

# パラメータサーバを用いた 並列機械学習システムにおける 耐故障性のシミュレーション

黎 明曦 (筑波大学)

谷村 勇輔 (産総研)

中田 秀基 (産総研)

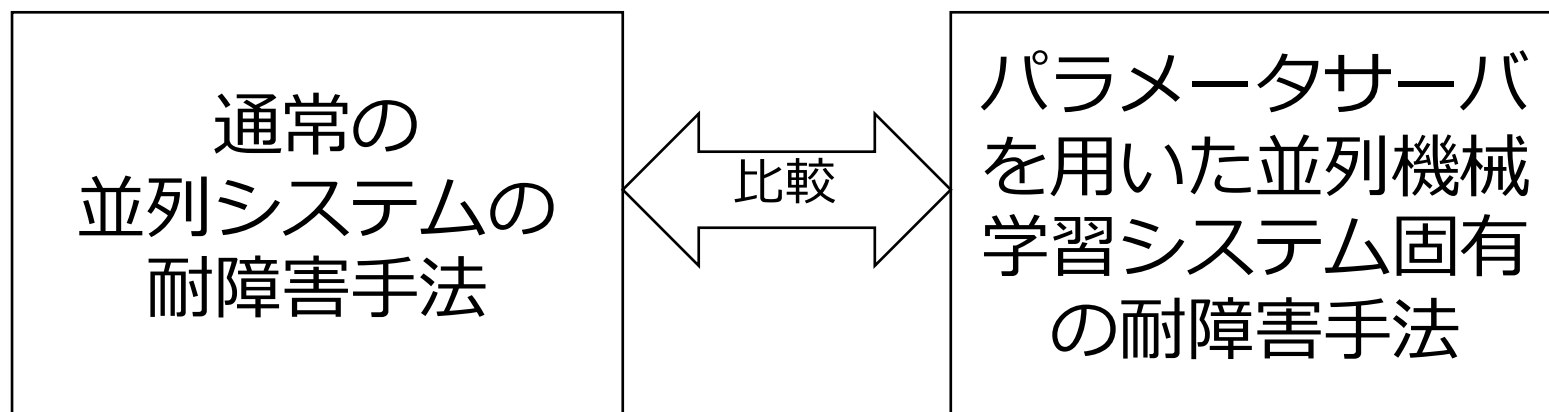
# 背景

- 機械学習：大量のデータの処理
  - 並列化による高速化は必要
- データ並列機械学習システム
  - 数百以上の計算機
  - 耐故障性は必須
- パラメータサーバ
  - サーバに類似した情報が多数存在する
  - 復旧は容易
  - 性能面の解析は不十分

# 研究の概要及び結果

概要：

- 並列シミュレータ「SIMGRID」



結果：

- パラメータサーバ上の情報を用いて復旧：通常の耐障害手法よりもコストを大幅に低減
- 収束への悪影響を許容：通信コストも低減

# 発表の概要

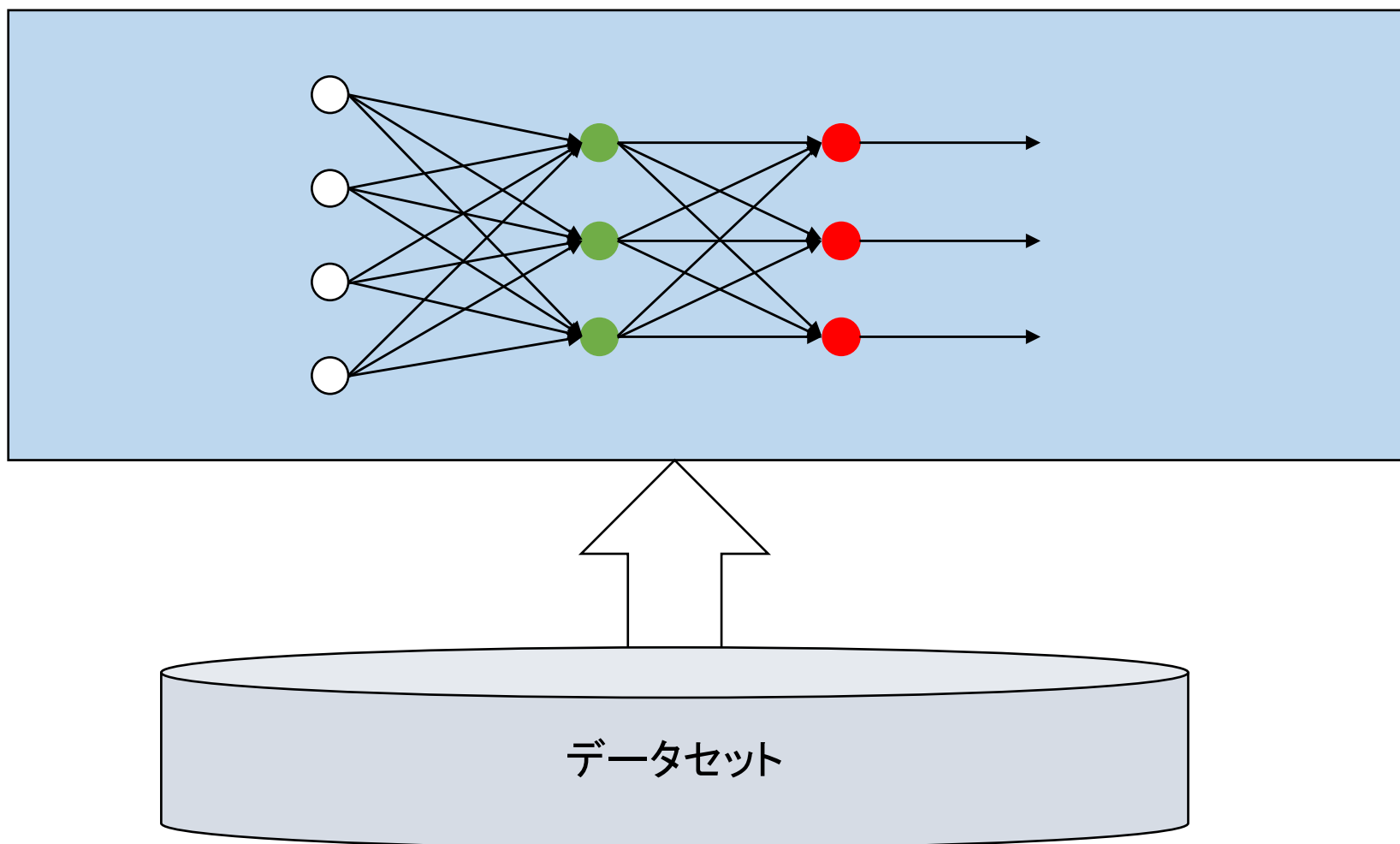
- パラメータサーバとは
- SIMGRIDとは
- 耐障害手法の実装
- シミュレーション
- まとめ

# 発表の概要

- パラメータサーバとは
- SIMGRIDとは
- 耐障害手法の実装
- シミュレーション
- まとめ

# 機械学習の基本

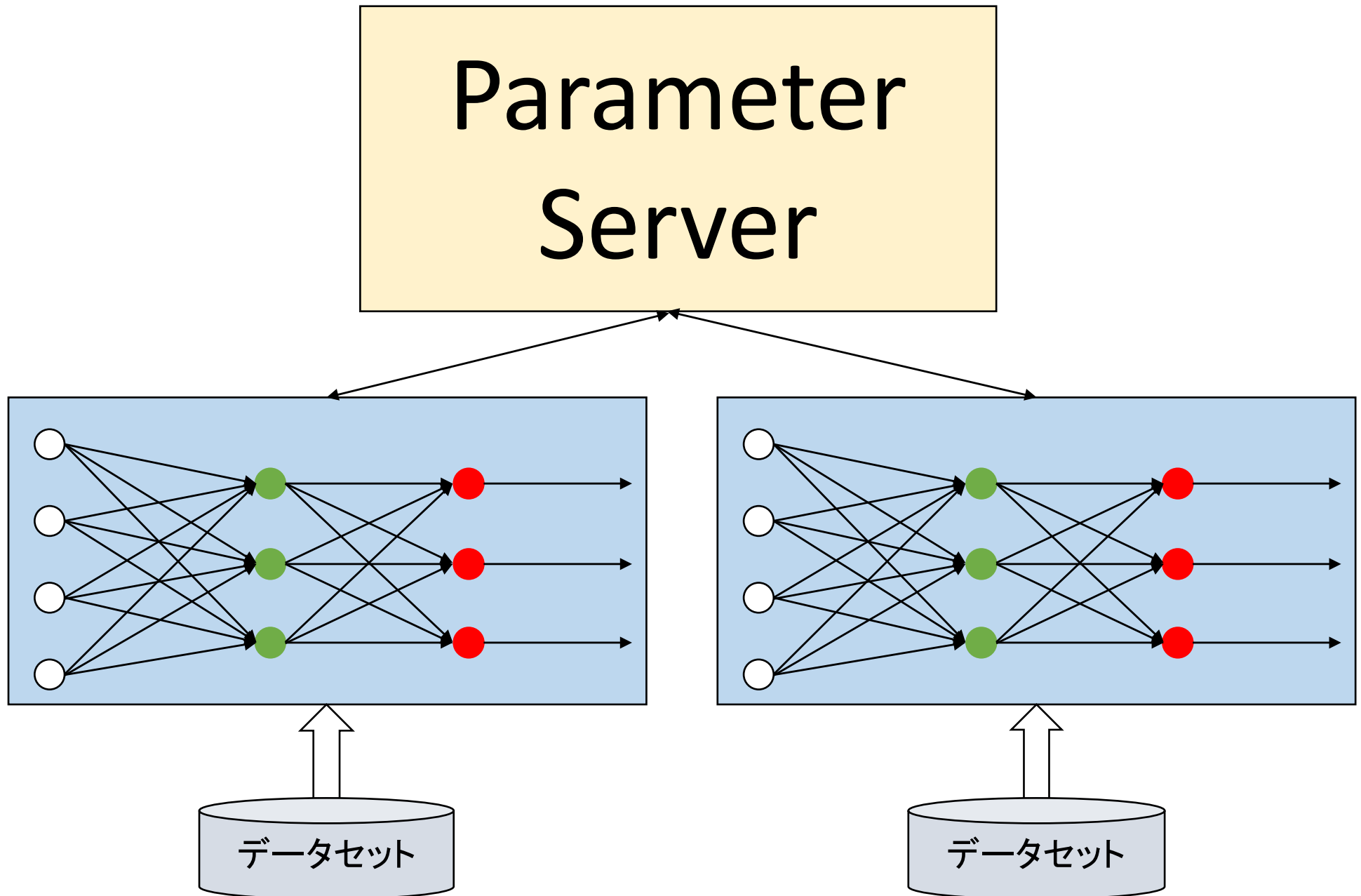
- データセットを用いてパラメータを更新



# 機械学習の並列化

- 機械学習の並列化は大別2種類
  - モデル並列
  - データ並列
- パラメータサーバ
  - データ並列機械学習システムのフレームワーク

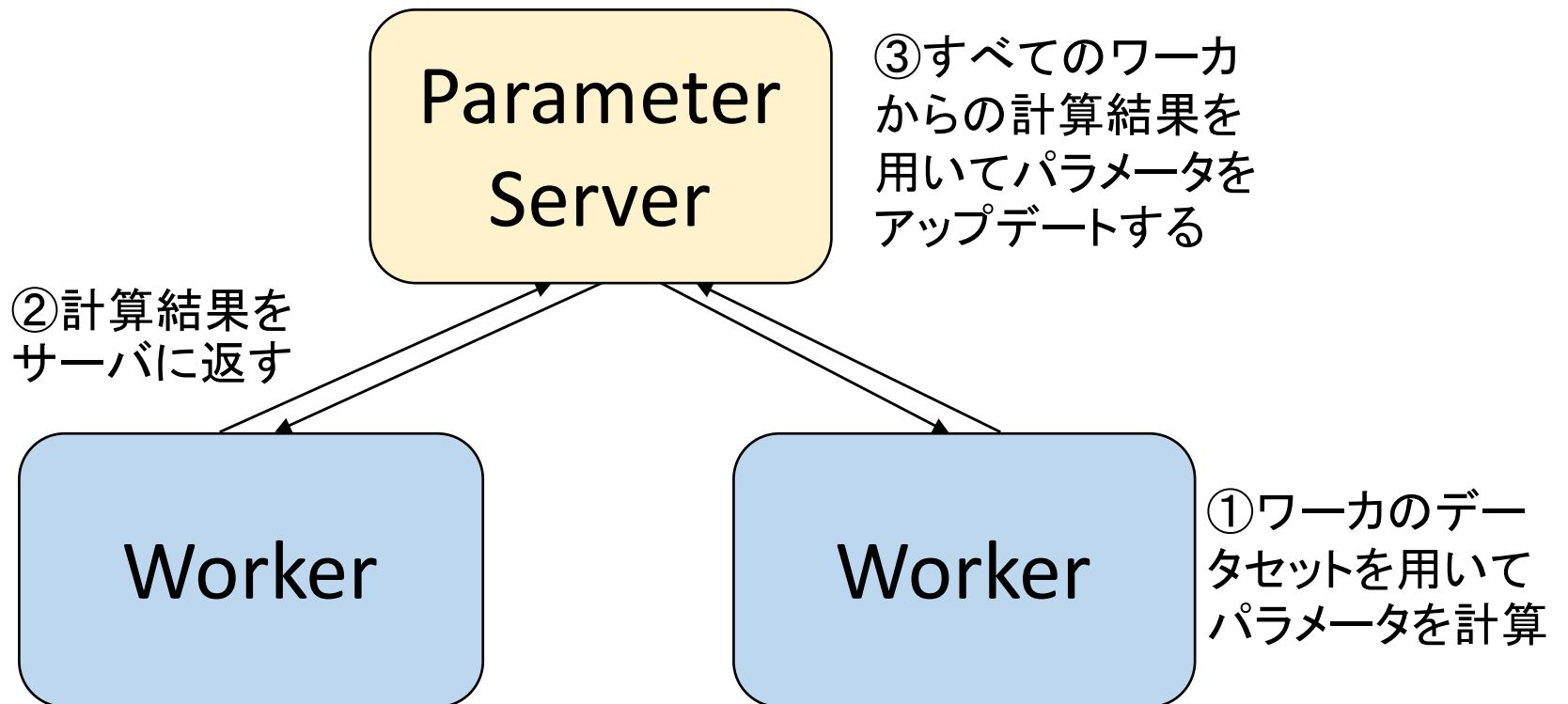
# データ並列





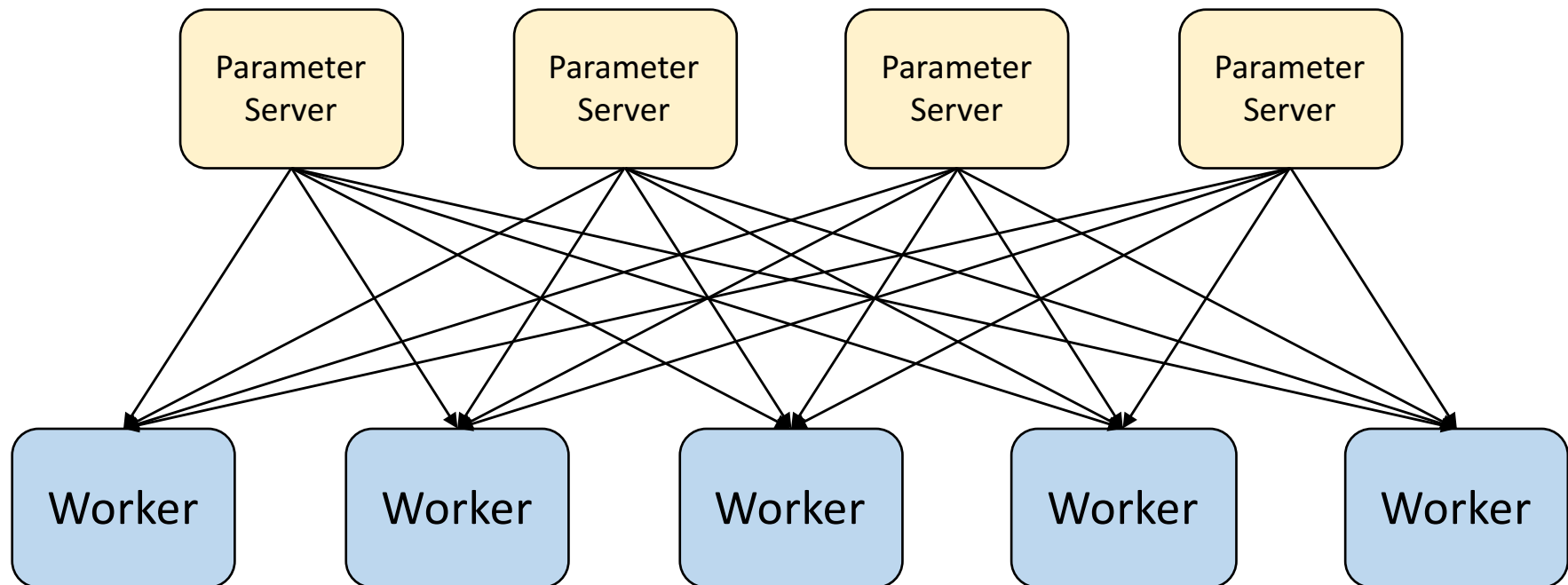
# パラメータサーバとは

- データ並列サーバ機構
- 大規模な機械学習システム向け



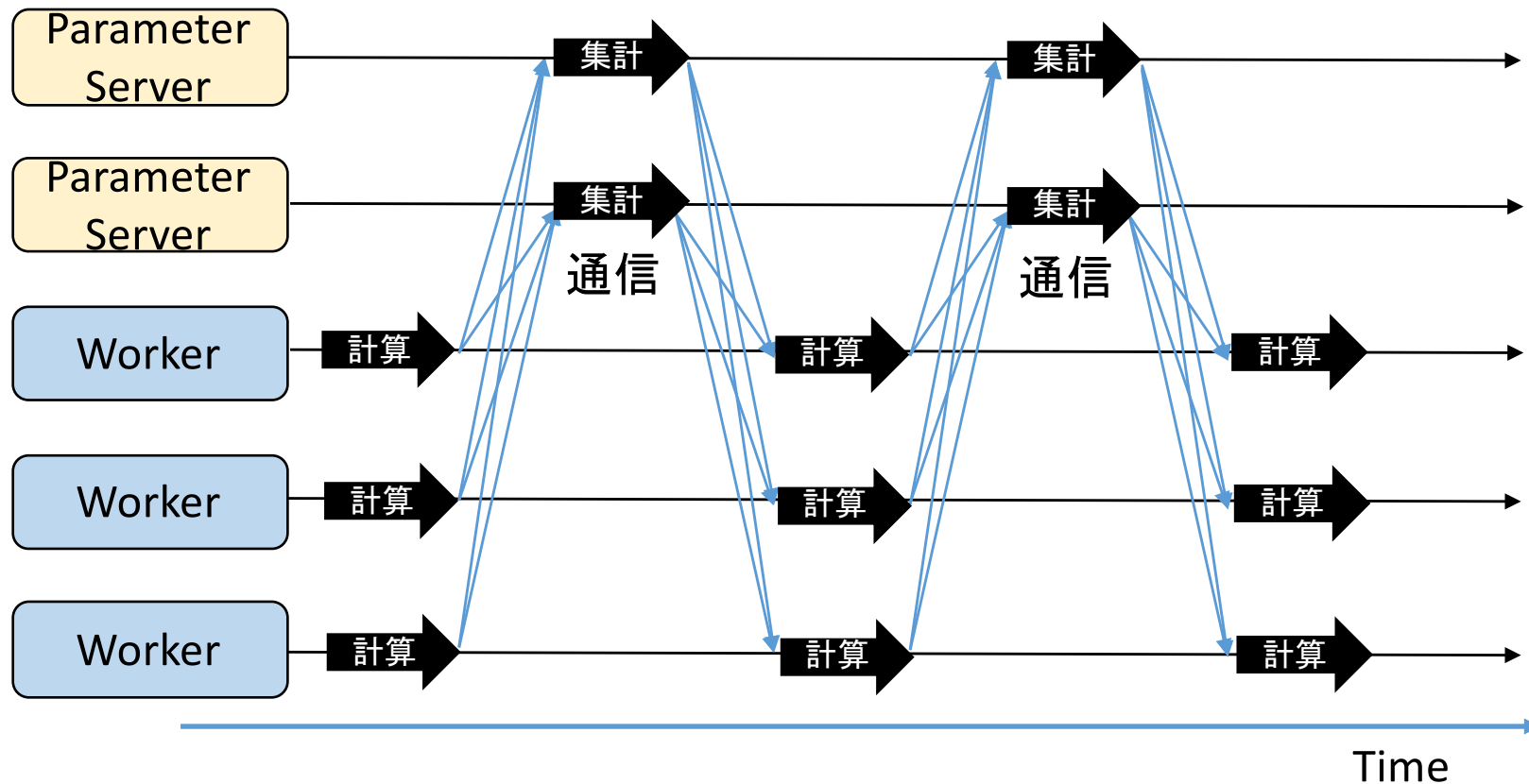
# パラメータサーバの構成

- サーバノードは複数
  - ボトルネック
  - フォールトトレランス



# パラメータサーバの通信パターン

- パラメータサーバ群とワーカ群間の通信パターン



# 発表の概要

- パラメータサーバとは
- SIMGRIDとは
- 耐障害手法の実装
- シミュレーション
- まとめ

# SIMGRIDとは

- シミュレーションフレームワーク
  - 並列計算アプリケーション向け
- アプリケーションシミュレータ
  - イベントの実際発生 **×**
  - イベントのコストだけ抽出 **✓**
- 並列機械学習システムにおける耐障害手法についてシミュレーションと考察

# 発表の概要

- パラメータサーバとは
- SIMGRIDとは
- 耐障害手法の実装
- シミュレーション
- まとめ

# 耐障害手法

- 障害の種類：
  - ネットワーク障害
  - ノード障害
    - パラメータサーバノード障害
    - ワーカーノード障害（確率が一番高い）
  - etc.

## 障害復旧の前提：バックアップノード

- ① システムがワーカーノードの故障を察知
- ② システムがバックアップノードを起動
- ③ バックアップノードが障害ワーカーの状態を引き継ぐ
- ④ 作業再開

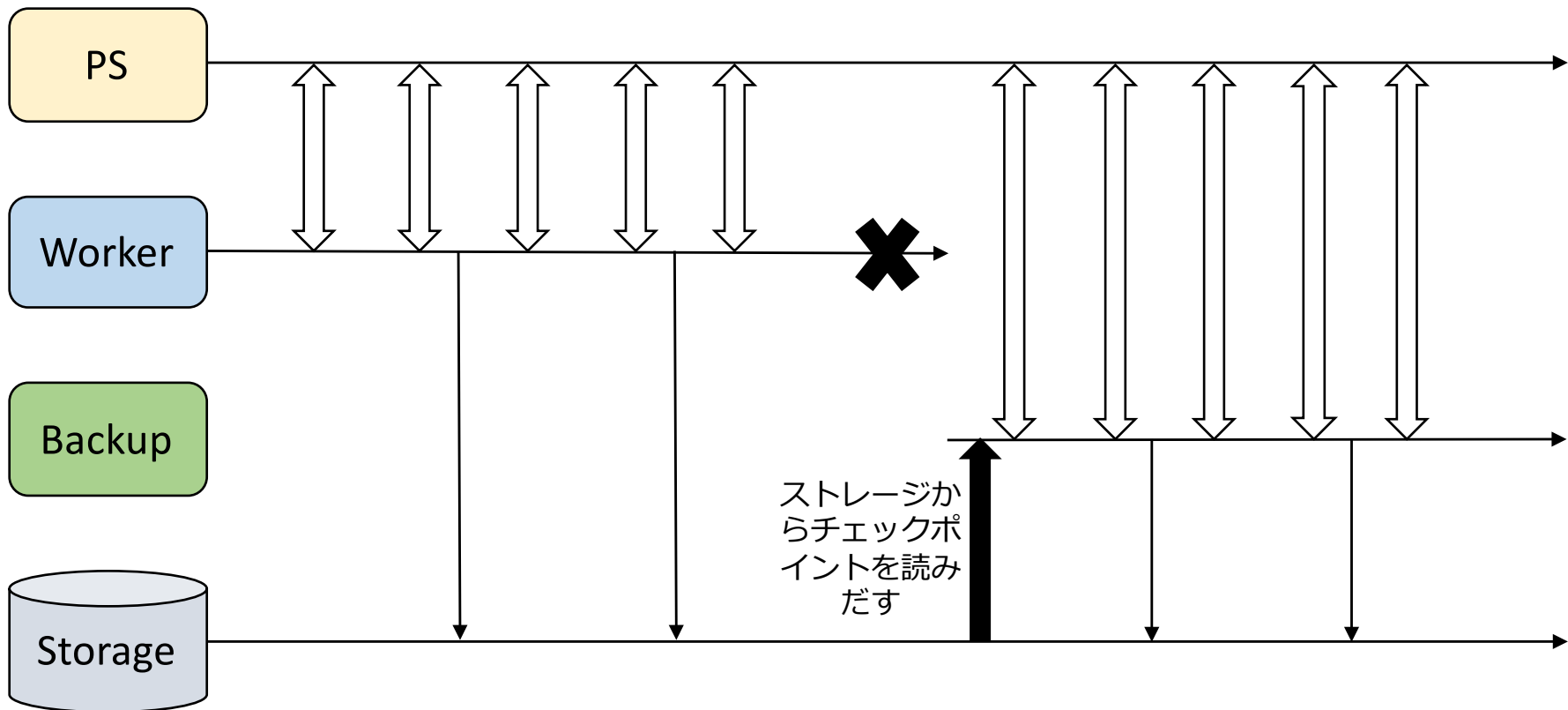
# 耐障害手法

- 3種類の耐障害手法：
  - ◆一般的な並列システムに使われている手法
    - CKPT法
  - ◆パラメータサーバの特性を生かした手法
    - PS法
    - IGNR法



# CKPT法

- チェックポイントを用いる手法
  - 多くの並列システムに使われている



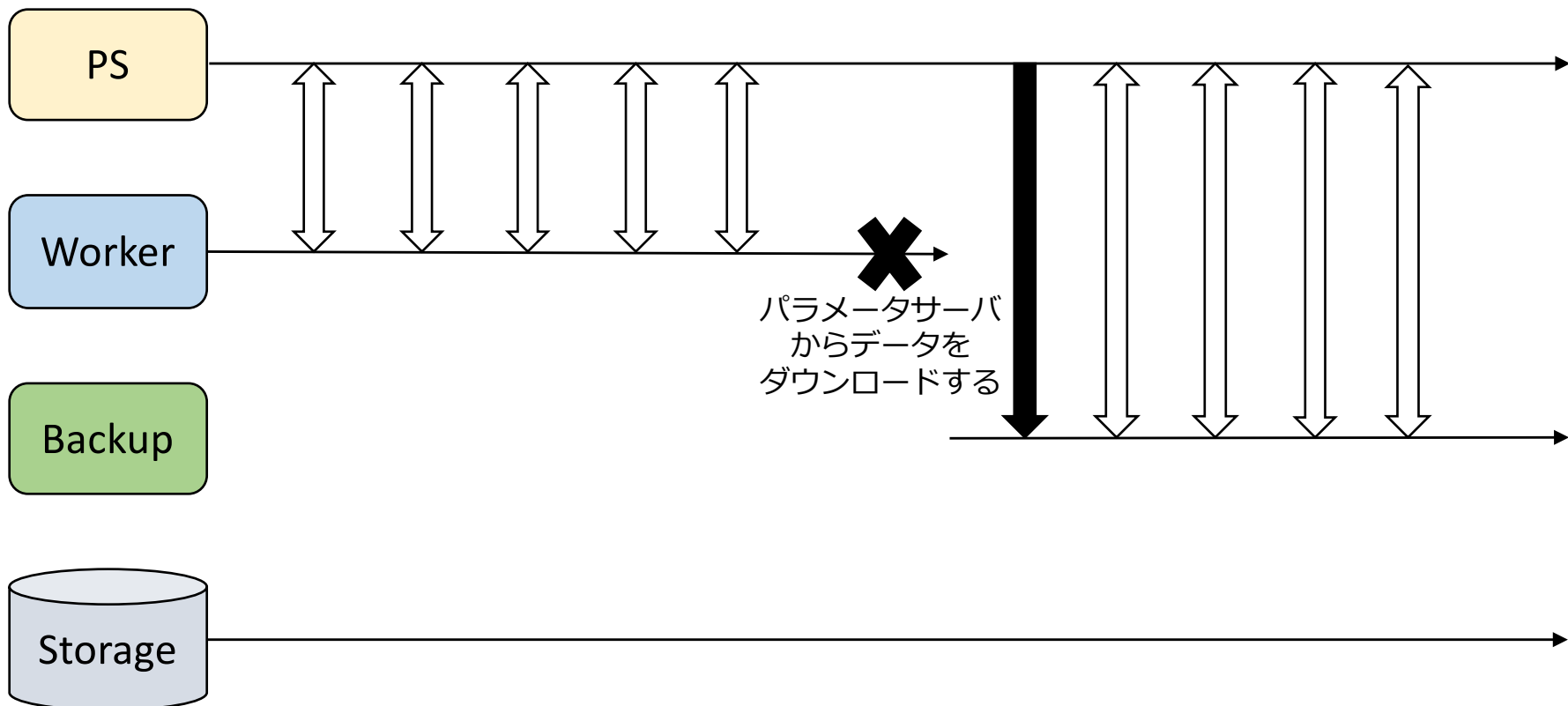
# CKPT法

## CKPT法の問題：

- ストレージにI/O操作が必要
  - 記録間隔が短い：I/Oコストが大幅に増加
  - 記録間隔が長い：障害による巻き戻しの被害大
- 他のワーカノードの状態の巻き戻し
  - 機械学習システムにとって不必要

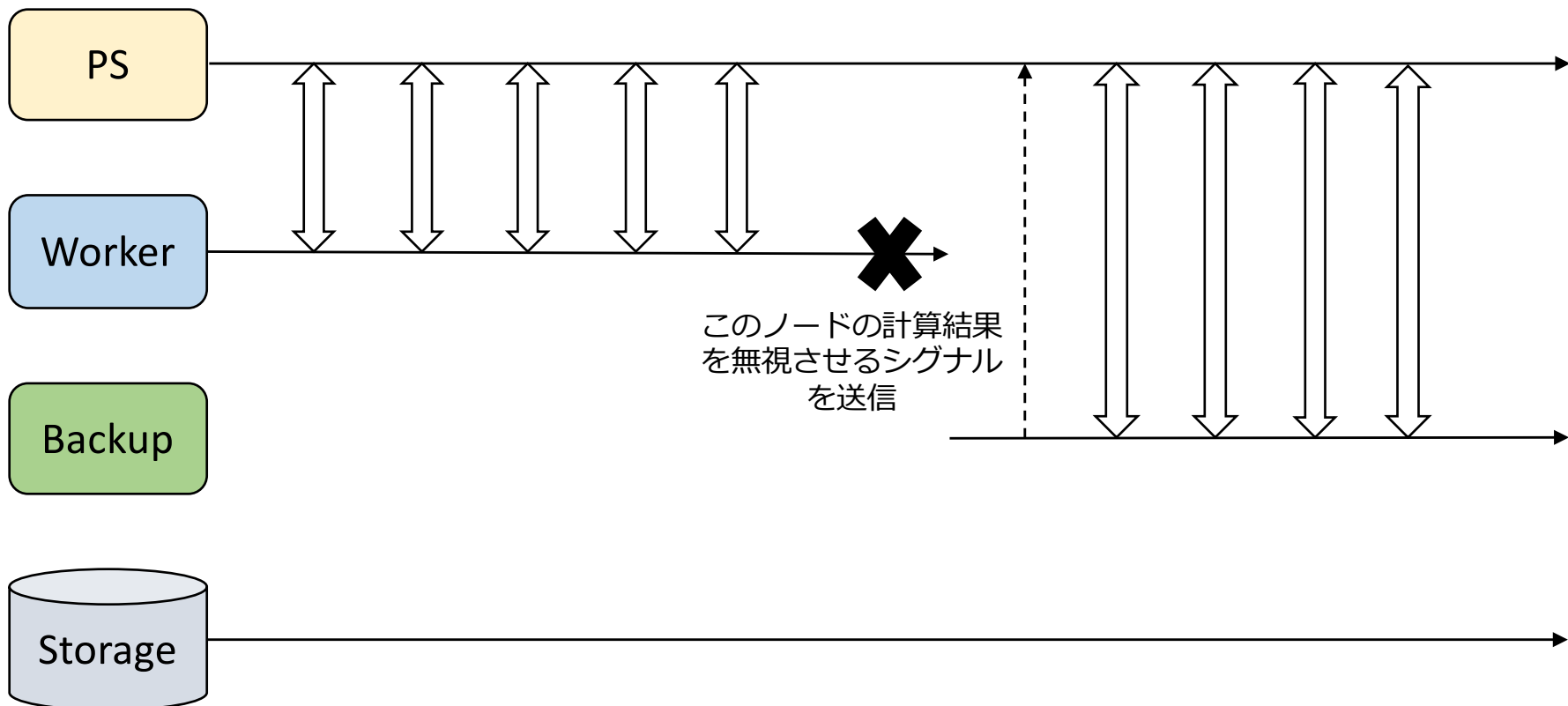
# PS法

- ワーカーのパラメータはパラメータサーバ群のデータと同一
  - 高冗長性
  - パラメータサーバ → チェックポイントサーバ



# IGNR法

- 更新情報：多数のワーカから  
一つのワーカの更新情報がなくても一回の学習に問題がない？

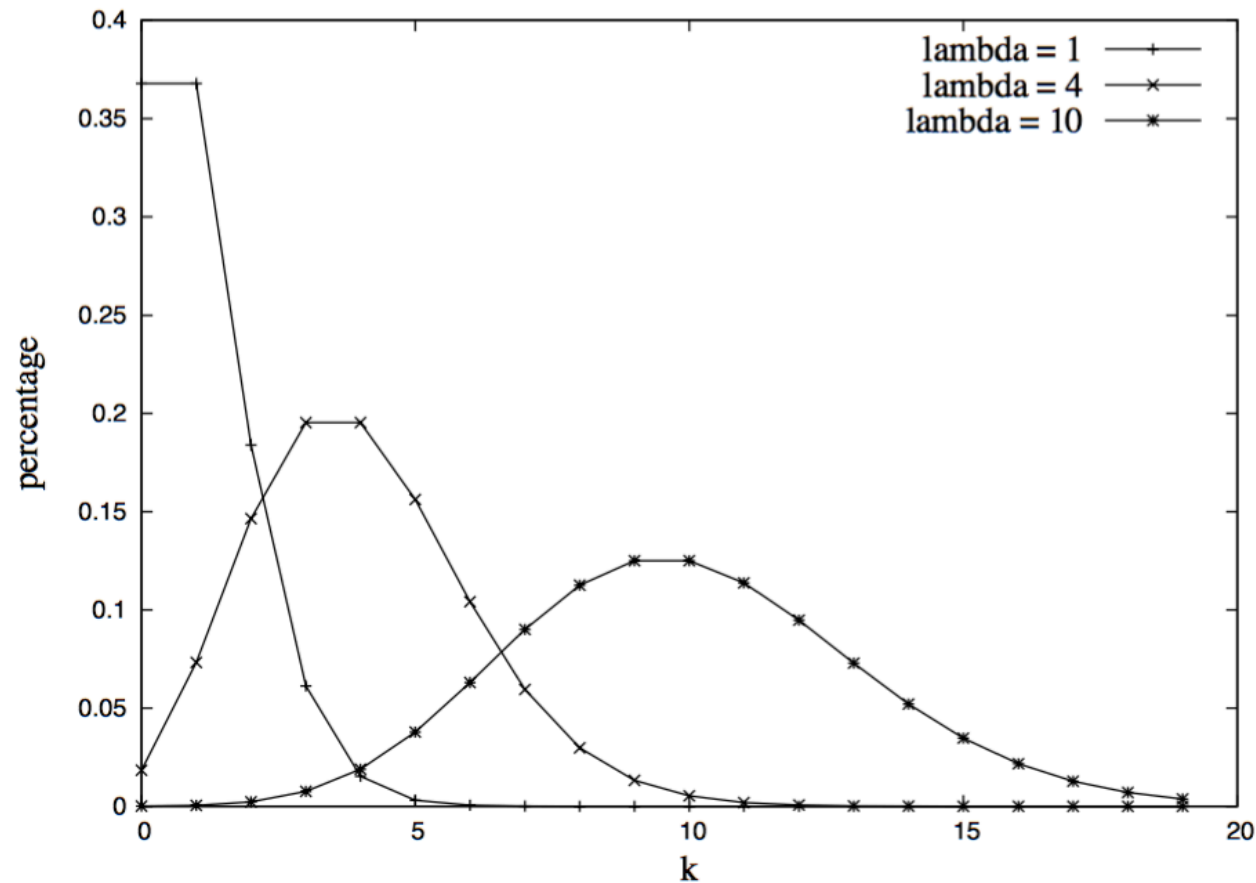


# 発表の概要

- パラメータサーバとは
- SIMGRIDとは
- 耐障害手法の実装
- シミュレーション
- まとめ

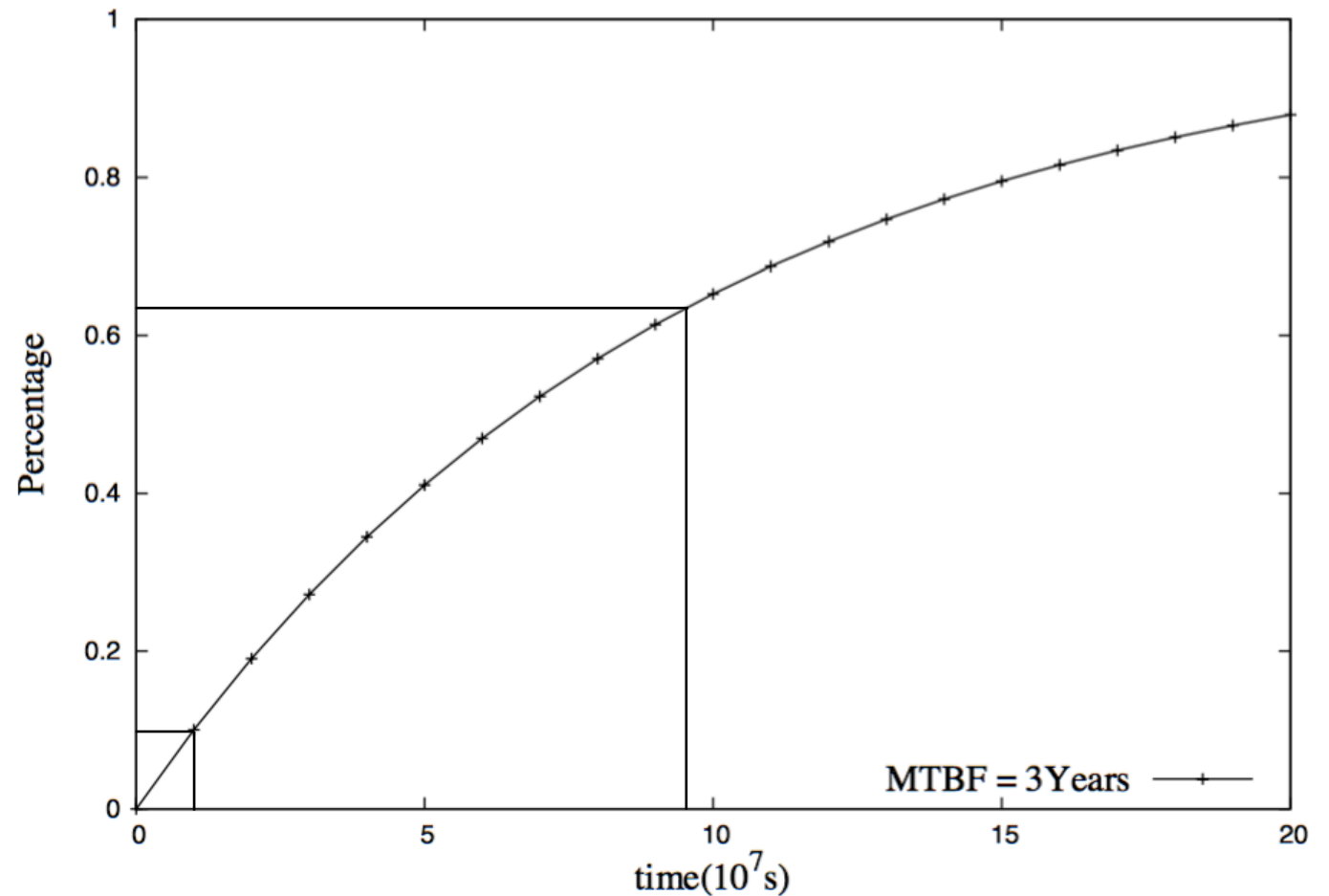
# 故障の発生確率

- 計算機の故障発生：ポアソン分布
  - 発生回数： $\lambda$ のポアソン分布
  - 発生間隔： $\lambda$ の指数分布



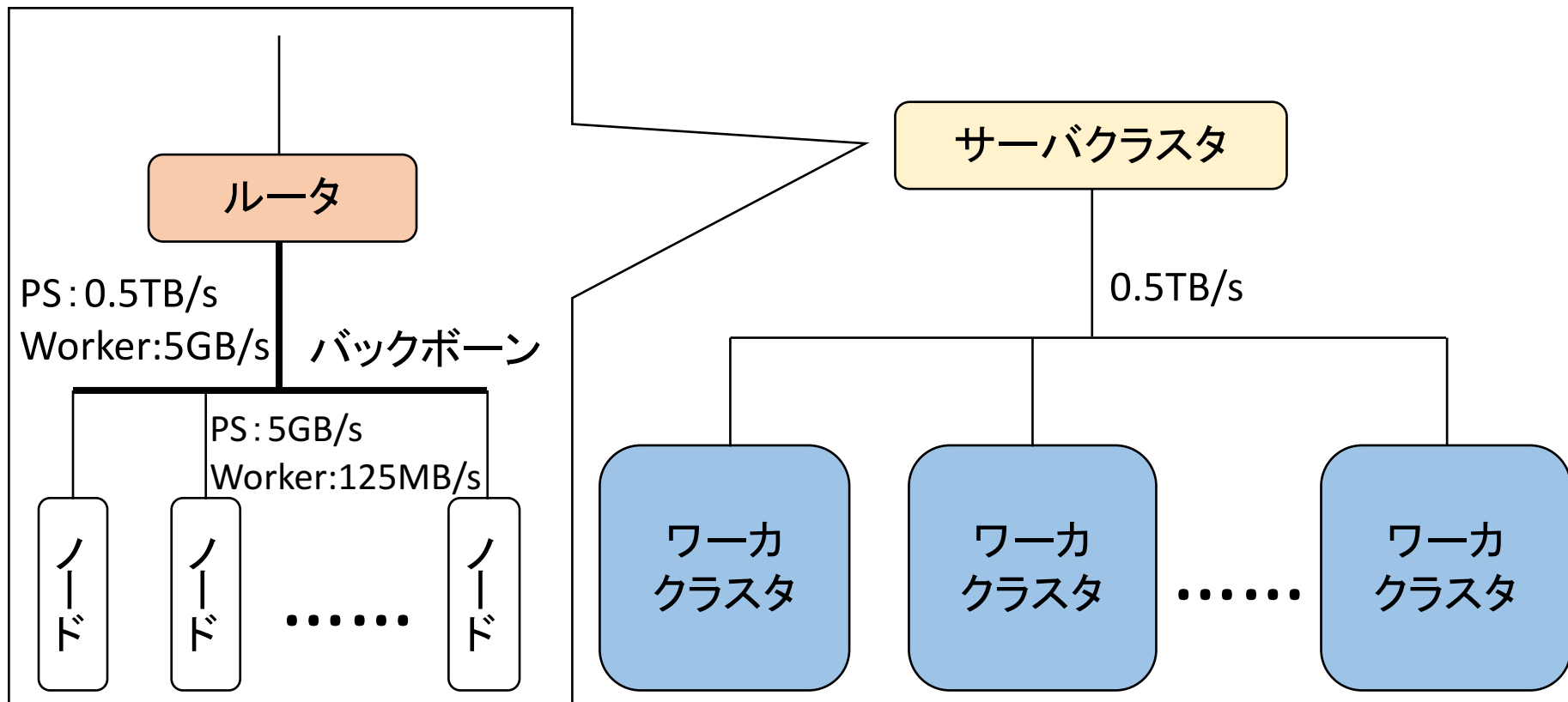
# 故障の発生間隔

- 指数分布の期待値 = 故障の平均間隔 =  $1 / \lambda$
- $\lambda = 1 / \text{故障の平均間隔}$
- MTBF



# 評価環境の構築

- サーバクラスタ : 1個
- ワーカクラスタ : 100個
- 各クラスタのノード数 : 100個





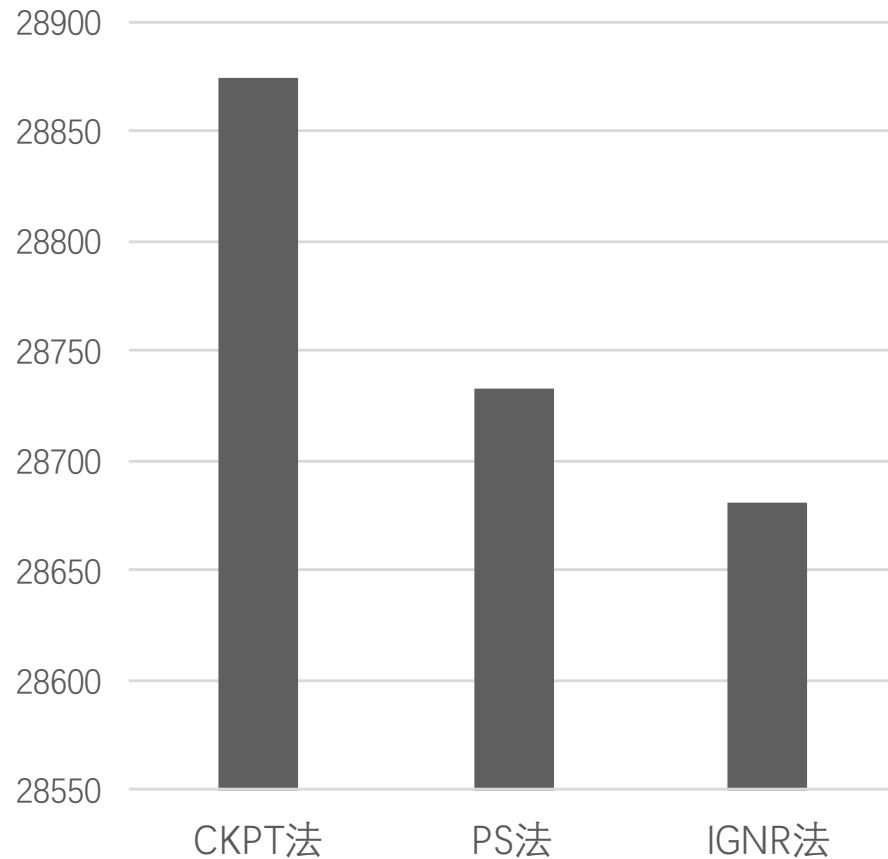
# 評価環境の構築

- ストレージはノードに付属すると仮定
  - ネットワークを経由せず読み書きが可能

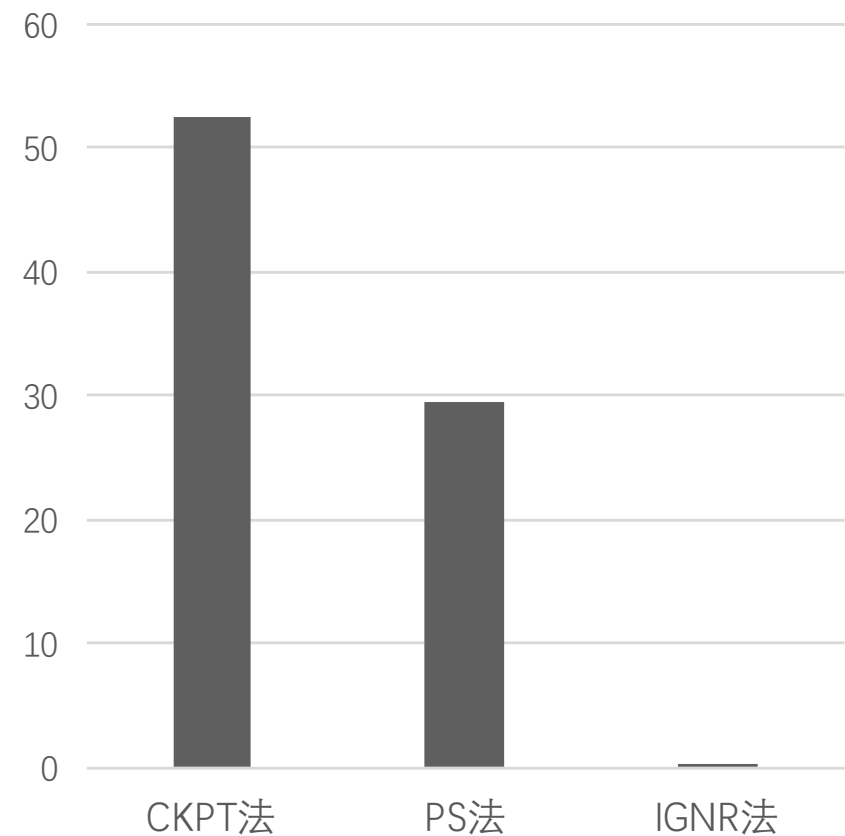
項目	詳細
学習タスク 1 個の処理時間	100s
学習モデルのサイズ	1GB
チェックポイントストレージの I/O 速度	1GB/s
チェックポイントのデータサイズ	8GB
ノードの起動時間	5s
計算機一機につきのMTBF	3Years = 94608000s

# 評価結果

三種類の手法に対してそれぞれ学習回数200回のシミュレーションを10回行った。

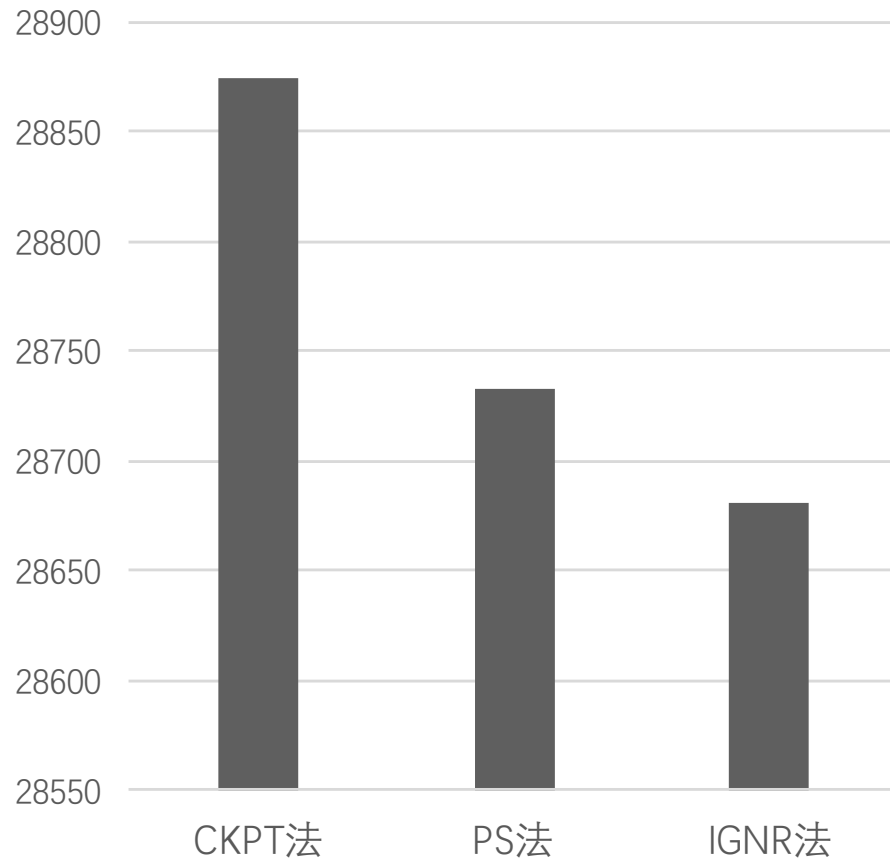


三つの手法の平均学習時間 (s)



三つの手法の学習時間の標準偏差 (s)

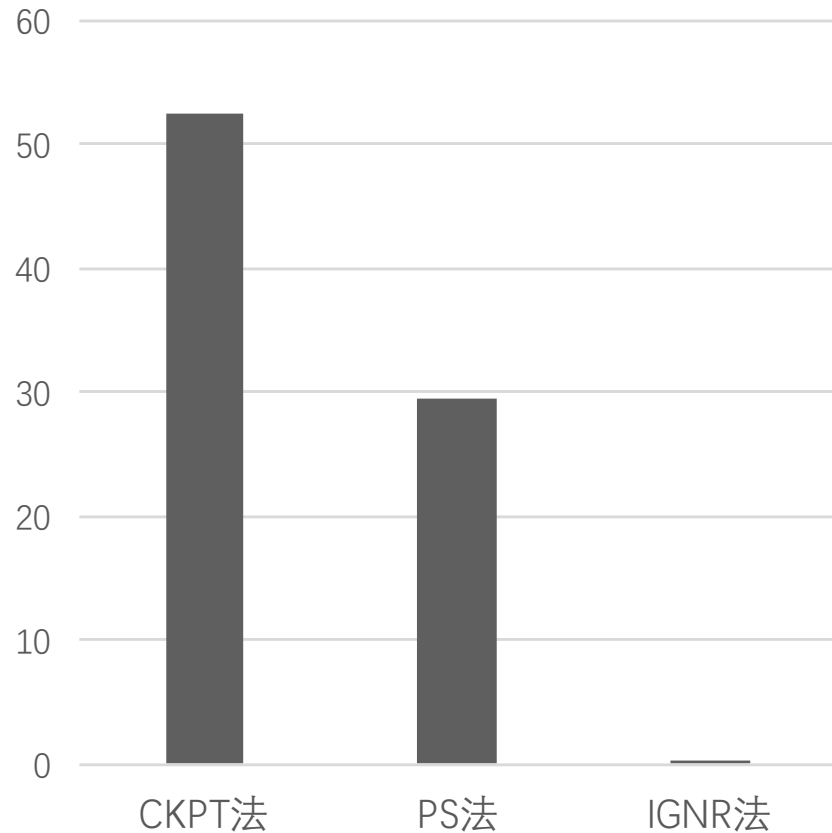
# 評価結果



CKPT法は学習時間が最も高い

- I/O操作コスト大
- システム全体の再起動による更新したパラメータの紛失

# 評価結果



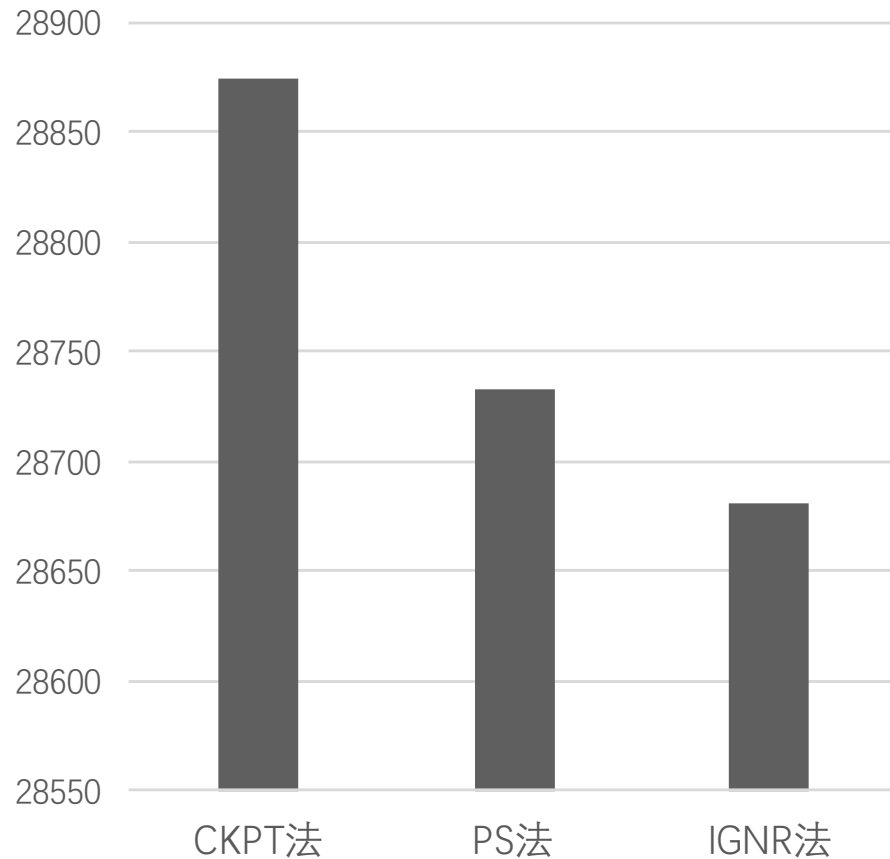
CKPT法の標準偏差値も一番高い

- 障害の復旧時間長
- 障害によるデータの紛失量は障害の起こるタイミングによる影響大

# 評価結果

一回の障害による学習時間の増加

手法	CKPT法	PS法	IGNR法
増加時間 (s)	78.1	12.6	-0.5123



IGNR法の学習総時間は障害が起きると下降

- パラメータの通信コストがない
- 学習に影響？

# 発表の概要

- パラメータサーバとは
- SIMGRIDとは
- 耐障害手法の実装
- シミュレーション
- まとめ

# 結論

- パラメータサーバ用いた並列機械学習システムに固有の冗長性を利用した耐故障性手法
  - 一般に広く用いられている定期的なチェックポイントを用いる手法と比較
  - コストの大幅な低減を確認
- 復旧のパラメータ通信コストも低減
  - 解の収束への悪影響の可能性

# 今後の課題

- ワーカーノードだけではなく、サーバノードとネットワークの障害を考慮
- より現実的なチェックポイントストレージ設定での性能解析
- さまざまなネットワーク環境での評価
- IGNR法がもたらす、学習全体の遅延の定量的解析



# 謝辞

この成果の一部は、国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の委託業務の結果得られたものです。本研究は JSPS 科研費 JP16K00116 の助成を受けたものです。

ご清聴ありがとうございました。