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)*

CRC–TR 211
Strong-interaction matter under extreme conditions

New developments in the studies of the QCD phase diagram @ ECT* Trento

September 13, 2024

Outline

I. Motivation – The chiral phase transition

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

III. The Machine Learning model – Masked Autoregressive Flows

IV. (new) $N_f = 5$ using unimproved staggered

V. Results on density estimation

VI. Summary and outline

The chiral transition in lattice QCD

- Nature of the chiral transition in the chiral limit is of much research interest - although many emerging results indicate possibility of a second order transition [F. Cuteri et.al., *JHEP* 21, S. Sharma et al PRD 22 & PhD thesis 21, see talks O. Philipsen and Y. Zhang Wednesday, and more]

- On lattice, such a study necessarily requires extrapolation to zero quark mass - simulating even close to this limit is numerically challenging

- 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)

- One of the (many) results of this study was to find the Z2 boundary separating the first order and the crossover region at finite lattice spacings as a function of N_f

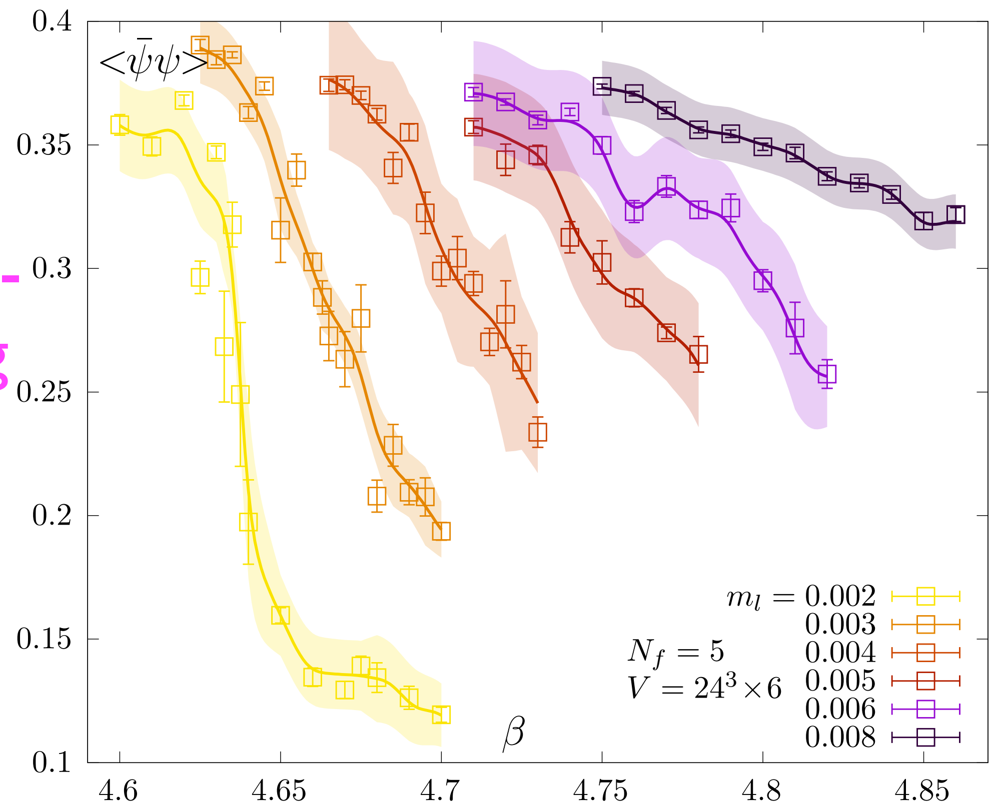
- In M. Neumann et.al., *PoS* LATTICE2022 (2023), the authors studied $N_f = 5$ degenerate quarks and determined the Z2 boundary - replacing some of the finite size scaling analysis with novel Machine Learning (ML) techniques

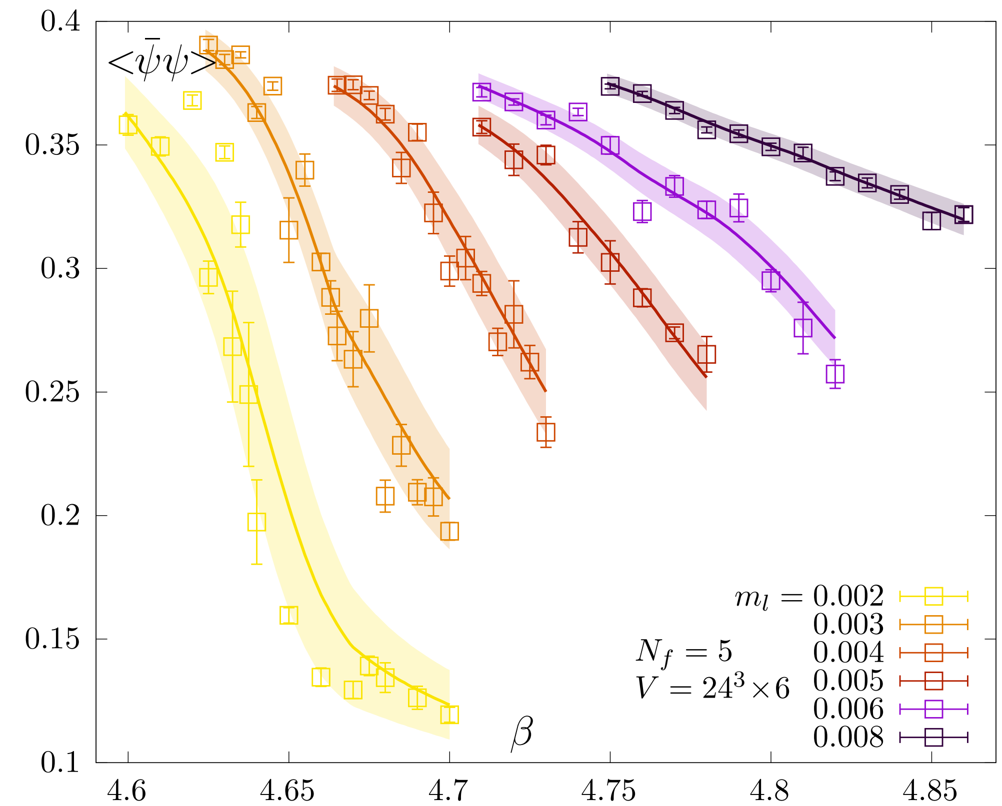
- The goal of the present work is to apply this analysis to data published in F. Cuteri et.al., *JHEP* 11 (2021) to see if the ML analysis can reproduce their results

Z2 boundary for Nf=5 HISQ

- The ML technique used in this work aims at the joint probability densities $p\left(\bar{\psi}\psi, S\right)$ conditioned on lattice parameters like N_σ, m_l, β
- Learning such a density correctly allows interpolation in the dimensions of the conditional inputs - avoiding some expensive lattice simulations
- Interpolation in the gauge coupling already exits (β re-weighting) - can this ML technique do better?

Classical β - reweighting





ML - β interpolation

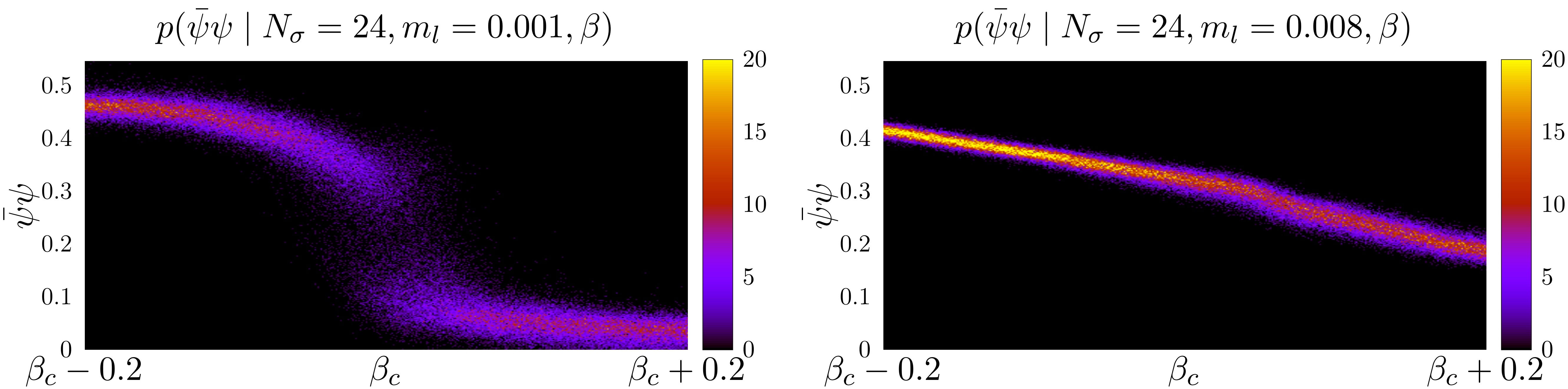
M. Neumann et.al., *PoS* LATTICE2022 (2023)

Z2 boundary for Nf=5 HISQ

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

Neumann M (2023) PhD Thesis Universität Bielefeld

- First step : Density estimation followed by β , m_l, N_σ extrapolation using *Masked Autoregressive flows*
- Second step : Using the marginal probability $p\left(\bar{\psi}\psi \,|\, N_\sigma, m_l, \beta\right)$ to identify first-order regions along the β and m_l axes
- In the original analysis a further classification algorithm was used to compute the critical mass that separates the first-order regions from the crossover
- Alternatively, one should compute Binder cumulants like B_3 and B_4 to determine β_c and $m_{l,c}$ [O. Philipsen PoS LATTICE2019 (2019) 273]



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

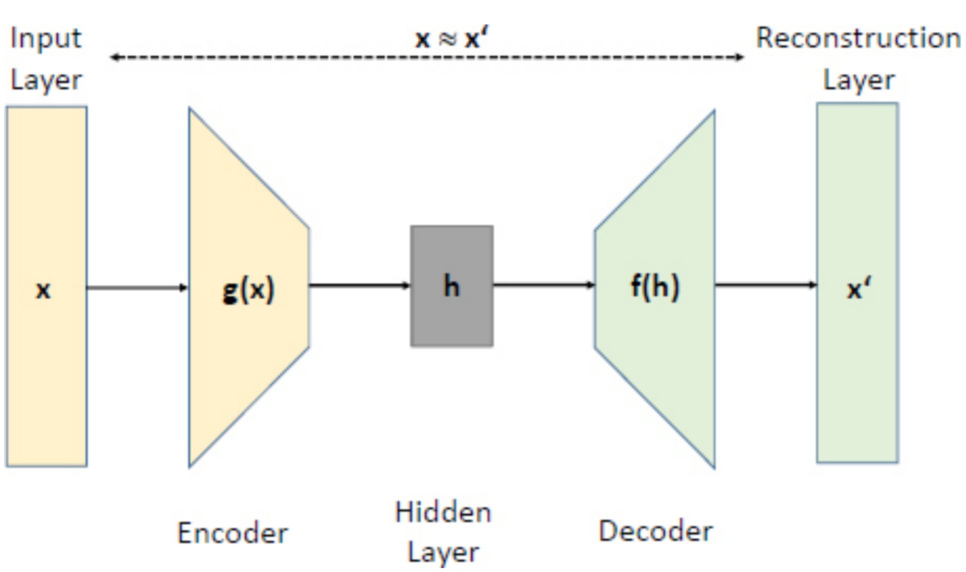
Density estimation using MADE

- 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 Neural Network as conditional probabilities
- Why : $p(x_1, x_2 \ldots x_D) = p(x_N | x_1, \ldots x_{N-1})\, p(x_{N-1} | x_1, \ldots x_{N-2}) \ldots p(x_1)$

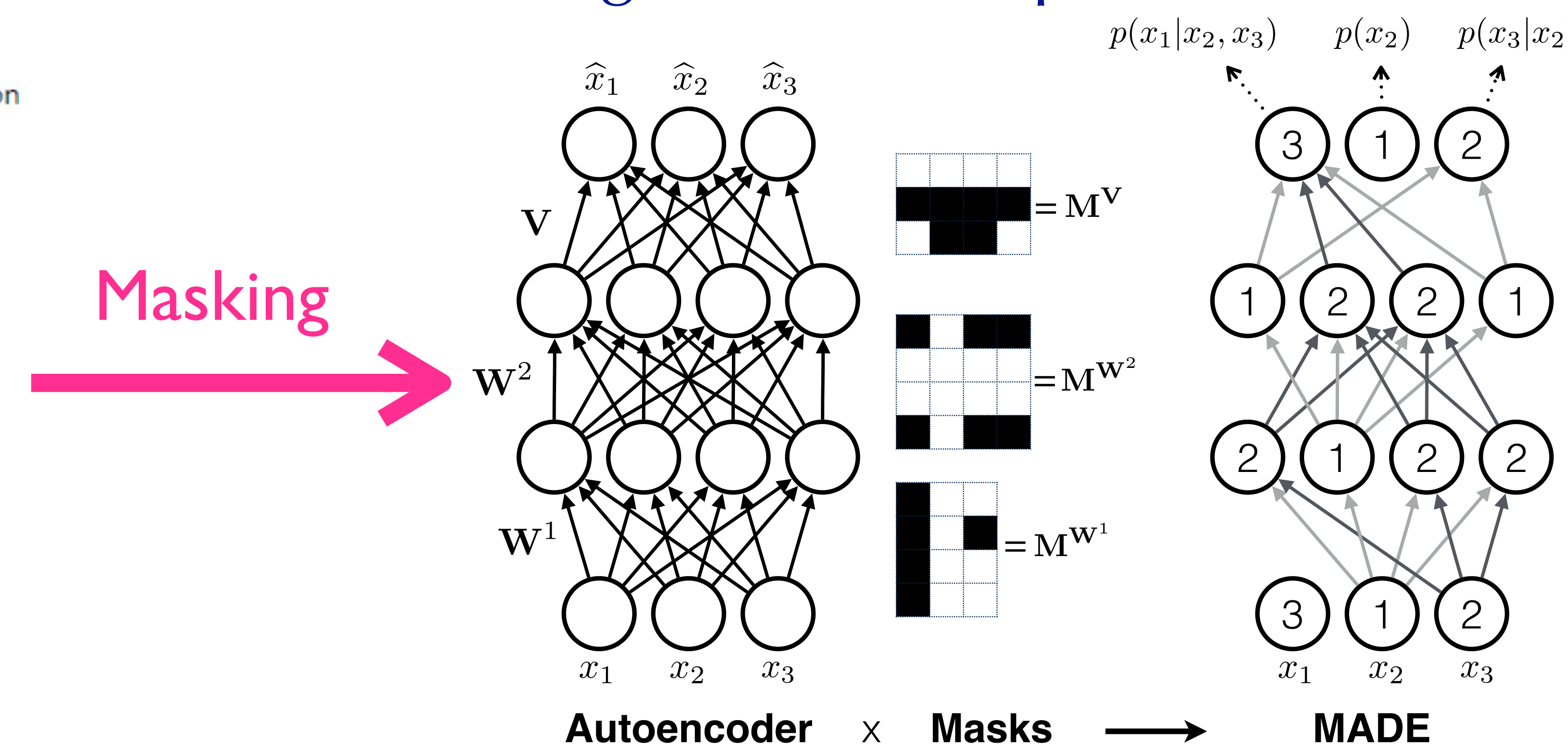
MADE: Masked Autoencoder for Distribution Estimation

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

- The authors used **masking** of connections in an Autoencoder to implement the **autoregressive property** needed for constructing conditional probabilities :



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



Masked Autoregressive Flows

- Next step : Combine these MADE blocks as a chain to make a **Masked Autoregressive Flow**

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

- A flow is then constructed by MADE blocks in a chain - more blocks add complexity to the estimated density - each of whose random numbers modelled by the previous block

- 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 $u_i \sim \mathcal{N}(0,1)$

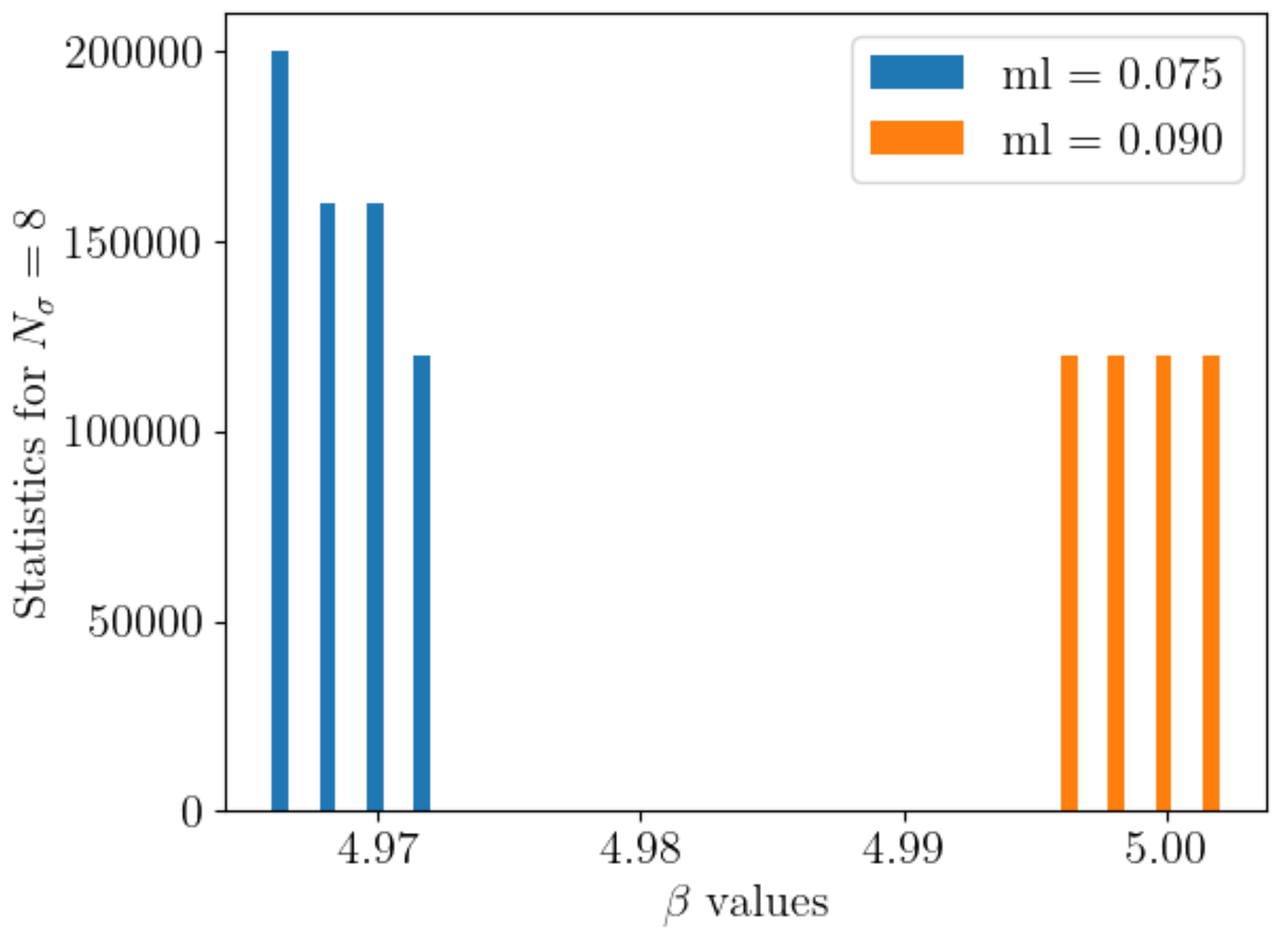
- Goal : Maximise the log-likelihood of the data under the NN model

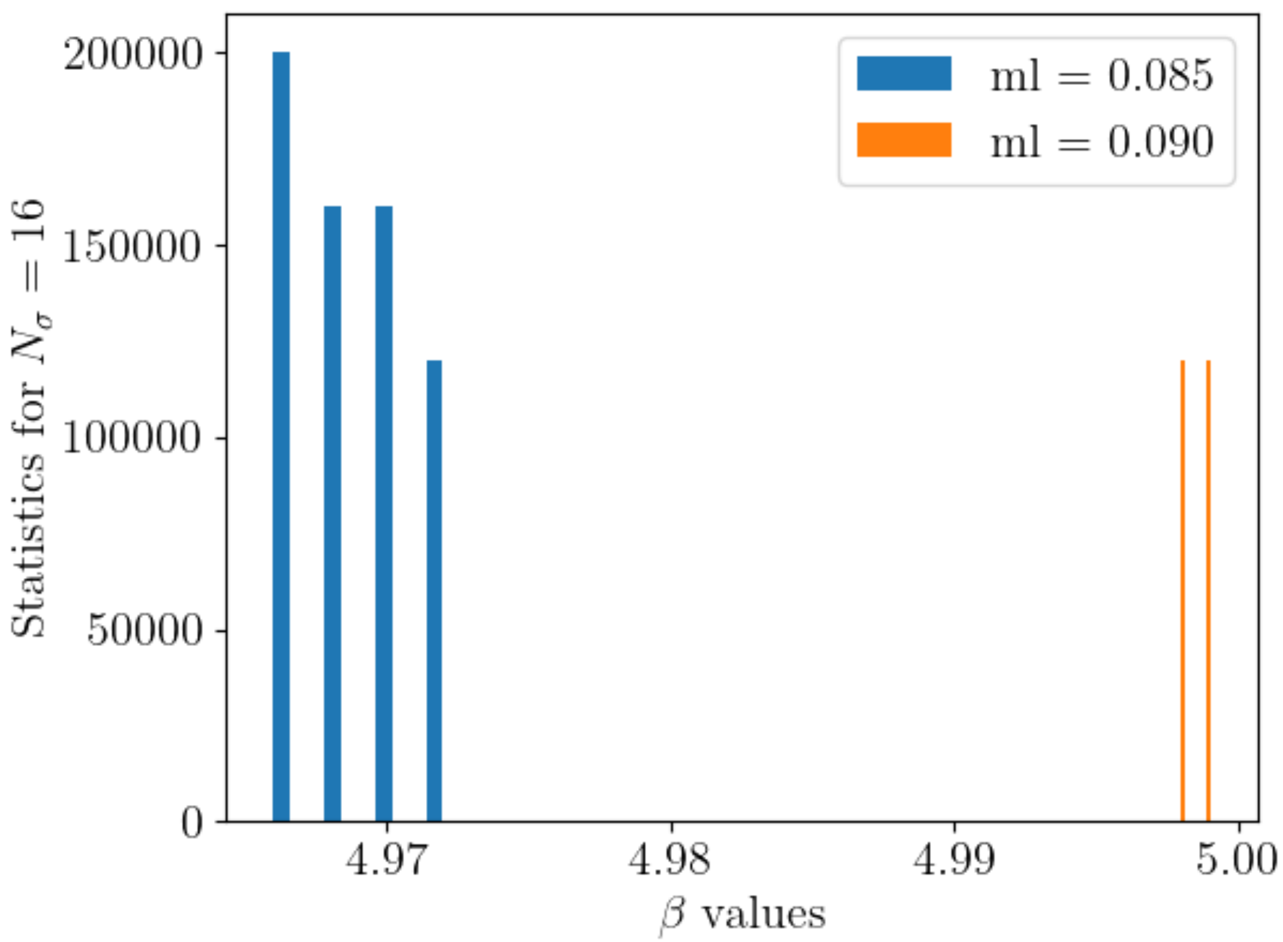
Goal 1: Test the procedure by removing data

- 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 $m_l \in \{0.075, 0.080, 0.085, 0.090\}$ [!Frankfurt Data!]

- Initially trained only on $N_\sigma \in \{8, 16\}$, total training data ~3.4 million values for $(\bar{\psi}\psi, S)$

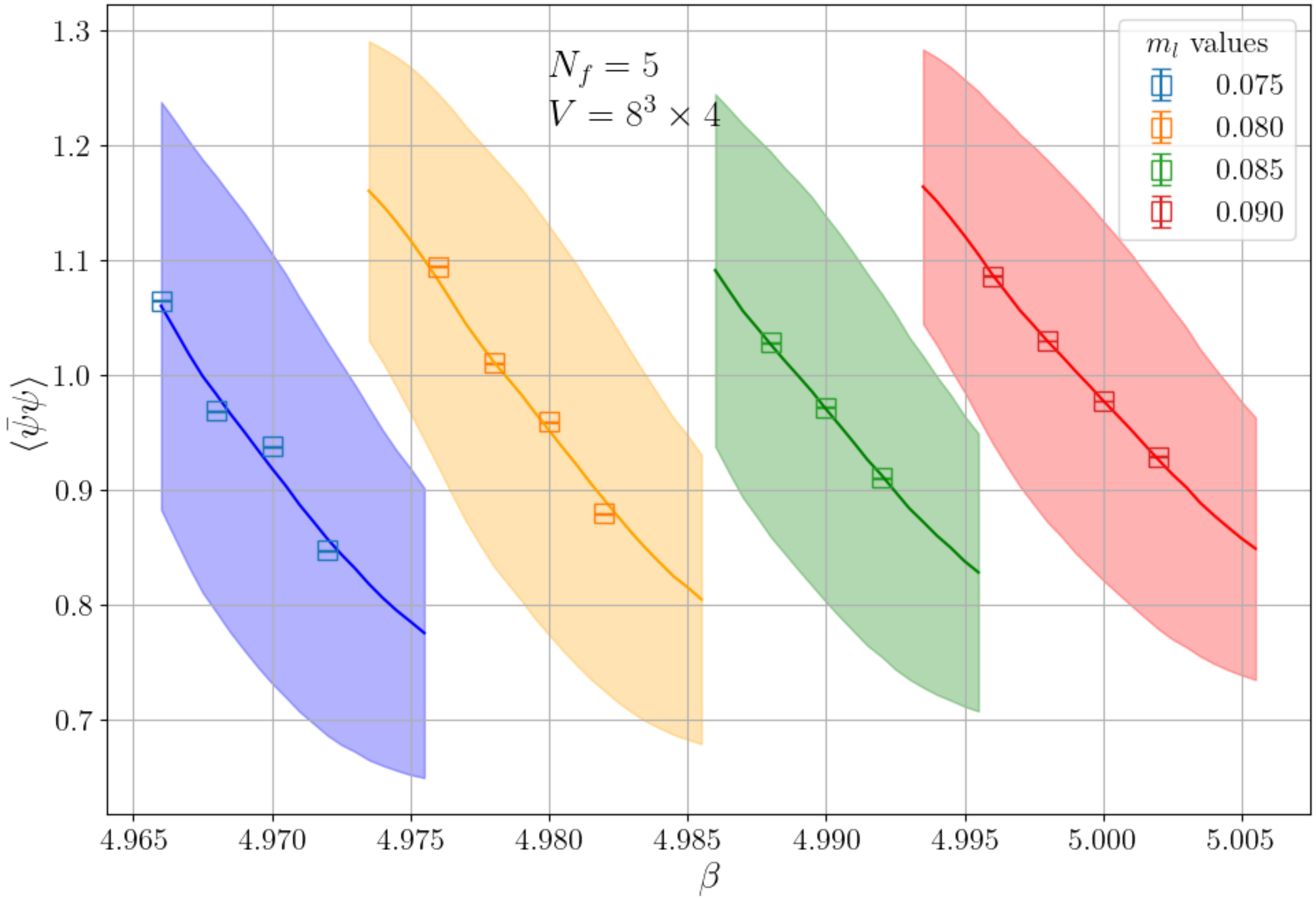




Results : $\langle \bar{\psi}\psi \rangle$ for $N_\sigma = 8$

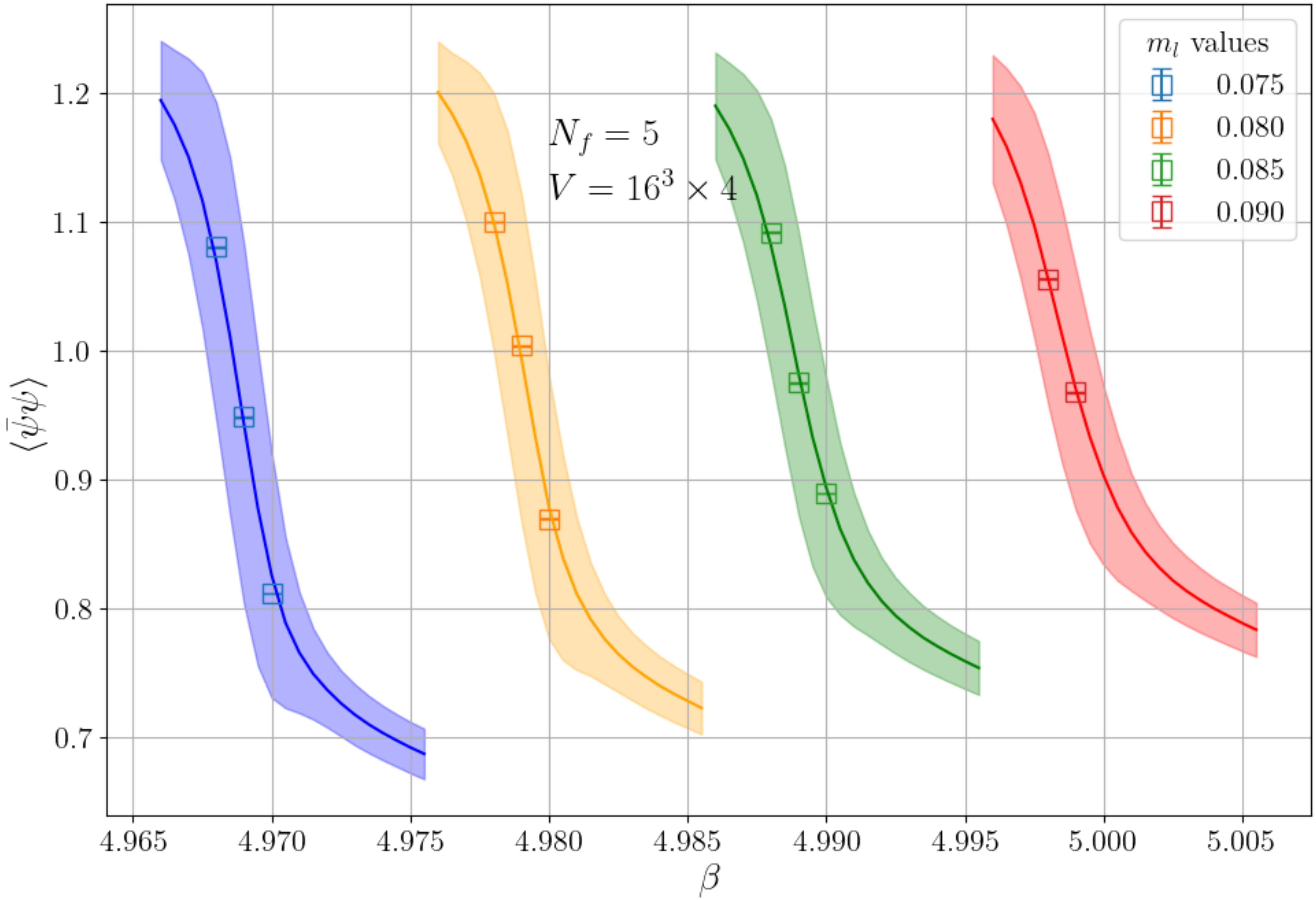
- Training done by removing **all** $N_\sigma = 12$ data - training time ~ 4hr 30 minutes
- Quantity obtained : $p\left(\bar{\psi}\psi, S \,|\, N_\sigma, m_l, \beta\right)$
- Results for 100K evaluations of the model

MAF prediction for the β interpolation on training set



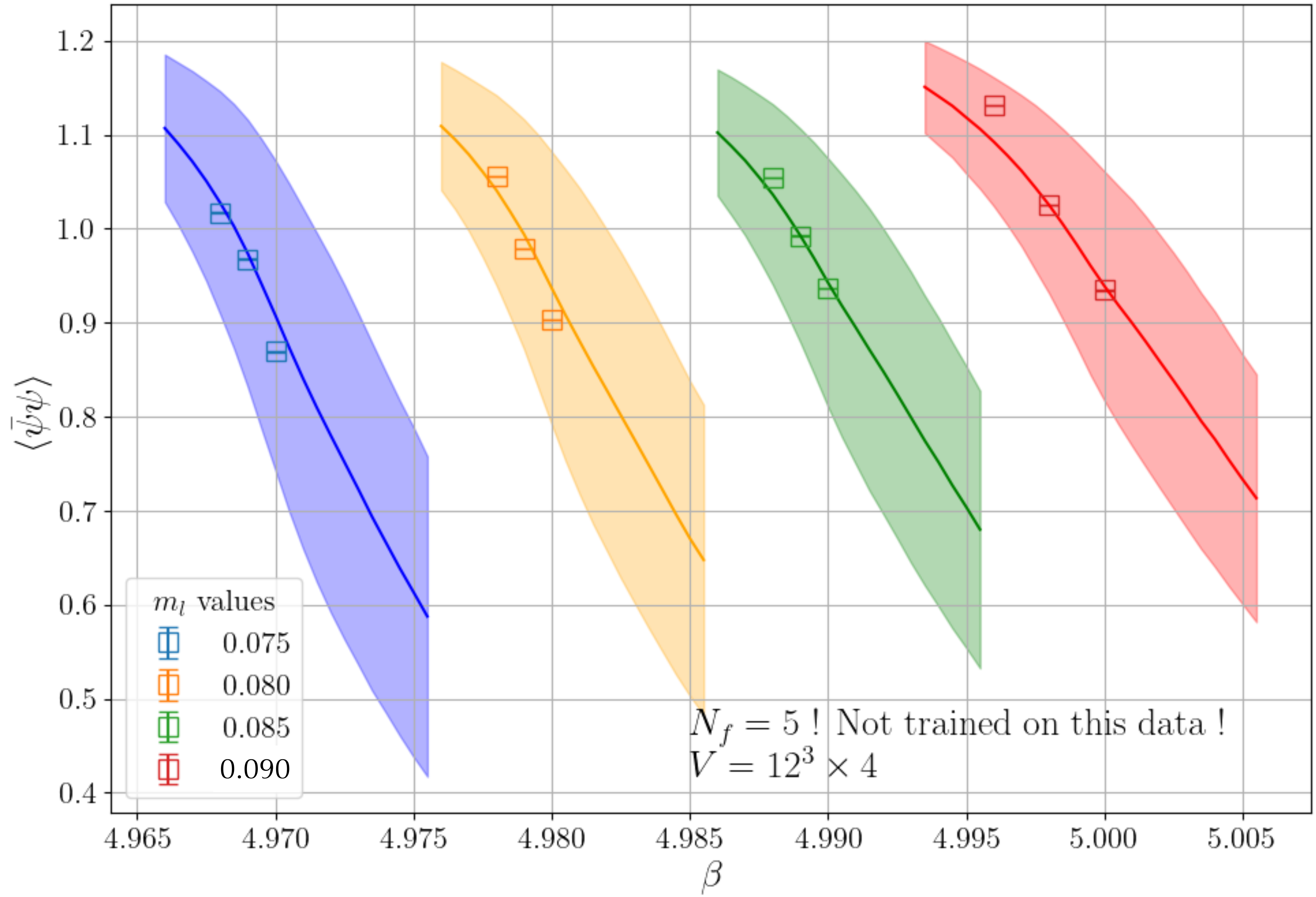
Results for $\langle\bar{\psi}\psi\rangle$ for $N_\sigma = 16$

MAF prediction for the β interpolation on training set



Results for $\langle\bar{\psi}\psi\rangle$ for $N_\sigma = 12$

MAF prediction for volume, β and mass interpolation
$N_\sigma = 12$ (genuine prediction !)



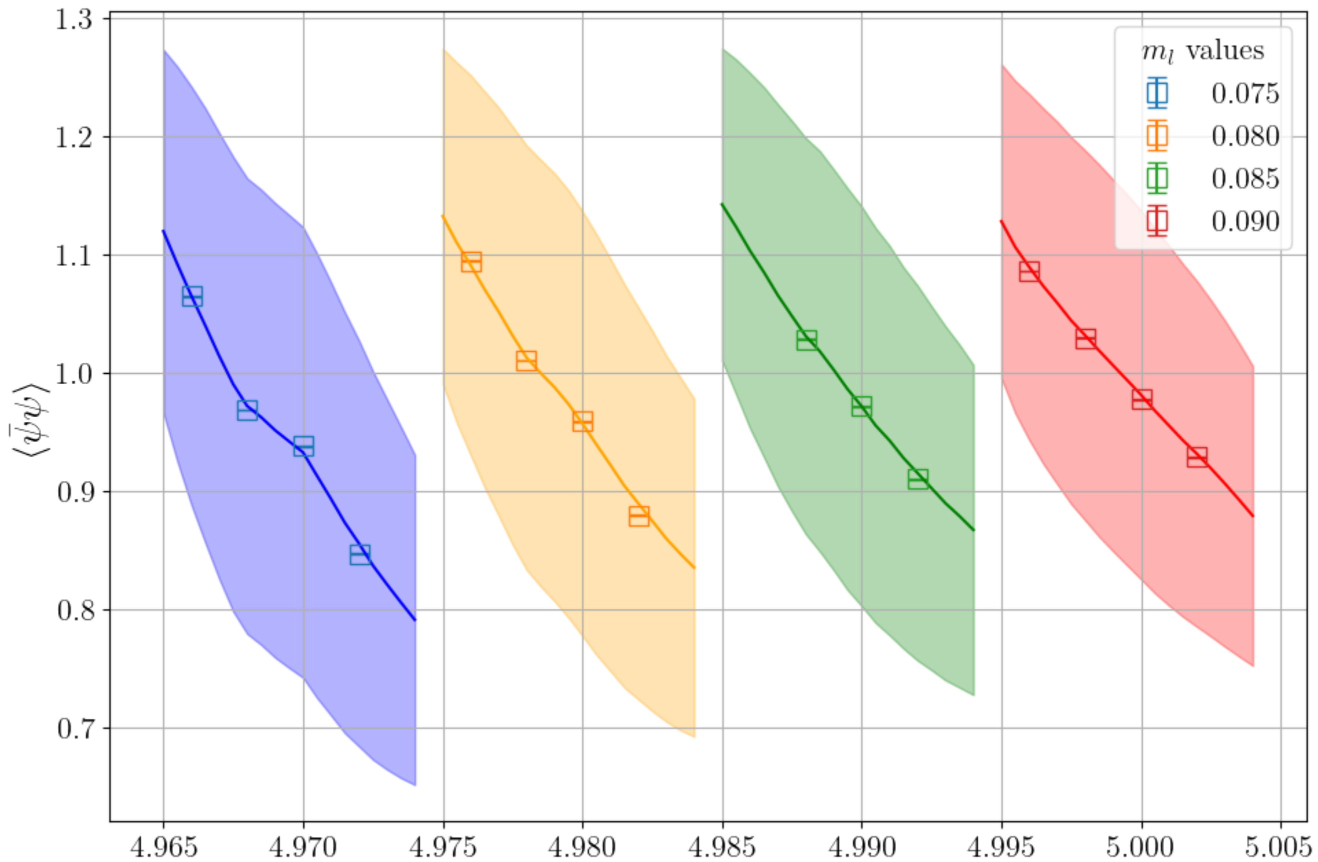
MAF applied to the entire data

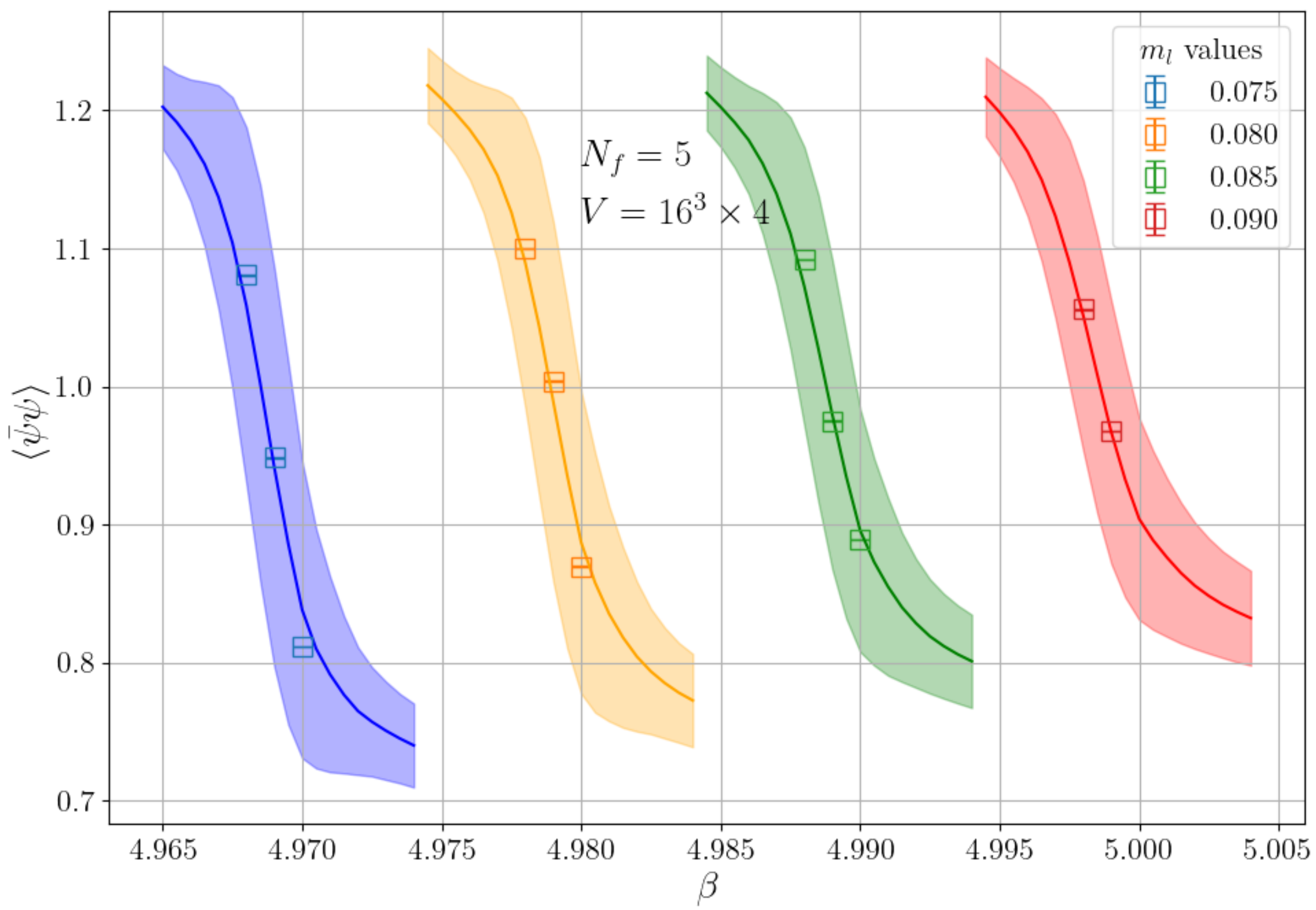
- Goal II : Training the model on all data in order to estimate the Z2 boundary for $N_f = 5$ and $N_\tau = 4$ in accordance with F. Cuteri et.al., *JHEP* 11 (2021) :

What we want to reproduce with the ML analysis

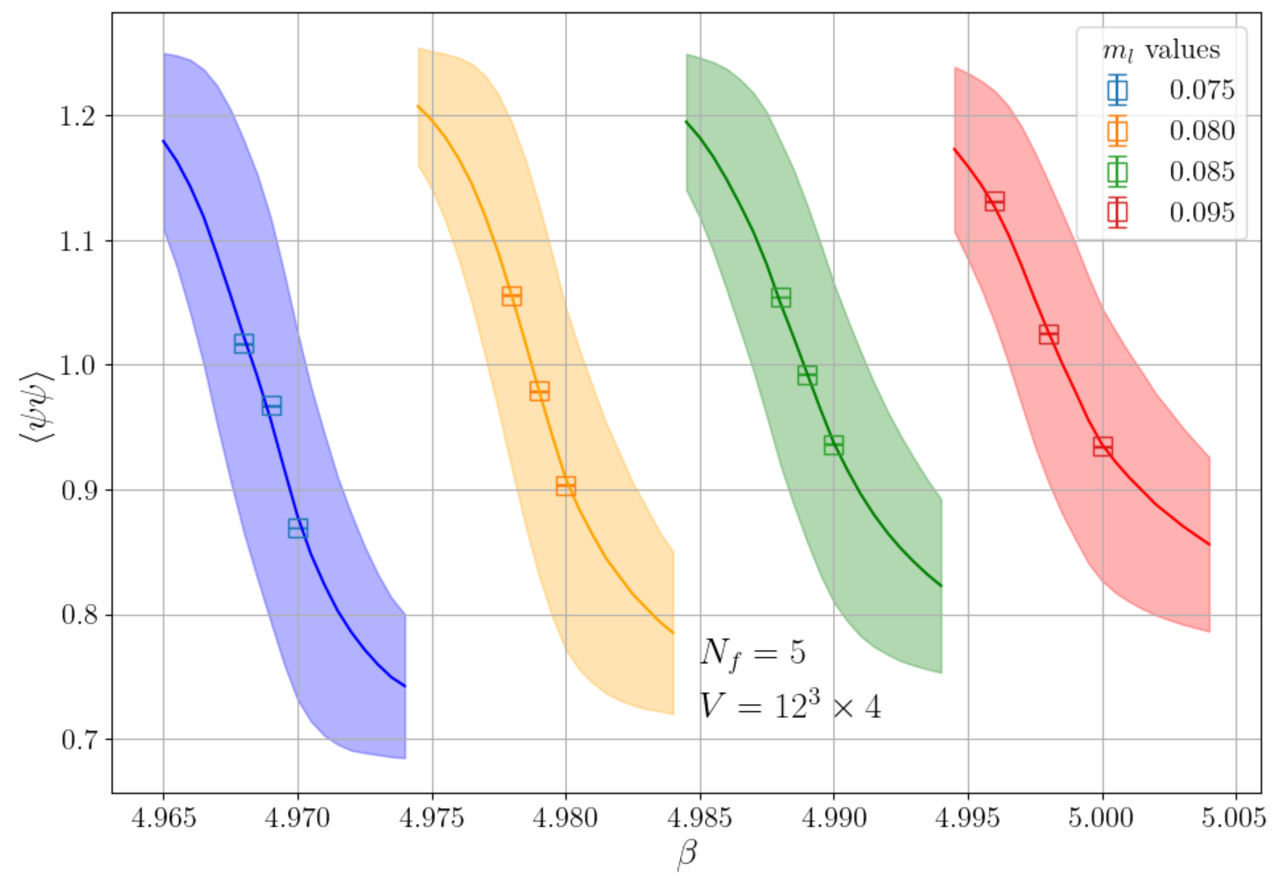
N_τ	N_f	am_min	am_max	am_c	d.o.f.	$\chi^2_\mathrm{d.o.f.}$	$Q[\%]$	β_c at am_c
	2.1	0.0015	0.0045	0.00343(14)	9	0.173	99.7	5.2363(3)
	2.2	0.0025	0.01	0.00579(15)	10	0.257	99	5.2238(3)
	2.4	0.0075	0.015	0.01088(19)	13	0.603	85	5.2006(4)
	2.6	0.0125	0.02	0.01577(23)	10	0.230	99	5.1779(5)
	2.8	0.0175	0.025	0.02106(25)	10	0.270	99	5.1568(5)
4	3	0.0225	0.3	0.0264(5)	10	0.164	99.8	5.1368(9)
	4	0.05	0.065	0.0551(7)	10	0.365	96	5.0529(13)
	5	0.07	0.09	0.0820(8)	12	0.734	72	4.9828(15)
	6	0.1	0.12	0.1078(6)	7	1.148	33	4.9234(12)
	7	0.12	0.14	0.1308(8)	7	0.874	53	4.8692(18)
	8	0.14	0.17	0.1539(11)	7	0.668	7	4.8233(24)

MAF applied to the entire data

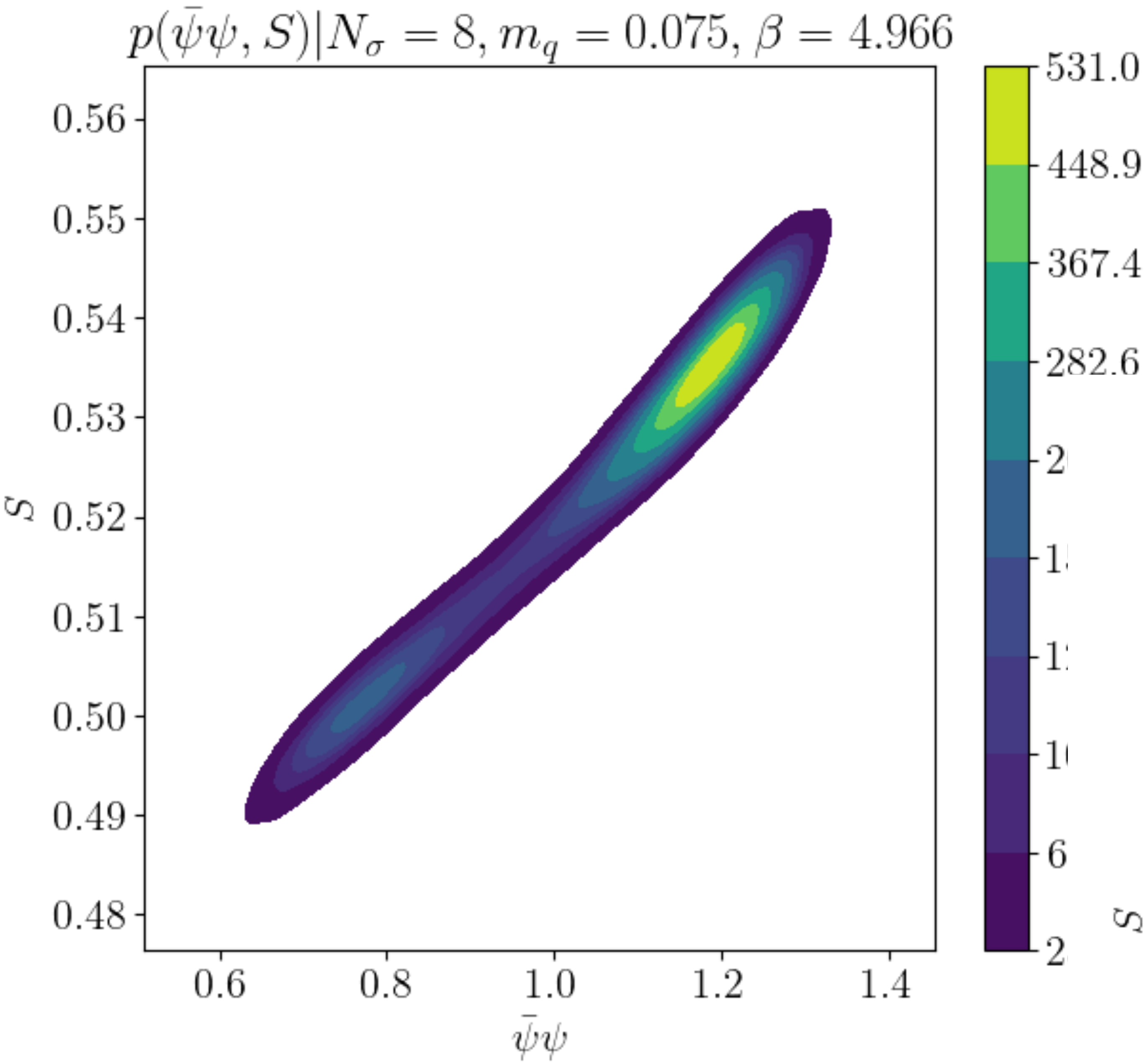




Check that the model still reproduces the data !



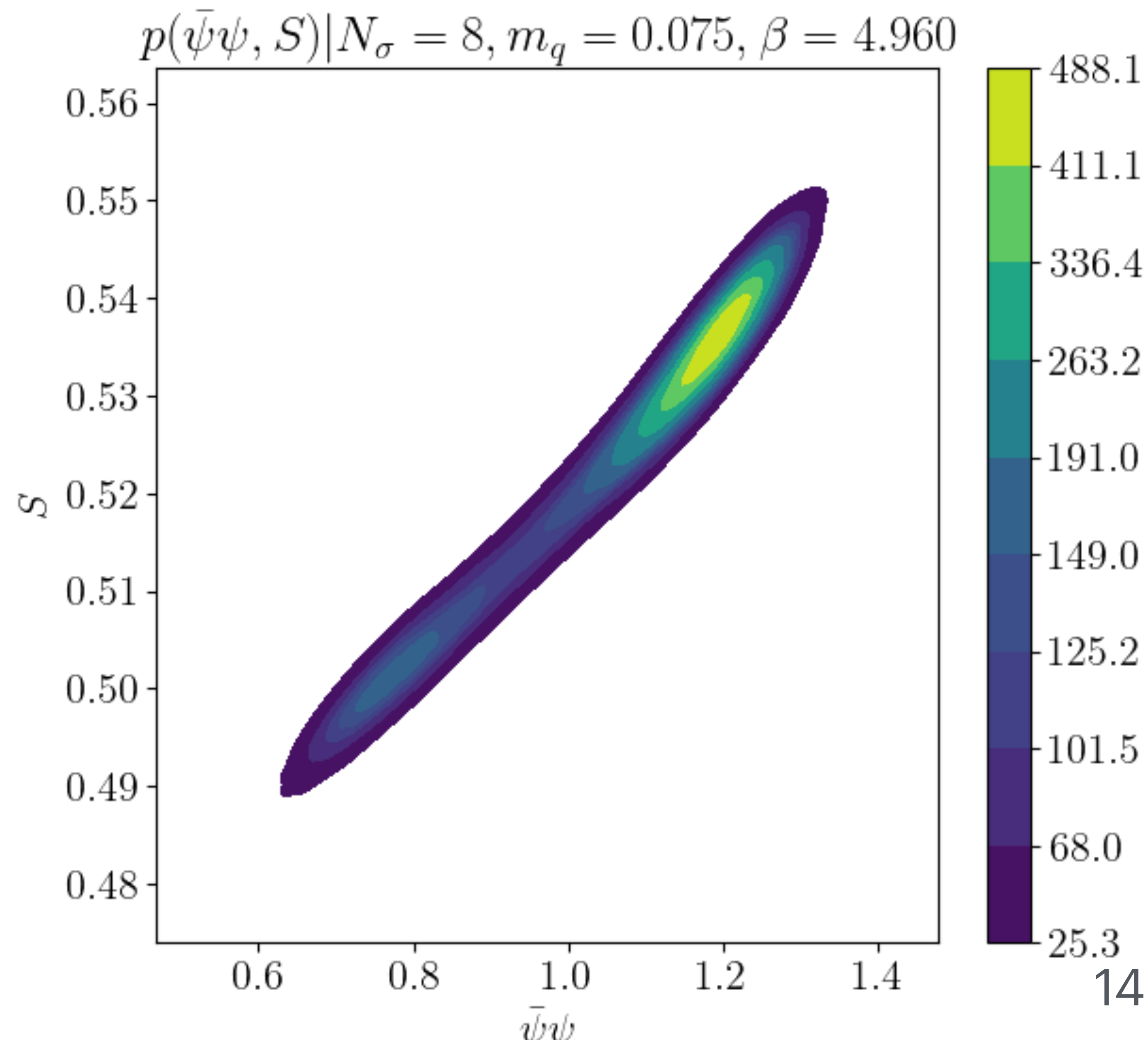
Results for $p\left(\bar{\psi}\psi, S\right)$ for some N_σ, m_l, β



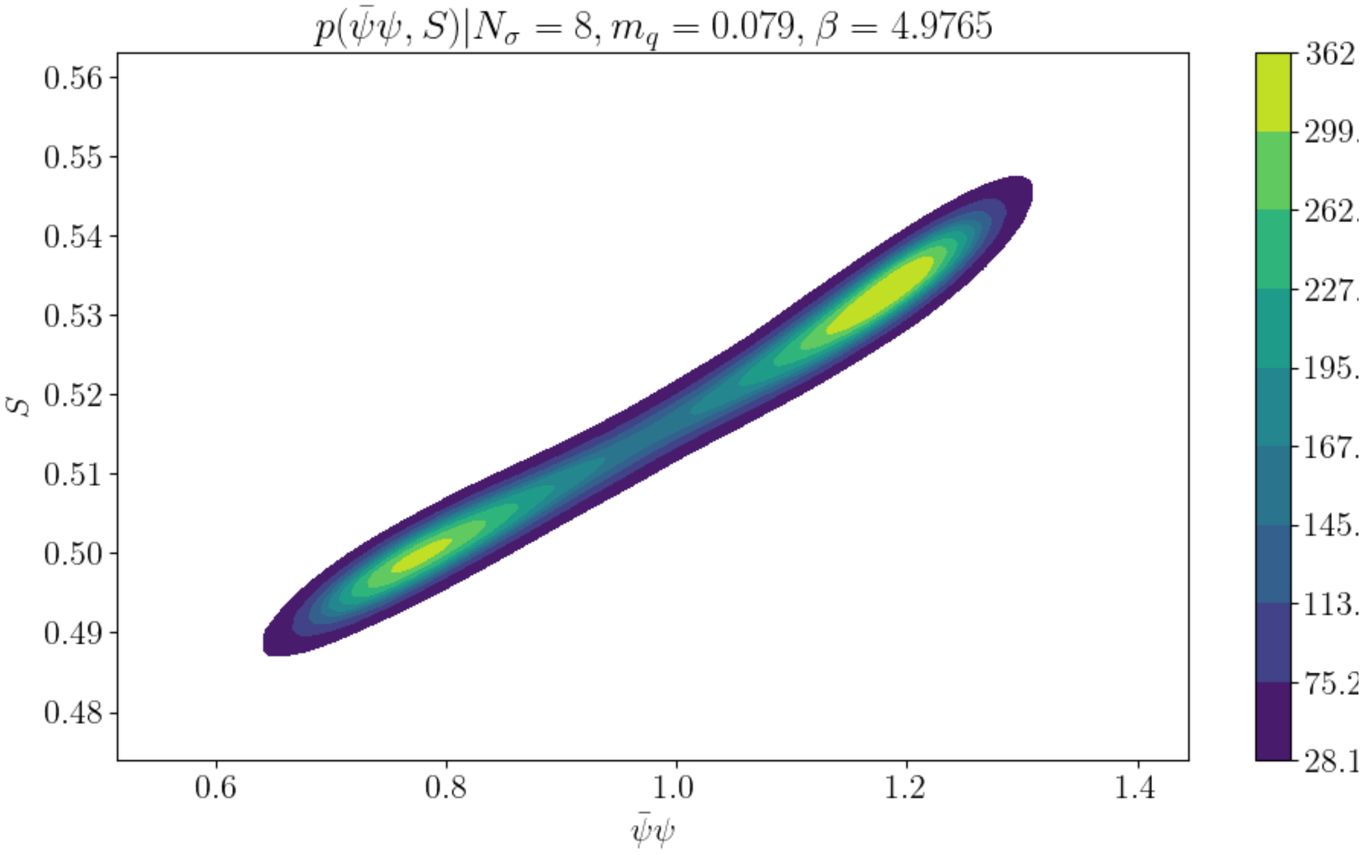
From the data

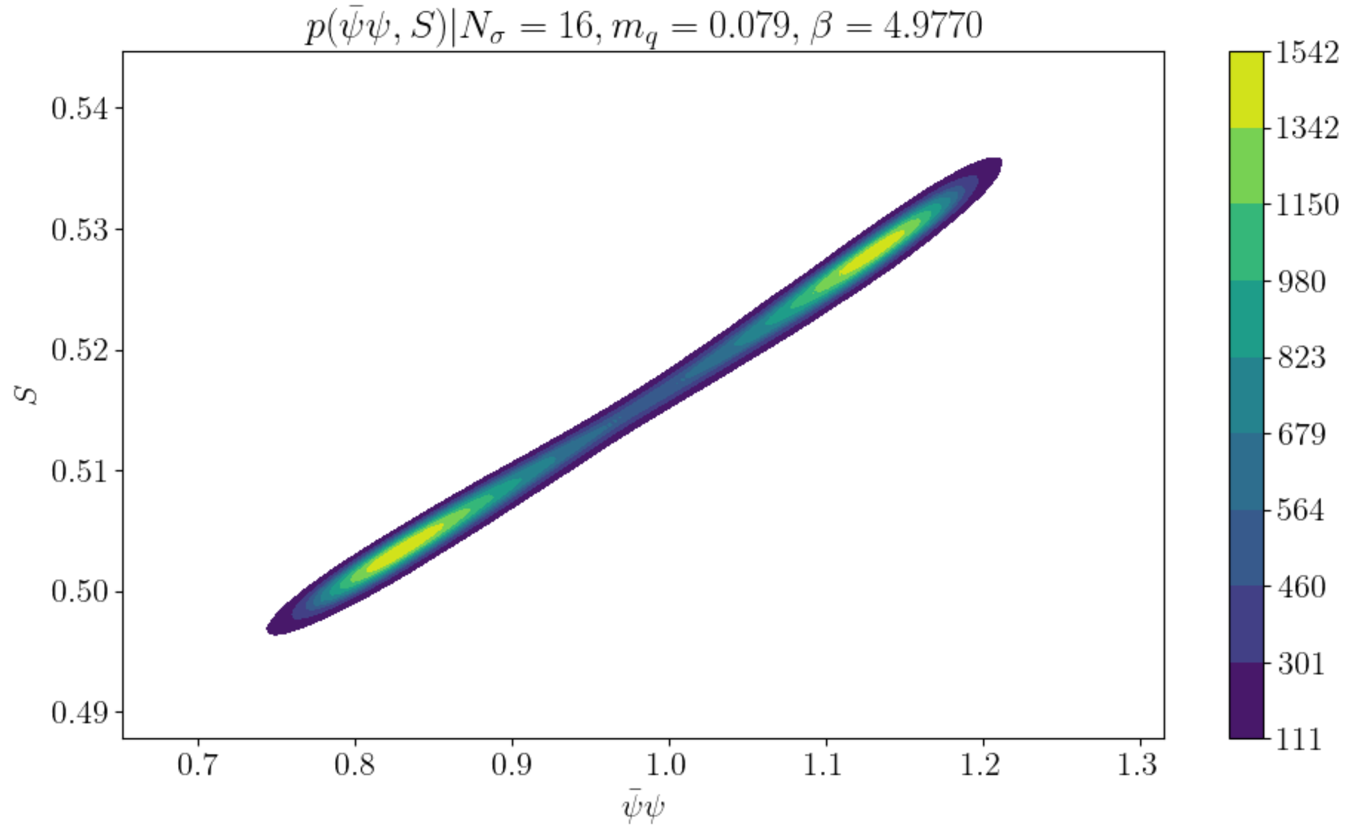
Another indication of learning the correct density

MAF prediction



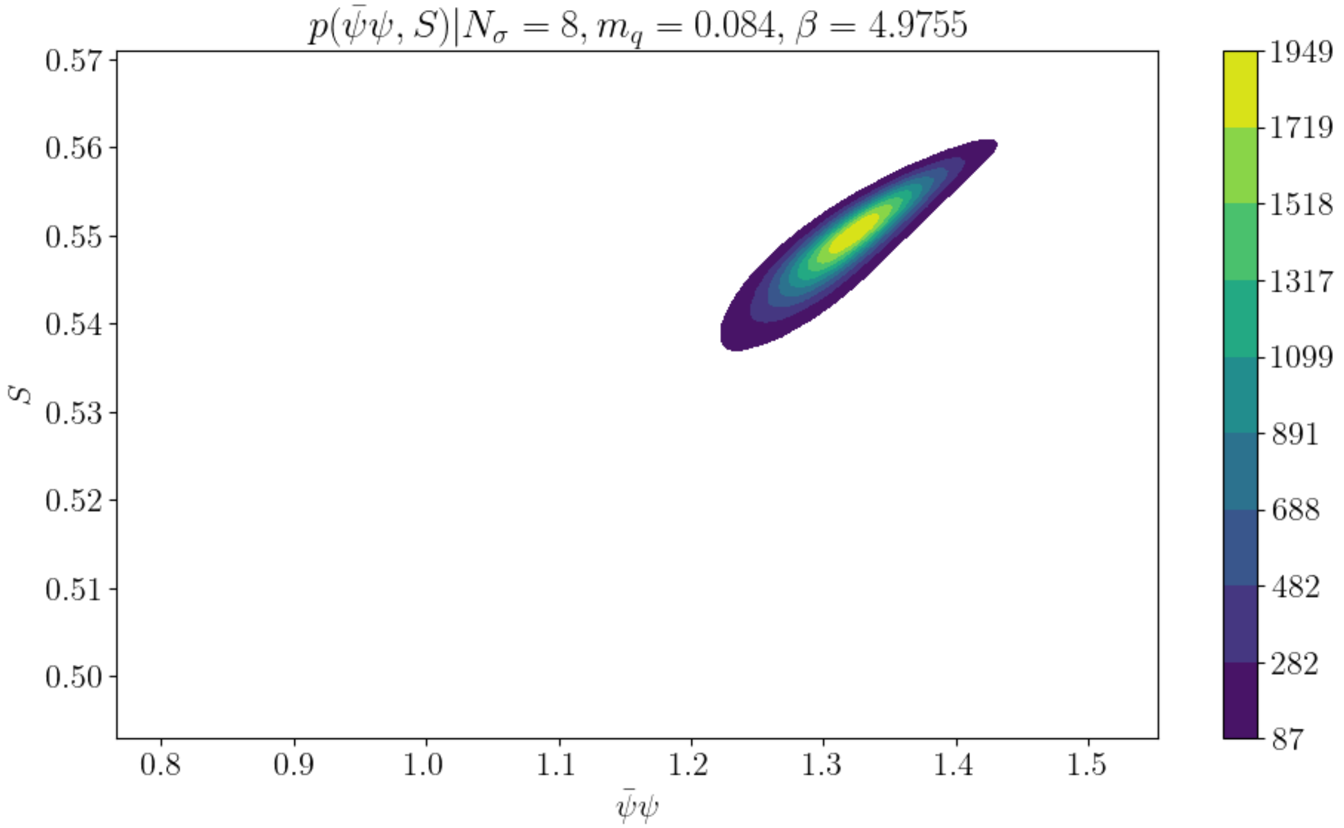
Results for $p\left(\bar{\psi}\psi, S\right)$ for some N_σ, m_l, β





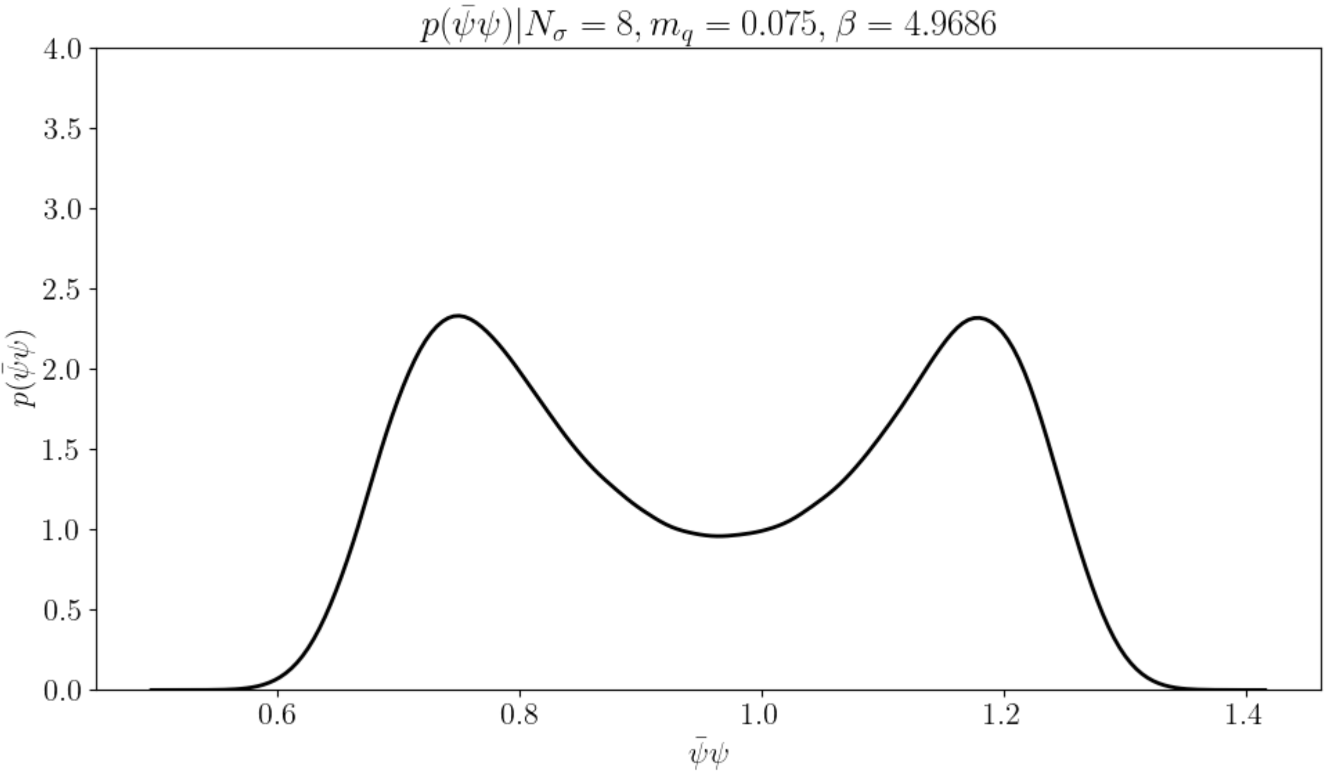
All MAF predictions - recall

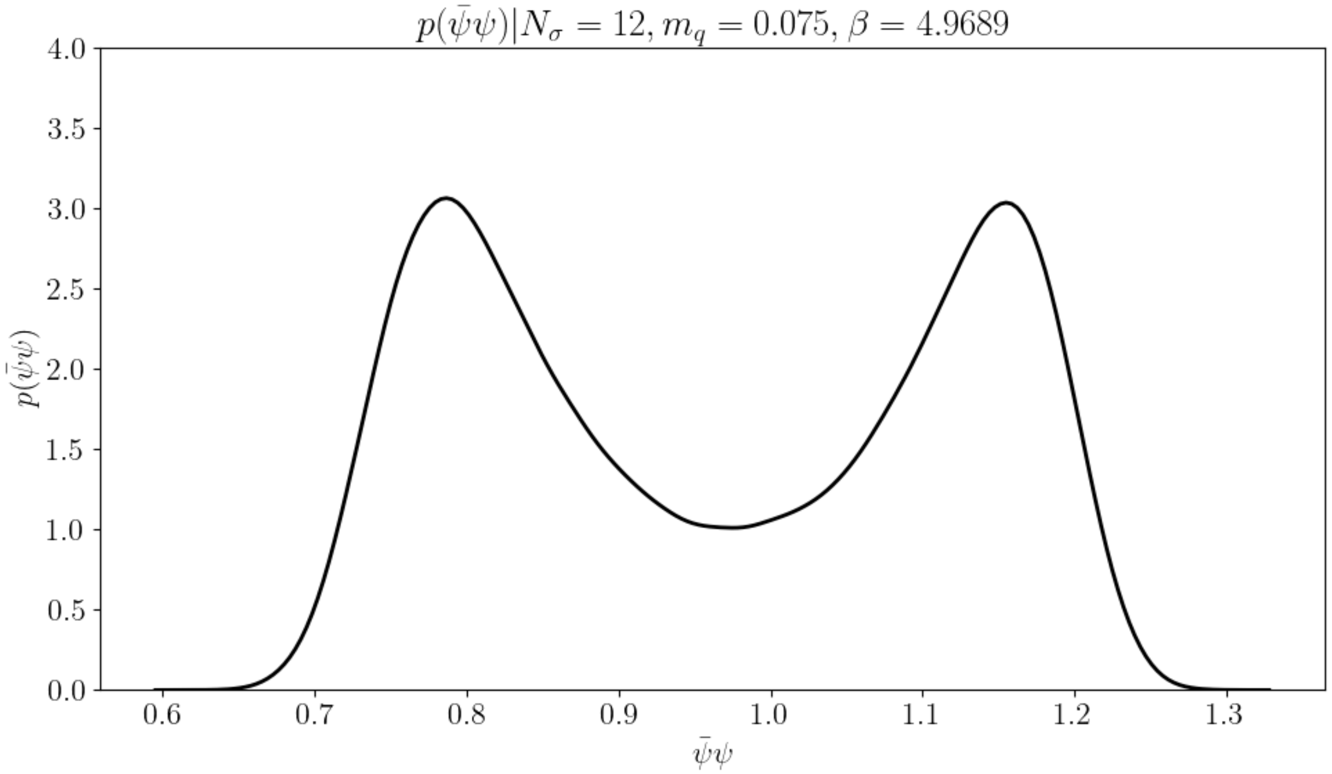
$\beta_c \sim 4.9828$ and $am_c \sim 0.082$



MAF applied to the entire data

- Picture when we should be in the first order region

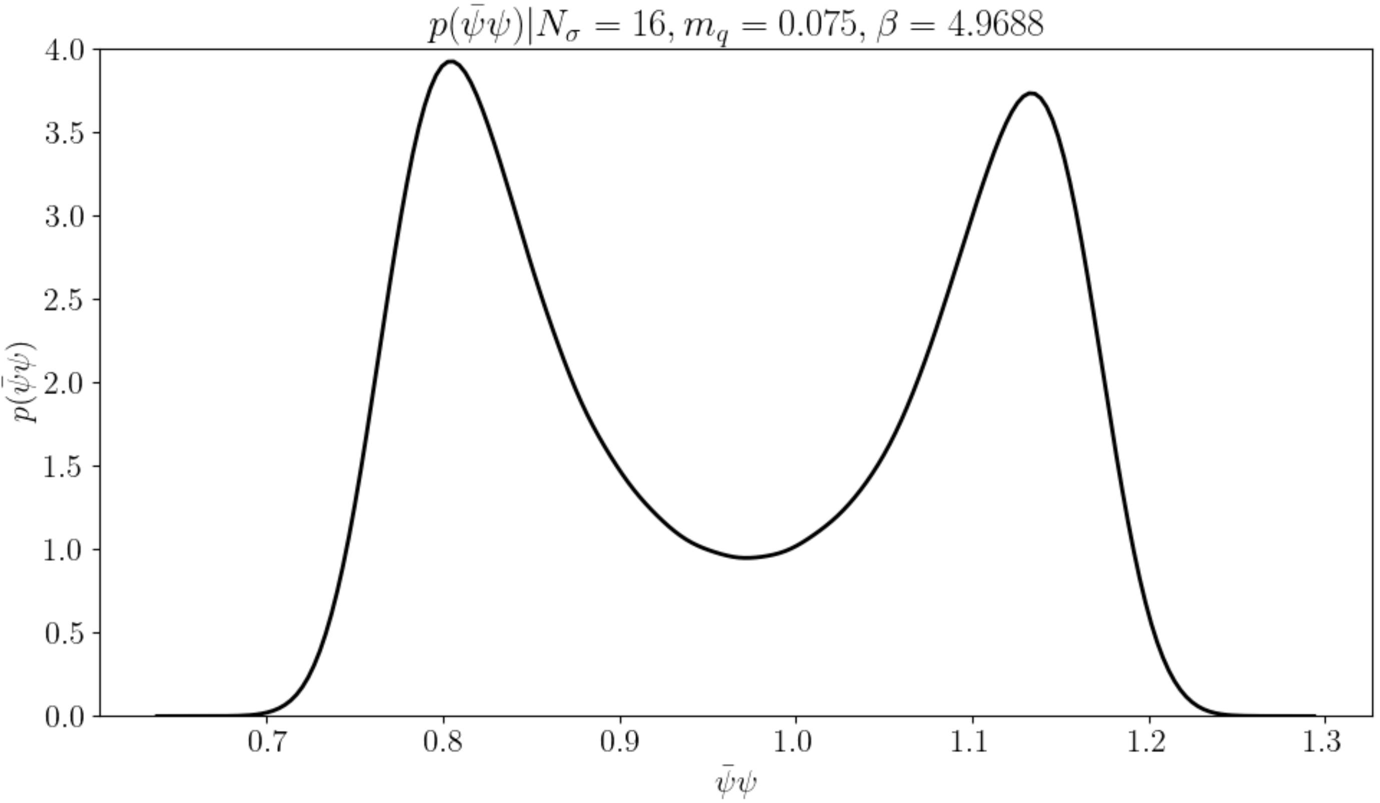




$am_l = 0.075$

recall -

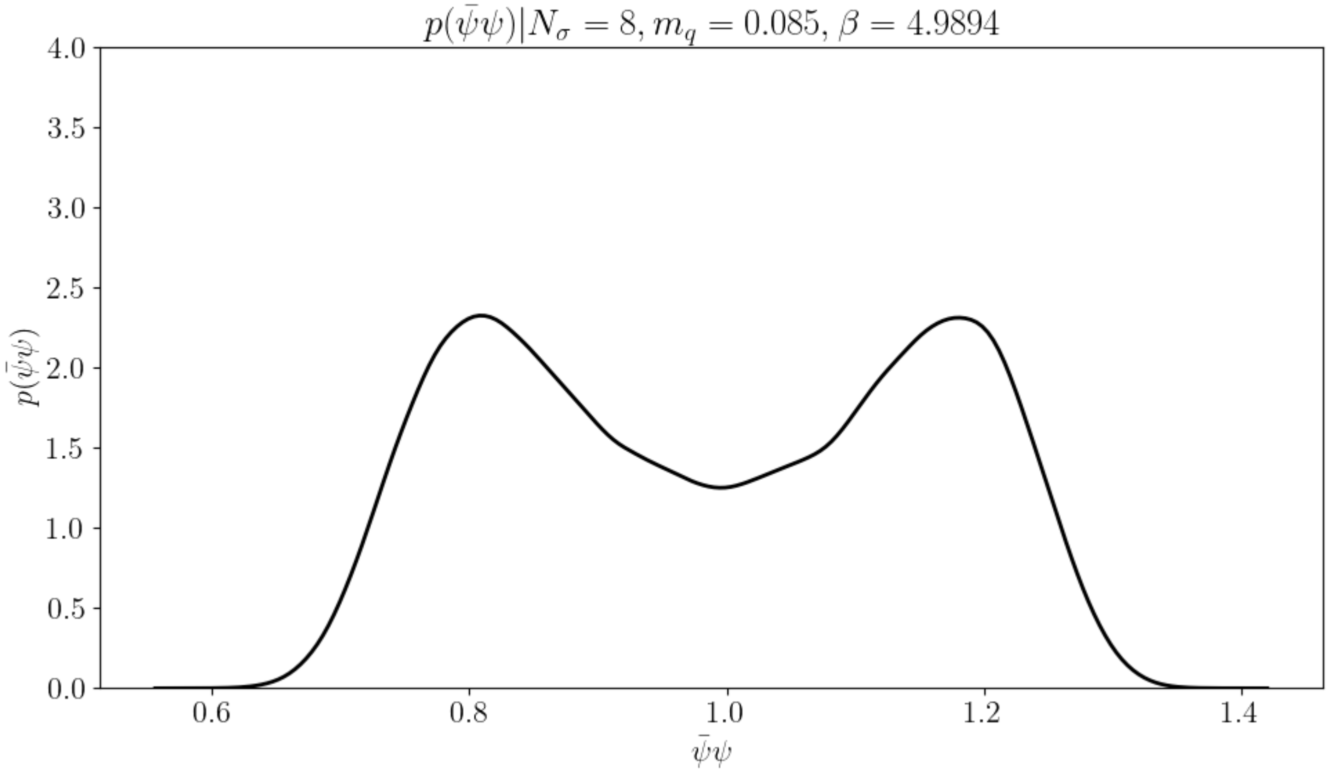
$am_c \sim 0.082$

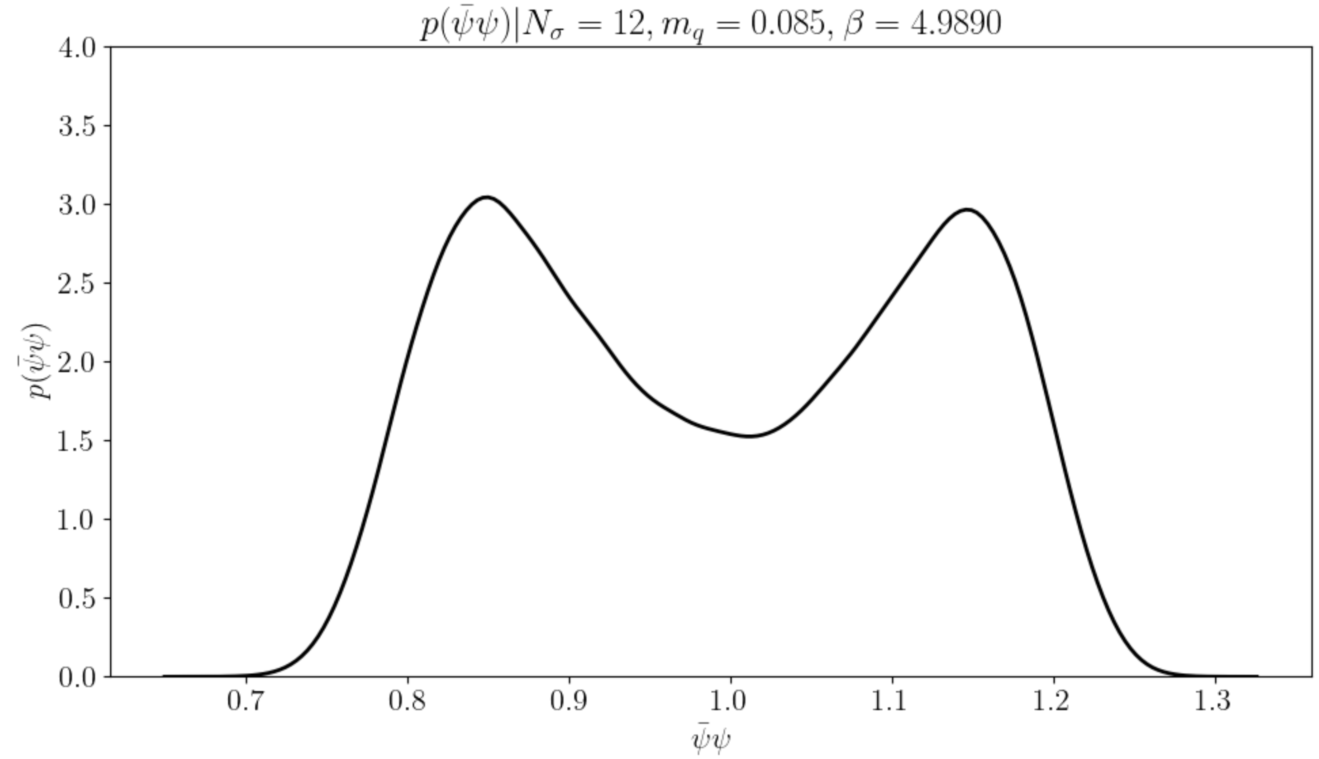


Peaks appear to resolve with increasing volume

MAF applied to the entire data

- Picture when we (should be) in the crossover region

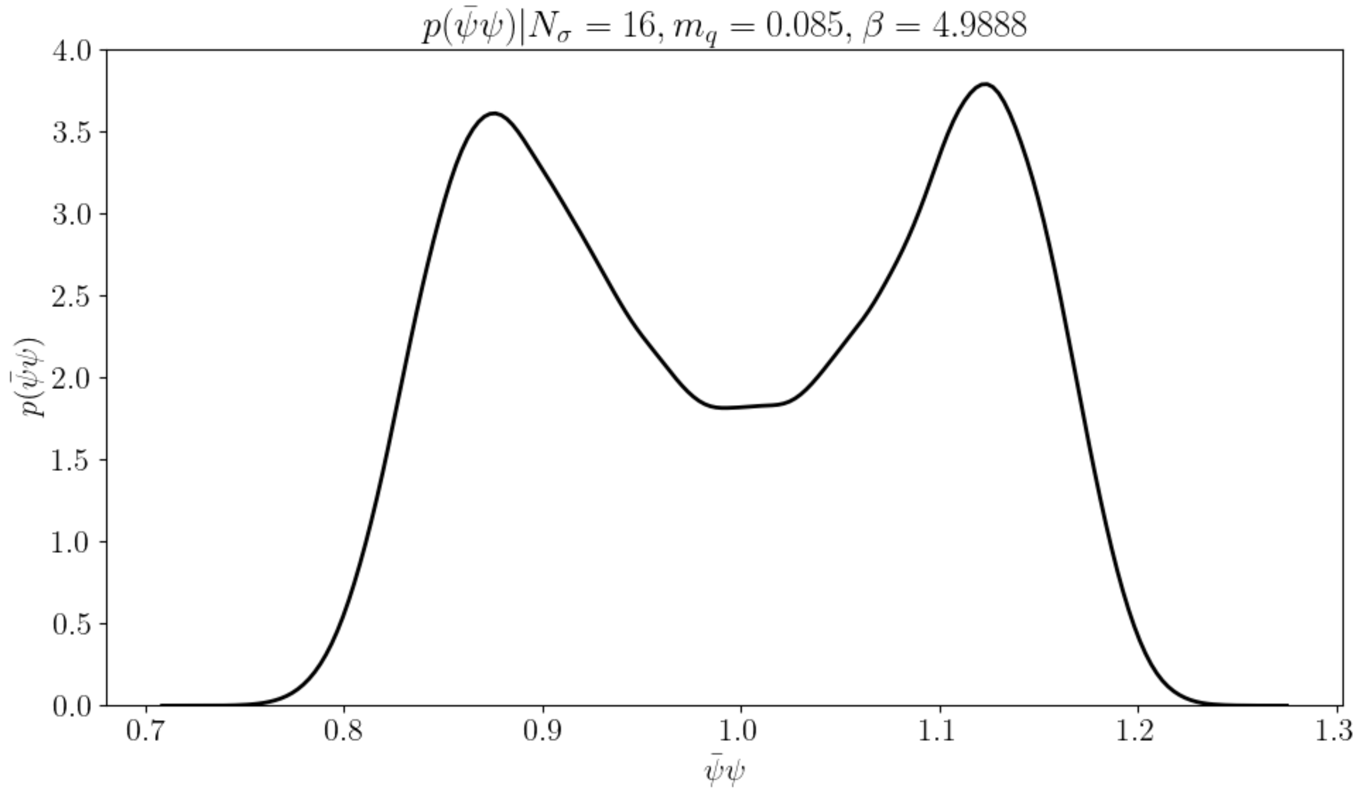




$am_l = 0.085$

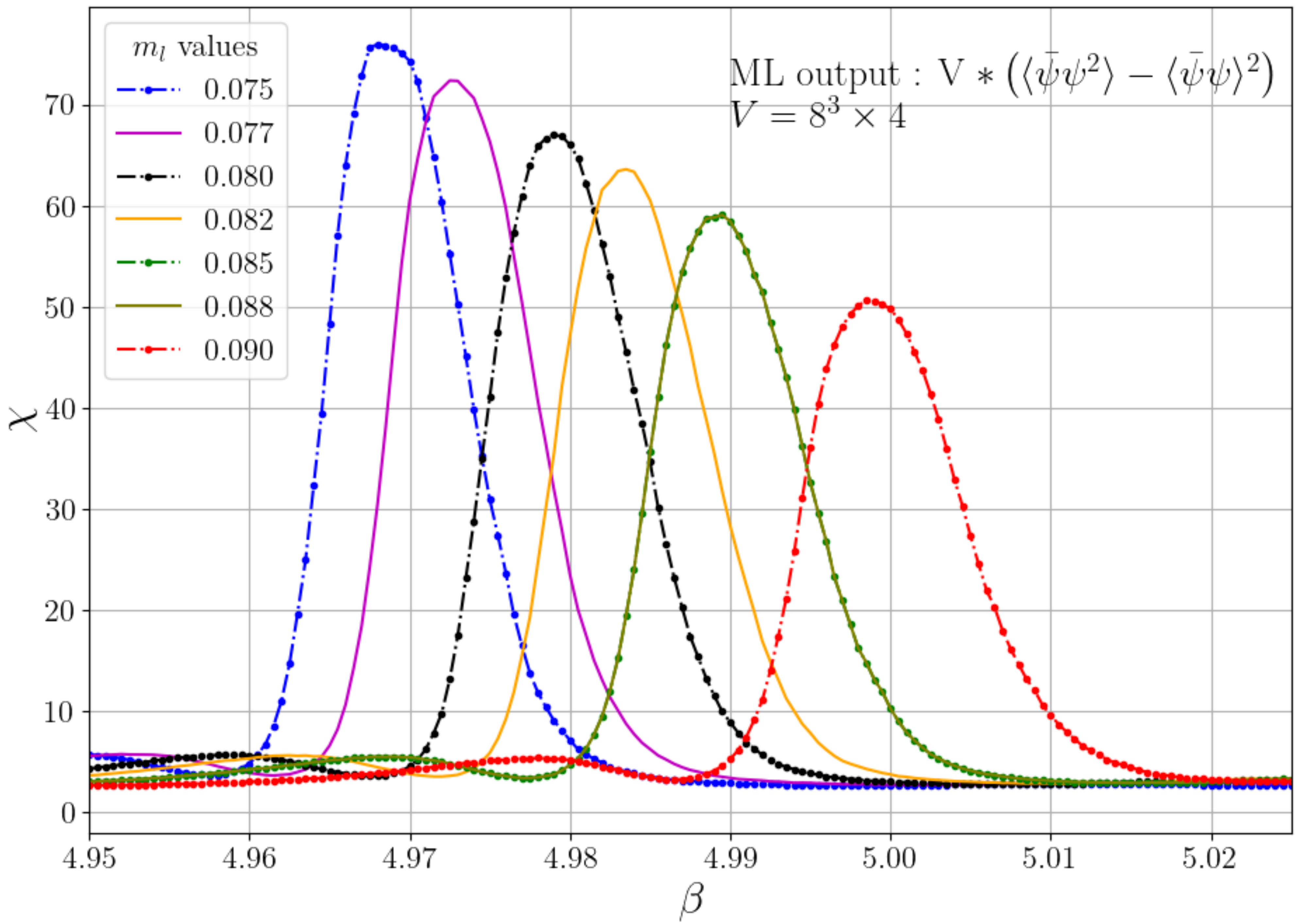
recall -

$am_c \sim 0.082$



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

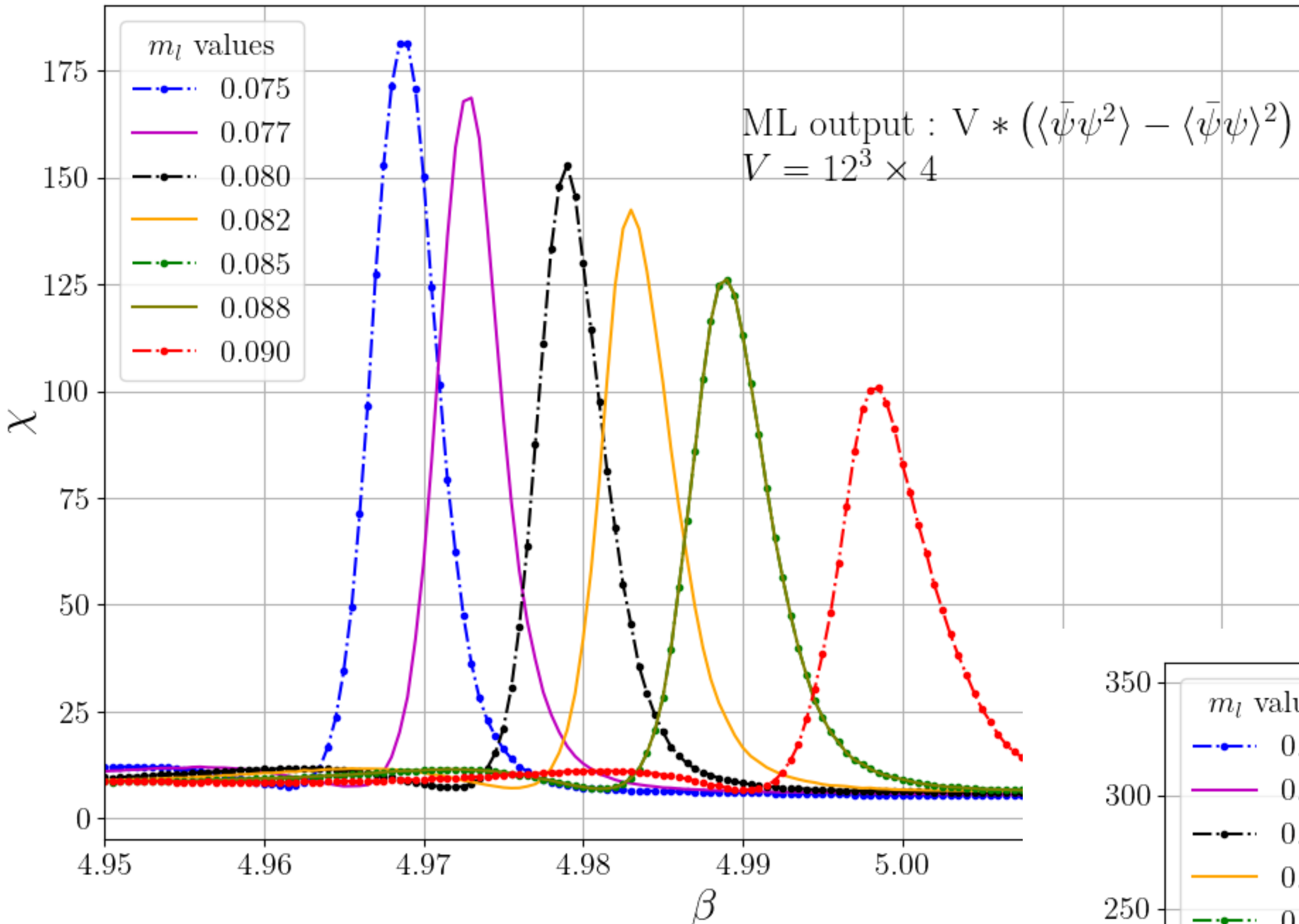
- With $p\left(\bar{\psi}\psi, S \middle| N_\sigma, m_l, \beta\right)$ - we are free to compute higher moments !
- We see scaling of peak height, width, location from ML prediction



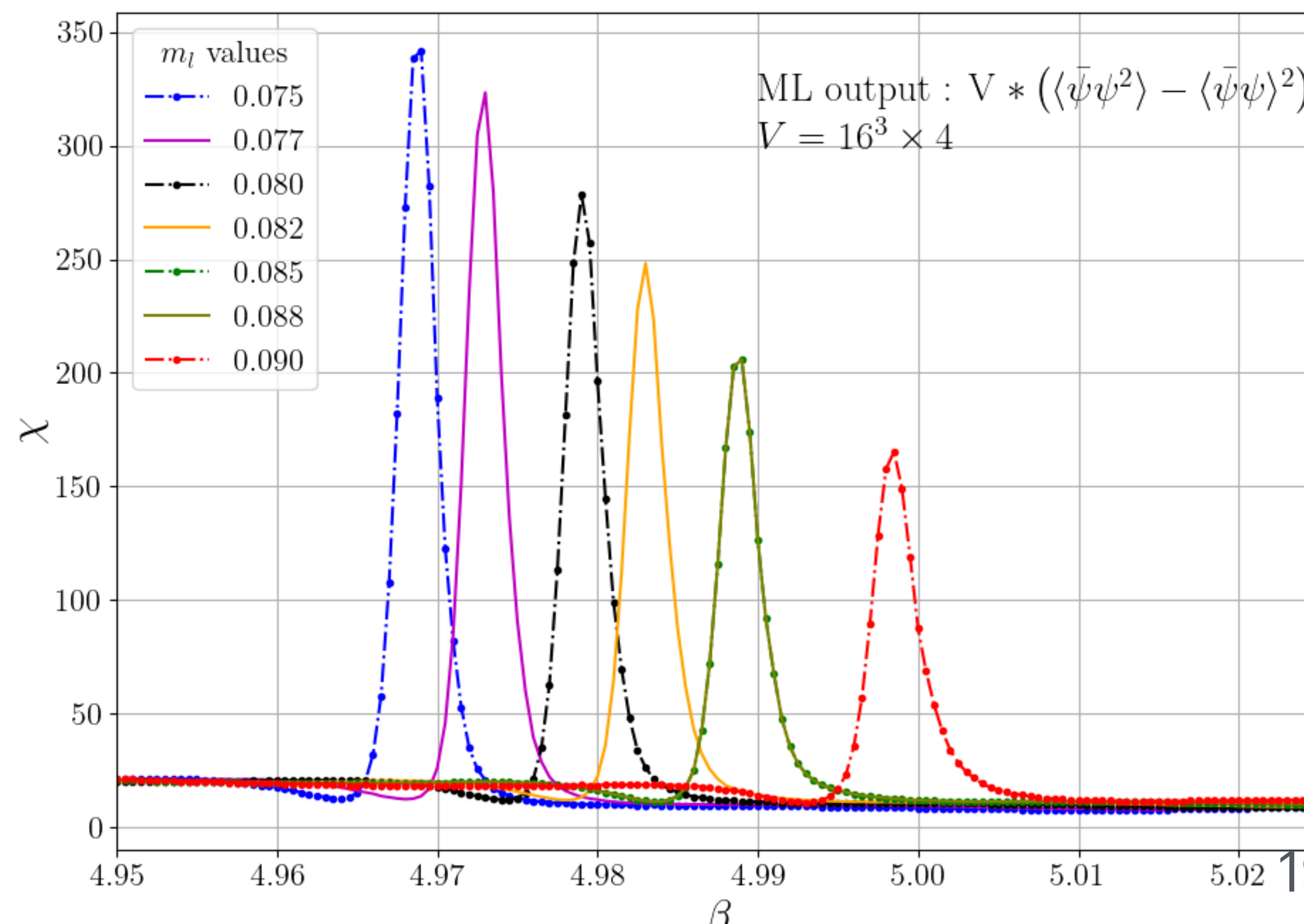
mass values in original data

Interpolated mass values

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



With **increasing lattice volume** and **decreasing bare quark** we see narrowing and shifting of peaks



Pending analysis with of fitting peak positions to obtain β_c and then computing $B_4(\beta_c, am_l, N_\sigma)$ to find the $Z2$ boundary …

Summary & Outlook

- The MAF model tuned for the HISQ data seems to work ``out of the box'' when applied to Frankfurt data
- Results on interpolation appear consistent with actual lattice data - at least at the level of the chiral condensate
- Model evaluations are cheap ~ 0.5 seconds for 1M evaluations
- One fit for all data - but with 7333 trainable parameters
- Yet to determine critical mass and gauge coupling in agreement with F. Cuteri et.al., *JHEP* 11 (2021)
- Plan I : Study the systematics of ML model - may not always converge to the same fit parameters - some kind of bootstrap needed ?
- Plan II : is to also include data for different N_τ to be able to interpolate in that direction - reproduce the tri-critical scaling

Backup slides

Some numbers and parameters

Some Training Statistics

- With $N_\sigma = 12$ removed - on 1 GPU with Approximately 30.8 GB of GPU memory used at peak with training time ~ 4hr 30 mins
- With all data - on 1 GPU with Approximately 32.9 GB of GPU memory used at peak with training time ~ 5hr 20 mins

Some Evaluation Statistics

- Time for 1 M for each β, m_l, N_σ evaluation : ~ 0.45 seconds

Model parameters used from Marius Neumann's Thesis

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
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 $(S, \bar{\psi}\psi)$ ~ 4.3 M
optimizer	Adam

MAF Inference on probability