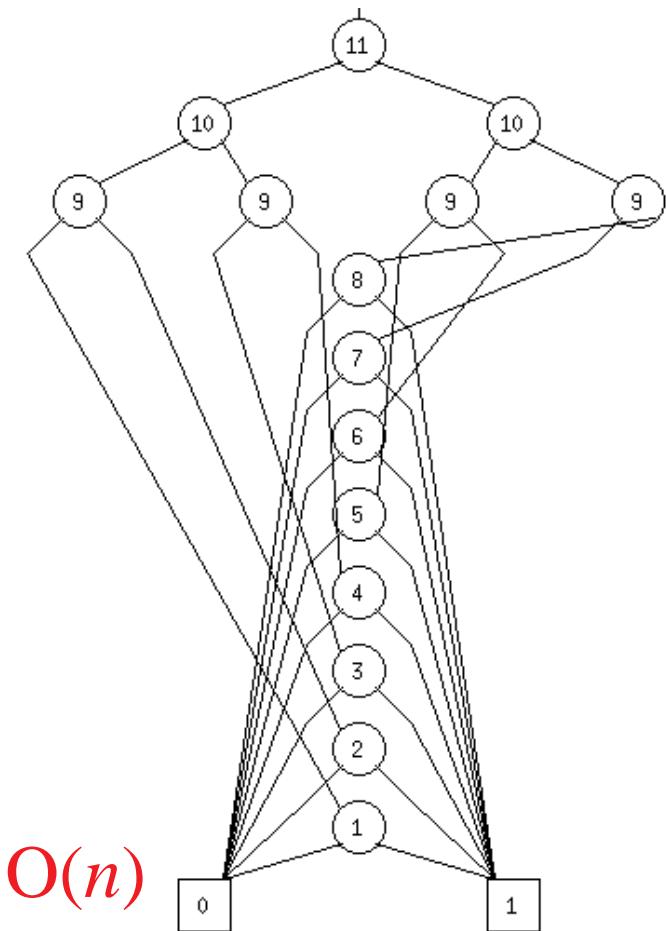
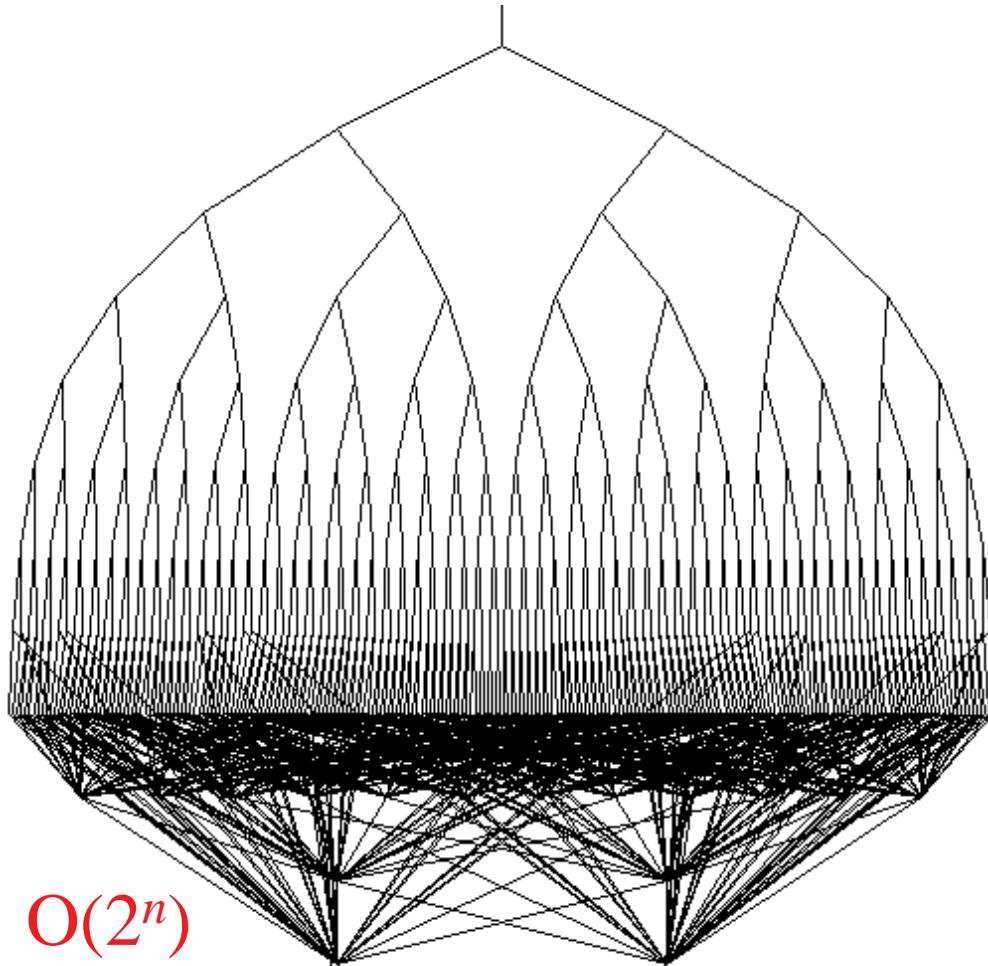
1986年に画期的なBDD演算
(**Apply演算アルゴリズム**)を提案。
以後急速にBDD技術が発達。
(長期間、情報科学の全分野
での最多引用文献となった)



近年のPC主記憶の大規模化により、
BDDの適用範囲が拡大
(特に2000年以降)

BDD簡約化の効果

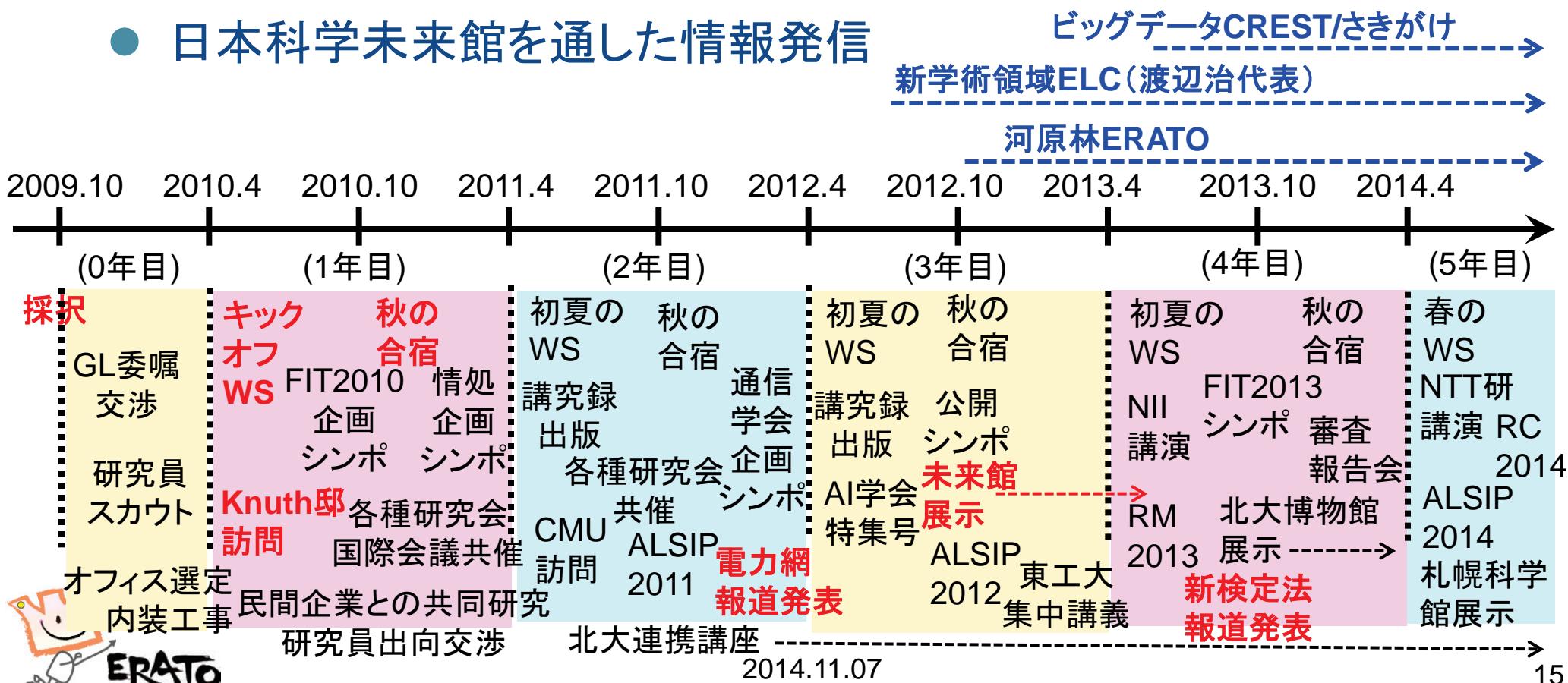
- 特定の問題では、指数関数的な圧縮効果が得られる。
 - 例題に依存するが、多くの実用的な問題で効果がある。



ERATO湊プロジェクトの活動経過

■ 2010年度より研究活動を開始して、約5年が経過。

- 今年度が最終年度(来年度1年の延長が決定-軟着陸期間)
- 開放的な組織で、多様な研究者が集まる「場」を形成
- 基礎技術・工学的応用の両面での研究成果
- 日本科学未来館を通した情報発信



電力網への応用(2012年2月プレスリリース)

■ 林 泰弘 教授(早稲田大)との共同研究

- 電力系スマートグリッド業界のリーダ的存在
(経産省スマートハウス標準化検討会座長、他多数の要職)
- 電力網最適化の研究で**1990年代より湊と協力関係**

■ 大震災後、より緊急性の高い研究課題に

- 今後長期的に不足する電力を自然エネルギーで補うために必須の電力網解析・制御技術を支援

情報科学の研究者集団として
我が国の苦境を克服するため
できる限り貢献したい。

→ ERATOプロジェクト
での取り組みを加速



RIANT 先進グリッド技術研究所
Research Institute of Advanced Network Technology

ホーム > 研究所案内

研究所案内

所長ご挨拶 先進グリッド技術研究所 研究テーマと分野 研究概要

所長ご挨拶



林泰弘 教授

地球温暖化や環境破壊問題など地球規模の深刻な課題が取り上げられない日はありません。現代のエネルギー問題の解決には、この地球環境保全を考えつつエネルギーをいかに安定して供給するか、加えて経済成長をも調和させる取り組みが急務となっています。その究極の解決策になると期待されているのがスマートグリッドです。

スマートグリッドは電力エネルギーネットワークにITの技術や方法を応用していく、より効率的な電力利用の実現を目指します。太陽光発電や風力発電などの再生可能エネルギー源の大規模導入、ヒートポンプ給湯器などの熱エネルギー利用の電化、電気自動車の普及による移動手段の電化などを通じて、CO₂排出量が少なく環境にやさしい電化社会へすむわざ無く低炭素電化社会の到来に私たちが注目しているといえます。

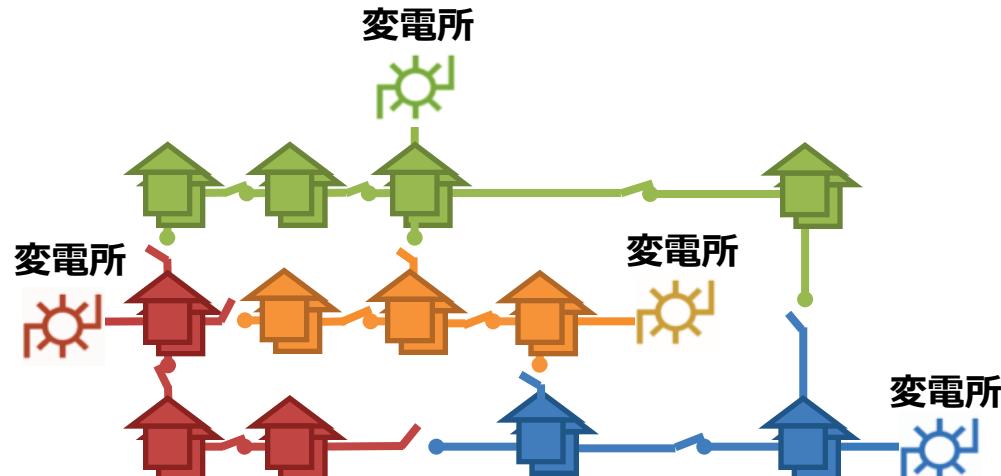
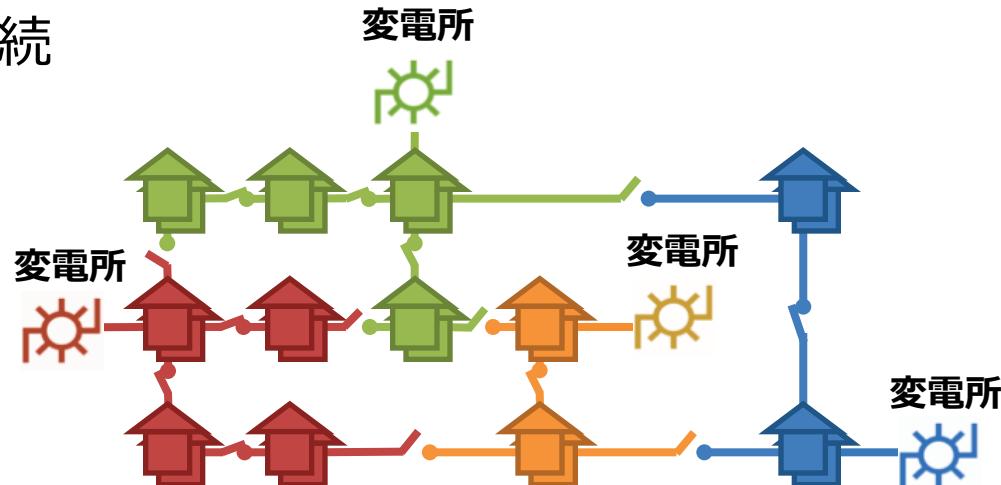
電力網のスイッチ制御

- ・各区域が必ず 1 ケ所の変電所設備に接続
- ・停電しない
- ・異なる変電所系統をつながない
- ・電流が多過ぎると電線が焼ける
- ・遠くに伝えると電圧が下がる

膨大な数の構成が存在

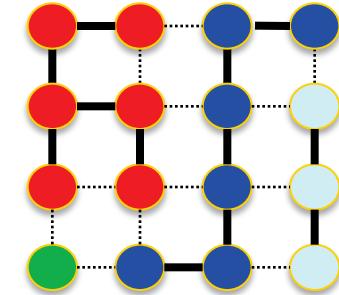
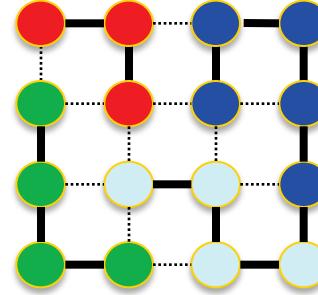
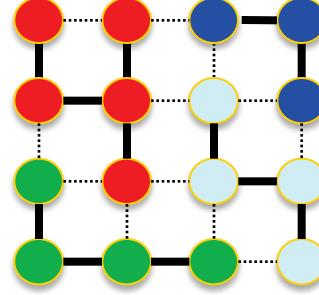
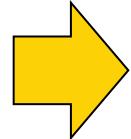
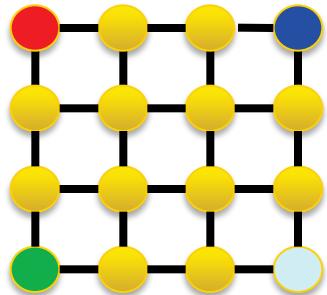
右のような13スイッチの小さな例題でも
8192通りの組合せの中から
210通りを探し出す問題

標準的な配電網の例では
468個のスイッチがあり、
10の140乗の組合せから
探し出す問題に



電力網のスイッチ制御

- ・ZDDを使った新しい高速アルゴリズムを開発。
- ・標準的な電力網モデル（スイッチ468個）でグラフ的制約と電気的制約を共に満たす解を全て求め、圧縮して表現することに成功（世界初）
- ・電力の損失を最小にする組合せを求める事にも成功



圧縮データ: 約110万個 (779MB)

解の個数: 約 10^{63} (2136那由他8201阿僧祇3834恒河沙8532極9116載8261正
2214潤8049溝560穰9817杼8392垓4438京5235兆3981億8952万1540) 通り

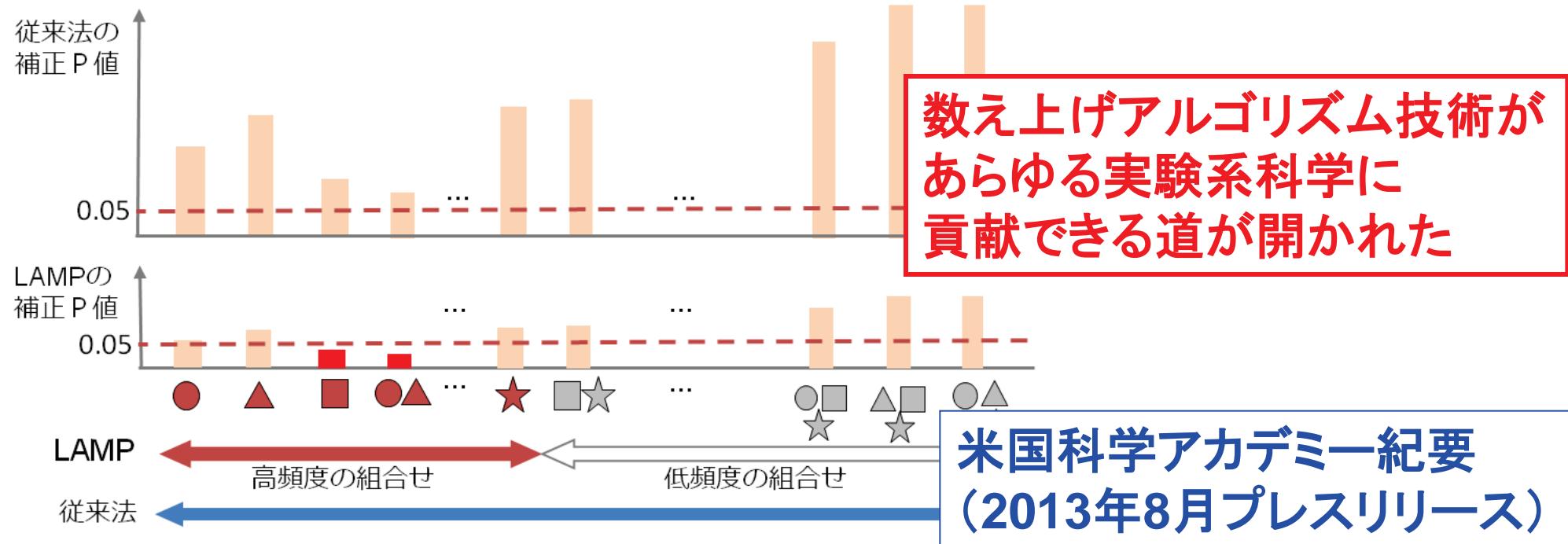
実行時間: 数十分

IEEE(米国電気電子学会)
スマートグリッド分野
専門論文誌 2編掲載

電力網のほかにも様々な応用がある

- ・避難所の配置問題
- ・選挙の区割り問題
- ・道路、鉄道、ガス、水道、通信
- ・建物のフロアプラン

LAMP: ビッグデータから科学的発見をもたらす統計手法



Statistical significance of combinatorial regulations

Aika Terada^{a,b,c}, Mariko Okada-Hatakeyama^d, Koji Tsuda^{c,e,1}, and Jun Sese^{a,b,1}

^aDepartment of Computer Science and ^bEducation Academy of Computational Life Sciences, Tokyo Institute of Technology, Meguro-ku, Tokyo 152-8550, Japan; ^cMinato Discrete Structure Manipulation System Project, Exploratory Research for Advanced Technology, Japan Science and Technology Agency, Sapporo, Hokkaido 060-0814, Japan; ^dLaboratory for Integrated Cellular Systems, RIKEN Center for Integrated Medical Sciences (IMS-RCAI), Yokohama, Kanagawa 230-0045, Japan; and ^eComputational Biology Research Center, National Institute of Advanced Industrial Science and Technology, Koto-ku, Tokyo 135-0064, Japan

Edited by Wing Hung Wong, Stanford University, Stanford, CA, and approved July 3, 2013 (received for review February 4, 2013)

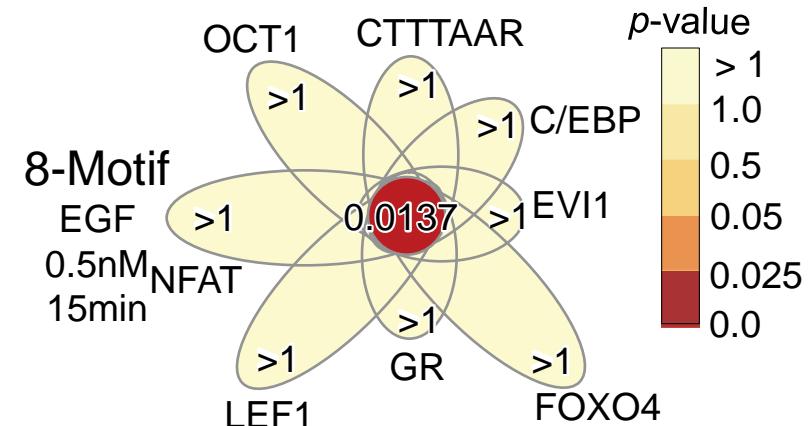
More than three transcription factors often work together to enable cells to respond to various signals. The detection of combinatorial regulation by multiple transcription factors, however, is not only

deliberately excluding such tests. Here, we propose an efficient branch-and-bound algorithm, called the “limitless arity multiple-testing procedure” (LAMP). LAMP counts the exact number of 19

LAMP (Limitless Arity Multiple testing Procedure)

- 実験科学では、統計検定値(p値)が有意でなければ、科学的発見として認められない
- これまでの検定法は単一因子の発見には有効だったが、複数因子の組合せでは実用的な精度で検定できなかった。
- 近年、組合せ因子の発見が重要になってきた
 - 例えばiPS細胞は4つの因子の組合せにより実現
- LAMP法:超高速な数え上げアルゴリズム技術により、組合せ因子を高精度で検定

これまで見過ごされてきた、乳がん細胞の分化に関する転写因子の組合せを発見！



日本科学未来館での成果展示

■ 日本科学未来館で本プロジェクトの成果展示「フカシギの数え方」を開催

- 2012年8月1日～2013年4月25日
(期間中、夏・冬・春休み1回ずつ)
- 合計23万人が来場

■ 好評につき再展示

- 北大総合博物館(2013年7月～2014年4月)
- 札幌市青少年科学館
(2014年春の大型連休期間)
- 今回のサイエンスアゴラ2014で
里帰り展示中(1階・大地球儀の真下)



Miraikan



なぜ未来館で？

■ JST本部からの紹介(具体的な指示ではなく可能性の打診のみ)

- 約1年周期で最新のメディアアートを展示するコーナー
- 昔、ERATO五十嵐プロジェクトの展示をした前例があった。
(ただしメディアアート系の研究だったので成果＝展示作品)
- アルゴリズム技術のような実体のない研究の展示実績は無し

■ 未来館スタッフにプロジェクト内容を紹介したところ、非常に興味を持ってもらえて、展示することに。

- 「那由他」や「不可思議」のような巨大数が遊びでなく出てくる世界
- 「組合せ爆発のすごさとアルゴリズム技術の威力」を見せる
というコンセプト
- ただし、何もないところから展示を作り上げる必要があった。



アルゴリズム技術の展示の難しさと新しさ

■ 未来館は大きく分けると

「宇宙地球科学」「生命科学」「情報科学」の3分野

- 宇物地物系：宇宙船・潜水艇・ロケットなど置くだけで迫力
- 生物系：脳の断面図とかDNA構造とか研究対象が見える
- 情報系：何となく未来的だが何に挑戦しているのか伝わりにくい
(ロボットはわかりやすいが、それ以外は一般に難しい)

■ アルゴリズムという実体のない物をどう見せるか

- 「新しい展示手法の開発」は未来館のミッションの1つ
- 展示開発課を中心に熱心に取り組んでいただいた。
- 実績のある空間デザイナーやアートクリエーターの紹介
- プロジェクトの若手研究者はなるべく本務の研究に専念させて
主に研究総括が中心となって準備作業にあたった

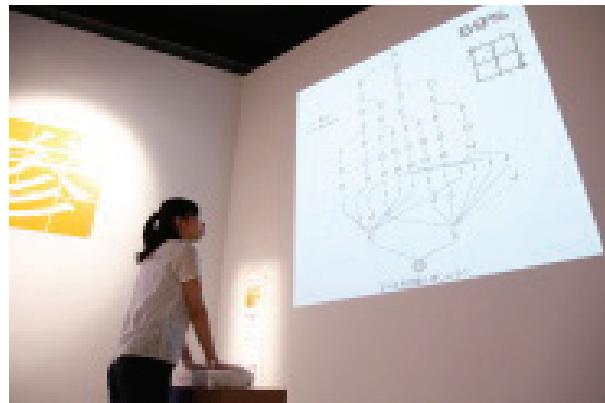
展示の工夫



インタラクティブ展示
(ペンや手でなぞってみる)



巨大な数表
(視野の広さで実感)



体感展示
(体重をかけて圧縮)



顔写真と説明文
(研究者が語りかける)

展示の工夫



動画展示
(計算時間の経過を実感)

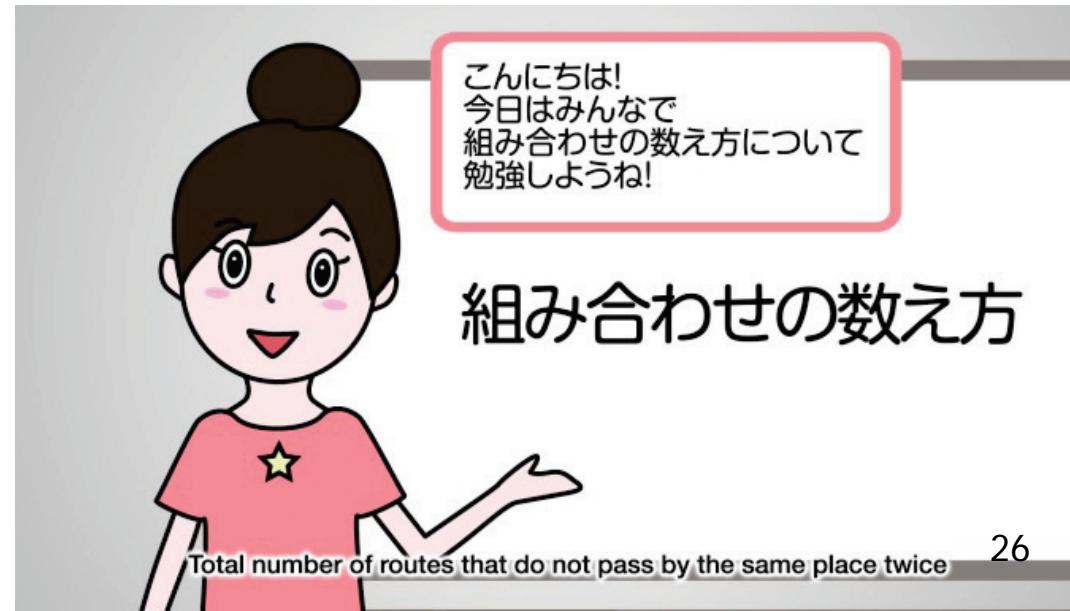


コミュニケーションスペース
(畳の上でゆっくり遊び学べる)

サイエンスアゴラ2014で実物の再現展示をご覧になれます。
(未来館1階・巨大地球儀の真下のブース002)

アニメーション動画の反響

- YouTubeの未来館公式チャンネルに登録・公開
 - サイエンス系コンテンツでは極めて異例の大ヒットを記録
→ **これまでの再生回数158万回** (ニコ動にも転載され66万回を記録)
 - ネット系ニュースメディアで多数取り上げられた
 - 海外からも大きな反響(世界各国語で多数のコメント)
 - 完全オリジナル作品で著作権問題の心配なし
→ 国内外の高校・大学の授業でしばしば使われている
(今でも毎日、数十～数百回
再生されている)
 - 未来館とネット系メディア
の連動企画の成功例に



アップロード済み

人気の動画

グリッド



『つかぎの数え方』おねえさんといっしょ！みんなで数え...
再生回数 1,586,344 回 2年前

CC



「つかぎの数え方」同じところを2度通らない道順の数
再生回数 199,641 回 2年前



常設展示『アナグラのうた』～消えた博士と残された装置～
再生回数 24,991 回 2年前



常設展示『アナグラのうた』寺田克也氏による壁画ライブペイ...
再生回数 22,449 回 1年前



企画展「トイレ？行っとトイレ！～ボクらのうんちと地球のみ...’
再生回数 18,332 回 4か月前



『整腸ラップ』～SAY! Cho-oh Rap～
再生回数 13,009 回 3か月前



企画展『世界の終わりのものがたり～もはや逃れられない73...』
再生回数 7,093 回 2年前



日本科学未来館 -展示案内-
再生回数 6,523 回 2年前



企画展「THE 世界一展～極める日本！モノづくり～」
再生回数 5,991 回 8か月前



サイエンティスト・トーク「3.11の地震はまだ終わっていない」
再生回数 5,977 回 1年前



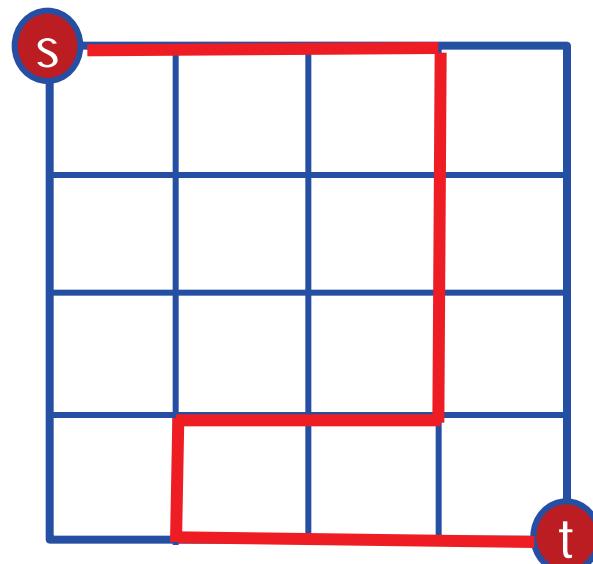
新型ASIMOを用いた自律型説明ロボット実証実験
再生回数 5,827 回 1年前



宇宙から見た地球 / Earth from Space
再生回数 5,768 回 1年前

“self-avoiding walks”の数え上げ

- 最短経路の数え上げは簡単
($\rightarrow {}_{2n}C_n$; 高校で習う問題)
- 最短でない経路を許すと突然難しくなる。
(計算式や漸化式は見つかっていない)
 - おねえさんが25万年かかった結果がアニメの中に表示されているため、「計算式教えて」というコメント多数。
 - 残念ながら計算式は知られていない。
効率良く数え上げるしかない。



「おねえさんの問題」の世界記録

- ERATO研究員の岩下氏が世界記録を更新(2013年12月)
 - $n=26$ までの数え上げに成功(1.736×10^{163} 通り)
 - ノルウェーの大学が記録した $n=24$ (2013年2月)を2段階上回る。
- 数列のオンライン百科事典(OEISサイト)に公式登録
 - 世界記録に実用的意味はあまりないが、子供に夢を与える効果

<i>n</i>	The number of paths
1	2
2	12
3	184
4	8512
5	1262816
6	575780564
7	789360053252
8	3266598486981642
9	41044208702632496804
10	1568758030464750013214100
11	182413291514248049241470885236
12	64528039343270018963357185158482118
13	69450664761521361664274701548907358996488
14	227449714676812739631826459327989863387613323440
15	2266745568862672746374567396713098934866324885408319028
16	68745445609149931587631563132189232821587945968099457285419306
17	634481461123796397131029754079552440044943986866480693646369387855336
18	1782112840842065129893384946652325275167838065704767655931452474605826692782532
19	1523344971704879993080742810319229690899454255323294555776029866737355060592877569255844
20	3962892199823037560207299517133362502106339705739463771515237113377010682364035706704472064940398
21	31374751050137102720420538137382214513103312193698723653061351991346433379389385793965576992246021316463868
22	755970286667345339661519123315222619353103732072409481167391410479517925792713631234987038883317634987271171404439792
23	55435429355237477009914318489061437930690379970964331332556958646484008407334885544566386924020875711242060085408513482933945720
24	12371712231207064758338744862673570832373041989012943539678727080484951695515930485641394550792153037191858028212512280926600304581386791094
25	8402974857881133471007083745436809127296054293775383549824742623937028497898215256929178577083970960121625602506027316549718402106494049978375604247408
26	17369931586279272931175440421236498900372229588288140604663703720910342413276134762789218193498006107082296223143380491348290026721931129627708738890853908108906396

未来館展示を通して感じたこと

■ 未来館の展示は、成功しても失敗しても影響が大きい

- うまく行けば研究に対する社会的理 解が大きく広がる
 - プロジェクト研究員の士気向上(家族や友人の評判)
 - 関連学科や研究室の人気上昇(将来の人材の獲得)
- 面白くなれば研究費の浪費と言われるおそれ
 - 説明員の人工費、展示装置の製作費など、大きな費用

■ 「大型研究プロジェクトでは納税者への広報活動は重要」と常々言われているが…

- 実際やろうとすると結構たいへん。
- 中途半端にやっても効果が目に見えない。
(大学の業績調書では「その他の業績」にしかならない。)
- やるなら本気でやるべき。うまく当たれば大きなメリットがある。
- ERATOのおかげで予算・人員を機動的に投入することができた。

情報系研究者から見た日本の科学技術展示の課題

■ 日本科学未来館と国立科学博物館の役割分担

- 未来館は「未来のこと」、科博は「過去のこと」と「普遍的なこと」
- 科博は**「収蔵」がメインの仕事**で、「展示」はオーソドックス
- 未来館は「展示」は強力だが**収蔵機能が全くない。**

■ 未来館と科博の谷間がだんだん広がっている印象

- 科博は伝統的な博物学は強いが、新興学問分野は弱い。
特に情報系は時代の変化が速いので、橋渡しが難しい
- 科学技術の発展の展示は、江戸初期の「塵劫記」から始まって、からくり人形、自動織機、アナログ計算機、みどりの窓口1号機（昭和39年）までで終わってしまっている。**(50年分の空白)**
- 普遍的法則の展示では、「メートル」「グラム」「ボルト」のような物理単位の説明はあるが、**情報量「ビット」の説明は見当たらない。**
- 科博にも未来館にも間を埋める予算・人員の余裕はなさそうだが何とかしないといけないのでないか？

ご清聴ありがとうございました。