



# 分散連合機械学習： 基礎，動向および無線設計から見た課題

2022年5月18日(水)

IEICE IT/EMM合同研究会@岐阜

佐藤 光哉

電気通信大学 人工知能先端研究センター

E-mail: [k\\_sato@aix.uec.ac.jp](mailto:k_sato@aix.uec.ac.jp)

HP: <https://www.ksatolab.aix.uec.ac.jp/>

※本スライドは当日の発表スライドの一部を改変・削除したものである

## 本講演の目的

「**本分野の新規参入者を増やす**」

## 講演内容

- 分散型の学習方式の概要
- 通信面での課題
- 研究事例：伝送レート適応化による高速高精度なDecentralized FL
- 研究事例：(web公開版未収録)
- 研究動向

チュートリアル的な位置付けでの講演となります

## 主題

- 分散機械学習とは？
- 無線設計からみた課題は？
- 関連する研究動向は？
- 本分野に興味を持った大学院生を想定して書きました
  - 本講演内容の多くも本論文がベース
- 関係者各位に感謝します

解 説 論 文

### 無線設計の問題として見る 分散連合機械学習

Decentralized Federated Learning over Wireless Channels: A Review

佐藤光哉 Koya SATO

アブストラクト 機械学習の発展は目覚ましく、理論と実用双方から幅広い研究が行われている。一方、機械学習のアプリケーション拡大に伴い、データプライバシーや通信コストに対する懸念も顕著になっている。本稿では、こういった問題の抜本的解決とビッグデータ解析の両立に向けた方式として近年注目されている、連合機械学習 (FL: Federated Learning) をはじめとした分散的な機械学習方式について概説する。特に、集中制御サーバを設けない FL である分散連合機械学習 (DFL: Decentralized FL) に着目し、無線通信路における問題とその解決に向けた方策を述べる。

キーワード 分散連合機械学習, 無線ネットワーク, 伝送レート適応化

**Abstract** Machine learning has made remarkable progress, and a wide range of research has been conducted from theoretical and practical perspectives. However, the expansion of machine learning applications has raised novel concerns, such as data privacy and communication costs. This article outlines distributed machine learning methods such as federated learning (FL), which has been attracting attention in recent years as a method that can both fundamentally solve these problems and analyze big data. Then, we focus on decentralized FL (DFL), which does not have a centralized server. We describe the problems in wireless communication channels and countermeasures to solve them.

**Key words** Decentralized federated learning, wireless networks, rate adaptation

1. はじめに

単独の計算機での機械学習について考える。具体的に、パラメータベクトル  $\mathbf{x}$  から構成されるニューラルネットワーク (NN: Neural Network) のような学習器を、学習データセット  $\mathcal{D}$  から構築する。データサンプル  $\xi \in \mathcal{D}$  をこの学習器へ入力した際の損失関数を  $L(\mathbf{x}; \xi)$  と定義すると、本機械学習は、以下の問題を解くようパラメータベクトル  $\mathbf{x}$  を調整する問題と表すことができる。

$$\min_{\mathbf{x} \in \mathbb{R}^m} \left[ f(\mathbf{x}) := \frac{1}{|\mathcal{D}|} \sum_{i=1}^{|\mathcal{D}|} L(\mathbf{x}; \xi_i) \right] \quad (1)$$

ここで、 $m$  は学習器中のパラメータ数、 $|\mathcal{D}|$  は学習データ数、 $\xi_i$  は  $\mathcal{D}$  中の  $i$  番目のデータである。

本問題に対する最も基本的なアプローチは、確率的勾配降下法 (SGD: Stochastic Gradient Descent) であろう。本手法では、学習データのランダムサンプリングと損失関数に対する勾配の計算、モデルパラメータの更新を繰り返すことで問題 (1) を解く (アルゴリズム 1)。データ 1 つ 1 つに対する演算を繰り返すことで、データ数が膨大であったり、後から学習データを追加取得した場合でも効率よく学習できる。

機械学習における手法はほとんどがこのような繰り返し演算に基づくといっても過言ではなく、クラウド上でのビッグデータ解析についても、やはりどこかで何らかの計算機がデータセット  $\mathcal{D}$  を解析することになる。

ここで、繰り返しにはなるが、以上のステップは単独の計算機上の演算によるものである。クラウドを活用したデータ解析においては、学習対象であるデータセット  $\mathcal{D}$  は無数のユーザから収集したデータから構成されるものであることから、自ずと以下のような状態が生じる (図 1)：

**データプライバシー** 学習対象が画像やログといったプライベートな情報であるほど、クラウドからの情報流出に対するリスクは大きくなる。

**Algorithm 1** SGD による機械学習

Require: モデルパラメータの初期ベクトル  $\mathbf{x}_1$ , 学習率  $\eta$ , 総学習回数  $K$ .

- 1: for  $k = 1, 2, 3, \dots, K$  do
- 2: データセット  $\mathcal{D}$  からランダムに学習サンプルを 1 つ抽出し、 $\xi_k$  とする。
- 3: 損失関数  $L(\mathbf{x}_k; \xi_k)$  を計算する。
- 4: 勾配ベクトル  $\nabla L(\mathbf{x}_k; \xi_k)$  を計算する。
- 5:  $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k - \eta \nabla L(\mathbf{x}_k; \xi_k)$  によりパラメータベクトルを更新する。
- 6: end for

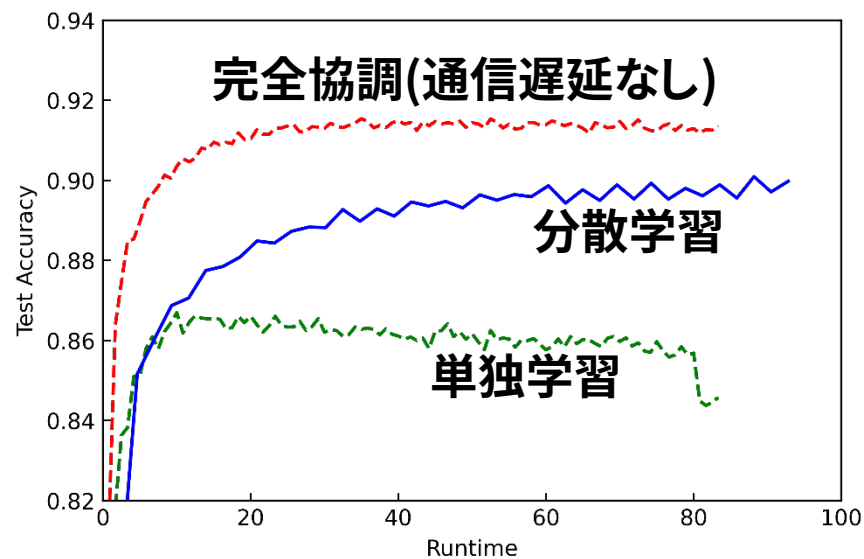
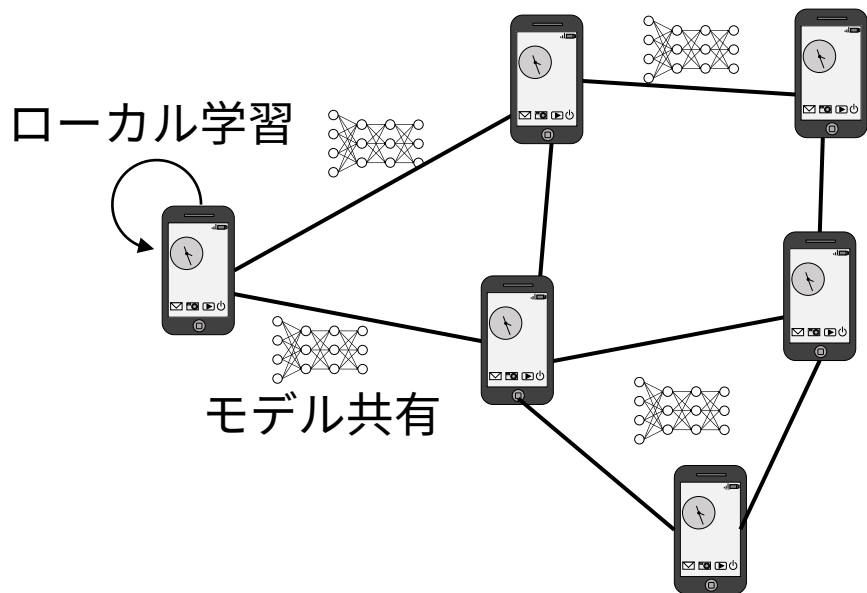
佐藤光哉 正員 電気通信大学人工知能先端研究センター  
E-mail: k\_sato@ieice.org  
Koya SATO, Member (Artificial Intelligent eXploration Research Center, University of Electro-Communications, 1-5-1 Chofugaoka, Chofu-shi, 182-8585 Japan).

電子情報通信学会 基礎・境界ソサイエティ  
Fundamentals Review Vol.xx No.xx pp.1-10 xxxxx 年 xx 月  
© 電子情報通信学会 xxxxx

IEICE Fundamentals Review Vol.xx No.xx 1

佐藤, "無線設計の問題として見る分散連合機械学習," 2022年7月公開予定

# 主要テーマ1: (分散)連合機械学習の設計 (本講演, 2019-) 4

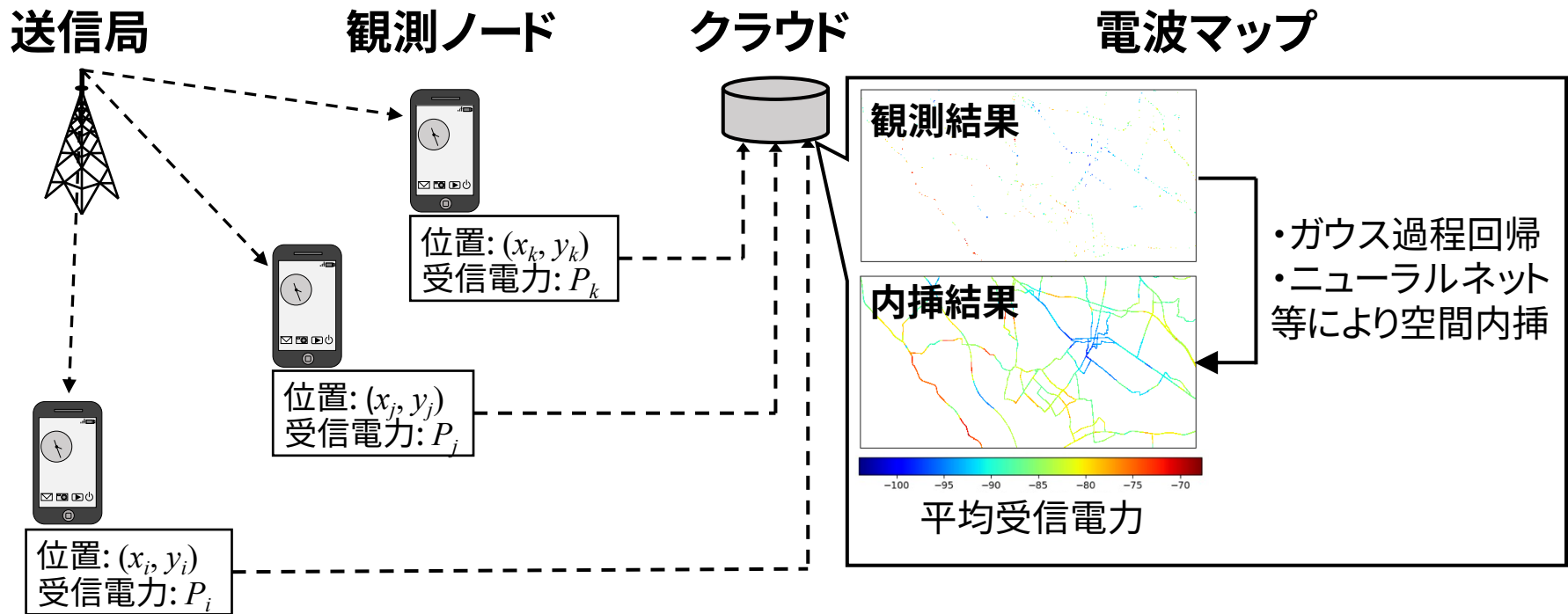


## ローカル学習と周辺端末との学習機の共有を繰り返す

- 端末がデータを持ち寄り、狭いネットワーク上で事象を学習
- 特に、高速高精度な通信/学習設計に従事

## 関連発表

- K. Sato+, IEEE Trans. Cogn. Commun. Netw., Dec. 2021.
- K. Sato+, IEEE ICC 2020, June 2020.



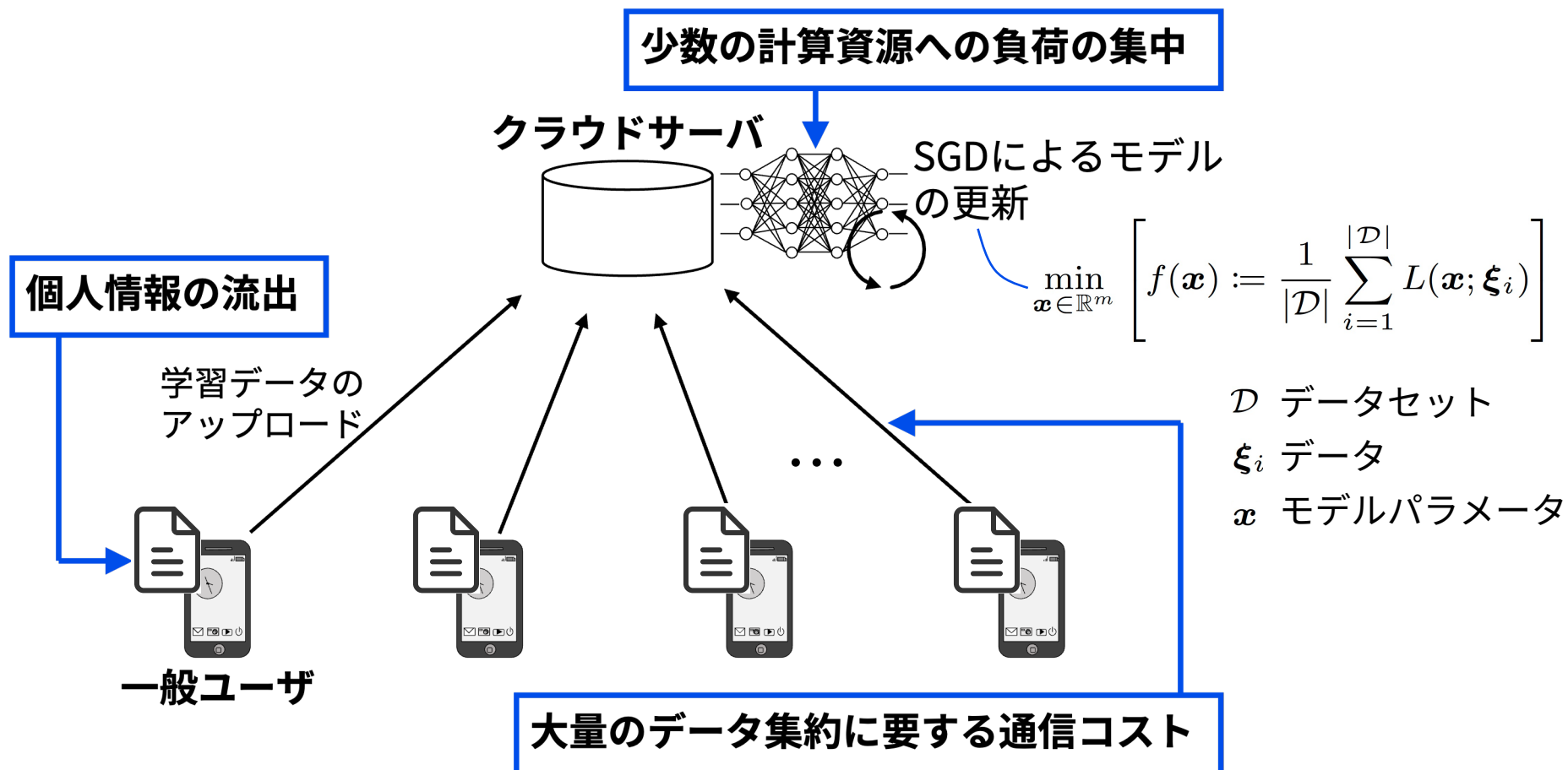
端末が移動観測した結果を集約し**受信電力を可視化**

➤ 応用例: 無線リソース最適化, UAVの経路設計, 位置指紋, ...

## 関連発表

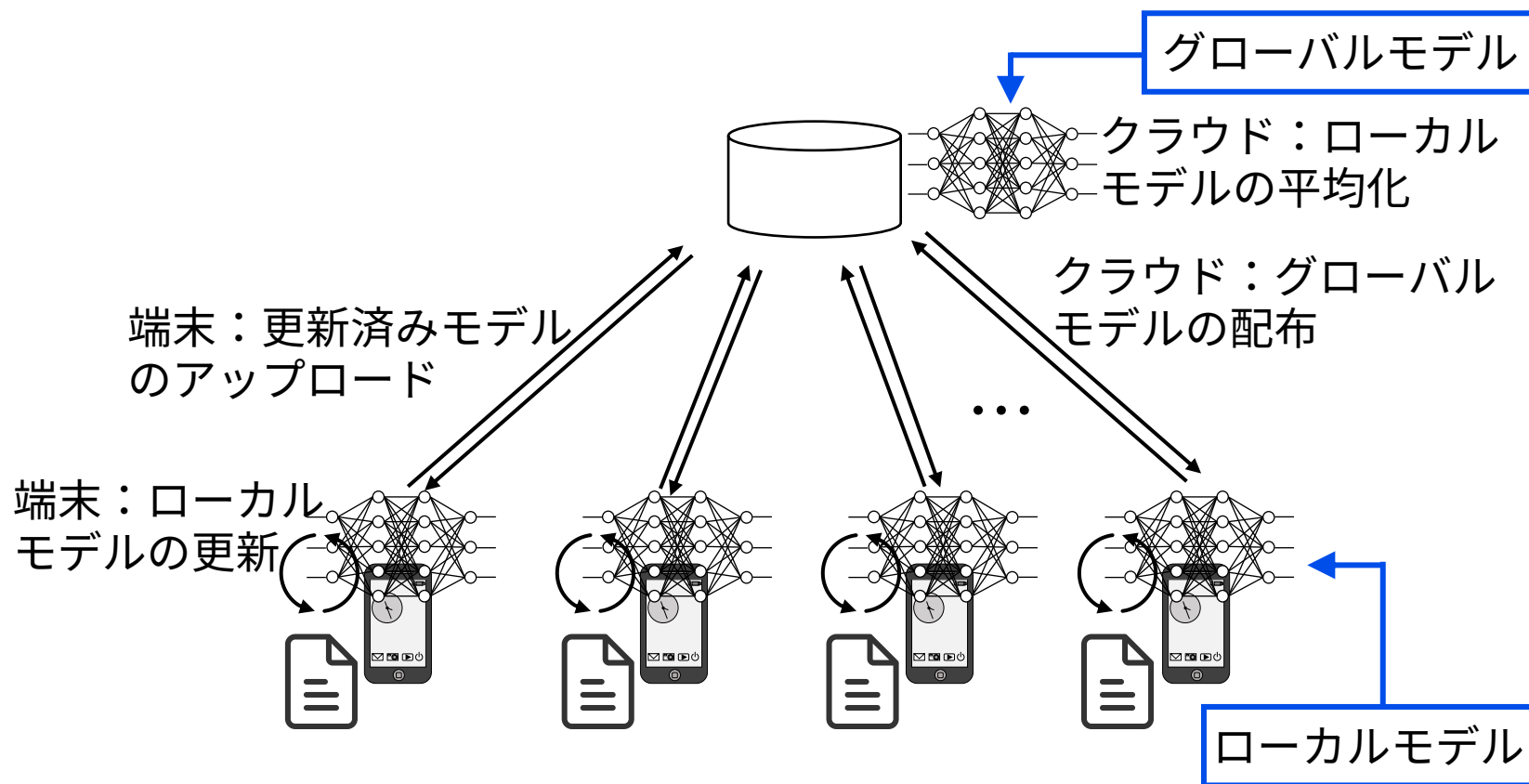
- 佐藤光哉, 招待講演, IEICE 信号処理研究会, Aug. 2021.
- K. Sato+, IEEE Trans. Veh. Technol., Jan. 2021.
- K. Sato+, IEEE Trans. Cogn. Commun. Netw., March 2017.

- 分散型の学習方式の概要
- 通信面での課題
- 研究事例：伝送レート適応化による高速高精度なDecentralized FL
- 研究事例：(web公開版未収録)
- 研究動向
- 終わりに



SGD: Stochastic Gradient Descent

上記の問題をクリアしつつデータ解析を実現する方法はないか？

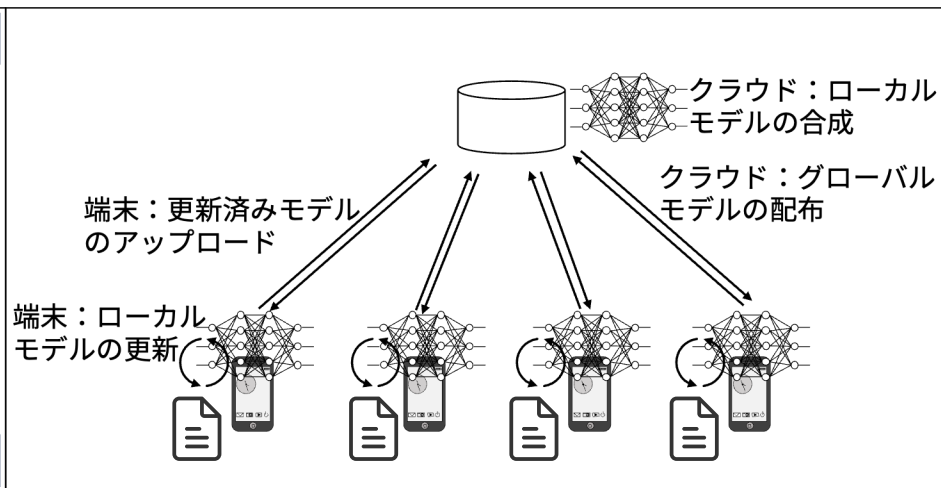
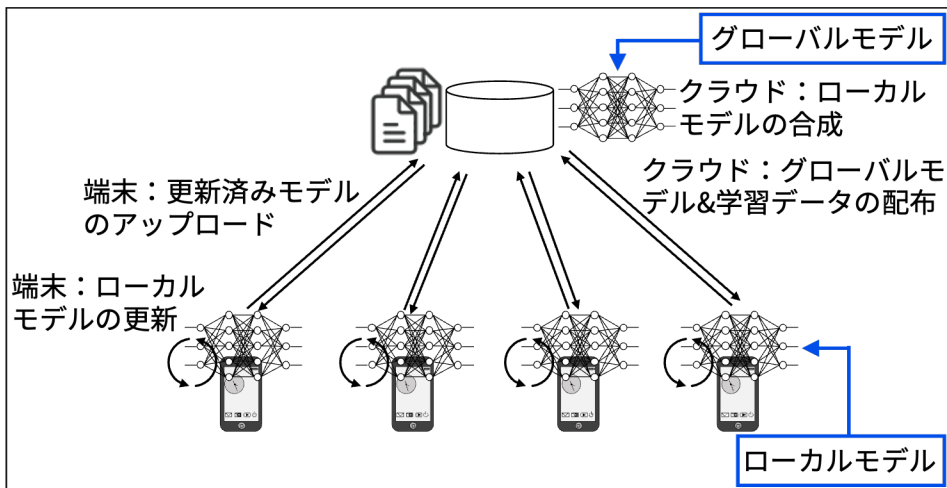


- AISTATS2017で提唱された学習コンセプト
  - [McMahan+, AISTATS2017] [Kairouz+, FTML2021]
- **端末ローカルでのSGDとクラウド側でのモデル合成を繰り返す**
  - **端末からのデータ開示を避ける**ことで前述の問題を対策

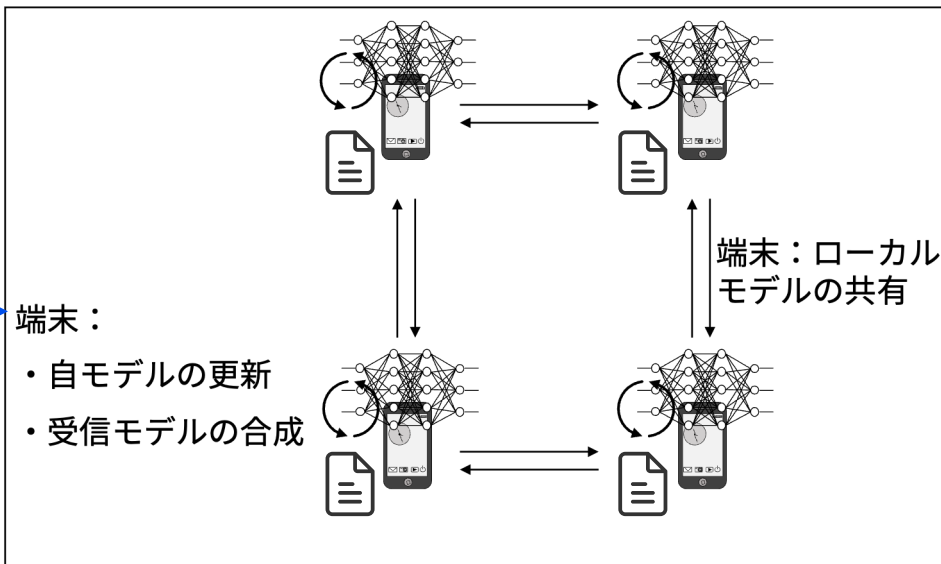


## 分散機械学習 (Distributed ML)

## 連合機械学習 (FL)



## 分散連合機械学習 (Decentralized FL)



- サーバからデータを配布
- 単独計算機上での複数GPUを利用した学習もこれに該当

- データの開示なし
- 集中制御サーバもなし

## Distributed ML

- Parallelized SGD: [Zinkevich+, NeurIPS2010]

## FL

- FedAvg: [McMahan+, AISTATS2017] **本分野のブームの火付け役**
- FedProx: [Li+, MLSys2020]

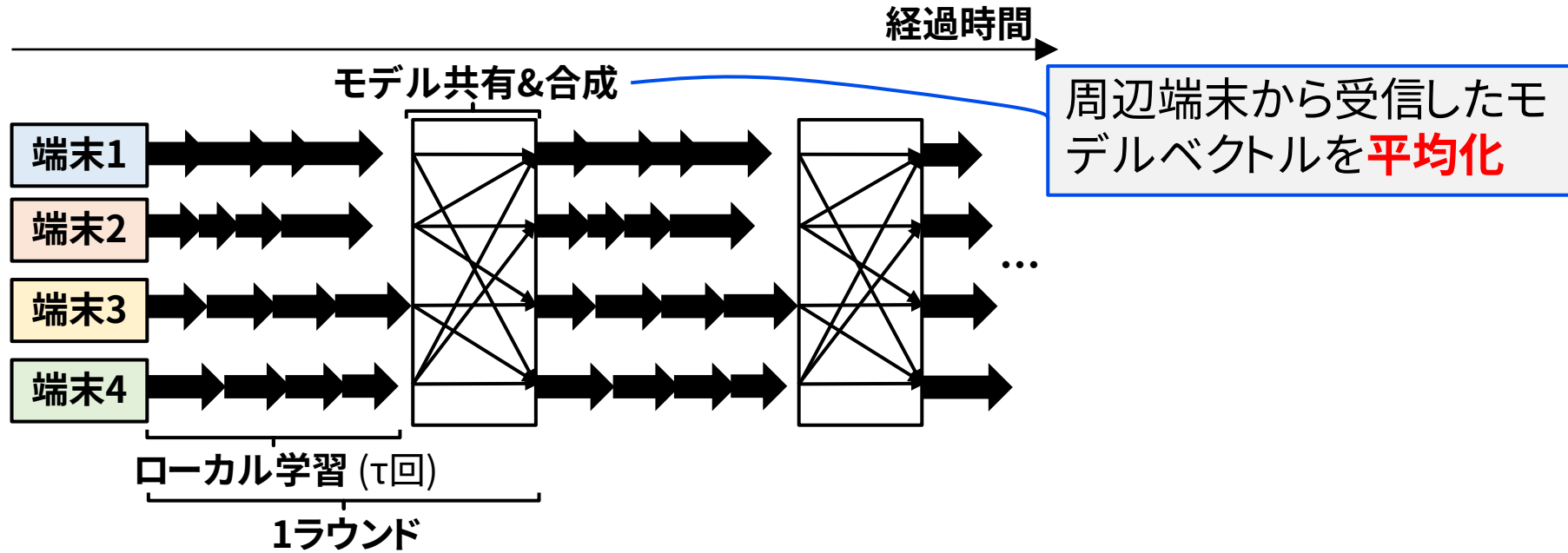
## Decentralized FL

- Decentralized Parallel SGD: [Lian+, NeurIPS2018]
- Asynchronous D-PSGD: [Lian+, ICML2018]

## 各種Centralized/Decentralizedアルゴリズムの統一フレームワーク

- Cooperative SGD: [Wang+, ICML Workshop2019]
- いずれも**端末でのSGDとモデルの合成の繰り返し**が基本

# Cooperative SGD(c-SGD)によるDecentralized FL 11

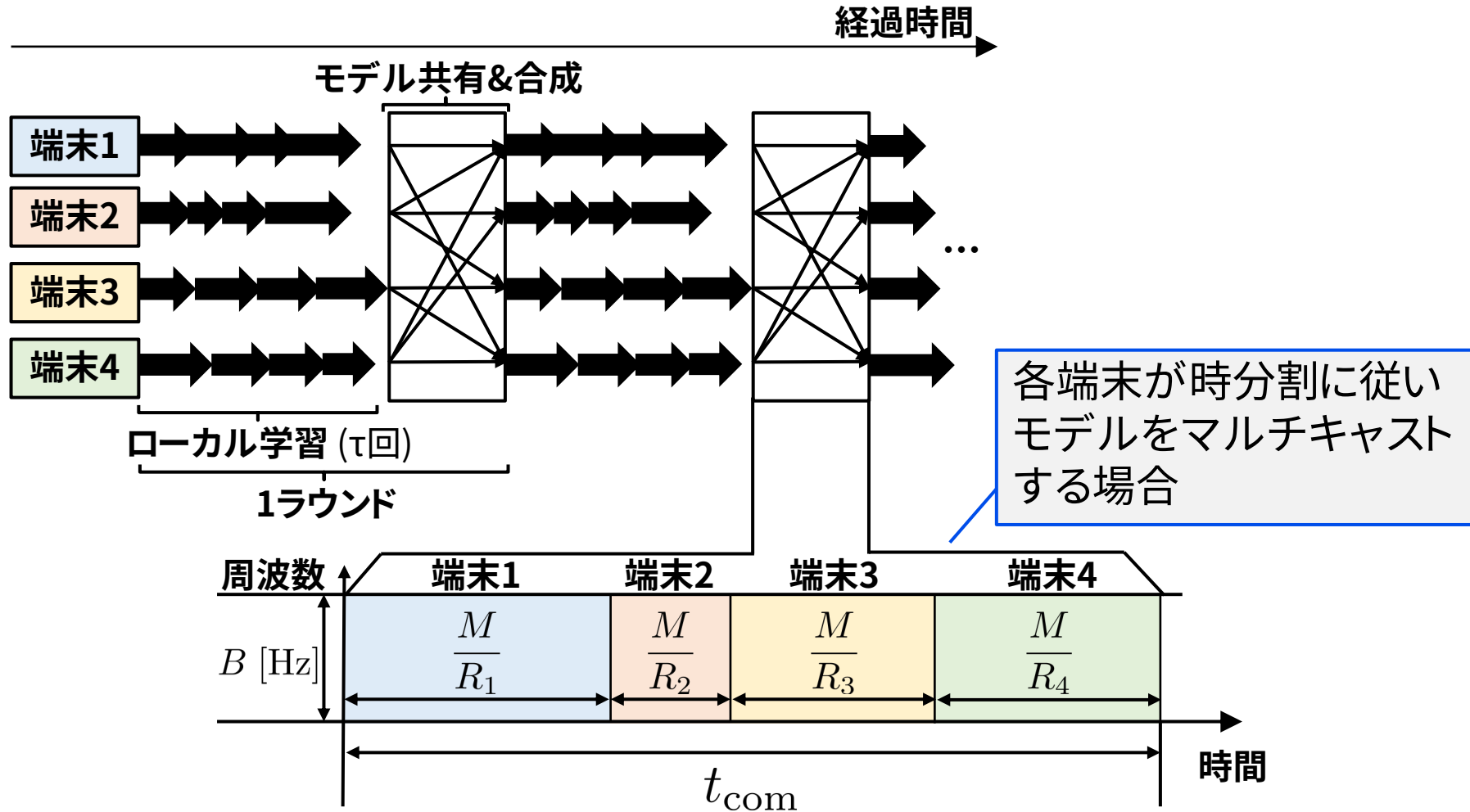


$\tau$  ラウンドごとのローカルSGDの回数

$M$  学習器のモデルサイズ[bits]

$t_{\text{com}}$  1ラウンドに要する通信時間

# Cooperative SGD(c-SGD)によるDecentralized FL 12



$\tau$  ラウンドごとのローカルSGDの回数

$M$  学習器のモデルサイズ[bits]

$t_{com}$  1ラウンドに要する通信時間

- 分散型の学習方式の概要
- 通信面での課題
- 研究事例：伝送レート適応化による高速高精度なDecentralized FL
- 研究事例：(web公開版未収録)
- 研究動向
- 終わりに

- 深層学習器の表現力は学習パラメータ数に依存

Model Name	Input Shape	Params(M)
VGG16	(224, 224)	<b>138.4</b>
VGG19	(224, 224)	<b>143.7</b>
ResNet50v2	(224, 224)	<b>25.6</b>
Resnet152v2	(224, 224)	<b>60.4</b>
InceptionV3	(299, 299)	<b>23.9</b>
InceptionResNetV2	(299, 299)	<b>55.9</b>
DenseNet201	(224, 224)	<b>20.2</b>
MobileNet	(224, 224)	<b>4.3</b>
MobileNetV2	(224, 224)	<b>3.5</b>
EfficientNetB0	(224, 224)	<b>5.3</b>
EfficientNetB7	(600, 600)	<b>66.7</b>

参考：<https://qiita.com/T-STAR/items/59527f2b5484b9c22d1e>

- これらを都度共有するのか？

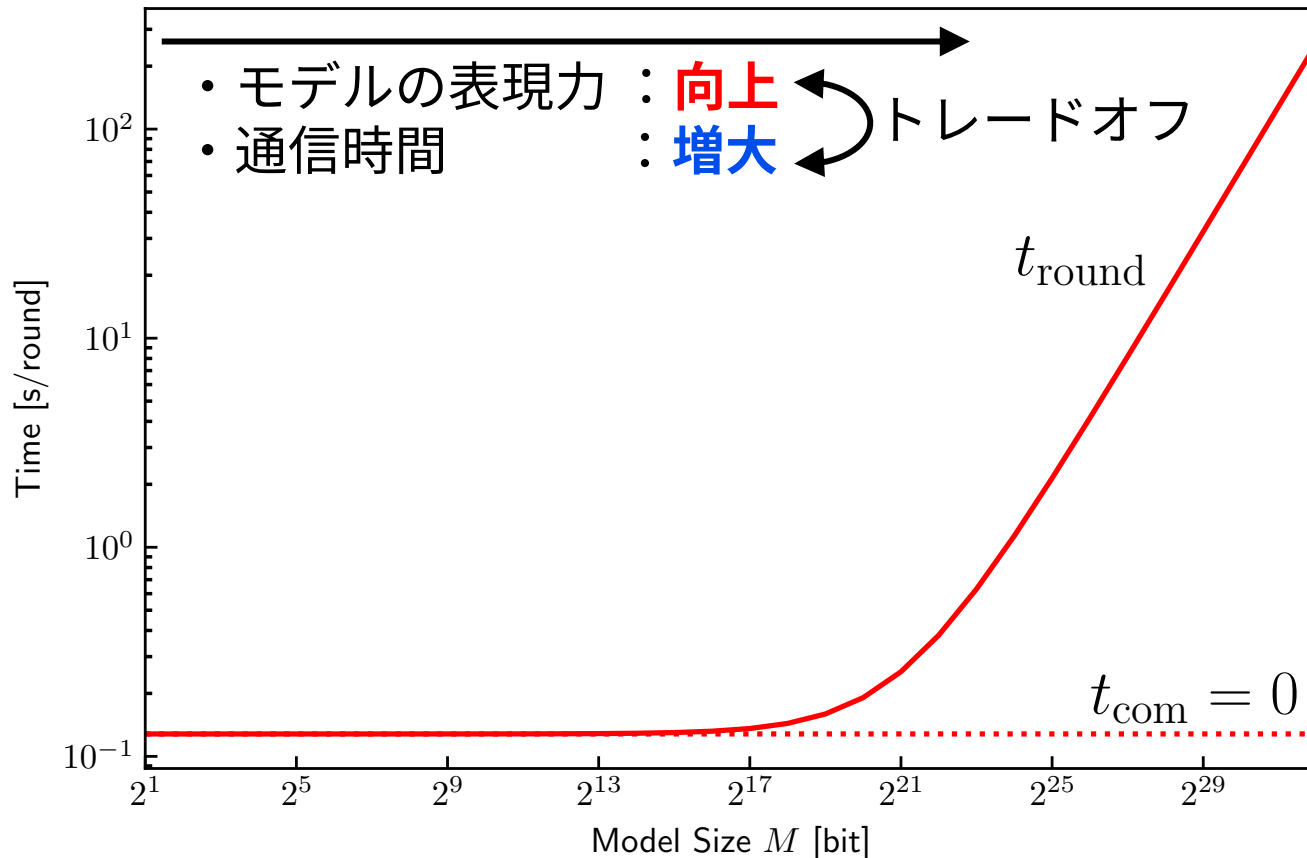
$n$ 台の端末が時分割に従うモデル共有によりC-SGDを行う場合の学習時間:

$$t_{\text{round}} = \tau t_{\text{cal}} + t_{\text{com}}$$

1ラウンド内でのローカル学習回数

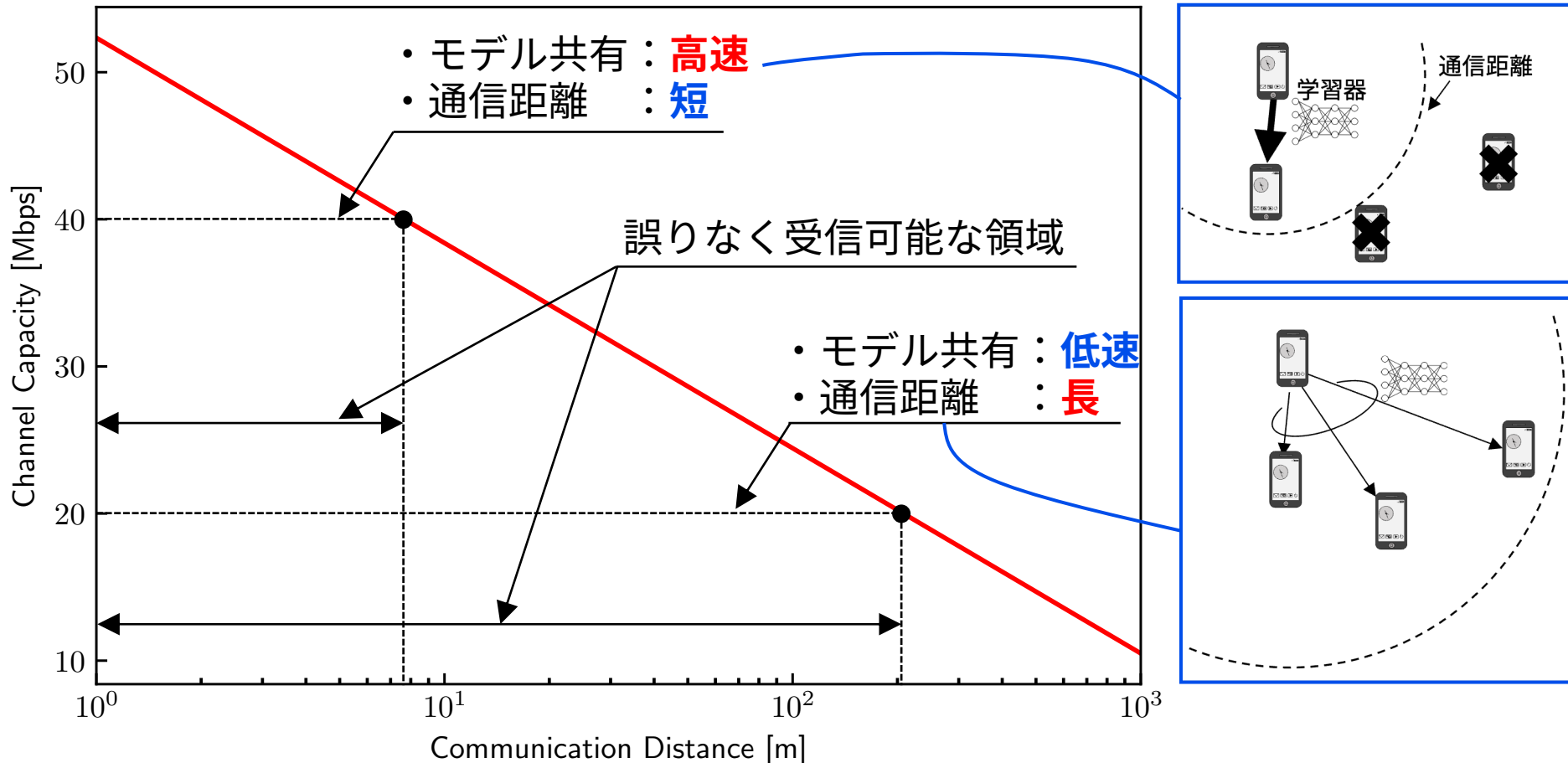
$$= \tau t_{\text{cal}} + \sum_{i=1}^n \frac{M}{R_i}$$

SGD1回に要する計算時間 [sec]  
 モデルサイズ [bits]  
 伝送レート [bps]



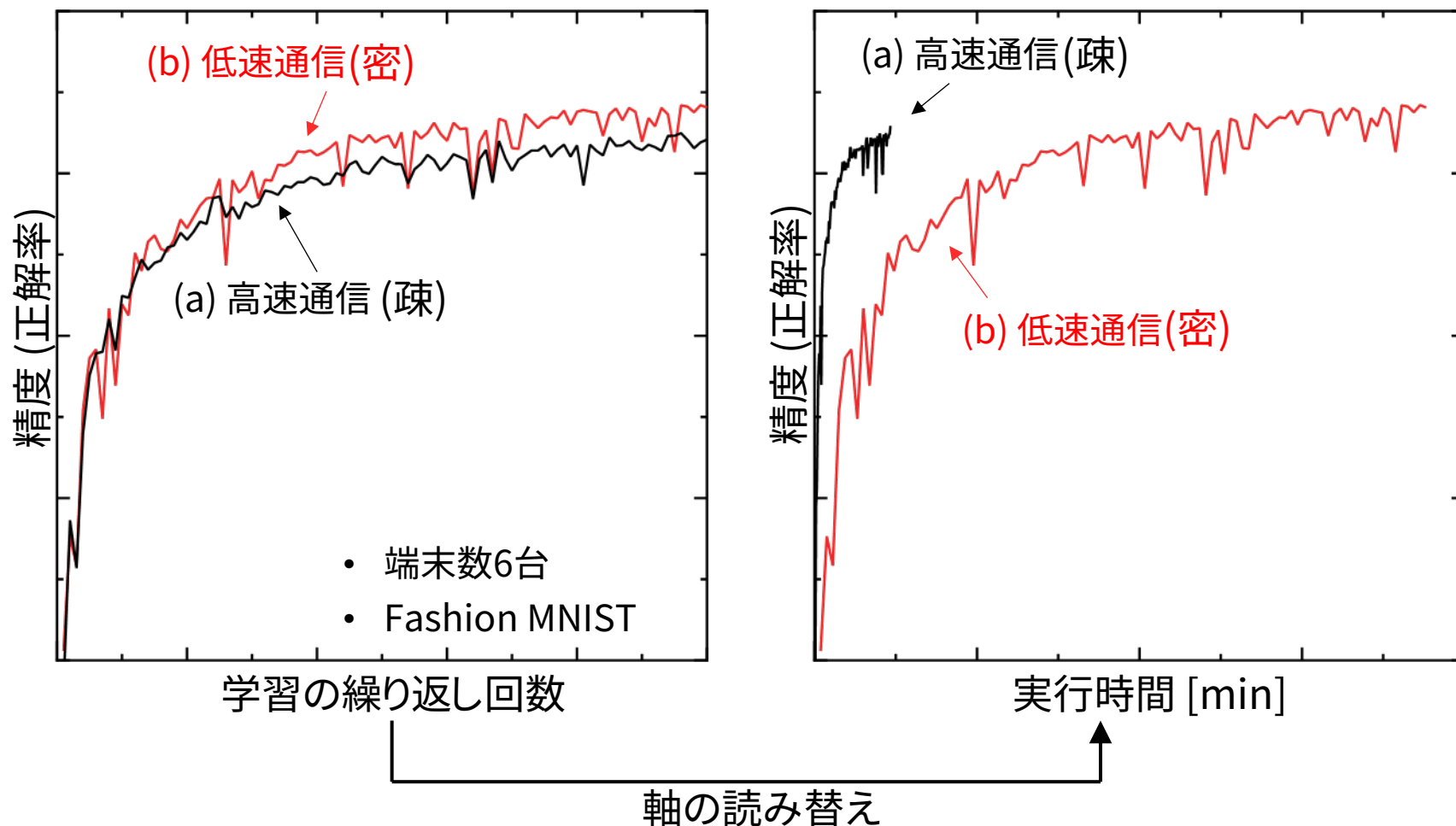
$\tau = 128$   
 $t_{\text{cal}} = 1.0 \times 10^{-3}$   
 $n = 6$   
 $R_i = 1.0 \times 10^9$

例: 無線LANベースでの適応変調に基づくマルチキャスト



- 距離減衰+AWGNを仮定
- 通信路容量が伝送レートを上回ればモデル共有成功と仮定
  - 端末間の協調効果-モデル共有時間には**トレードオフ**





- 十分学習させた後の学習精度と実行時間特性の間にはトレードオフ

## ここでの目的

本特性を踏まえて通信ボトルネックを解消する送信レート設計法の確立

- 各ノードは同サイズのi.i.d.データセットを持つ
- 各ノードは同一構成の学習器を同一のシードで初期化
- 相互に位置関係を把握済み+時間同期済み+送信レートが互いに既知 (後述)

減衰係数 i.i.d.: independent and identically distributed

瞬時受信電力

$$P(d) = P_{\text{Tx}} \underbrace{G_A G_f}_{\substack{\text{フェージング利得} \\ \text{(i.i.d.レイリー環境)}}} \left( \frac{d}{d_0} \right)^{-\epsilon} \quad [\text{mW}]$$

## 通信路容量

$$C(d) = \underbrace{B}_{\text{帯域幅}} \log_2 \left( 1 + \frac{P(d)}{\underbrace{N_0 B}_{\text{AWGN}}} \right) \quad [\text{bps}]$$

## モデル共有の成功条件

$$R_i \leq C(d)$$

## 送信レート: 学習に先立ち最適化

※符号長 (=学習器の規模)は十分大きいものと仮定

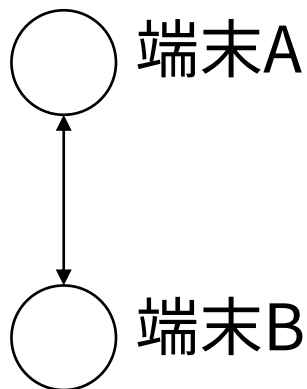
※以下は固定&既知

- 送信電力
- 帯域幅
- 平均AWGN
- 距離減衰係数

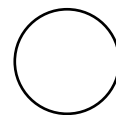
$k+1$ 回目の学習でモデル共有を行う場合、ネットワーク全体の学習の動作:

$$\underline{X}_{k+1} = (\underline{X}_k - \eta \underline{G}_k) \underline{W}_k$$

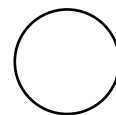
学習率  
 損失関数に対する  
 勾配ベクトル  
 $n \times n$ の**平均化行列**  
 各端末のモデルパラメータを  
 格納したベクトル



$$W = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$$



$$W = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$



$k+1$ 回目の学習でモデル共有を行う場合、ネットワーク全体の学習の動作:

$$\underline{X}_{k+1} = (\underline{X}_k - \eta \underline{G}_k) \underline{W}_k$$

学習率  
 損失関数に対する  
 勾配ベクトル  
 $n \times n$ の**平均化行列**  
 各端末のモデルパラメータを  
 格納したベクトル

- 平均化行列の2番目に大きい固有値 $\lambda$ :
  - $W$  の非ゼロ要素多: 0に近づく
  - 非ゼロ要素少: 1に近づく
 (直観的には)**ネットワークのスパース度**を表現
- 学習精度の繰り返し回数特性:  $\lambda$ の関数として表現可能

## アイデア

$\lambda$ を制約することで、学習結果を保ちつつ通信時間を短縮できないか？

ネットワークトポロジの密度と連結性を制約に通信時間を最小化する送信レートを選択:

$$\begin{aligned}
 & \min_{\mathbf{R}} t_{\text{com}} && \text{1回のモデル共有における通信時間} \\
 & \text{s.t. } \Pr[\lambda \leq \lambda_{\text{target}} | (\text{topology is strongly connected})] \geq 1 - p_{\text{out}} && \text{成功確率} \\
 & R_i \geq 0 \quad \forall i && \text{ネットワークトポロジの密度に対する制約}
 \end{aligned}$$

$\mathbf{R} = [R_1, R_2, \dots, R_n]$  : 送信レートベクトル

$\lambda_{\text{target}}$  :  $\lambda$ のターゲット値 (ハイパーパラメータ)

(topology is strongly connected) : “有向グラフにおいて全ノードが相互に行き来可”

$p_{\text{out}}$  : アウテージ確率 (通信失敗起因)

※学習に先立ち各ノードで独立に本問題を解く。位置が共有済みの場合同一の解になるため、自動的に送信レートを相互に共有できる (=分散環境でも時分割できる)

## 評価概要

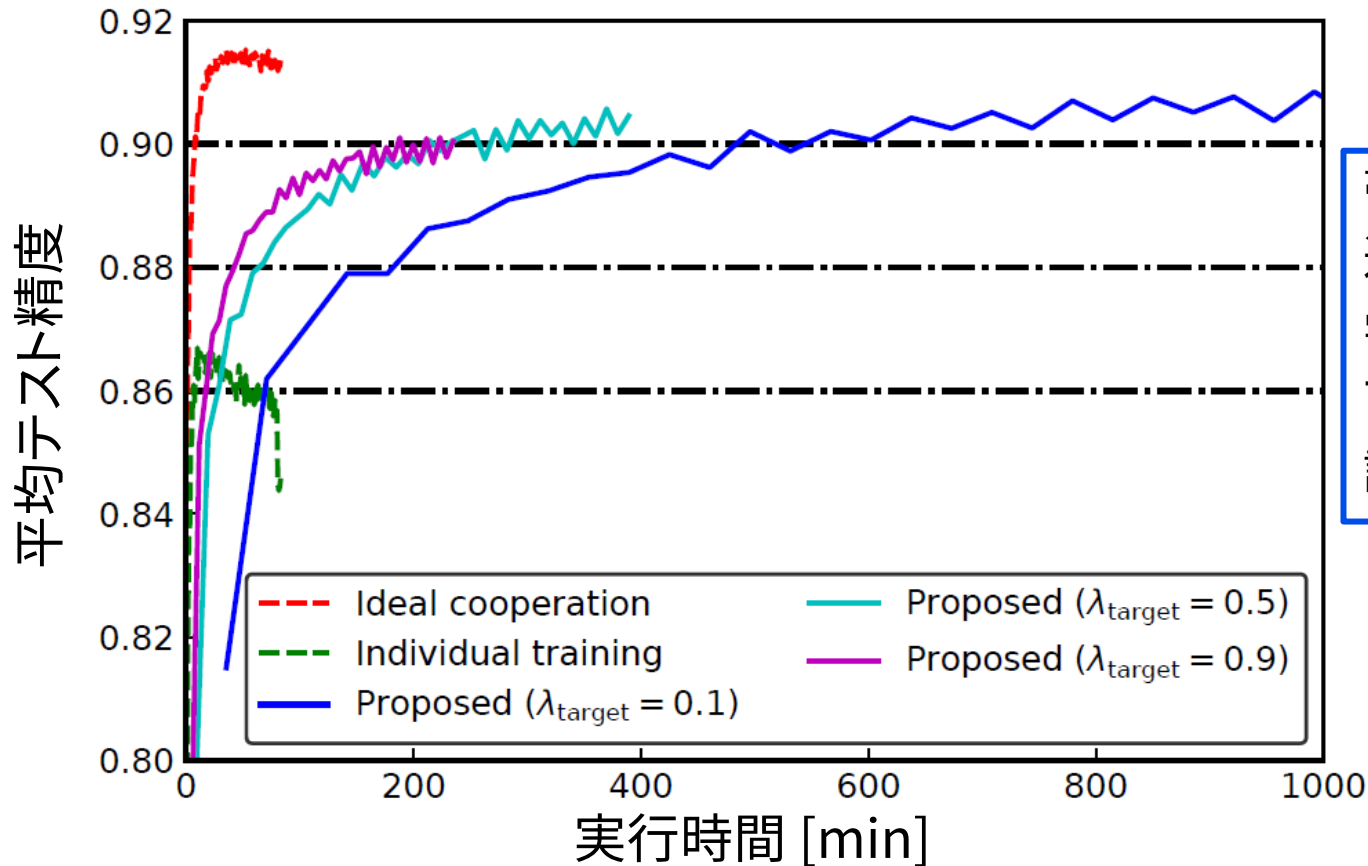
- データセット: FashionMNIST (右図)
- ノード数6台
  - 500m四方のエリアに分散配置
  - 60000枚の学習データをシャッフルの上重複なく分配 (10000枚/ノード)
- 各ノードの正解率の平均を評価
  - 10回独立に試行し、その結果を最終的な評価結果とした



## 使用モデル

- 畳み込み層×2, 全結合層×3
- 活性化関数: ReLU
- 損失関数: 交差エントロピー
- 総パラメータ数: 1475338
  - **47Mbits in 32-bit float**

- 総学習回数200000回
  - 5000回ごとに各端末所有のモデルの精度を評価→その平均を評価、というシミュレーションを独立に10回試行した

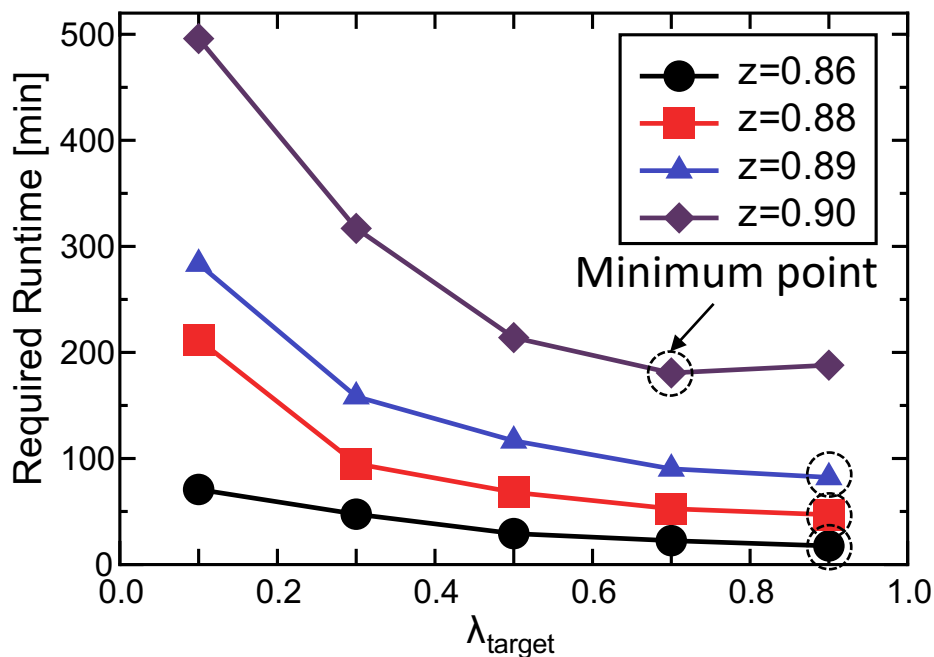


帯域幅: 1.4[MHz]  
送信電力: 0[dBm]  
学習回数 $\tau$ : 1000  
アウトージ確率: 0.50  
計算時間: 0.01[s/loop]

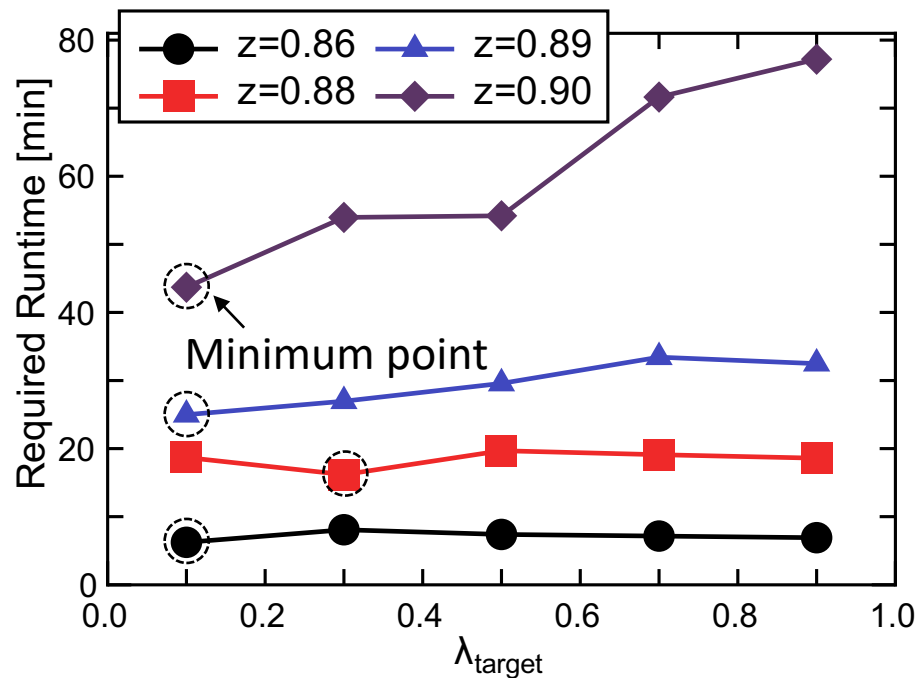
- 疎にするほど十分な回数繰り返した後の精度は落ちる
- 一方で「所望精度を達成する時間」の観点では疎のトポロジが有利なケース多数

評価指標: ”ある所望精度を達成する時間”

### 実行時間特性(減衰係数4)



### 実行時間特性(減衰係数3)



- 減衰係数大: 疎に取った方が有利
- 減衰係数小: 密に取った方が有利 (という傾向)
- 通信路の特性に応じたネットワークの疎密の調整が有効



Device-to-Device(D2D)環境における通信パラメータ設計を考える

## 典型的な無線リソース割当て

$$\max \sum_{i=1}^l C_i$$

通信リンク数

s.t.(commun. resource)

- 注水定理に基づく好条件のリンクへの優先的割り当てが基本
- 通信相手は固定

## (Decentralized) FL×時分割

$$\min \sum_{i=1}^n \frac{M}{R_i}$$

注水定理では足を引っ張る端末が発生

s.t.(indicator for target acc.)

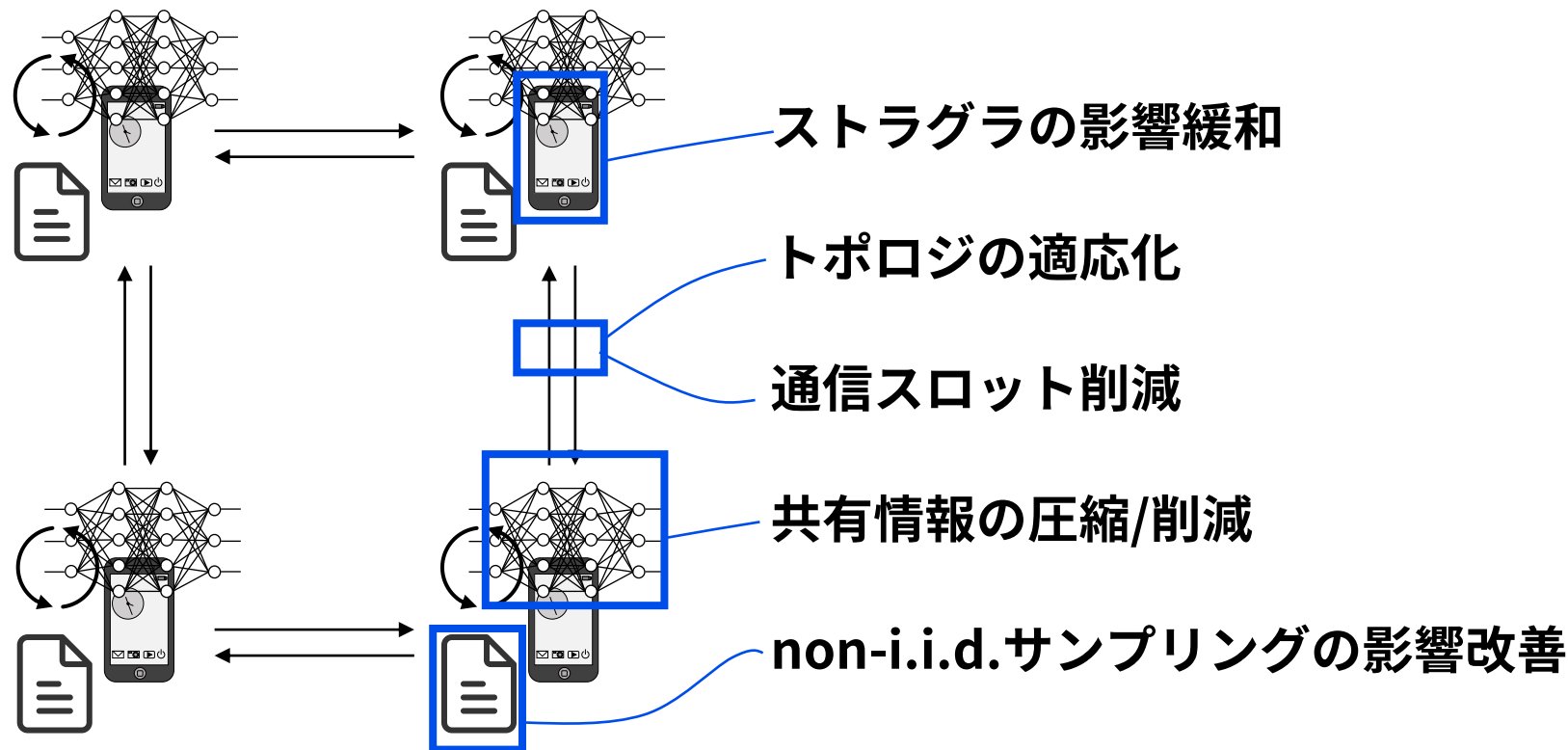
- 条件が悪いリンクにも一定数リソースを割り当てる必要
  - もしくはリンクを切る
- **学習結果が改善されればトポロジや通信内容はなんでもよい**

- リソース最適化の知見は活用可 (例: max-min SINR)
- ただし目的が「高速高精度な学習」であることを念頭に置くことが重要

- 分散型の学習方式の概要
- 通信面での課題
- 研究事例：伝送レート適応化による高速高精度なDecentralized FL
- 研究事例：(web公開版未収録)
- 研究動向
- 終わりに

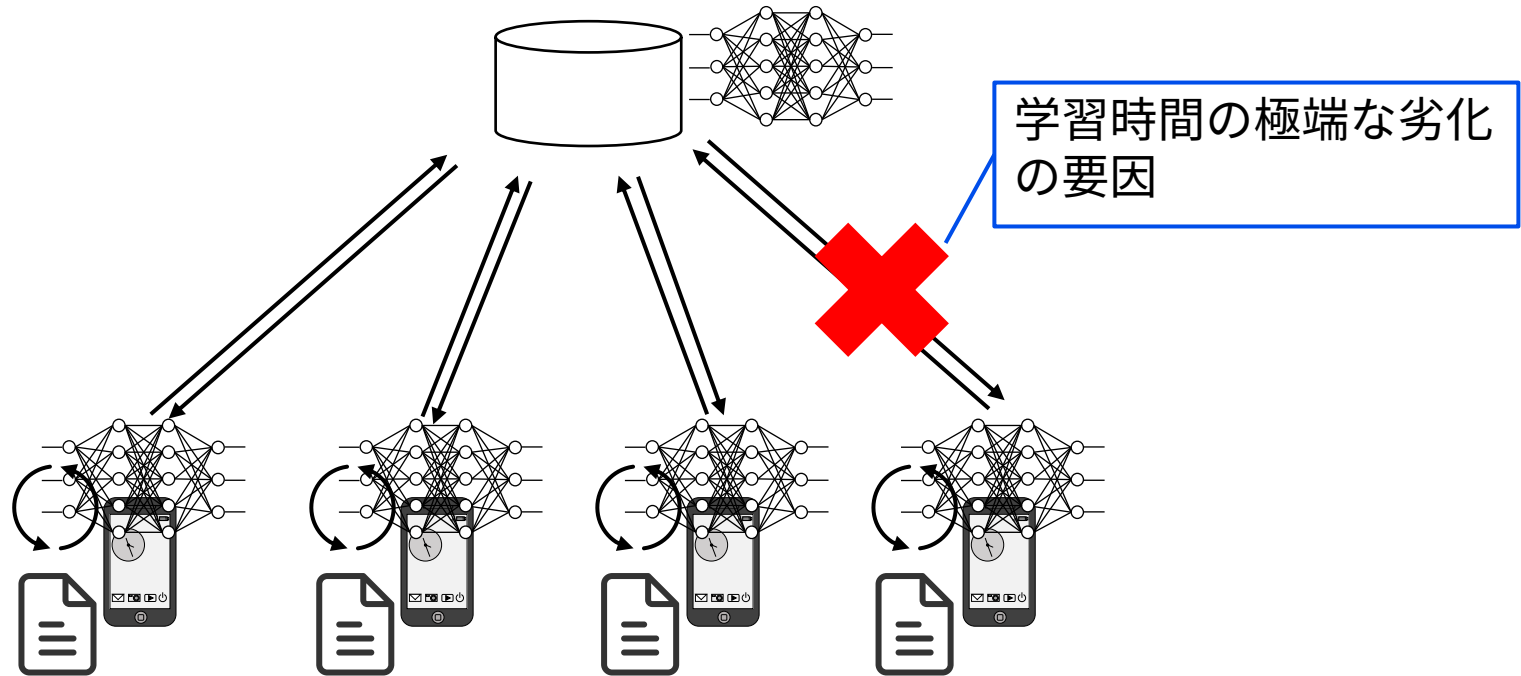
- 分散型の学習方式の概要
- 通信面での課題
- 研究事例：伝送レート適応化による高速高精度なDecentralized FL
- 研究事例：(web公開版未収録)
- **研究動向**
- 終わりに

- 要素ごとに学習特性の劣化要因が存在。その数だけ方策あり
  - 学習データ、共有情報、計算時間、トポロジなど



- FL/DFL双方を対象に関連する話題を紹介する

ストラグラ：極端に通信/計算が遅い端末。**確率的に発生する恐れ**



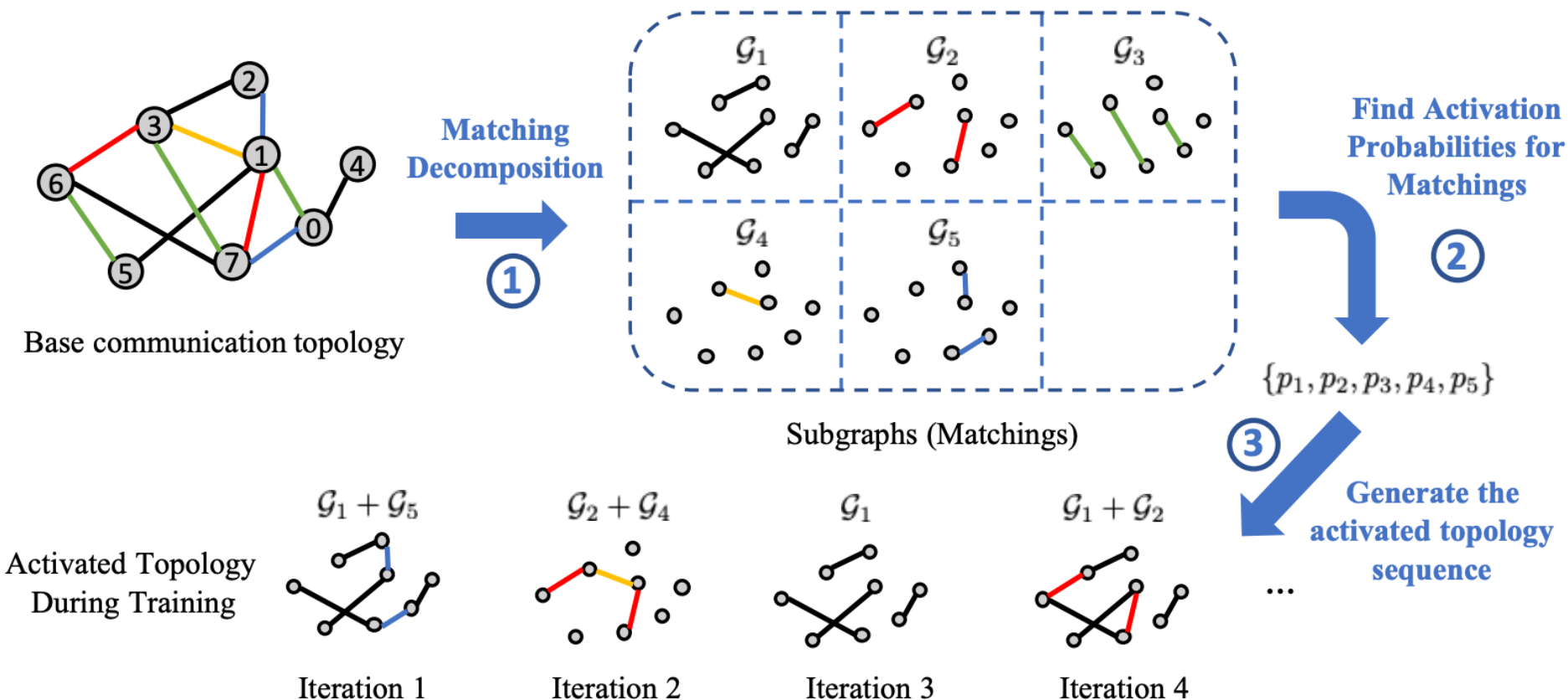
## クライアント選択の適応化 (FL)

- [Nishio+, ICC2019]
- [Yu+, IoTJ2022]

## 符号化計算 (FL)

- [Prakash+, JSAC2021]

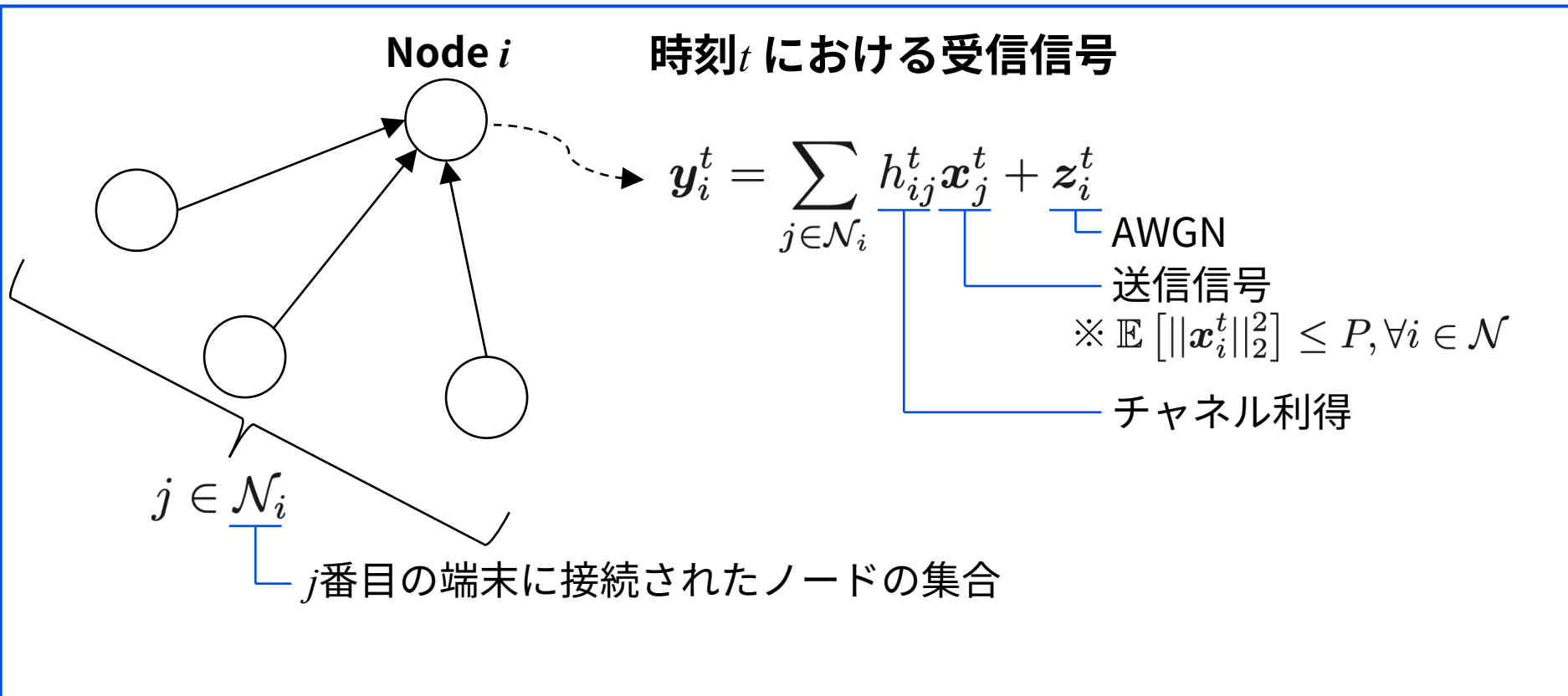
- 例：グラフ彩色の枠組みで、限られたスロットで密な通信を狙う



出典：[Wang+, IndianCC2019]

- arXiv: <https://arxiv.org/abs/1905.09435>
- “MATCHA”フレームワークとしてよく比較対象となる

- 先に紹介した伝送レート適応化もこの話題に関連

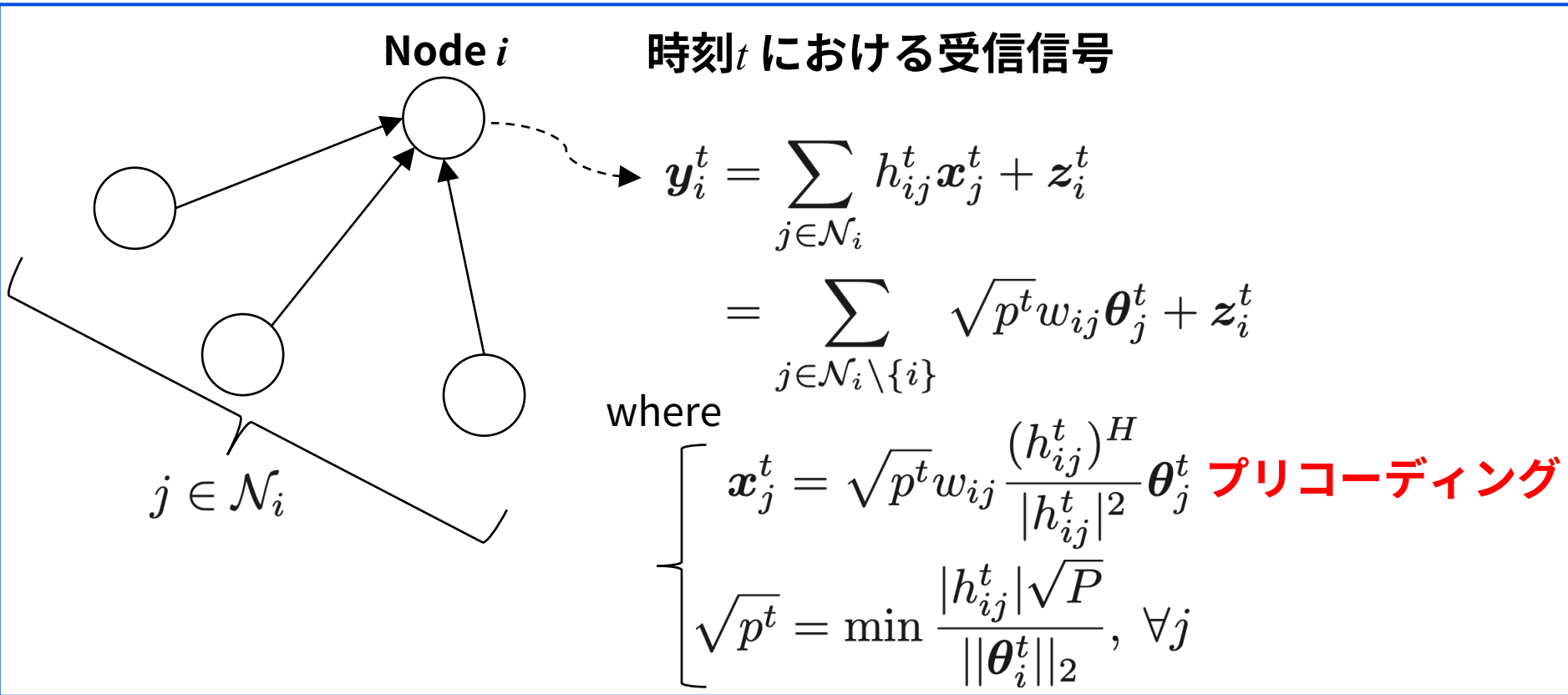


アナログ変調信号の合成結果からモデルを表現することで同時送信を実現

- 計算方法/変数の表記は[Shi+, ISIT2021]に準拠

DFL: [Shi+, ISIT2021]

FL: [Yang+, TWC2020]



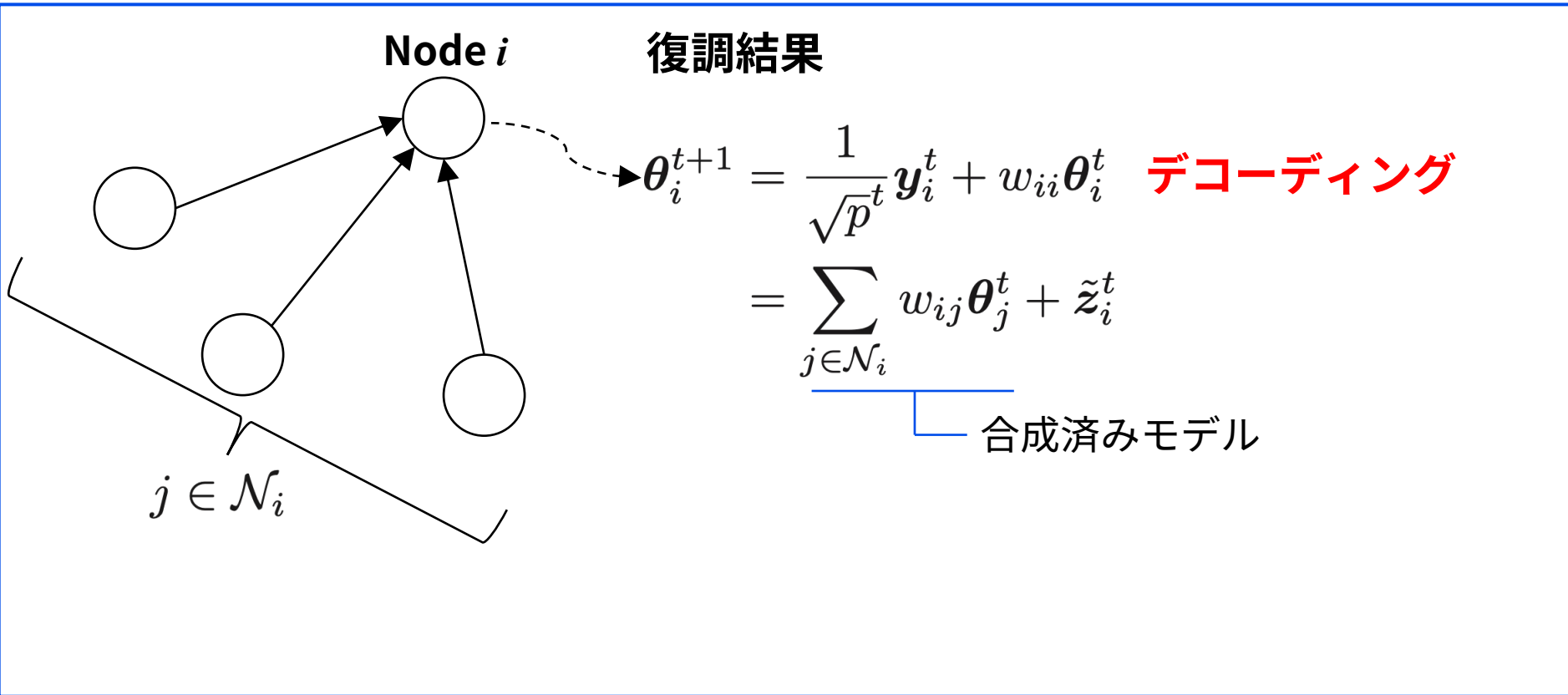
アナログ変調信号の合成結果からモデルを表現することで同時送信を実現

- 計算方法/変数の表記は[Shi+, ISIT2021]に準拠

DFL: [Shi+, ISIT2021]

FL: [Yang+, TWC2020]





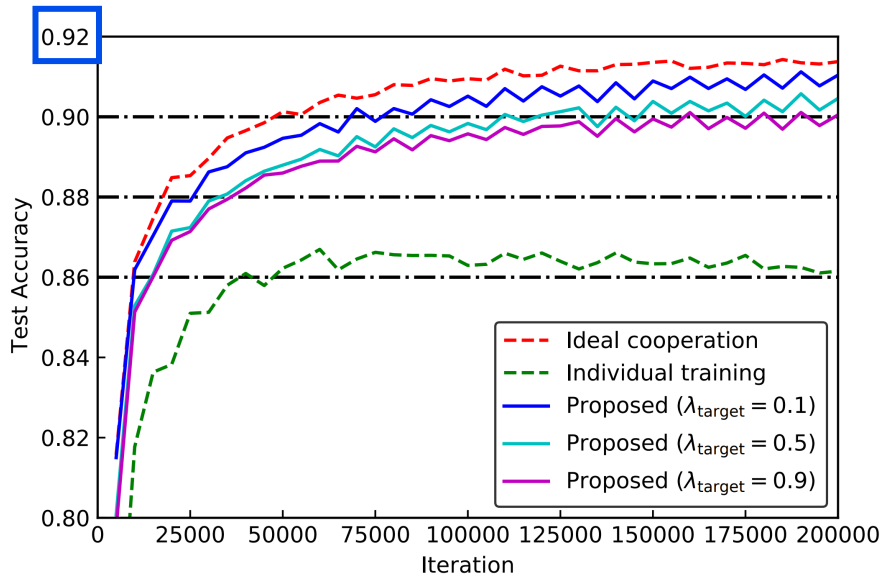
アナログ変調信号の合成結果からモデルを表現することで同時送信を実現

- 計算方法/変数の表記は[Shi+, ISIT2021]に準拠

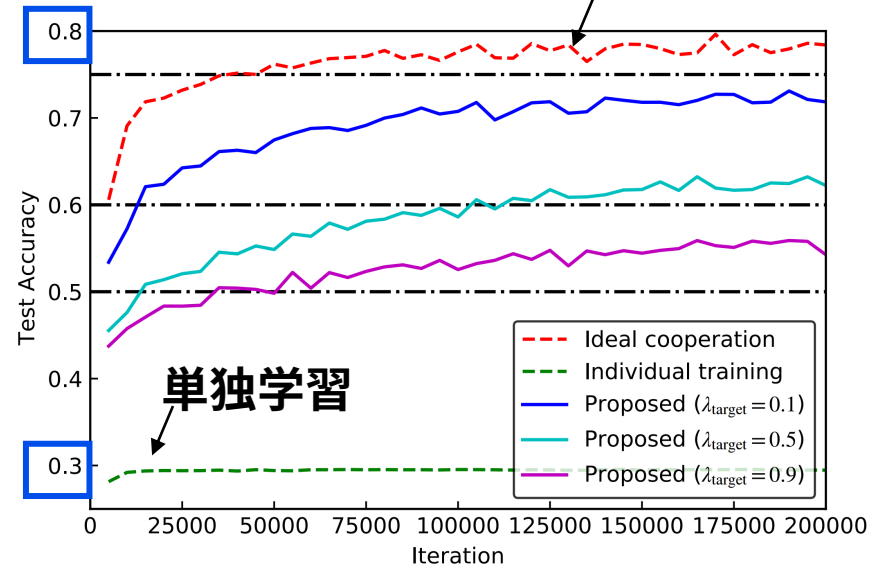
DFL: [Shi+, ISIT2021]

FL: [Yang+, TWC2020]

- 端末6台でのC-SGDにおける平均精度 (Fashion MNIST) **理想協調(全結合トポロジ)**



i.i.d. setting (10ラベルから抽出)



non-i.i.d. (各端末3ラベルのみ)

## non-i.i.d.環境における学習収束性の解析 (FL)

- [Li+, ICLR2020]

## 一部のデータの公開を許容したハイブリッド (FL)

- [Yoshida+, ICC2020]

## 関連動向のサーベイ (FL)

- [Zhu+, Neurocomputing2021]

- 学習パラメータの量子化に基づく方式
  - [Taheri+, ICML2020] (DFL)
    - Push-Sumアルゴリズムを用いた分散端末間の合意により量子化前と同等の学習収束特性が得られるとの指摘
- 蒸留に基づく方式
  - [Jeong+, arXiv2018]
    - **端末-サーバ間の共有情報をモデルパラメータではなく学習器の出力情報とすること**で、通信オーバーヘッドを26倍改善可能であるとの報告

- Decentralized FLを中心とした話題について述べた
- 特に情報理論分野との結びつきが強い研究課題の例：
  - ネットワークトポロジの適応化
    - 学習精度-通信時間にはトレードオフ。この関係を簡易な式で表現することは可能か？
  - 学習モデルの圧縮/削減
    - (モデルを共有する場合)符号長 $\div$ 学習パラメータ数 $\div$ 表現力
      - 通信パラメータ/学習器/共有アルゴリズムの一体設計は可能か？
  - ストラグラ対策
    - 符号化計算で確実な計算 or 消失を許容して粗く高速に繰り返す？

以下の資料も閲覧いただければ幸いです：

- 佐藤光哉, "無線設計の問題として見る分散連合機械学習," IEICE Fundamentals Review, 2022年7月公開予定

- [McMahan+, AISTATS2017] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-efficient learning of deep networks from decentralized data,” in Proc. AISTATS, Fort Lauderdale, Florida, USA, Apr. 2017.
- [Kairouz+, FTML2021] P. Kairouz et al., "Advances and open problems in federated learning", Foundations and Trends in Machine Learning: Vol. 14: No. 1–2, pp 1-210, 2021.
- [Zinkevich+, NeurIPS2010] M. Zinkevich, M. Weimer, L. Li, and A. Smola, “Parallelized stochastic gradient descent,” in Proc. NeurIPS, Vancouver, Canada, 2010, pp. 1–9.
- [Li+, MLSys2020] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, “Federated optimization in heterogeneous networks,” in Proc. MLSys 2020, virtual conference, 2020, pp. 429–450.
- [Lian+, NeurIPS2018] X. Lian, C. Zhang, H. Zhang, C.-J. Hsieh, W. Zhang, and J. Liu, “Can decentralized algorithms outperform centralized algorithms? a case study for decentralized parallel stochastic gradient descent,” in Proc. NeurIPS, Long Beach, CA, USA, Jan. 2018, pp. 5330–5340.
- [Lian+, ICML2018] X. Lian, W. Zhang, C. Zhang, and J. Liu, “Asynchronous decentralized parallel stochastic gradient descent,” in Proc. ICML, Stockholm, Sweden, 2018.
- [Wang+, ICML Workshop2019] J. Wang and G. Joshi, “Cooperative SGD: A unified framework for the design and analysis of communication-efficient SGD algorithms,” in Proc. ICML Workshop, Long Beach, CA, USA, 2019.
- [Sato+, TCCN2021] K. Sato and D. Sugimura, “Rate-adapted decentralized learning over wireless networks,” IEEE Trans. Cogn. Commun. Netw., vol. 7, no. 4, pp. 1412–1429, 2021.
- [Nishio+, ICC2019] T. Nishio and R. Yonetani, “Client selection for federated learning with heterogeneous resources in mobile edge,” in Proc. IEEE ICC, Shanghai, China, May 2019, pp. 1–7.
- [Yu+, IoTJ2022] L. Yu, R. Albelaihi, X. Sun, N. Ansari, and M. Devetsikiotis, “Jointly optimizing client selection and resource management in wireless federated learning for Internet of Things,” IEEE Internet Things J., vol. 9, no. 6, pp.4385-4395, March 2022.

- [Prakash+, JSAC2021] S. Prakash et al., "Coded computing for low-latency federated learning over wireless edge networks," IEEE J. Sel. Areas Commun., vol. 39, no. 1, pp. 233-250, Jan. 2021.
- [Wang+, IndianCC2019] J. Wang, A. K. Sahu, Z. Yang, G. Joshi, and S. Kar, "MATCHA: Speeding up decentralized SGD via matching decomposition sampling," in Proc. 2019 Sixth Indian Control Conference (ICC), Telangana, India, 2019, pp. 299–300.
  - arXiv: <https://arxiv.org/abs/1905.09435>
- [Shi+, ISIT2021] Y. Shi, Y. Zhou and Y. Shi, "Over-the-air decentralized federated learning," IEEE ISIT, 2021, pp. 455-460.
- [Yang+, TWC2020] K. Yang, T. Jiang, Y. Shi and Z. Ding, "Federated learning via over-the-air computation," IEEE Trans. Wireless Commun., vol. 19, no. 3, pp. 2022-2035, March 2020.
- [Li+, ICLR2020] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, "On the convergence of FedAvg on non-IID data," in Proc. ICLR, 2020, pp.1–26.
- [Yoshida+, ICC2020] N. Yoshida, T. Nishio, M. Morikura, K. Yamamoto and R. Yonetani, "Hybrid-FL for wireless networks: Cooperative learning mechanism using non-IID data," in Proc. IEEE ICC, 2020, pp. 1-7.
- [Zhu+, Neurocomputing2021] H. Zhu, J. Xu, S. Liu, and Y. Jin, "Federated learning on non-IID data: A survey," Neurocomputing, vol. 465, pp. 371–390, 2021.
- [Jeong+, arXiv2018] E. Jeong, S. Oh, H. Kim, J. Park, M. Bennis, and S.-L. Kim, "Communication-efficient on-device machine learning: Federated distillation and augmentation under non-iid private data," arXiv, 2018. [Online] Available: <https://arxiv.org/abs/1811.11479>