

HPC による計算社会科学

産業技術総合研究所
野田五十樹

計算社会科学とHPC

●マルチエージェント社会シミュレーションを中心として (人工知能アプローチ)

▶ 事例

- 交通シミュレーション
- 人流シミュレーション
- 市場シミュレーション

▶ 計算社会科学で必要なもの

まとめ

●計算社会科学をより実用的にするには、

- ▶ 非常に多数の試行を効率的かつ気軽に実施。
 - 多岐にわたる条件を網羅する必要。
 - 繰り返しデータ同化・パラメータ探索の必要性。
 - 真の人のモデル・社会の初期条件の不可知性を補完。
- ▶ 離散データ、構造化データをnativeでサポート。
 - 表現の柔軟性。
 - データとしてのプログラム。

●次世代HPCに望まれる要件

- ▶ 数値計算偏重でない開発環境。
 - 多様なプログラム言語サポート。
 - 通常のPCとのシームレスなエコシステム。
- ▶ 柔軟なジョブ管理。
 - ブロック単位でのジョブ割り付けを柔軟化。
 - ジョブ実行の仮想化？

交通 シミュレーション

オンデマンド交通サービス

●Smart Access Vehicle(SAV) System

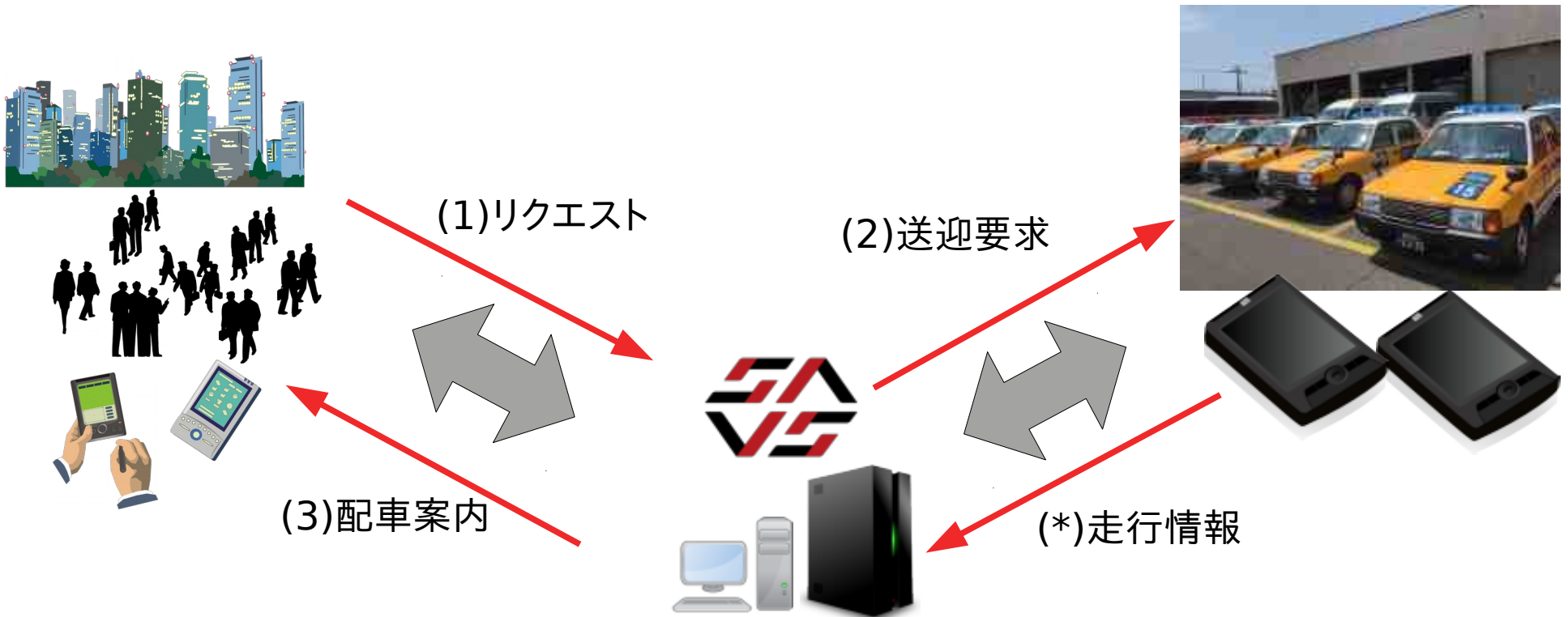
- ▶ 主目的: 都市生活者へのモビリティの提供
 - 公共交通の充実
- ▶ 特徴
 - バス+タクシー(乗り合いタクシー or オンデマンドバス)
 - 全自動
 - シミュレーションと連動
- ▶ 共同研究
 - はこだて未来大学
 - 産業技術総合研究所
 - 名古屋大学



Smart Access Vehicle Service

システム概要

- (1) リクエスト (乗降地・希望時刻・人数)
- (2) 送迎要求 (経由点リスト、利用者リスト)
- (3) 配車案内 (予定時刻・号車)
- (*) 走行情報 (現在地等)



解くべき問題

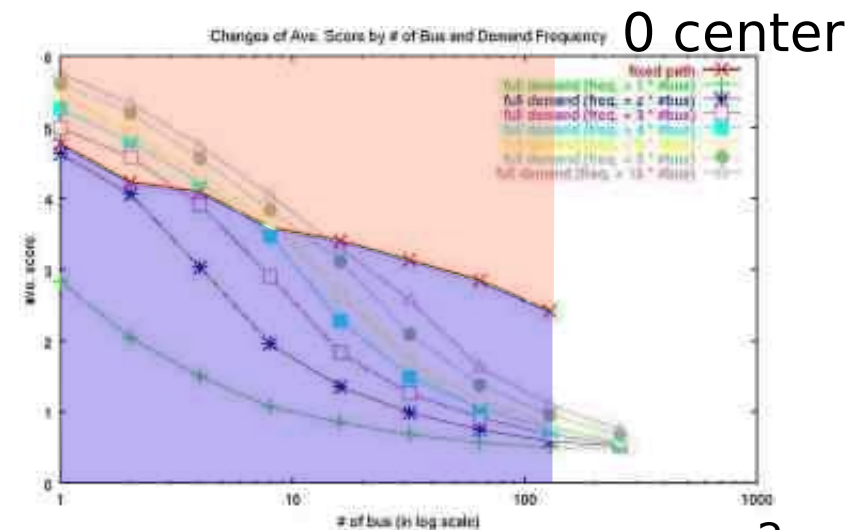
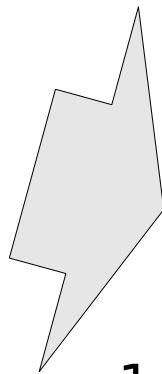
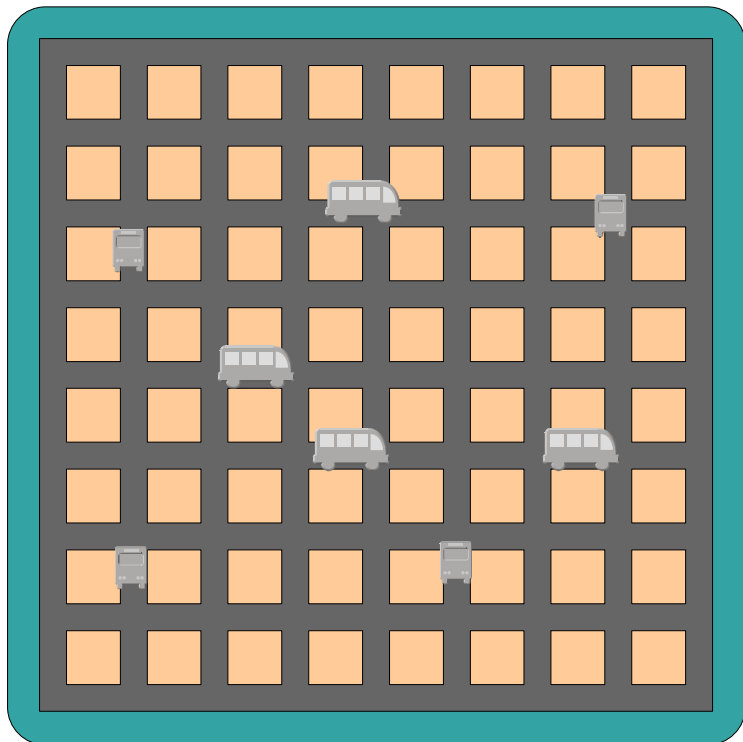
- そもそも、オンデマンド型交通サービスは成り立つのか？
 - ▶ どのような場合にメリットがあるのか？
- 最適配車？
- メリットを最大化する条件は？
 - ▶ 何台、運行すべきか？
 - 採算性とサービスの質の関係は？
 - ▶ どれくらいの範囲をカバーすべきか？
 - どれくらいの地域と人を対象とするのか？

計算社会科学での共通の視点

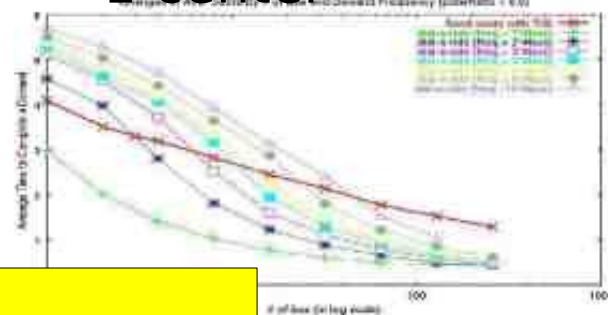
On-demand Bus Simulation

● Comparison of fixed-route and on-demand bus systems

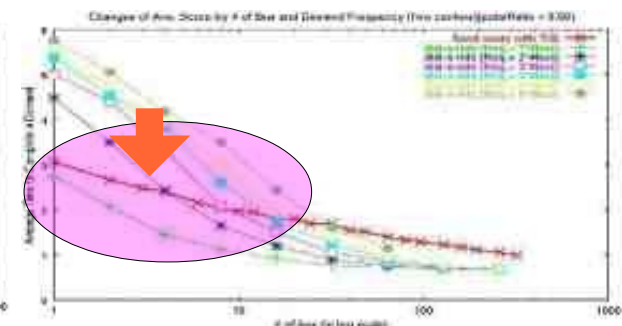
- ▶ scalability
- ▶ city types



1 center



2 centers

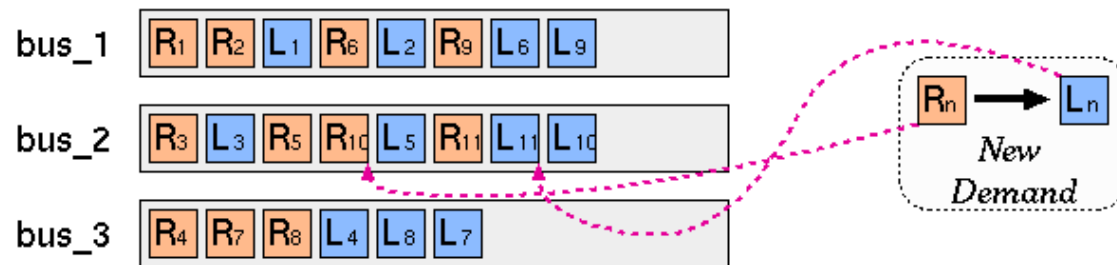


Scale Merit of On-demand Bus

Comparison under Optimum Operation

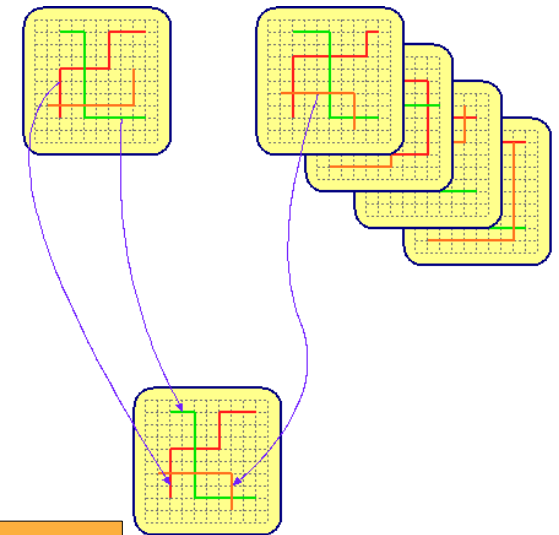
● on-demand bus

- ▶ “successive best insertion”
 - to find the best bus assignment by auction



● fixed-route bus

- ▶ genetic algorithm
 - to find the best route assignments

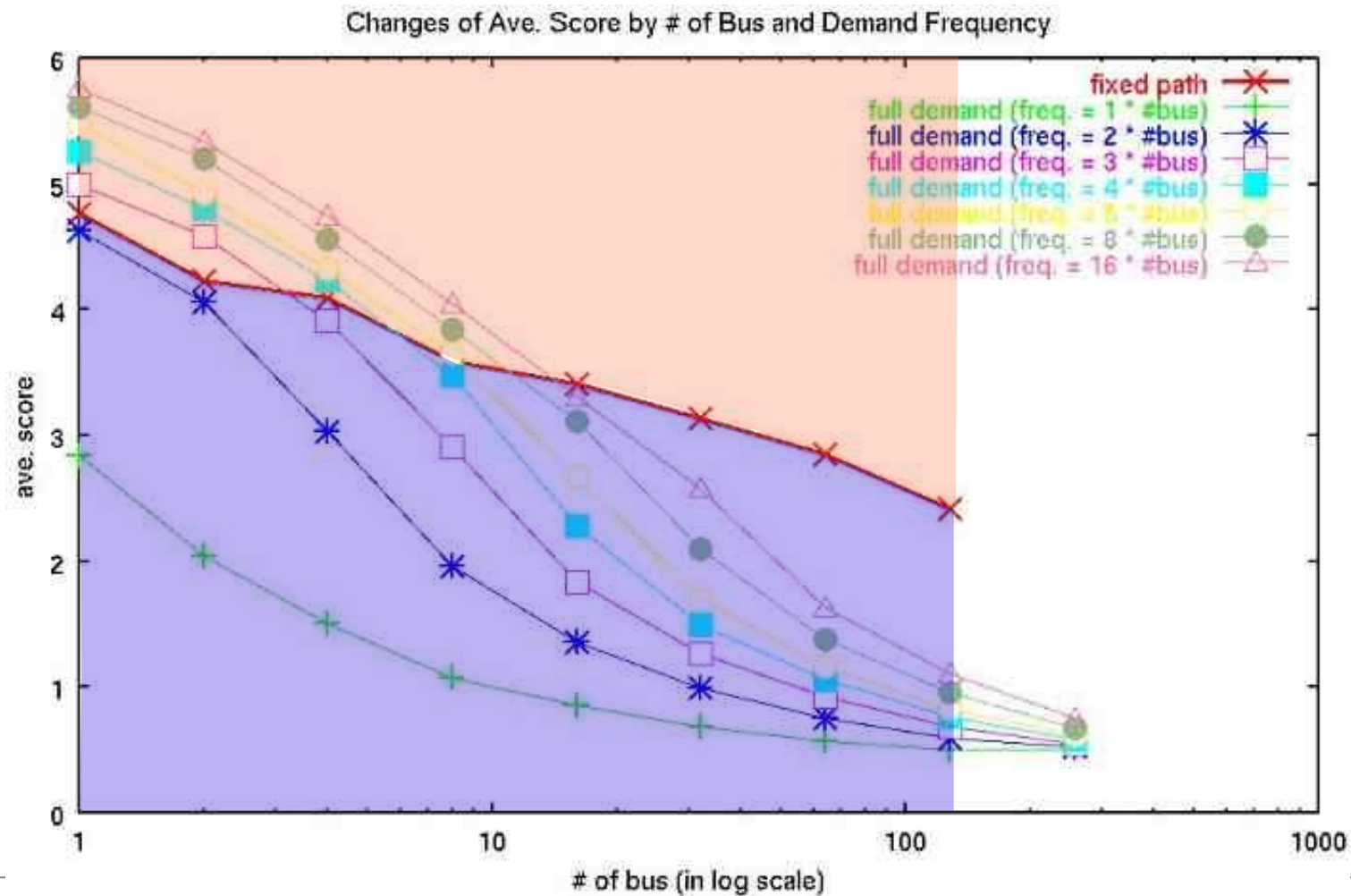


Compare travel time under

- same number of bus
- same number of demand

結果1

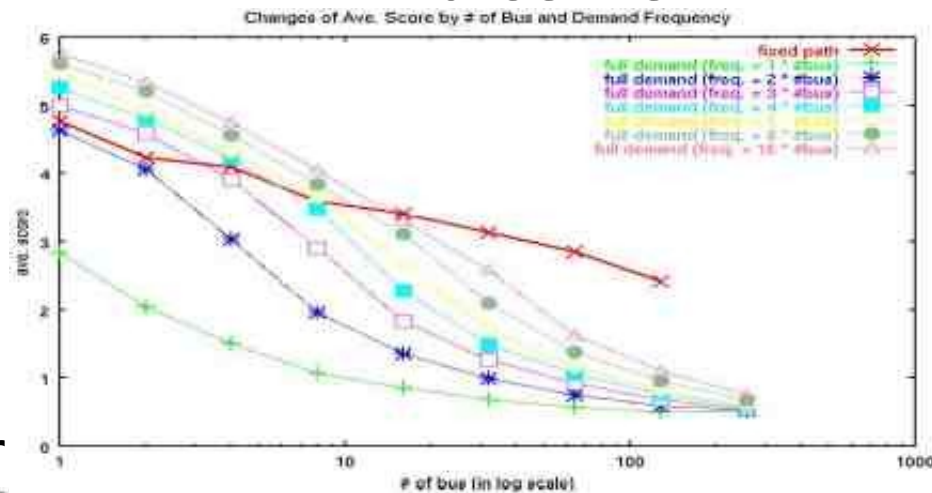
●利用者数に応じてバス台数を増減



結果2

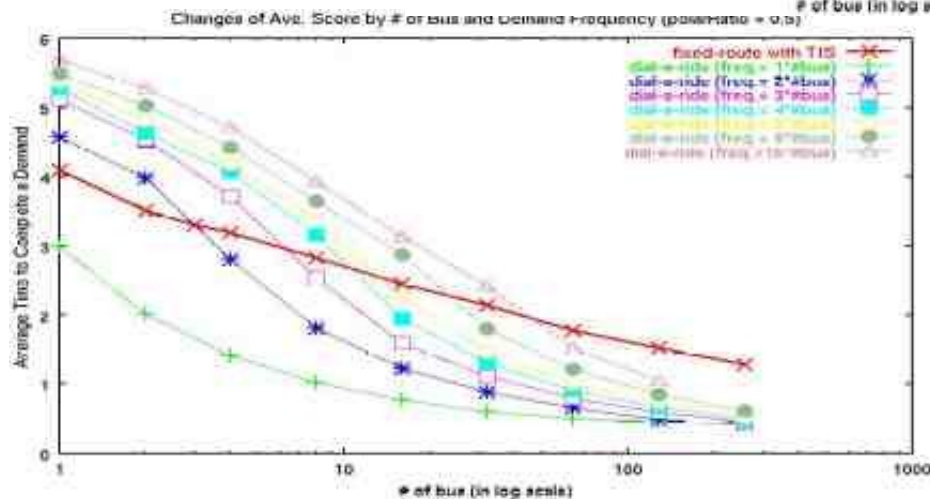
●一極集中・二極集中の場合

0 center

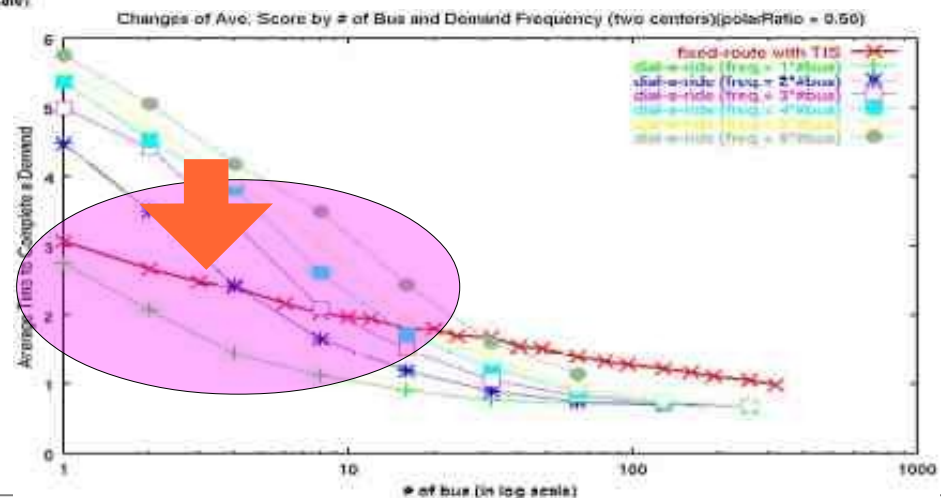


- 条件により得手・不得手がある。
- その条件は必ずしも自明ではない。

1 center



2 centers

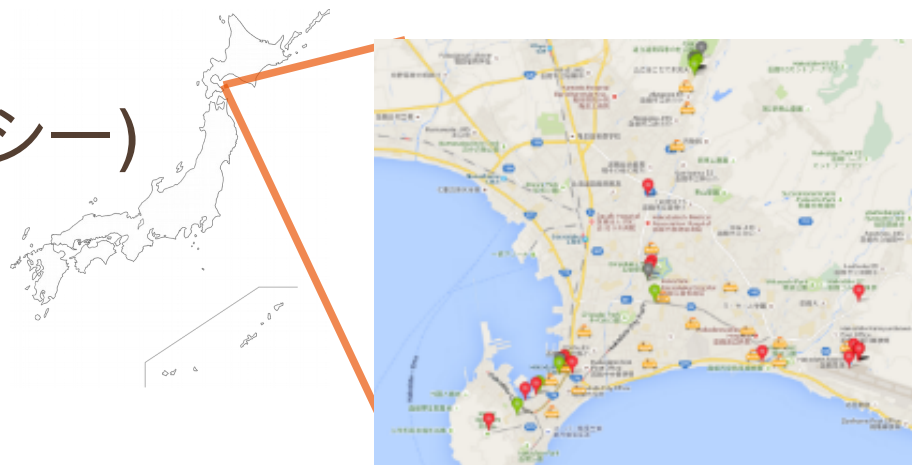


函館スマートシティプロジェクト

はこだて未来大・産総研・名大

●SAVS (Smart Access Vehicle Service)

- ▶ オンデマンドバス (乗り合いタクシー)
- ▶ サービス連携も可能
 - ◆ 病院
 - ◆ ショッピングモール
 - ◆ 観光



●シミュレーションによる利便性総合評価

- ▶ 多様な状況での評価
- ▶ 普及過程も評価



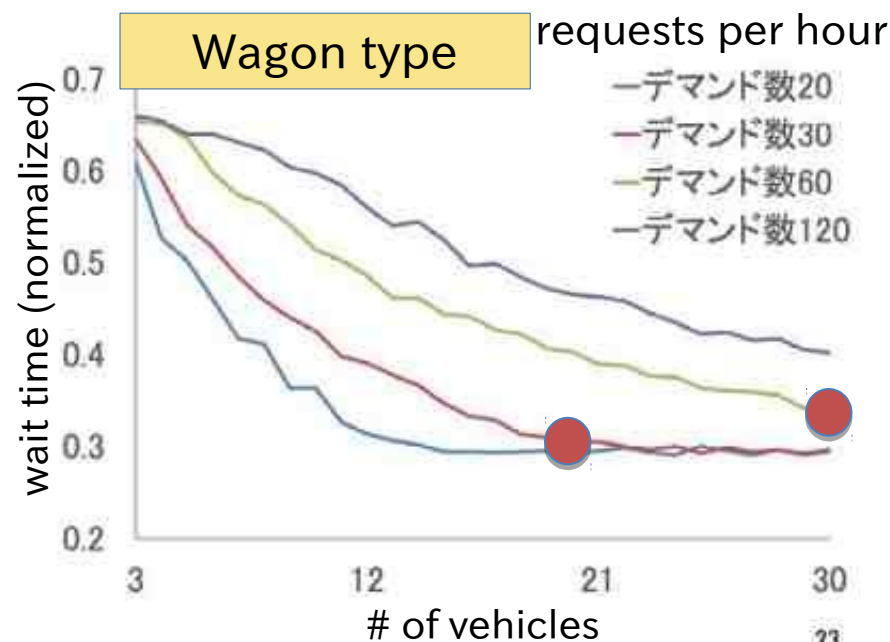
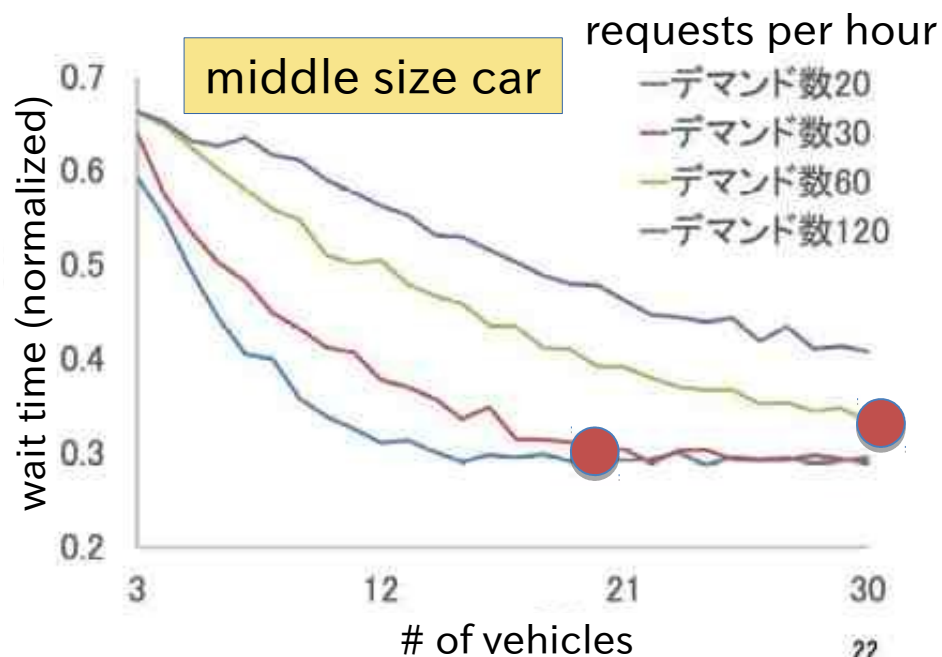
シミュレーション評価と運用規模の決定

●人工知能学会@函館にむけて

▶ 1000人の参加者に何台必要か？

■ いろいろな条件でシミュレーション評価（待ち時間評価）

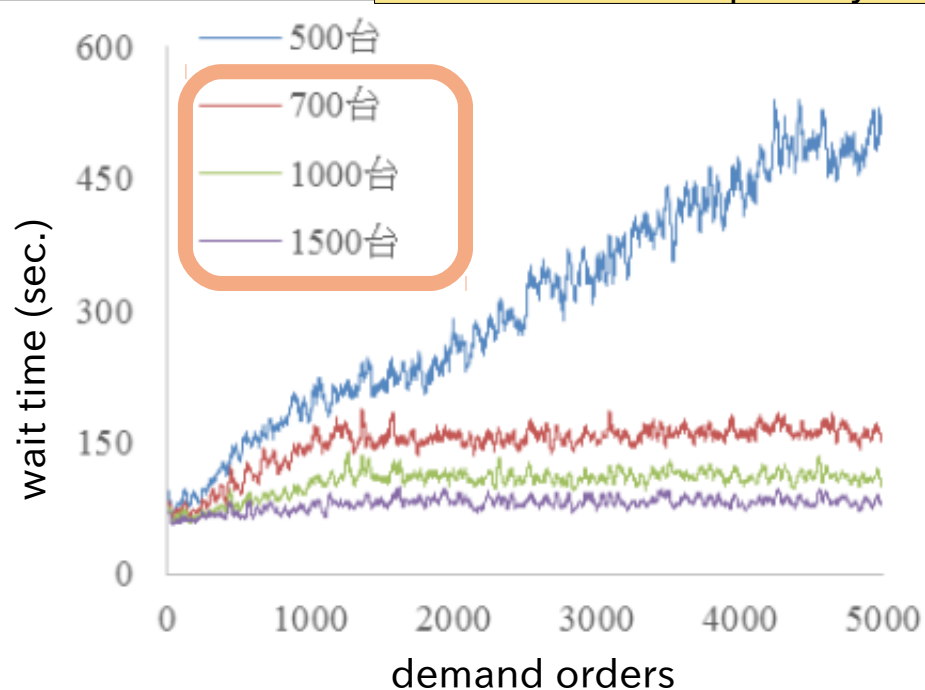
- ◆ 結果：通常日で20台、繁忙日で30台。
- ◆ 実データと比較：ほぼシミュレーションと同等の結果。



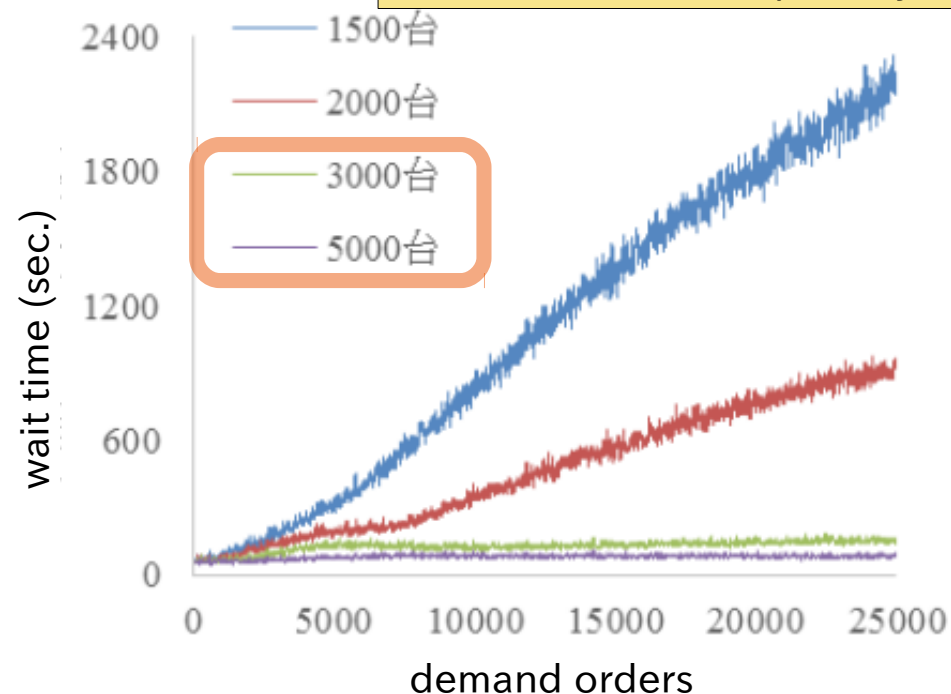
市民全員がSAVSを使うと？

- 全函館市民が自家用車を捨て、SAVSを利用すると、何台必要か？

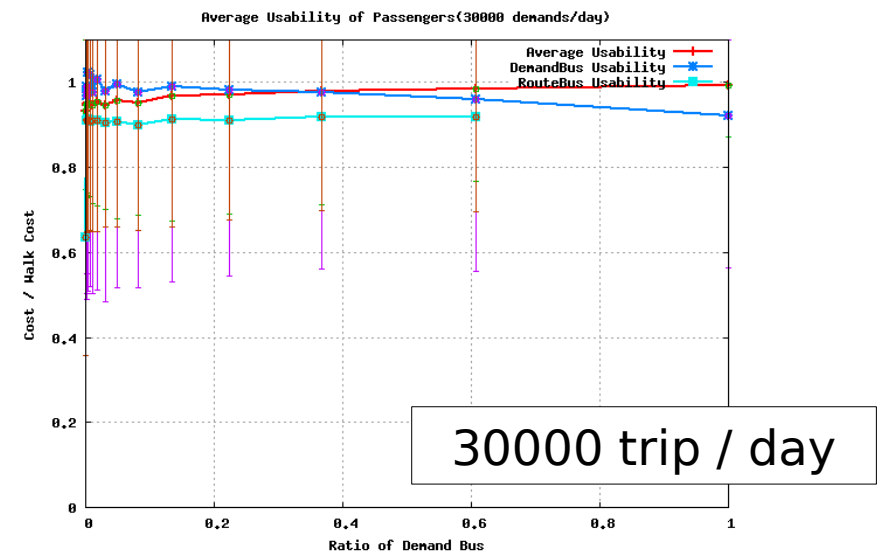
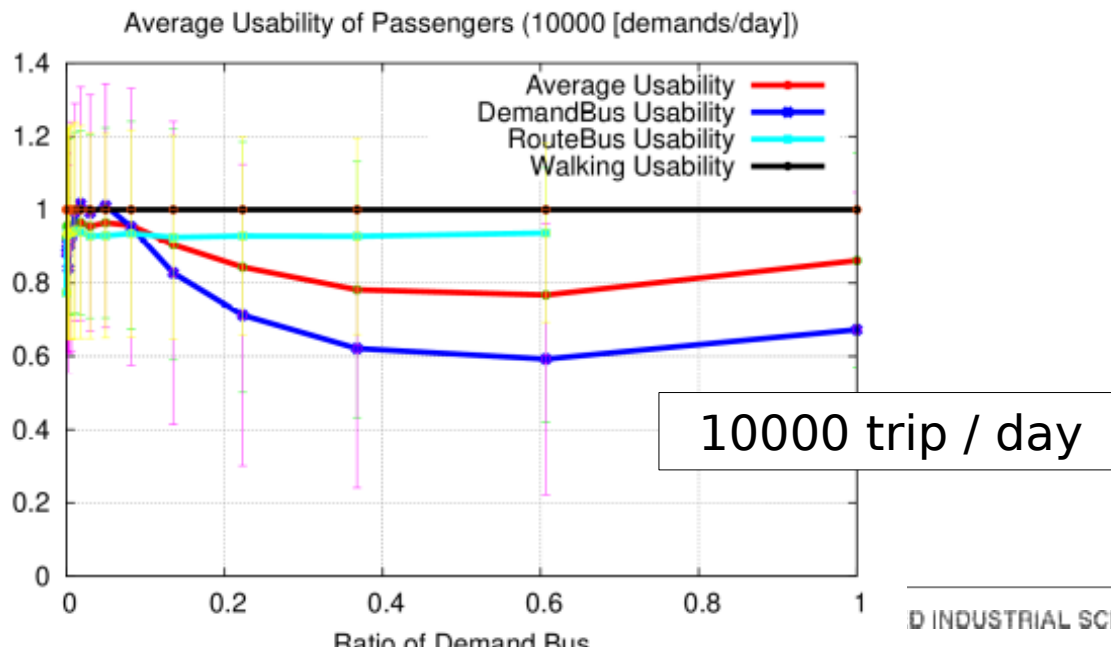
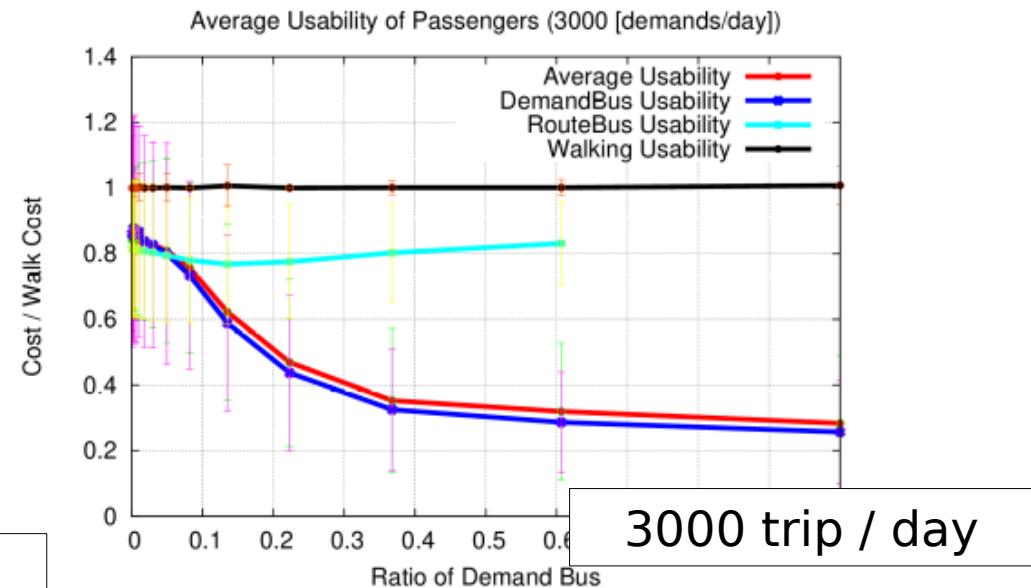
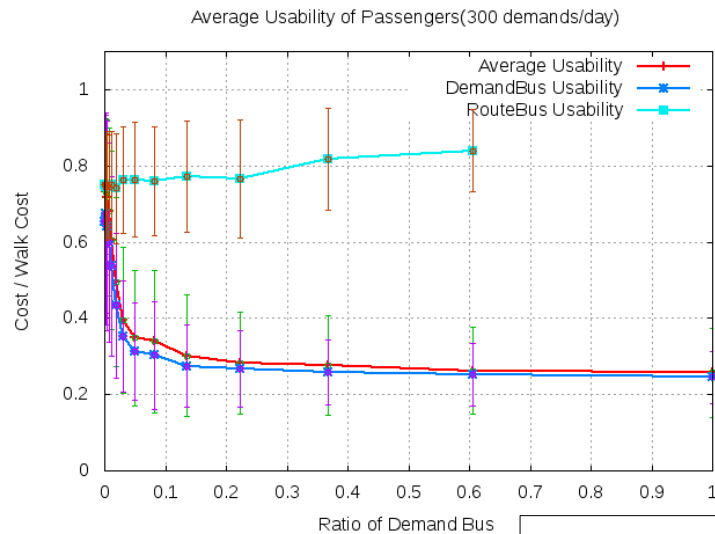
5000 demands per day



25000 demands per day



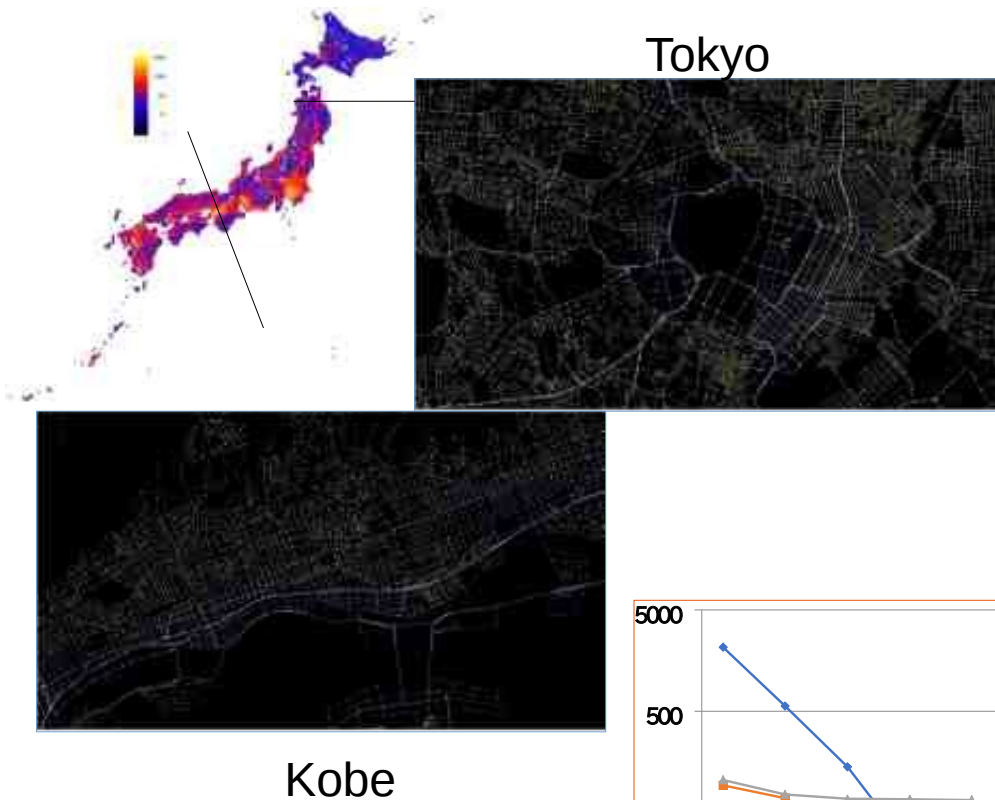
漸近的普及と利便性



Whole Japan · Whole World Traffic Simulation

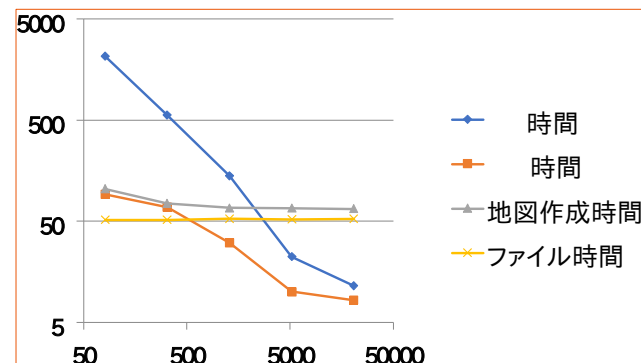
●Whole Japan

- 8 millions links, 12 millions vehicles



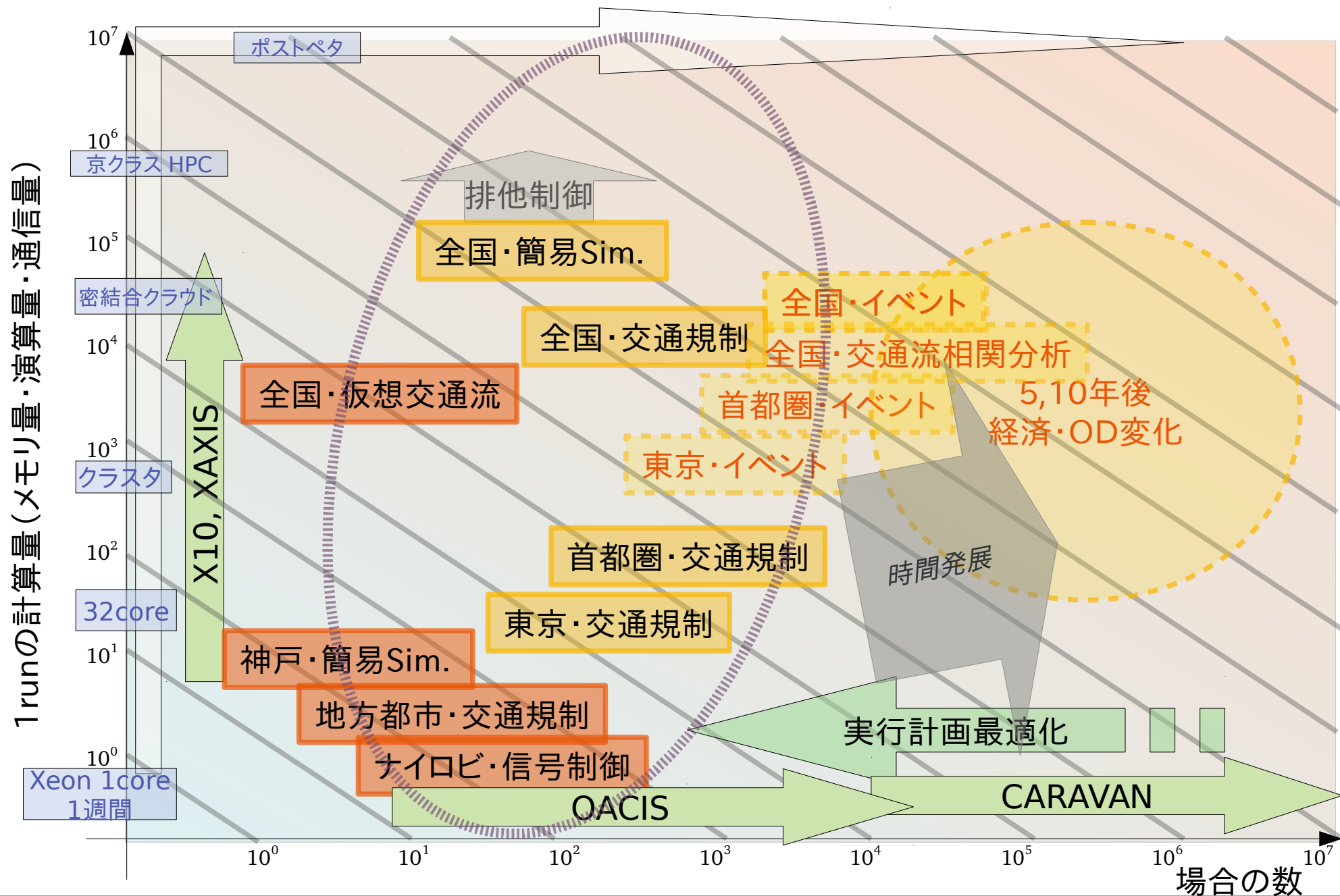
●Whole World

- 100 millions links, 100 millions vehicles
- real time simulation
 - using 20K nodes of K-Computer
 - no intelligence in agents



Performance of Parallelism

計算量ロードマップ(交通制御・分析)



経済 シミュレーション

Large Scale Market Simulation using K-comp. and X10

[東大 和泉潔先生]

和泉 潔

TOKYO
STOCK EXCHANGE

**HPC CHALLENGE:
SIMULATION**



+



Final Target: simulation of whole Tokyo Stock Market (2-3 K stocks, options)

APPLICATION FOR MARKET DESIGN

- Markets show complex dependency
- Hard to do experiments in real markets
→ *Computer simulation approach*

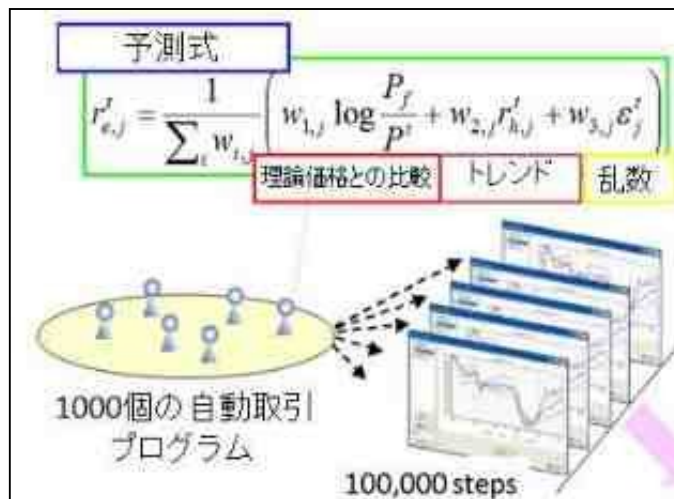
REAL EXAMPLES

- NASDAQ: new ticksize (Darley & Outkin 2007)
- Tokyo SE: new ticksize

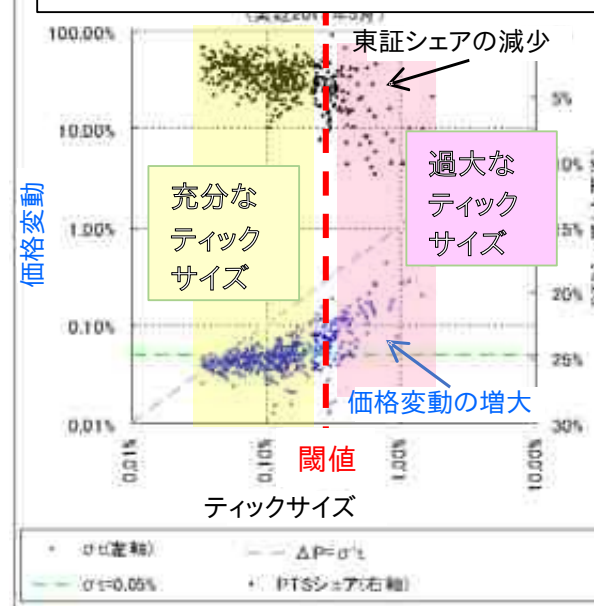
市場シミュレーションによる市場システム評価

市場間競争において、マーケットシェアの維持に必要な tic サイズを、マルチエージェント市場シミュレーションの網羅実行により同定。

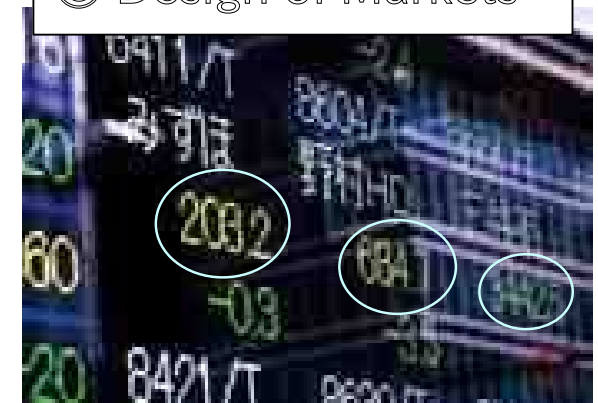
① Simulation



② Evaluation by Real Data



③ Design of Markets

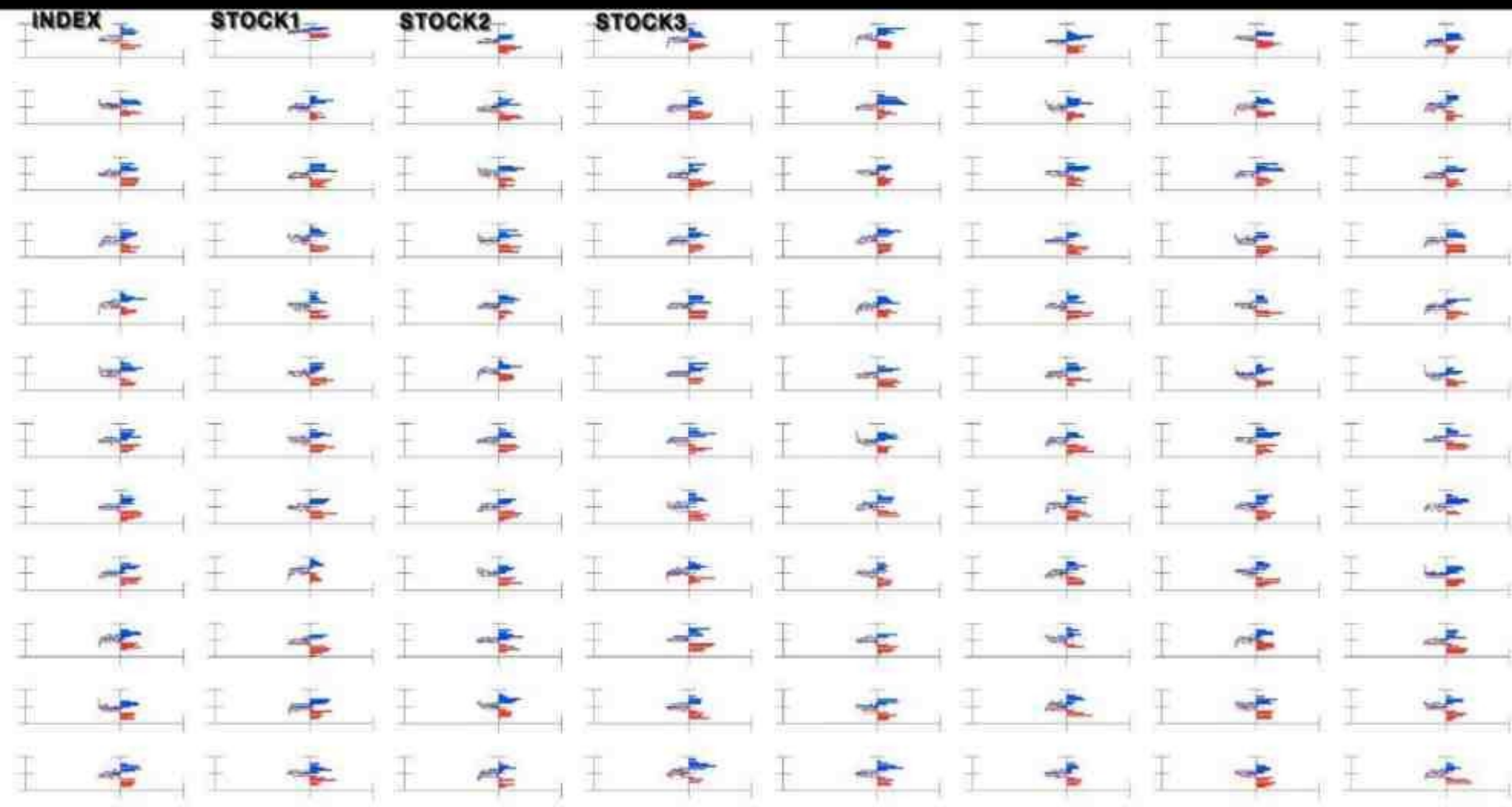


In 2014, different tick sizes are introduced based on our simulation results

分析結果は、東証のシステム改革の判断材料として活用。

Market Simulation of Flush Crash.

How the shock on stock-1 spread to whole markets?



Pham: Platform for large-scale and high-frequency artificial market

Background

New trading technologies may fluctuate financial markets.

Flash Crash



- High-frequency trading (HFT)
- Portfolio trading
- Financial derivatives



if(p) buy()
else sell()

Artificial market simulator

① handle orders(List[Order])

ORIGINAL, FULL INFORMATION

Market Market Market Market



Information
may not be latest;
may be delayed

Update Info
Orders

Update Info
Orders

COPIED, PARTIAL INFORMATION



Agents HFT agents

② place orders(List[Market])

COPIED, PARTIAL INFORMATION

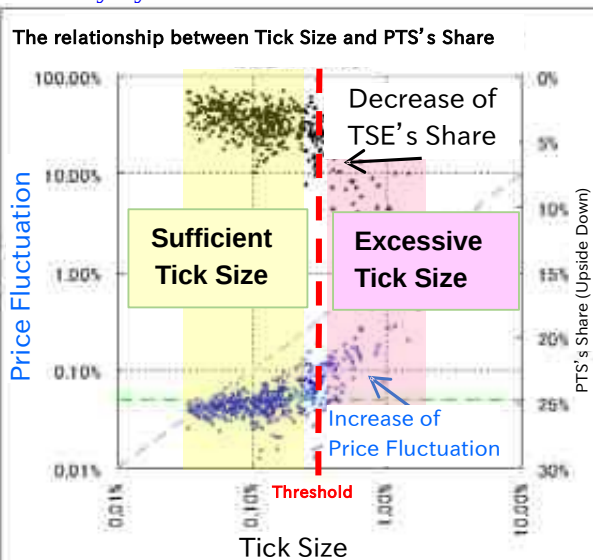


HFT agents Agents

② place orders(List[Market])

Result 1: Tick size reduction

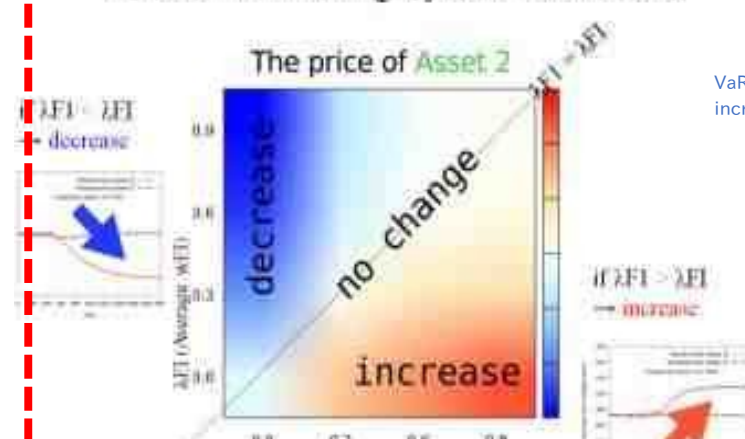
We found a threshold of the tick size of market stability by simulation.



Result 2: Shock transfer by HFT arbitrage

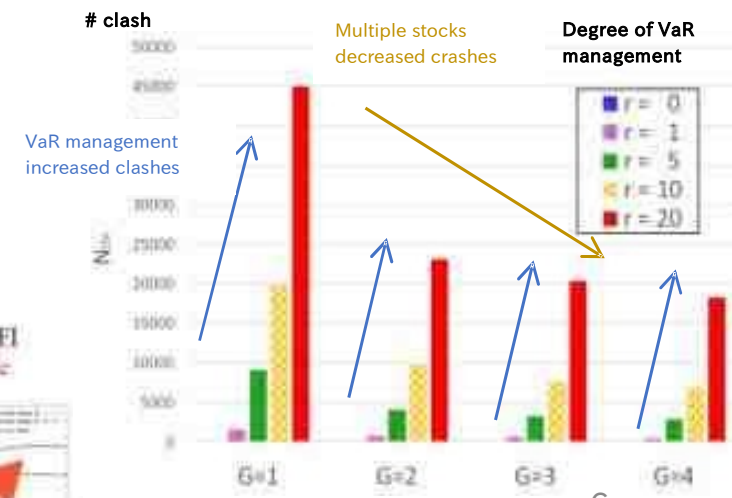
The impact of shock transfer is determined by the local traders' reaction speed to the shock

Effects from trading styles of other assets



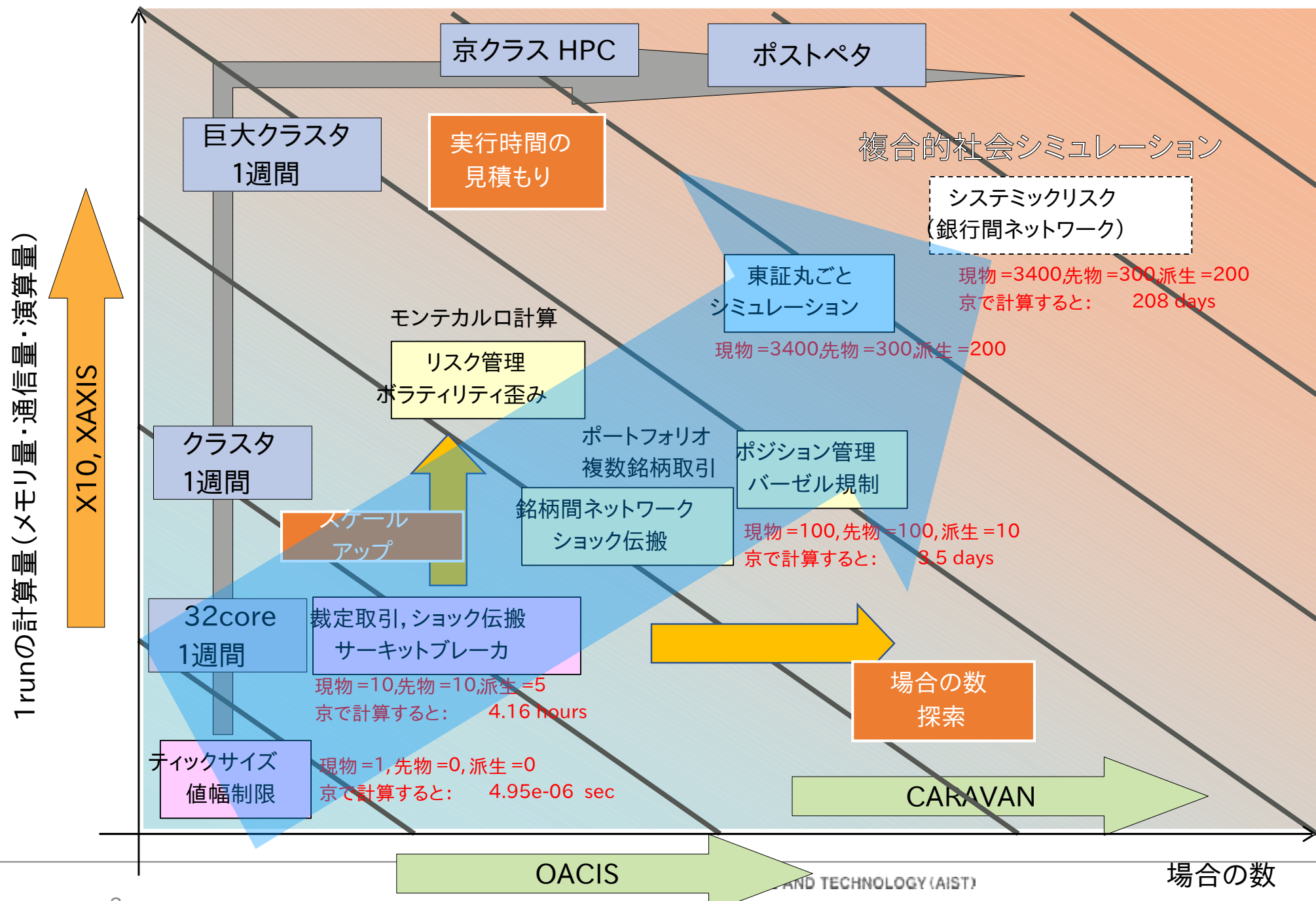
Result 3: VaR shock simulation

Introduction of traders with Value-at-Risk management may destabilize a market



ロードマップ

(時系列的な場合の数)



人流 シミュレーション

大阪西淀川区津波避難

(ネットワークフロー最大化(藤沢チーム)との比較を目指して。)

●スケール

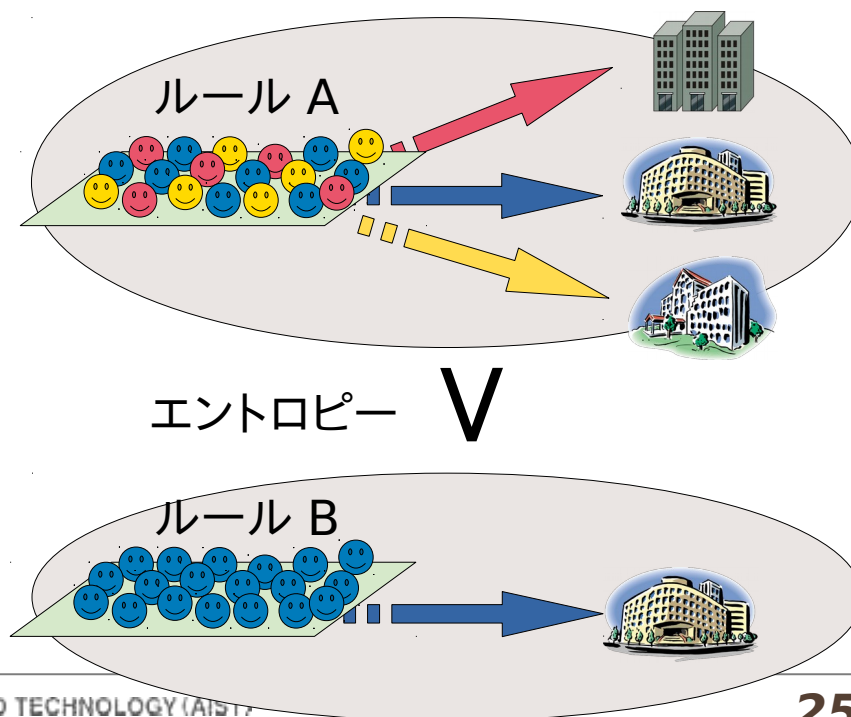
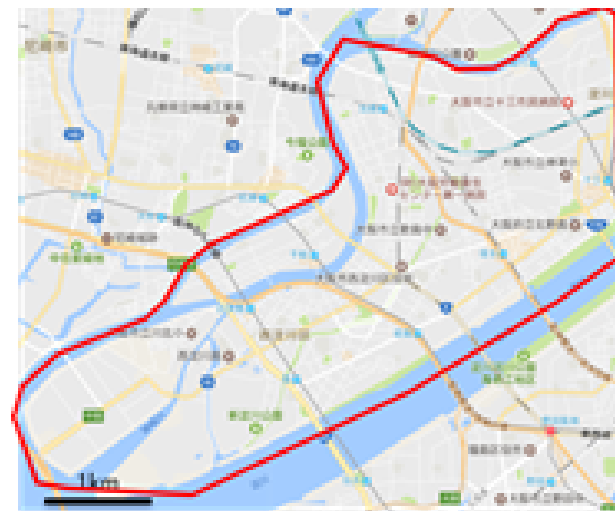
- ▶ 地図: 2933ノード、8924リンク
- ▶ 避難者: 49276人
- ▶ OD: 73の町丁目から経由点(533箇所から1つ)を経て86か所の避難場所へ。
 - 1町丁目につき、2つのグループに分割。

●全パラメータ数: 365

- ▶ 探索空間は $R^{73} \times 533^{146} \times 86^{146}$

●評価軸:

- ▶ 避難時間
- ▶ 避難誘導ルールの複雑さ
 - = 同じ地区の逃げ方のバリエーションの多さ
 - = 同じ地区からの異なる逃げ方の人数比の情報エントロピー



改良型NSGA-IIモジュール

●NSGA-II:

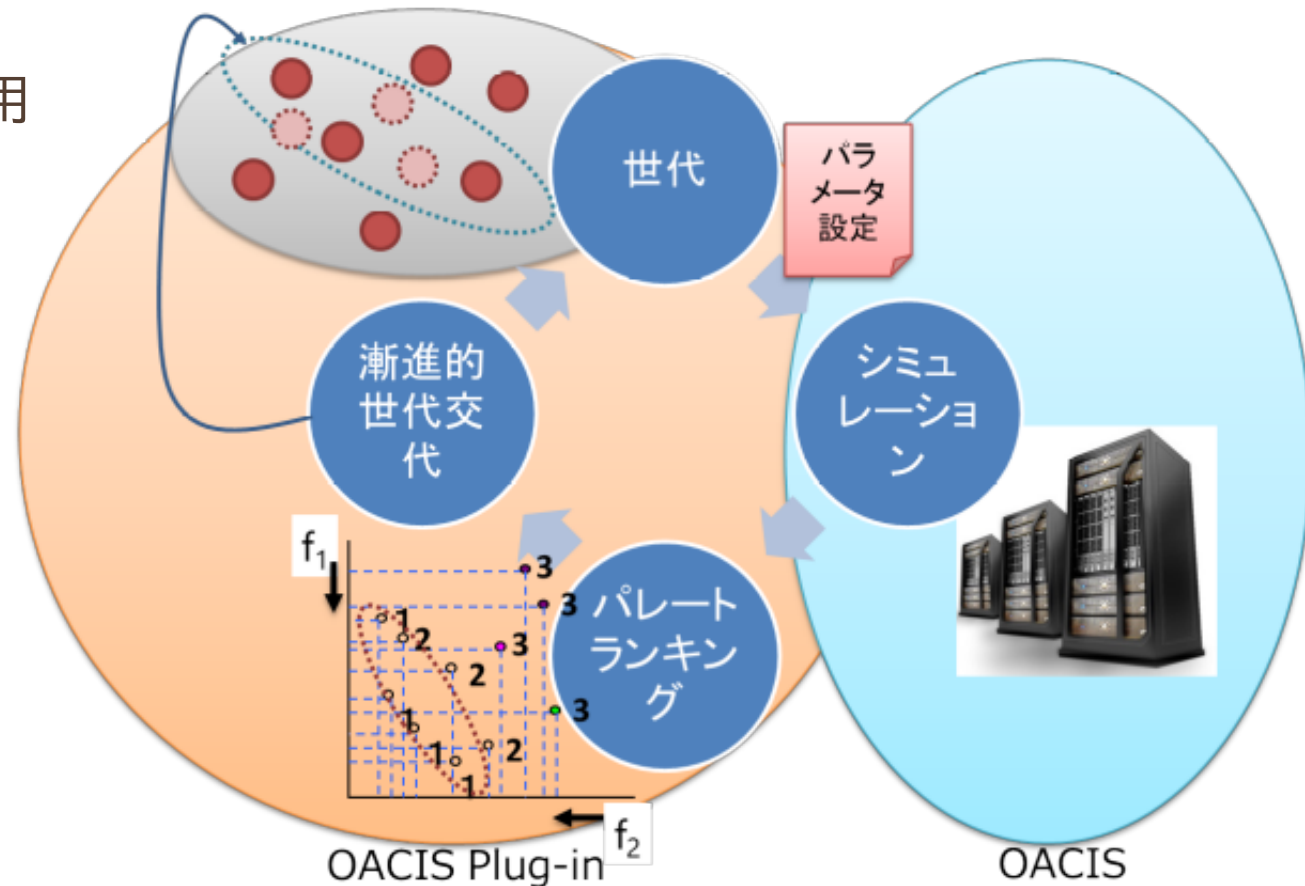
▶ 遺伝的アルゴリズムによる多目的最適化手法

■ パレートランキング

■ 混雑度による子孫選択

●部分的世代交代:

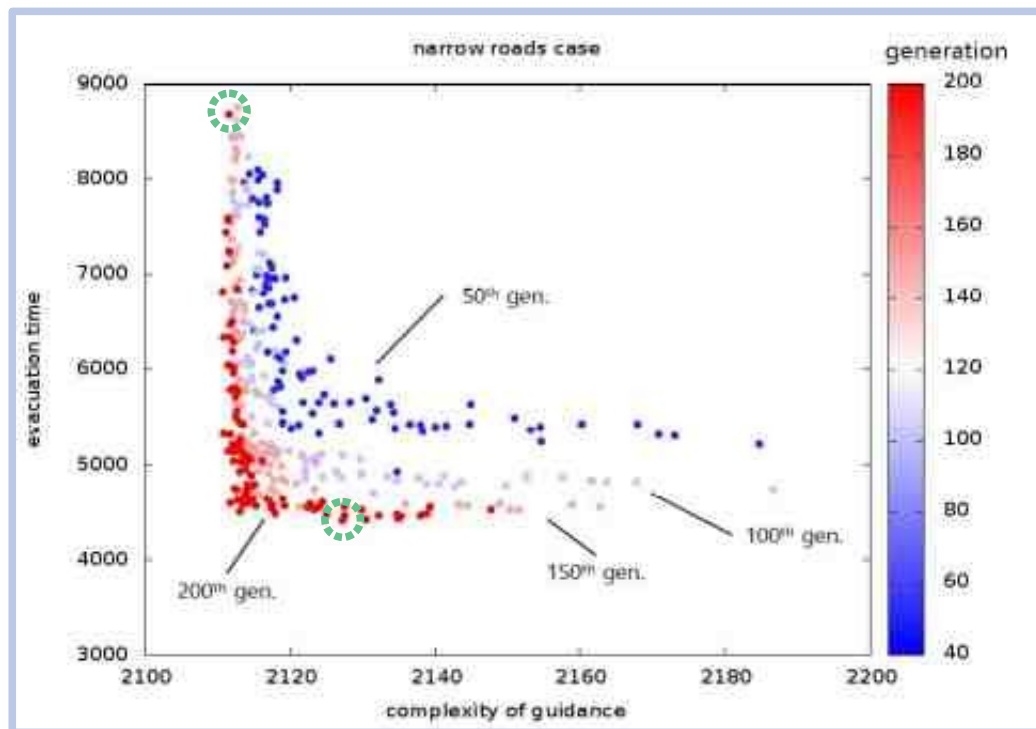
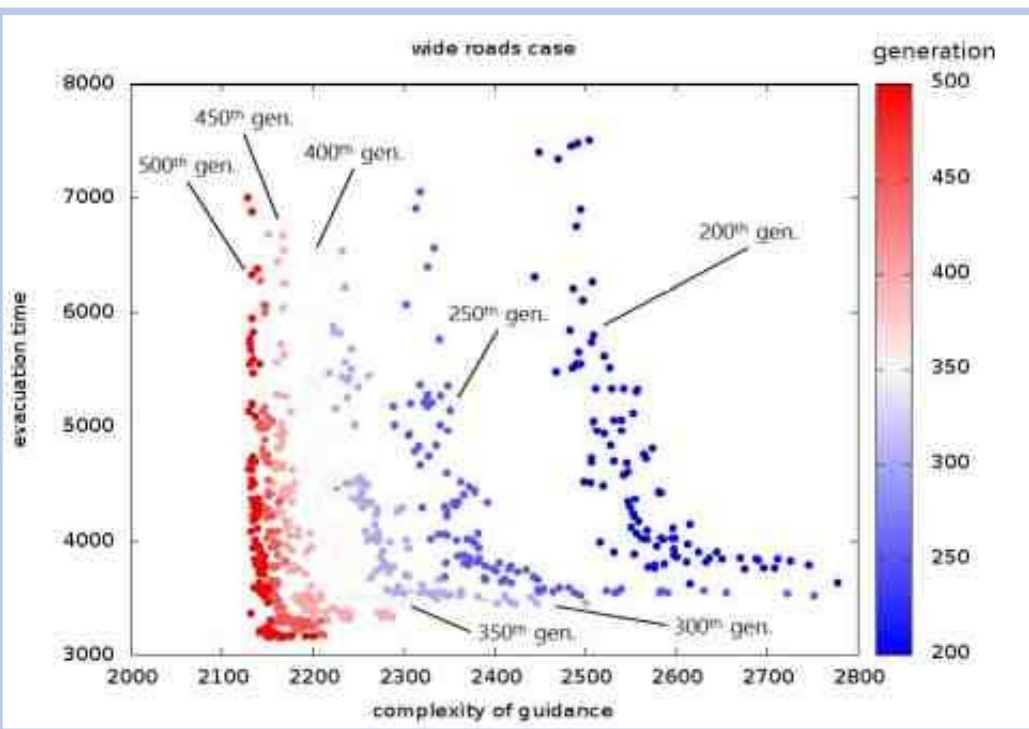
▶ HPCの並列性の有効活用



世代交代とパレート解

●大阪市淀川区での避難

- ▶ 車道も使ったシミュレーション(wide)
- ▶ 歩道だけのシミュレーション(narrow)

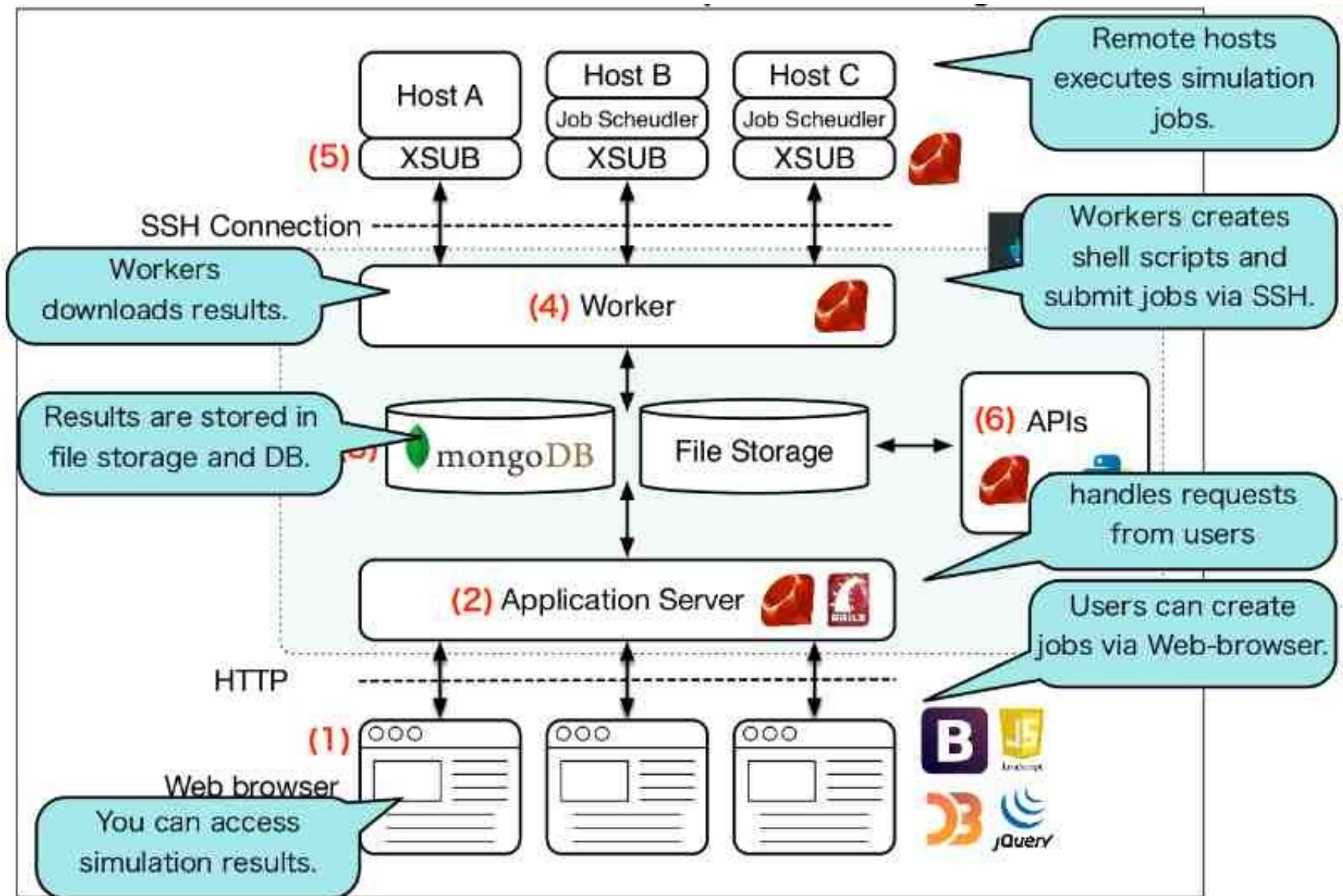


Job Scale and Frameworks

	class A	class B	class C	class D
# of jobs	$10^0 - 10^2$	$10^3 - 10^5$	$10^6 - 10^9$	$10^9 -$
# operations/job	$10^{19} - 10^{17}$	$10^{16} - 10^{14}$	$10^{13} - 10^{10}$	$10^9 -$
flops/jobs	$10^{18} - 10^{16}$	$10^{15} - 10^{13}$	$10^{12} - 10^9$	$10^8 -$
# of cores/job	$10^8 - 10^6$	$10^5 - 10^3$	$10^2 - 10^{-1}$	$10^{-2} -$
framework	capability comp.	OACIS xcrypt	CARAVAN Map-Reduce	Map-Reduce capability comp.
granularity of a job	MPI-para	MPI-para	SMP-para / job	function
size of results	- 1 TB	- 1 GB	- 1 kB	- 100 B
how to analyze	by hand	by hand / statistics	statistics	statistics
application example	stencil calc.	replica ex. MC data assim. for weather	genome analy.	MCMC optimization
social simulation	/	pedestrian/traffic in Tokyo Tokyo Stock Market	city-level traffic single market	data-driven model

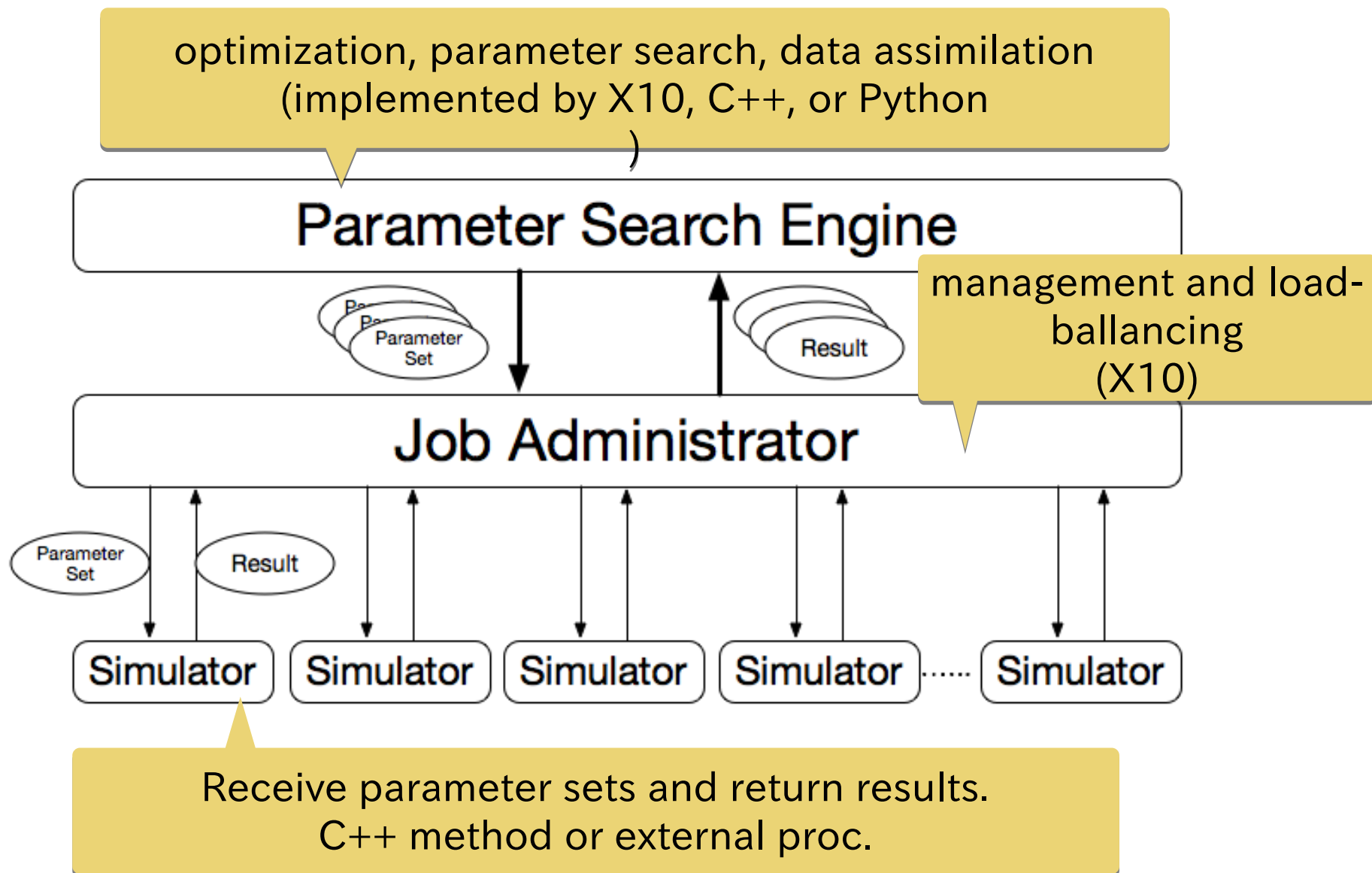
OACIS

Organizing Assistant for Comprehensive and Interactive Simulations



●Ruby on Rails + MongoDB, Unix-based OS

CARAVAN



OACIS/CARAVANによる並列実行性能

●OACIS/CARAVAN:

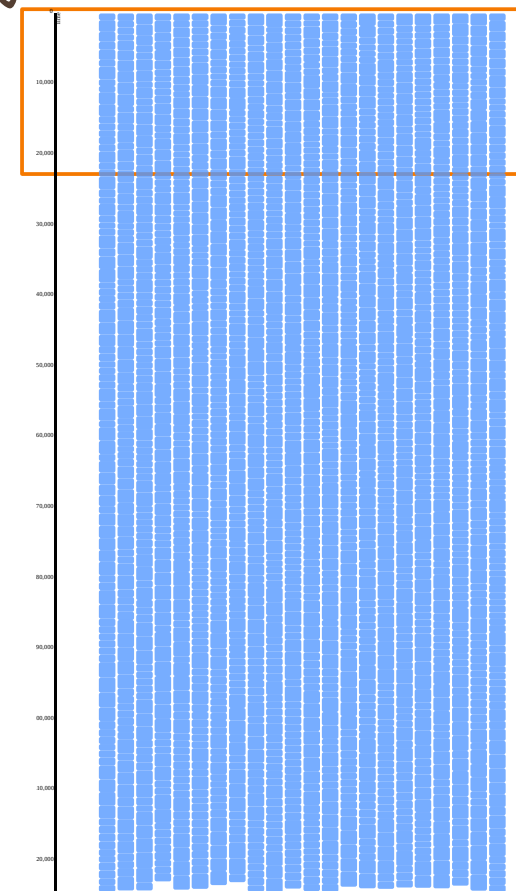
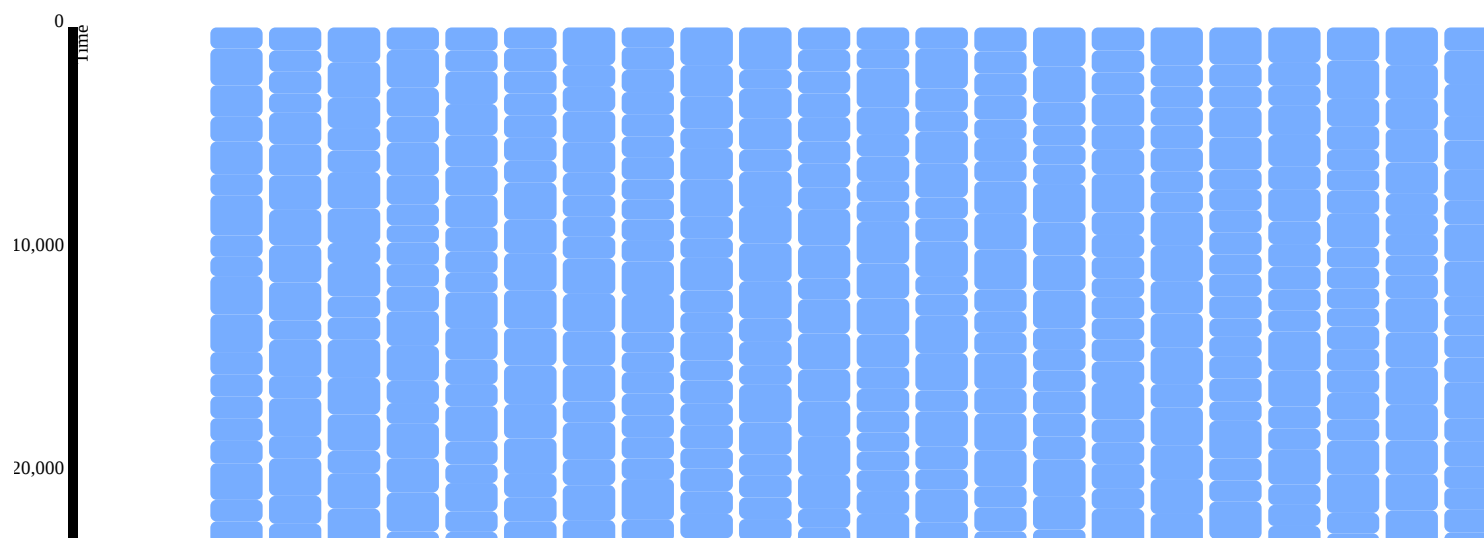
- ▶ 社会シミュレーションの大規模網羅的実行支援フレームワーク

●OACIS/CARAVAN+NSGA-II+漸進的世代交代

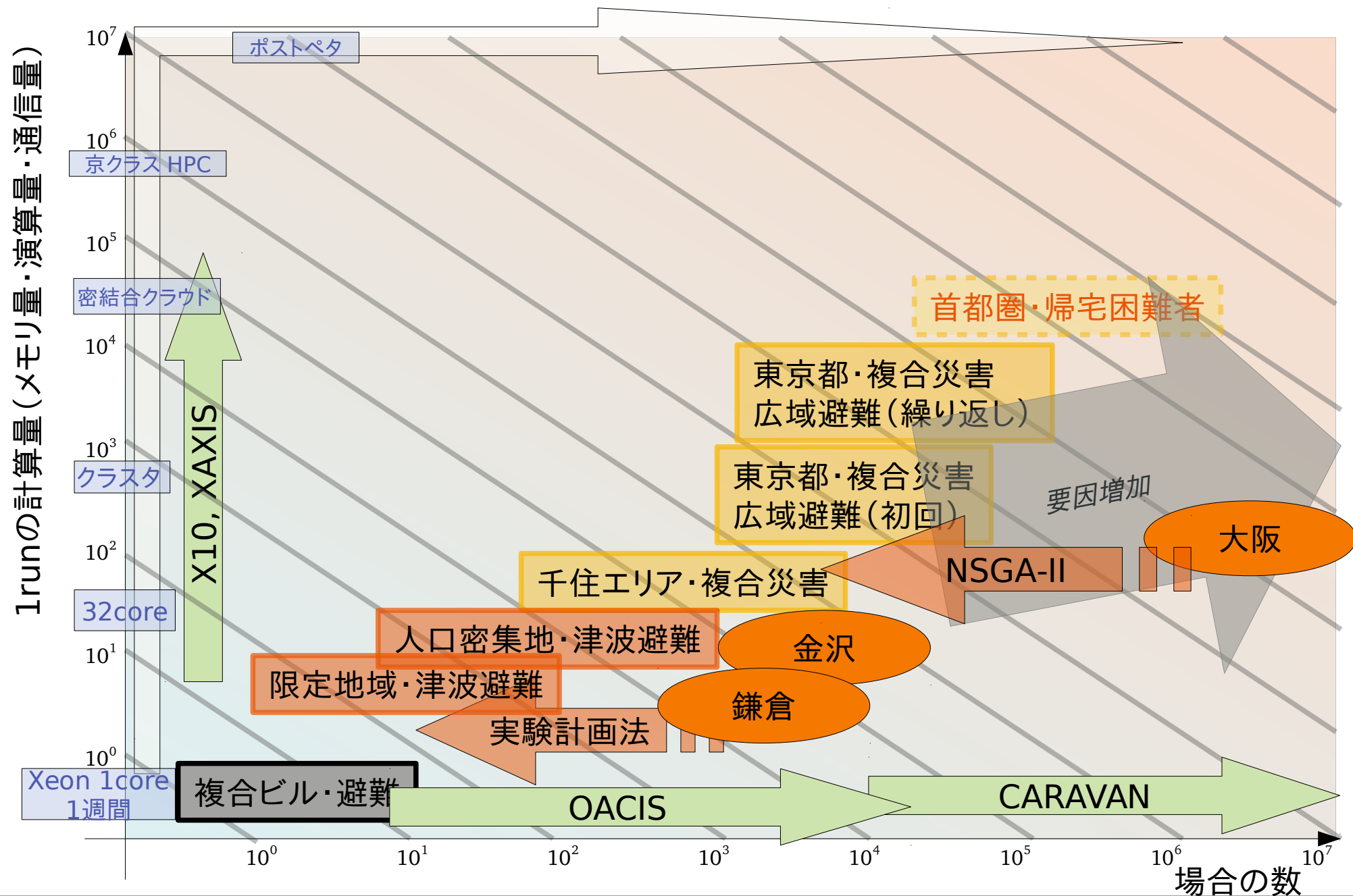
- ▶ 人流シミュレーション(CrowdWalk)の並列実行

- 1実行: 10分 - 30 分

●京の上での充足率: 95.3%



人流・避難誘導



計算社会科学のためのHPC

●開発環境

- 言語: java, ruby, python, ...
 - Fortran は使わない。
- テスト環境 = 実行環境
 - docker 機能のサポート
- 超並列クラウド計算機環境
 - MPI より ssh
 - 1コアあたりのメモリが多数。(知識や学習に必要)

●並列性

- 必要なのは、バカパラ的性能。
 - ある意味、性能の無駄遣いかもしれない。
 - HPC の専門家は、フルスペック性能ばかり求める。
 - ◆ メモリ効率・CPU効率・通信効率
- 計算社会科学では、計算モデル探求が目的。
 - モデルの変更が研究そのもの。
 - ◆ モデルを固定して最適化する時間がない。

●運用面

- ジョブのタイプの自由度
 - コア数の動的配分、長時間実行

計算対象としての社会現象の特徴

●モデルが未確定

- ▶ 人間行動や思考のモデルは開発途上
- ▶ 数値化・定量化の難しさ
- ▶ 場合の数によるデータ同化

●曖昧な現象境界

- ▶ 非独立事象・相互依存
- ▶ 規模効果
- ▶ 経年変化

●不完全なビッグデータ

- ▶ 不均一・不定形・不正確なデータ
 - データの不備を計算で補完する必要。
 - シンボル+パターン

試行錯誤・
インタラクティブな
分析

例えば、
数時間～数週間の
turn-around time

まとめ

●計算社会科学をより実用的にするには、

- ▶ 非常に多数の試行を効率的かつ気軽に実施。
 - 多岐にわたる条件を網羅する必要。
 - 繰り返しデータ同化・パラメータ探索の必要性。
 - 真の人のモデル・社会の初期条件の不可知性を補完。
- ▶ 離散データ、構造化データをnativeでサポート。
 - 表現の柔軟性。
 - データとしてのプログラム。

●次世代HPCに望まれる要件

- ▶ 数値計算偏重でない開発環境。
 - 多様なプログラム言語サポート。
 - 通常のPCとのシームレスなエコシステム。
- ▶ 柔軟なジョブ管理。
 - ブロック単位でのジョブ割り付けを柔軟化。
 - ジョブ実行の仮想化？