TOSHIBA

第2回量子アニーリング及び関連技術に関する研究会 2024/02/20 15:50~16:25

シミュレーテッド分岐マシンのリアルタイムシステムへの応用-金融・車載の事例-

辰村光介

株式会社 東芝 研究開発センター 情報通信プラットフォーム研究所 コンピュータ&ネットワークシステムラボラトリー

Contents

はじめに

シミュレーテッド分岐マシンのリアルタイムシステムへの応用 -金融・車載の事例-

01 量子インスパイアード最適化技術をリアルタイムシステムへ応用する

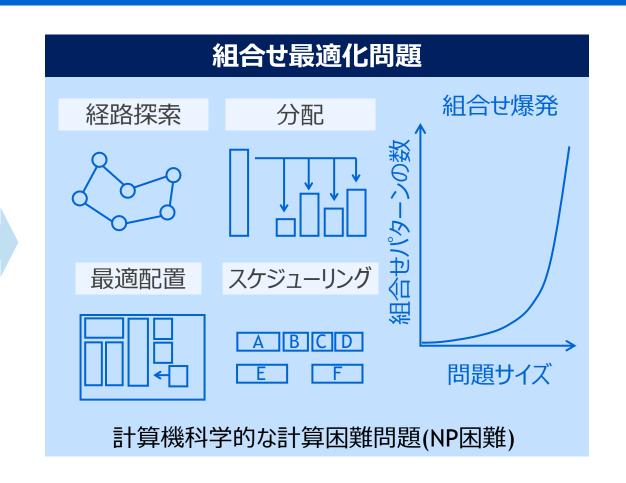
02 シミュレーテッド分岐マシン(SBM)の理論と実装、性能

SBMのリアルタイムシステムへの応用:金融と車載の事例

組合せ最適化問題とは

経済的に重要, しかし, 計算するのが困難





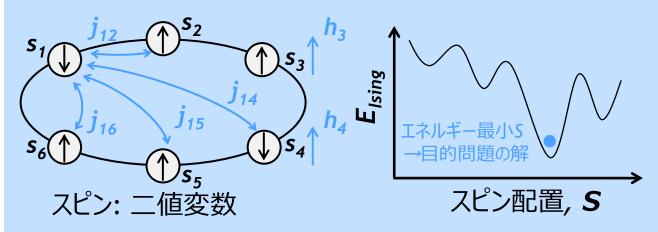
組合せ最適化問題を解く = その状況において最も合理的な判断を下すこと

組合せ最適化(二次離散最適化)問題の高速求解のための特定目的計算機

イジング問題

イジングモデルの最小エネルギー状態を探索する

$$E = -\sum_{ij} j_{ij} s_i s_j + \sum_{ij} h_i s_i$$



マッピング#1

組合せ最適化(二次離散最適化)

イジングマシン

Quantum annealer*1 (2011)



Memristor HNN*6

CMOS annealer*2

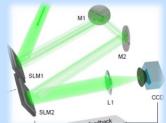






Optical Ising machines*4,5





and more ...

+ Simulated bifurcation machine (2019)

- *1 https://www.dwavesys.com/d-wave-two-system
- *2 https://www.hitachi.co.jp/New/cnews/month/2019/02/0219.html
- *3 https://www.fujitsu.com/global/about/resources/ news/press-releases/2018/0515-01.html
- *4 https://www.ntt.co.jp/news2017/1711e/171120a.html *5 D. Pierangeli, et al., Phys. Rev. Lett. 122, 213902 (2019).
- *6 F. Cai, et al., Nature Electronics 3, 409 (2020).

量子インスパイアード最適化計算機

- *2 H. Goto et al., Sci. Adv. 7, eabe7953, '21
- *3 K. Tatsumura et al., Nat. Ele. 4, 208–217, '21

シミュレーテッド分岐マシン (Simulated Bifurcation Machine, SBM)

アルゴリズム

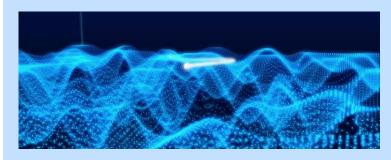
量子インスパイアード

量子分岐マシン (量子原理)



シミュレーテッド分岐アルゴリズム

(新しい解探索原理)

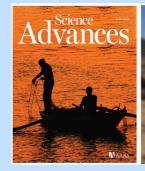


高い並列化可能性 Highly parallelizable

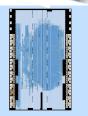
計算機実装

高速な求解性能*1,2

シングルチップ

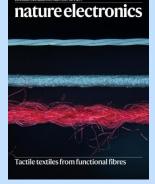






高い拡張性*3

マルチチップ





応用

高い実用性

エッジ/組込み

クラウド











ex. リアルタイムシステム



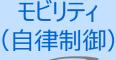
組込みSBM技術(Embeddable SBM)

従来困難だった、より合理的な判断をするリアルタイムシステムを実現

リアルタイムシステム

- ・時間的制約のあるシステム:決められた時間までに応答しなければならない
- ・"高速"リアルタイムシステム:応答時間は典型的にミリ秒程度以下
- ・厳しい時間制約のため、単純な条件判定に基づき応答するものが多い

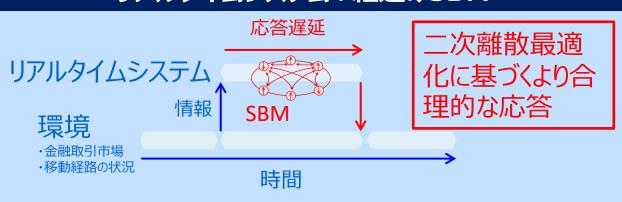
金融トレーディング

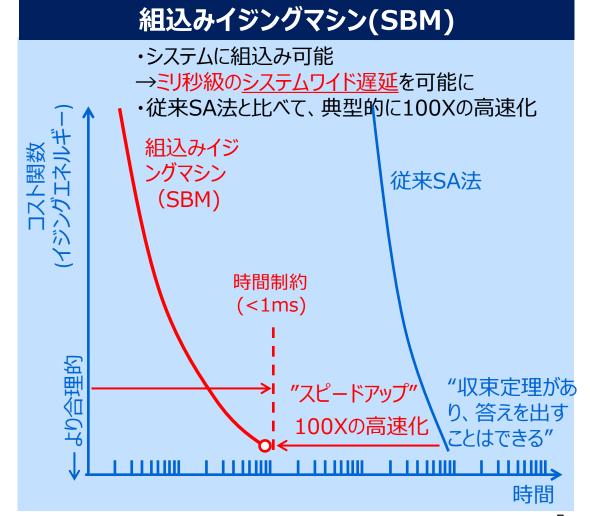






リアルタイムシステム+組込みSBM





Contents

はじめに

シミュレーテッド分岐マシンのリアルタイムシステムへの応用 -金融・車載の事例-

01 量子インスパイアード最適化技術をリアルタイムシステムへ応用する

02 シミュレーテッド分岐マシン(SBM)の理論と実装、性能

SBMのリアルタイムシステムへの応用:金融と車載の事例

- *2 H. Goto et al., Scientific Reports 6, 21686 (2016)
- *3 H. Goto et al., Science Advances 5, eaav2372 (2019)

量子分岐マシンの古典対応物からシミュレーテッド分岐マシンは誕生*1

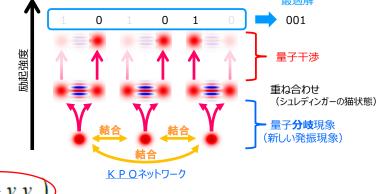
量子分岐マシン(2016)*2

非線形振動子ネットワークの断熱分岐過程を記述するハミルトニアン

高速化のためのアルゴリズム改良

$$H_{q}(t) = \hbar \sum_{i=1}^{N} \left[\frac{K}{2} a_{i}^{\dagger 2} a_{i}^{2} - \frac{p(t)}{2} \left(a_{i}^{\dagger 2} + a_{i}^{2} \right) + \Delta a_{i}^{\dagger} a_{j} \right] - \hbar \xi_{0} \sum_{i=1}^{N} \sum_{j=1}^{N} J_{i,j} a_{i}^{\dagger} a_{j}$$
状態変数の古典化
$$H_{c}(\mathbf{x}, \mathbf{y}, t) = \sum_{i=1}^{N} \left[\frac{K}{4} \left(x_{i}^{2} + y_{i}^{2} \right)^{2} - \frac{p(t)}{2} \left(x_{i}^{2} - y_{i}^{2} \right) + \frac{\Delta_{i}}{2} \left(x_{i}^{2} + y_{i}^{2} \right) \right] - \frac{\xi_{0}}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} J_{i,j} \left(x_{i} x_{j} + y_{i} y_{j} \right)$$

原理:量子断熱定理 (量子でしか成り立たない定理)



シミュレーテッド分岐マシン(2019)*3

そもそも、量子でしか成り立たない定理で動く量子分岐マシンを 古典化して、どうして機能するのだろうか?

→そこには新しい発見があった

- *1 H. Goto et al., Science Advances 5, eaav2372 (2019)
- *2 H. Goto et al., Science Advances 7, eabe7953 (2021)
- *3 T. Kanao et al., Comm. Phys. 5, 153 (2022)
- *4 T. Kanao et al., Applied Physics Express 16, 014501(2023)

新しい古典的メカニズム:断熱探索とエルゴート探索

エネルギーランドスケープ の動的な変化

単一・自明な最小点



分岐

(断熱過程)

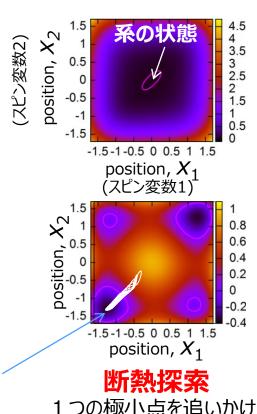


多数の極小点が現れる (コスト関数の関数曲面)

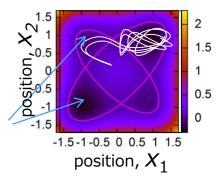
最良解

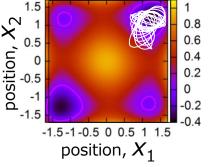
(-1,-1)

エネルギーランドスケープ(N_{spin} =2)



エネルギー的に許される 領域に複数の極小点





エルゴート探索

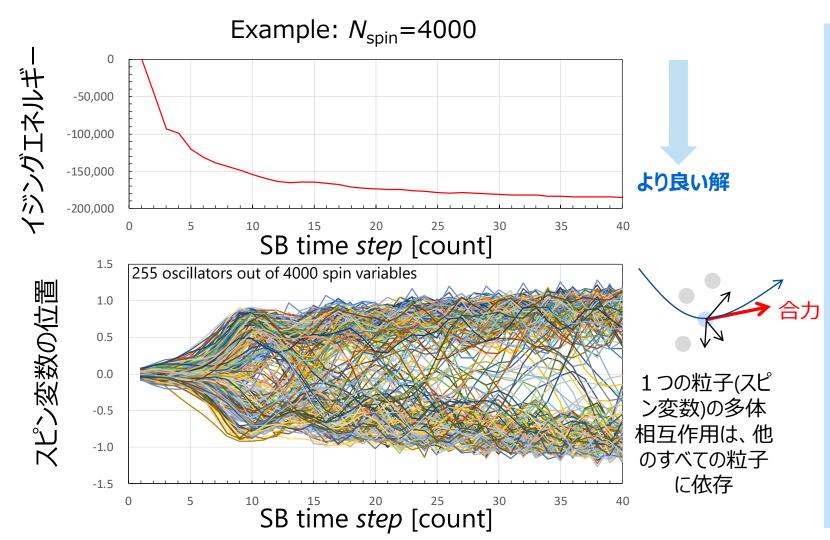
より良い極小点をより高い確率で

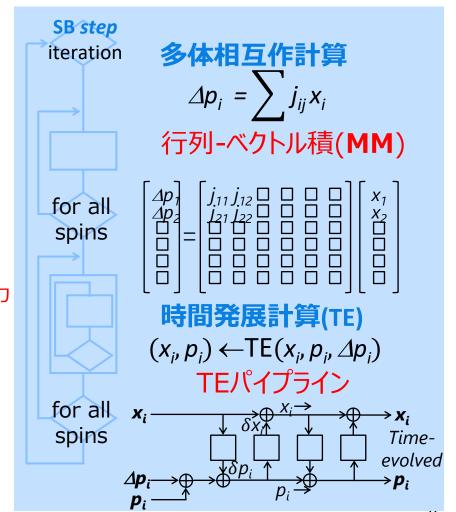
1つの極小点を追いかける

量子断熱定理に対応する古典断熱定理(多体系)の存在が示唆されるが、その数学的厳密証明は未だなされていない。 シミュレーテッド分岐アルゴリズムの理論拡張(第2世代*2、熱揺らぎアシスト法*3、高次コスト関数化*4、など)が続く。

SBアルゴリズムとその構造

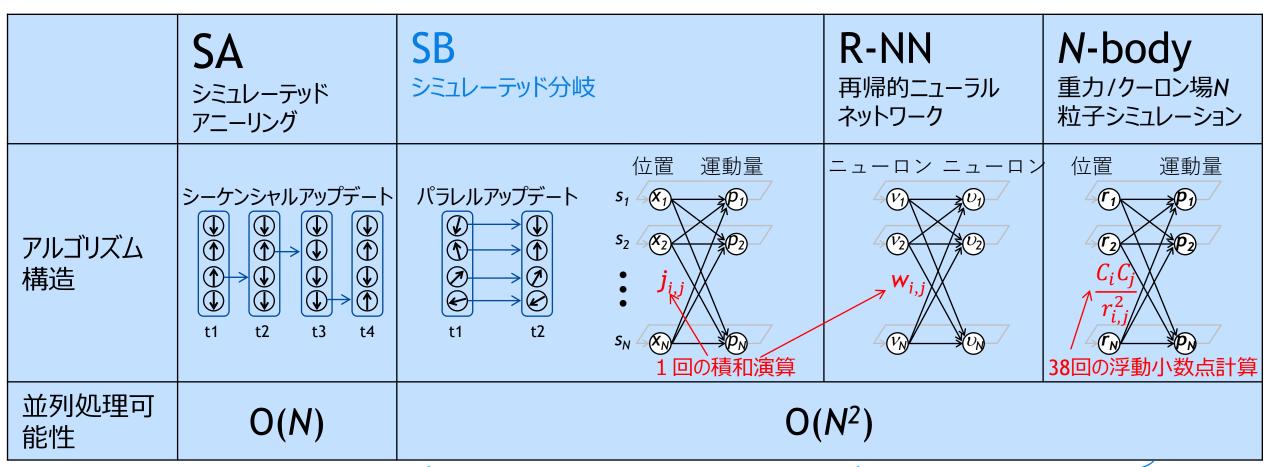
"N-body"型のアルゴリズム構造*1 → 高い並列化可能性 SAよりも並列性が高い





SBアルゴリズムの特徴と実装

SBは市販FPGA/GPU/NPUで加速可能(ASICに限定されない)*1 →SBMの高い実用性・可用性・経済合理性



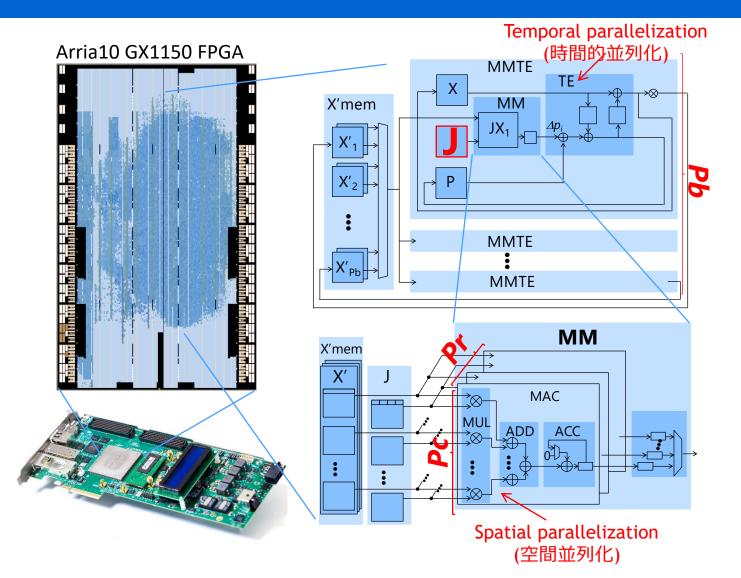
More parallelizable (より並列化可能)

高頻度なメモリアクセス J/W matrix (NxN matrix) →超並列・超高速実装のポイント とても似ている

チップ当たりより多くのPE数 PE: 一体相互作用計算のための演算要素

SBアルゴリズムのハードウェア処理(ハードワイヤ化)

FPGAベースSBM(SBアルゴリズムの超並列処理回路)*1

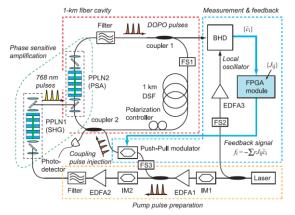


問題	complete-graph MAX-CUT	
マシンサイズ	4,096 spins	
アーキテクチャ		
並列度Pr/Pc/Pb	32/32/8	
積和PE数	8,192	
回路活性化率	92 %	
使用計算機リソース		
ALM	40%	
BRAM	56%	
DSP	7%	
クロック周波数	[MHz]	
Fsys	269	

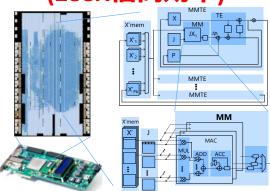
SBM(第1世代) の発表*1 (2019)

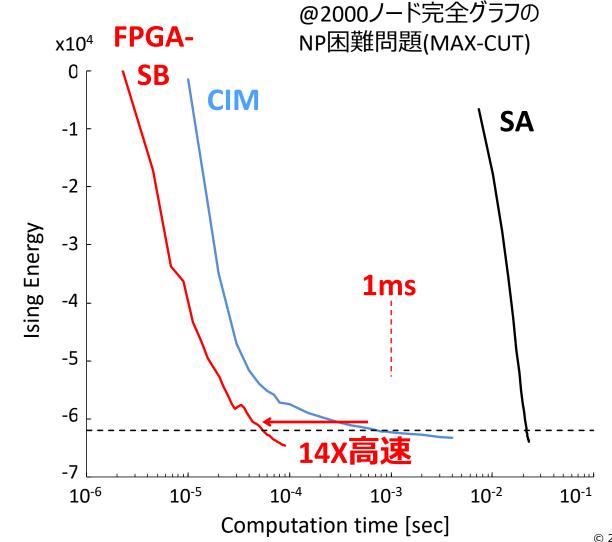
SA/先端イジングマシンより高速・高効率、高速リアルタイムシステム適合(<1ms)

Coherent Ising Machine 800 GMAC/s @ 1000 W



FPGA-SB 1,873 GMAC/s @ 49 W (288X倍高効率)



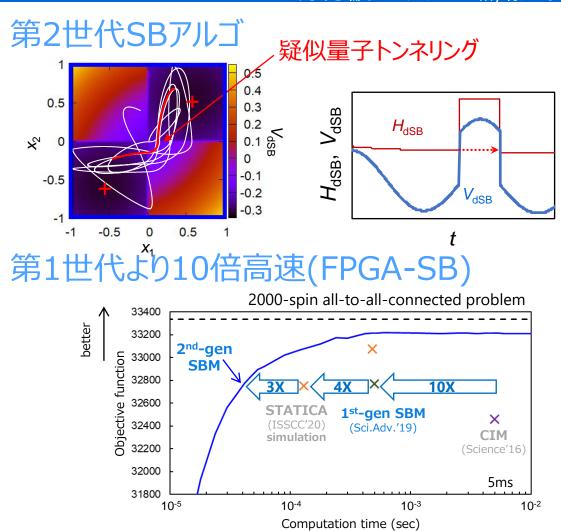


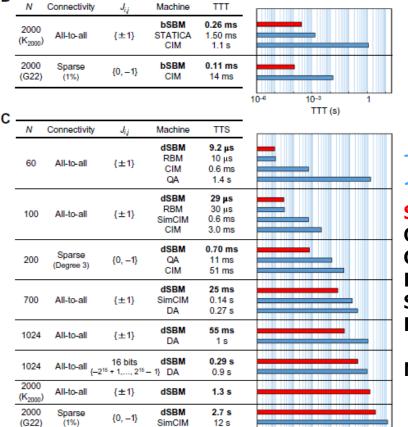


SBM(第2世代) の発表*1 (2021)

包括的な性能比較 → 強い競争力を実証

対 先端イジングマシンの群/様々な学術ベンチ問題





10⁻³

TTS (s)

TTS (Time-To-Solution)



比較対象 イジングマシン

SB: Simulated bifurcation

QA: Quantum annealer

CIM: Coherent Ising machine

DA: Digital annealer

SimCIM: Simulated CIM

RBM: Restricted Boltzmann

machine

MA: Momentum annealing

Contents

はじめに

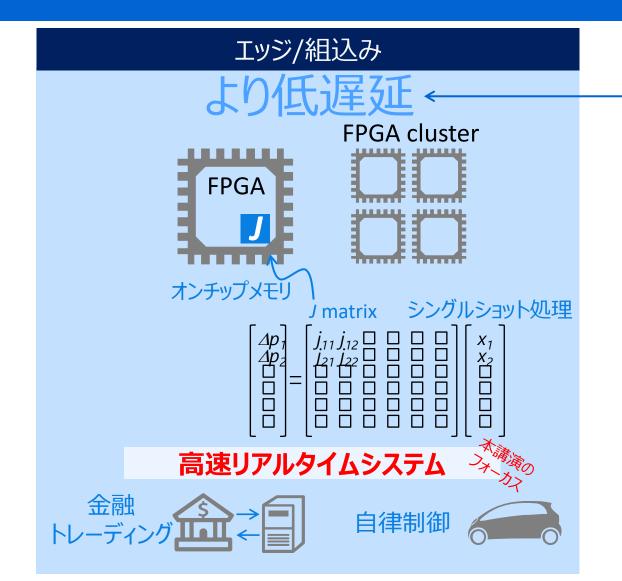
シミュレーテッド分岐マシンのリアルタイムシステムへの応用 -金融・車載の事例-

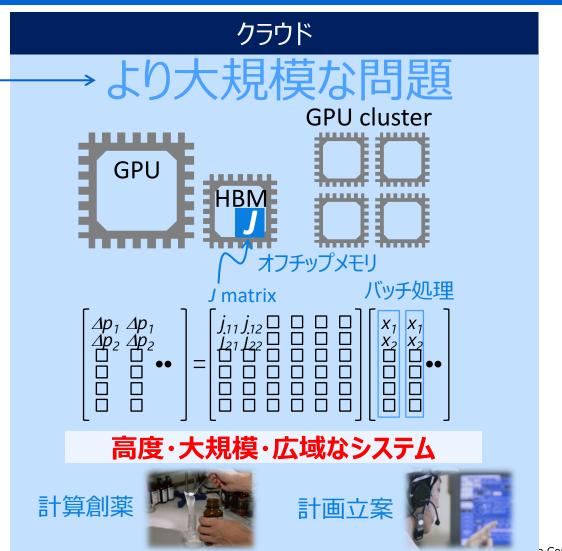
01 量子インスパイアード最適化技術をリアルタイムシステムへ応用する

02 シミュレーテッド分岐マシン(SBM)の理論と実装、性能

SBMのリアルタイムシステムへの応用:金融と車載の事例

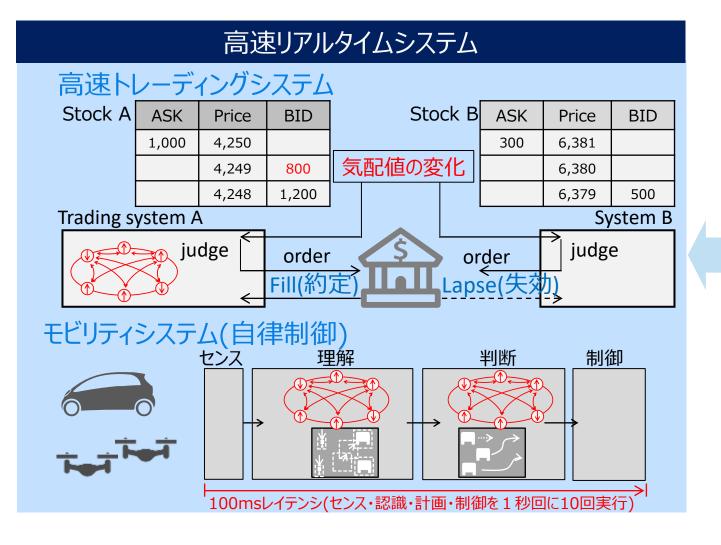
低遅延FPGA実装→エッジ/組込み、大規模GPU実装→クラウド





FPGA実装組込み型SBM (回路IP)

高速リアルタイムシステムがNP困難問題に基づく高度判断を実行可能に*1



FPGA実装組込み型SBM

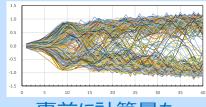
超低遅延 (サブミリ秒)

決定論的遅延



1.組込み(混載)可能

- フ_{_}カスタムI/Fを具備できる
- 3 カスタム回路(決定論的遅延、ソフトウェア割り込み無し)

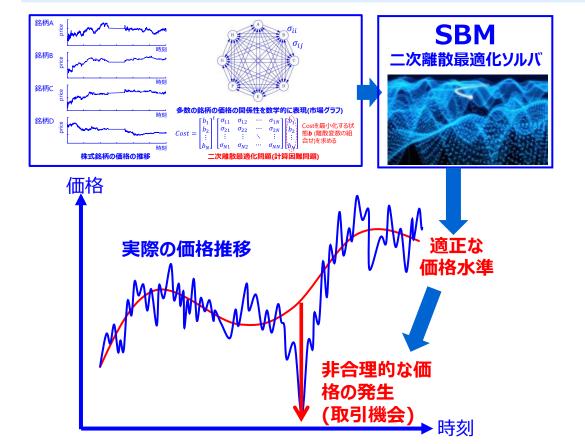


事前に計算量を 決定できる

- シン」を用いて執行するシステムを実証"
- *1 https://doi.org/10.1109/ISCAS45731.2020.9181114
- *2 https://doi.org/10.1109/ACCESS.2023.3316727
- *3 https://doi.org/10.1109/ACCESS.2023.3326816
- *4 https://doi.org/10.1109/ACCESS.2023.3341422

N銘柄の関係性の高速分析(二次離散最適化)による 取引機会の検知(適正価格推定とそれからの乖離)を特徴とする新しい戦略

どんなに優れた戦略であっても、同じ戦略を実行する競合がいれば、勝率は大きく低下する。 トレーダーは本質的にテクノロジーハングリーで、新規技術で人と違う取引機会を検知・執行したい(ハーディングの回避)。



SBMの金融応用#1

Type	高速リアルタイムトレーディング			資産運用
Target	外国為替	株式		
Strate gy	外国為替裁定 取引*1	拡張ペアトレー ディング*2	高速離散ポート フォリオ最適化*3	参ります。 タ様化ポート フォリオ運用*4
Optimi zation	任意サイズ閉 路探索	任意閉路の連続タブーサーチ	離散ポートフォリ オ最適化	最大独立集合 問題
Paper	A Currency Arbeitrage Machine, based on the Strendstool Biffurcation Appetritum for Ultrafast Description of Optimal Opportunity Control of the Control of	The second secon	BEEL ACTION THE STANDAY AND T	THE CONTROL OF THE CO
Date	Oct. 12, 2020	Sep. 18, 2023	Oct. 23, 2023	Dec. 12, 2023

高速離散ポートフォリオ最適化*1

*1 K. Tatsumura et al., "Real-time Trading System based on Selections of Potentially Profitable, Uncorrelated, and Balanced Stocks by NP-hard Combinatorial Optimization," IEEE Access 11, pp 120023 - 120033 (2023).

*2 VWAP:株式価格が売買高加重平均価格

シャープレシオ(リターン/リスク)を最大化させるため、リターン大・非相関・デル タニュートラルの複数銘柄を選択し、同時に取引する高速トレーディング戦略

取引可能 銘柄群♪





売買銘 柄群 N_b

(瞬時)期待リターン

・各銘柄iにつき、VWAP*2と株価の 乖離を瞬時期待値(ΔPi)と定義

・株式価格がVWAPに収れんする方 向の売買需要が存在することに注目



離散性

各銘柄は売買対象にするか否かの2状態。全銘柄の取 A_{trans} 引額(A_{trans})が揃うように、当日基準値(P_i^b)と最小取引 $L_i =$ $S_{min}p_i^b$ シェア(S_{min})を考慮して、銘柄毎注文ロット数(L_i)を決定

リスクの最小化

- ・選択銘柄群の相関性の合計の最小化(多様化ポートフォリオ)
- ・デルタニュートラル(買い/空売りの合計額をゼロ)

QUBO定式化

イジングビット

銘柄iを売買対象にするか否か
$$m{b} = [b_1 \quad b_2 \quad \cdots \quad b_N] \quad b_i \in \{0, 1\}$$

$$H_{QUBO} = \sum_{i} \sum_{j} Q_{i,j} = H_{cost} + H_{penalty}$$
 相関係数 瞬時期待リターン

組合せ最適化の トータルコスト関数
$$H_{QUBO} = \sum_{i}^{N} \sum_{j}^{N} Q_{i,j} = H_{cost} + H_{penalty}$$
 相関係数 瞬時期待リターン
$$A_{Cost} = \begin{bmatrix} b_1 & b_2 & \cdots & b_N \end{bmatrix} \begin{bmatrix} \Delta P_1 & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \Delta P_2 & \cdots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \Delta P_N \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix}$$

ペナルティ関数

$$H_{penalty} = c_2 \left(\left(\sum_{i}^{N} b_i \right) - N_s \right)^2 + c_3 \left(\sum_{i}^{N} sgn(\Delta p_i) b_i \right)^2$$

制約条件①:

・N銘柄中 N_h 個のビットのみ1

$$\sum_{i}^{N} b_{i} = N_{s}$$

制約条件②:

・売/買の銘柄は同数

$$\sum_{i}^{N} sgn(\Delta p_i)b_i = 0$$

高速離散ポートフォリオ最適化*1

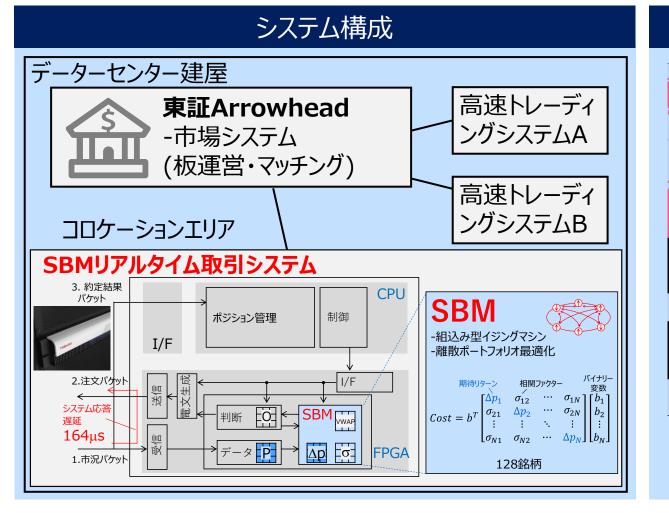
SBM高速リアルタイム取引システム(システムワイド応答時間:164us)を、 東証コロケーションエリアに設置

SBM

module

RX

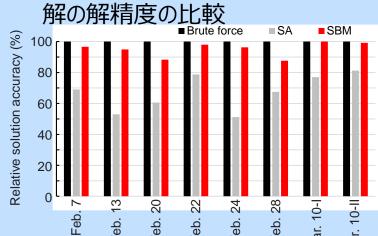
iudge



組込みSBM



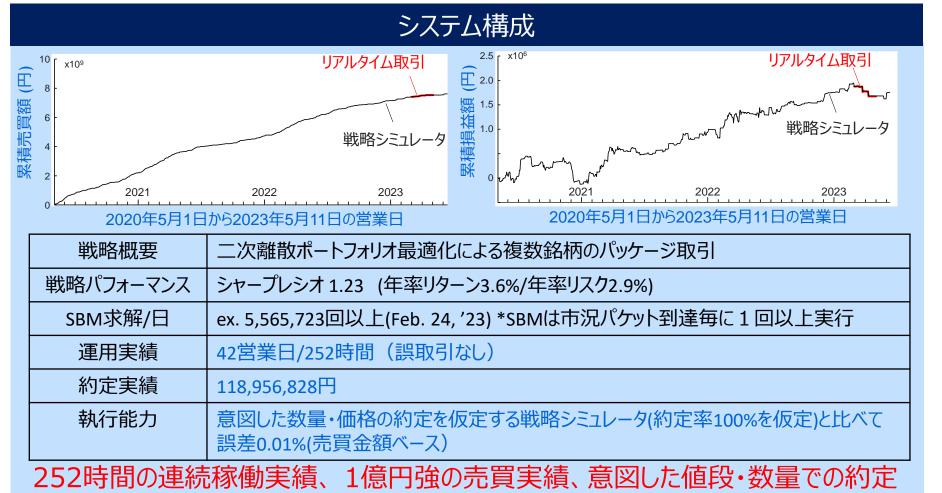
同等計算量SAとのシングルショット求解の解集度の比較



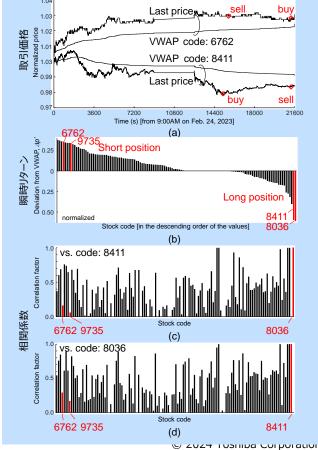
Market situations from Feb. 7, 2023 to Mar. 10, 2023

高速離散ポートフォリオ最適化*1

東京証券取引所のリアルタイム・リアルマネー取引データをもって、二次離散最適化に基づく戦略が執行可能(意図した値段・数量で約定)であることを実証



一回の取引行動(典型例)



量子インスパイアード二次離散最適化処理コプロセッサー

(組込みSBM)

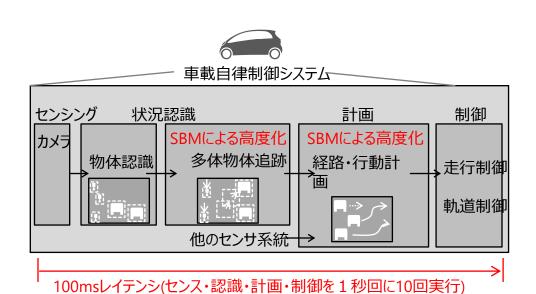
https://doi.org/10.11351/jsaeronbun.54.1216

汎用プロセッサー

周辺状況に即座に最適応答・適合する高度な自動運転/運転支援システムへ

車載自律制御は、センス・認識・計画・制御のサイクルは1秒間に10回以上繰り返す、リアルタイム・ミッションクリティカルシステム。

二次離散最適化による高度な認識・計画の機能を、厳しい時間制約のもと、組込みSBM技術で実現する。



Sensors
(Cam, LiDAR...)

CPU GPU

Quantum inspired
Processing Unit,
(QiPU)

Implementation: IP

DRAM

*NPU: Neural Processing Unit

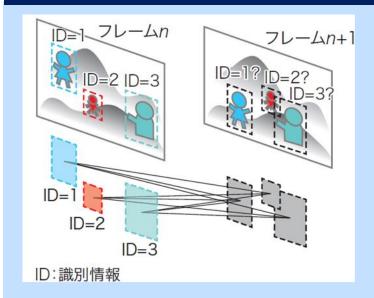
AI処理コプロセッサー

Fig.2 Architecture of automotive computing platform with

Quantum inspired Processing Unit 量子インスパイアード車載コンピュータプラットフォームの提案*1 https://doi.org/10.11351/jsaeronbun.54.1216

SBMの低遅延・広域な解探索能力を使って、従来の多体物体追跡法を、 多数・長期オクルージョンに対応できるように拡張したシステム

多体物体追跡を拡張



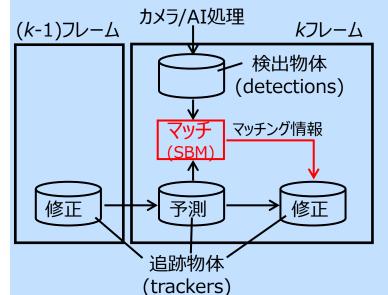
- ・従来SORT法をSBMの低遅延・広域な解探索能を用い、長期のオクルージョンに対応できるように拡張
- ・オクルージョン:追跡物体が他の物体などによって一時的に隠される現象を指す

QUBO定式化のポイント

- ・検出物対追跡物のマッチング処理をQUBOとして組込みSBMで解く
- ・ダブルマッチ禁止のペナルティ項のウェイト(c)を大小2通りで2回SBMを実

行し、オクルージョンの発生の有無および、

オクルージョン発生個所を特定



 $H_{QUBO} = \sum_{i}^{N} \sum_{j}^{N} O$ 重なり最大化 ングの禁止 $Q_{i,j} = H_{cost} + cH_{penalty}$

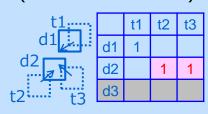
検出物と追跡物

オクルージョン無し (#検出物=#追跡物)



一対一マッチング

オクルージョン有り (#検出物 **<**#追跡物)



一対多マッチング を柔軟に許す

ダブルマッチ

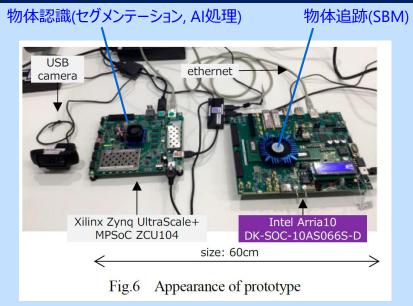
SBM特有機能を持つ多体物体追跡*1

*1 大矢/藤本(ミライズテクノロジーズ), 濱川/山崎/辰村(東芝), "量子インスパイアード車載プラットフォームの提案と試作", 自動車技術会論文集 54巻6号, pp. 1216-1221 (2023).

https://doi.org/10.11351/jsaeronbun.54.1216

車載搭載可能ボードを使って、SBM特有オクルージョン対策機能を持つ 多体物体追跡システム(~20FPSのスループット)を実証

車載搭載可能ボードへの実装



SBアルゴ	弾道SB (bSB)		
マシンサイズ	全結合512スピン		
MAC-PE数	2,048		
1回のマッチング	284µ秒		

スループット(>10FPS)

Table 3. Tracking performance of SORT with matching methods of Hungarian and SBM on MOT benchmark sequences (19)

MOT algorithm	Matching	MOTA [↑]	HOTA [↑]
Original SORT	Hungarian	48.77	44.80
Original SORT	SBM	48.76	44.83



Rectangles indicates Detection result, and Vectors in Rectangles indicates Tracking result.

*This image is processed for privacy protection.

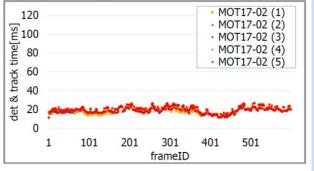


Fig.7 Evaluation result of processing time of prototype

- ・通常時追跡精度(w/o長期オクルージョン)は、SBM拡張法と従来法で同等
- ・システムスループットは20FPS程度

SBM特有機能

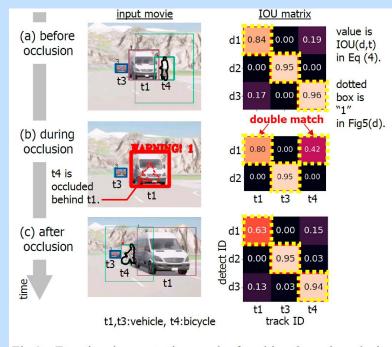


Fig.8 Function demonstration result of tracking through occlusion

・SBMの全解空間探索機能を活用したフレキ シブルマッチングよって、多数・長期オクルージョ ンに対応

シミュレーテッド分岐マシンのリアルタイムシステムへの応用 -金融・車載の事例-

シミュレーテッド分岐マシン(SBM)

- ・SBアルゴリズムは、量子インスパイアード、超並列処理可能な二次(高次)離散最適化のためのヒューリスティック
- ・SBアルゴリズムは、市販FPGA/GPU/NPUで加速可能 → 高い実用性・経済合理性
- ・FPGA実装組込み型SBMを、他の機能モジュールと混載することで、高速リアルタイムシステムを構築できる

二次離散最適化に基づく高度な判断を実行する高速リアルタイムシステム

- ・応答時間の制約のもと、これまで不可能だったより合理的な判断をするシステムを構築できる
 - →テクノロジーハングリーな金融・モビリティなどの分野で、排他的優位性をもたらしうる
- ・金融:二次離散最適化に基づく戦略が執行可能であること(応答速度の観点で、意図した値段・数量で約定できる)を、東京証券取引所のリアルタイム・リアルマネー取引データをもって実証
 - →これまで対象とされていない取引機会の検知および執行を可能に
- ・車載:イジングマシンの全解空間探索機能を活用したフレキシブルマッチングよって、多数・長期オクルージョンに対応できる物体追跡システムを、車載搭載可能ボードを使って実証(AIとの協調動作も鍵)
 - →二次離散最適化やAI処理のコプロセッサーを搭載する次世代車載プラットフォームへ

TDSL SQBM+™ オフィシャルサイト

https://www.global.toshiba/jp/products-solutions/ai-iot/sbm.html

SQBM+TM:シミュレーテッド分岐マシンを核とする量子インスパイアード最適化ソリューション(商用サービス名)

シミュレーテッド分岐マシン(SBM):学術名

「SQBM+」で検索

論文リンク、報道発表リスト、マニュアル(クラウド版、組込み版)

