

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

© Osaka Metropolitan University All Rights Reserved.

阪公大理^A, 阪公大NITEP^B, 阪大RCNPC^C, 阪大IDS^D, 九大理^E, KEK素核研^F
加藤睦代^A, 岩崎昌子^{A, B, C, D}, 長原 一^{C, D}, 末原大幹^E,
山田 悟^F, 中島悠太^{C, D}, 武村紀子^{C, D}, 中野貴志^{C, D}

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

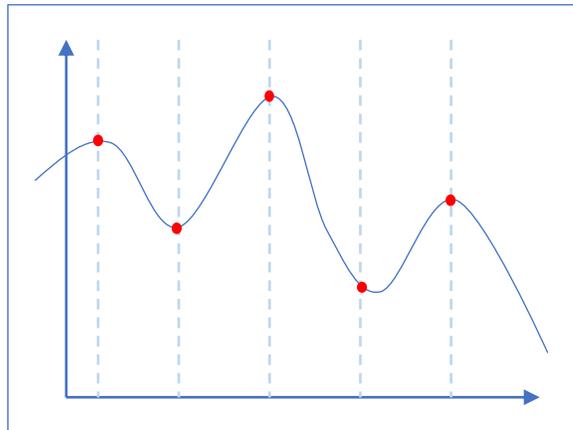


**機械学習を用いたスパースサンプリングを導入し
リアルタイムにデータサイズの低減を目指す**

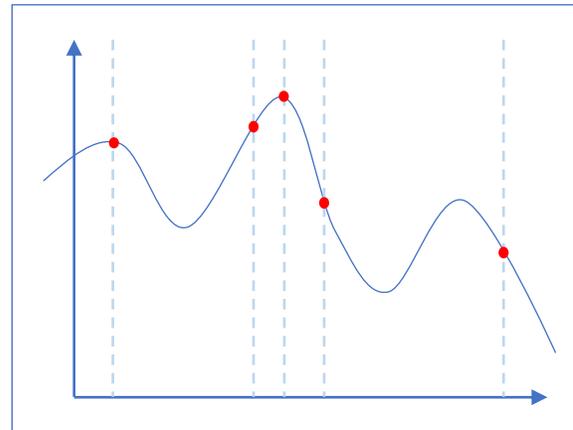
スパースサンプリング・復元による信号観測

元信号に対して**観測点が少ない**、劣決定問題を解く手法

…ブラックホールを「可視化」した手法として有名



通常サンプリング



スパースサンプリング

・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。

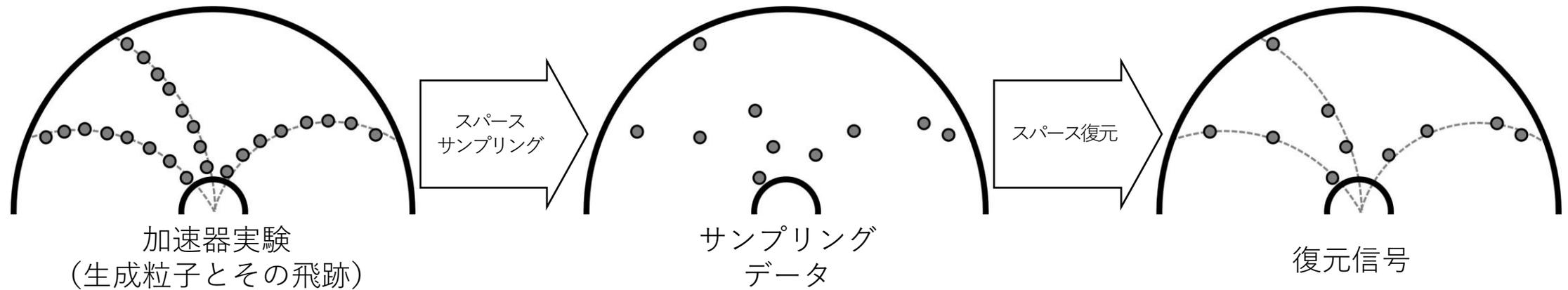
→多数ある観測点から本質的な観測点の選別も可能

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

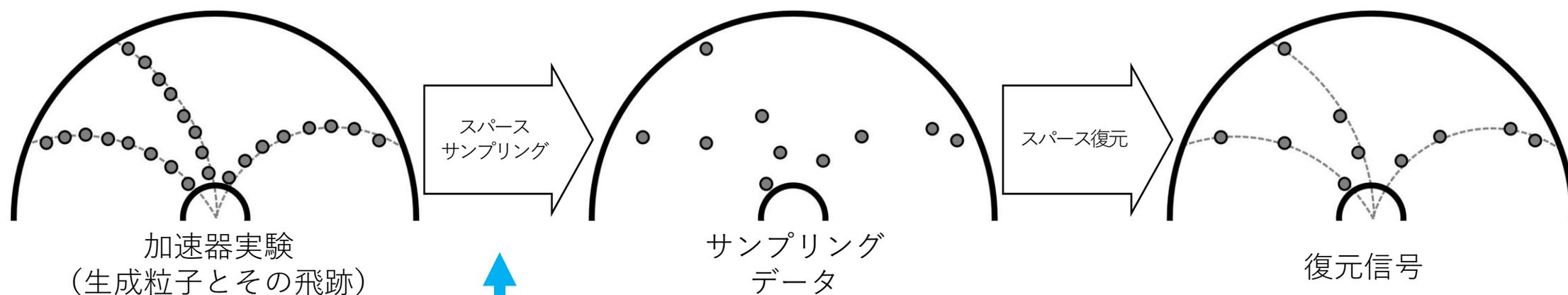
素粒子実験に、スパースサンプリング・スパース復元を適用させることを目指す。



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。

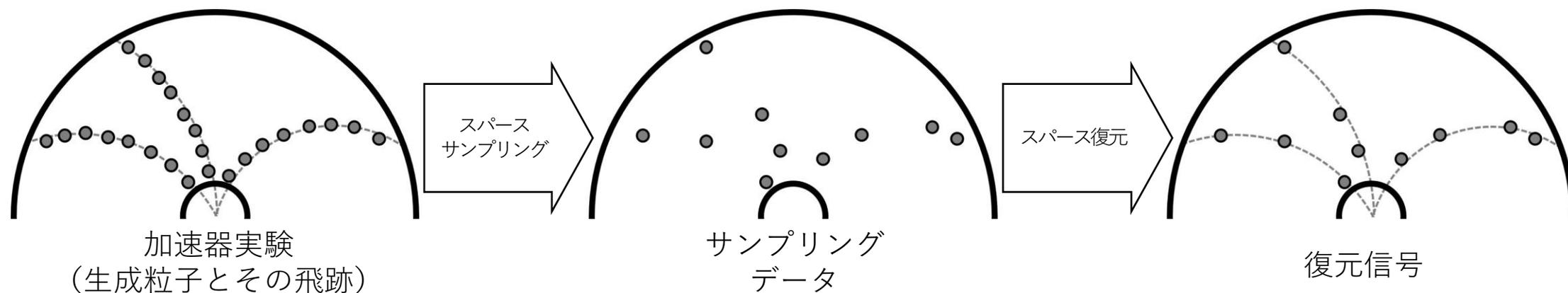


**DAQ時に必要な処理は「あらかじめ決められたパターンにしたがうサンプリング処理」
→ リアルタイムにデータ削減可能**

2. スパースサンプリング・復元の原理

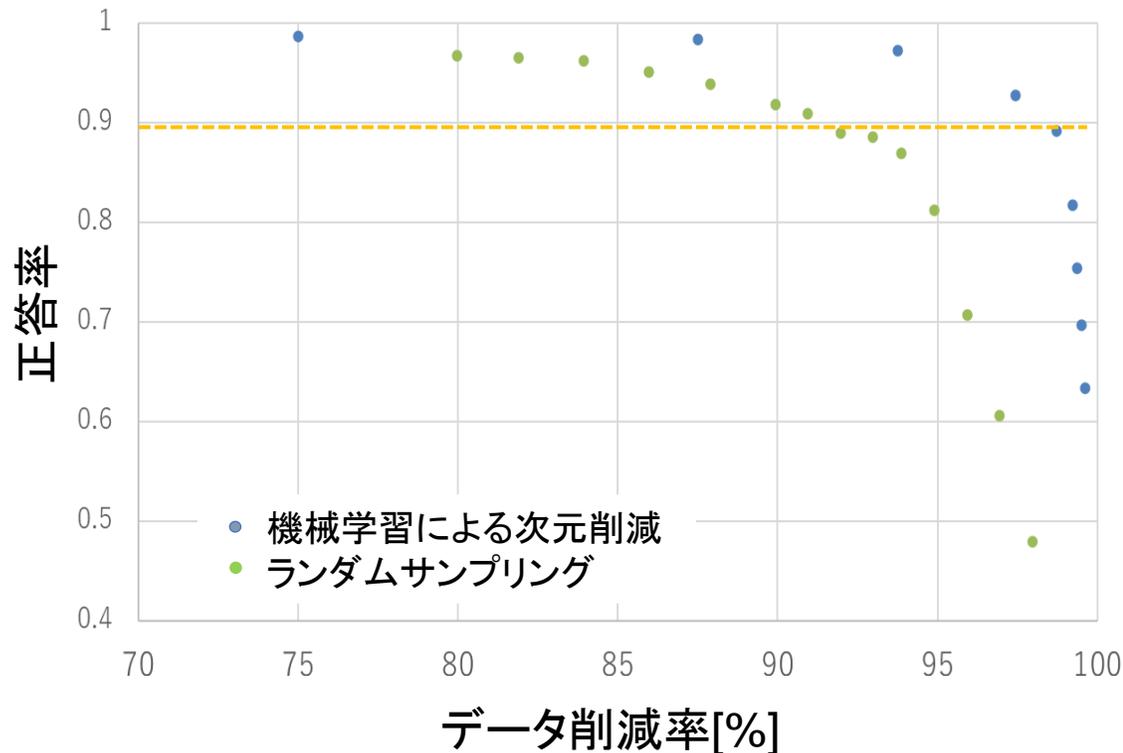
素粒子実験へのスパースサンプリングの適用

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基礎的な評価を行った

機械学習による次元削減を用いた文字分類（予備実験）



データを
1/10(ランダムサンプリング)
1/50(機械学習による次元削減)
に圧縮しても、
90%の識別性能を維持

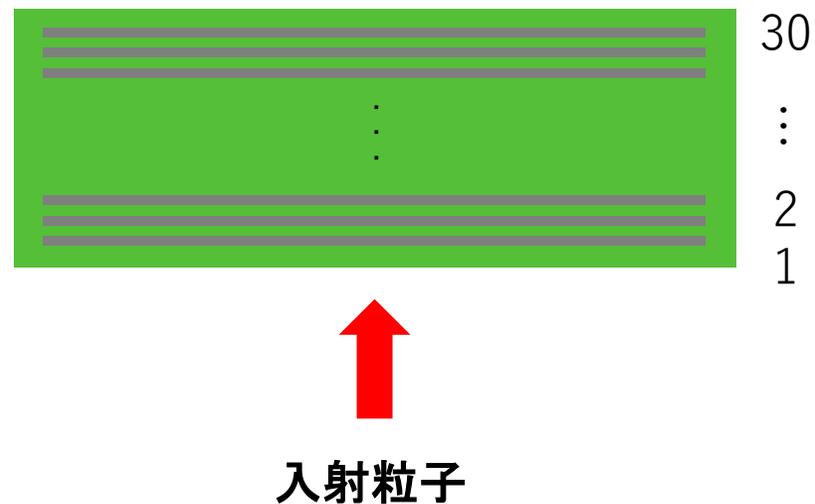
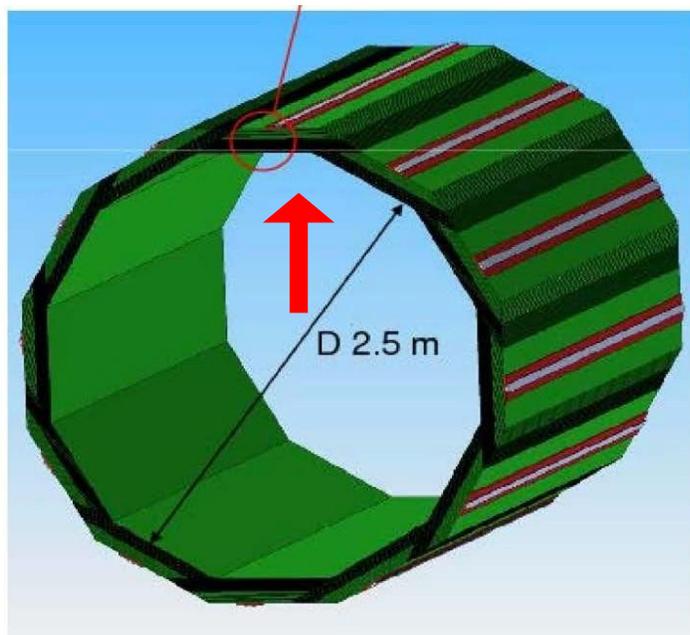


ECLのエネルギー較正でも
機械学習を用いたデータ削減を目指す

3-1. ILC SiD電磁カロリメータのエネルギー較正

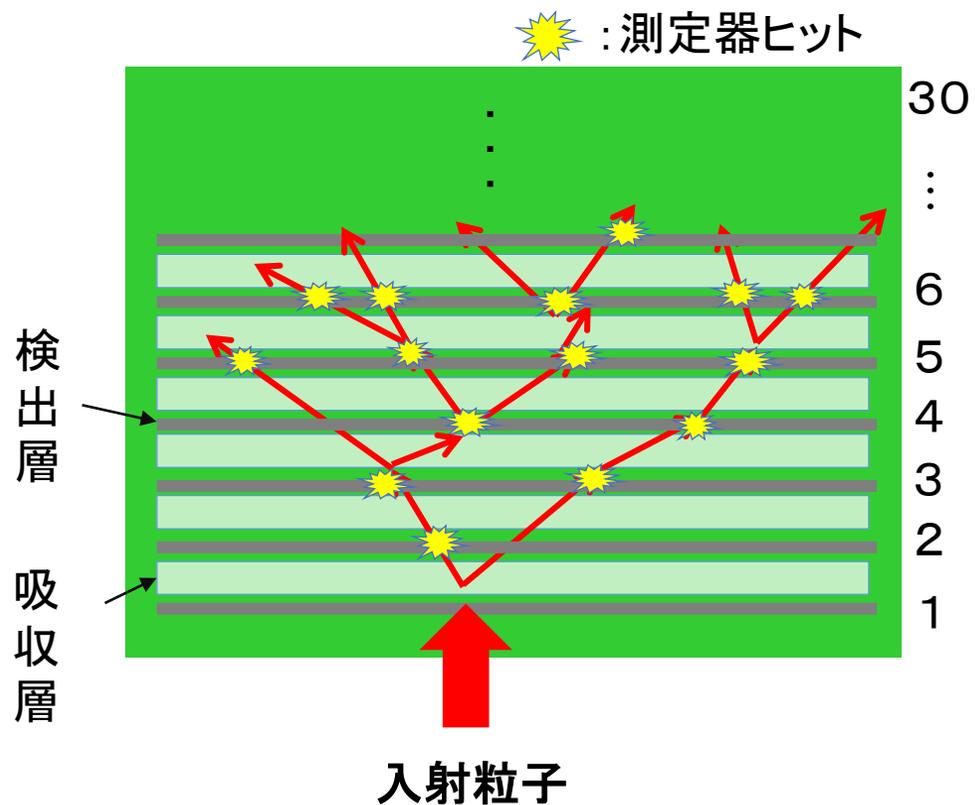
ILC SiD 電磁カロリメータ(ECL)

- 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) **$(17/\sqrt{E \oplus 1})\%$**



ILC TDR, vol.4,Page 89 [arXiv:1306.6329\[physics.ins-det\]](https://arxiv.org/abs/1306.6329)

ECLでのエネルギー測定方法



- 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
→測定器ヒットとして測定

同一入射粒子に由来する測定器ヒットを
集めてクラスタリングした。

3-1. ILC SiD電磁カロリメータのエネルギー較正

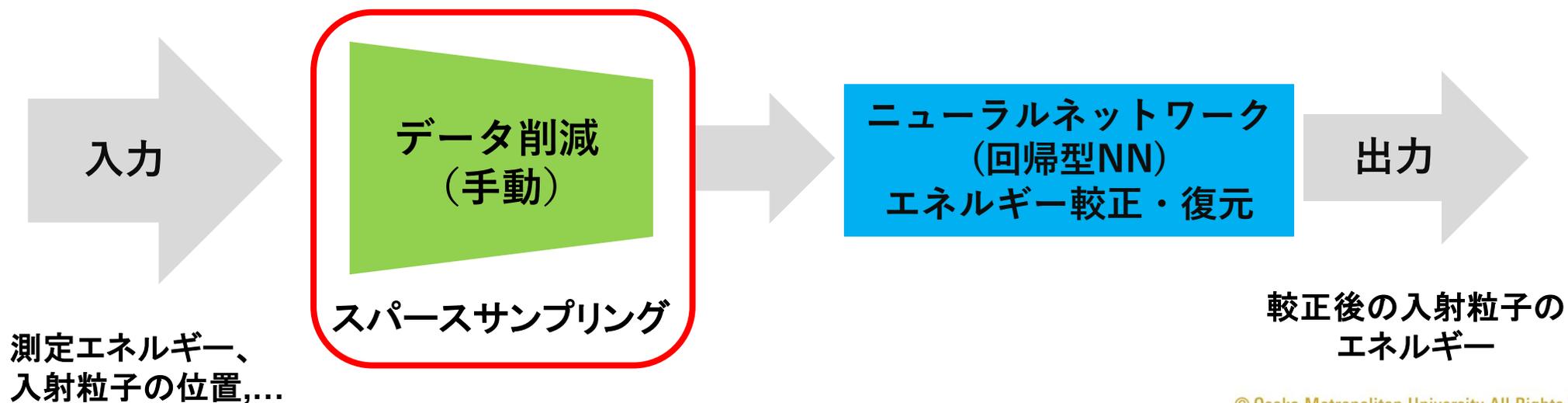


- [1] 中 祐介,他8名, 日本物理学会第75回年次大会 ,2020年03月
- [2] 中 祐介,他 11 名, 日本物理学会 2020 年秋季大会 ,2020年9月

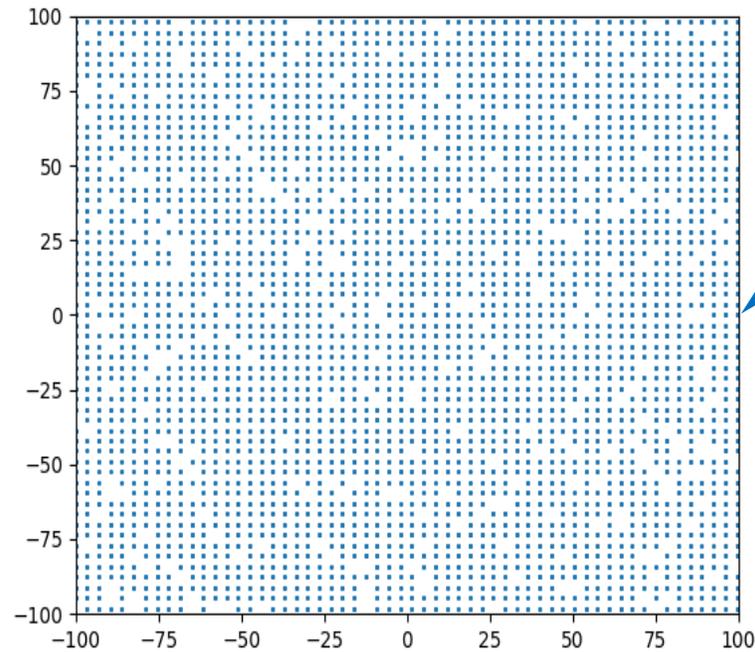
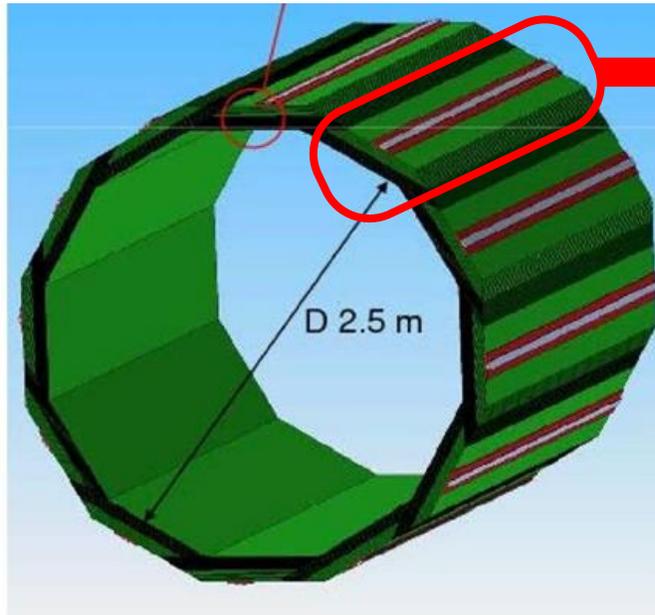
3-1. ILC SiD電磁カロリメータのエネルギー較正



- [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



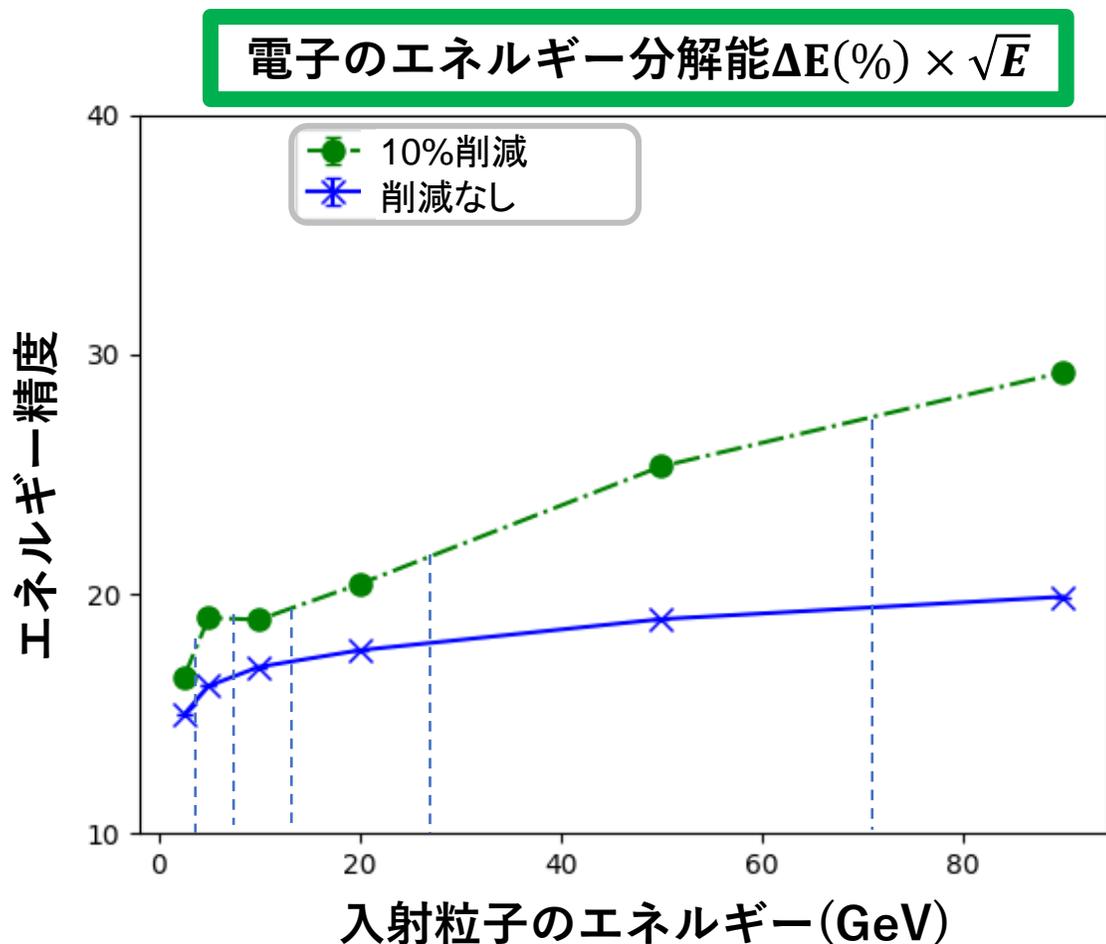
ECLにおけるデータ削減方法



ランダムに10%の
センサーを削減

今回は、削減のパターンとしてランダムサンプリングを採用した。

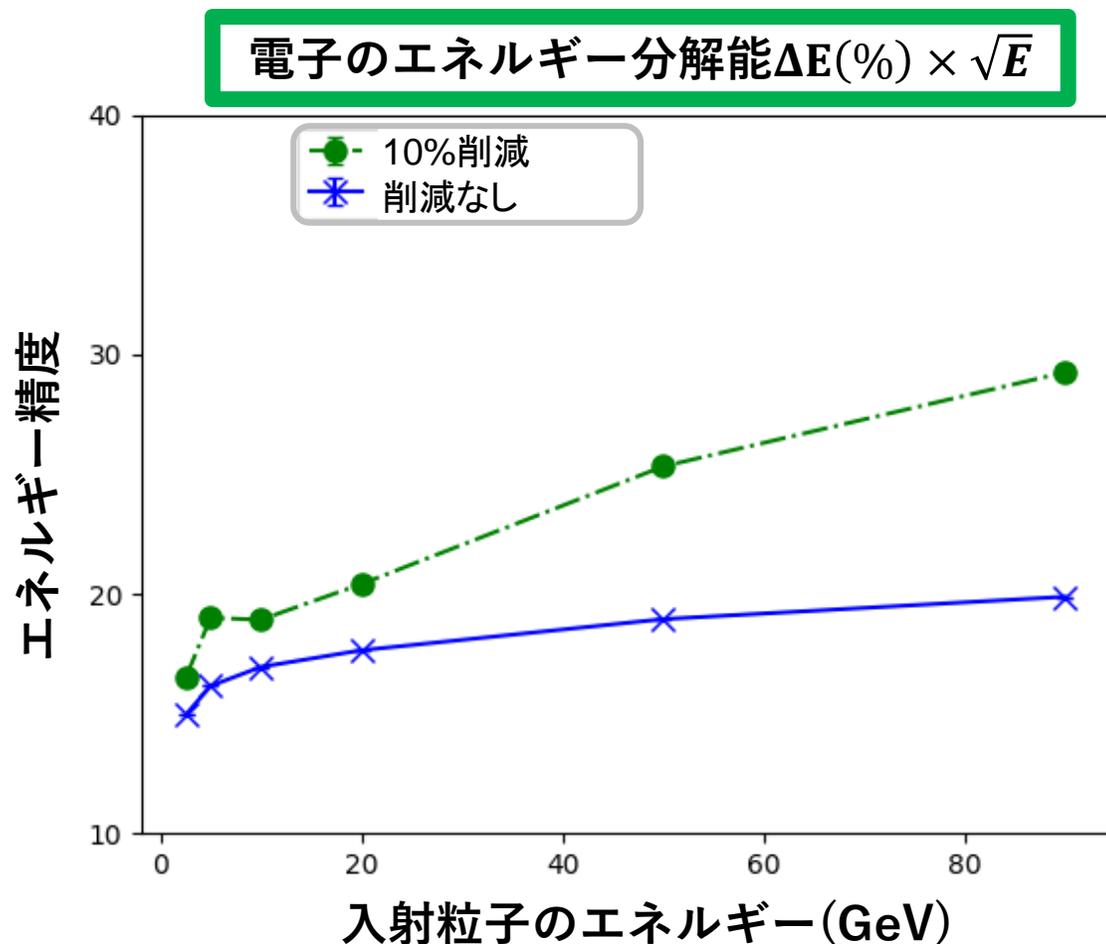
NN(回帰)を使用したエネルギー較正



電子のエネルギー回帰範囲を5個に分割して回帰を行った。

エネルギーの範囲	$\bar{E}(\text{GeV})$	$\Delta E(\text{GeV})$	#hit
2~3GeV	2.5	1.0	100
3.5~6.5GeV	5.0	3.0	177
7~13GeV	10.0	6.0	280
13.5~26.5GeV	20.0	13.0	590
27~73GeV	50.0	46.0	873
80~100GeV	90.0	20.0	1270

NN(回帰)を使用したエネルギー較正



高エネルギー領域では、

- ヒット数が多い
→ 入力が大きいため、高性能のNNが必要
- ΔE が大きいため
→ より学習用データ量が必要

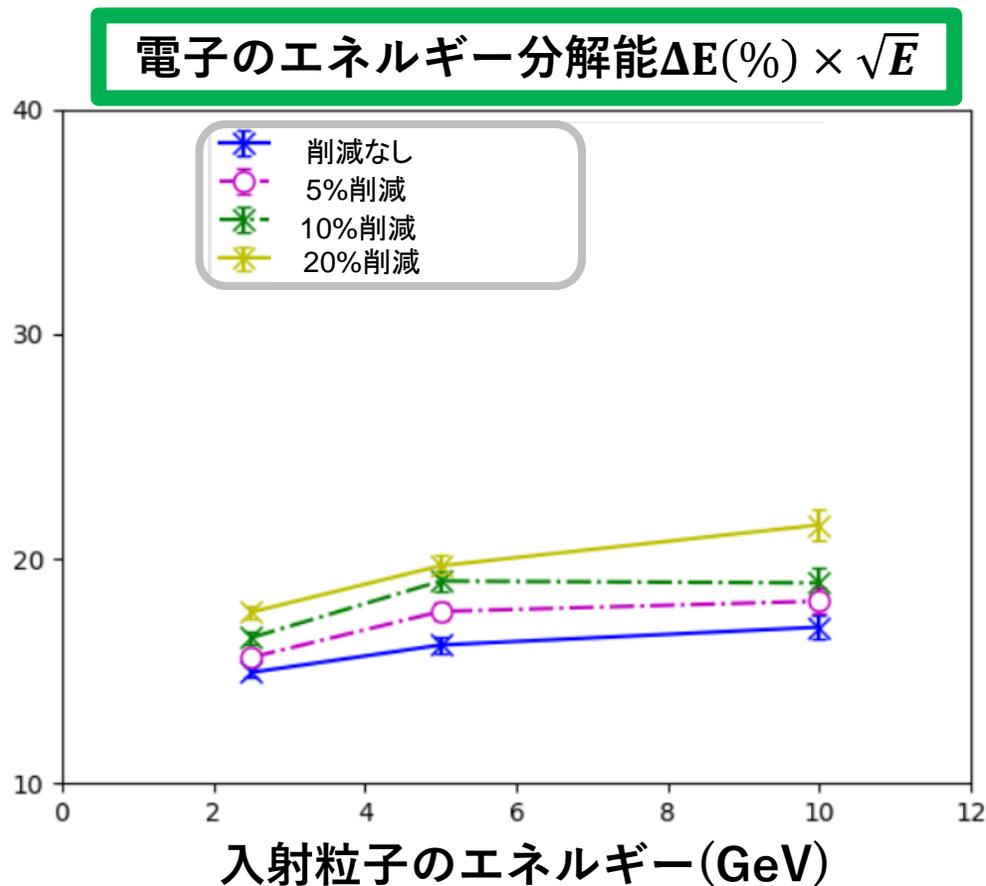
➡ 高エネルギー領域では評価が難しい。

10GeV以下では、精度の悪化がデータ削減による統計誤差(5%)程度となった。

今後2~10GeVの範囲のデータを使用して評価する。

NN(回帰)を使用したエネルギー較正

エネルギー精度



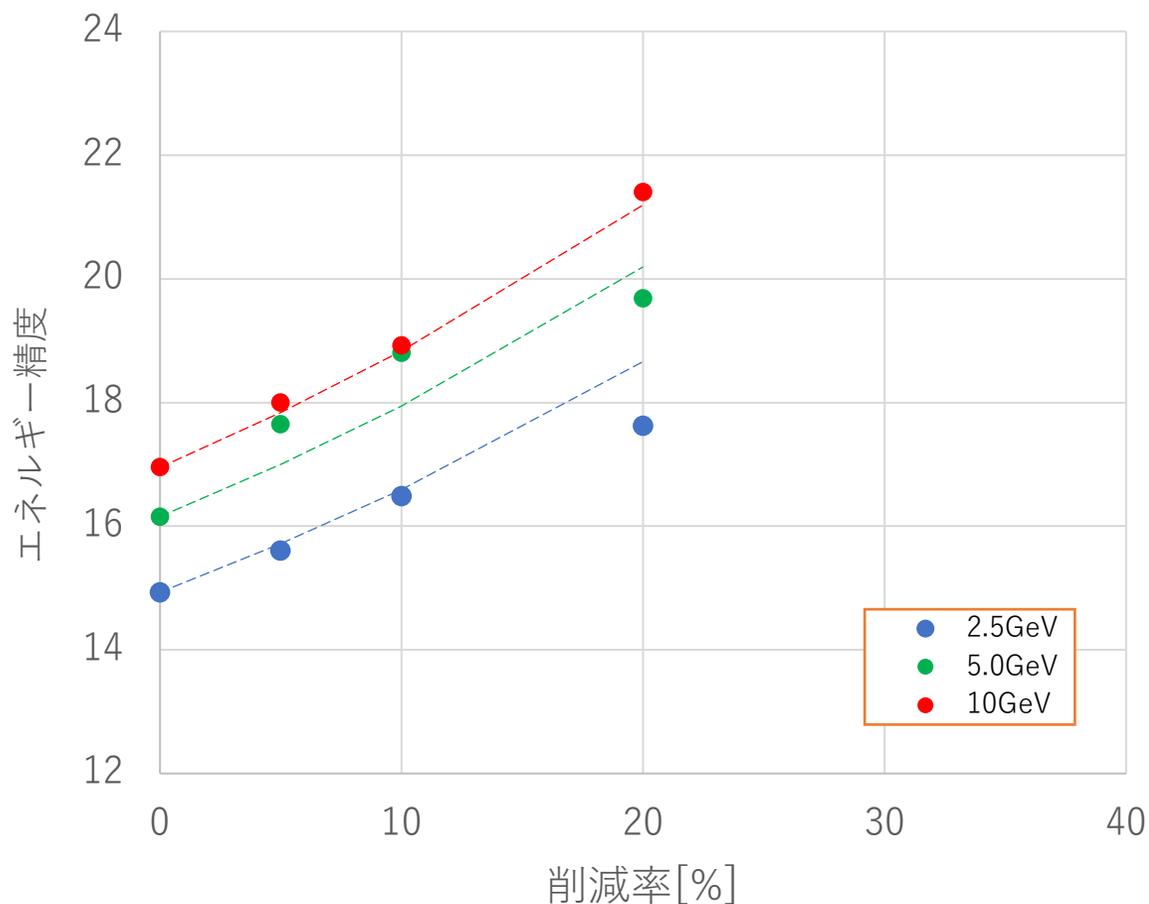
2~10GeVの測定器ヒットデータに 削減率5%, 10%, 20%でランダムサンプリングを適用し、削減なしの場合のエネルギー較正性能と比較し、評価を行った。

削減率に対するエネルギー分解能

	2.5GeV	5.0GeV	10GeV
削減なし	14.93	16.15	16.95
5%削減	15.60	17.65	18.10
10%削減	16.48	19.00	18.92
20%削減	17.62	19.68	21.50

NN(回帰)を使用したエネルギー較正

データ削減後のエネルギー較正精度



→手動でランダムにデータ削減を行った場合、2~10GeVの範囲であれば、データ削減による統計誤差の悪化と比較して同等か少し良い精度が得られた。

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でさらに削減を行っても精度を維持できる可能性がある



**機械学習を用いた
サンプリングパターンの最適化を行う。**

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基礎的な評価を行った

- ・ サンプリング手法として、**ランダムサンプリング**を用いた。

ランダムサンプリング後のエネルギー較正の結果

2~10GeVの範囲であれば、データ削減による統計誤差の悪化と比較して同等か少し良い精度が得られた。

今後の展望

- ・ 高統計のデータを使用して評価の精度を上げる。
- ・ 機械学習を用いたサンプリングパターンの最適化を行う。

4. 機械学習を用いたパターン最適化の手法

