Machine Learning Estimation on the Trace of Inverse Dirac operator

Hiroshi Ohno[1]

In Collaboration with Benjamin J. Choi[1], Takayuki Sumiomto[2] and Akio Tomiya[3]

[1]University of Tsukuba, [2]FLECT Co., [3]Tokyo Woman's Christian University

Program for Promoting Researches on the Supercomputer Fugaku

Large-scale lattice QCD simulation and development of AI technology

ML PhYs

German Japanese Workshop 2024

Johannes Gutenberg-Universität Mainz

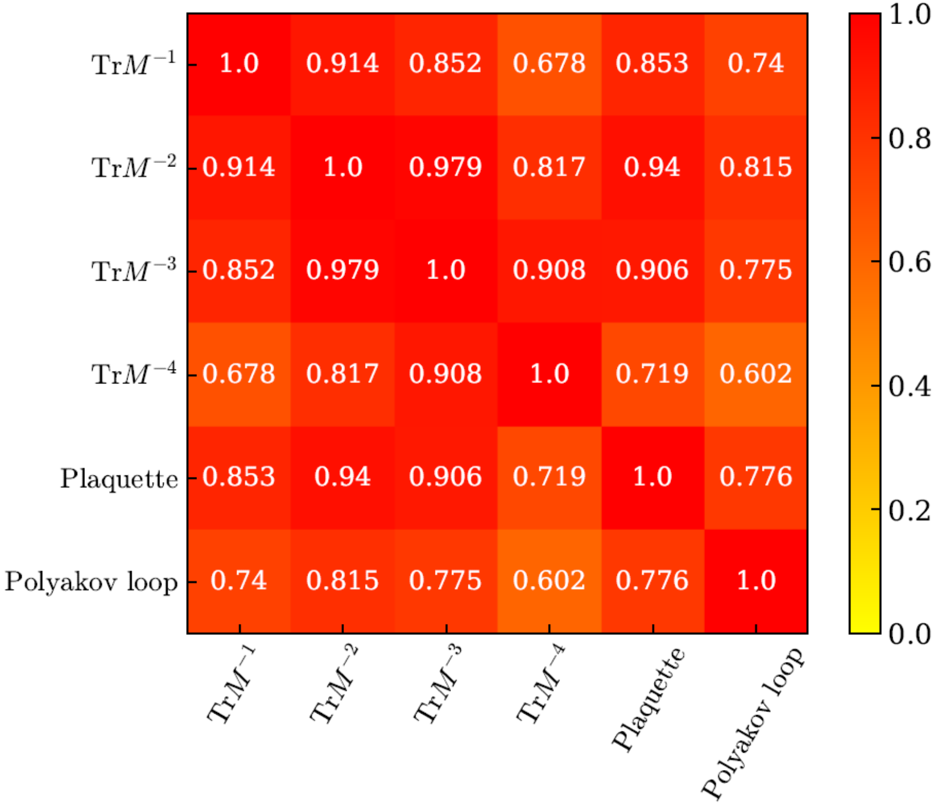
September 27, 2024

Motivation

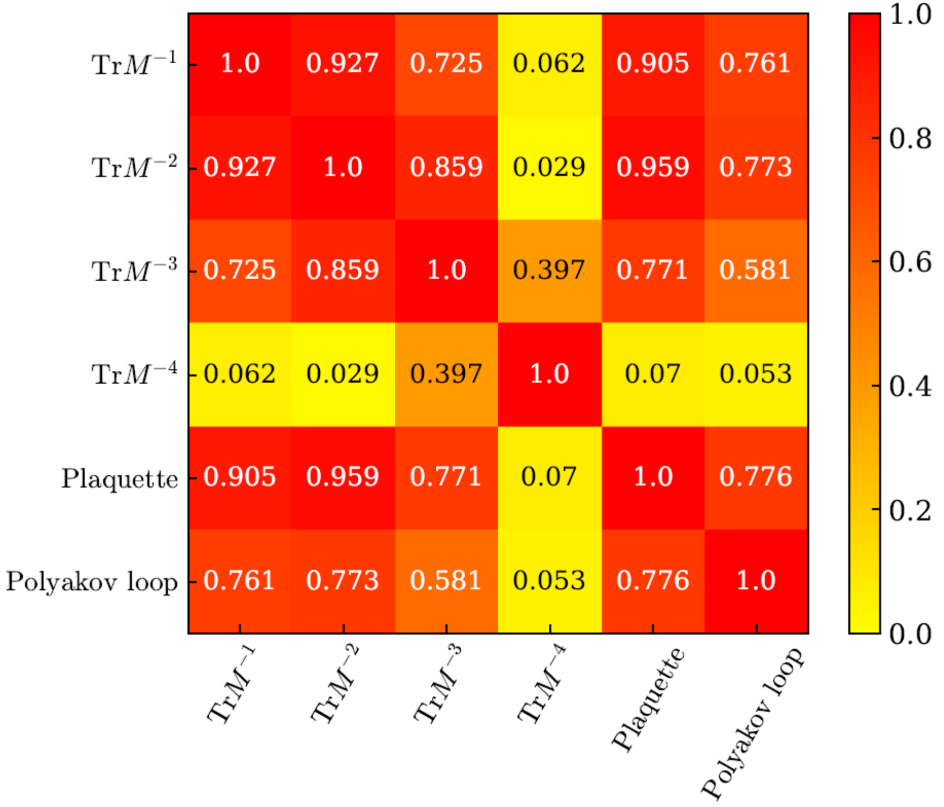
- The trace of operators ($\mathrm{Tr}O$) are often needed to calculate observables
  - e.g. $O = M^{-n}$, where M is the Dirac operator
- Calculating $\mathrm{Tr}M^{-n}$ on the lattice: **computationally demanding**
  - M: a large sparse matrix
  - The inversion of M → many matrix-vector multiplications within a CG solver
  - Stochastic trace estimation → many matrix inversions
- We study estimation of $\mathrm{Tr}M^{-n}$
  - from other observables which can be given **with a cheaper computational cost** such as plaquette, Polyakov loop, $\mathrm{Tr}M^{-m}$ where $m < n$
  - **using a machine learning technique**
  - Is this possible? → Let's see correlations between observables.

Correlations between observables

- An observable can be estimated from another
  - if there is sufficiently strong correlations between them
    → Sometimes this is true.



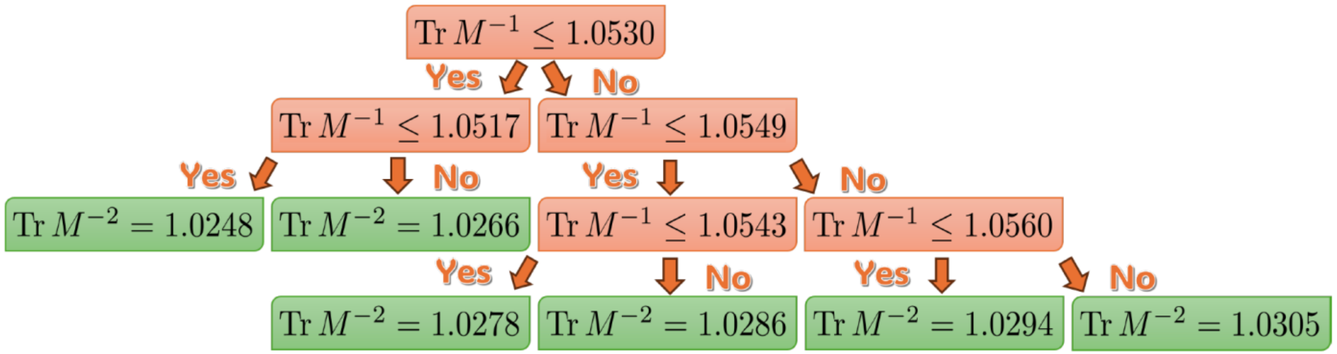
Heavier quark mass data



Lighter quark mass data

Machine learning estimation

- Similar to Yoon *et al.*, PRD **100** 014504 (2019).
- Modeling observables
  - input: X, target: Y, model: $f(X) = Y^P \cong Y$
  - The functional form of f is not known → Machine learning
  - Our choice: **Gradient Boosting Decision Tree**



Boosting

Iteratively creating an ensemble of improved models
→ Giving an averaged model

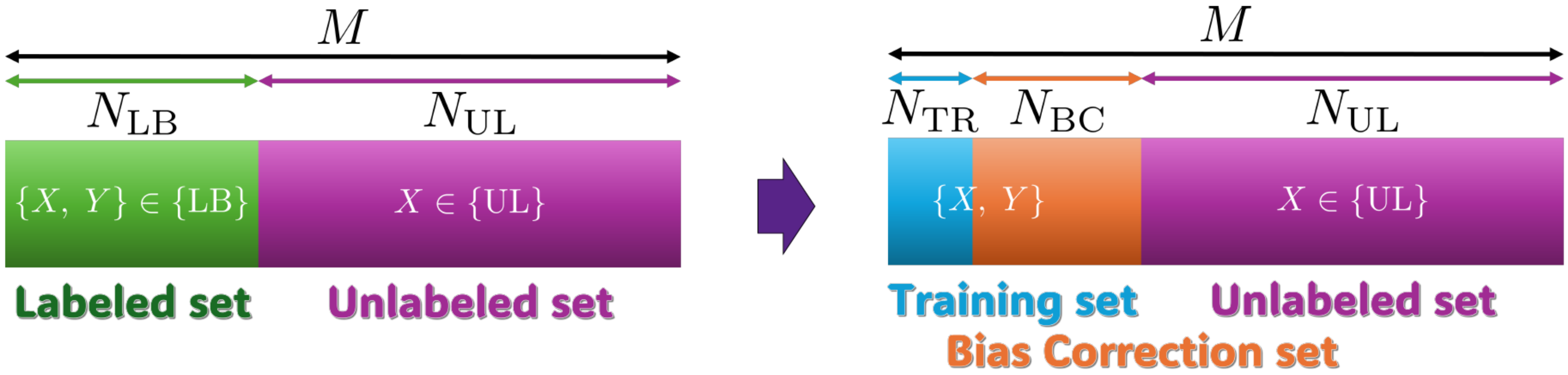
$$f_1 \to f_2 \to \cdots \to f_m$$

  - Depth, Ensemble size (boosting stage) : hyperparameters
- Correcting the prediction bias

$$\langle Y \rangle_{\mathrm{BC}} = \langle Y^P \rangle + \langle Y - Y^P \rangle$$

Data handling

- Splitting data into two sets
  - Labeled set (N_{LB}): input and target data **for training (N_{TR}) and bias correction (N_{BC})**
  - Unlabeled set (N_{UL}): input data **for estimation**
  - $M = N_{\mathrm{LB}} + N_{\mathrm{UL}}$
  - $N_{\mathrm{LB}} = N_{\mathrm{TR}} + N_{\mathrm{BC}}$



Prediction

- $\mathcal{P}1$: bias corrected ML prediction

$$\bar{Y}_{\mathcal{P}1} = \frac{1}{N_{\mathrm{UL}}} \sum_{i \in \{\mathrm{UL}\}} Y_i^P + \frac{1}{N_{\mathrm{BC}}} \sum_{j \in \{\mathrm{BC}\}} \left(Y_j - Y_j^P \right)$$

- $\mathcal{P}2$: weighted average of $\mathcal{P}1$ and target data of the labeled set
  - to improve the statistical precision

$$\bar{Y}_{\mathcal{P}2} = \frac{N_{\mathrm{UL}}}{M} \bar{Y}_{\mathcal{P}1} + \frac{N_{\mathrm{LB}}}{M} \bar{Y}_{\mathrm{LB}}, \qquad \bar{Y}_{\mathrm{LB}} = \frac{1}{N_{\mathrm{LB}}} \sum_{i \in \{\mathrm{LB}\}} Y_i$$

- Statistical errors are given by the bootstrap method.

Numerical setup

- Using data from HO *et al*. PoS **LATTICE2018** (2018) 174.
  - Iwasaki gauge and $N_f = 4$ Wilson clover fermion actions
  - Plaquette, Polyakov loop, $\mathrm{Tr}M^{-n}$ $(n = 1,2,3,4)$

ID	$L^3 \times T$	β	κ	c_{SW}	N_{conf}
0	$16^3 \times 4$	1.60	0.13575	2.065	5500
1	$16^3 \times 4$	1.60	0.13577	2.065	5500
2	$16^3 \times 4$	1.60	0.13580	2.065	5500
3	$16^3 \times 4$	1.60	0.13582	2.065	5500
4	$16^3 \times 4$	1.60	0.13585	2.065	5500

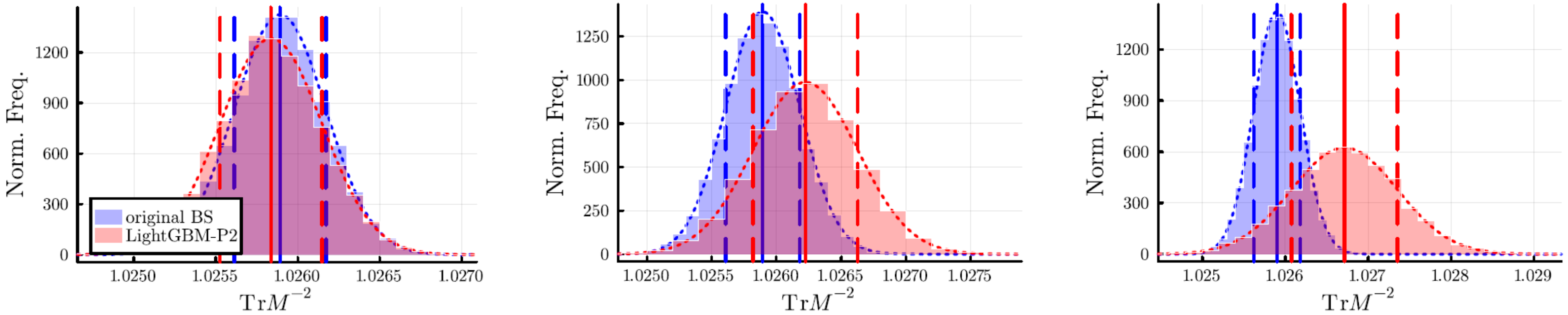
← 1st order phase transition

- Gradient boosting Decision Tree
  - using LightGBM framework (https://lightgbm.readthedocs.io/en/stable/)
  - via JuliaAI/MLJ.jl package (https://github.com/JuliaAI/MLJ.jl)
  - boosting stage = 40: empirically determined by monitoring the L2 loss.
  - depth = 3, learning rate = 0.1, subsampling = 0.7: same as Yoon et al., PRD **100** 014504 (2019)
- All calculations were done by Benjamin J. Choi (CCS, U. of Tsukuba)

Evaluations of results

- Evaluation 1

  Original data
  ML prediction



(a) **Score 2** (b) **Score 1** (c) **Score 0**

Score	Evaluation criteria ($\mathcal{X} = \mathcal{P}1$ or $\mathcal{P}2$)
2	**Both of** $\bar{Y}_{\mathrm{Orig.}}$ and $\bar{Y}_{\mathcal{X}}$ agree with $1\,\sigma$ level.
1	**Only one** of $\bar{Y}_{\mathrm{Orig.}}$ or $\bar{Y}_{\mathcal{X}}$ comes into the other's $1\,\sigma$ error.
0	$\bar{Y}_{\mathrm{Orig.}}$ and $\bar{Y}_{\mathcal{X}}$ **do not agree with** $1\,\sigma$ level each other.

- Evaluation 2
  - If a ML result get Score 2 and give $\sigma_{\chi}/\sigma_{\mathrm{Orig.}} \approx 1$

  → The ML result imitates the original data as well as possible

Searching for optimal ratio of labeled/training set

- To reduce of the computational cost

  → need to know sufficiently minimal ratio of the labeled set to the entire data

$$R_{\mathrm{LB}} = \frac{N_{\mathrm{LB}}}{M} = 5, 10, \ldots, 50\%$$

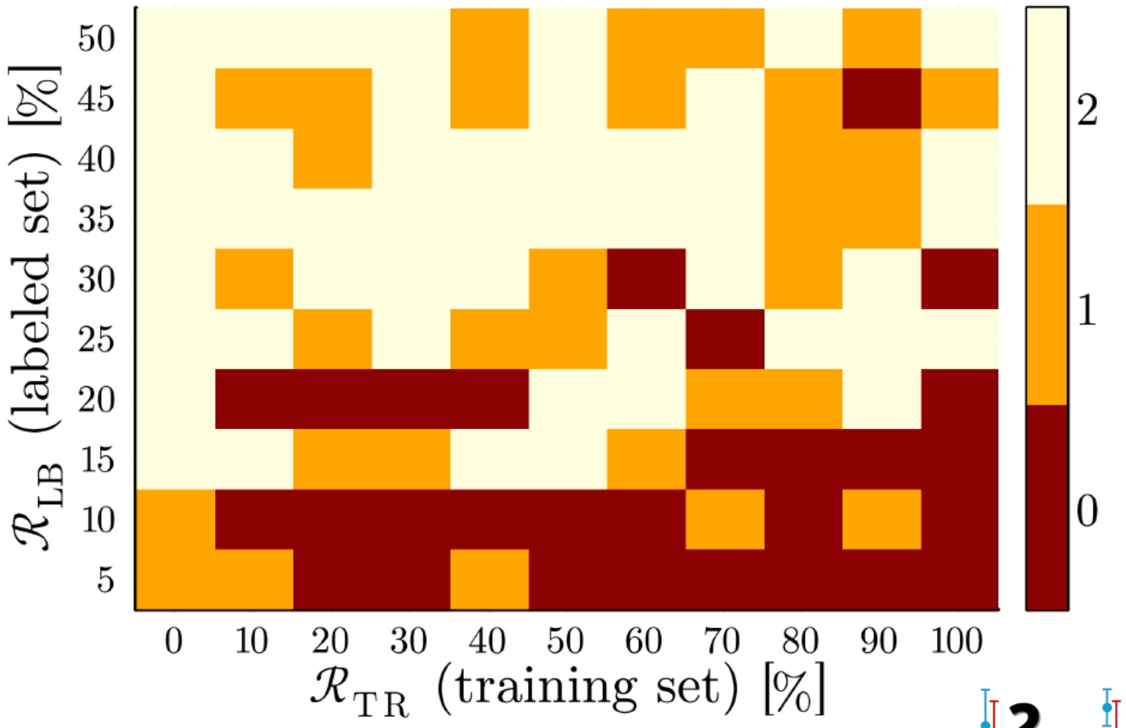
- For sufficient statistical precision of the bias correction

  → need to know sufficiently maximal ratio of the training set to the labeled set

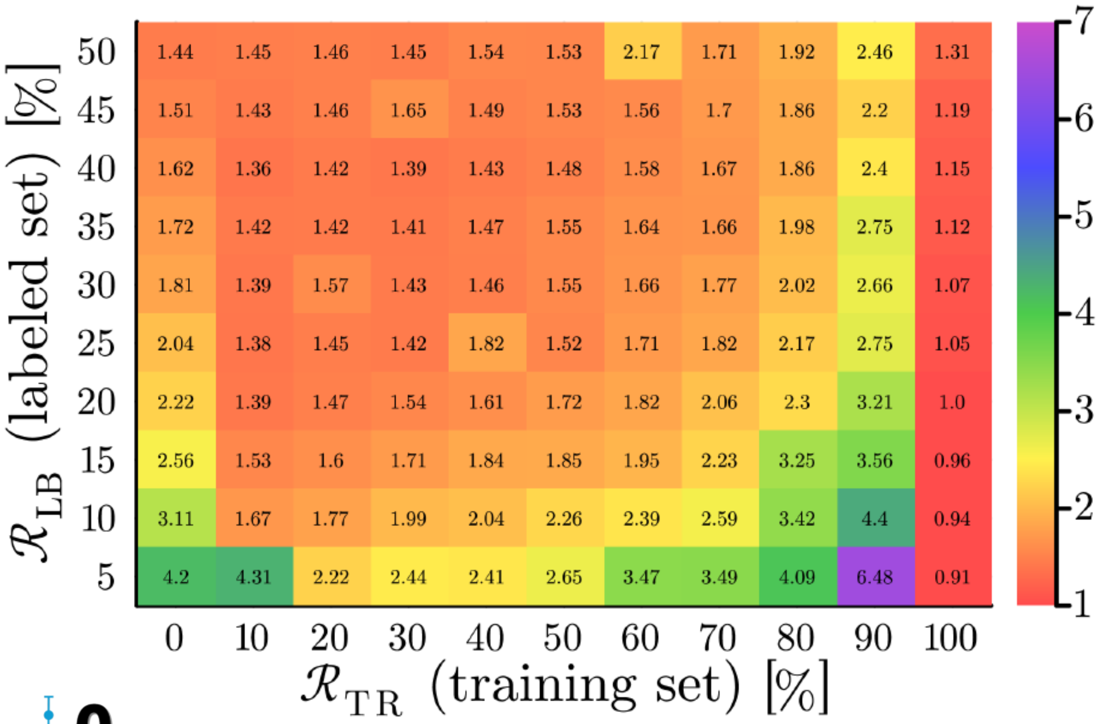
$$R_{\mathrm{TR}} = \frac{N_{\mathrm{TR}}}{N_{\mathrm{LB}}} = 10, 20, \ldots, 90\%$$

- Also
  - $R_{\mathrm{TR}} = 0\%$ to check the statistical precision of the labeled set
  - $R_{\mathrm{TR}} = 100\%$ to check the result without the bias correction

Example result: plaquette → $\mathrm{Tr}M^{-3}$ ($\mathcal{P}1$, ID-0)



(a) $\bar{Y}$ (central value) check

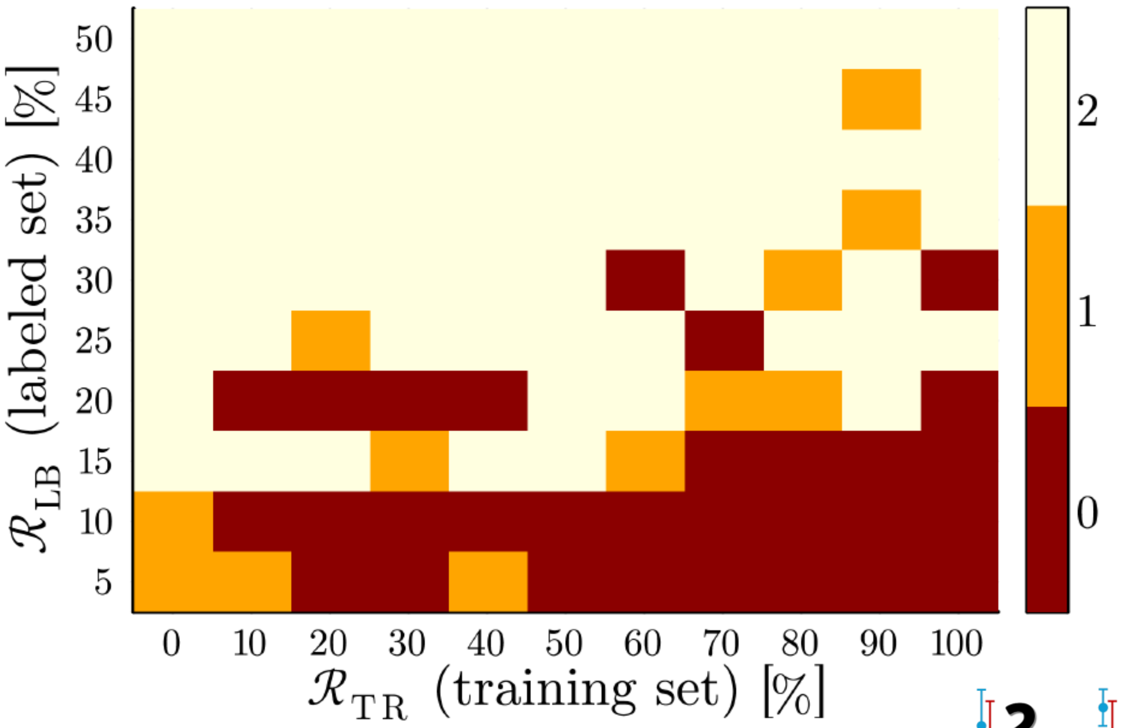


(b) Magnitude of $\sigma_{\mathcal{P}1}/\sigma_{\mathrm{Orig.}}$

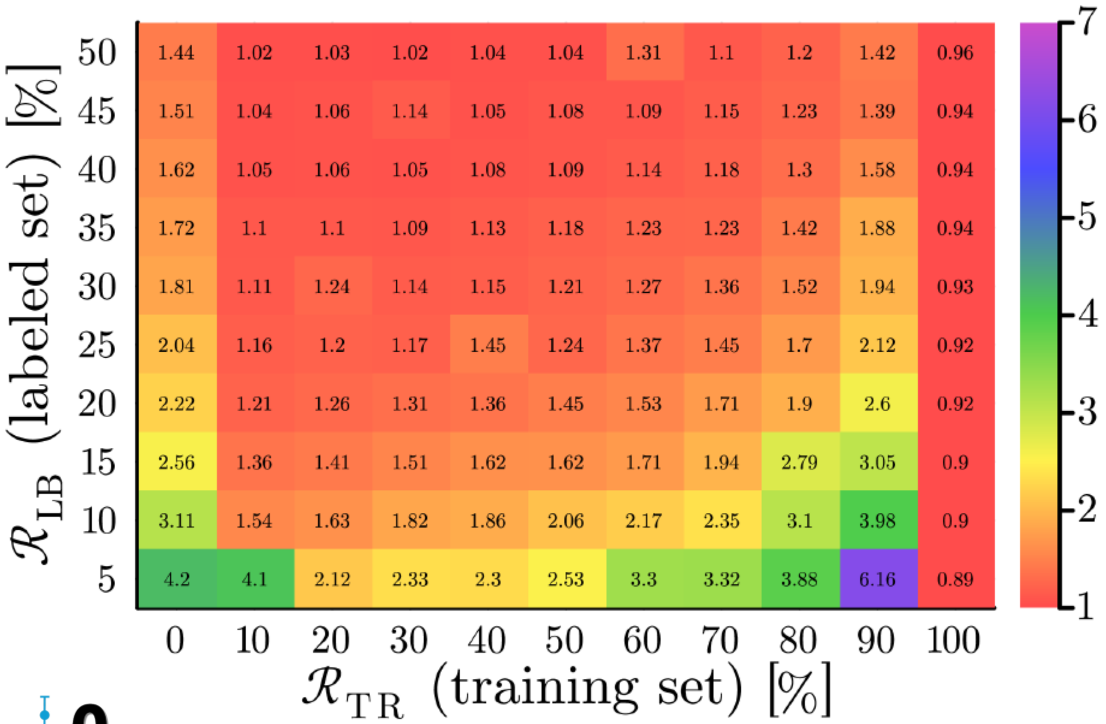
Evaluation 1.
cannot find consistently score-2 region

Evaluation 2.
Roughly $\sigma_{\mathcal{P}1}/\sigma_{\mathrm{Orig.}} \approx 1.5$
in $R_{\mathrm{LB}} \geq 30\%, R_{\mathrm{TR}} \leq 50\%$

Example result: plaquette → $\mathrm{Tr}M^{-3}$ ($\mathcal{P}2$, ID-0)



(a) $\bar{Y}$ (central value) check



(b) Magnitude of $\sigma_{\mathcal{P}2}/\sigma_{\mathrm{Orig.}}$

Evaluation 1.
consistently score-2 region
in $R_{\mathrm{LB}} \geq 30\%, R_{\mathrm{TR}} \leq 50\%$

Evaluation 2.
Roughly $\sigma_{\mathcal{P}1}/\sigma_{\mathrm{Orig.}} \approx 1.1$
in $R_{\mathrm{LB}} \geq 30\%, R_{\mathrm{TR}} \leq 50\%$

Preliminary results: ID-0 (heaviest quark mass)

- $\mathcal{P}2$ results
- {sufficiently minimal R_{LB}, sufficiently maximal R_{TR}}

ID-0	Plaquette	Polyakov loop	$\mathrm{Tr}M^{-1}$	$\mathrm{Tr}M^{-2}$	$\mathrm{Tr}M^{-3}$
$\mathrm{Tr}M^{-1}$	{**30**, 50}	{**35**, 40}			
$\mathrm{Tr}M^{-2}$	{**30**, 50}	{**35**, 40}	{**25**, 40}		
$\mathrm{Tr}M^{-3}$	{**30**, 50}	{**35**, 40}	{**40**, 40}	{**15**, 70}	
$\mathrm{Tr}M^{-4}$	{**45**, 60}	{**45**, 60}	{**40**, 40}	{**35**, 50}	{**25**, 40}

- $\mathrm{Tr}M^{-n}$ $(n = 1,2,3)$ can be accurately estimated from plaquette and Polyakov loop with $R_{\mathrm{LB}} \geq 30\%$ and $R_{\mathrm{LB}} \geq 35\%$, respectively.
- Not good at estimating $\mathrm{Tr}M^{-4}$: $R_{\mathrm{LB}} \geq 45\%$.

Preliminary results: ID-2 (close to the 1st order PT)

- $\mathcal{P}2$ results
- {**sufficiently minimal R_{LB}**, **sufficiently maximal R_{TR}**}

ID-2	Plaquette	Polyakov loop	$\mathrm{Tr}M^{-1}$	$\mathrm{Tr}M^{-2}$	$\mathrm{Tr}M^{-3}$
$\mathrm{Tr}M^{-1}$	{**10**, 80}	{**10**, 80}			
$\mathrm{Tr}M^{-2}$	{**10**, 90}	{**10**, 80}	{**10**, 90}		
$\mathrm{Tr}M^{-3}$	{**10**, 90}	{**10**, 70}	{**10**, 80}	{**10**, 90}	
$\mathrm{Tr}M^{-4}$	{**40**, 40}	{**40**, 80}	{**40**, 80}	{**40**, 80}	{**40**, 80}

- $\mathrm{Tr}M^{-n}$ $(n = 1,2,3)$ can be accurately estimated from plaquette and Polyakov loop with $\boldsymbol{R_{\mathrm{LB}} \geq 10\%}$.
- Not good at estimating $\mathrm{Tr}M^{-4}$: $\boldsymbol{R_{\mathrm{LB}} \geq 40\%}$.

Preliminary results: ID-4 (lightest quark mass)

- $\mathcal{P}2$ results
- {**sufficiently minimal R_{LB}**, **sufficiently maximal R_{TR}**}

ID-4	Plaquette	Polyakov loop	$\mathrm{Tr}M^{-1}$	$\mathrm{Tr}M^{-2}$	$\mathrm{Tr}M^{-3}$
$\mathrm{Tr}M^{-1}$	{**30**, 40}	{**35**, 40}			
$\mathrm{Tr}M^{-2}$	{**30**, 50}	{**40**, 50}	{**30**, 40}		
$\mathrm{Tr}M^{-3}$	N.A.	N.A.	N.A.	N.A.	
$\mathrm{Tr}M^{-4}$	N.A.	N.A.	N.A.	N.A.	N.A.

- N.A.: cannot find proper $\{R_{\mathrm{LB}}, R_{\mathrm{TR}}\}$.
- $\mathrm{Tr}M^{-n}$ $(n = 1,2)$ can be accurately estimated from plaquette and Polyakov loop with $\boldsymbol{R_{\mathrm{LB}} \geq 30\%}$ and $\boldsymbol{R_{\mathrm{LB}} \geq 40\%}$, respectively.

Summary and outlook

- The ability of machine learning estimation of $\mathrm{Tr}M^{-n}$ from plaquette, Polyakov loop and $\mathrm{Tr}M^{-m}$ $(m < n)$ was checked by using LightGBM via JuliaAI/MLJ.jl.
- The ML estimation works better for heavier quark masses and especially around the 1st order phase transition point.
- We also found that the bias correction works well.

- Extending the analysis to other ensembles with different parameters:
  - larger volume, finer lattice, …
- Calculating cumulants of the chiral order parameter from the estimated traces and checking the accuracy.

Backup slides

Preliminary results: ID-1 (second heaviest quark mass)

- $\mathcal{P}2$ results
- {**sufficiently minimal R_{LB}**, **sufficiently maximal R_{TR}**}

ID-1	Plaquette	Polyakov loop	$\mathrm{Tr}M^{-1}$	$\mathrm{Tr}M^{-2}$	$\mathrm{Tr}M^{-3}$
$\mathrm{Tr}M^{-1}$	{**30**, 40}	{**35**, 40}			
$\mathrm{Tr}M^{-2}$	{**20**, 50}	{**35**, 40}	{**15**, 40}		
$\mathrm{Tr}M^{-3}$	{**25**, 50}	{**35**, 40}	{**30**, 60}	{**20**, 80}	
$\mathrm{Tr}M^{-4}$	{**45**, 40}	{**50**, 40}	{**45**, 50}	{**45**, 50}	{**25**, 50}

- $\mathrm{Tr}M^{-n}$ $(n = 1,2,3)$ can be accurately estimated from plaquette and Polyakov loop with $\boldsymbol{R_{\mathrm{LB}} \geq 30\%}$ and $\boldsymbol{R_{\mathrm{LB}} \geq 35\%}$, respectively.
- Not good at estimating $\mathrm{Tr}M^{-4}$: $\boldsymbol{R_{\mathrm{LB}} \geq 45\%}$.

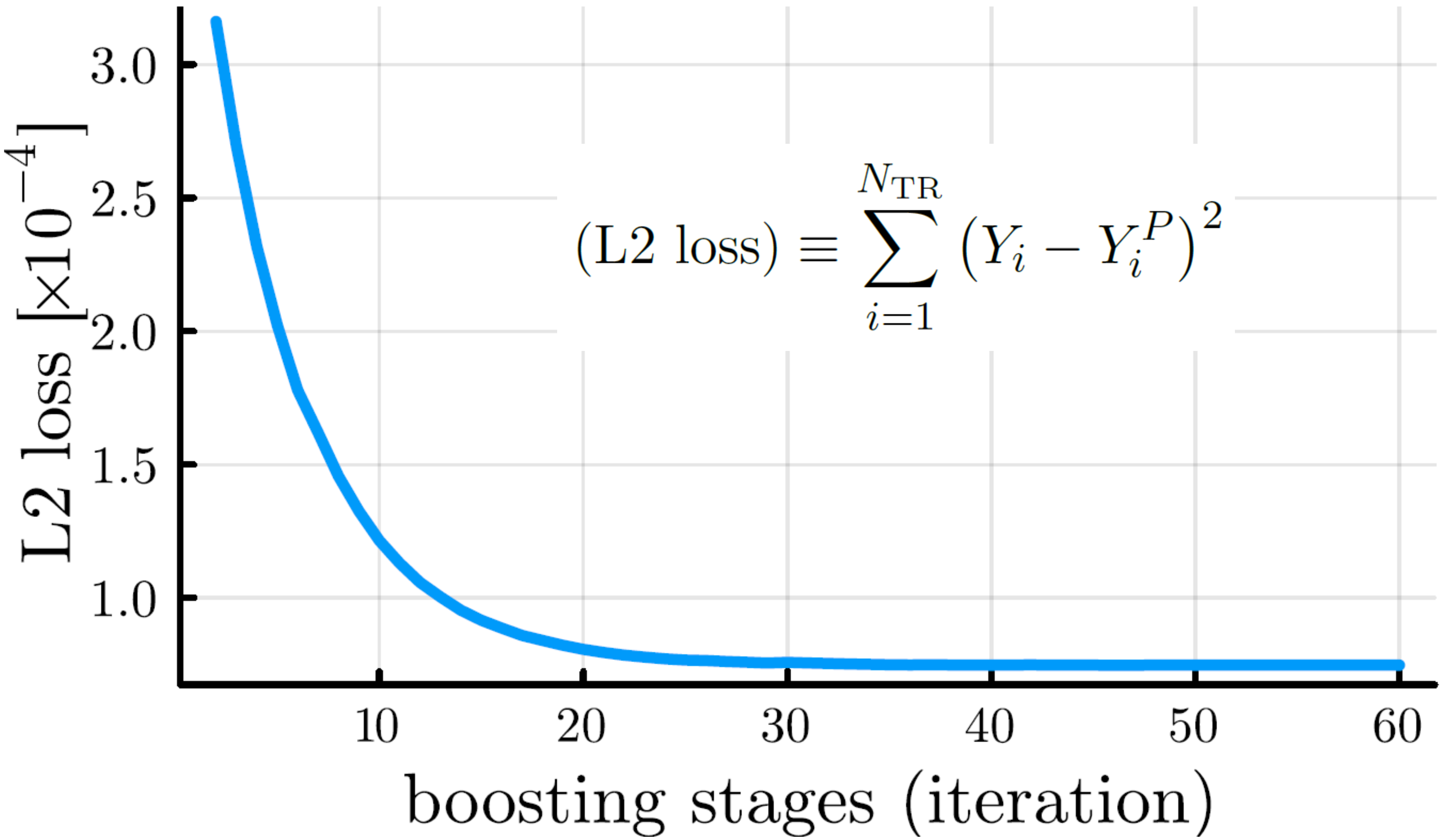
Preliminary results: ID-3 (second lightest quark mass)

- $\mathcal{P}2$ results
- {sufficiently minimal R_{LB}, sufficiently maximal R_{TR}}

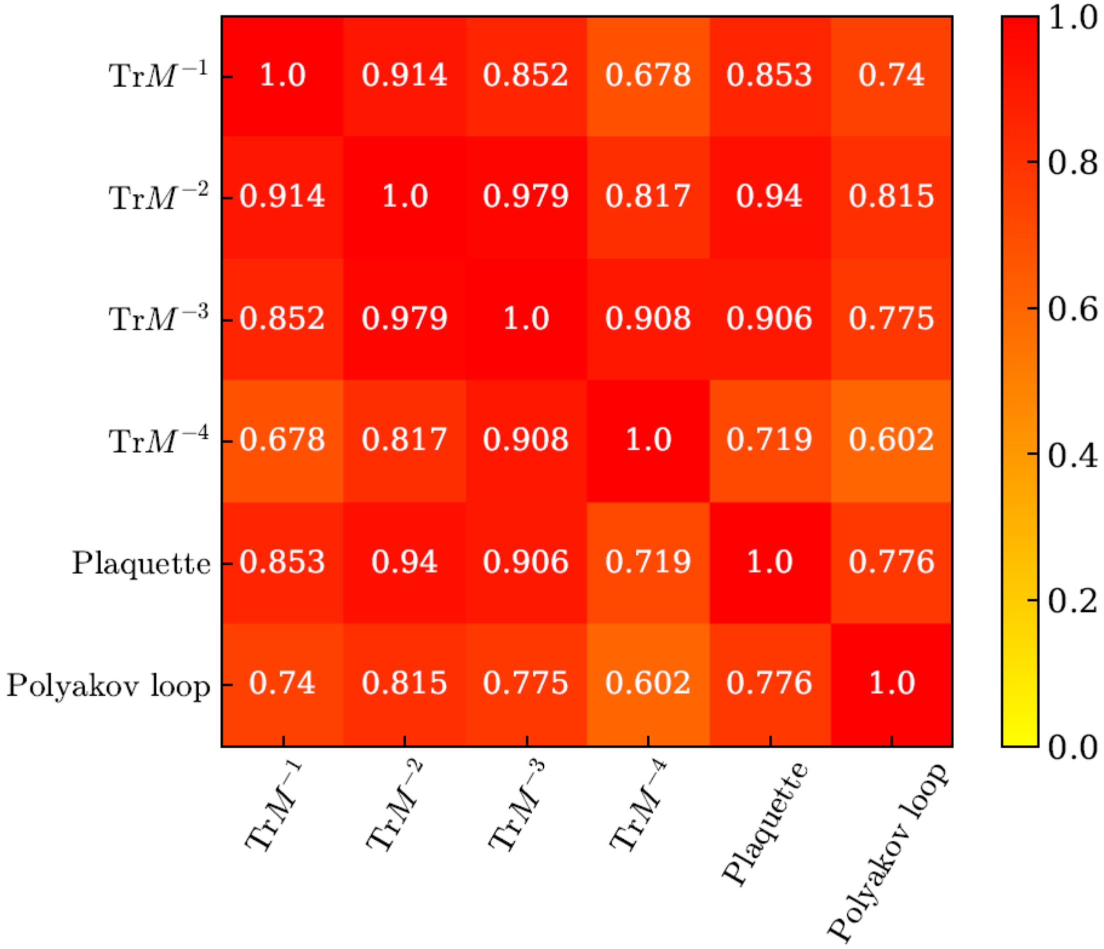
ID-3	Plaquette	Polyakov loop	$\mathrm{Tr}M^{-1}$	$\mathrm{Tr}M^{-2}$	$\mathrm{Tr}M^{-3}$
$\mathrm{Tr}M^{-1}$	{10, 60}	{20, 80}			
$\mathrm{Tr}M^{-2}$	{10, 90}	{10, 60}	{10, 80}		
$\mathrm{Tr}M^{-3}$	{10, 70}	{10, 50}	{15, 60}	{10, 90}	
$\mathrm{Tr}M^{-4}$	{40, 50}	{40, 50}	{40, 50}	{40, 50}	{40, 50}

- $\mathrm{Tr}M^{-n}$ $(n = 1,2,3)$ can be accurately estimated from plaquette and Polyakov loop with $R_{\mathrm{LB}} \geq 10\%$.
- Not good at estimating $\mathrm{Tr}M^{-4}$: $R_{\mathrm{LB}} \geq 40\%$.

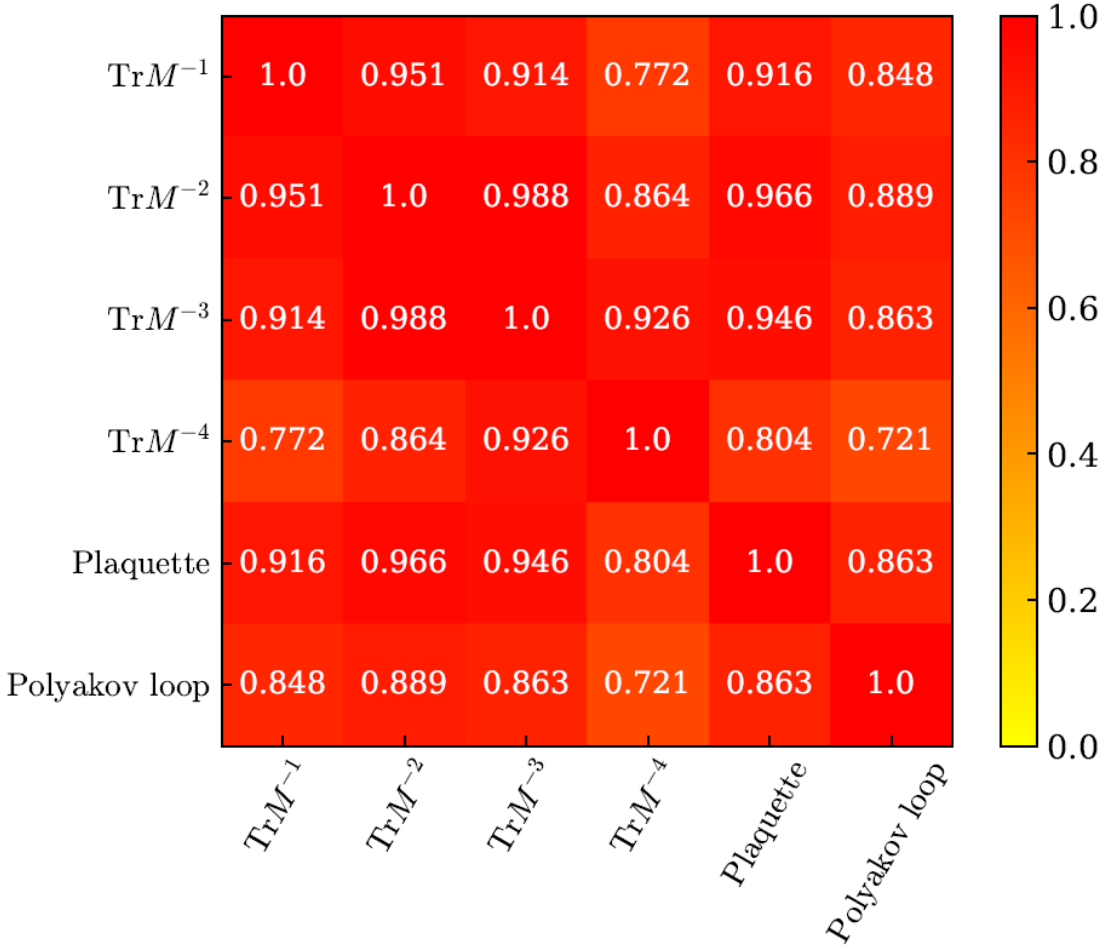
Optimal boosting stage



Correlation map: ID-0 and ID-1

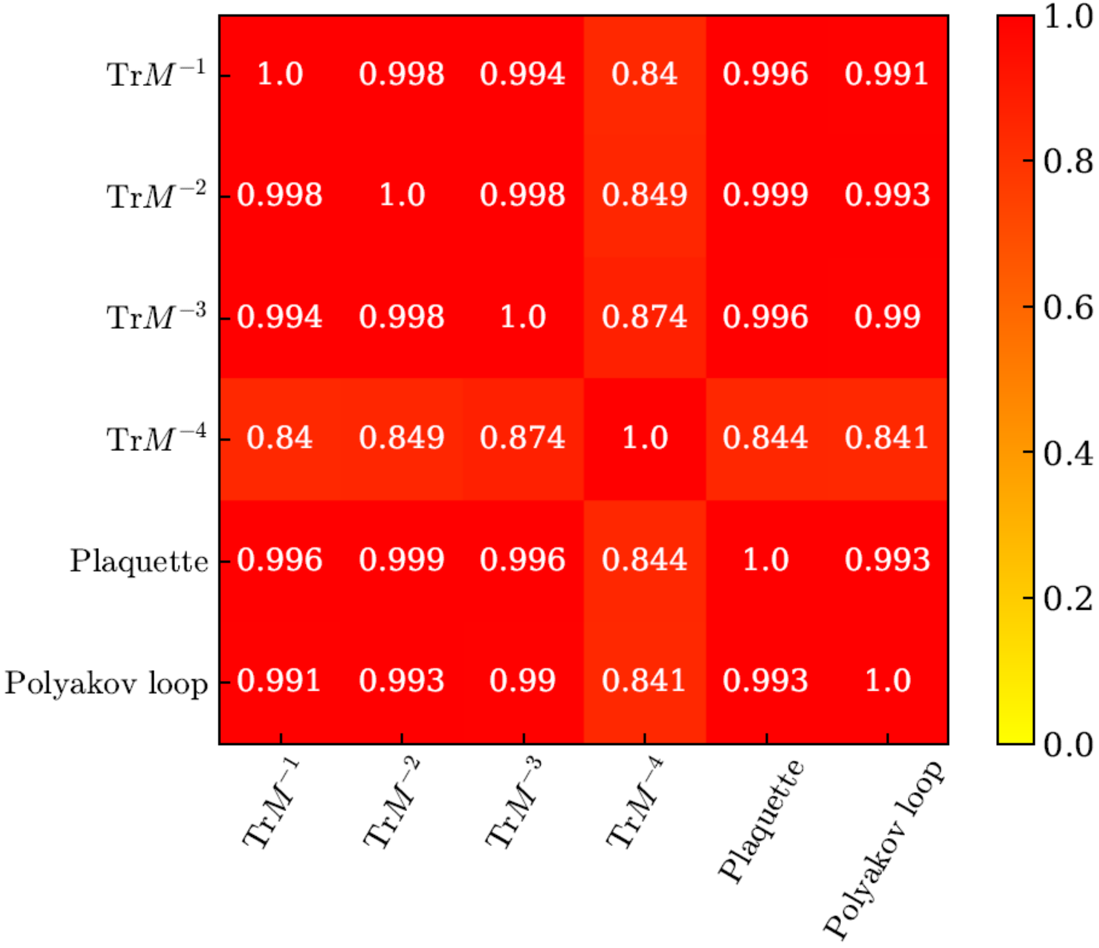


(a) $\kappa = 0.13575$, `ID 0`

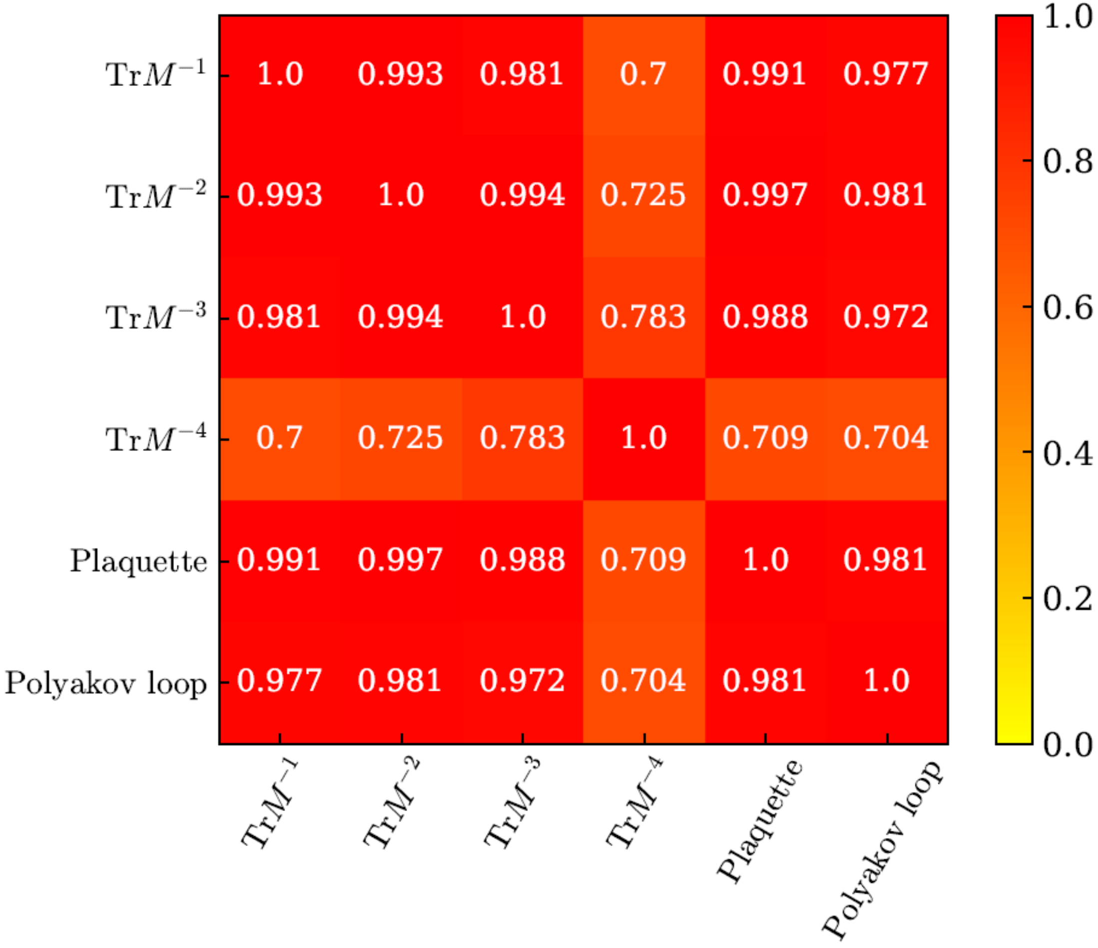


(b) $\kappa = 0.13577$, `ID 1`

Correlation map: ID-2 and ID-3

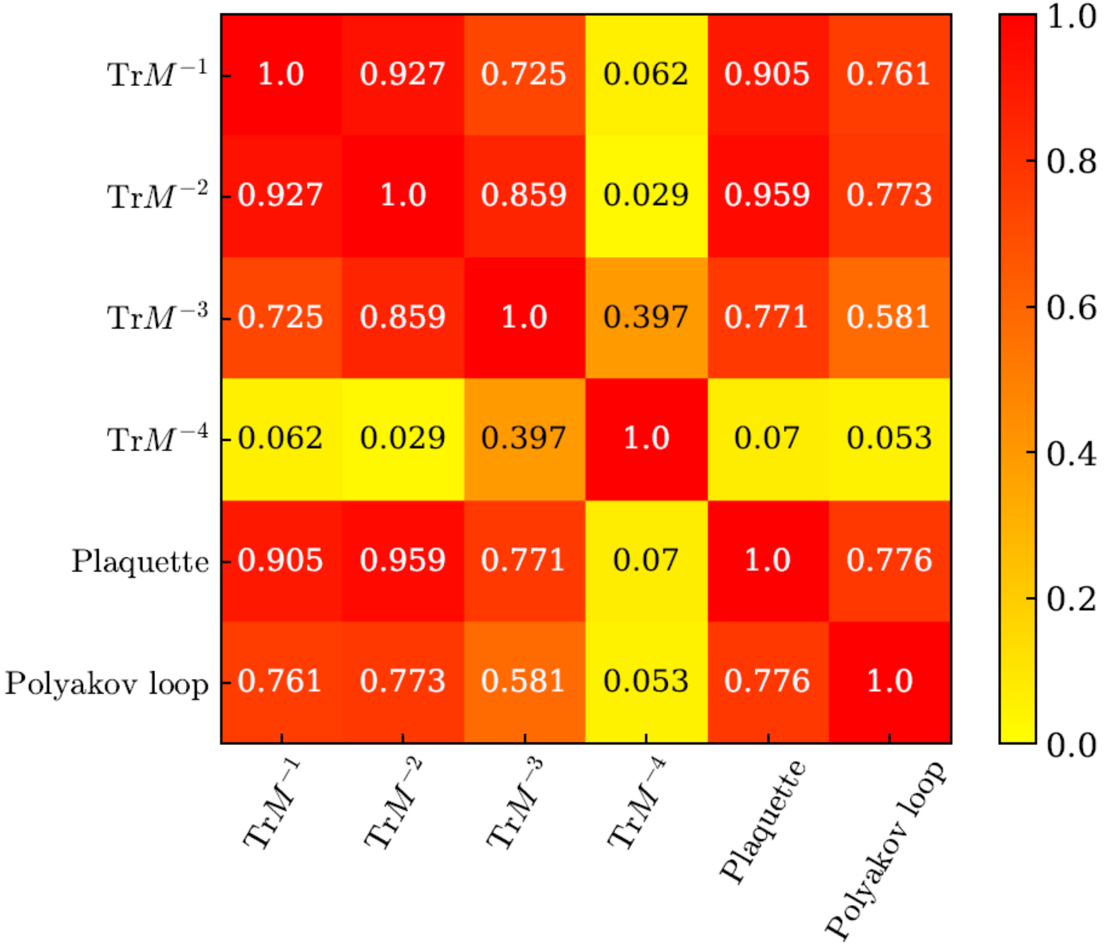


(c) $\kappa = 0.13580$, ID 2



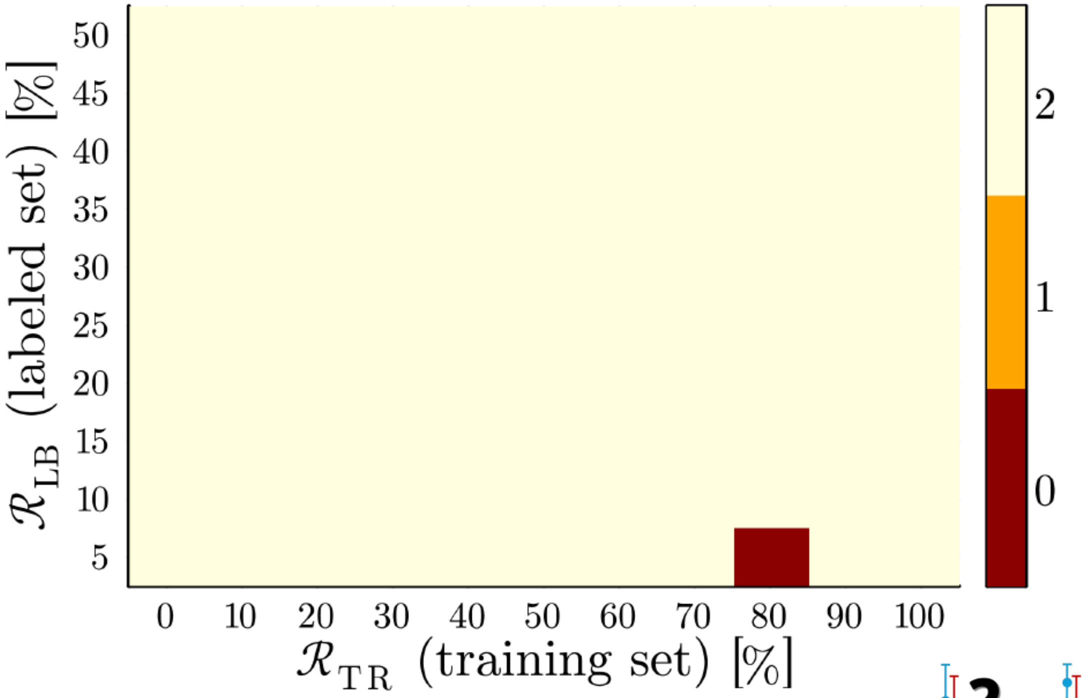
(d) $\kappa = 0.13582$, ID 3

Correlation map: ID-4

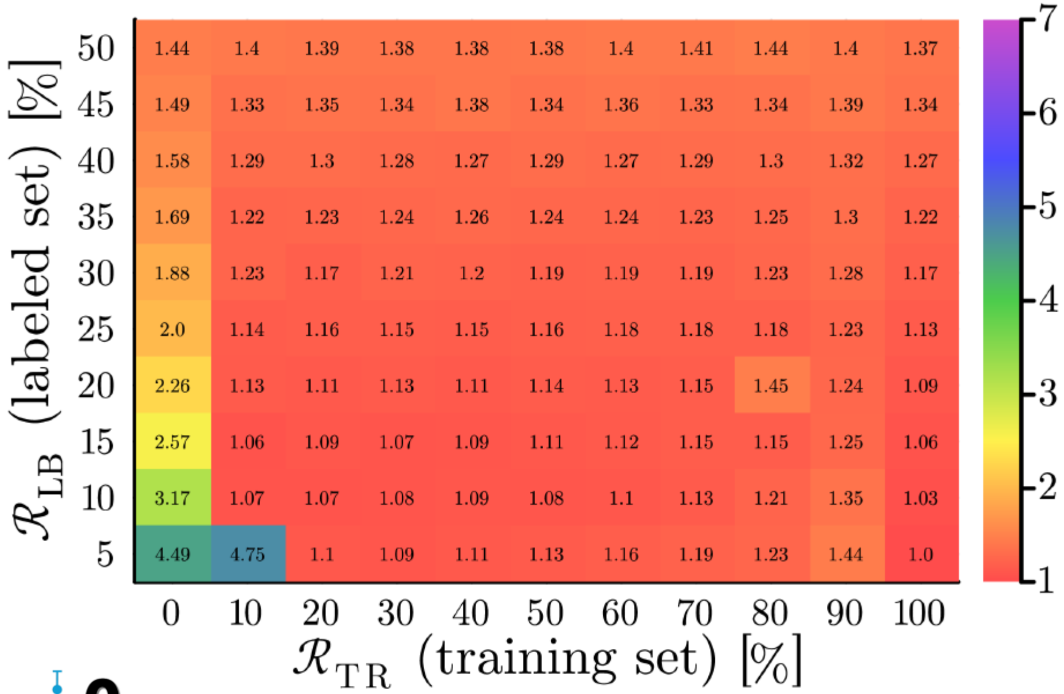


(e) $\kappa = 0.13585$, `ID 4`

Example result: plaquette → $\mathrm{Tr}M^{-3}$ ($\mathcal{P}1$, ID-2)



(a) $\bar{Y}$ (central value) check



(b) Magnitude of $\sigma_{\mathcal{P}1}/\sigma_{\mathrm{Orig.}}$

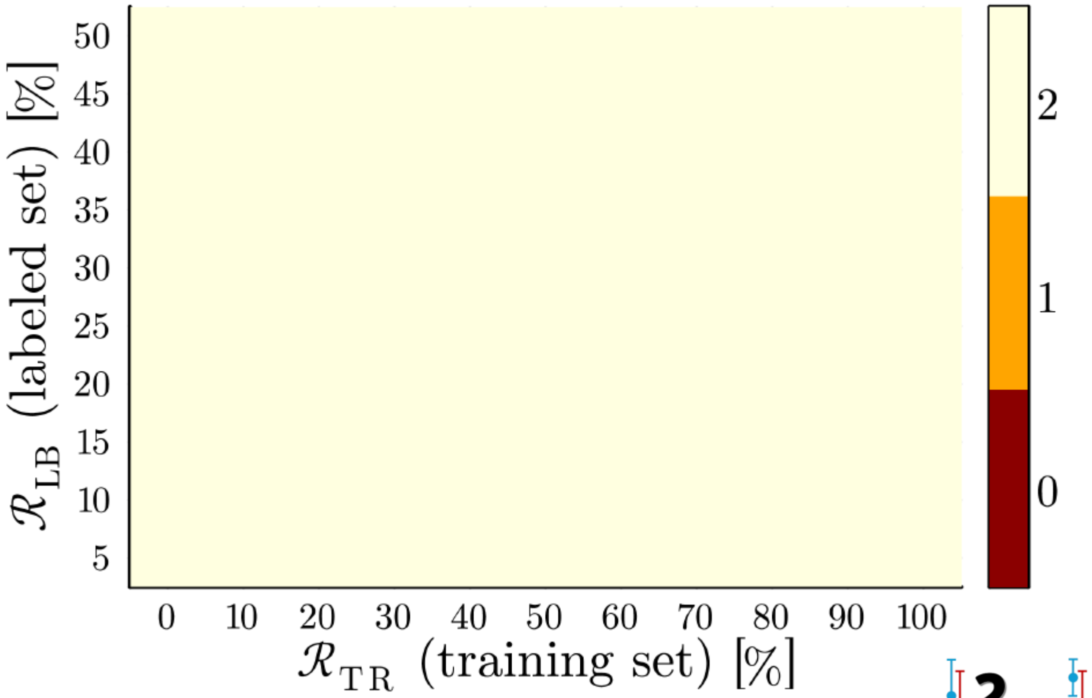
Evaluation 1.

consistently score-2 region
in $R_{\mathrm{LB}} \geq 10\%, R_{\mathrm{TR}} \leq 90\%$

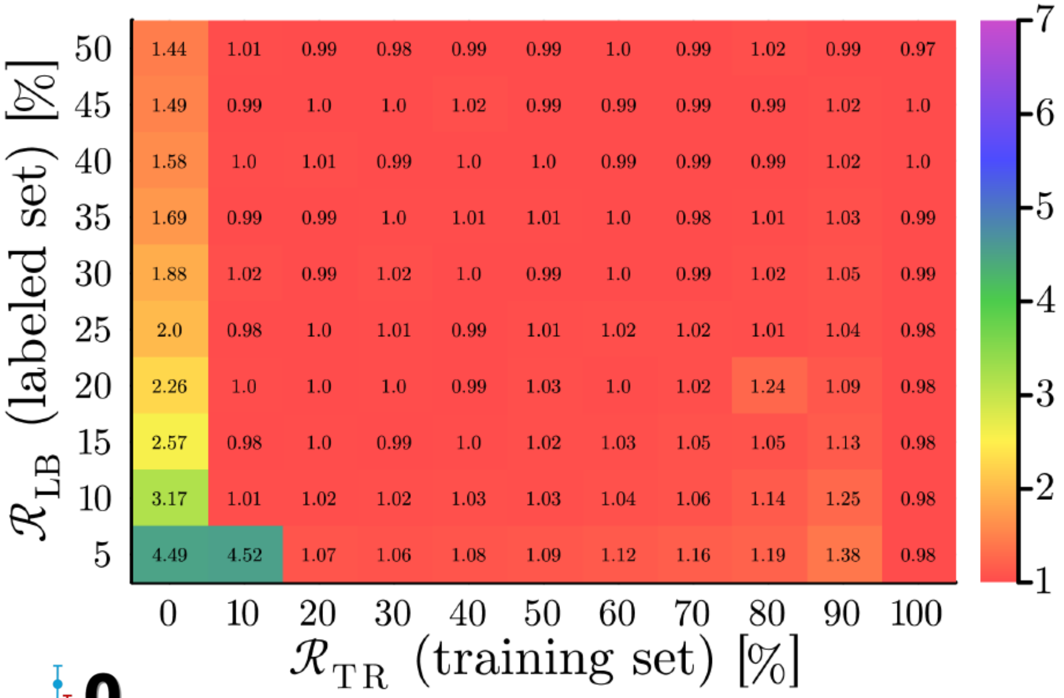
Evaluation 2.

Roughly $\sigma_{\mathcal{P}1}/\sigma_{\mathrm{Orig.}} \approx 1.3$
in $R_{\mathrm{LB}} \geq 10\%, R_{\mathrm{TR}} \leq 70\%$

Example result: plaquette → $\mathrm{Tr}M^{-3}$ ($\mathcal{P}2$, ID-2)



(a) $\bar{Y}$ (central value) check

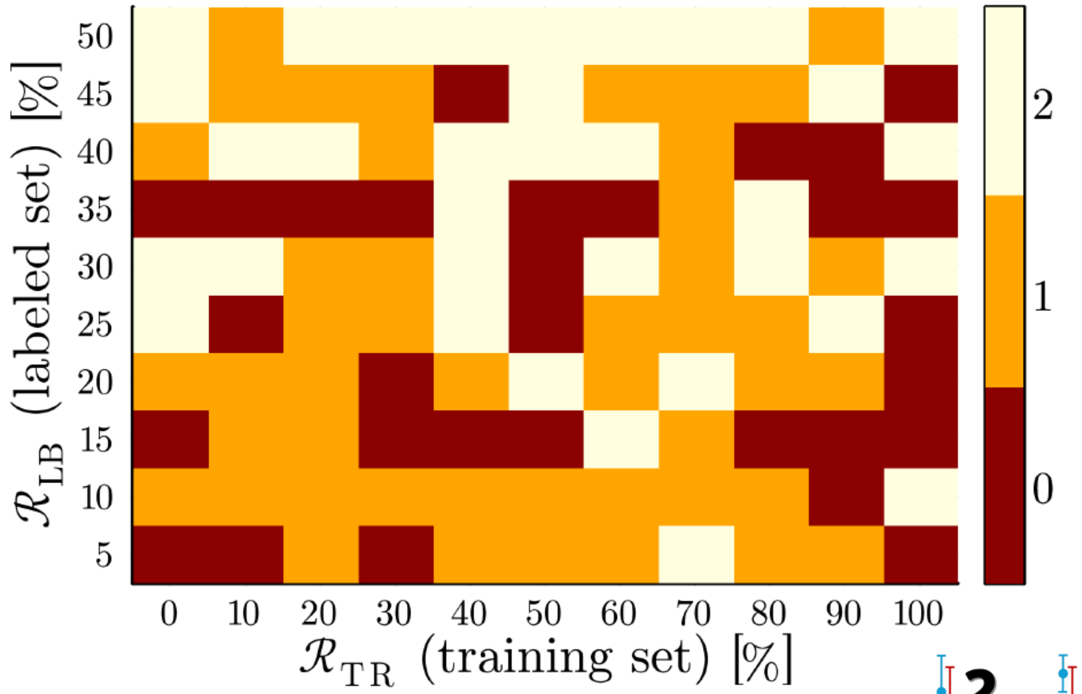


(b) Magnitude of $\sigma_{\mathcal{P}2}/\sigma_{\mathrm{Orig.}}$

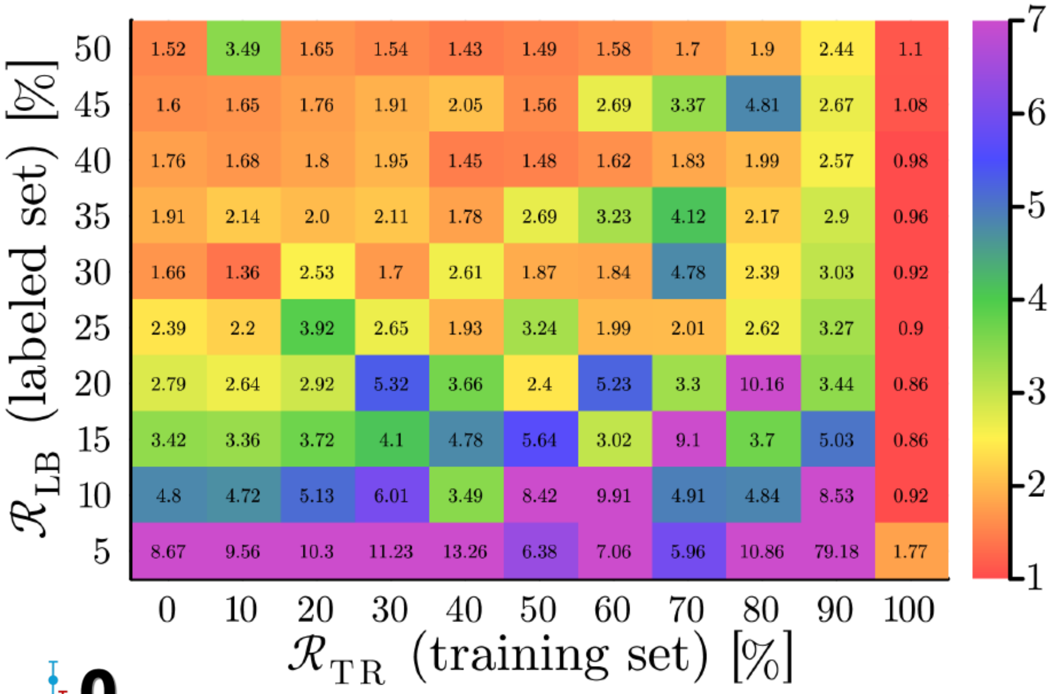
Evaluation 1.
consistently score-2 region
in $R_{\mathrm{LB}} \geq 5\%, R_{\mathrm{TR}} \leq 90\%$

Evaluation 2.
Roughly $\sigma_{\mathcal{P}1}/\sigma_{\mathrm{Orig.}} \approx 1$
in $R_{\mathrm{LB}} \geq 10\%, R_{\mathrm{TR}} \leq 70\%$

Example result: plaquette → $\mathrm{Tr}M^{-3}$ ($\mathcal{P}1$, ID-4)



(a) $\bar{Y}$ (central value) check

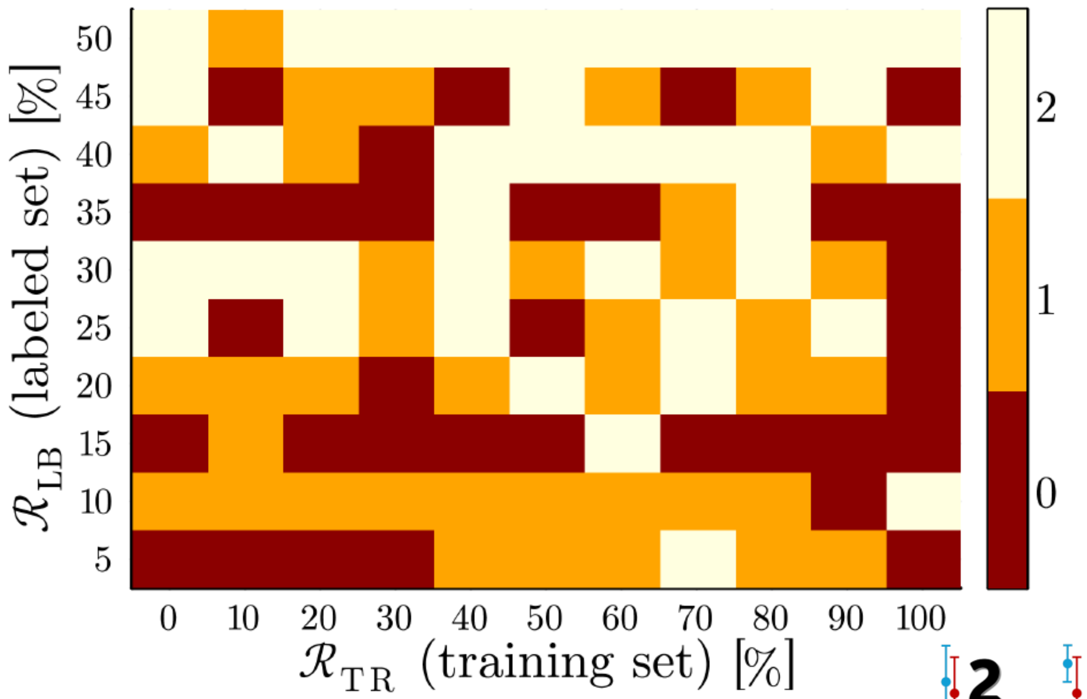


(b) Magnitude of $\sigma_{\mathcal{P}1}/\sigma_{\mathrm{Orig.}}$

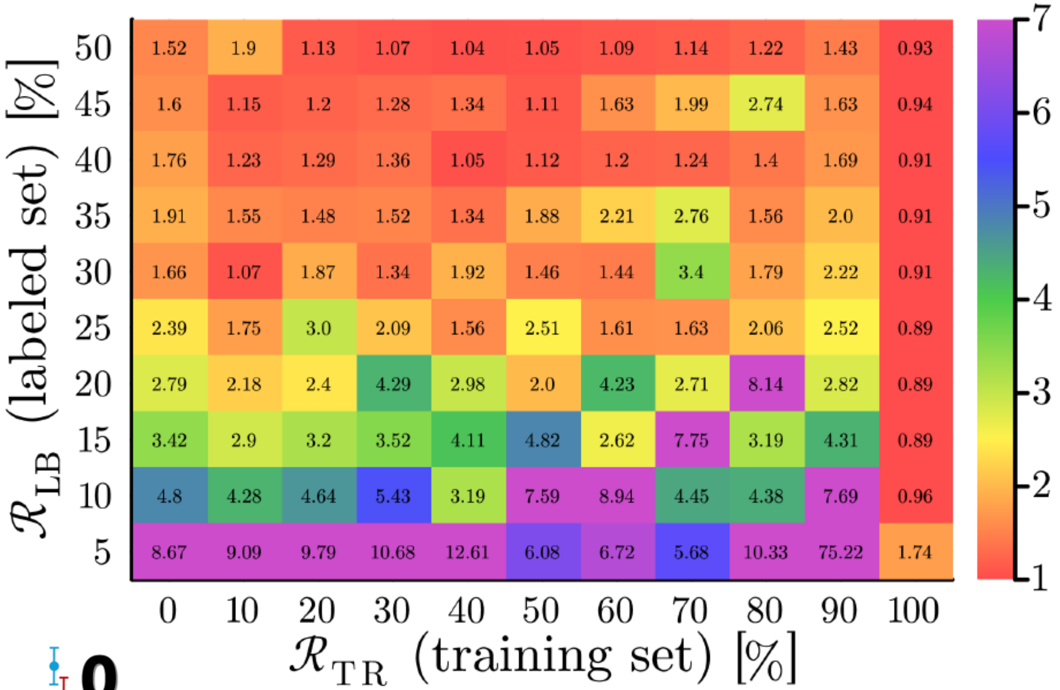
Evaluation 1.
cannot find consistently score-2 region

Evaluation 2.
Magnitude of $\sigma_{\mathcal{P}1}/\sigma_{\mathrm{Orig.}}$ is randomly distributed

Example result: plaquette → $\mathrm{Tr}M^{-3}$ ($\mathcal{P}2$, ID-4)



(a) $\bar{Y}$ (central value) check



(b) Magnitude of $\sigma_{\mathcal{P}2}/\sigma_{\mathrm{Orig.}}$

Evaluation 1.
cannot find consistently score-2 region

Evaluation 2.
Magnitude of $\sigma_{\mathcal{P}1}/\sigma_{\mathrm{Orig.}}$ is randomly distributed