MOLAPのための
多次元配列の実現方式

福井大学大学院工学研究科
一色　淳夫

研究の目的

MOLAP：**多次元配列を用いて**
OLAPシステムを実現

⇒ **配列が疎，次元依存性という問題**

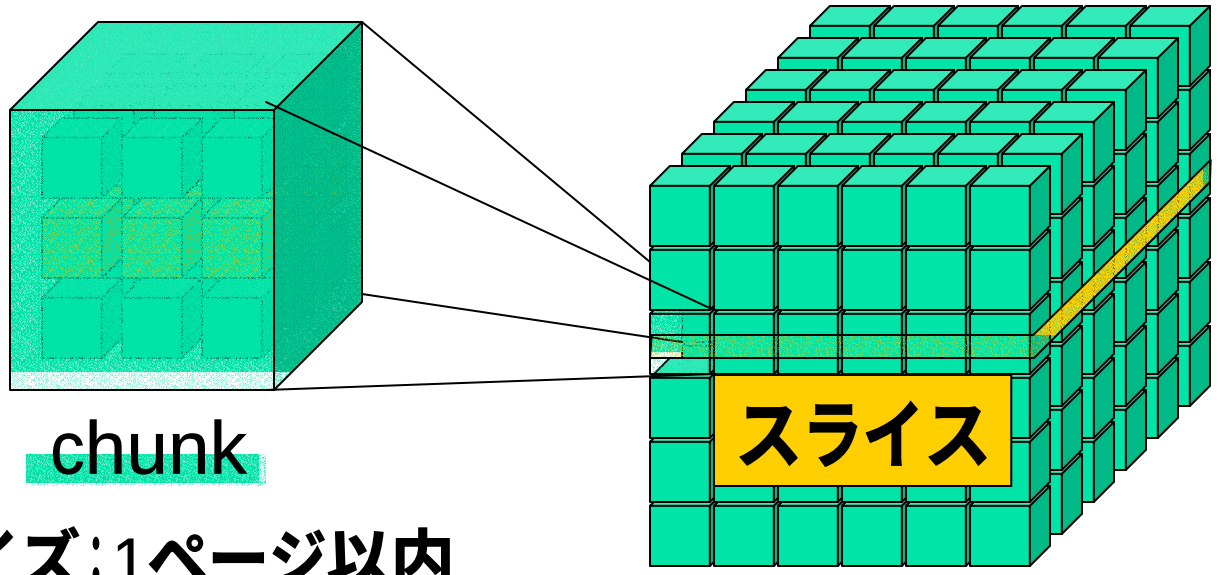
↓

検索時の読み込み回数と
次元軸によるばらつきの低減

MOLAPシステムの問題

- **疎配列問題**
  データの格納されていないセルが
  データベースの領域を占める
- **次元依存性**
  n次元データを二次記憶に線形に格納した場合，例えばスライス断面の集約演算レスポンスにばらつき

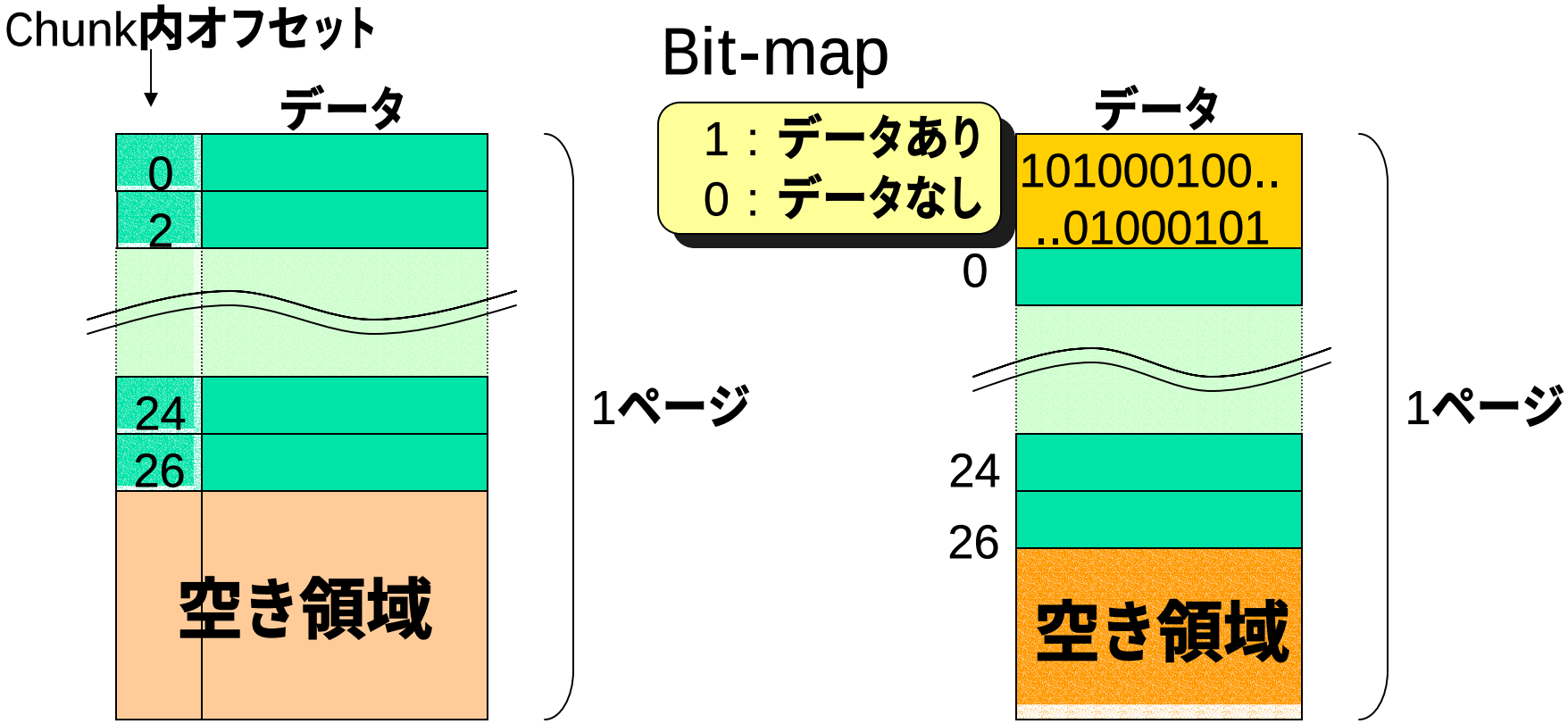
次元依存性の対策



サイズ：1ページ以内

多次元データベース

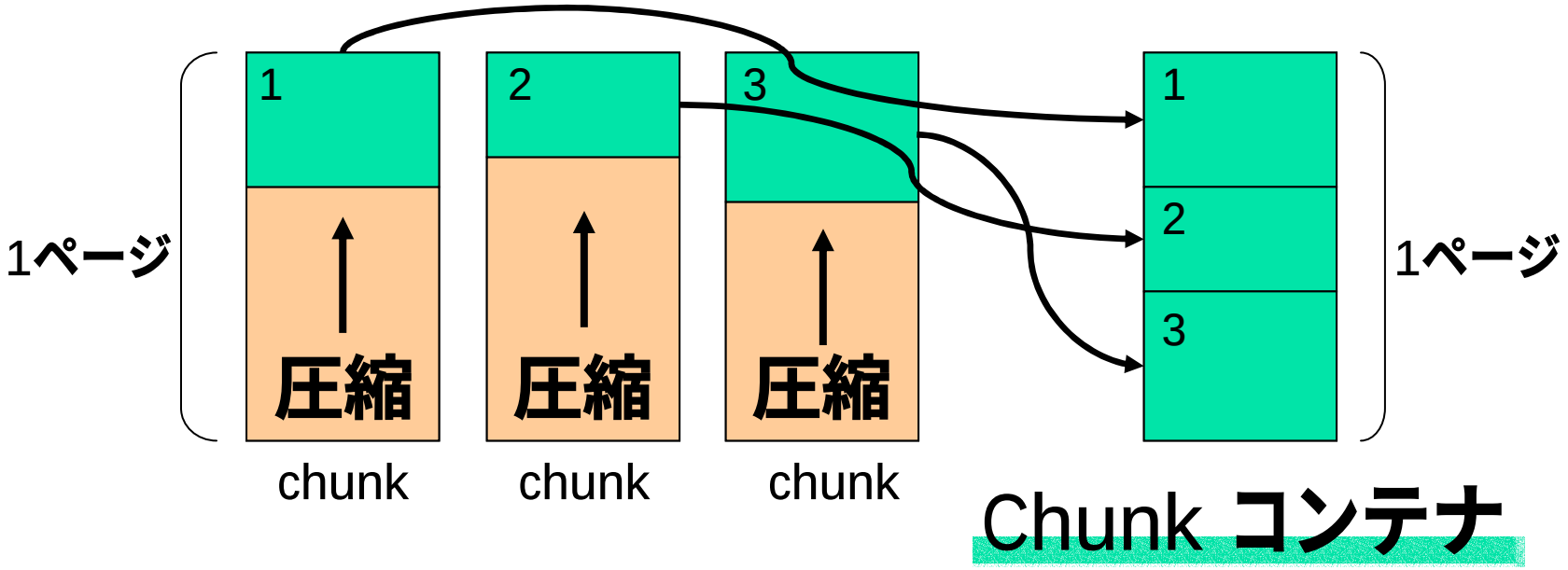
疎配列問題の解消



Chunk-offset compression

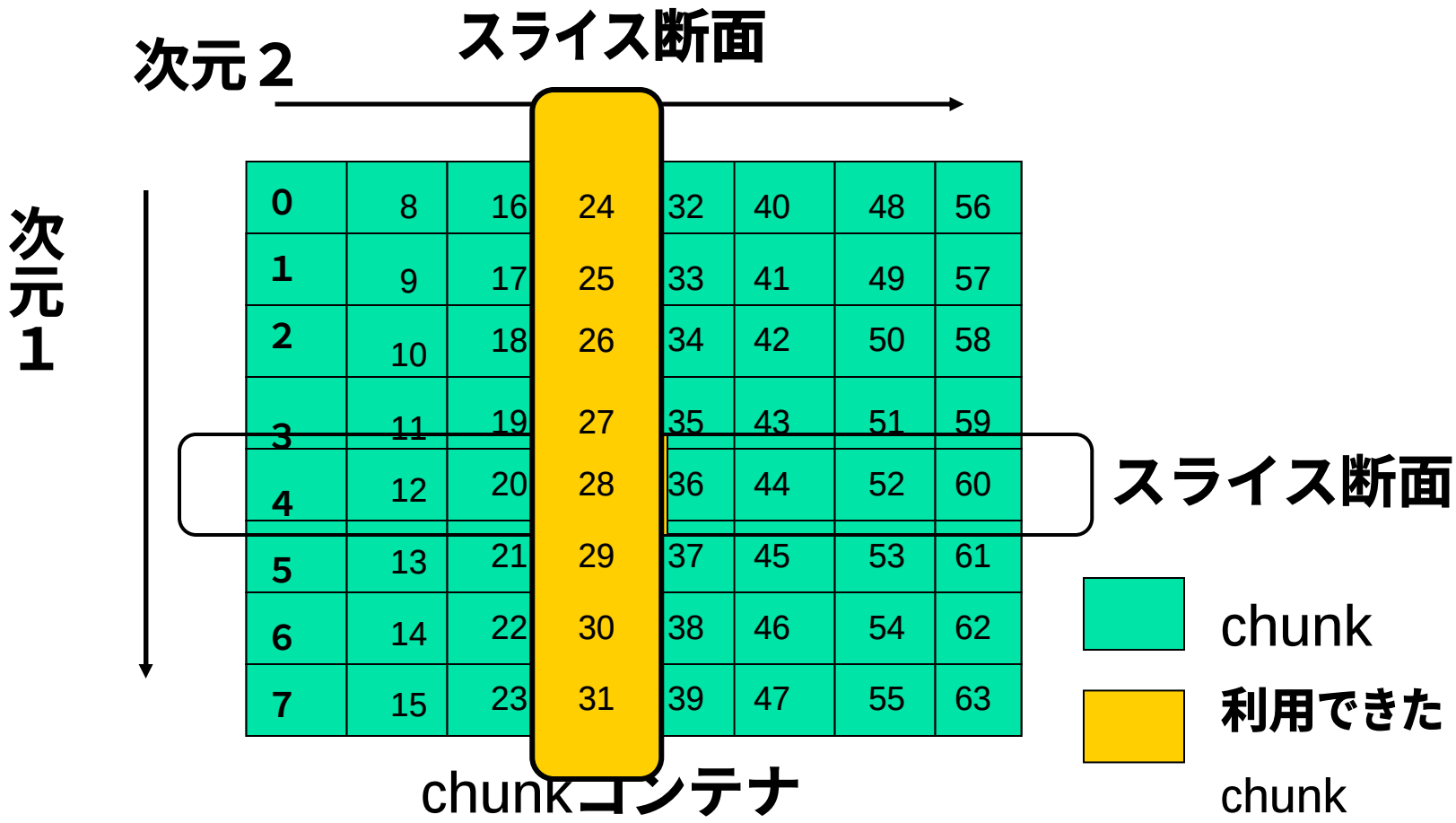
Bit-mapを使った圧縮法

Chunkのコンテナ化



- スライス時の読み込み回数の削減
- 次元依存性の発生

コンテナ格納方法による次元依存性

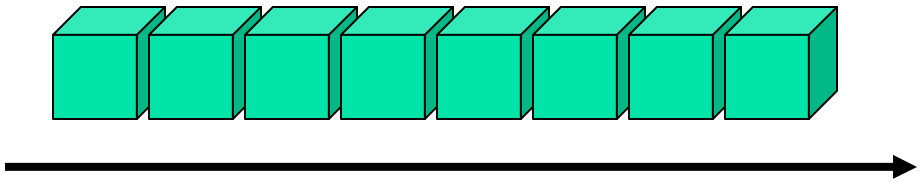


コンテナ格納時のchunkの選択条件

- コンテナ充填率を上げる
  - 一度の読み込みで得られるchunk増加
- 隣接chunkを同一コンテナに
  - スライス処理に有利
- コンテナ形状を超立方体に
  - 次元依存性の解消
  - コンテナ化時の干渉を防ぐ

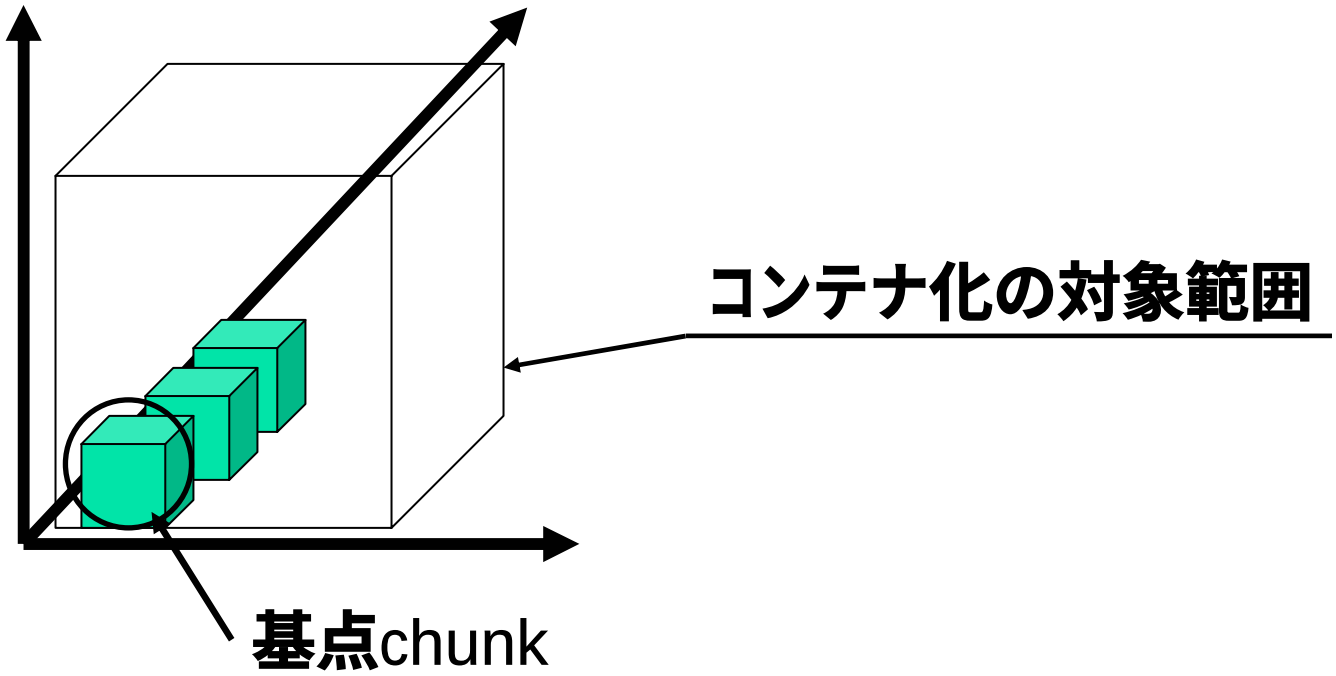
コンテナ化方式の提案　I

- 次元順
  - 通常の次元順にコンテナを作成

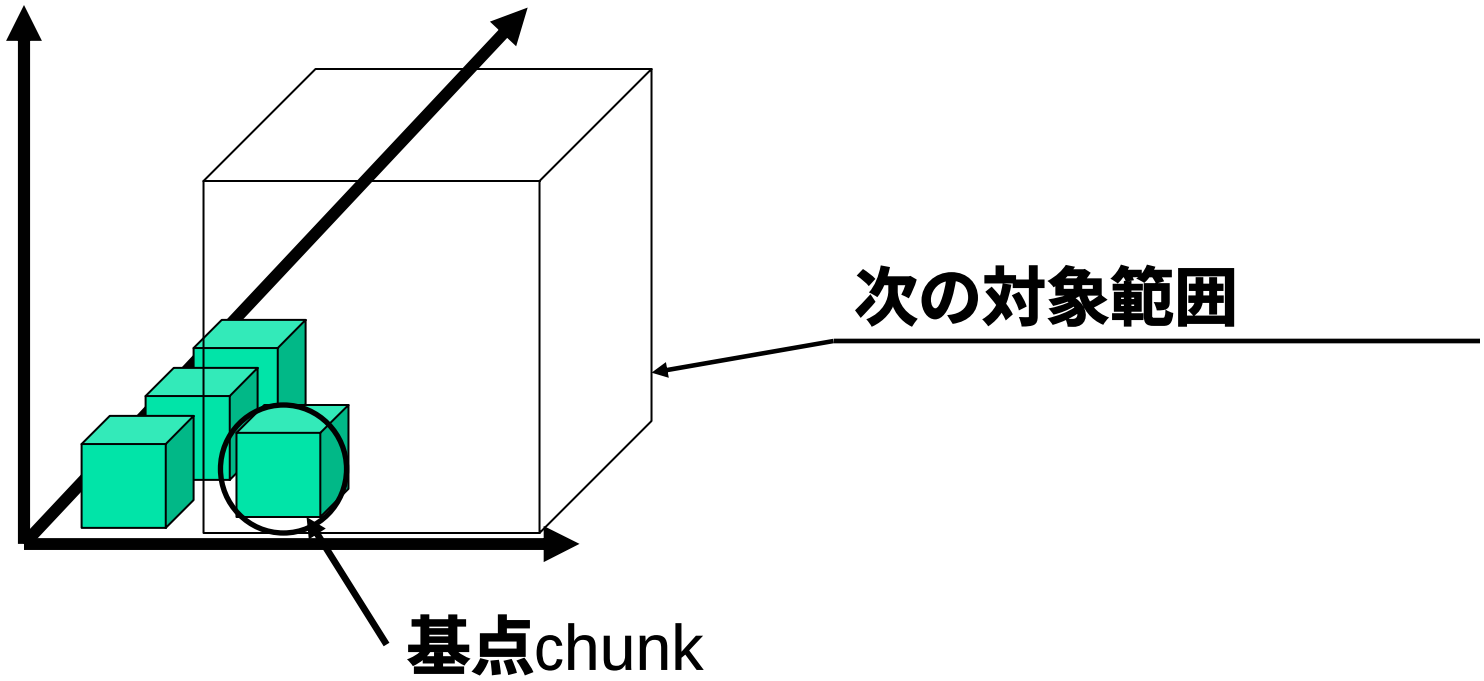


  - 次元依存性を考慮しない
    ⇒ 他の方式の比較基準

コンテナ化対象範囲と基点

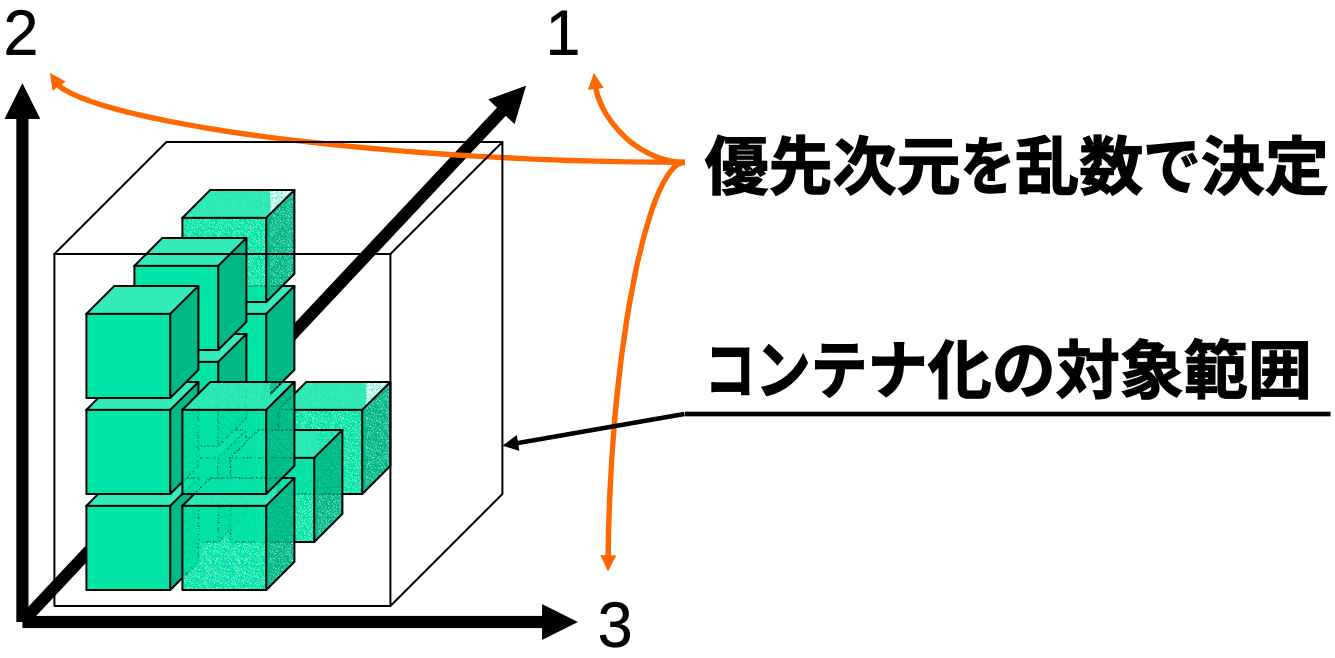


コンテナ化対象範囲と基点



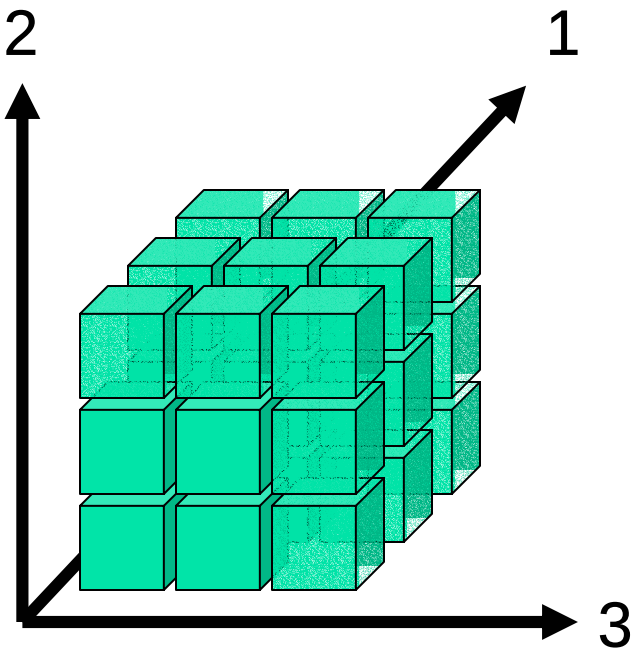
コンテナ化方式の提案　II

- **特定次元優先**　(PRI)



コンテナ化方式の提案 III

- **超直方体** (rect-n)

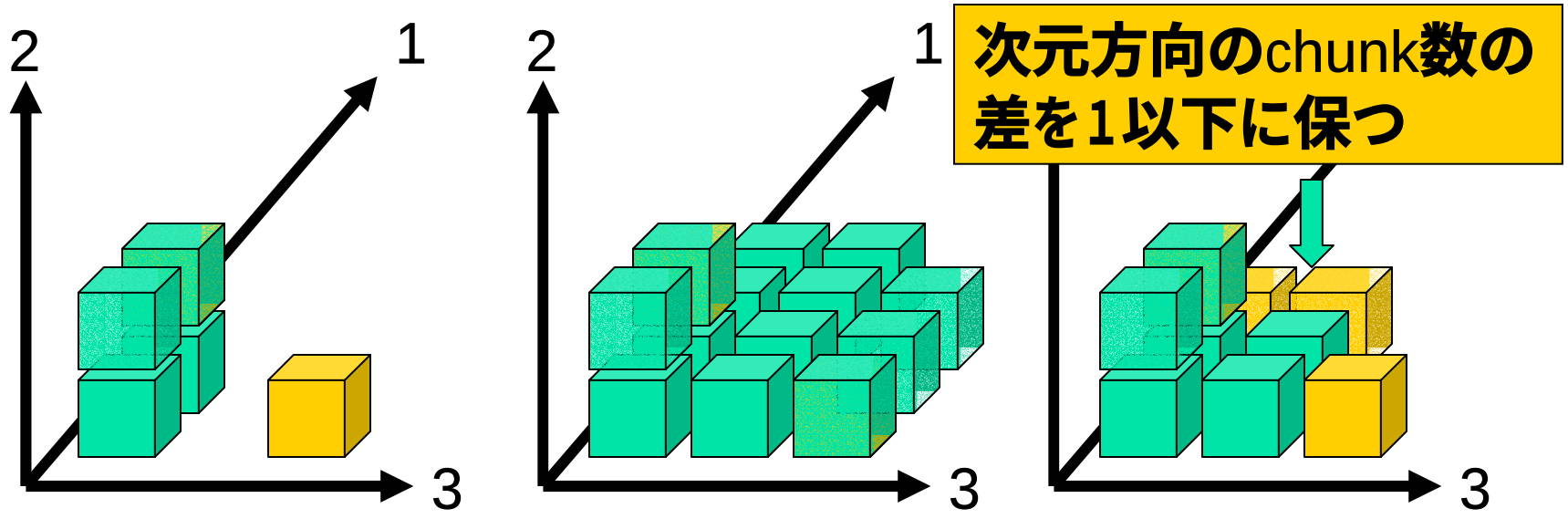


優先次元を乱数で決定

- rect1 **優先次元**
- rect2 **充填率**
- rect3 **形状**

コンテナ化方式の提案 III－2

- rect1
  優先次元
- rect2
  充填率
- rect3
  形状

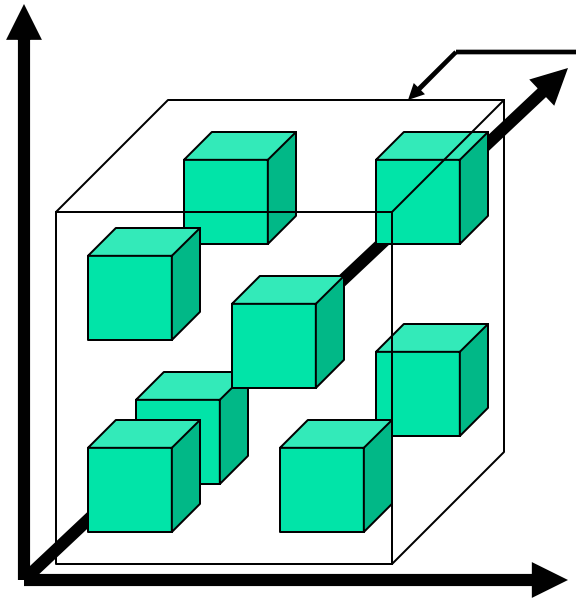


コンテナ化済み/高データ充填率

コンテナ化対象

コンテナ化方式の提案 IV

- **特定領域でランダム** (rand-n)

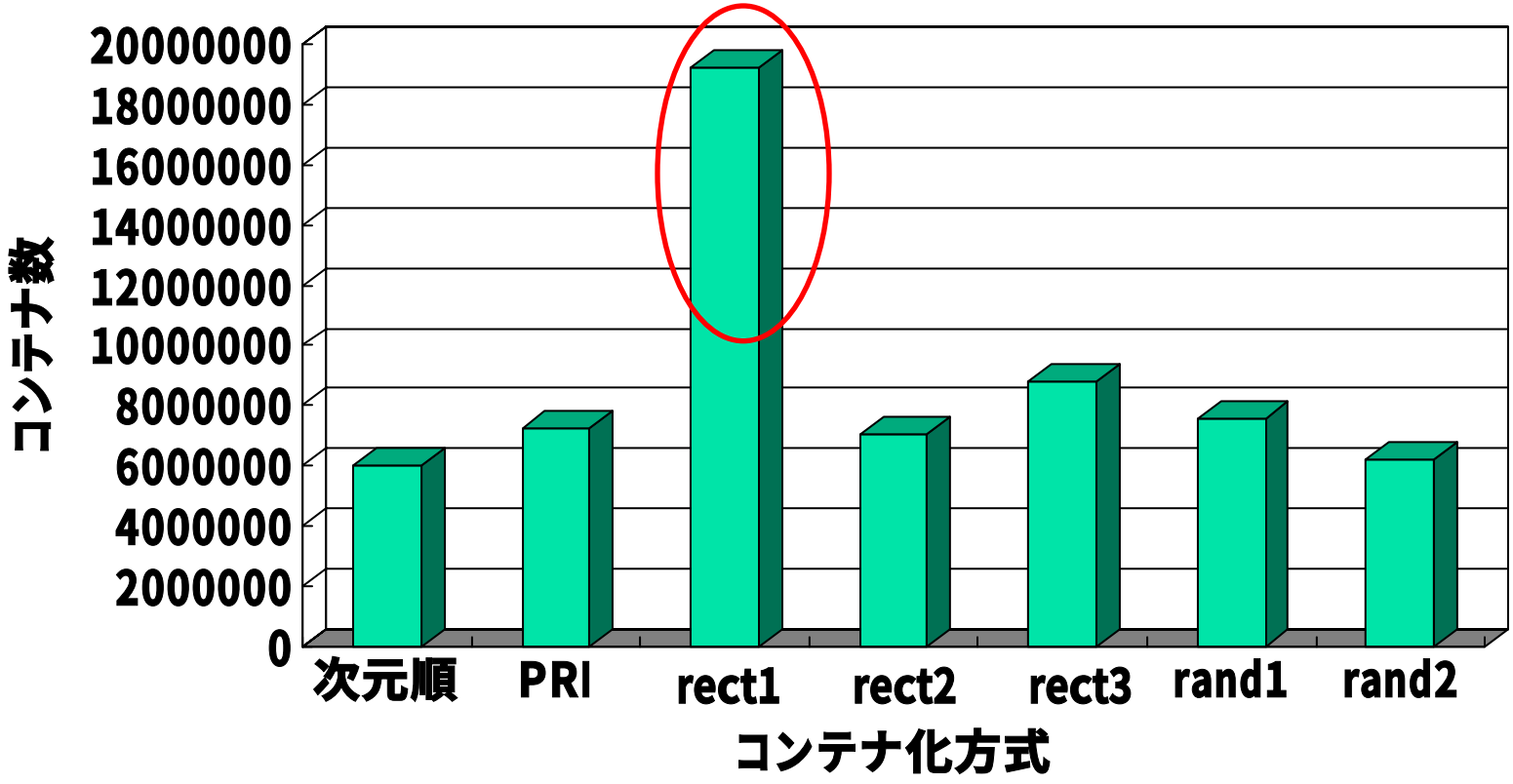


- rand1 **小さな領域**
  - **各次元方向 2**
- rand2 **大きな領域**
  - **各次元方向 3**

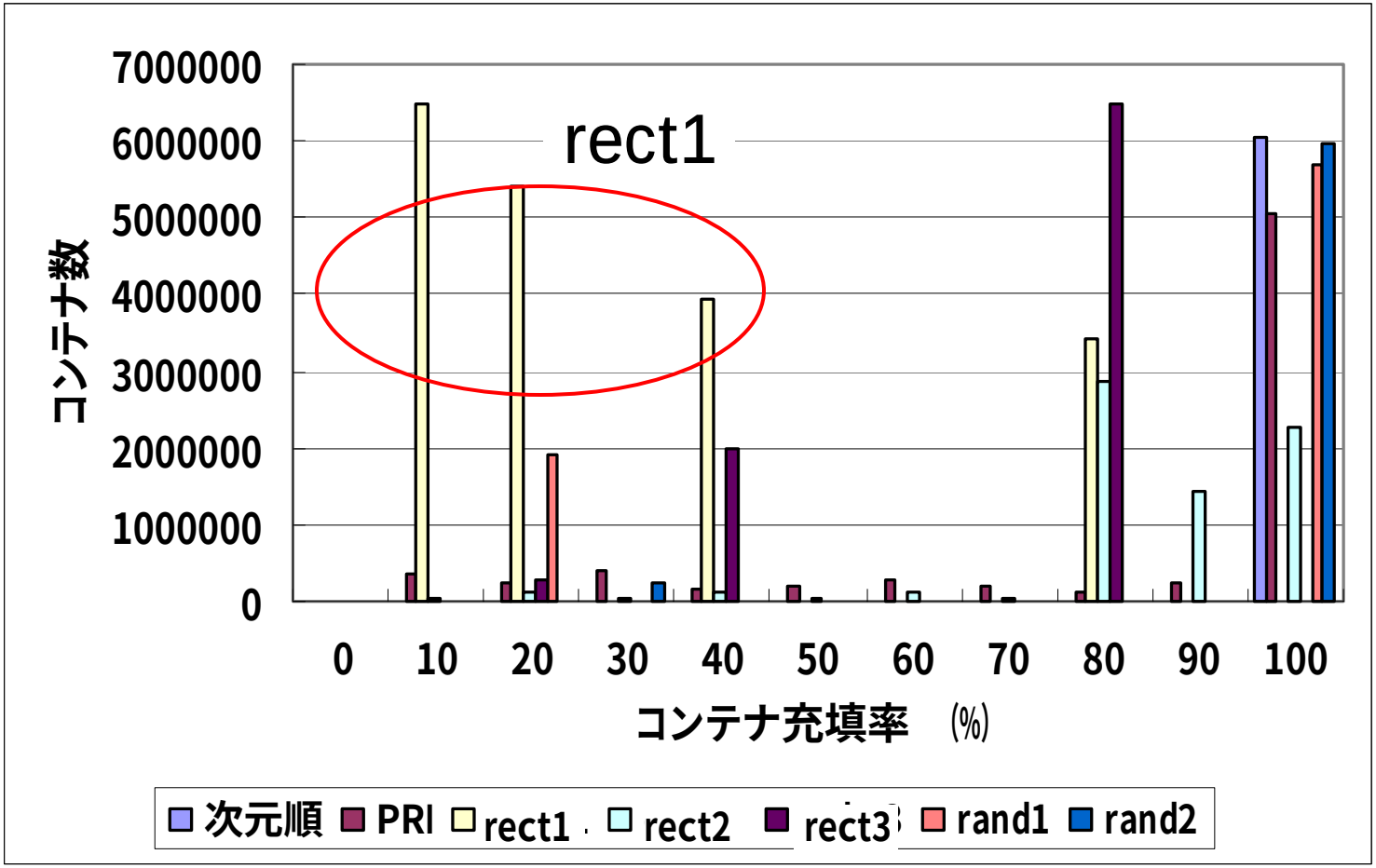
シミュレーション条件

- **次元数　5**
- **次元方向のchunk数**　36
- Chunk**総数**　36^5 =　60,466,176**個**
- Chunk**のデータ充填率　一様分布１０％**
- **各次元で，スライス時のコンテナ読み込み回数の平均と標準偏差を計測**

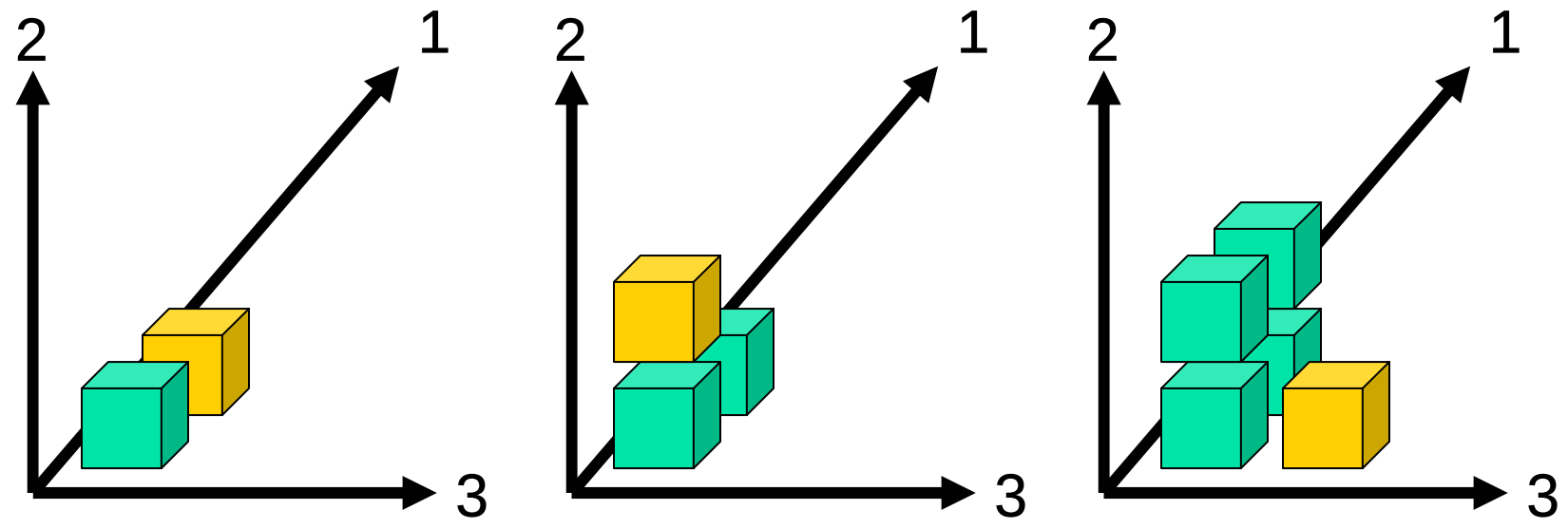
実験結果　(コンテナ総数)


20000000
18000000
16000000
14000000
12000000
10000000
8000000
6000000
4000000
2000000
0
コンテナ数
次元順
PRI
rect1
rect2
rect3
rand1
rand2
コンテナ化方式

実験結果　(コンテナ充填率)



rect1の低充填率の原因

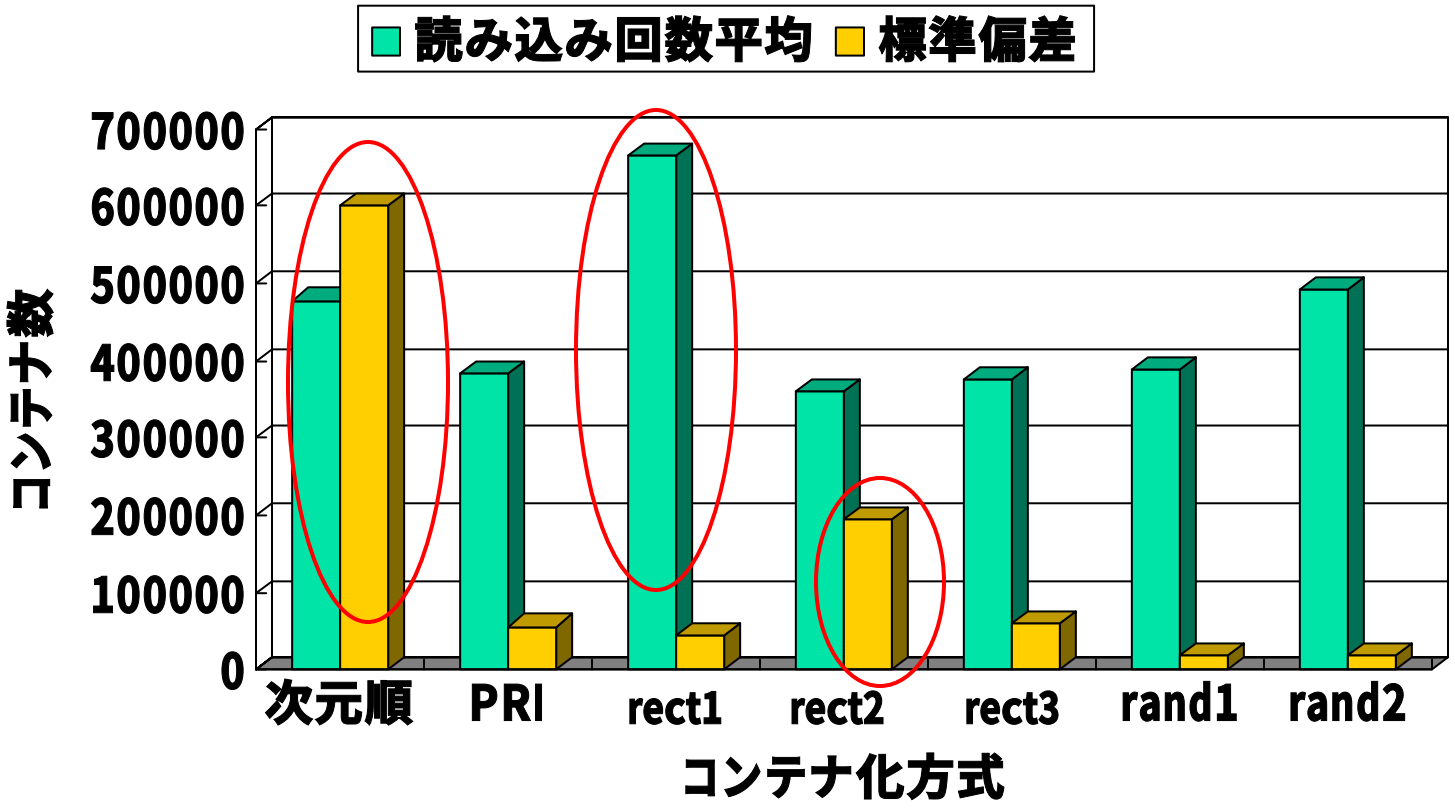


ただちにコンテナ化終了

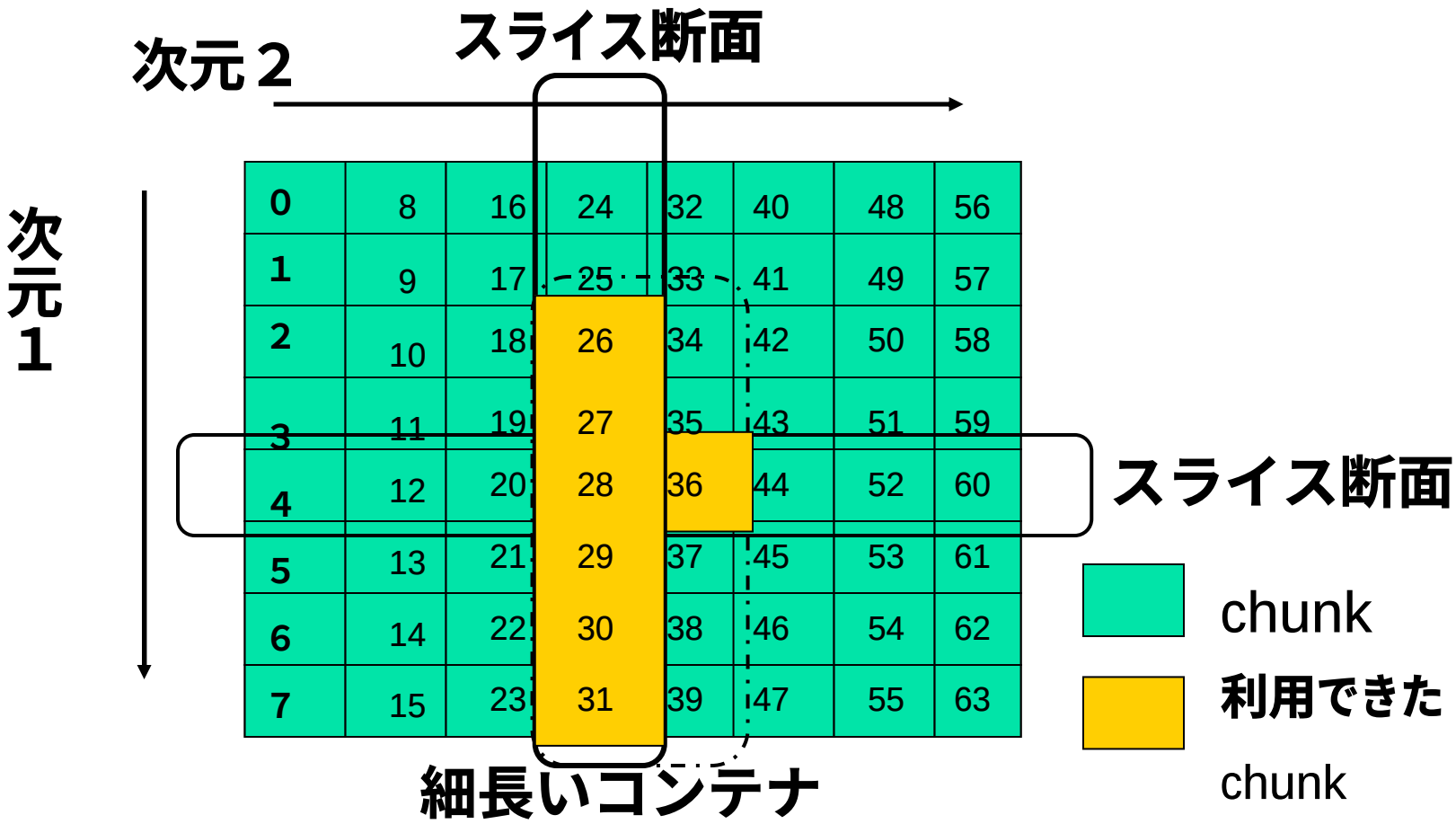
コンテナ化済み /高データ充填率

コンテナに格納

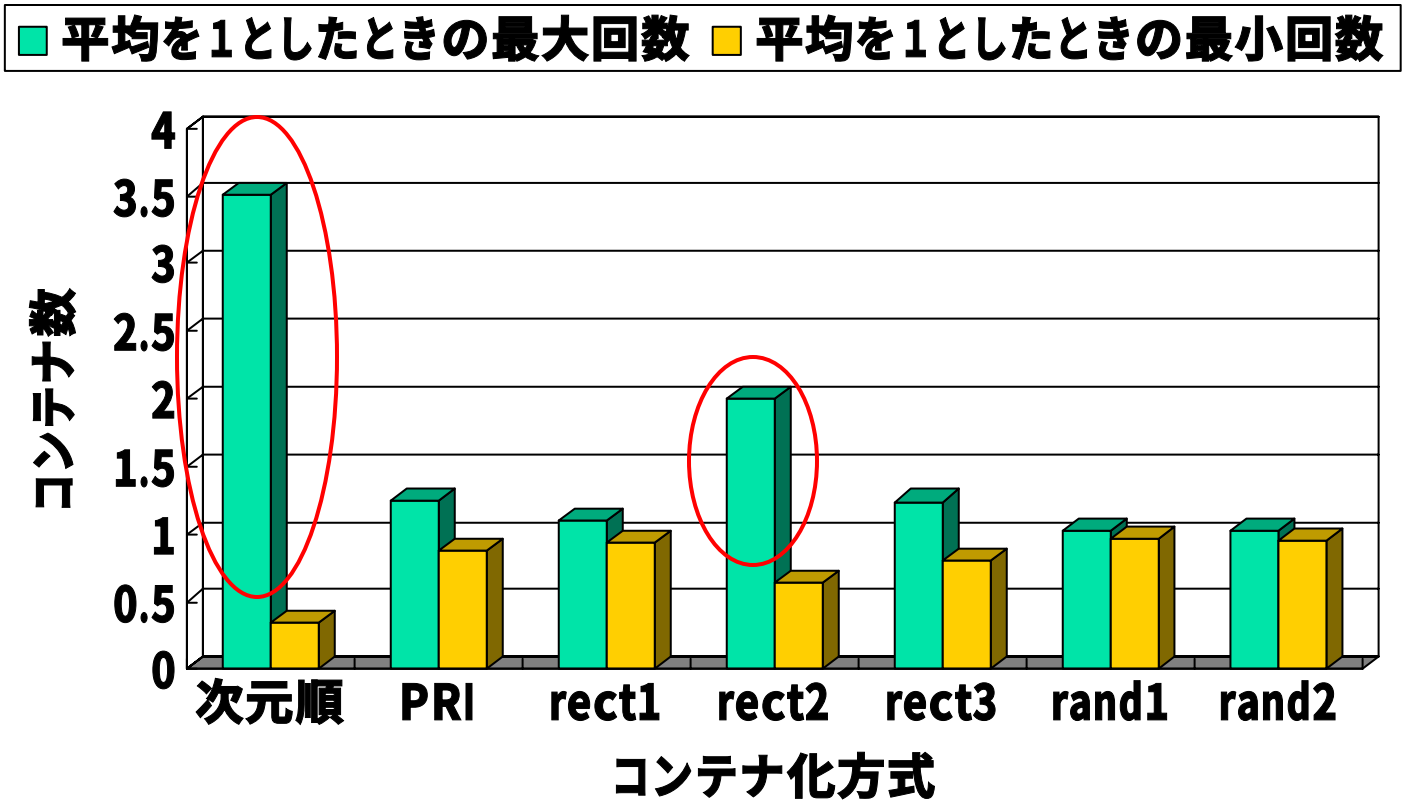
実験結果 (4次元スライス時)



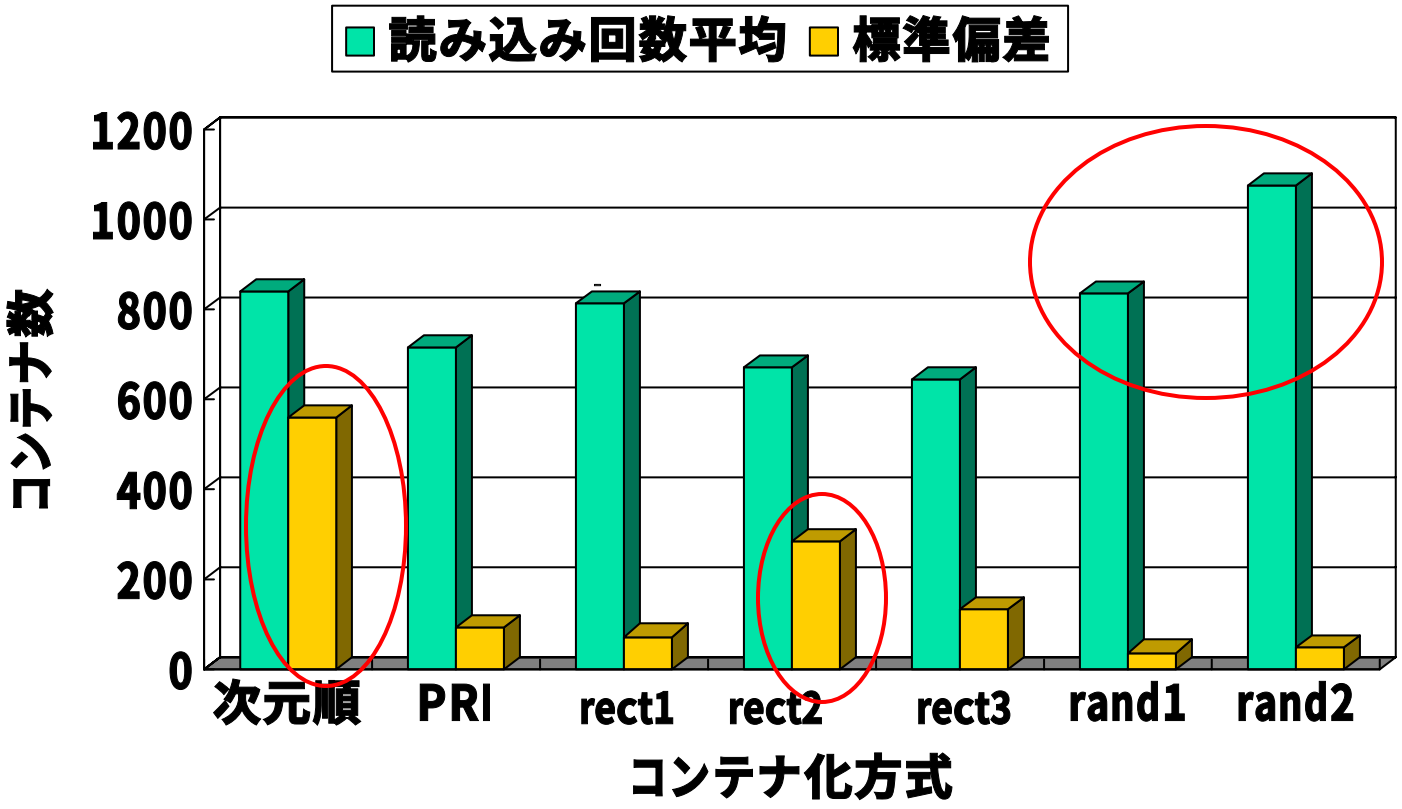
rect2の大きい標準偏差の原因



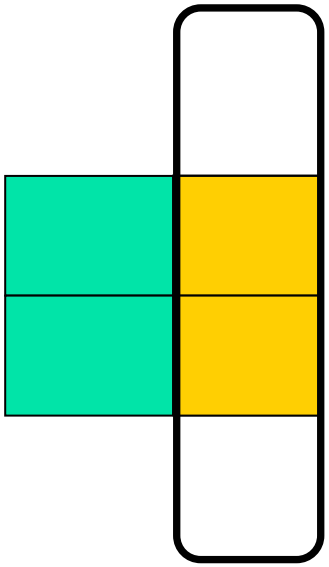
実験結果 (4次元スライス時の最大･最小読み込み回数)



実験結果　(2次元スライス時)

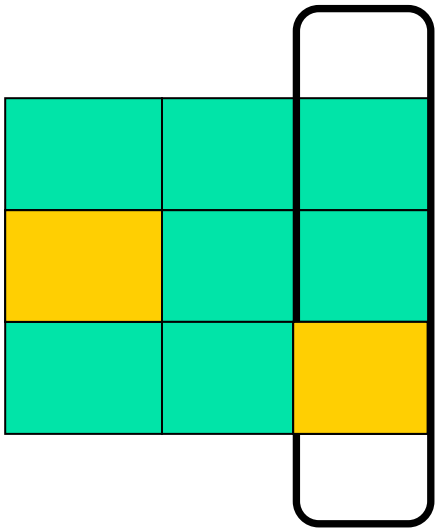


rand2の読み込み回数劣化原因



スライス断面

rand1



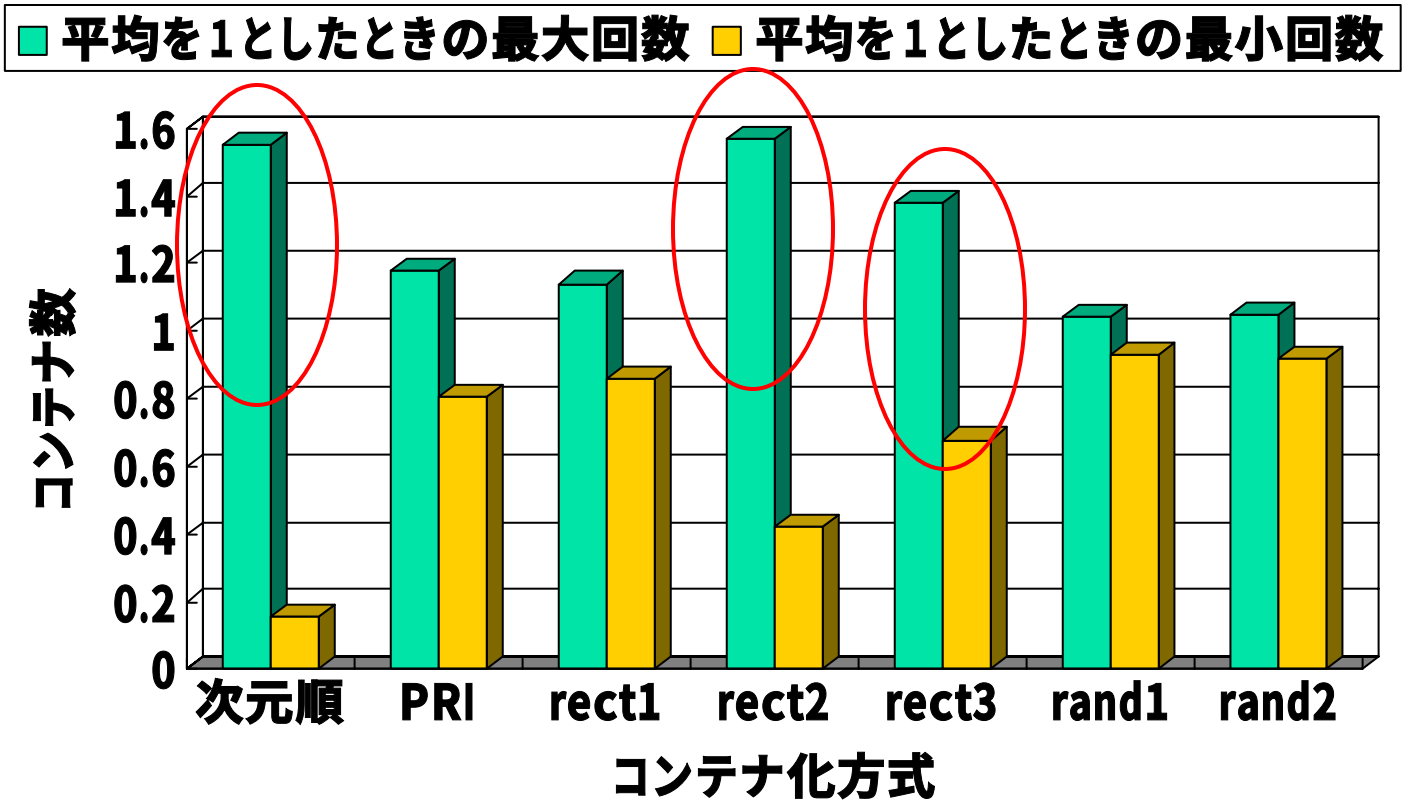
スライス断面

rand2

対象範囲内chunk

同一コンテナ内chunk

実験結果（2次元スライス時の最大･最小読み込み回数）



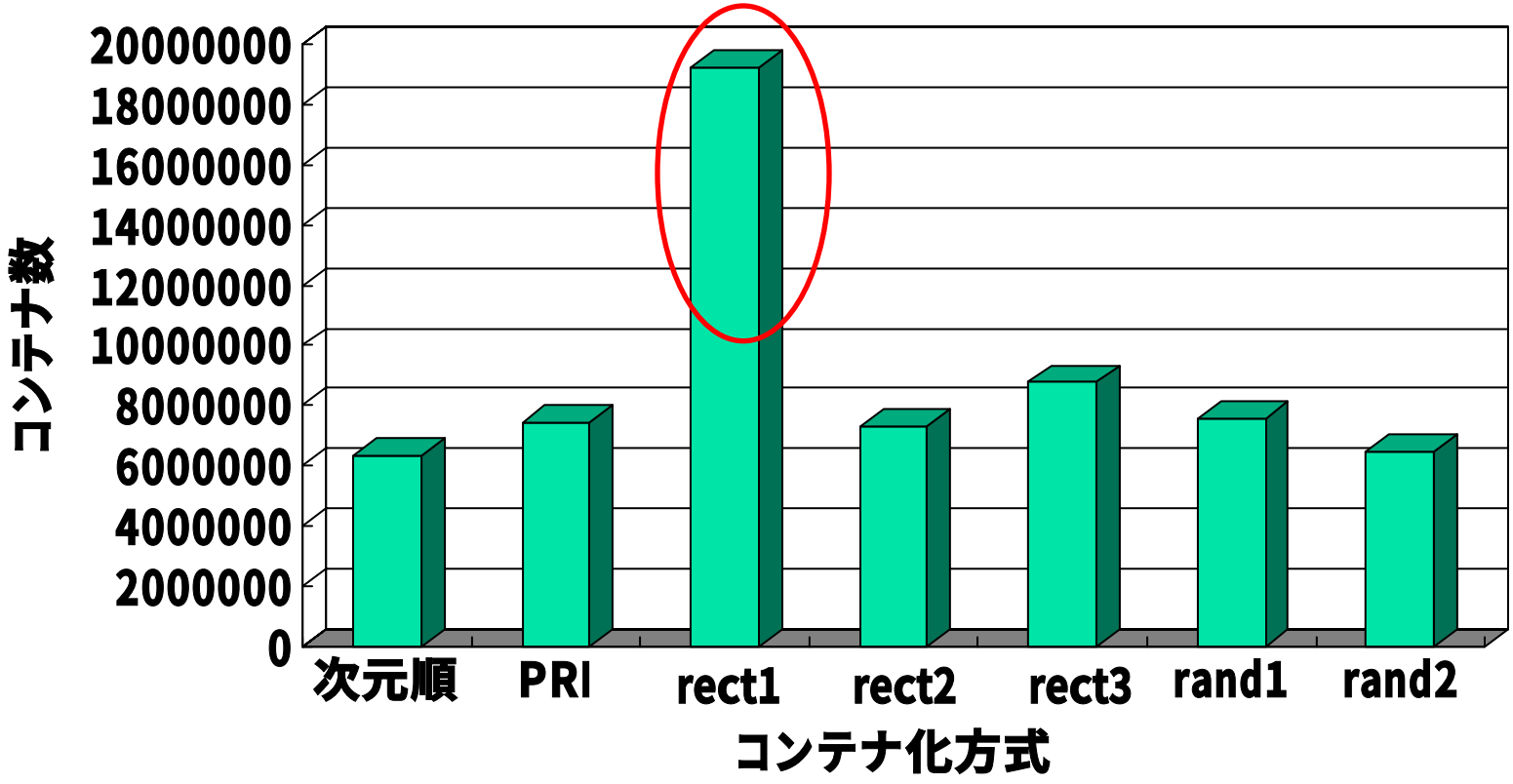
まとめと課題

- 疎配列と次元依存性の問題を“chunk”のコンテナ化を行うことで解消した.
- 方式 rect3, PRI でコンテナを作成した場合, 読み込み回数の平均, 標準偏差のどちらも低減できる.
- 今後, 配列の各次元の長さが異なる場合など、より一般化した評価が必要.

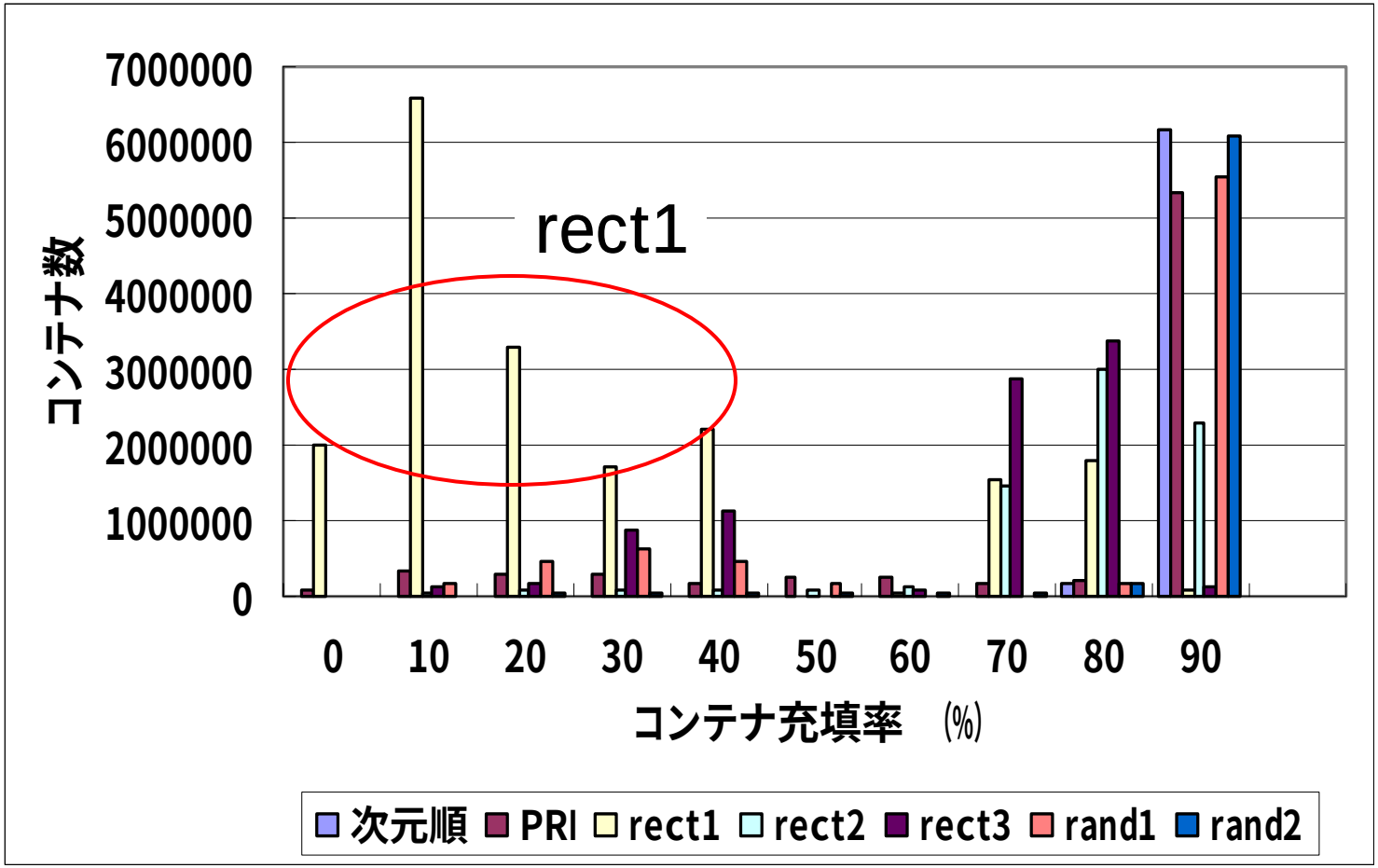
シミュレーション条件

- **次元数　5**
- **次元方向のchunk数　36**
- Chunk**総数**　36 ^{5} = 60,466,176**個**
- Chunk**のデータ充填率　正規分布平均1**
  0%
- **各次元で，スライス時のコンテナ**
  読み込み回数の平均と標準偏差を計測

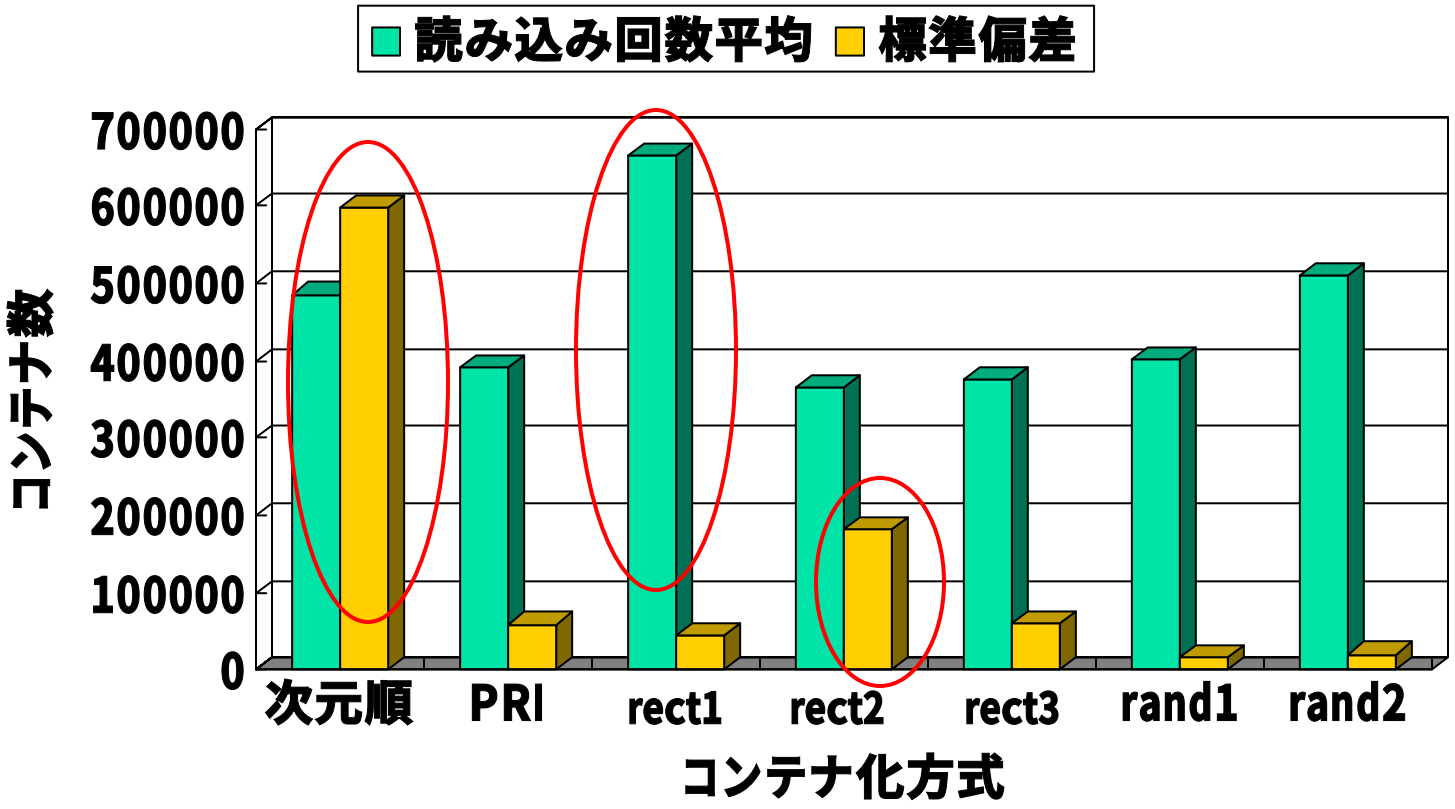
実験結果　(コンテナ総数)


20000000
18000000
16000000
14000000
12000000
10000000
8000000
6000000
4000000
2000000
0
コンテナ数
次元順
PRI
rect1
rect2
rect3
rand1
rand2
コンテナ化方式

実験結果　(コンテナ充填率)



実験結果 (4次元スライス時)



実験結果　(2次元スライス時)