UNIVERSITÄT BIELEFELD

Faculty of Physics

PUNCH 4NFDI

Testing machine learning against finite size scaling using MAFs

Simran Singh

University of Bielefeld

work done as part of **A01** *project in* **CRC Tr-211** *between* ***Bielefeld*** *(F. Karsch, C. Schmidt & S. Singh) and* ***Frankfurt*** *(O. Phillipsen, R. Kaiser, J.P. Klinger)*

ML meets LFT @ Swansea University

July 24-26, 2024

Outline

I. Motivation – Nature of the chiral transition

II. Nf = 5 project using MAFs and HISQ (old)

III. MADE and MAF

IV. (new) Nf = 5 using Stout smeared data

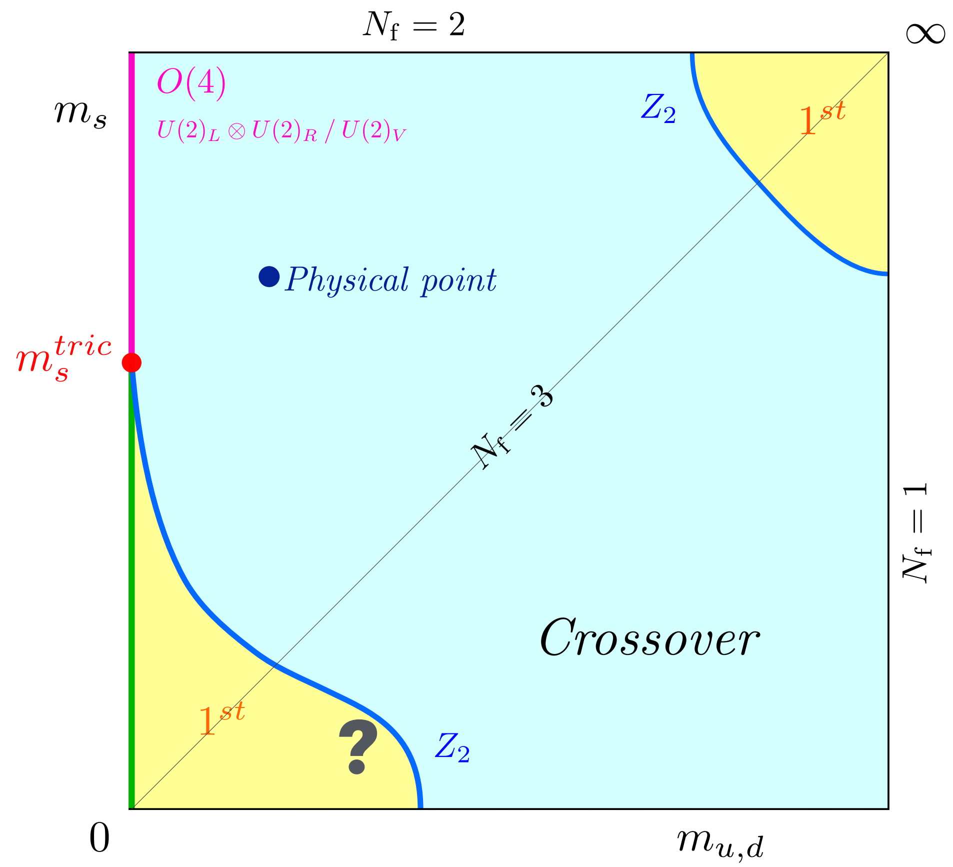
V. Results on density estimation

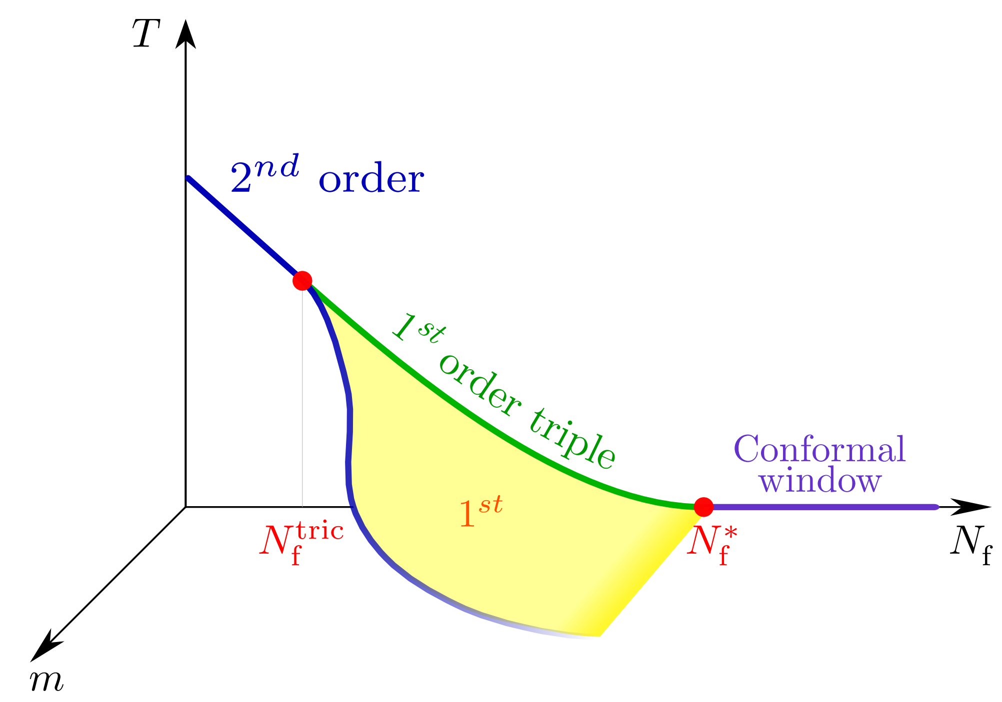
VI. Future directions

N_f and the chiral transition

- Nature of the QCD chiral phase transition is (strictly speaking) inconclusive
- Cannot simulate $m_f = 0$ on lattice - can only extrapolate from finite mass simulations
- Proposal for studying the critical surface that separates first-order regions from crossover as function of degenerate N_f quarks by [F. Cuteri et.al., *JHEP* 11 (2021)]
- Using the argument that the surface terminates in a tri-critical line, conclusions for the order of the chiral transition can be drawn.

[F. Cuteri et.al., *JHEP* 11 (2021)]





Z2 boundary for Nf=5 HISQ

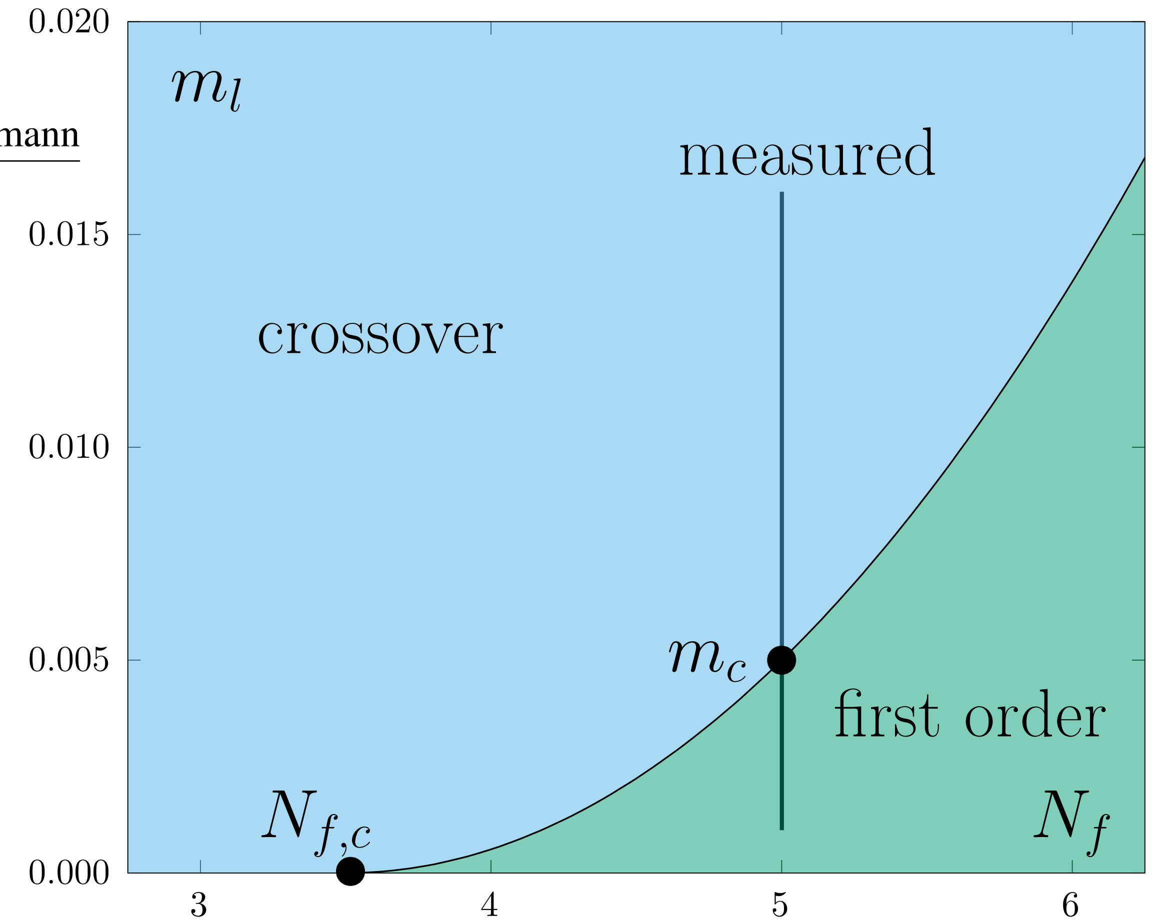
- The analysis of the above kind requires LOTS of lattice simulations - for EACH N_f, varying lattice volumes and spacings need to be studied
- In [M. Neumann et.al., *PoS* LATTICE2022 (2023)], the authors proposed using Machine Learning techniques for learning probability densities $p\left(\bar{\psi}\psi, S\right)$ conditioned on N_σ, m_l, β
- Learning such a density correctly allows interpolation in the dimensions of the conditional inputs - avoiding <u>some</u> expensive lattice simulations

A ML approach to many flavor QCD M. Neumann

N_σ	0.001	0.002	0.003	0.0035	0.004	0.0045	0.005
16	17201	18887	11526	0	18866	0	0
24	5294	83177	149885	25028	30571	19332	19352

N_σ	0.006	0.008	0.010	0.012	0.014	0.016
16	61382	61220	61456	61456	61256	61256
24	42762	82061	65140	13380	36574	36499

M. Neumann et.al., *PoS* LATTICE2022 (2023)



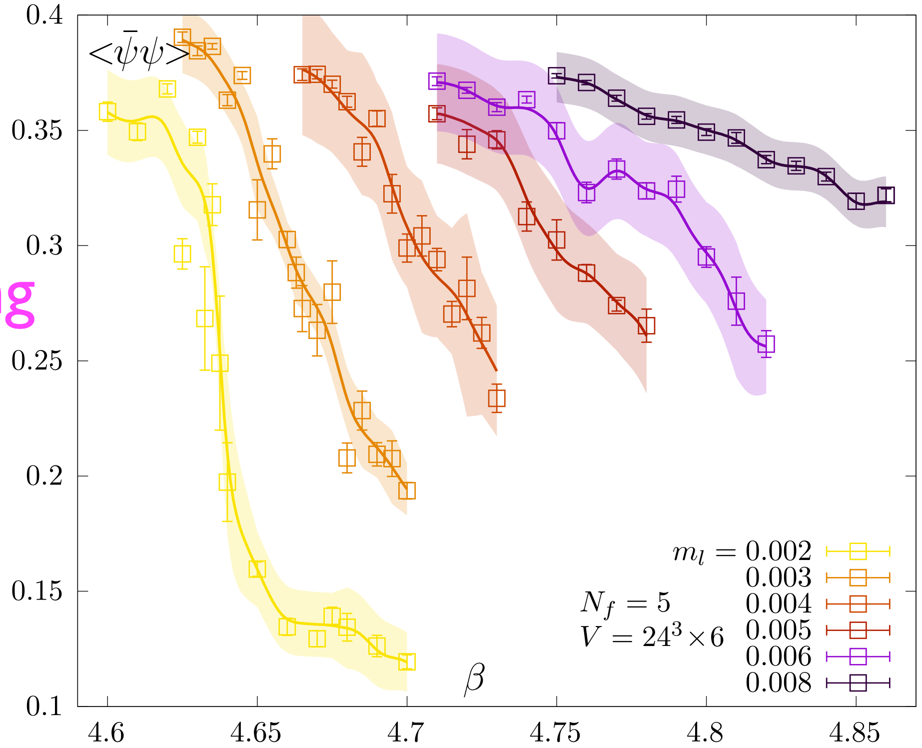
Z2 boundary for Nf=5 HISQ

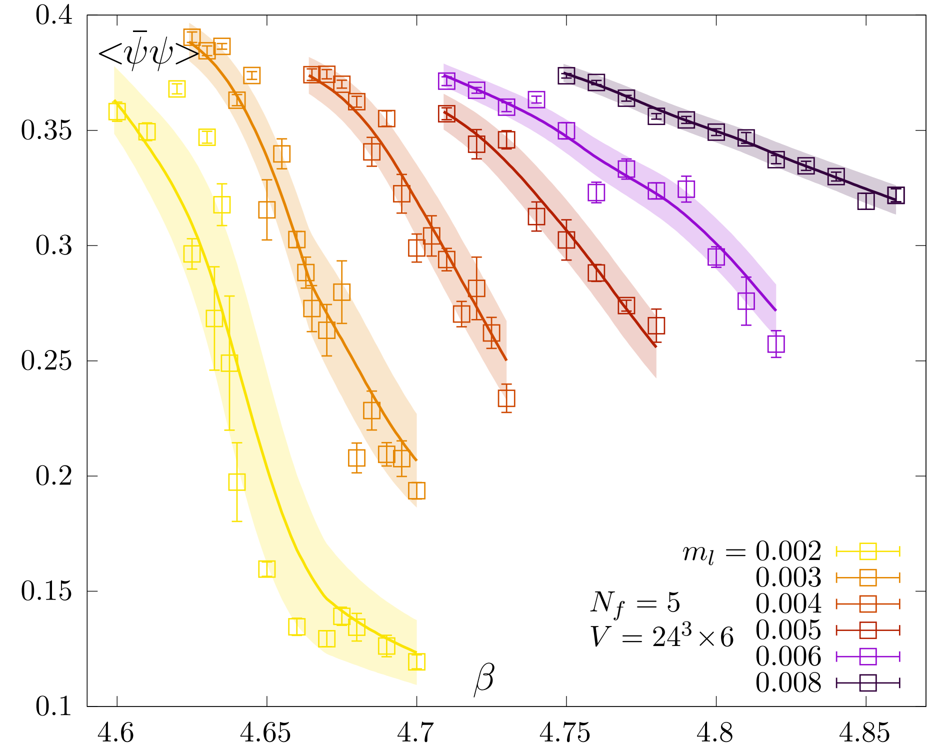
M. Neumann et.al., *PoS* LATTICE2022 (2023)

Neumann M (2023) PhD Thesis Universität Bielefeld

- First step (**Will discuss**) : Density estimation followed by β , m_l, N_σ extrapolation using *Masked Autoregressive flows*

β - reweighting

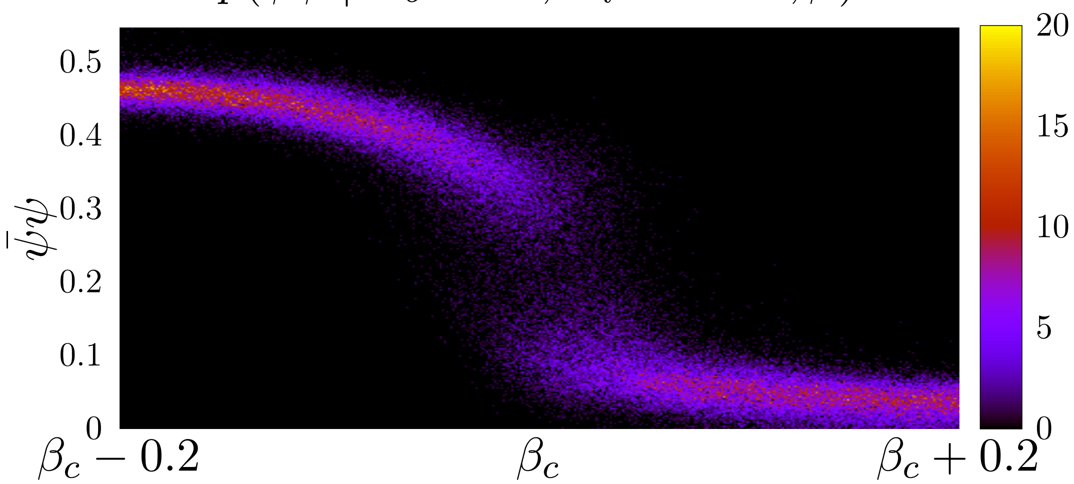




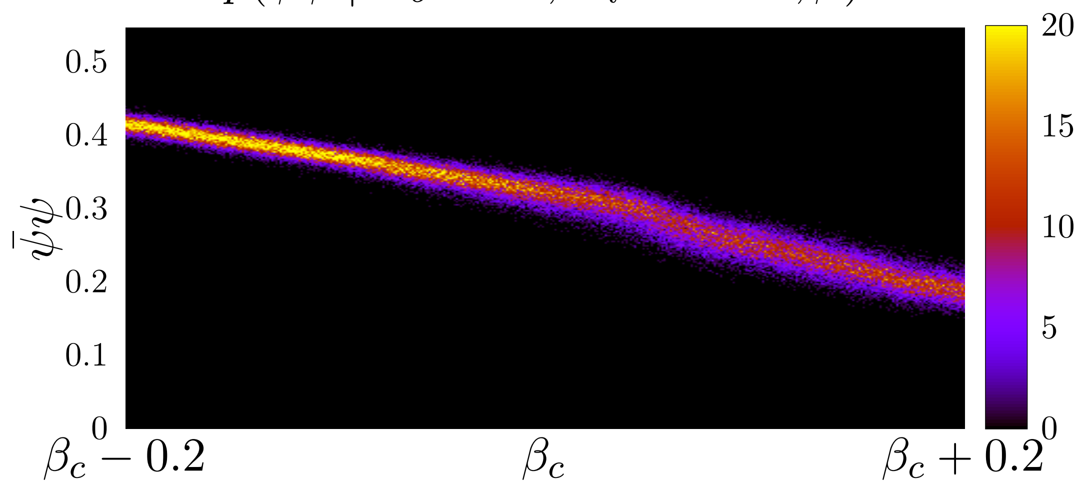
ML - β interpolation

- Second step (**Won't discuss**) : Classification of densities projected as "images" via vision transformers to nail down $m_{l,c}$ for $N_f = 5$

$p(\bar{\psi}\psi \mid N_\sigma = 24, m_l = 0.001, \beta)$



$p(\bar{\psi}\psi \mid N_\sigma = 24, m_l = 0.008, \beta)$



Density estimation using MADE

MADE: Masked Autoencoder for Distribution Estimation

Mathieu Germain **Karol Gregor** **Iain Murray** **Hugo Larochelle**

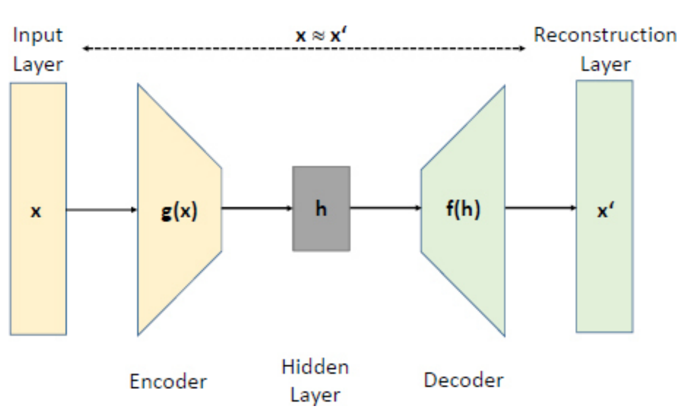
PoS of 32nd International Conference on Machine Learning, France, 2015

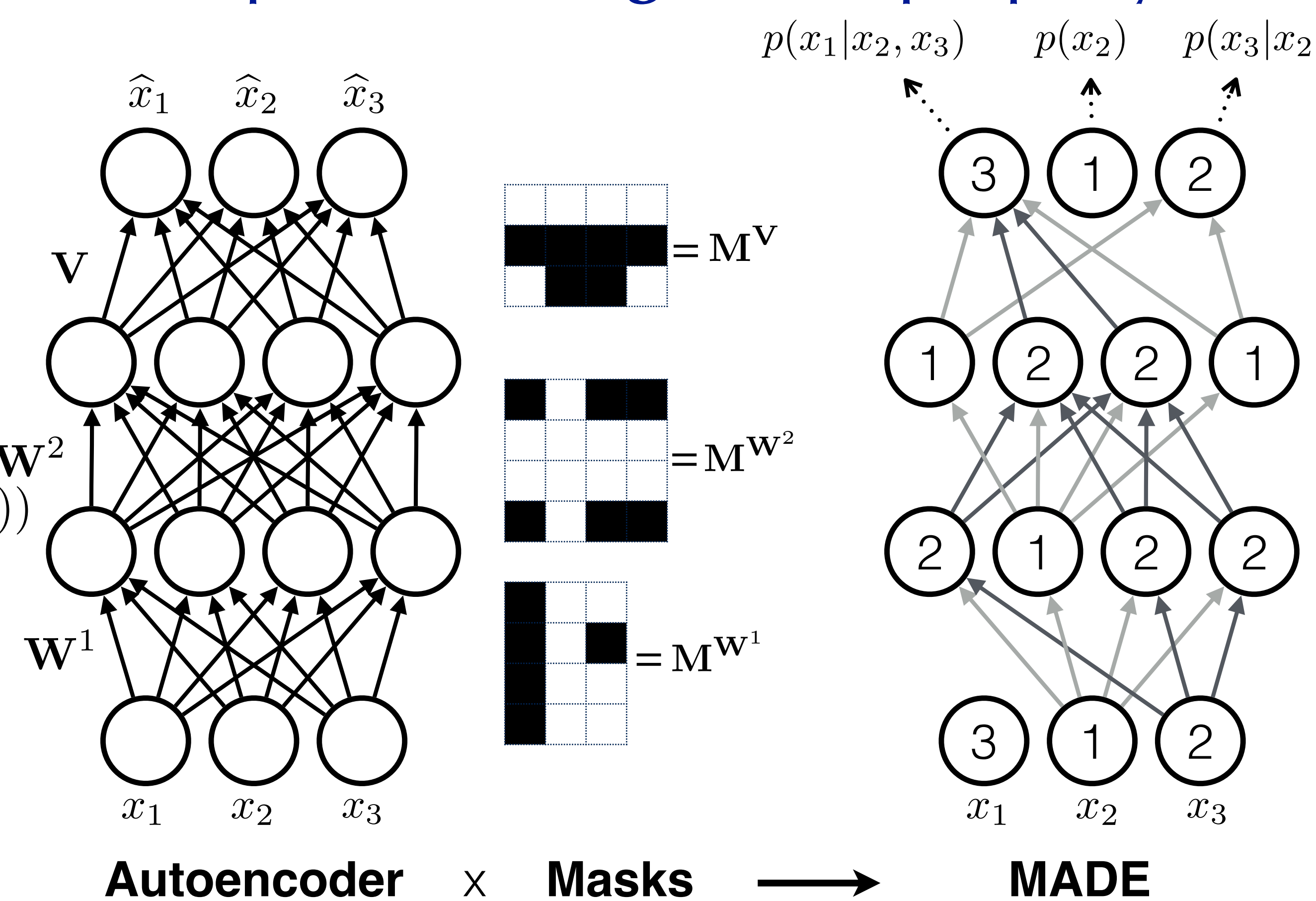
- Goal : Learn a probability density from examples of data $(\vec{x}, \vec{y}) \rightarrow p(\vec{x} | \vec{y})$
- How : Interpret the outputs of an autoencoder as valid probabilities
- Each output as conditional probability and product as joint probability
- Introduce **masks** on hidden layer units to impose **autoregressive** property

$$\mathbf{h}(\mathbf{x}) = \mathbf{g}(\mathbf{b} + \mathbf{W}\mathbf{x})$$
$$\widehat{\mathbf{x}} = \mathrm{sigm}(\mathbf{c} + \mathbf{V}\mathbf{h}(\mathbf{x}))$$

Masking

$$\mathbf{h}(\mathbf{x}) = \mathbf{g}(\mathbf{b} + (\mathbf{W} \odot \mathbf{M}^{\mathbf{W}})\mathbf{x})$$
$$\hat{\mathbf{x}} = \mathrm{sigm}(\mathbf{c} + (\mathbf{V} \odot \mathbf{M}^{\mathbf{V}})\mathbf{h}(\mathbf{x}))$$





Masked Autoregressive Flows

- **Autoregressive property** from conditionals

$$p(x_1, x_2 \dots x_D) = p\left(x_N | x_1, \dots x_{N-1}\right) p\left(x_{N-1} | x_1, \dots x_{N-2}\right) \dots p\left(x_1\right)$$

- Each conditional as a single Gaussian : $p\left(x_i | \vec{x}_{1:i-1}\right) = \mathcal{N}\left(x_i | \mu_i, (\exp(\alpha_i))^2\right)$ with $\mu_i = f_{\mu_i}\left(\vec{x}_{1:i-1}\right)$ and $\alpha_i = f_{\alpha_i}\left(\vec{x}_{1:i-1}\right)$

- Data generated via : $x_i = u_i \exp(\alpha_i) + \mu_i$ with $u_i \sim \mathcal{N}(0,1)$

- A flow is then constructed by MADE blocks in a chain

Masked Autoregressive Flow for Density Estimation

George Papamakarios
University of Edinburgh
g.papamakarios@ed.ac.uk

Theo Pavlakou
University of Edinburgh
theo.pavlakou@ed.ac.uk

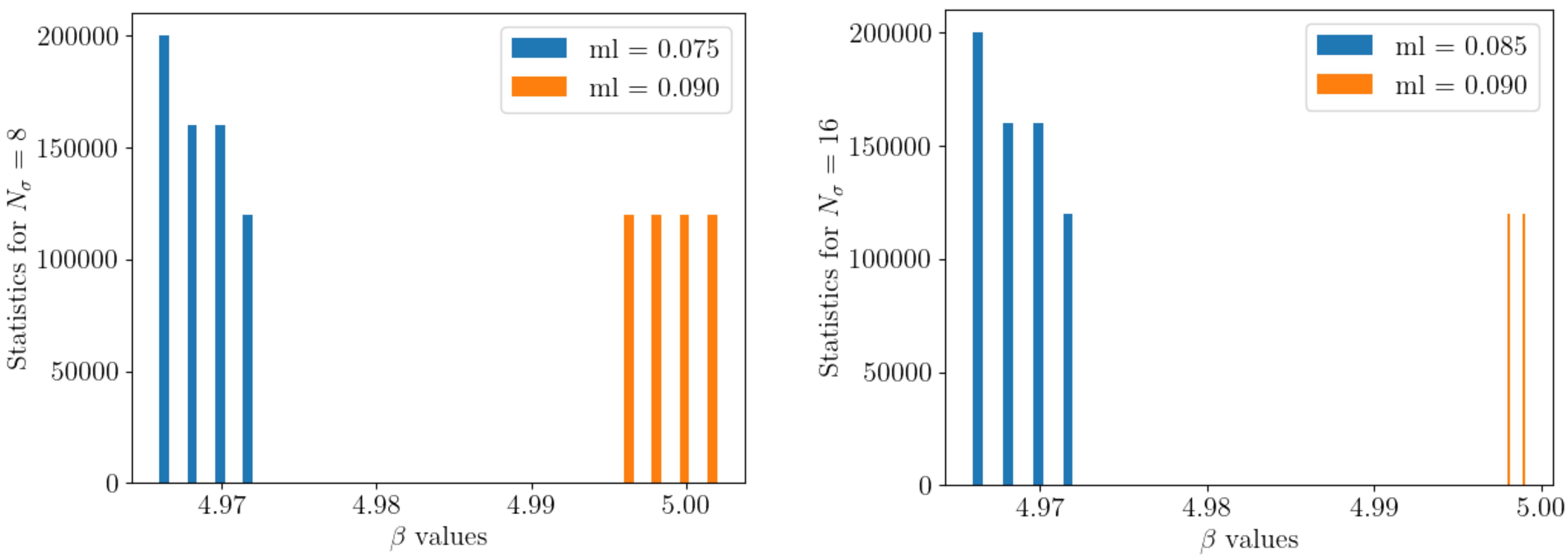
Iain Murray
University of Edinburgh
i.murray@ed.ac.uk

Relevant parameters for this analysis (see PhD thesis of M. Neumann)

MAF parameter	value	
kernel regulizer	L1L2	
L1	0.0001	
L2	0.0001	
loss function	- log prob	
number of MADE blocks	8	
number of samples	~~1000000~~	10K - 100K
number of epochs	500	
number of inputs	2 $(S, \bar{\psi}\psi)$	
number of conditional inputs	3 (β, m_l, N_σ)	
batch size	~~1024~~	2048
amount of training data	~~1.583.962~~ x $(S, \bar{\psi}\psi)$~ 3.400.000	
optimizer	Adam	

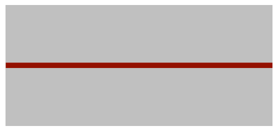
The goal : Test the procedure for different data

- Goal : To reproduce the Z2 critical boundary via ML for [F. Cuteri et.al., *JHEP* 11 (2021)]
- Un-improved staggered quarks $N_f = 5$, $N_\tau = 4$ with $N_\sigma \in \{8, 12, 16\}$ and $m_l \in \{0.075, 0.080, 0.085, 0.090\}$
- Trained only on $N_\sigma \in \{8, 16\}$, total training data ~3.4 million values for $\left(\bar{\psi}\psi, S\right)$

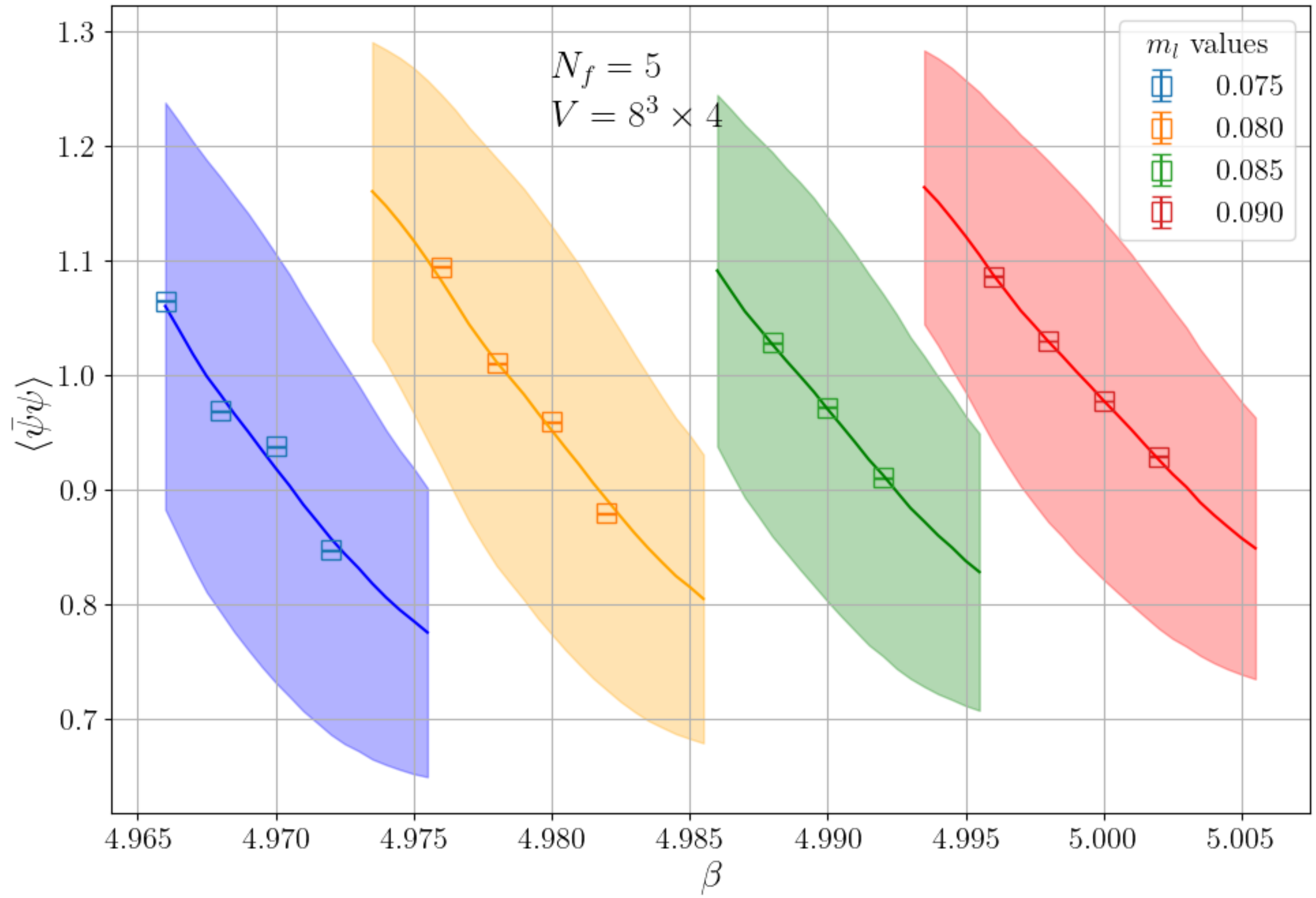


Results : $\langle \bar{\psi}\psi \rangle$ for $N_\sigma = 8$

- Training done by removing **all** $N_\sigma = 12$ data
- Quantity obtained : $p\left(\bar{\psi}\psi, S \,|\, N_\sigma, m_l, \beta\right)$
- Results for 100K evaluations of the model

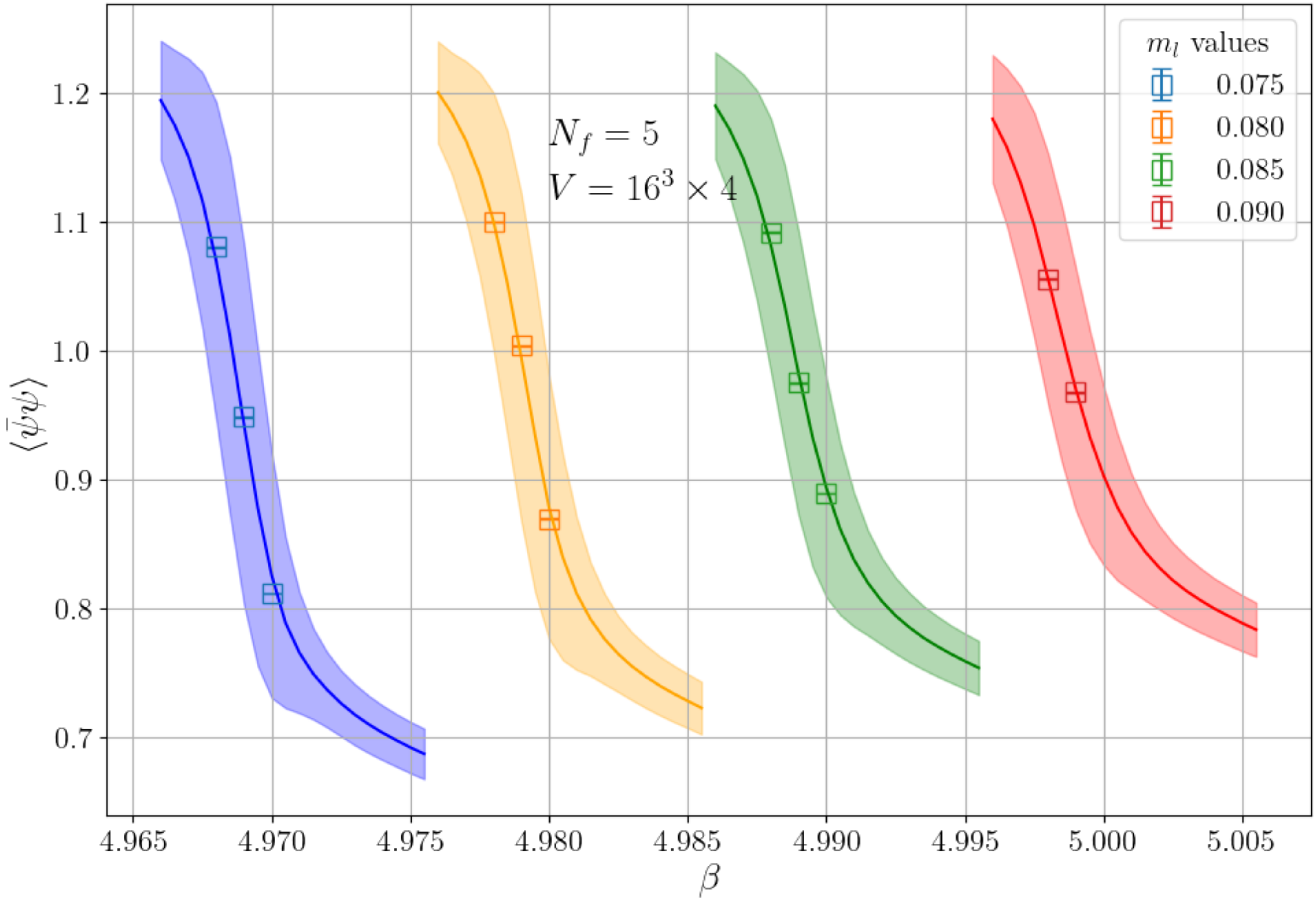


MAF prediction for the β interpolation on training set



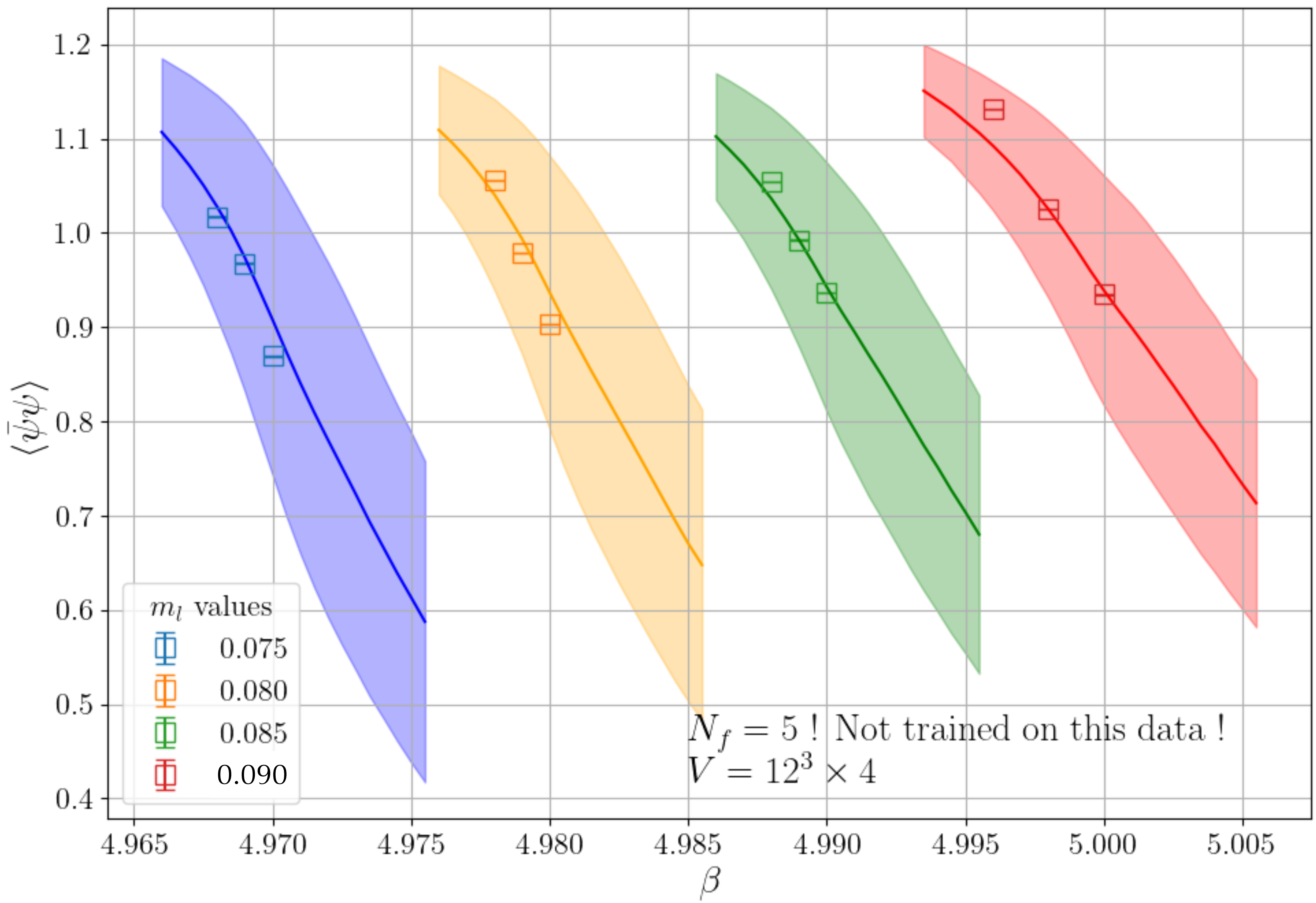
Results for $\langle \bar{\psi}\psi \rangle$ for $N_\sigma = 16$

MAF prediction for the β interpolation on training set



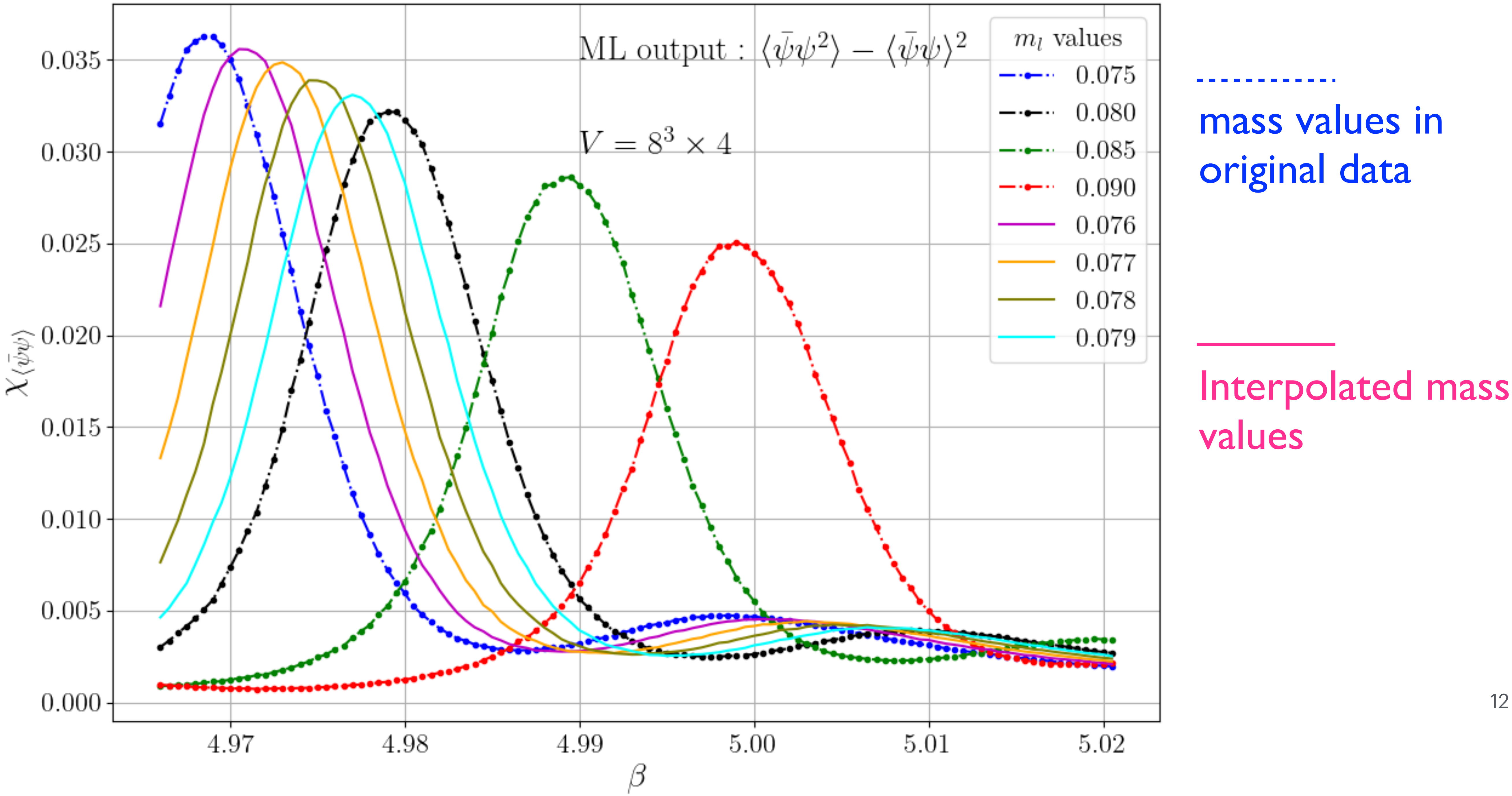
Results : $\langle \bar{\psi}\psi \rangle$

MAF prediction for $N_\sigma = 12$ (genuine prediction !)

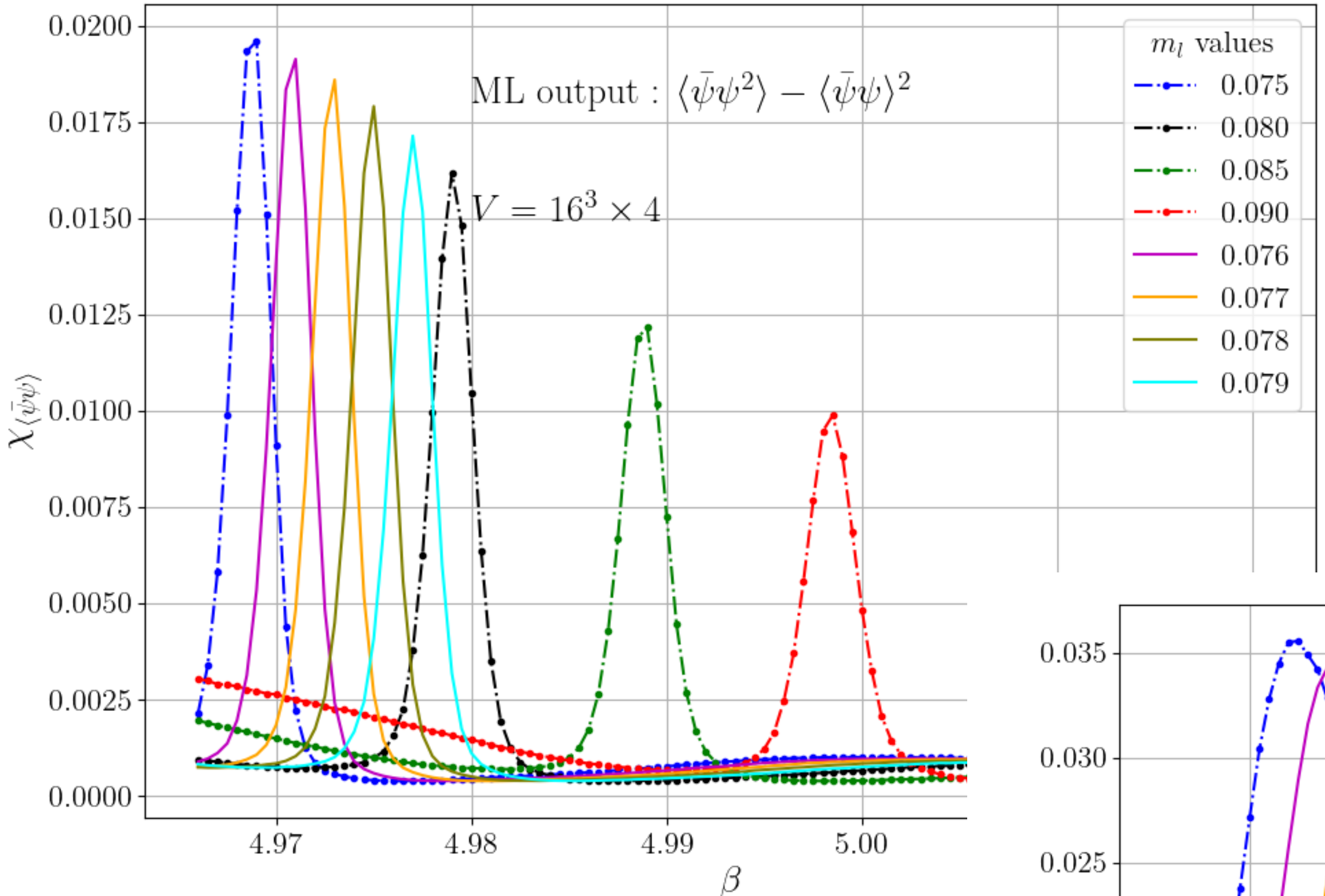


Results : $\chi_{\bar{\psi}\psi}$ for $8^3 \times 4$

- With $p\left(\bar{\psi}\psi, S \middle| N_\sigma, m_l, \beta\right)$ - we are free to compute higher moments !
- We see scaling of peak height, width, location from ML prediction

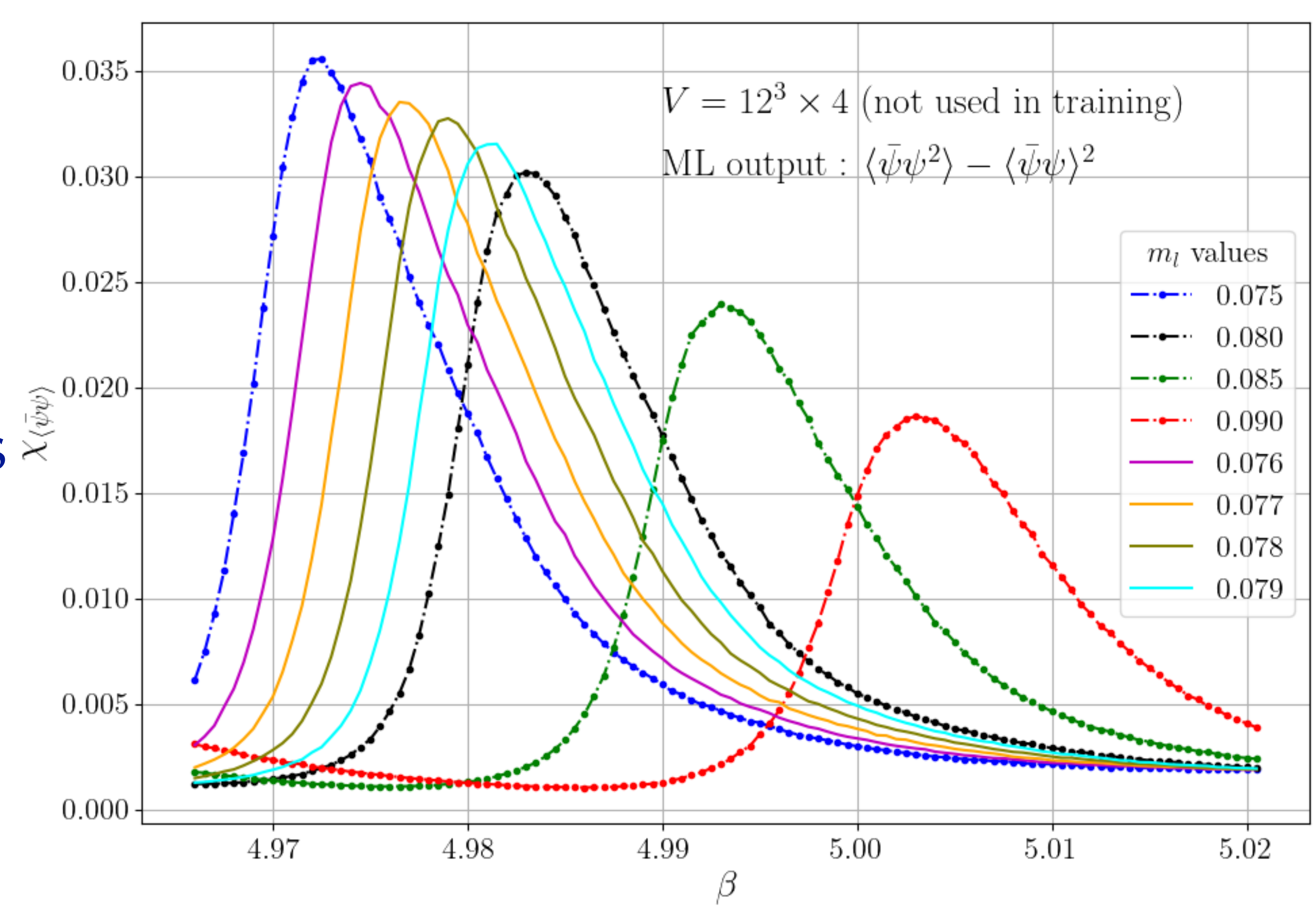


Results for $\chi_{\bar{\psi}\psi}$ for 16^3 , $12^3 \times 4$

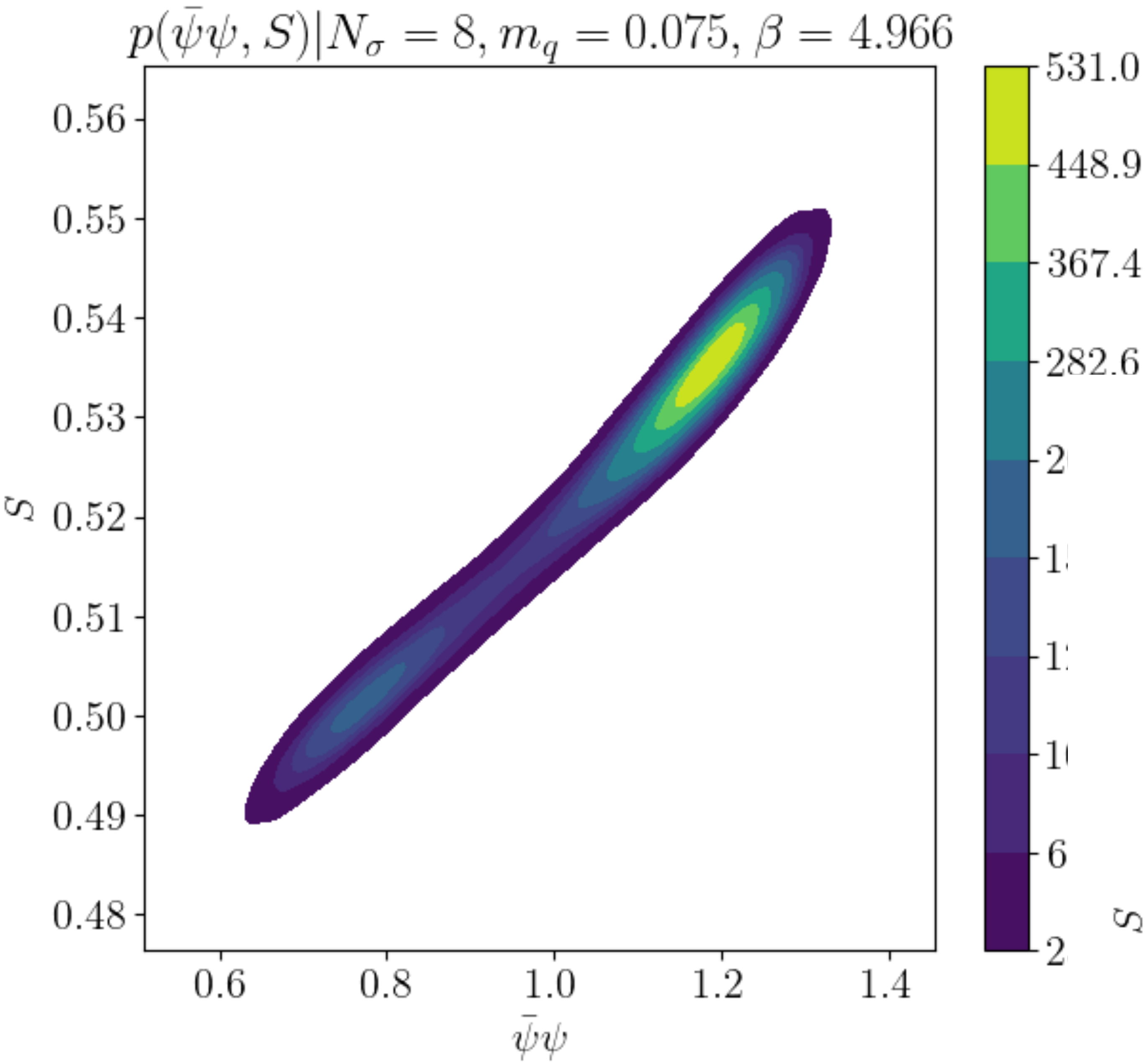


With **increasing lattice volume** we see shifting and narrowing of peaks - indicating a phase transition

With **decreasing bare quark** mass we see shifting and narrowing of peaks - indicating a phase transition



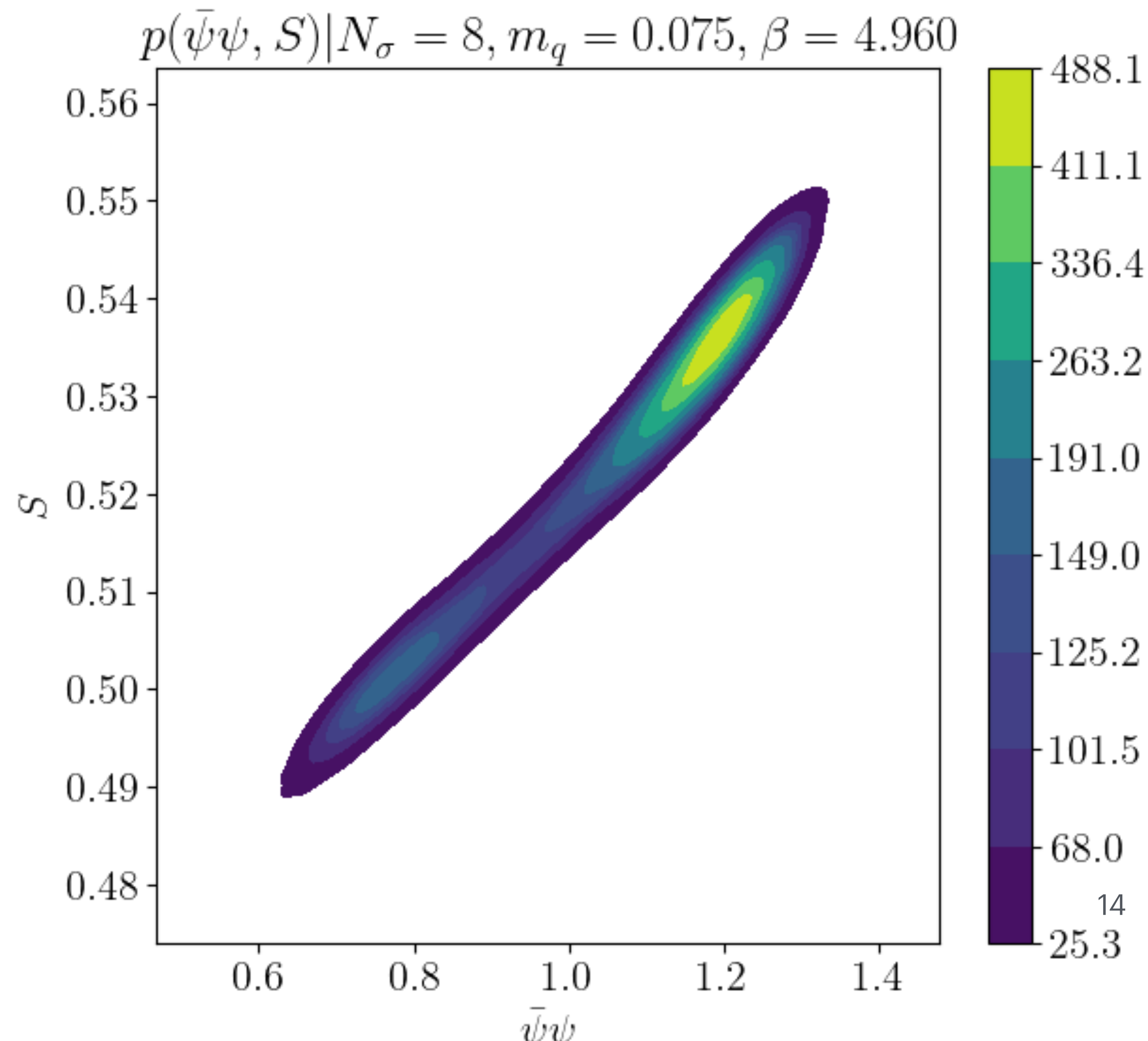
Results for $p\left(\bar{\psi}\psi, S\right)$ for some N_σ, m_l, β



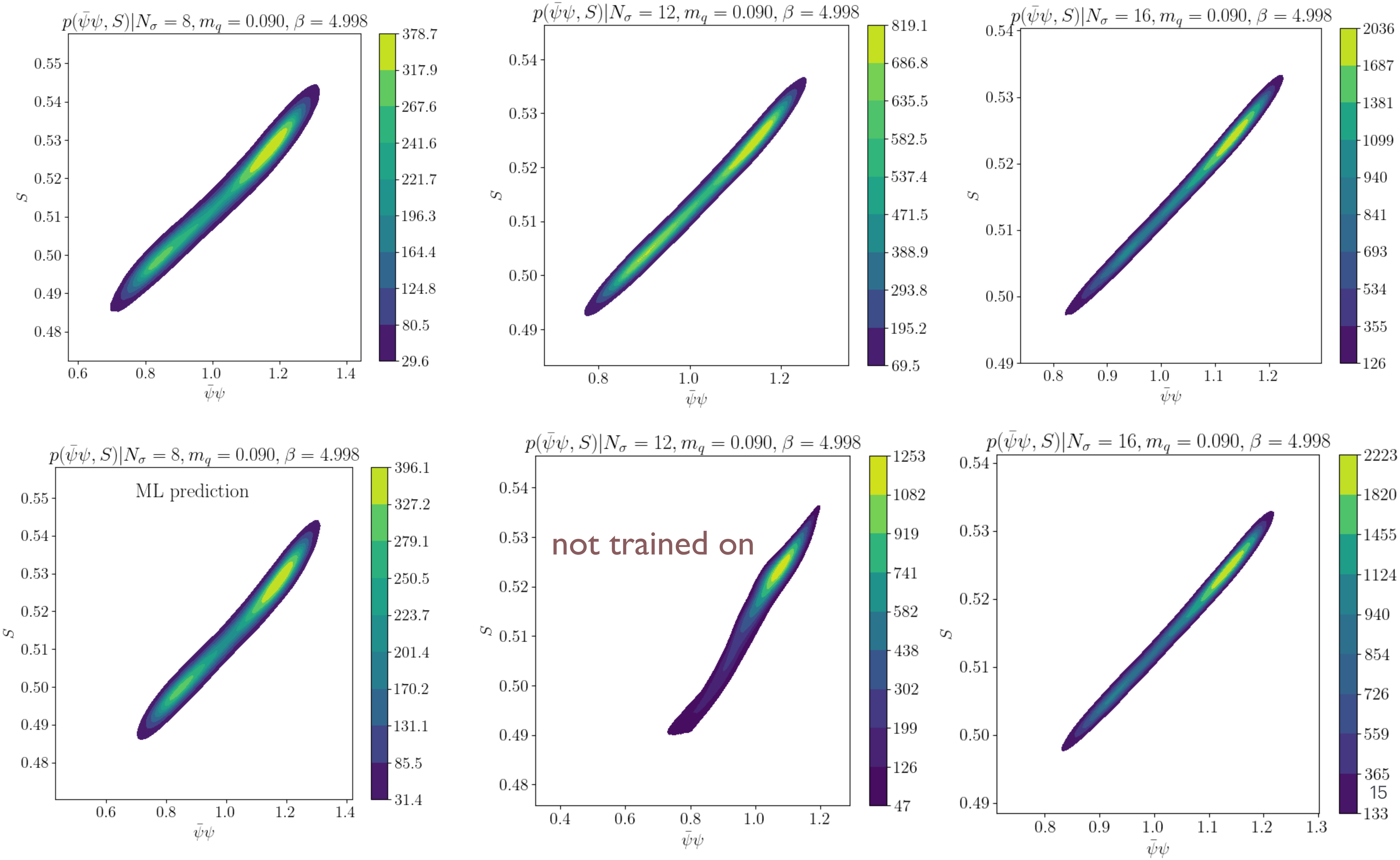
From the data

Another indication of learning the correct density

MAF prediction



MAF Inference vs standard finite size scaling from lattice studies



Ongoing steps

- Currently the model doesn't (want to) train on the $N_\sigma = 12$ data

- Only running on TensorFlow for CPUs - required package not compatible with current TF version & Model doesn't compile on new TF version

- Expand the conditionals to N_τ and N_f

- Explore in the direction of a statement made in [G. Papamakarios et. al., 1705.07057]

"… accurate densities do not necessarily imply good performance in other tasks, such as in data generation … Choice of method should be informed by whether the application at hand calls for accurate densities, latent space inference or high quality samples "

A possible direction ?

- Learning probability densities for correlators typically needed in spectral function reconstruction

- Can we make G_E "continuous" in T to get a better re-constructed spectral function ?

$$G_E(\tau, T) = \int_0^\infty \frac{\mathrm{d}\omega}{\pi} \rho_E(\omega, T) \frac{\cosh[\omega\tau - \omega/(2T))]}{\sinh[\omega/(2T)]},$$

L. Altenkort et. al., PHYSICAL REVIEW LETTERS **130**, 231902 (2023)