

機械学習を用いたスパースサンプリングによる データ処理技術の基礎開発(II)

© Osaka Metropolitan University All Rights Reserved.

阪公大理^A, 阪公大NITEP^B, 阪大RCNPC^C, 阪大IDS^D,
阪大情報^E, 九大理^F, KEK素核研^G, 九工大^H

加藤睦代^A, 岩崎昌子^{A, B, C, D}, 長原 一^{C, D}, 吉田道隆^E, 末原大幹^F,
山田 悟^G, 中島悠太^{C, D}, 武村紀子^{D, H}, 中野貴志^{C, D}

1. Introduction

2. スパースサンプリング・復元の原理

3. ILC SiD電磁カロリメータのエネルギー較正

4. パターン最適化NNを使用した削減パターンの最適化

4-1. 削減パターン最適化の手法

4-2. 最適化パターンによる削減後のILC SiD電磁カロリメータの
エネルギー較正と最適化パターンの評価

5. まとめ

1. Introduction

大型加速器実験では
超精密実験を目指して加速器を高輝度化し、測定器も高精度化した。
→大量データを高頻度で収集する必要がある。

しかし、データ収集帯域には限度がある
e.g. Belle2実験のデータ転送帯域：30GB/sec

トリガー条件を満たさないデータは捨てられている

…ダークマター事象候補のデータが捨てられている？

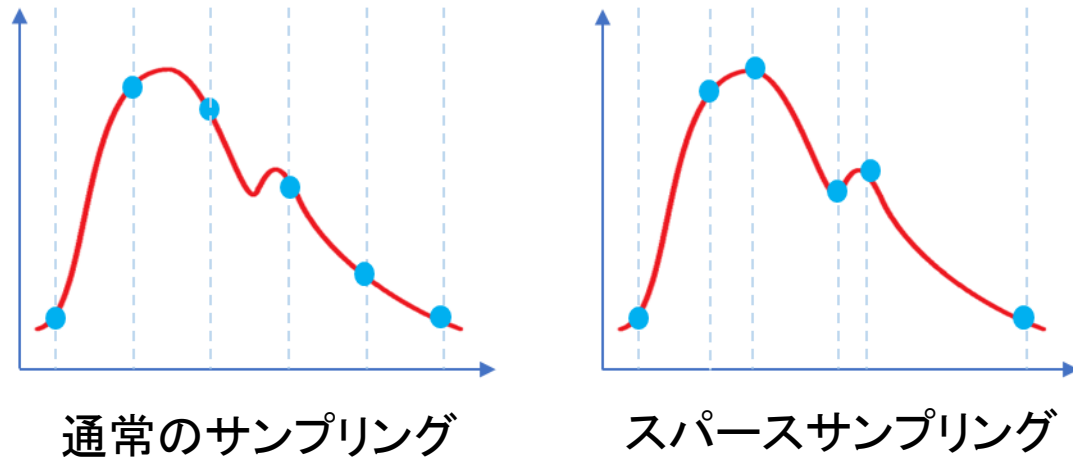
リアルタイムにデータサイズを低減できれば、より多くの事象数を取得することが可能になる



**機械学習を用いたスパースサンプリングを導入し
リアルタイムにデータサイズの低減を目指す**

スパースサンプリング・復元による信号観測

元信号に対して**観測点が少ない**、劣決定問題を解く手法



同じサンプリング数で、より広帯域に信号復元可能

…ブラックホールを「可視化」した手法として有名

・計測した**少数データ**から、信号のスパース性(時間や空間の近傍連続性)を仮定して元データ情報を**復元**する。

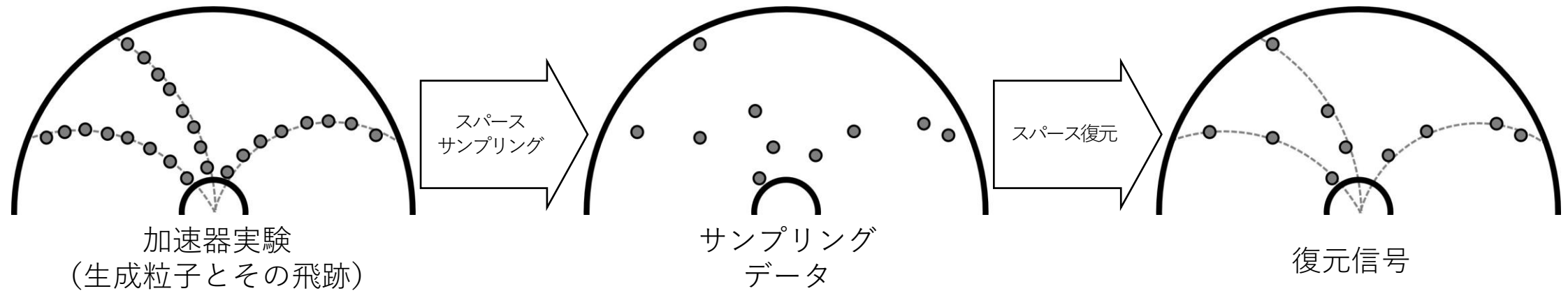
→多数ある観測点から本質的な観測点の選別も可能

信号が持つ**重要な情報を保つような**
サンプリングパターンを機械学習で最適化する

2. スパースサンプリング・復元の原理

素粒子実験へのスパースサンプリングの適用

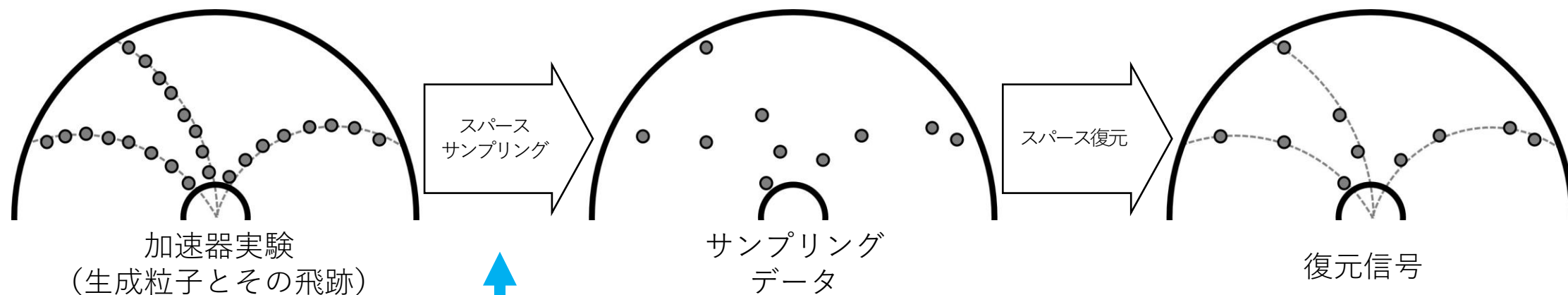
素粒子実験に、スパースサンプリング・スパース復元を適用させることを目指す。



2. スパースサンプリング・復元の原理

素粒子実験へのスパースサンプリングの適用

素粒子実験に、スパースサンプリング・スパース復元を適用させることを目指す。

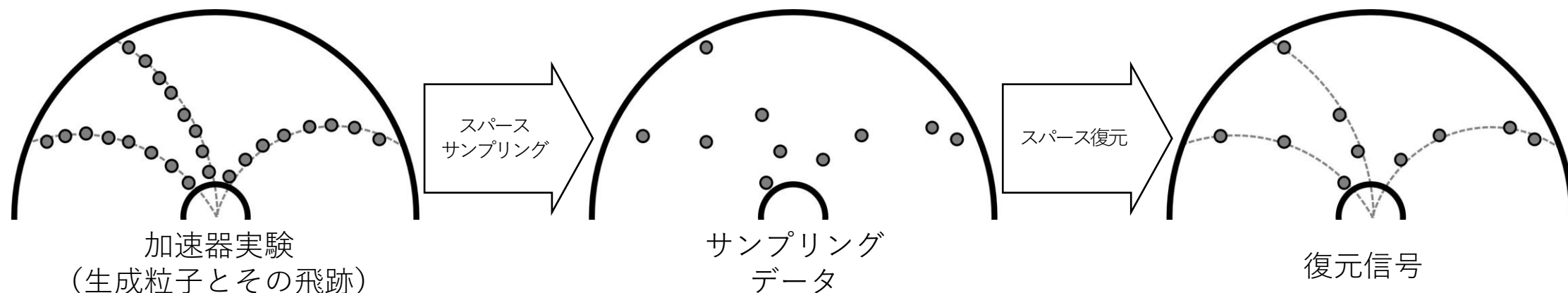


**DAQ時に必要な処理は「あらかじめ決められたパターンにしたがうサンプリング処理」
→ リアルタイムにデータ削減可能**

2. スパースサンプリング・復元の原理

素粒子実験へのスパースサンプリングの適用

素粒子実験に、スパースサンプリング・スパース復元を適用させることを目指す。

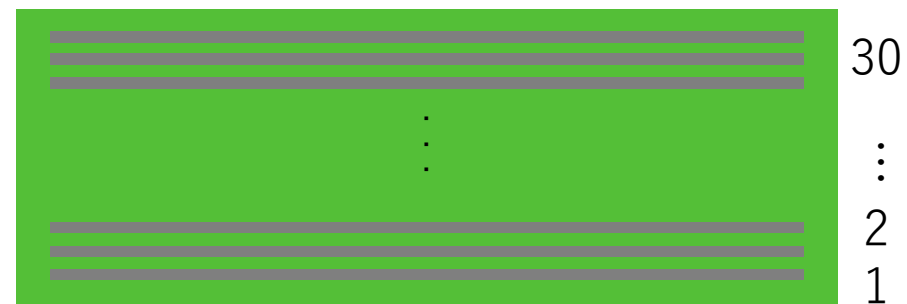
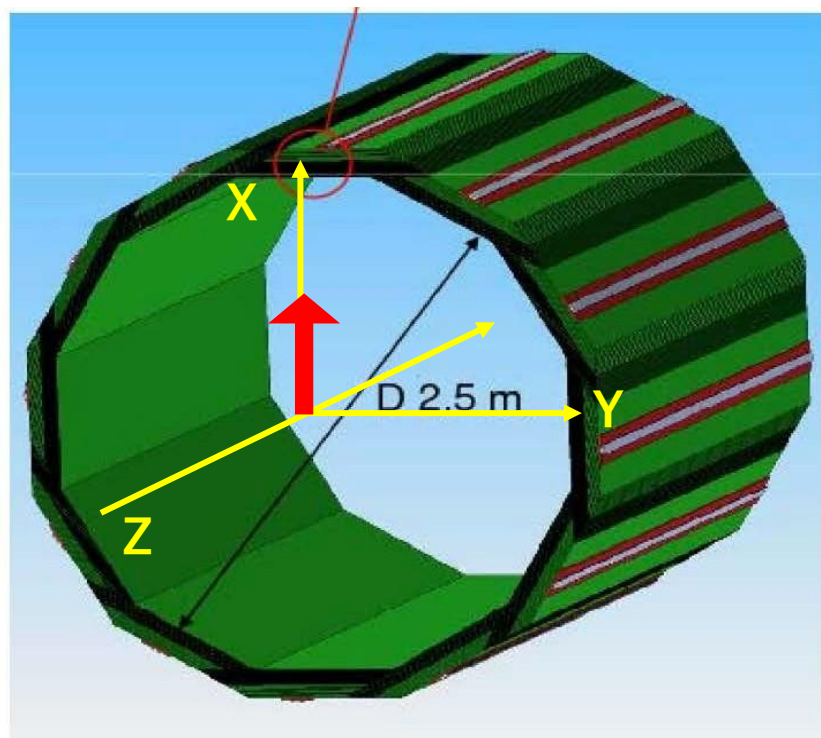


本研究では、ILC SiD 電磁カロリメータのエネルギー較正にスパースサンプリングを適用し、エネルギー較正性能について評価を行った

3. ILC SiD電磁カロリメータのエネルギー較正

ILC SiD 電磁カロリメータ(ECL)

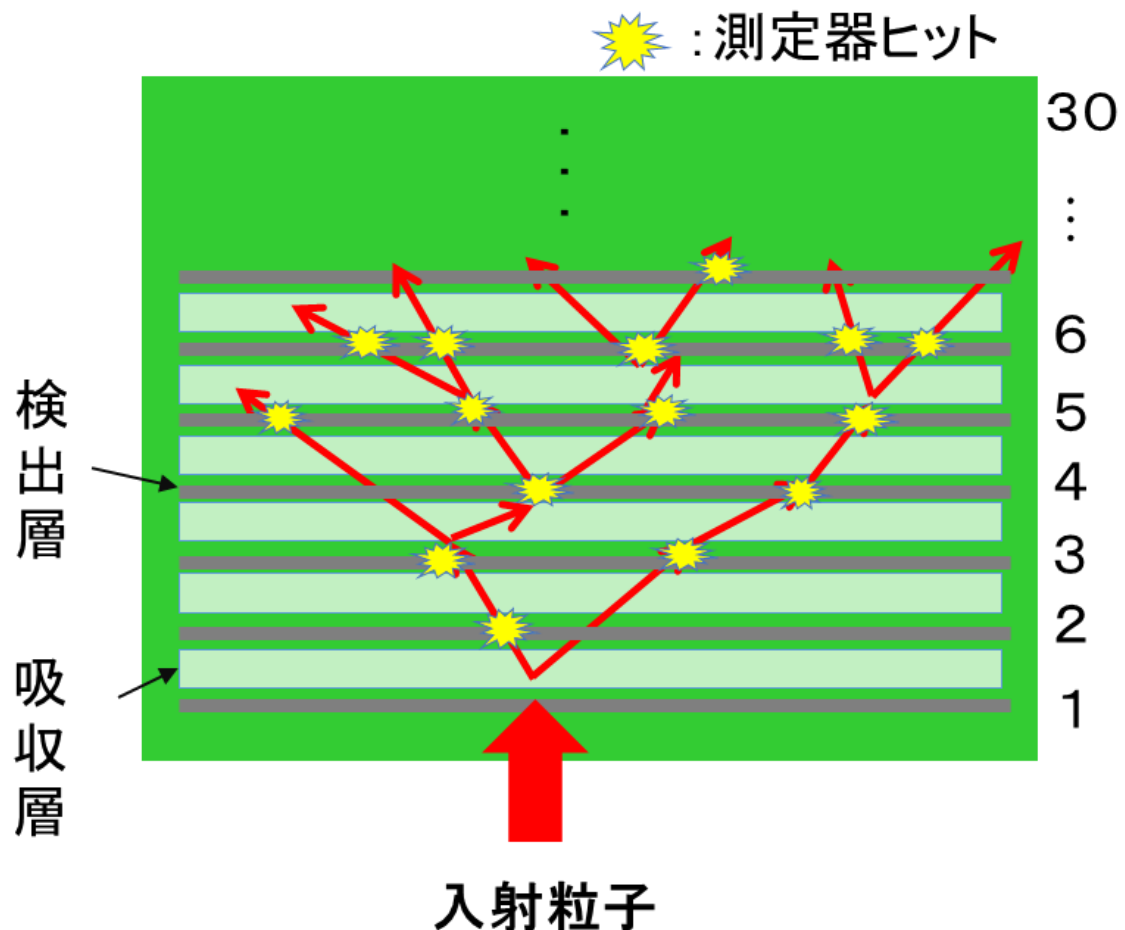
- 検出層(Si)と吸収層(W)が交互に30層重なる**サンプリング型**
- 物質質量(total) $\sim 26X_0$
- エネルギー分解能(デザイン値) $(17/\sqrt{E\oplus 1})\%$



入射粒子

3. ILC SiD電磁カロリメータのエネルギー較正

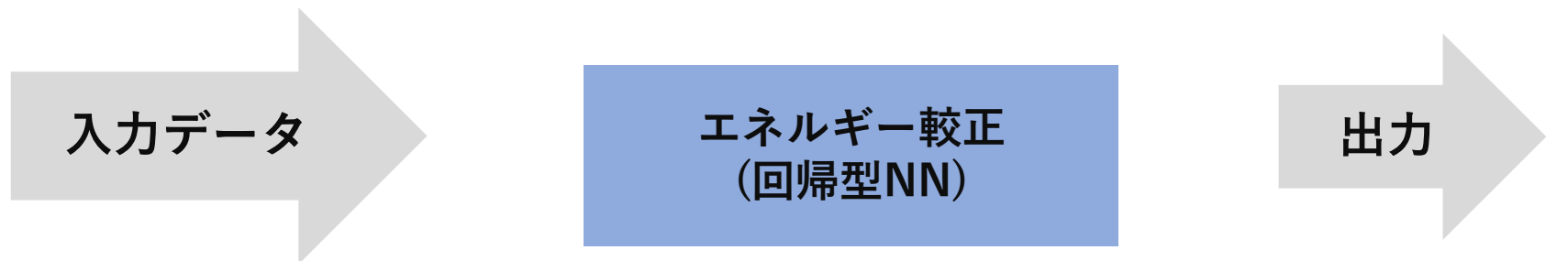
ILC SiD 電磁カロリメータ(ECL)



- 粒子が電磁カロリメータに入射すると電磁シャワーが生成
- 検出層で粒子がエネルギーを落とすとセンサーから測定器ヒット(位置とエネルギー)を得る
- 測定器ヒットデータから生成シャワーを集めてクラスターとして認識する
- 本研究では、同一入射粒子に由来する測定器ヒットデータを集めて1つのクラスターとした。

3. ILC SiD電磁カロリメータのエネルギー較正

ニューラルネットワーク(NN)を使用したエネルギー較正



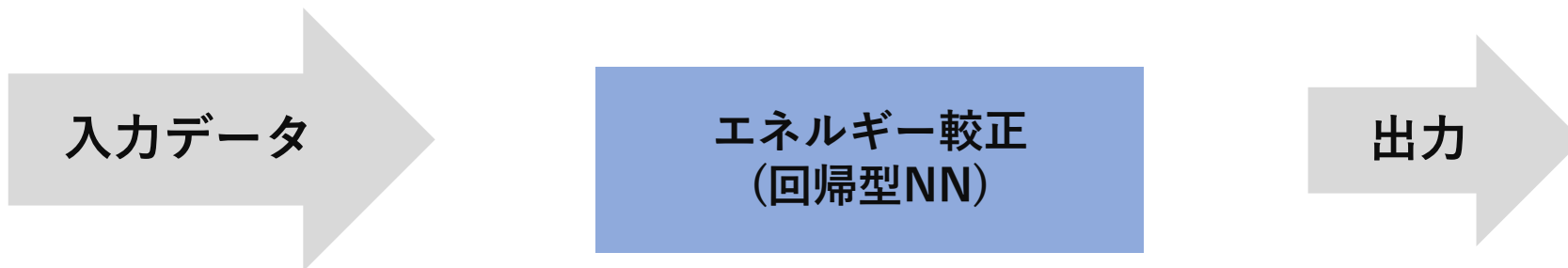
ヒットの位置情報,
測定エネルギー
(low level data)

較正後の入射粒子の
エネルギー

- [1] 中 祐介,他8名, 日本物理学会第75回年次大会 ,2020年03月
- [2] 中 祐介,他 11 名, 日本物理学会 2020 年秋季大会 ,2020年9月

3. ILC SiD電磁カロリメータのエネルギー較正

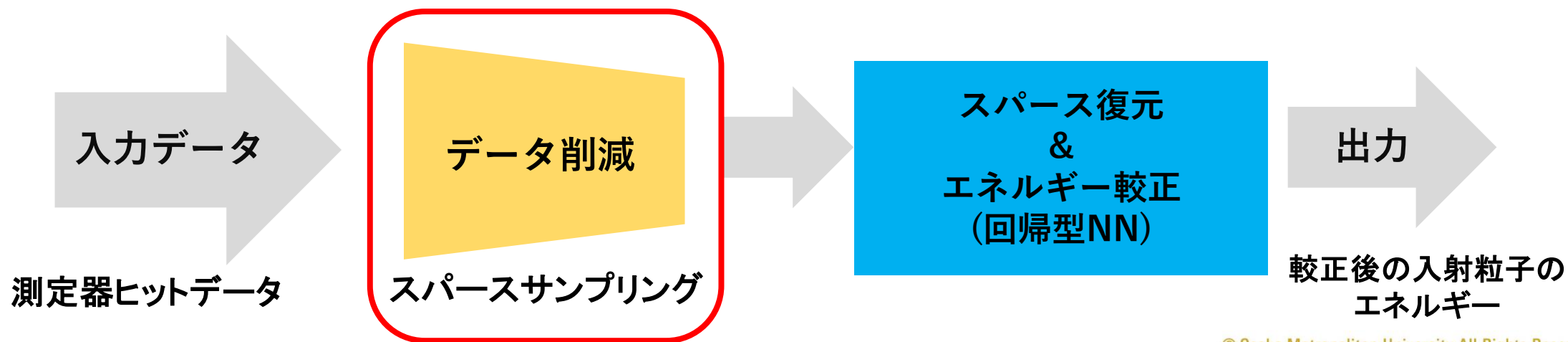
ニューラルネットワーク(NN)を使用したエネルギー較正



ヒットの位置情報,
測定エネルギー
(low level data)

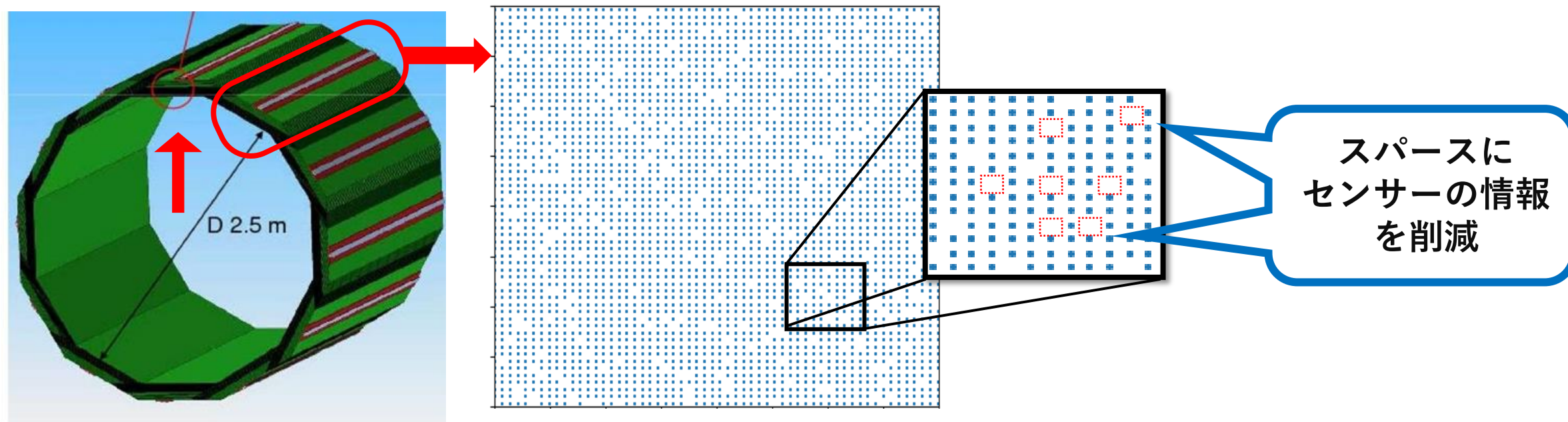
較正後の入射粒子の
エネルギー

- [1] 中 祐介,他8名, 日本物理学会第75回年次大会 ,2020年03月
- [2] 中 祐介,他 11 名, 日本物理学会 2020 年秋季大会 ,2020年9月



3. ILC SiD電磁カロリメータのエネルギー較正

ILC SiD電磁カロリメータにおけるデータ削減

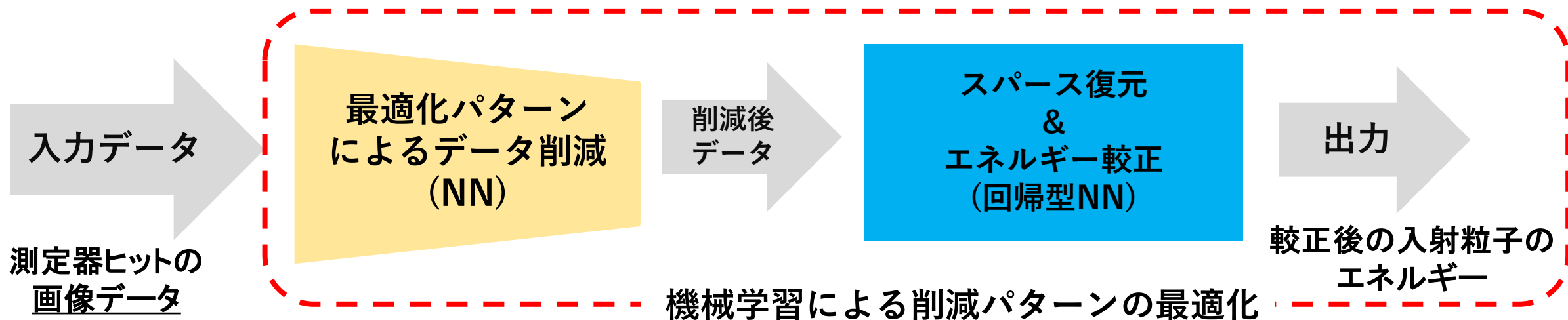


測定器ヒットデータを使用し、エネルギー較正における削減パターンの最適化を行った。

3. ILC SiD電磁カロリメータのエネルギー較正

削減パターンの最適化のための機械学習

削減パターン最適化機械学習(※)を使用して、パターンの最適化を行う。

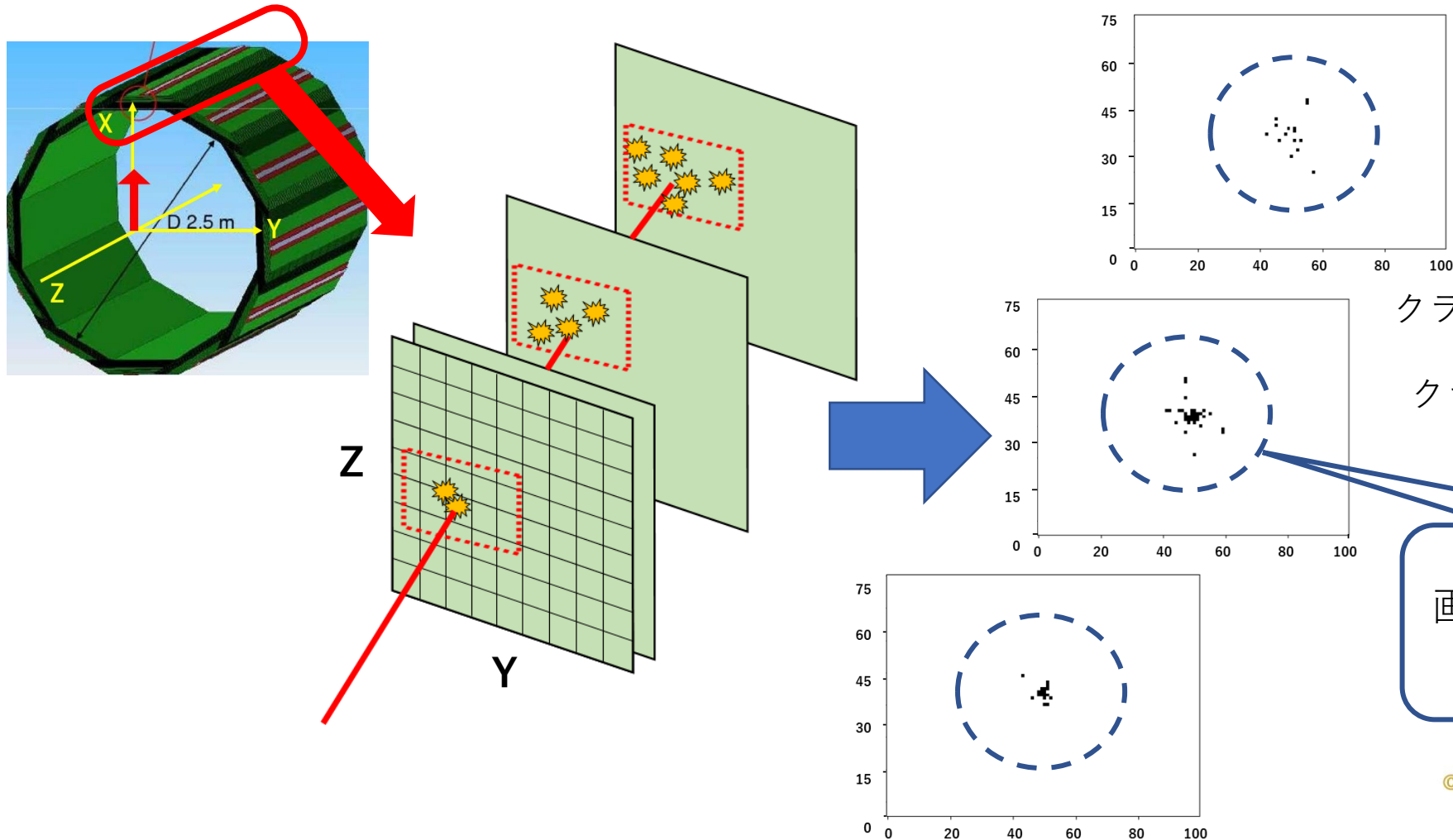


※ : M.Yoshida, H. Nagahara, et.al.,
"Joint optimization for compressive video sensing and reconstruction under hardware constraints",
The European Conference on Computer Vision (ECCV2018), 2018,

4. パターン最適化NNを使用した削減パターンの最適化

4-1. 削減パターン最適化の手法

削減パターン最適化機械学習を行うため、測定器ヒットデータを画像に変換した。



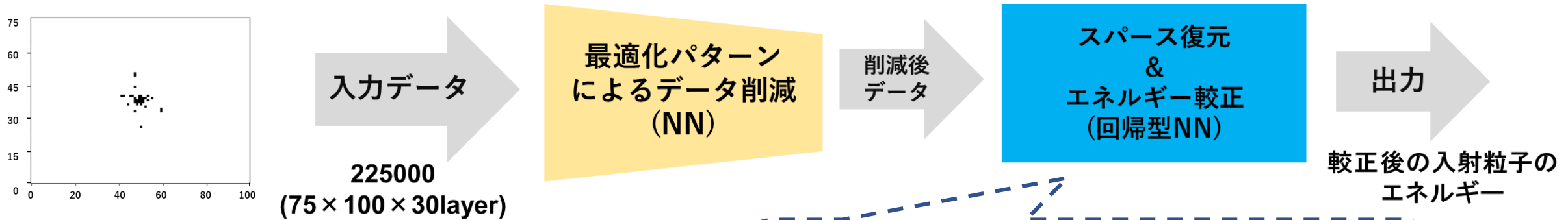
クラスター重心の位置座標が z y 平面の原点になるようにシフトさせ
クラスターの広がりをカバーする範囲
75 (z) × 100 (y) センサー
の領域を切り取る

画像の中心にクラスター重心
が来るように成形

3. ILC SiD電磁カロリメータのエネルギー較正

削減パターンの最適化のための機械学習

削減パターン最適化機械学習(*)を使用して、パターンの最適化を行う。



使用データ

電子の2~6.5GeVのデータを使用

学習用データ	~7万クラスター
評価用データ	~2万クラスター
テストデータ	~1万クラスター

中間層	4層
ノード数	100
入力パラメータ	225000
エポック数	10
オプティマイザー	Adam
活性化関数	ReLU

使用した計算機 CPUメモリ 512GB
GPU NVIDIA A100(GPUメモリ 80GB)

4. パターン最適化NNを使用した削減パターンの最適化

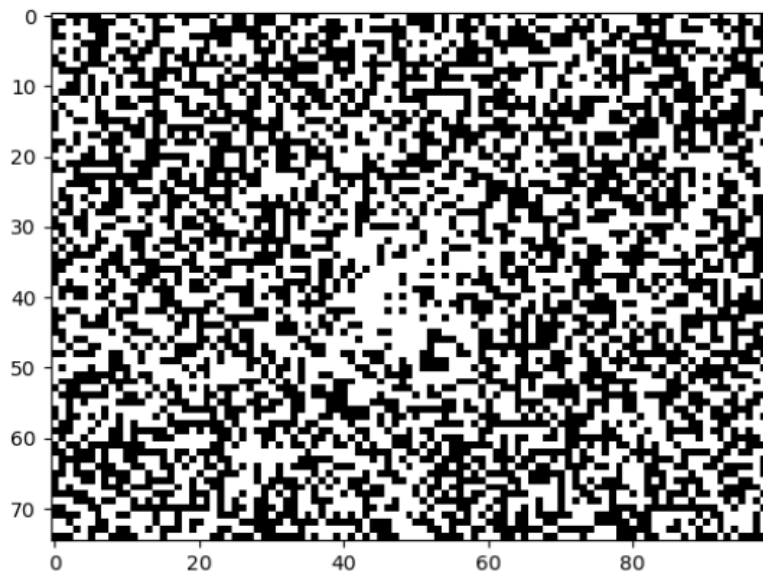
4-2. 最適化パターンによる削減後のILC SiD電磁カロリメータのエネルギー較正

使用データ 2~6.5GeVの電子のデータを使用
下のような削減パターンを30枚(層)得た。

1層



2層



3層

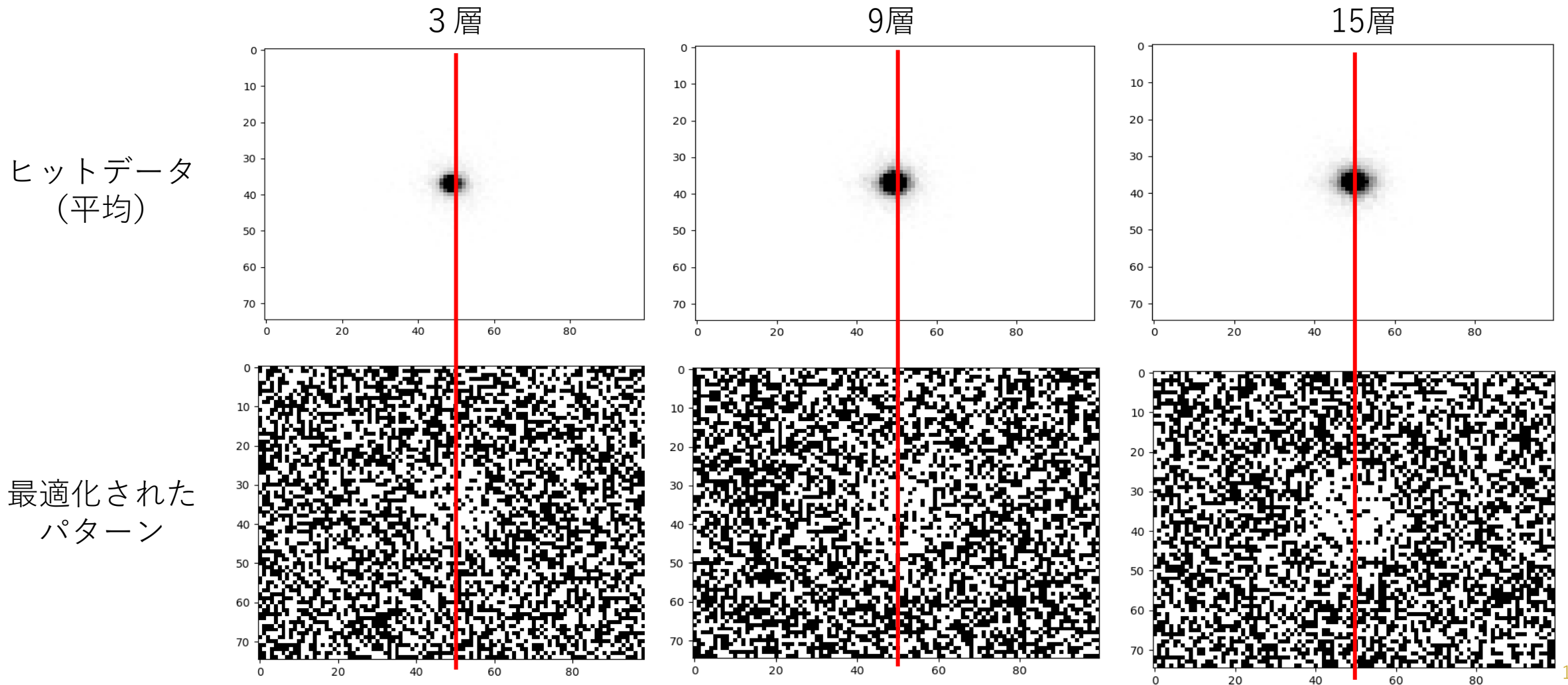
。 。 。 。



□ : データを保持
■ : データを削除

4. パターン最適化NNを使用した削減パターンの最適化

4-2. 最適化パターンによる削減後のILC SiD電磁カロリメータのエネルギー較正



4. パターン最適化NNを使用した削減パターンの最適化

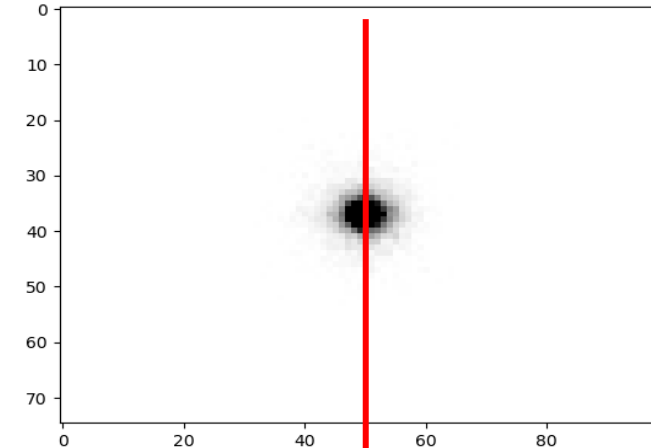
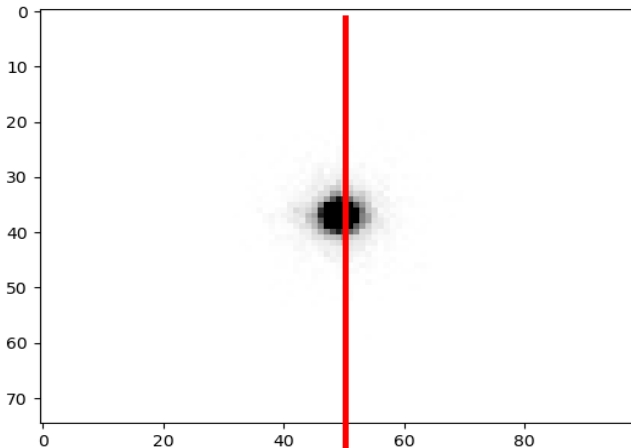
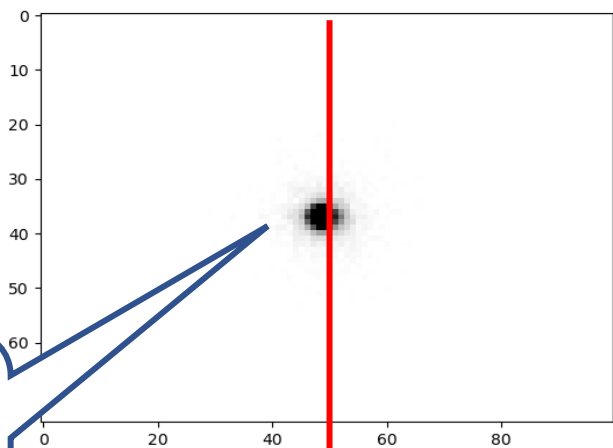
4-2. 最適化パターンによる削減後のILC SiD電磁カロリメータのエネルギー較正

3層

9層

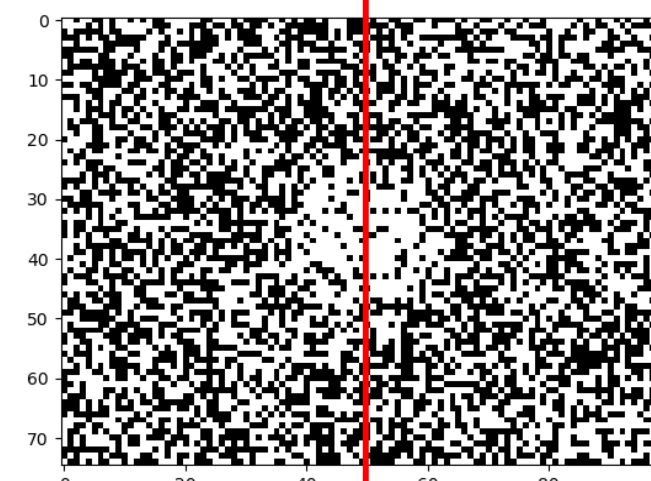
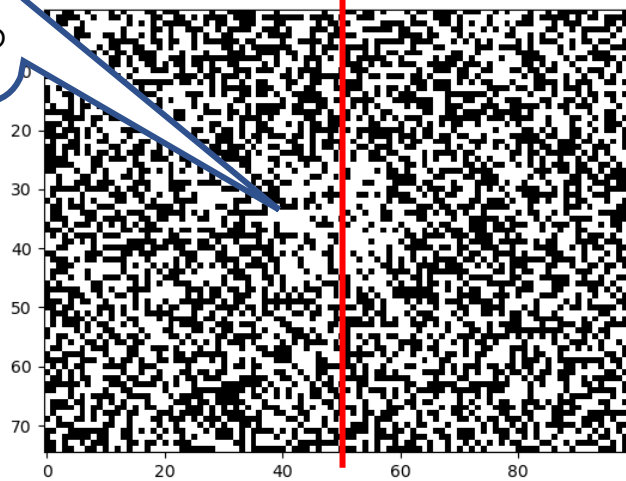
15層

ヒットデータ
(平均)



ヒットのある箇所
を保持するように
最適化されている

最適化された
パターン



4. パターン最適化NNを使用した削減パターンの最適化

4-2. 最適化パターンによる削減後のILC SiD電磁カロリメータのエネルギー較正

3層

9層

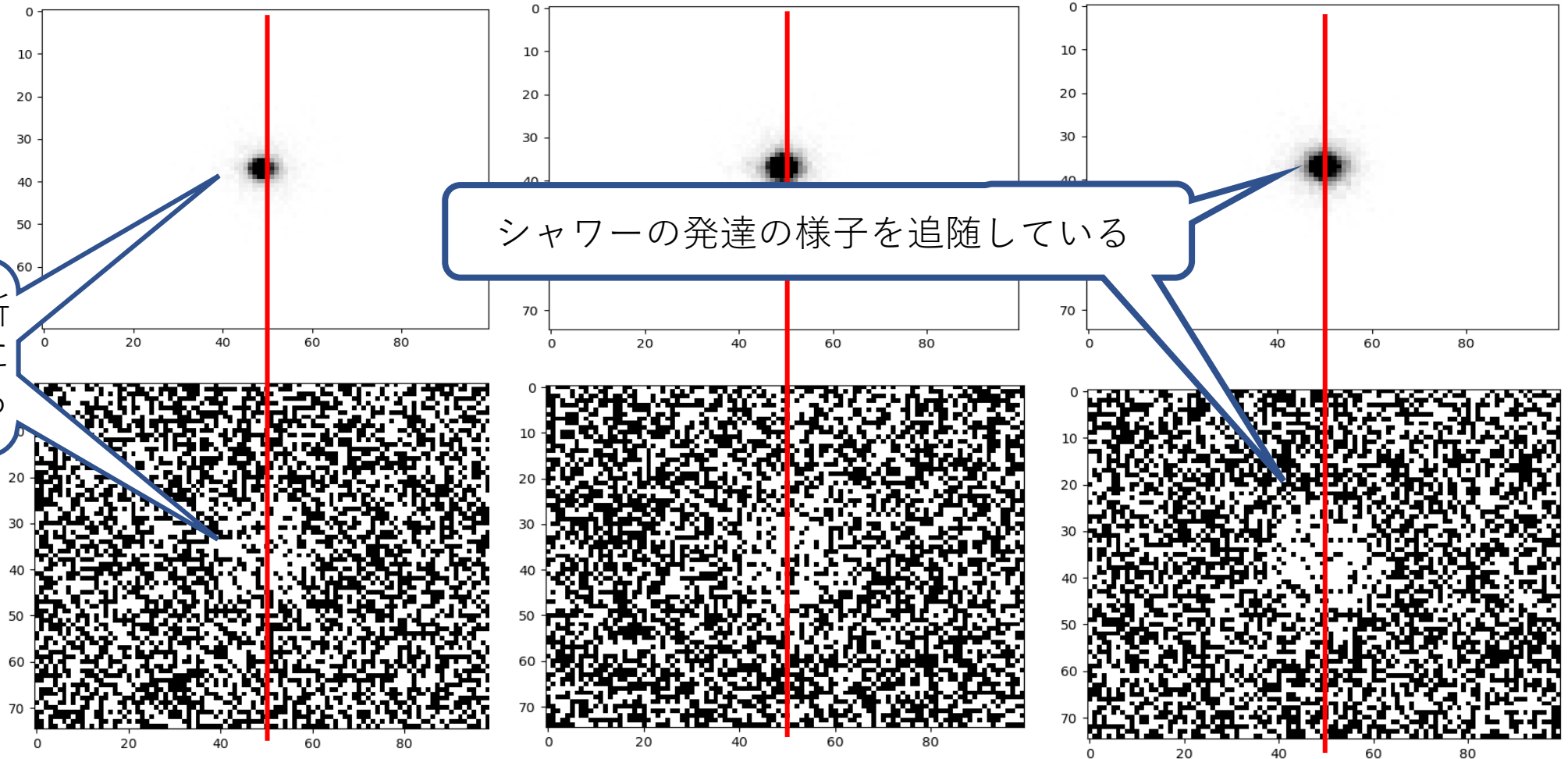
15層

ヒットデータ
(平均)

シャワーの発達の様子を追従している

ヒットのある箇所
を保持するように
最適化されている

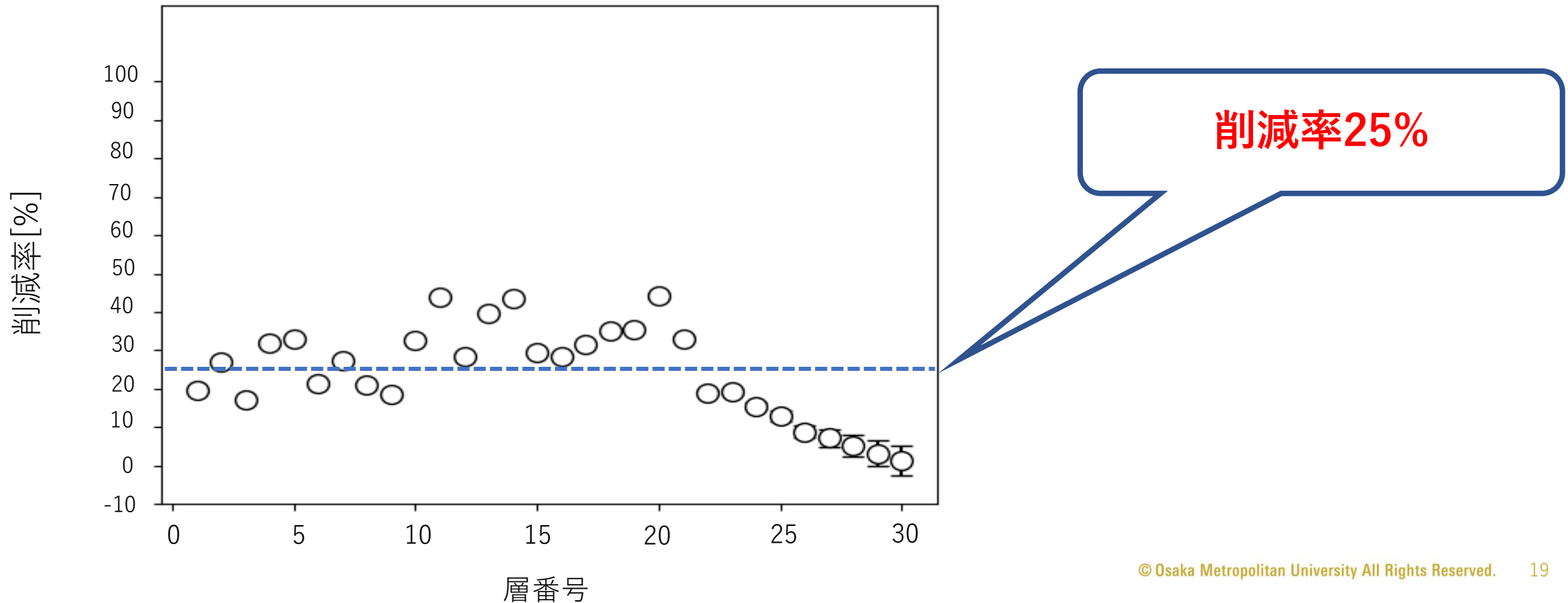
最適化された
パターン



4. パターン最適化NNを使用した削減パターンの最適化

4-2. 最適化パターンによる削減後のILC SiD電磁カロリメータのエネルギー較正 測定器ヒットに対する削減率

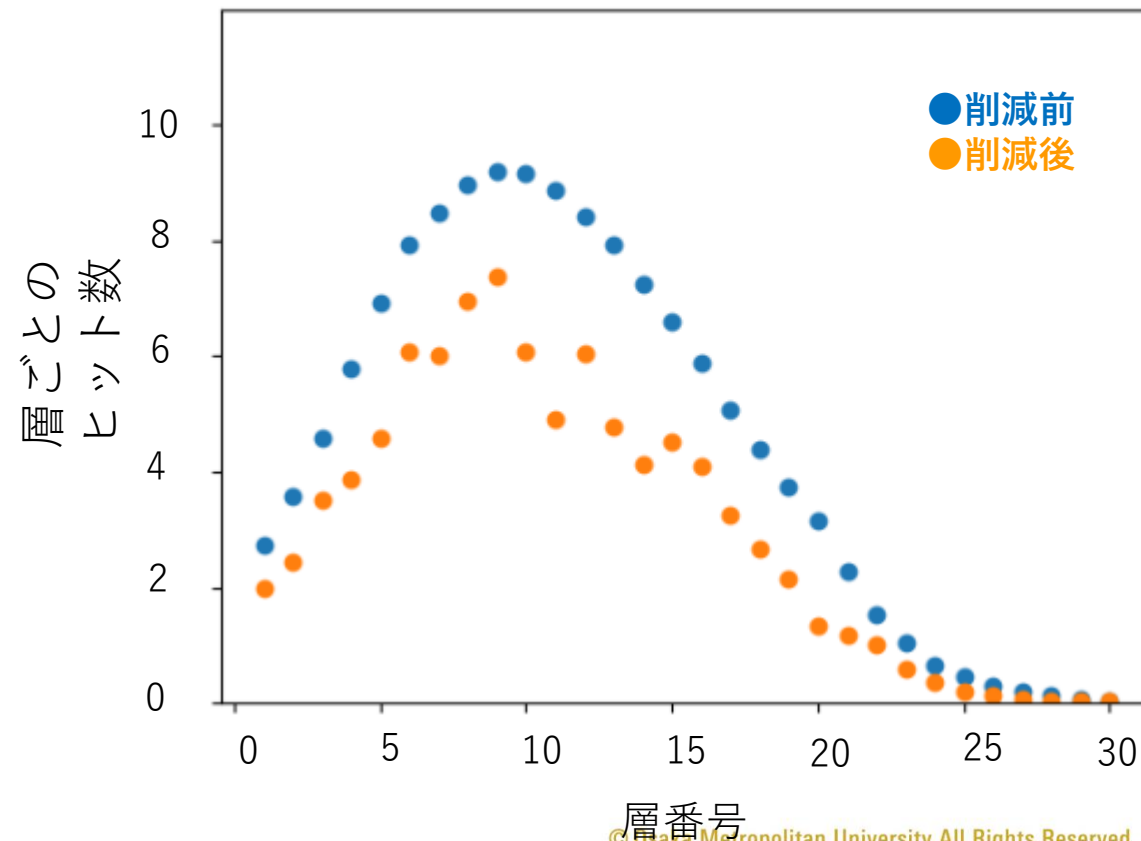
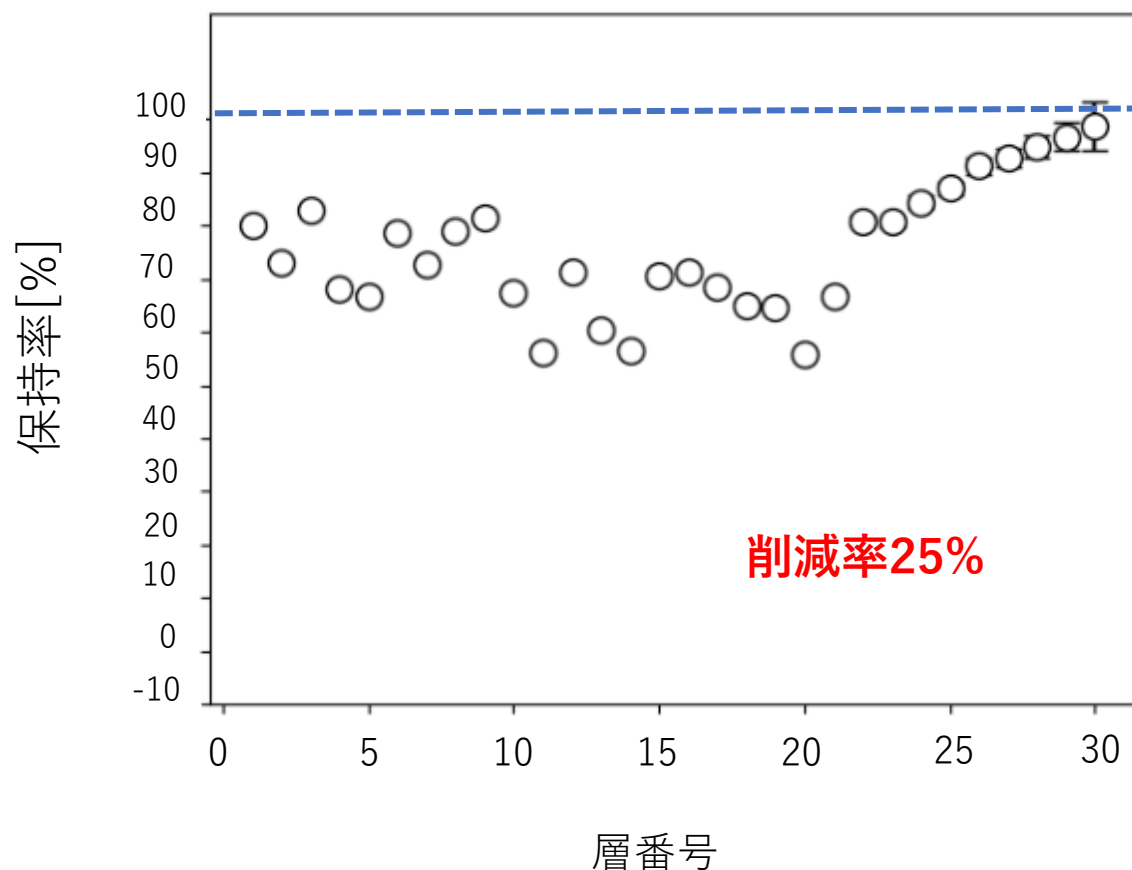
入射粒子のエネルギー2~6.5GeV



4. パターン最適化NNを使用した削減パターンの最適化

4-2. 最適化パターンによる削減後のILC SiD電磁カロリメータのエネルギー較正 測定器ヒットに対する保持率

入射粒子のエネルギー2~6.5GeV

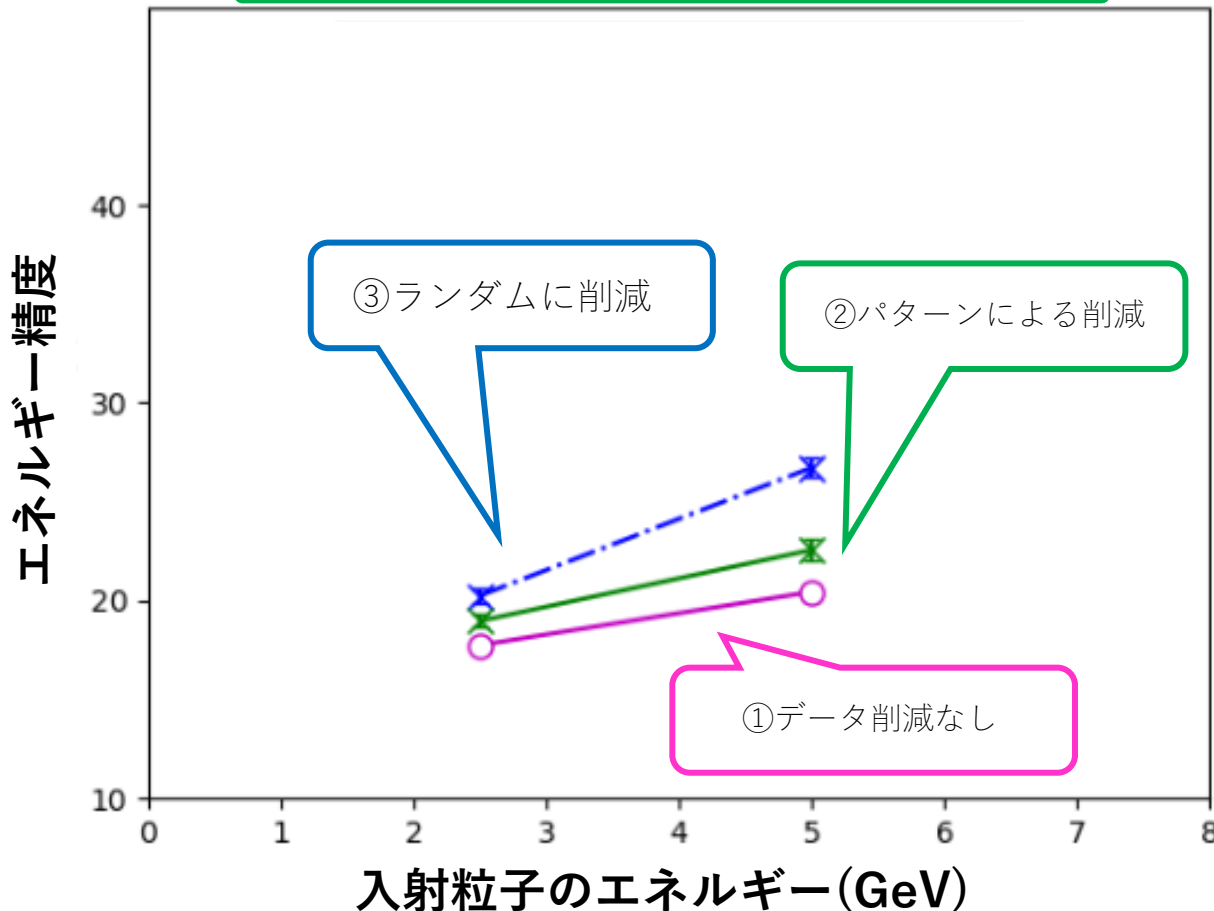


4. パターン最適化NNを使用した削減パターンの最適化

4-2. 最適化パターンによる削減後のILC SiD電磁カロリメータのエネルギー較正

ECLエネルギー較正精度の比較

電子のエネルギー分解能 $\Delta E(\%) \times \sqrt{E}$



使用データ

- ・電子の2~6.5GeVのデータを使用
- ・データ数 10万クラスター

- ①データ削減なし
- ②最適化した削減パターンによるデータ削減 (25%の情報を削除)
- ③ランダムに25%データを削除

エネルギー分解能

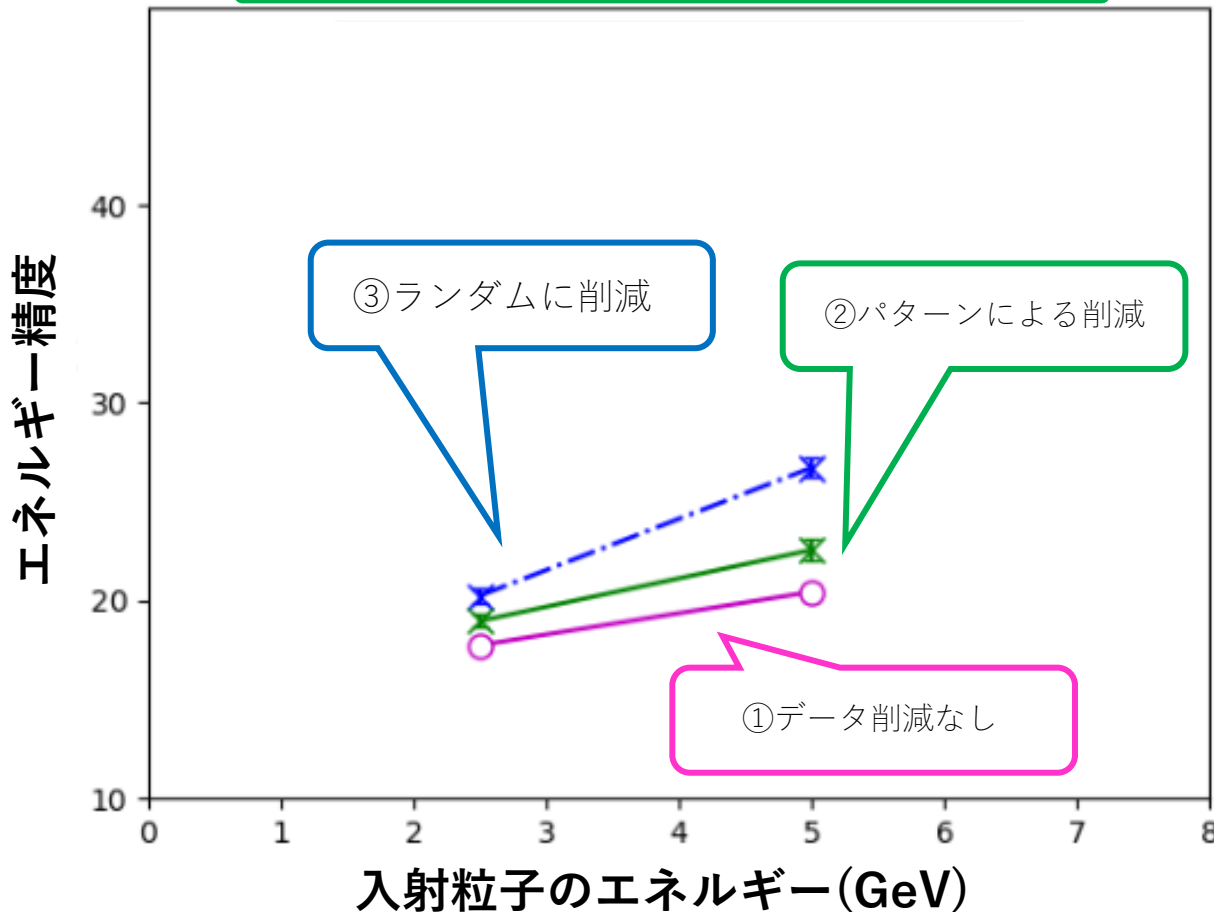
	2.5GeV	5.0GeV	
削減なし	17.74%	20.44%	↓ 7~9%
パターンによる削減	18.97%	22.55%	
ランダム削減	20.25%	26.71%	↓ 14~24%

4. パターン最適化NNを使用した削減パターンの最適化

4-2. 最適化パターンによる削減後のILC SiD電磁カロリメータのエネルギー較正

ECLエネルギー較正精度の比較

電子のエネルギー分解能 $\Delta E(\%) \times \sqrt{E}$



使用データ

- ・電子の2~6.5GeVのデータを使用
- ・データ数 10万クラスター

- ①データ削減なし
- ②最適化した削減パターンによるデータ削減
(**25%**の情報を削除)
- ③ランダムに**25%**データを削除

- ・機械学習で最適化したパターンによるデータ削減で削減なしの場合とほぼ同等(~8%劣化)の較正結果が得られた。
- ・ランダムにデータ削減を行うと精度が~20%劣化した。



機械学習を使用することで
較正精度の劣化を抑えてデータ削減できた

機械学習を用いたスパースサンプリングを適用し ILC SiD 電磁カロリメータのエネルギー較正性能の評価を行った

- 機械学習を用いて、サンプリングパターンを最適化することができた。
- 最適化パターンによるデータ削減を行った場合、(25%削減)
 - ・データ削減なしの場合と遜色ない(~8%の劣化)結果が得られた。
 - ・ランダム削減よりもエネルギー較正精度がよかった。

機械学習を用いたスパースサンプリングを適用し ILC SiD 電磁カロリメータのエネルギー較正性能の評価を行った

今後の展望

1. 適応手法の一般化

クラスター重心の位置座標が $z-y$ 平面の原点になるようにシフトさせ、画像化している。
→任意の場所に入力されたデータに対する最適化を試す

2. ネットワークデザインの改良

パターン最適化のために作成した画像化データで回帰をしている。
→画像化する前のヒットデータで回帰を行う。

ありがとうございました。

謝辞

- 本研究は、以下の助成により実施されました。

阪大RCNP研究プロジェクト

阪大IDS学際プロジェクト

阪大CMC公募型利用制度

学際大規模情報基盤共同利用(JHPCN) 公募型共同研究

文部科学省補助事業「ダイバーシティ研究環境実現イニシアティブ（牽引型）」

大阪公立大学 女性研究者 研究実践力強化支援プログラム(REPECT)

日本学術振興会 科学研究費助成事業 学術変革領域研究(A) 「学習物理」

- ILC SiD ECL用 MCデータを使用させて頂きました。