UNCERTAINTY QUANTIFICATION IN DEEP LEARNING

TOWARDS DETECTOR PHYSICS

EICUG 2ND DETECTOR MEETING
CENTER FOR FRONTIERS IN NUCLEAR SCIENCE
STONY BROOK UNIVERSITY
8 DECEMBER 2022

MICHELLE KUCHERA
DAVIDSON COLLEGE


Hot and Dense Nuclear Matter
Hadrons
u
d
d
Atomic Nucleus
Nuceli in the Cosmos
Fundamental Interactions
u c t γ H
d s b g
ν_e ν_μ ν_τ Z^0
e μ τ $W^\pm$
NUCLEAR THEORY
• Correlations and predictions
• Estimations and causations
ARTIFICIAL INTELLIGENCE
MACHINE LEARNING
Experimental Design
NUCLEAR EXPERIMENT
• Methods
• Tools
ACCELERATOR SCIENCE AND OPERATIONS
NUCLEAR DATA
• Databases
• Data Mining
• Visualization
DISCOVERY
APPLICATIONS
REVIEWS OF MODERN PHYSICS
Recent Accepted Authors Referees Search Press About Editorial Team
Access by
Colloquium: Machine learning in nuclear physics
Amber Boehnlein, Markus Diefenthaler, Nobuo Sato, Malachi Schram, Veronique Ziegler, Cristiano Fanelli, Morten Hjorth-Jensen, Tanja Horn, Michelle P. Kuchera, Dean Lee, Witold Nazarewicz, Peter Ostroumov, Kostas Orginos, Alan Poon, Xin-Nian Wang, Alexander Scheinker, Michael S. Smith, and Long-Gang Pang
Rev. Mod. Phys. **94**, 031003 – Published 8 September 2022

CONTENT

Call for controlled research into UQ for our purposes

- Epistemic, aleatoric
- Deeply Uncertain

Methods

- Bayesian Neural Networks, Implicit Quantile Networks, Normalizing Flows, Generative Adversarial Networks

Examples from particle physics

- jet simulation, likelihood inference

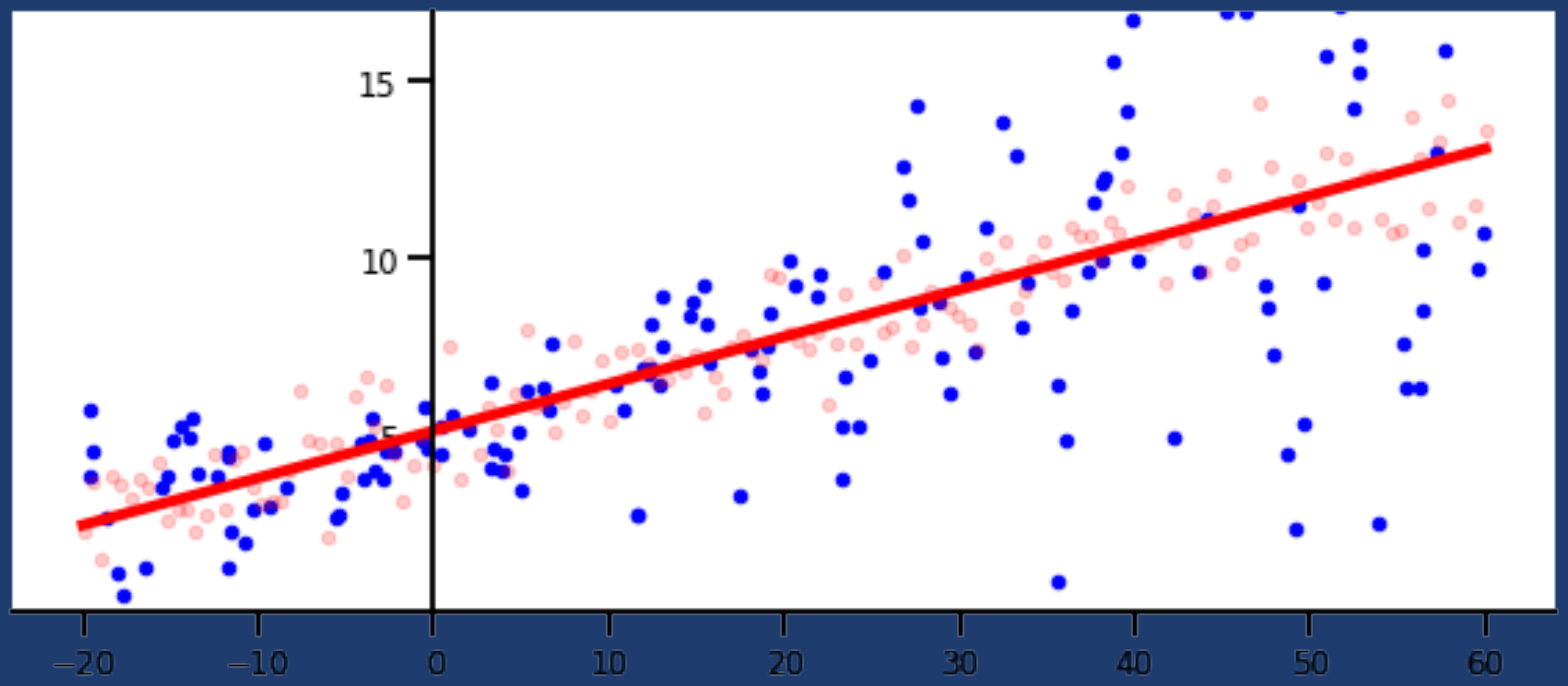
Epistemic:

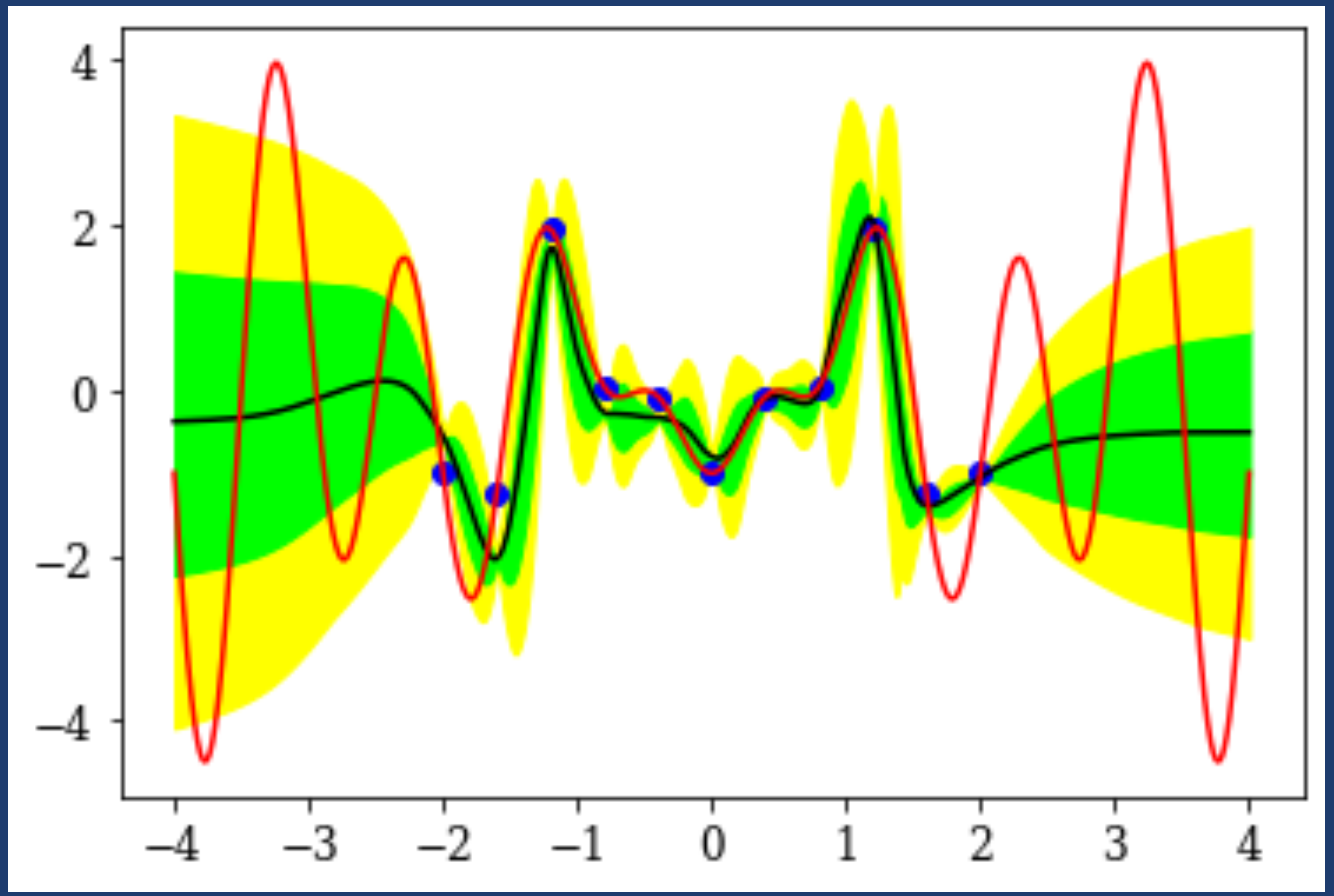
- model uncertainty
- reducible

Aleatoric:

- uncertainty in data
- irreducible

- statistical
- systematic





MACHINE LEARNING
Science and Technology

PAPER

Deeply uncertain: comparing methods of uncertainty quantification in deep learning algorithms

João Caldeira[1] and Brian Nord[1,2,3]

[1] Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510, United States of America
[2] Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, United States of America
[3] Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, United States of America

E-mail: caldeira@fnal.gov

Keywords: uncertainty quantification, neural networks, Bayesian inference, ensemble of neural networks

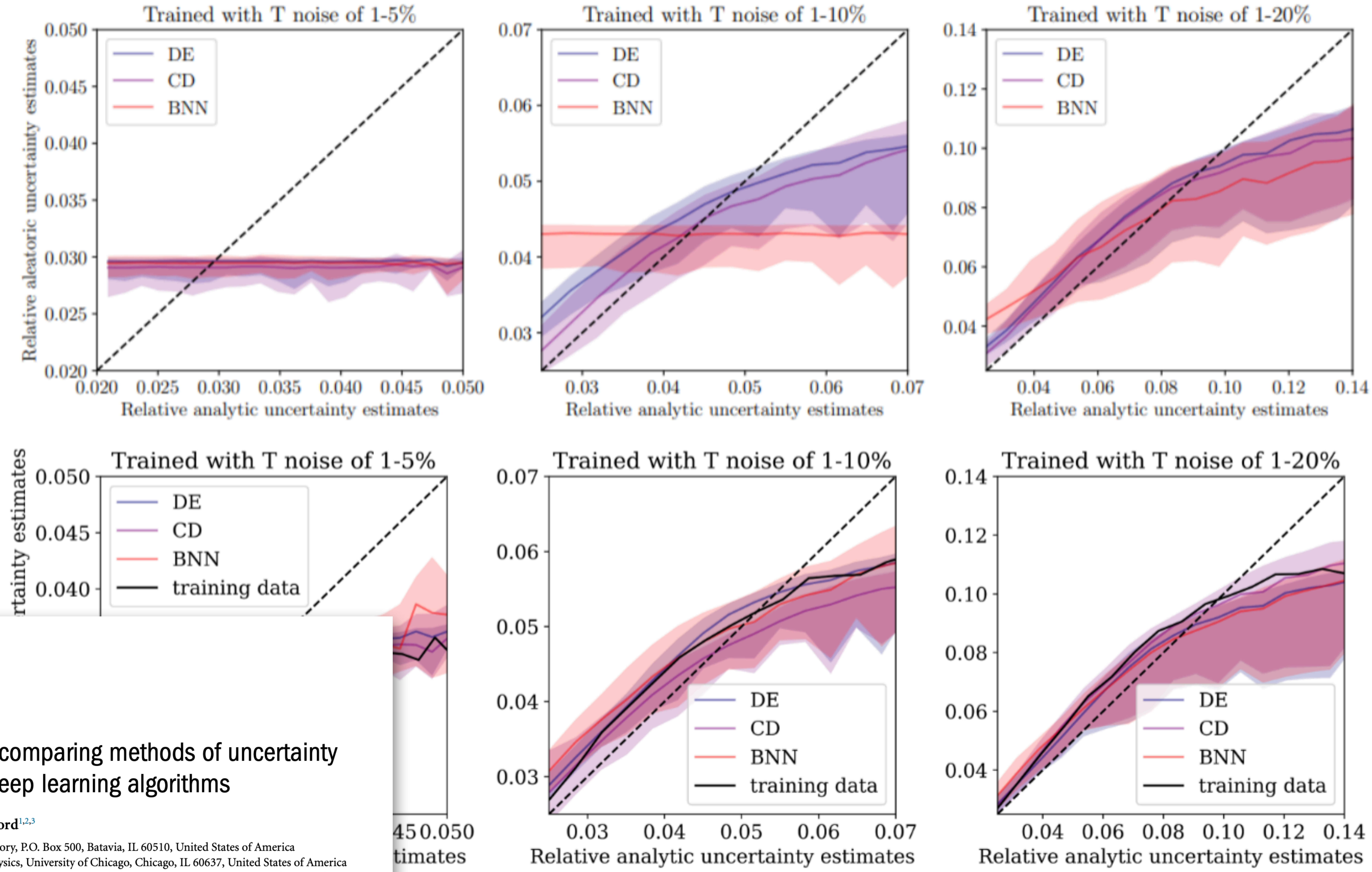
- *Aleatoric statistical* uncertainty can be included by adding noise in the 10 measurements of the period, T. For each data point in the training set, we draw the amount of measurement noise ν uniformly in some range, and then draw each measurement of the period from a normal distribution with standard deviation νT. The choice of the range for ν in the training set merits a longer discussion in section 3.
- *Aleatoric systematic* uncertainty exists if the single measurement of L also contains noise, as this is a source of uncertainty that cannot be statistically determined from the single measurement of L. Note that since there is no statistical way to determine this noise from the input data alone, the uncertainty must be determined from the typical noise seen in training. In our training and test sets, all measurements of L are drawn from a normal distribution with standard deviation $0.02L$.
- *Epistemic systematic* uncertainty reflects how uncertain the model is of its predictions. One way to test this is by looking at predictions far from the training set manifold. In this experiment, we train networks with $g \in (5, 15)$ m/s^2, and $L \in (0.2, 0.8)$ m. Either of these can be moved outside that range, and we will consider both cases below.

$$\hat{g} = \frac{1}{N}\sum_{i=1}^{N}\mu_i = \mathrm{mean}(\mu_i) \qquad \text{(gravitational constant mean)}$$

$$\sigma_{al} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2} = \sqrt{\mathrm{mean}(\sigma_i^2)} \qquad \text{(aleatoric uncertainty)}$$

$$\sigma_{ep} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\mu_i^2 - \hat{g}^2} = \mathrm{stdev}(\mu_i) \qquad \text{(epistemic uncertainty)}$$

$$\sigma_{pr} = \sqrt{\sigma_{al}^2 + \sigma_{ep}^2} \qquad \text{(total predictive uncertainty)}$$



MACHINE LEARNING
Science and Technology

PAPER

Deeply uncertain: comparing methods of uncertainty quantification in deep learning algorithms

João Caldeira[1] and Brian Nord[1,2,3]

1 Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510, United States of America
2 Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, United States of America
3 Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, United States of America

E-mail: caldeira@fnal.gov

Keywords: uncertainty quantification, neural networks, Bayesian inference, ensemble of neural networks

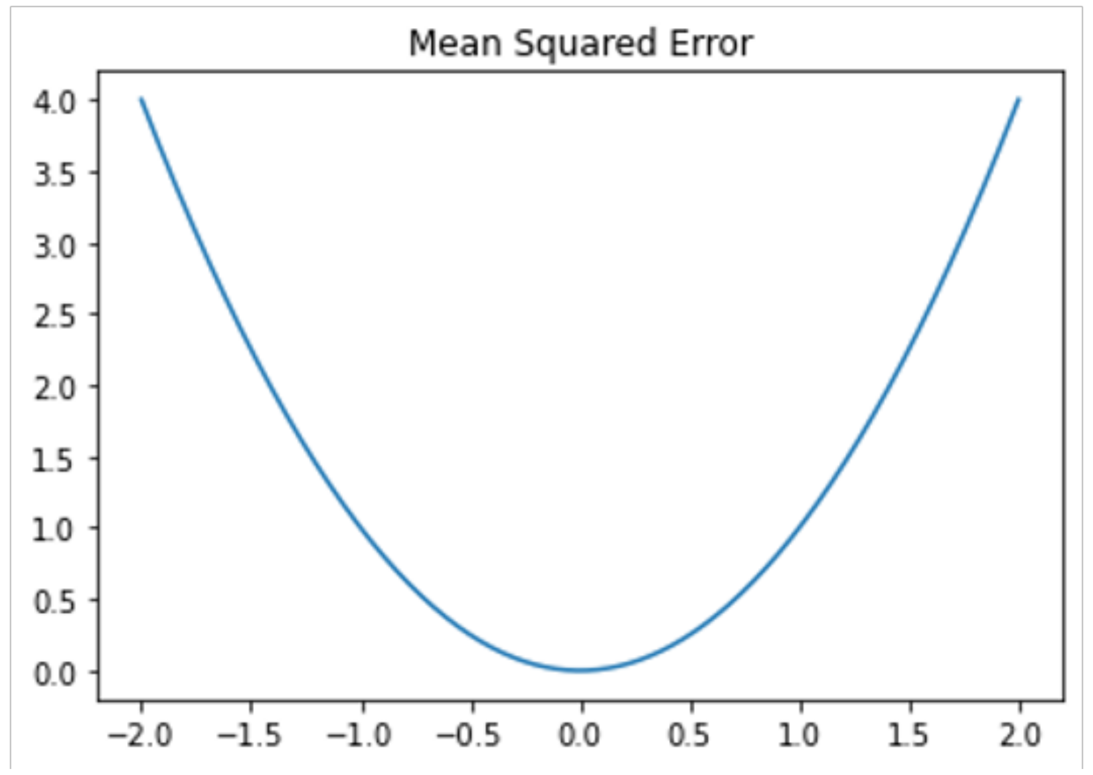
PLOTS COURTESY B. KRONHEIM, UMD

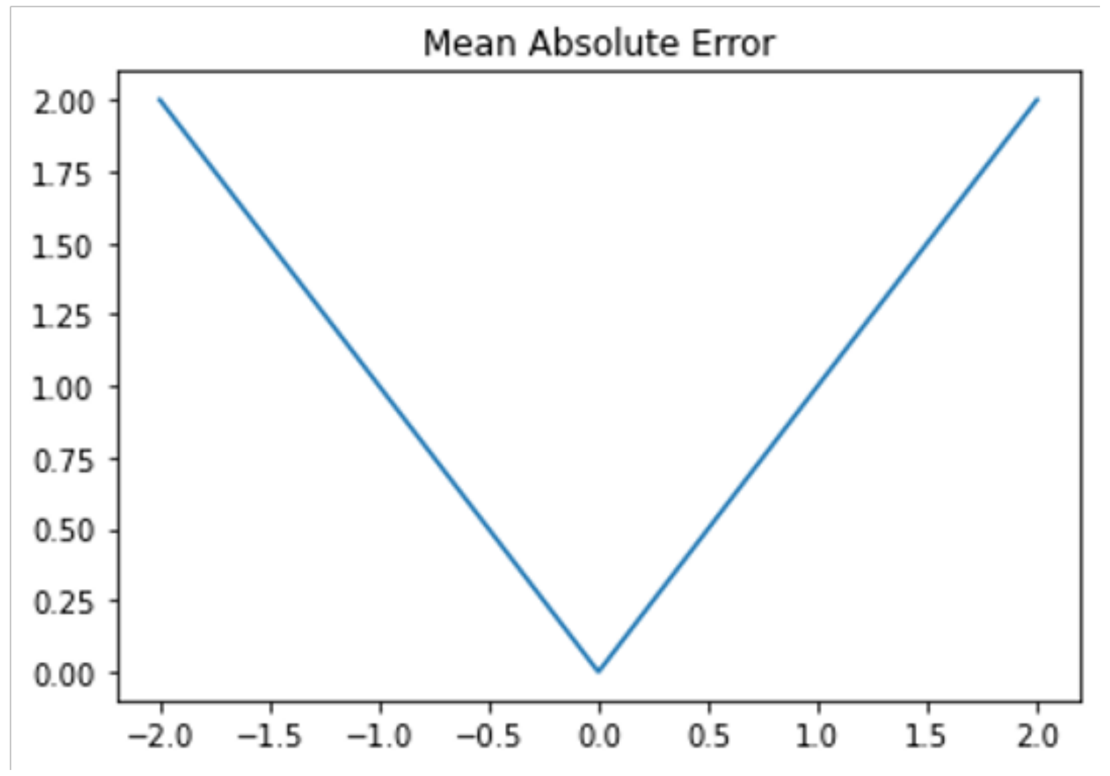
CONSIDERATIONS:

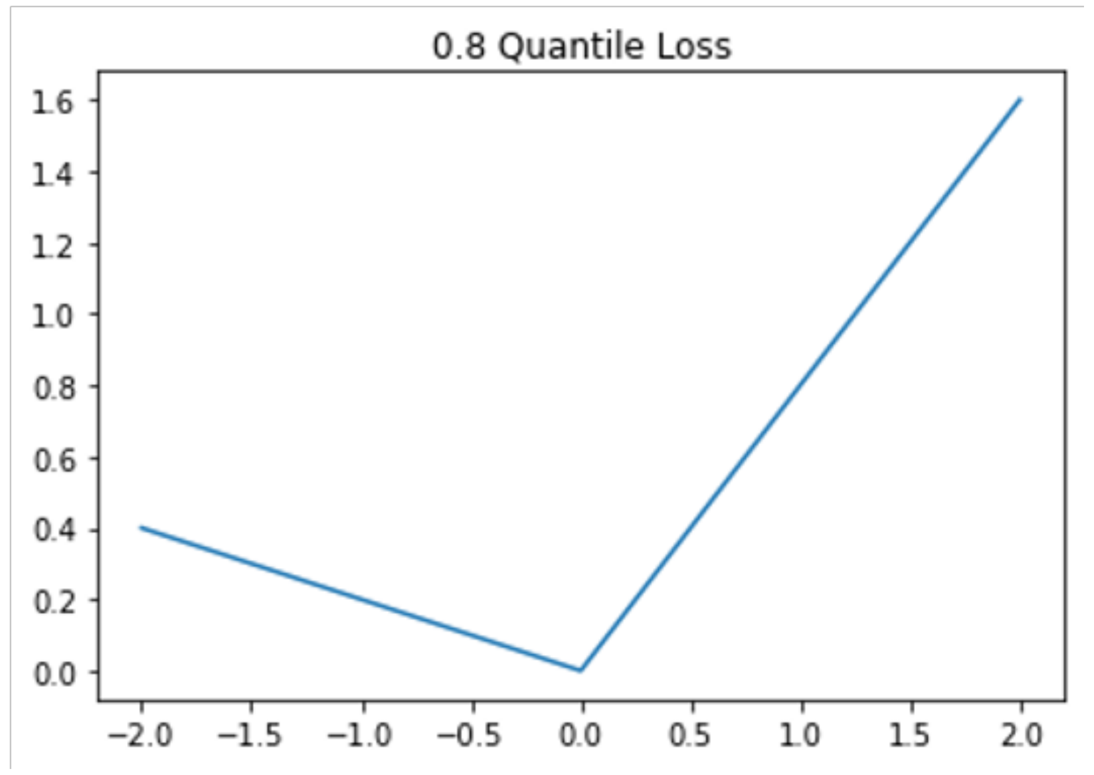
- LOSS FUNCTIONS
- TRAINING APPROACH / METHODS

Loss functions

- $\frac{1}{N}\sum_{i=1}^{N}(\hat{y}-y)^2$: mean
- $\frac{1}{N}\sum_{i=1}^{N}|\hat{y}-y|$: median
- $-\log p(y|\theta)$: maximum likelihood estimation
- $\mathcal{L}(f,x,y,\tau)=\begin{cases}\tau(y-f(x,\tau)) & y\geq f(x,\tau)\\(\tau-1)(y-f(x,\tau)) & y<f(x,\tau)\end{cases}$: quantile function





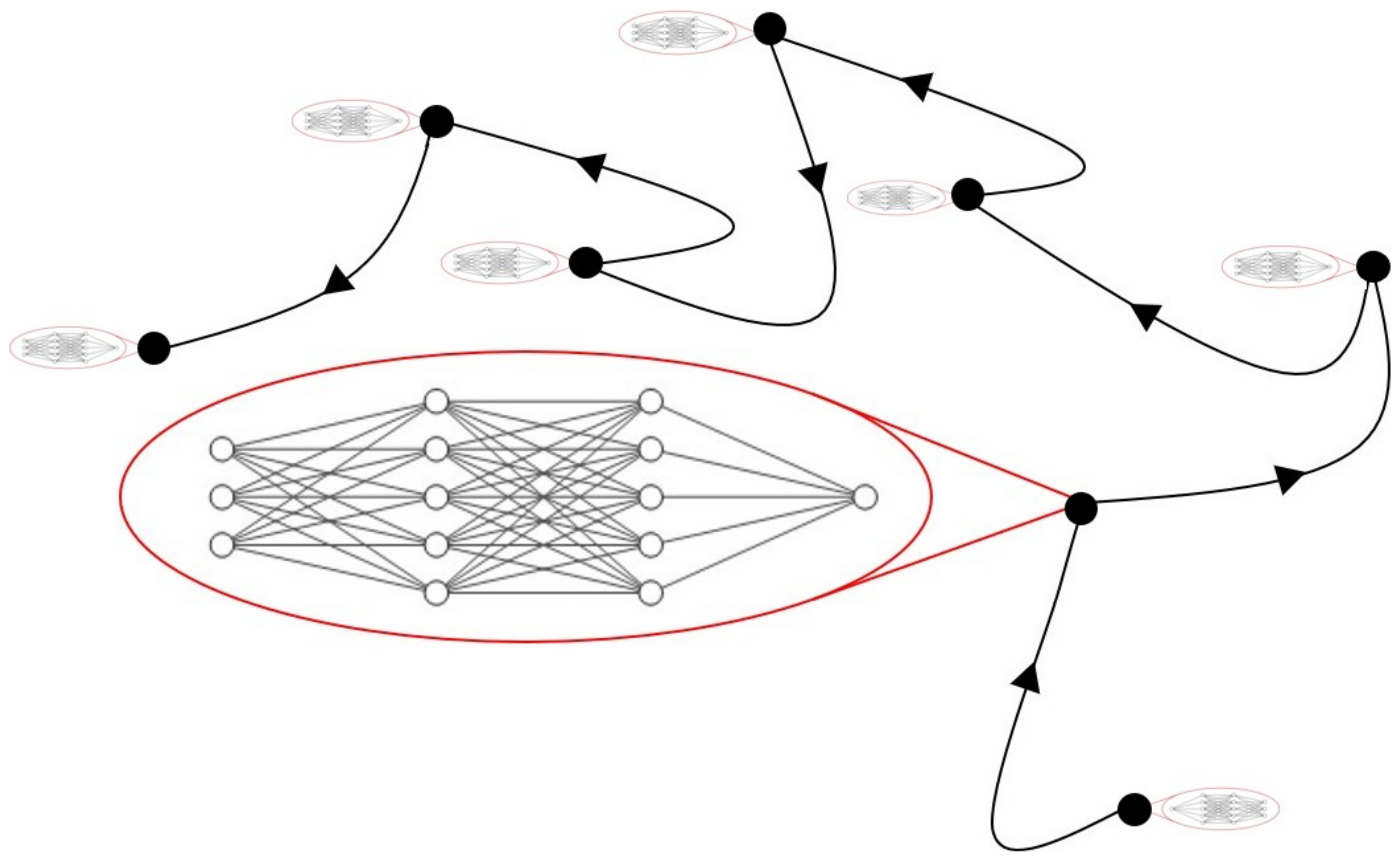


FAST MAPPING FROM THEORY TO OBSERVABLES

Bayesian Neural Networks

$$\log p(y \,|\, \theta)$$

Training — Bayesian inference



Can we make predictions with useful epistemic uncertainty estimates?

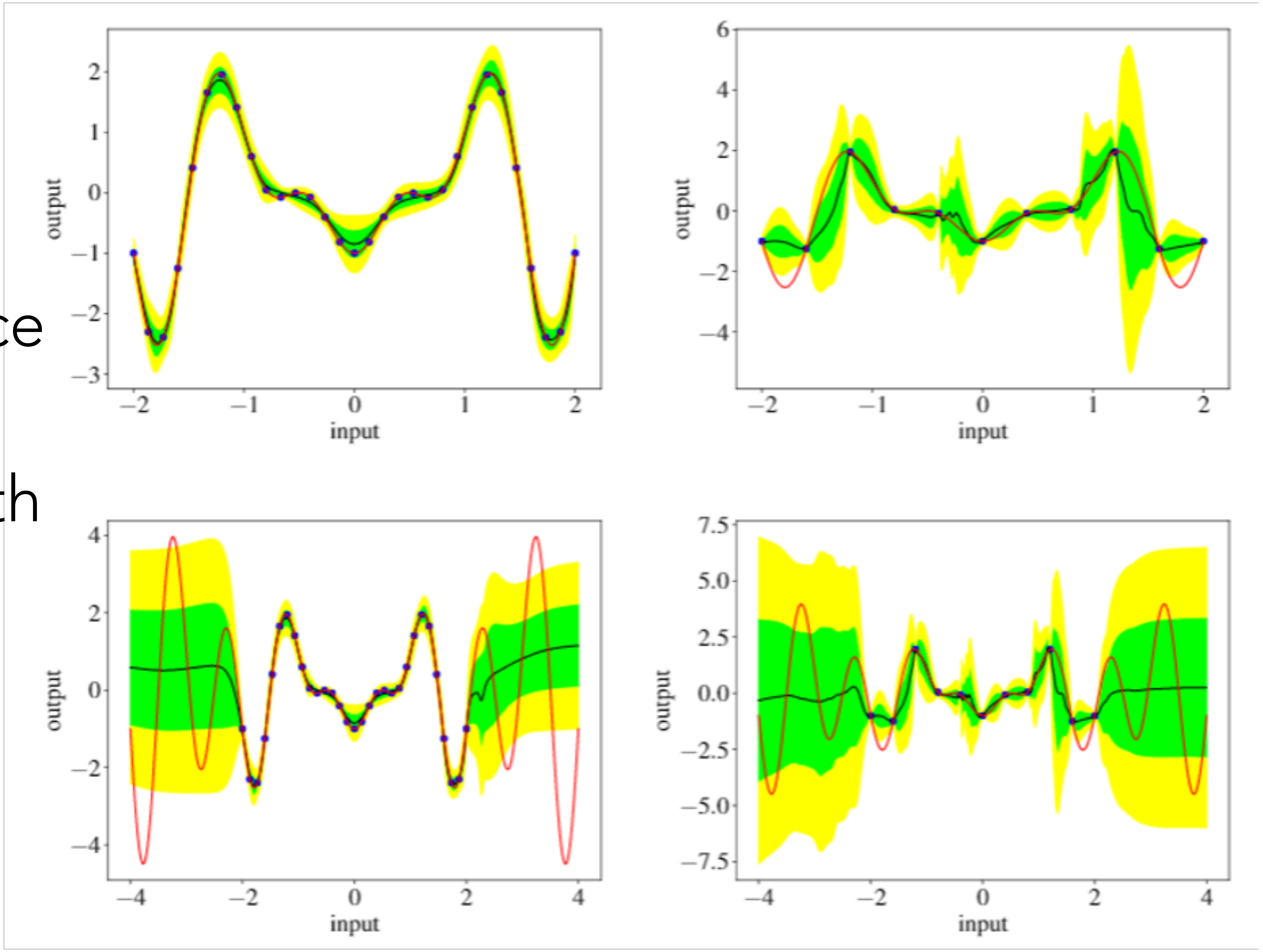
$$p(y \,|\, x, D) = \int \delta(y - f(x, \theta)) \, p(\theta \,|\, D) \, d\theta.$$

FAST MAPPING FROM THEORY TO OBSERVABLES

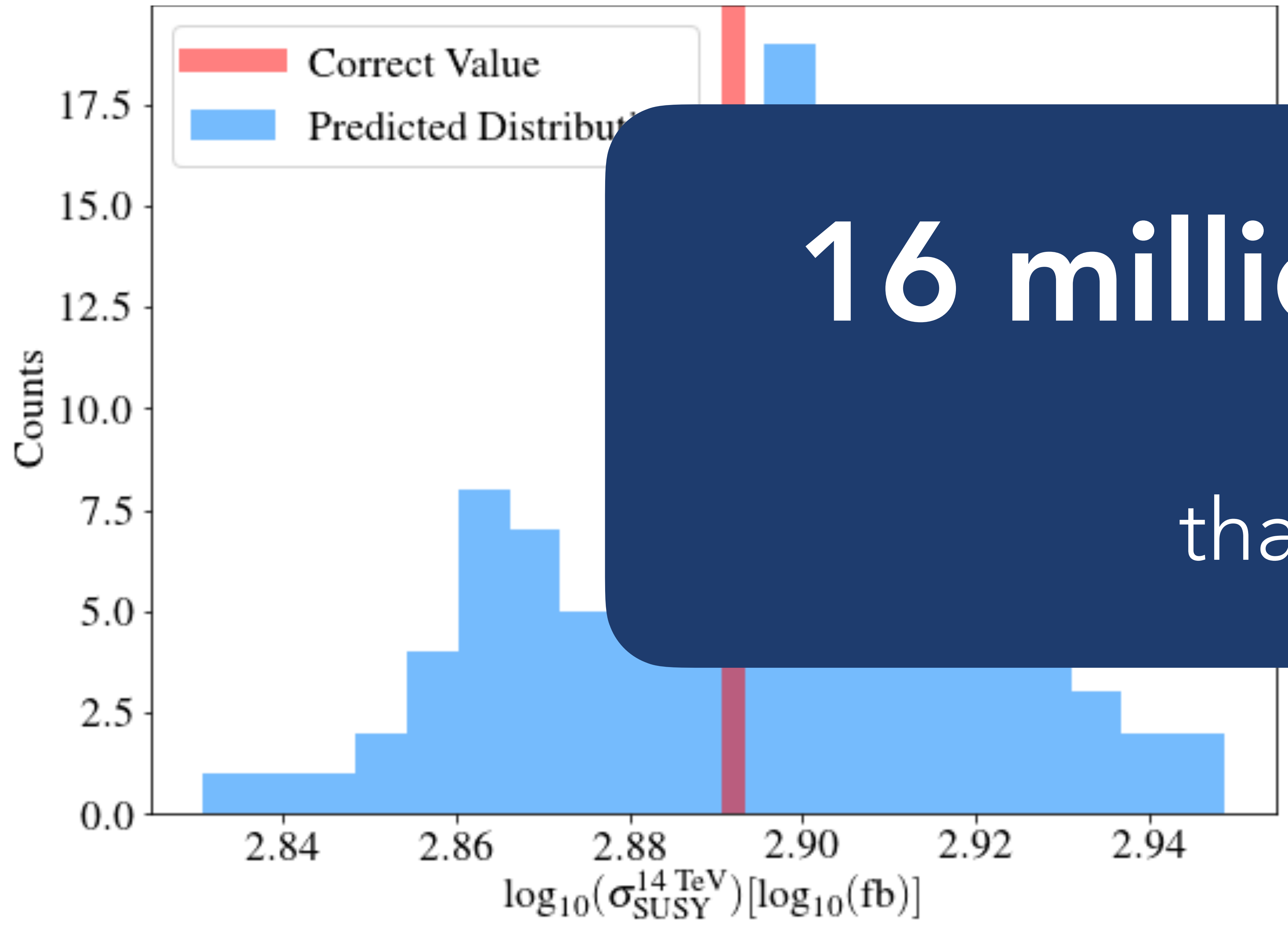
Bayesian Neural Networks

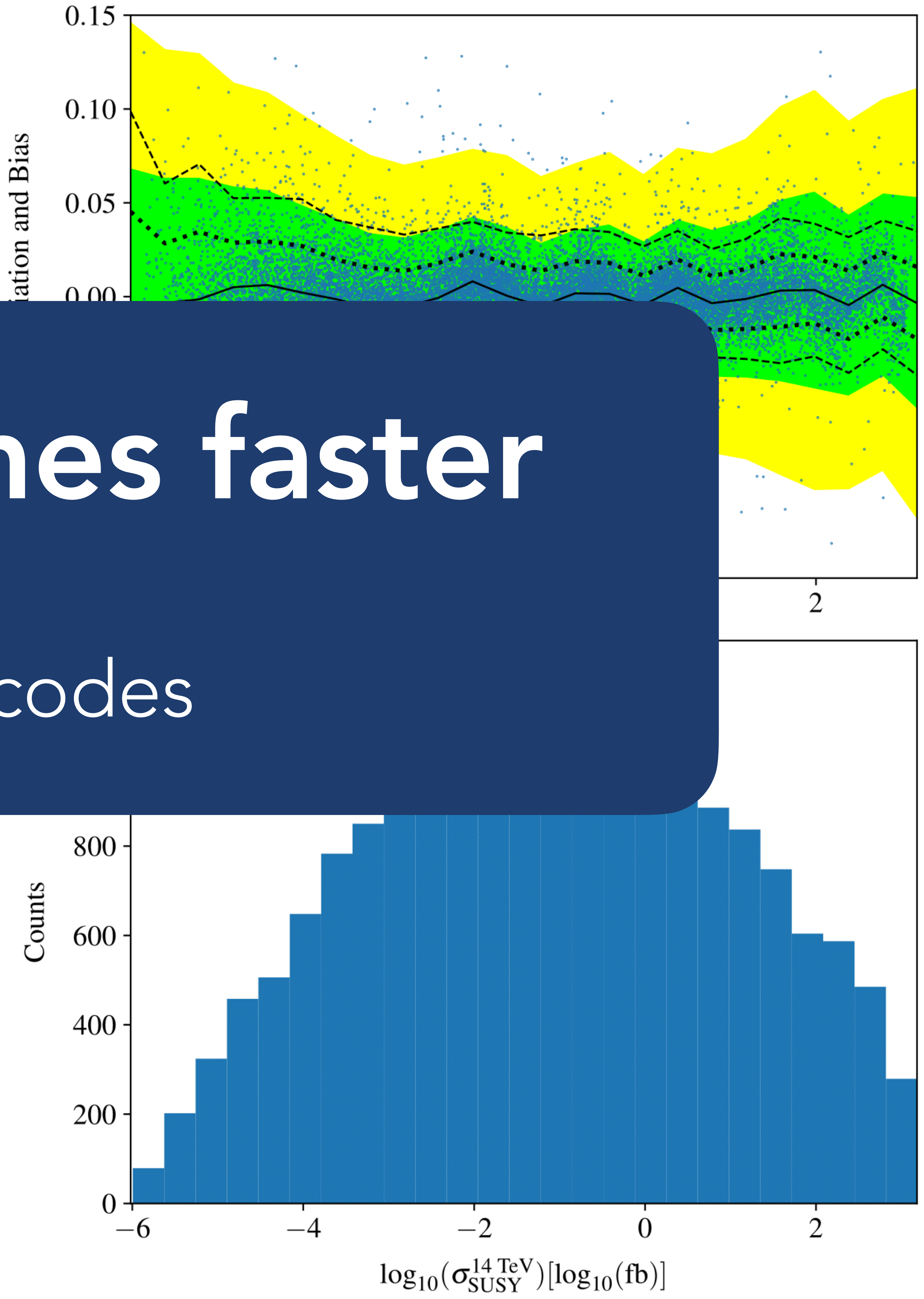
Training — Bayesian inference

Can we make predictions with useful epistemic uncertainty estimates?



FAST MAPPING FROM THEORY TO OBSERVABLES





16 million times faster

than theory codes

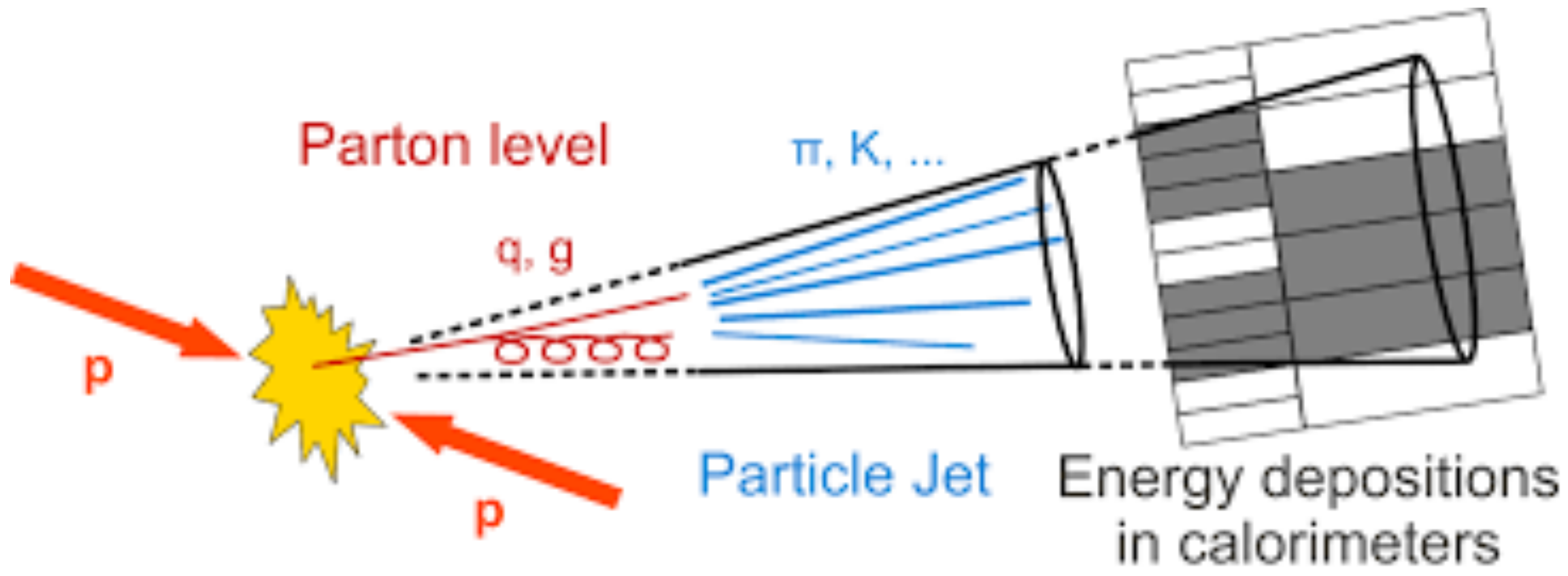
B.S. Kronheim, M.P. Kuchera, H.B. Prosper, A. Karbo, Bayesian neural networks for fast SUSY predictions, Physics Letters B, Volume 813, 2021, 136041, ISSN 0370-2693, https://doi.org/10.1016/j.physletb.2020.136041.

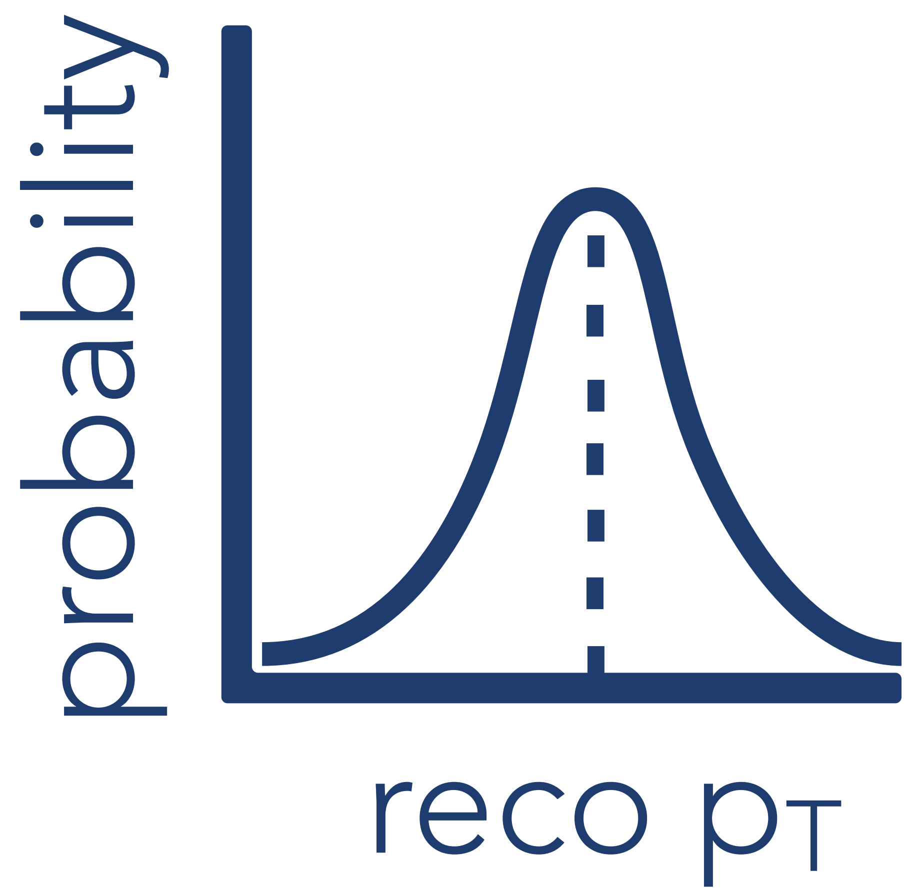
https://arxiv.org/abs/2009.14393

https://alpha-davidson.github.io/TensorBNN

JET SIMULATION AND CORRECTION

NEED FOR DISTRIBUTION PREDICTIONS

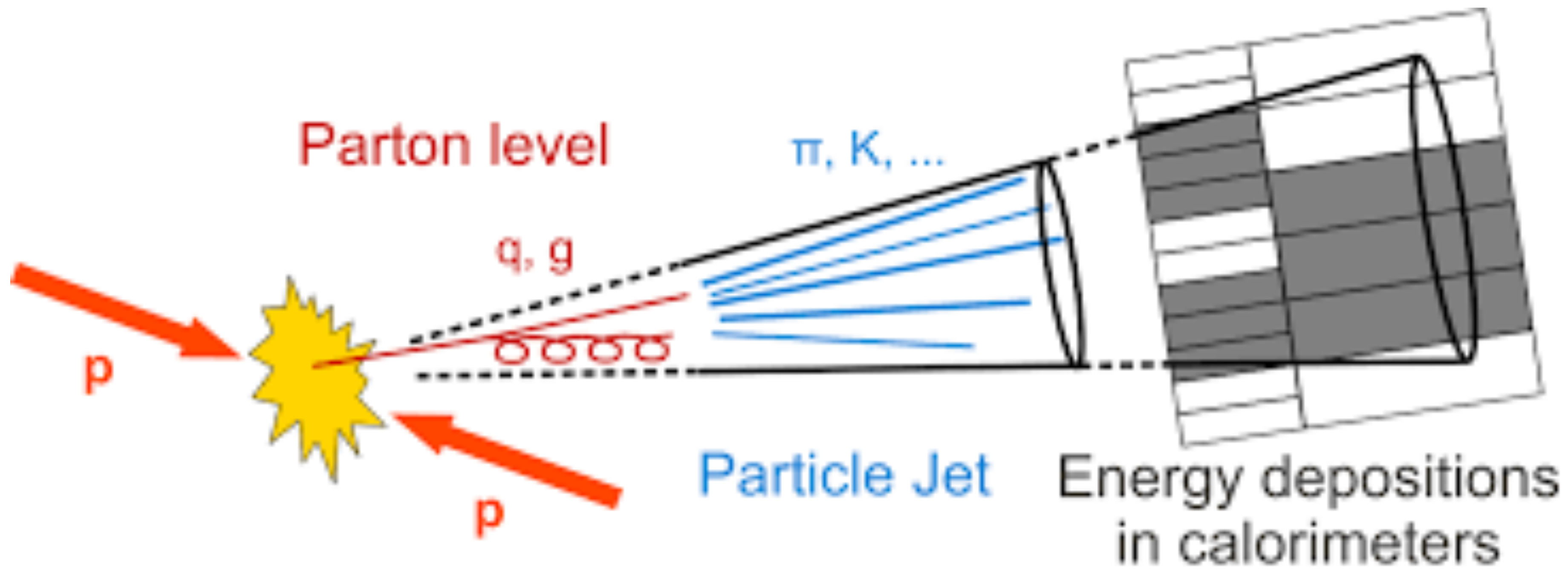




JET SIMULATION AND CORRECTION

J. BLUE, ET.AL., CHEP '21
EPJ WOC **251**, 03055 (2021) HTTPS://DOI.ORG/10.1051/EPJCONF/202125103055

EXISTING METHODS



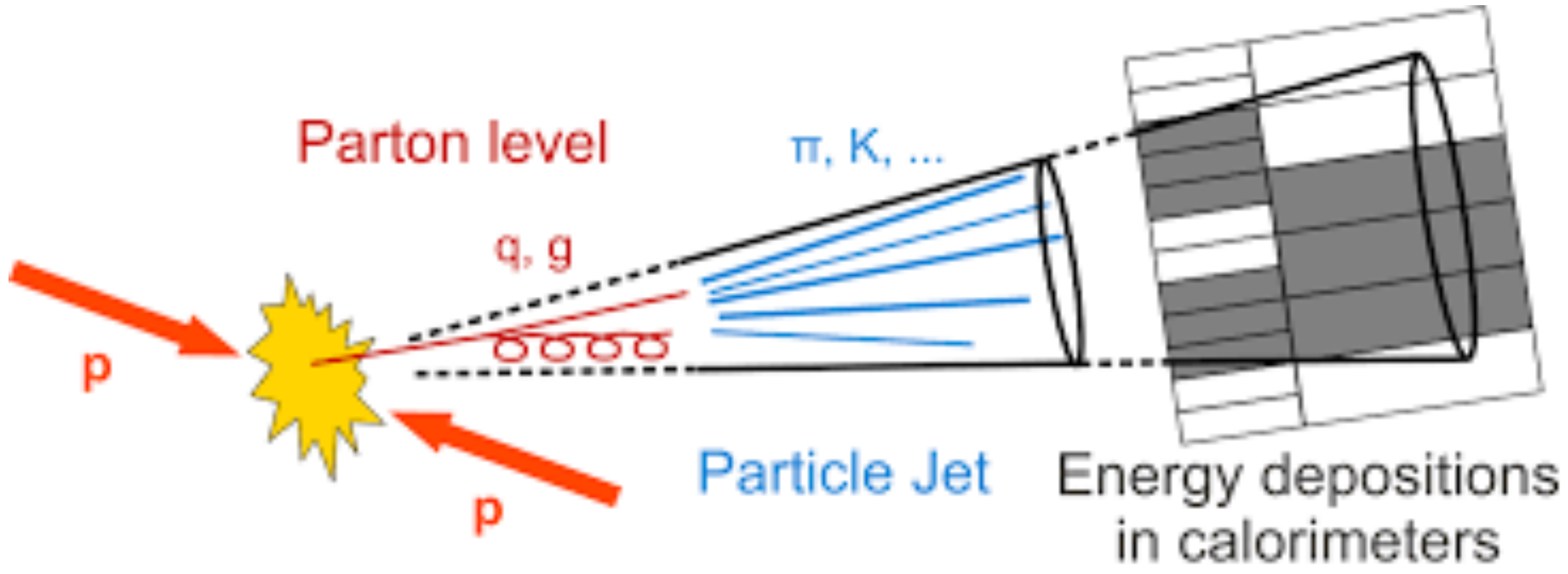
(conditional) generative adversarial networks

arXiv:1912.00477
arXiv:1807.01954
arXiv:1805.00850
arXiv:1712.10321

normalizing flows

arXiv:1904.12072
arXiv:2001.05486
arXiv:2001.10028
arXiv:2012.09873
arXiv:2106.05285

EXISTING METHODS



(conditional) generative adversarial networks

arXiv:1912.00477
arXiv:1807.01954
arXiv:1805.00850
arXiv:1712.10321

How to GAN away Detector Effects

Marco Bellagente[1], Anja Butter[1], Gregor Kasieczka[2], Tilman Plehn[1], and Ramon Winterhalder[1]

1 Institut für Theoretische Physik, Universität Heidelberg, Germany
2 Institut für Experimentalphysik, Universität Hamburg, Germany
bellagente@thphys.uni-heidelberg.de

Precise simulation of electromagnetic calorimeter showers using a Wasserstein Generative Adversarial Network

Martin Erdmann[a] **Jonas Glombitza**[a] **Thorben Quast**[a,b]

[a] *III. Physikalisches Institut A, Rheinisch Westfälische Technische Hochschule, Aachen, Germany*
[b] *EP-LCD, CERN, Geneva, Switzerland*

Fast and accurate simulation of particle detectors using generative adversarial networks

Pasquale Musella · Francesco Pandolfi

CaloGAN: Simulating 3D High Energy Particle Showers in Multi-Layer Electromagnetic Calorimeters with Generative Adversarial Networks

Michela Paganini,[1,2,*] Luke de Oliveira,[2,†] and Benjamin Nachman[2,‡]

[1] *Department of Physics, Yale University, New Haven, CT 06520, USA*
[2] *Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA*

(Dated: January 1, 2018)

Flow-based generative models for Markov chain Monte Carlo in lattice field theory

M. S. Albergo,[1,2,3] G. Kanwar,[4] and P. E. Shanahan[4,1]

[1] *Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada*
[2] *Cavendish Laboratories, University of Cambridge, Cambridge CB3 0HE, U.K.*
[3] *University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*
[4] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.*

`i-flow`: High-dimensional Integration and Sampling with Normalizing Flows

Christina Gao[1], Joshua Isaacson[1], and Claudius Krause[1]

[1] Theoretical Physics Department, Fermi National Accelerator Laboratory, Batavia, IL, 60510, USA

Event Generation with Normalizing Flows

Christina Gao,[1] Stefan Höche,[1] Joshua Isaacson,[1] Claudius Krause,[1] and Holger Schulz[2]

[1] *Fermi National Accelerator Laboratory, Batavia, IL, 60510, USA*
[2] *Department of Physics, University of Cincinnati, Cincinnati, OH 45219, USA*

Measuring QCD Splittings with Invertible Networks

Sebastian Bieringer[1], Anja Butter[1], Theo Heimel[1], Stefan Höche[2], Ullrich Köthe[3], Tilman Plehn[1], and Stefan T. Radev[4]

1 Institut für Theoretische Physik, Universität Heidelberg, Germany
2 Fermi National Accelerator Laboratory, Batavia, IL, USA
3 Heidelberg Collaboratory for Image Processing, Universität Heidelberg, Germany
4 Psychologisches Institut, Universität Heidelberg, Germany
heimel@thphys.uni-heidelberg.de

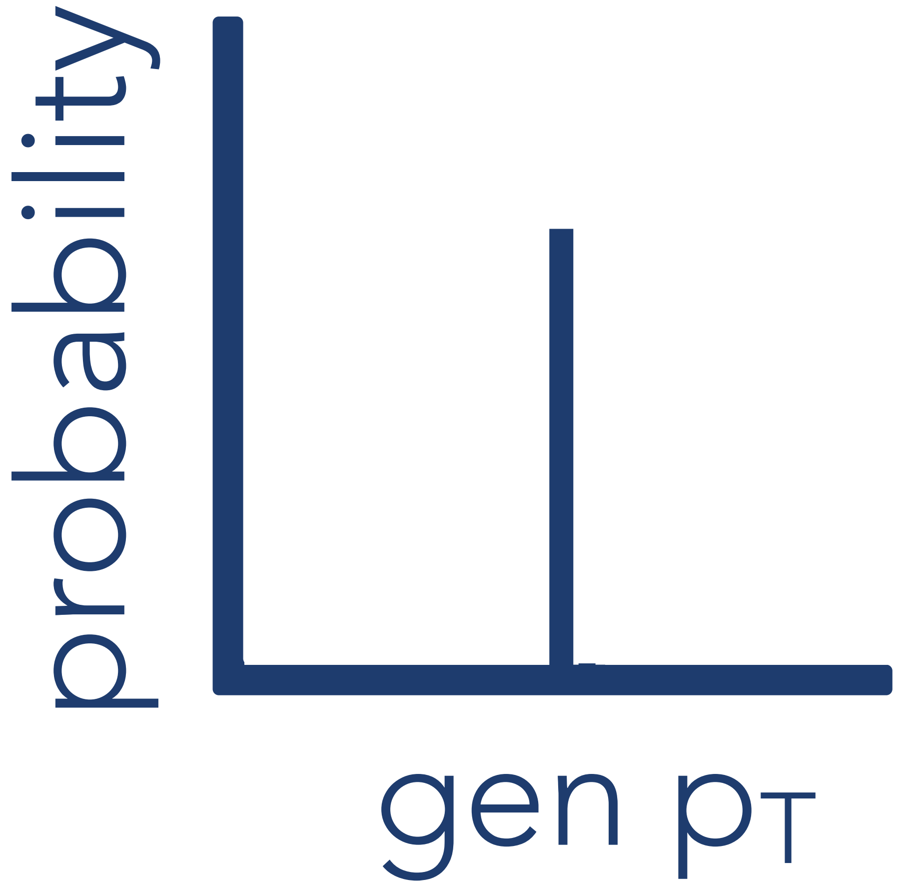
CaloFlow: Fast and Accurate Generation of Calorimeter Showers with Normalizing Flows

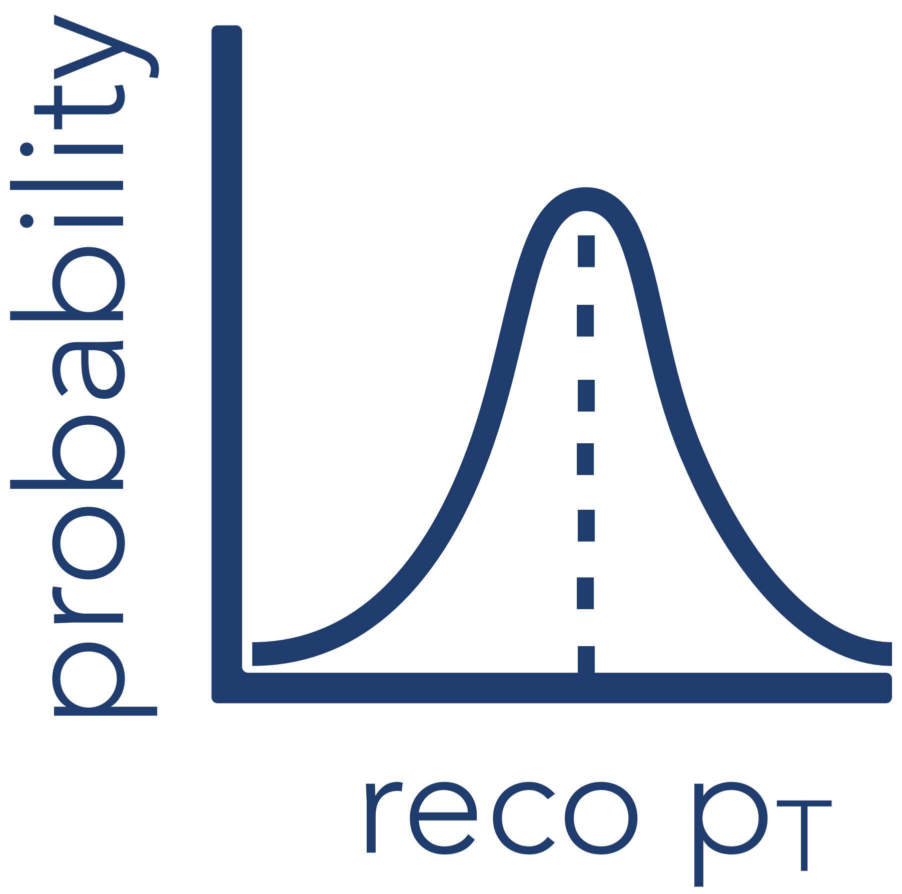
---

Claudius Krause and David Shih

NHETC, Dept. of Physics and Astronomy, Rutgers University, Piscataway, NJ 08854, USA

E-mail: Claudius.Krause@rutgers.edu, shih@physics.rutgers.edu





normalizing flows

arXiv:1904.12072

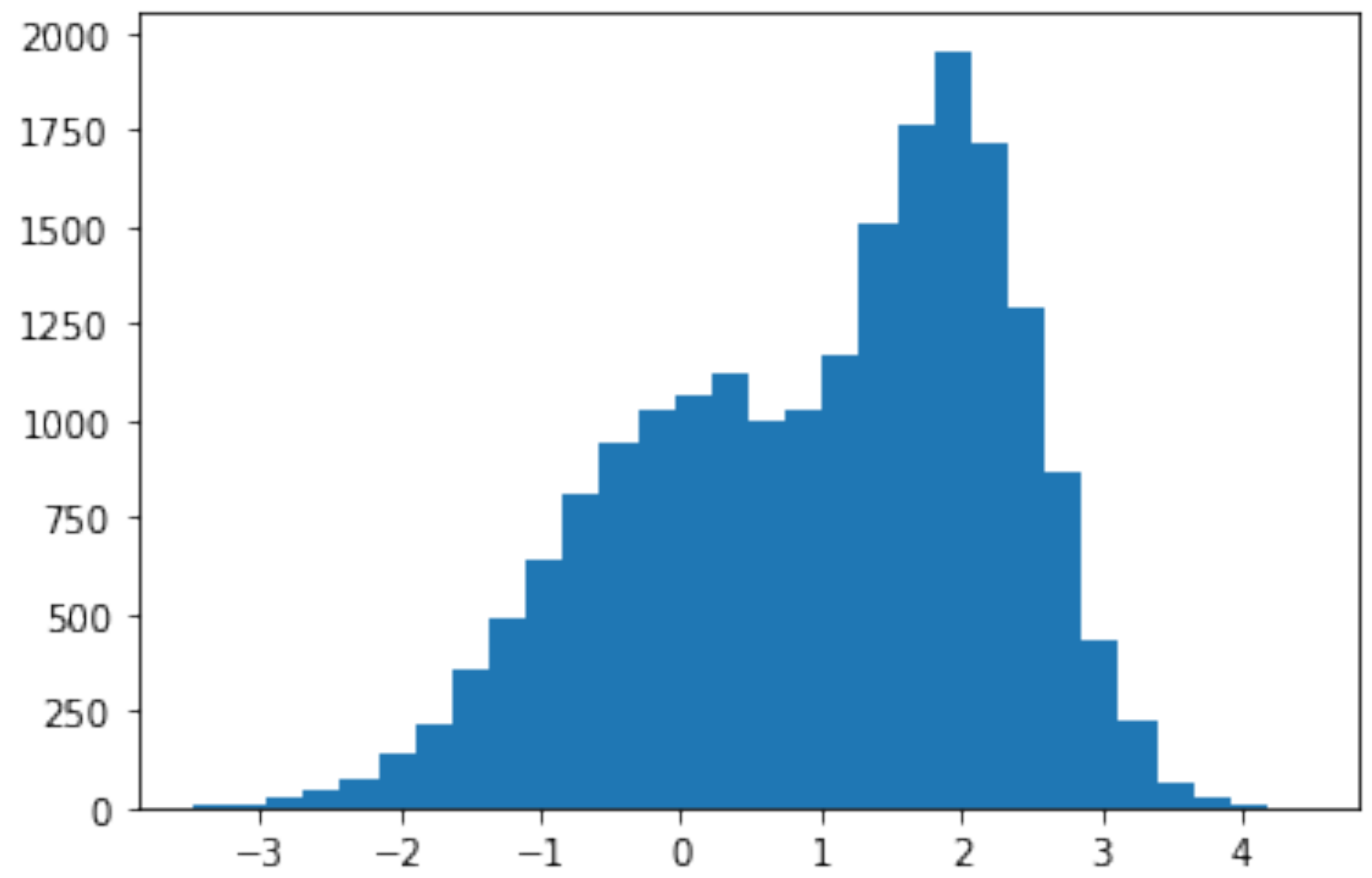
arXiv:2001.05486

arXiv:2001.10028

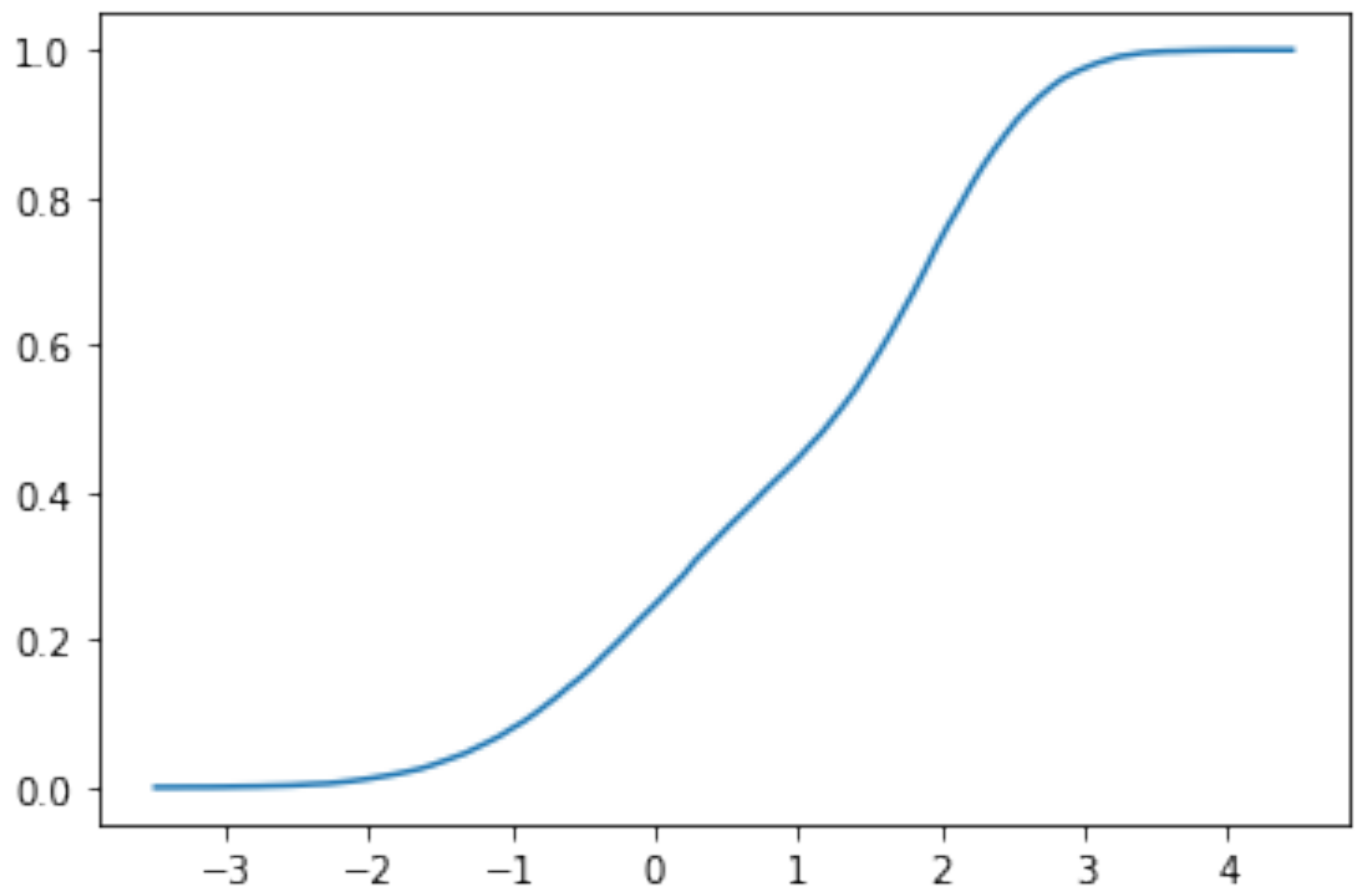
arXiv:2012.09873

arXiv:2106.05285

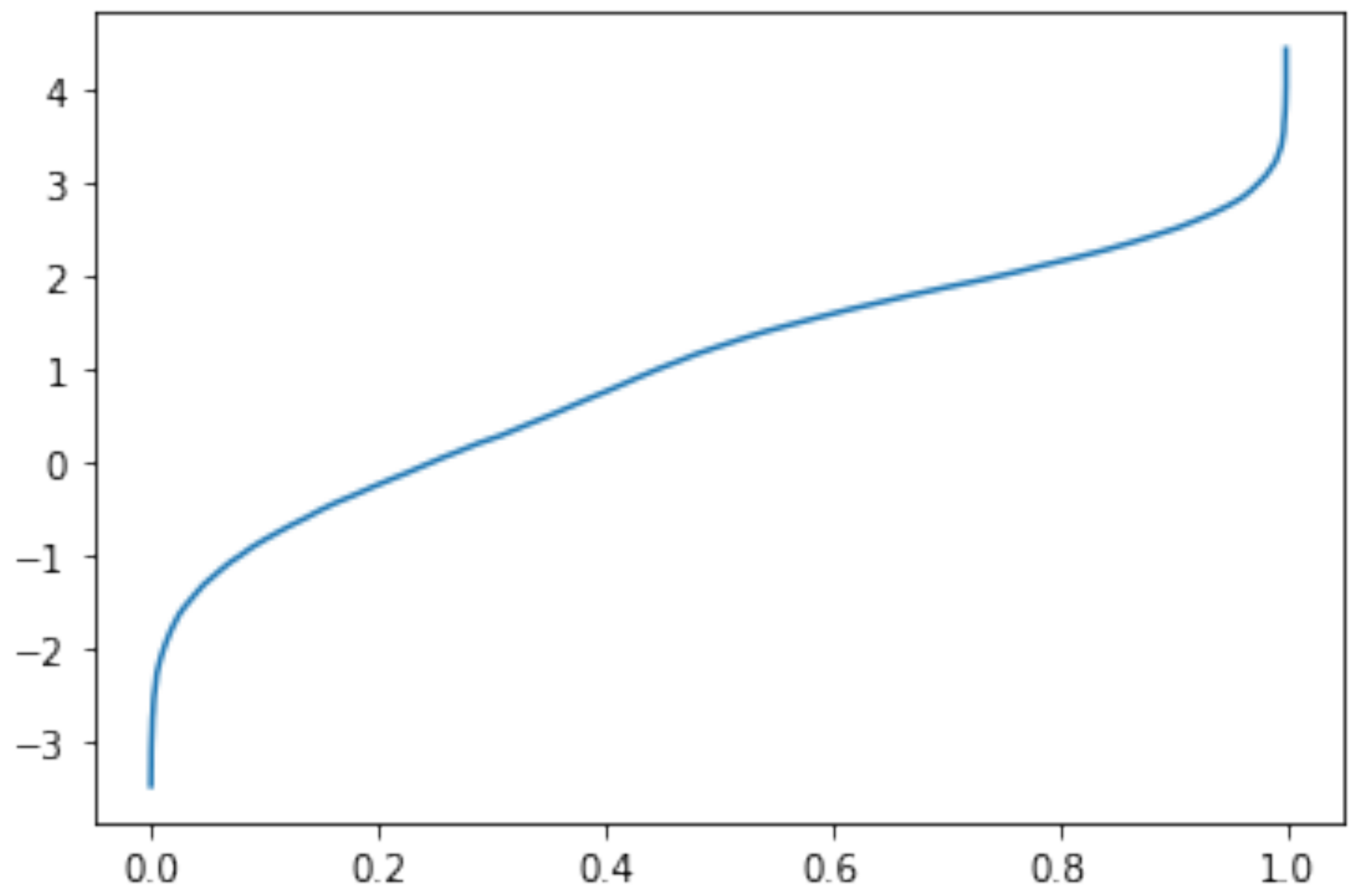
IMPLICIT QUANTILE NETWORKS STATS REVIEW



data

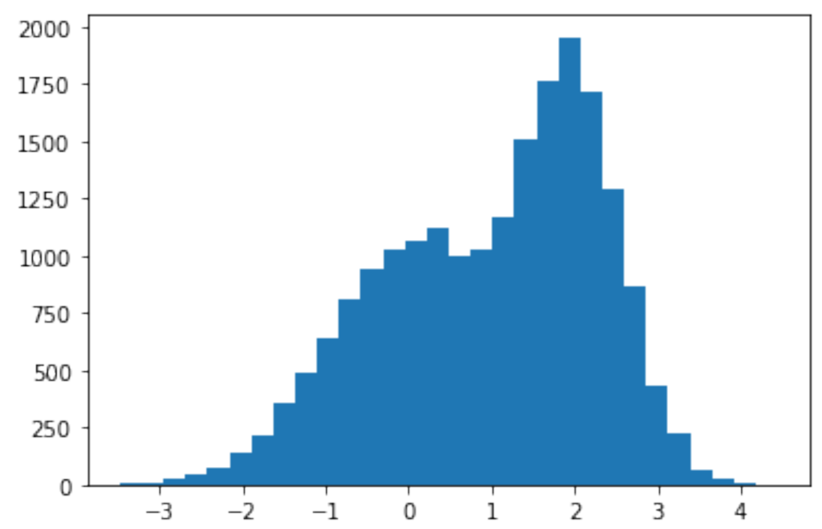


CDF

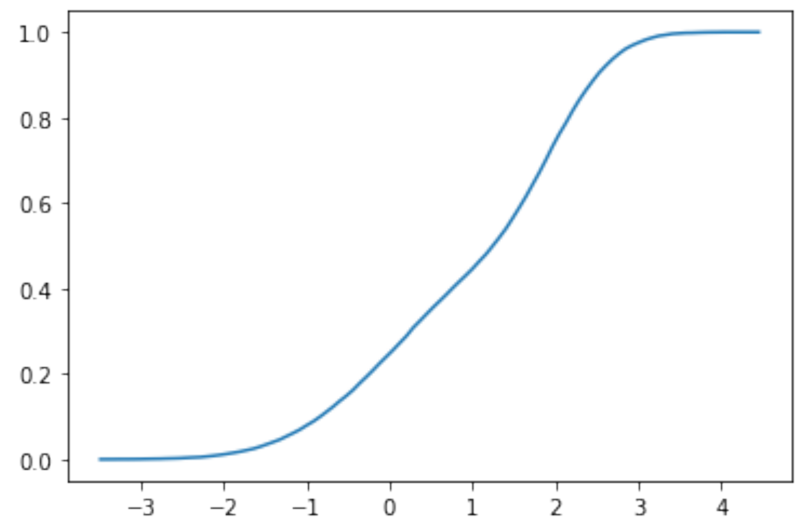


quantile function

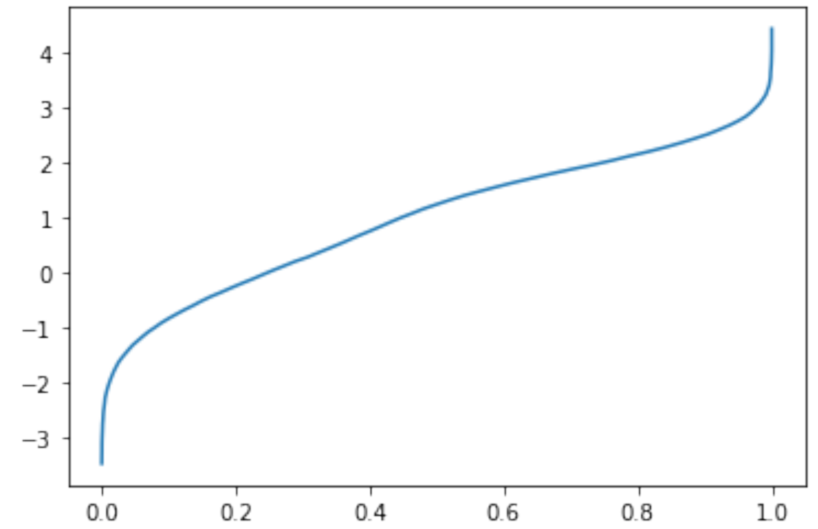
IMPLICIT QUANTILE NETWORKS STATS REVIEW



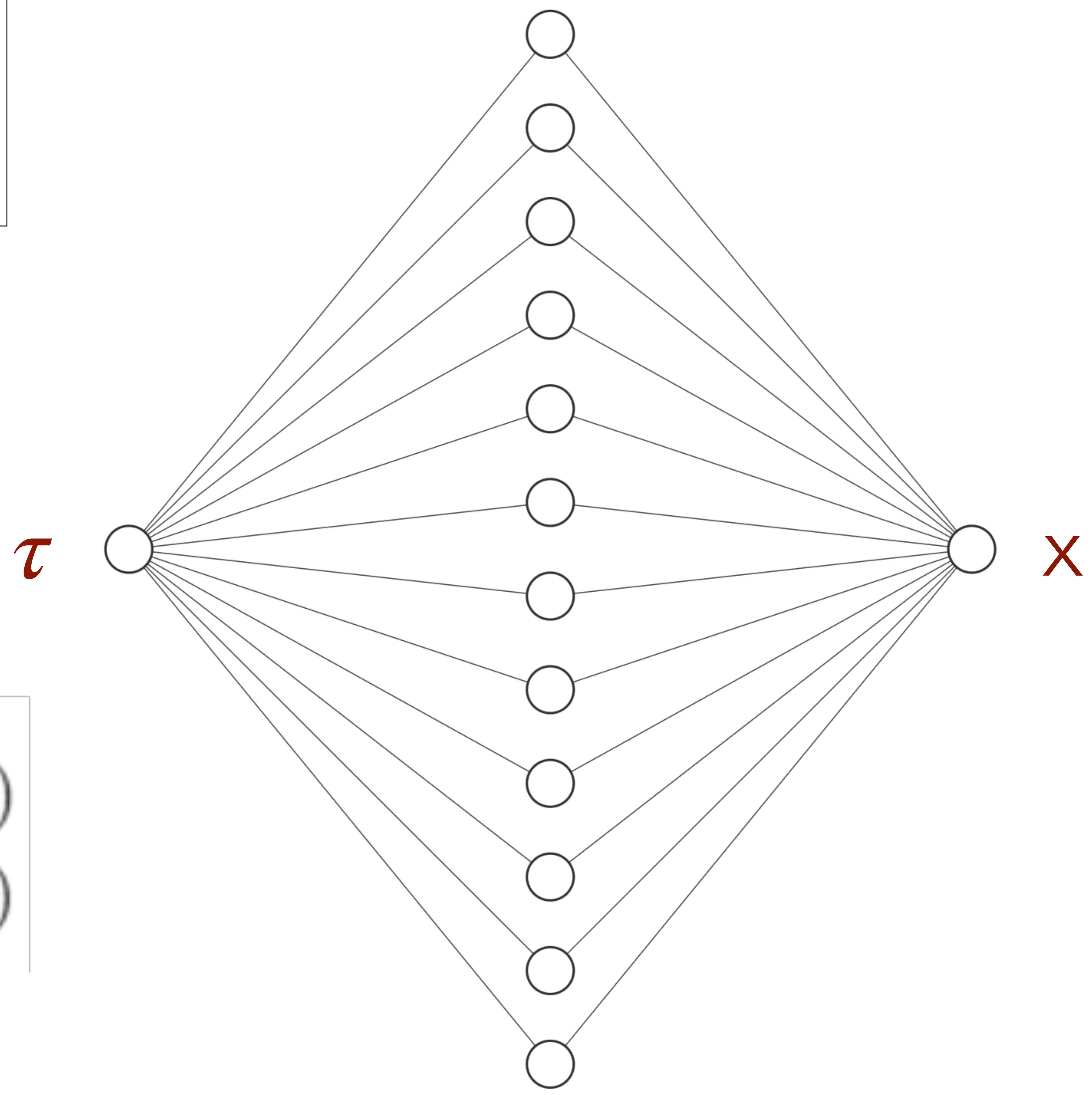
data (x)



CDF

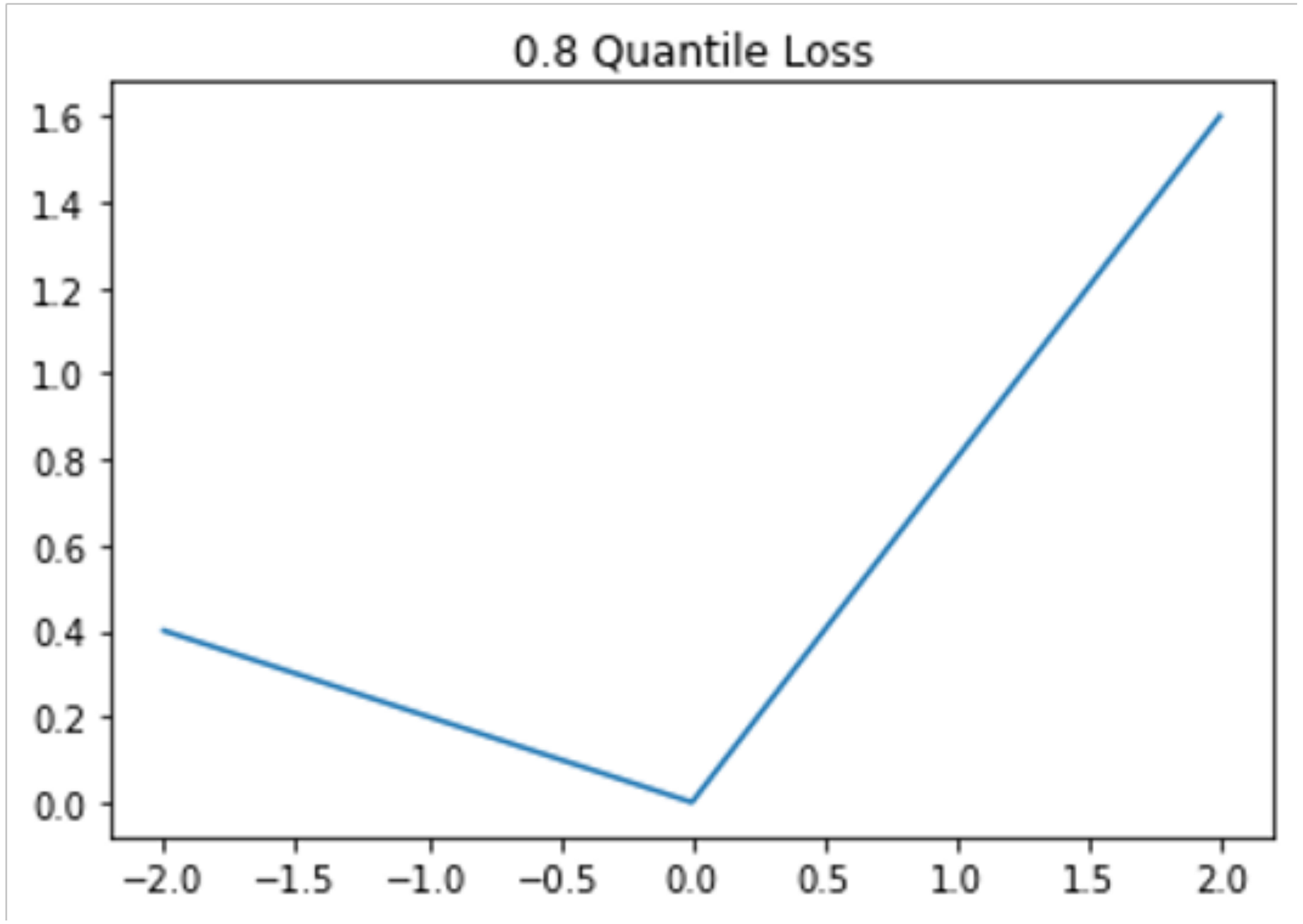


quantile function (τ)



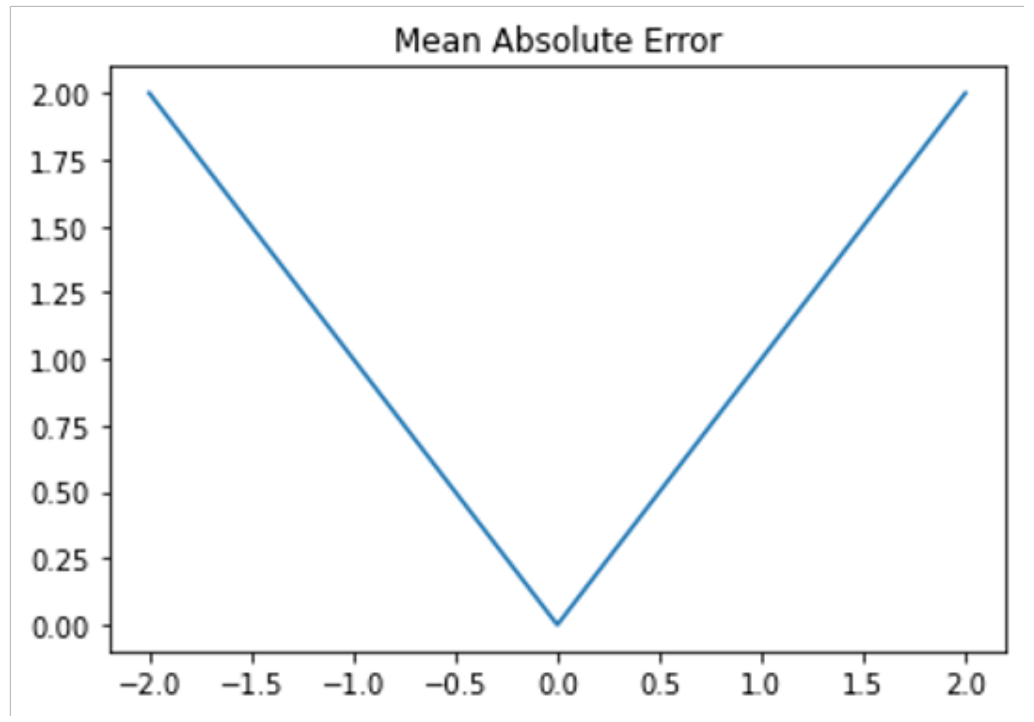
$$\mathcal{L}(f, x, y, \tau) = \begin{cases} \tau(y - f(x,\tau)) & y \geq f(x,\tau) \\ (\tau - 1)(y - f(x,\tau)) & y < f(x,\tau) \end{cases}$$

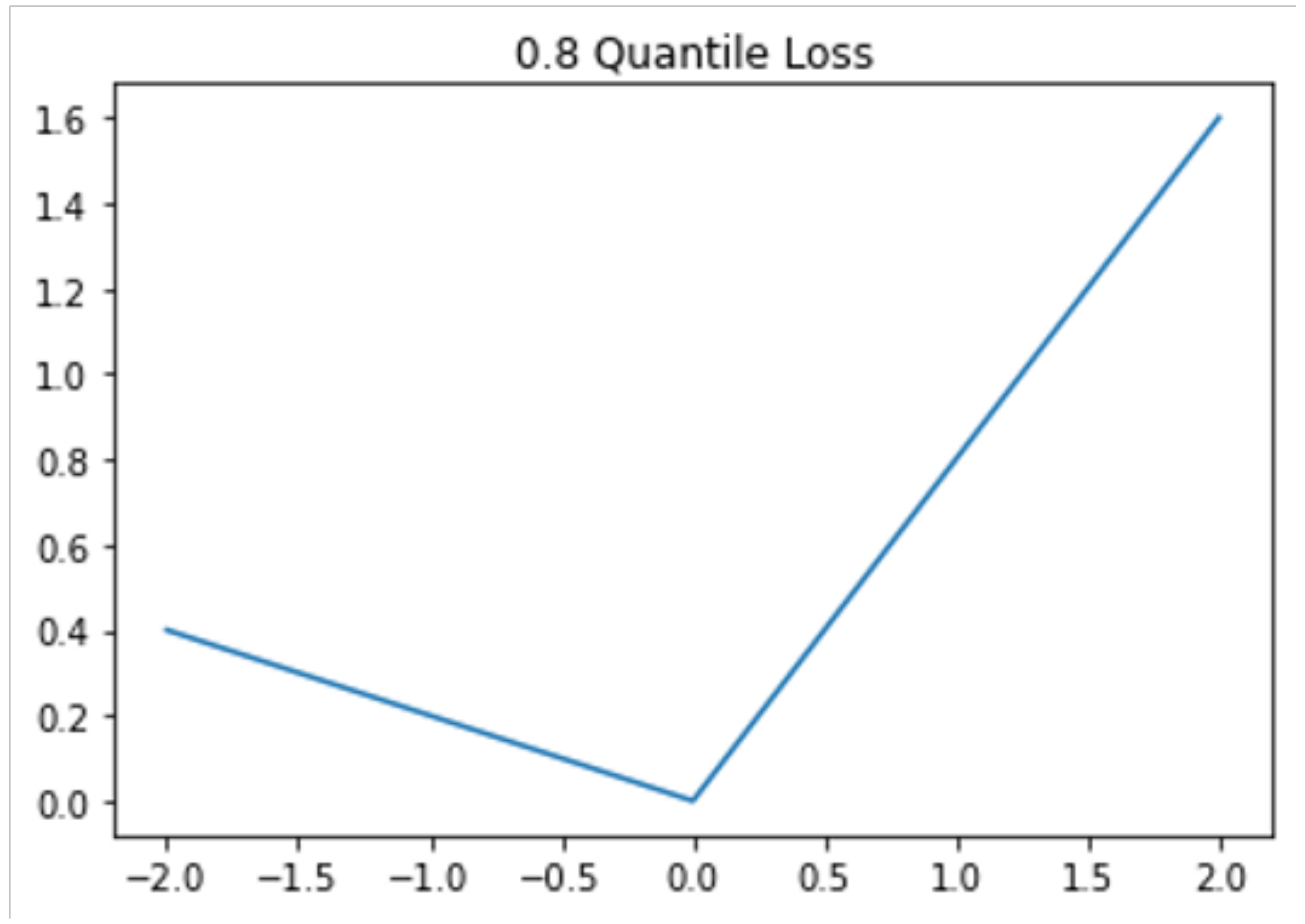
IMPLICIT QUANTILE NETWORKS LOSS FUNCTION

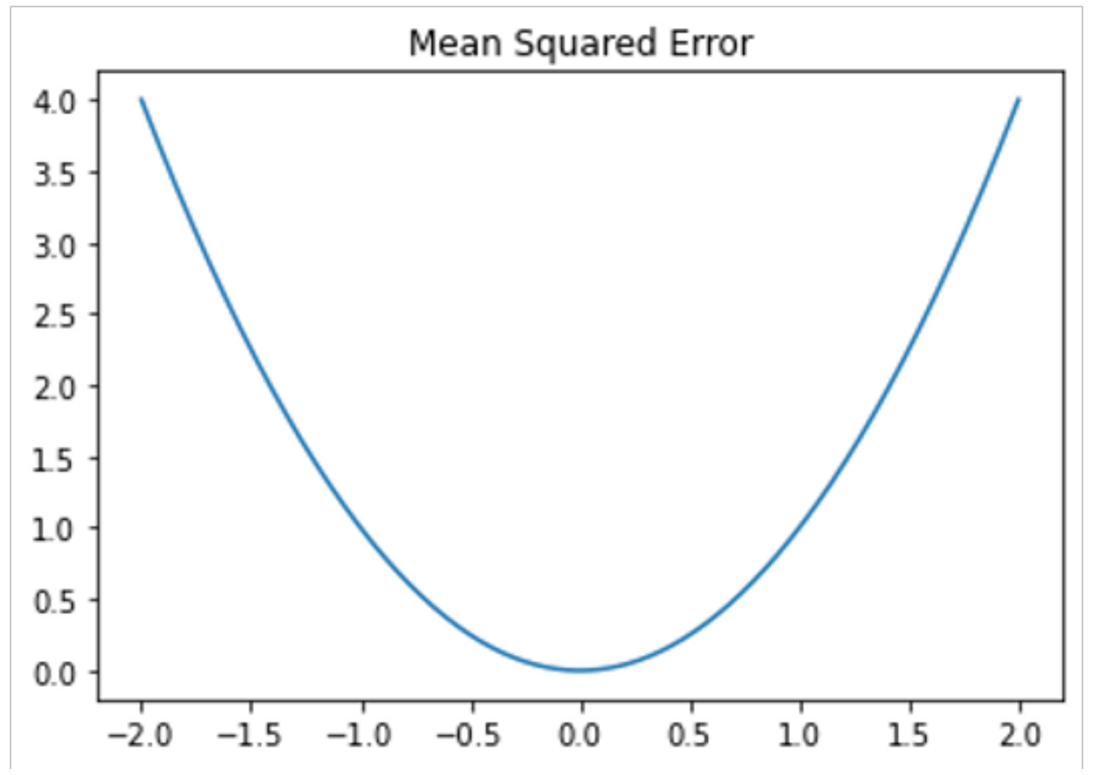


$$\mathcal{L}(f, x, y, \tau) = \begin{cases} \tau(y - f(x,\tau)) & y \geq f(x,\tau) \\ (\tau - 1)(y - f(x,\tau)) & y < f(x,\tau) \end{cases}$$

IMPLICIT QUANTILE NETWORKS LOSS FUNCTION

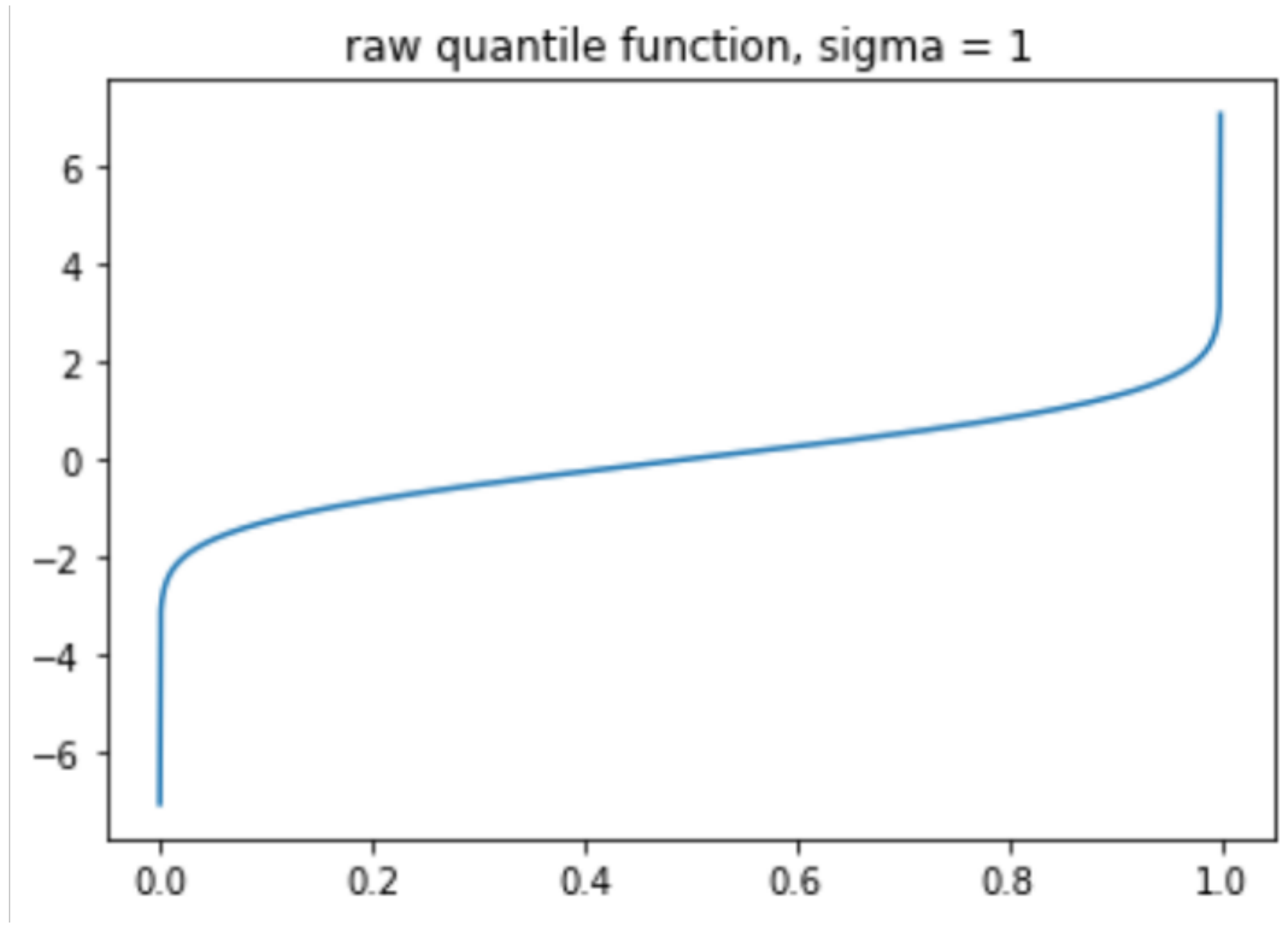


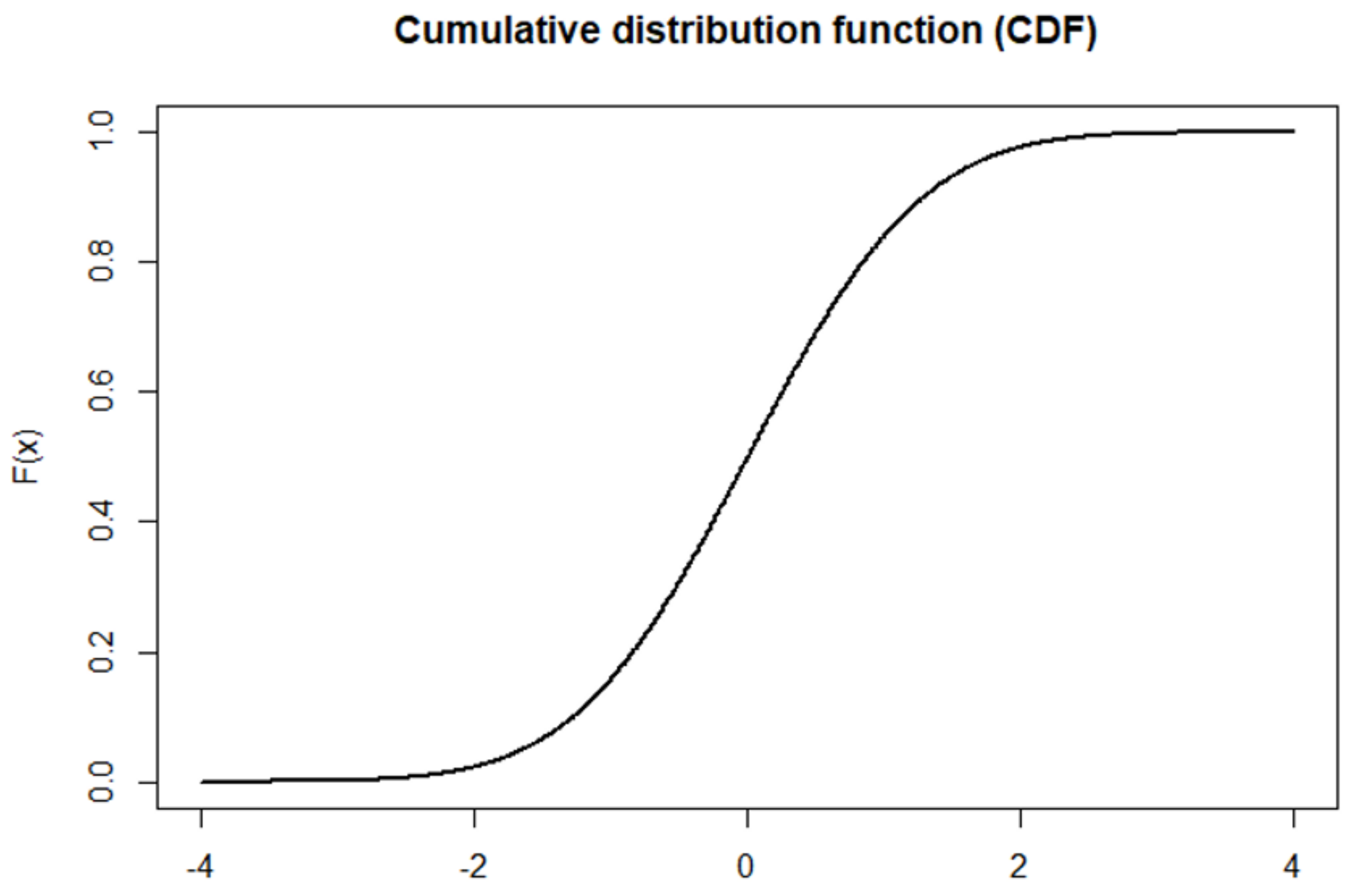


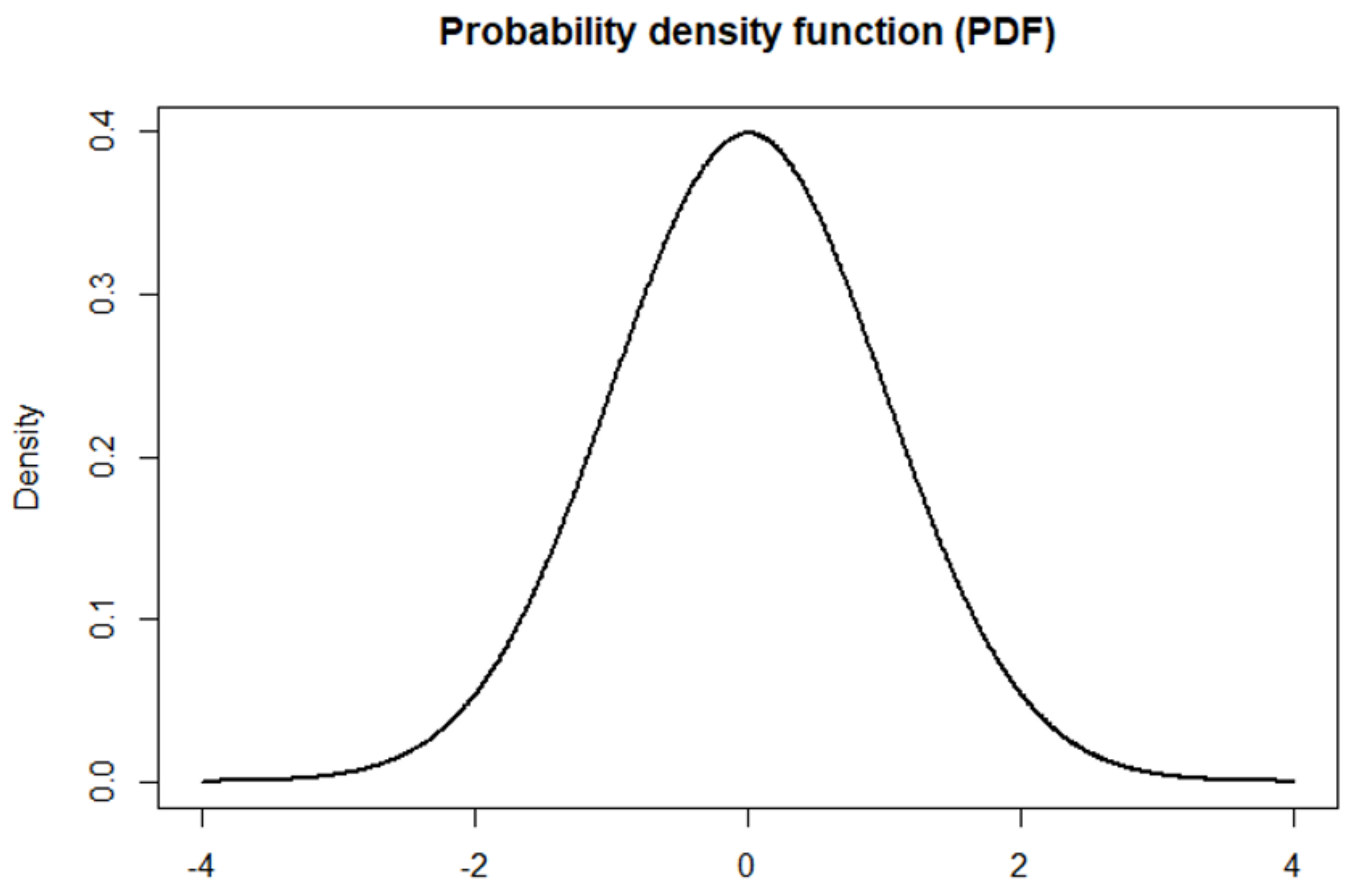


$$\mathcal{L}(f, x, y, \tau) = \begin{cases} \tau(y - f(x,\tau)) & y \geq f(x,\tau) \\ (\tau - 1)(y - f(x,\tau)) & y < f(x,\tau) \end{cases}$$

IMPLICIT QUANTILE NETWORKS LOSS FUNCTION







$$\mathcal{L}(f, x, y, \tau) = \begin{cases} \tau(y - f(x,\tau)) & y \geq f(x,\tau) \\ (\tau - 1)(y - f(x,\tau)) & y < f(x,\tau) \end{cases}$$

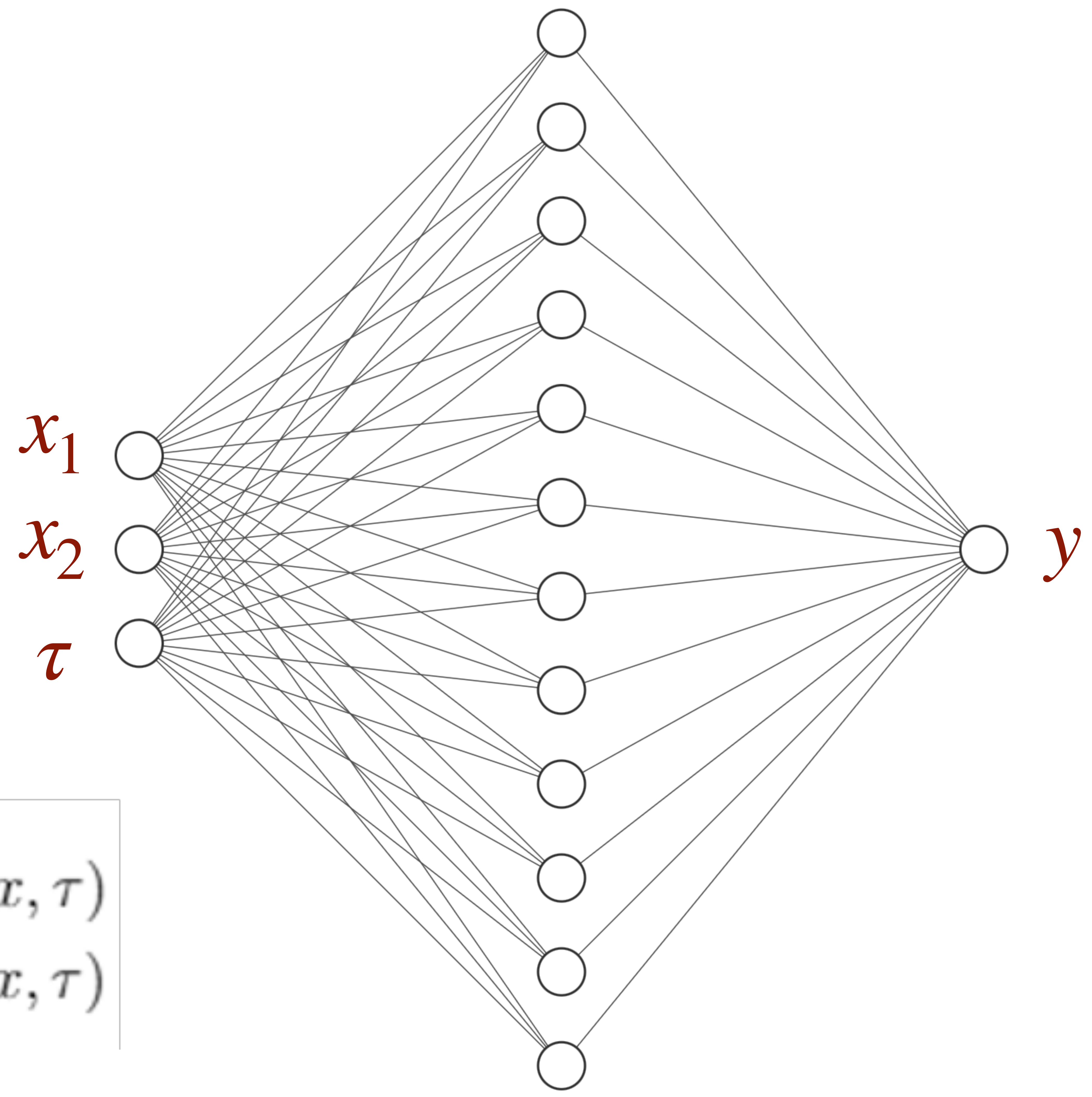
regularization $\begin{cases} \left(\frac{dy}{d\tau}\right)^2 & \frac{dy}{d\tau} < 0 \\ 0 & \frac{dy}{d\tau} \geq 0 \end{cases}$

IMPLICIT QUANTILE NETWORKS ARCHITECTURE

$p(y \mid x_1, x_2)$

$(x_1, x_2) \to y$

$$\mathcal{L}(f, x, y, \tau) = \begin{cases} \tau(y - f(x, \tau)) & y \geq f(x, \tau) \\ (\tau - 1)(y - f(x, \tau)) & y < f(x, \tau) \end{cases}$$



IMPLICIT QUANTILE NETWORKS ARCHITECTURE

$$p(p_T', \eta', \phi', m' \,|\, p_T, \eta, \phi, m)$$

$$(p_T, \eta, \phi, m) \to (p_T', \eta', \phi', m')$$

$$p(y^{(1)}, y^{(2)}, \ldots, y^{(n)} \,|\, \mathbf{x}) = p(y^{(1)} \,|\, \mathbf{x}) p(y^{(2)} \,|\, \mathbf{x}, y^{(1)}) \prod_{i=3}^{n} p(y^{(i)} \,|\, \mathbf{x}, y^{(1)}, \ldots, y^{(i-1)})$$

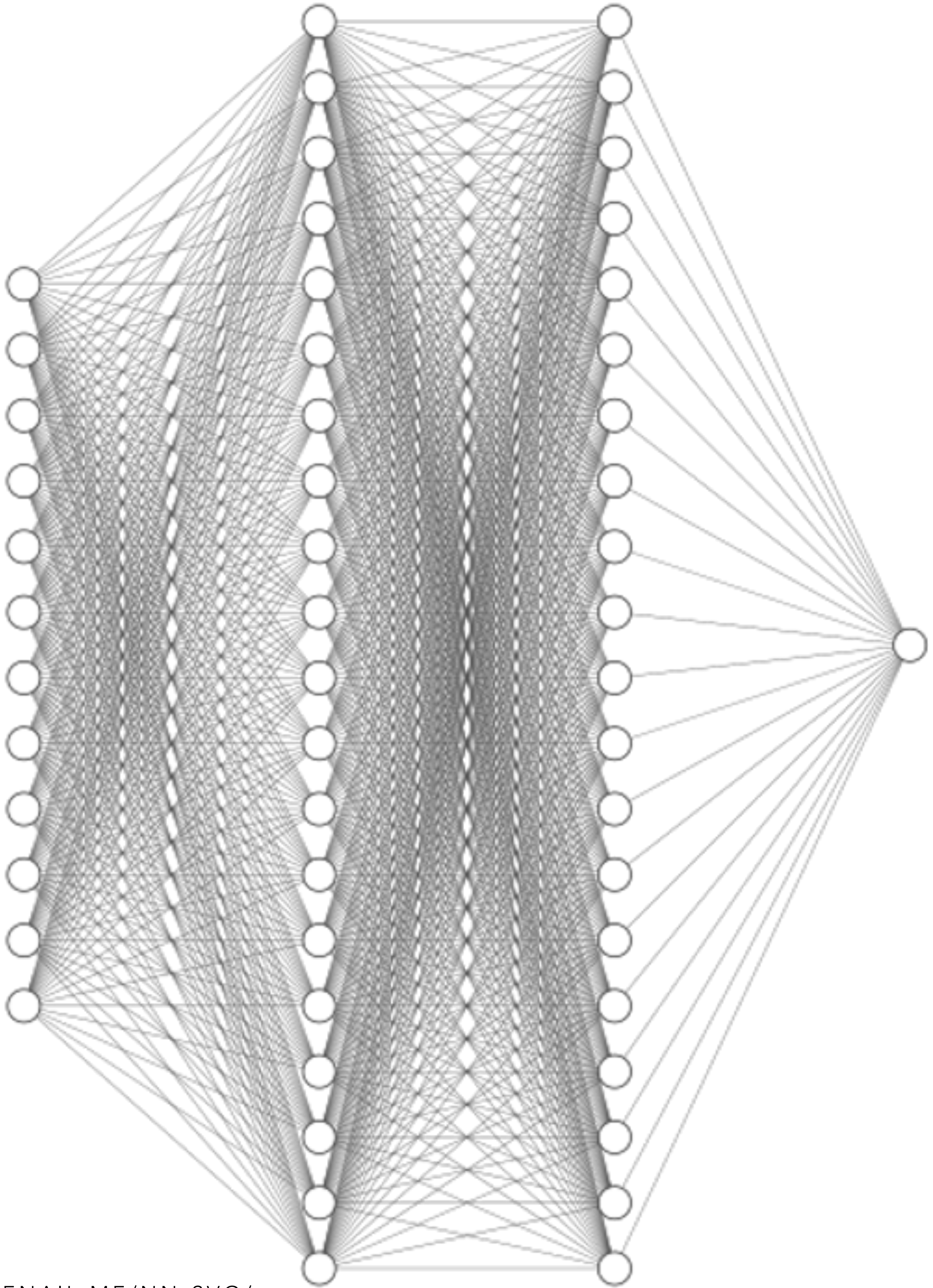
IMPLICIT QUANTILE NETWORKS ARCHITECTURE

$$p(p_T', \eta', \phi', m' \,|\, p_T, \eta, \phi, m)$$

$$(p_T, \eta, \phi, m) \rightarrow (p_T', \eta', \phi', m')$$

$$p(y^{(1)}, y^{(2)}, \ldots, y^{(n)} \,|\, \mathbf{x}) = \boxed{p(y^{(1)} \,|\, \mathbf{x})}\boxed{p(y^{(2)} \,|\, \mathbf{x}, y^{(1)})}\prod_{i=3}^{n}\boxed{p(y^{(i)} \,|\, \mathbf{x}, y^{(1)}, \ldots, y^{(i-1)})}$$

IMPLICIT QUANTILE NETWORKS ARCHITECTURE

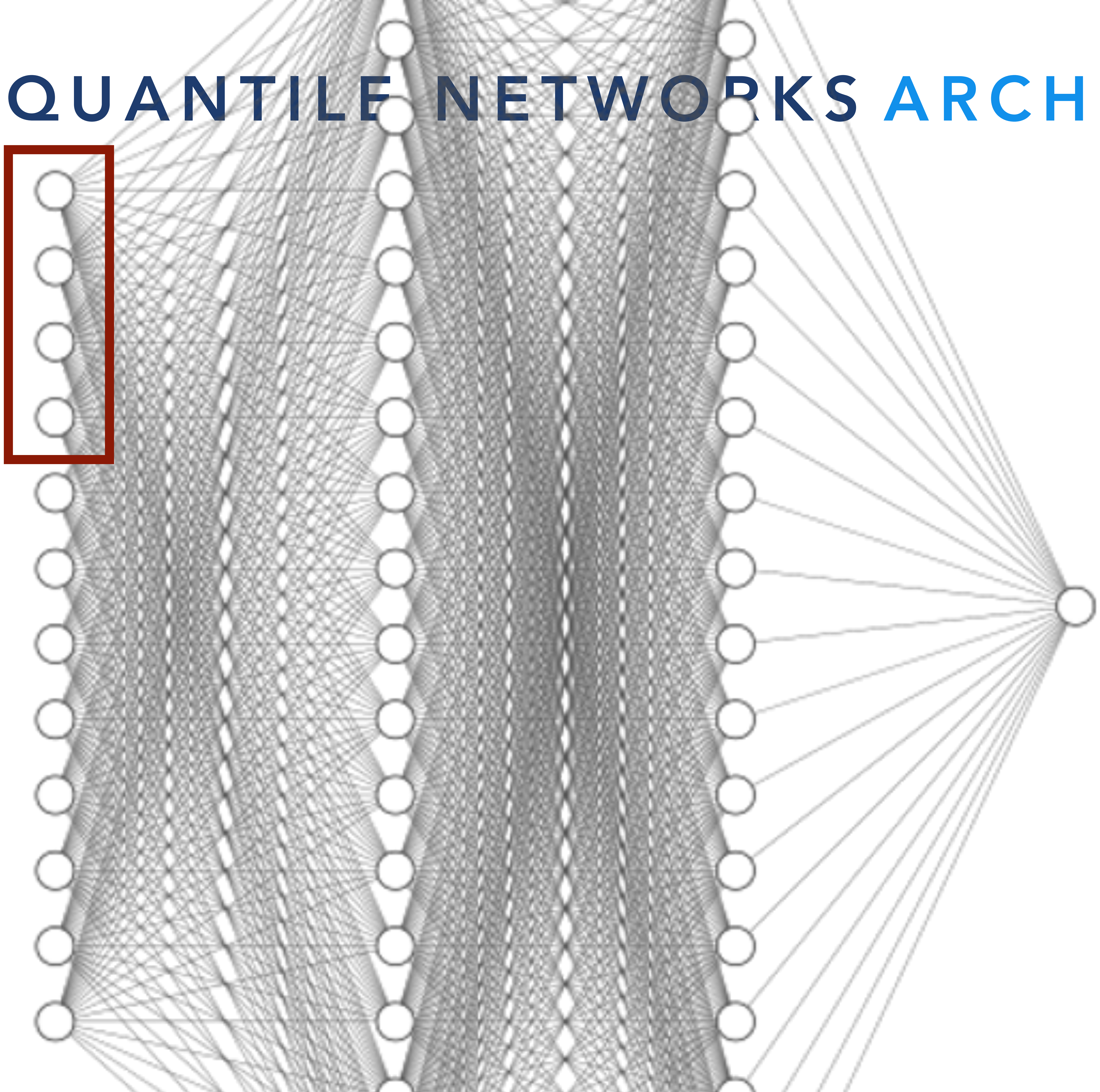


$$(p_T, \eta, \phi, m) \rightarrow (p_T', \eta', \phi', m')$$

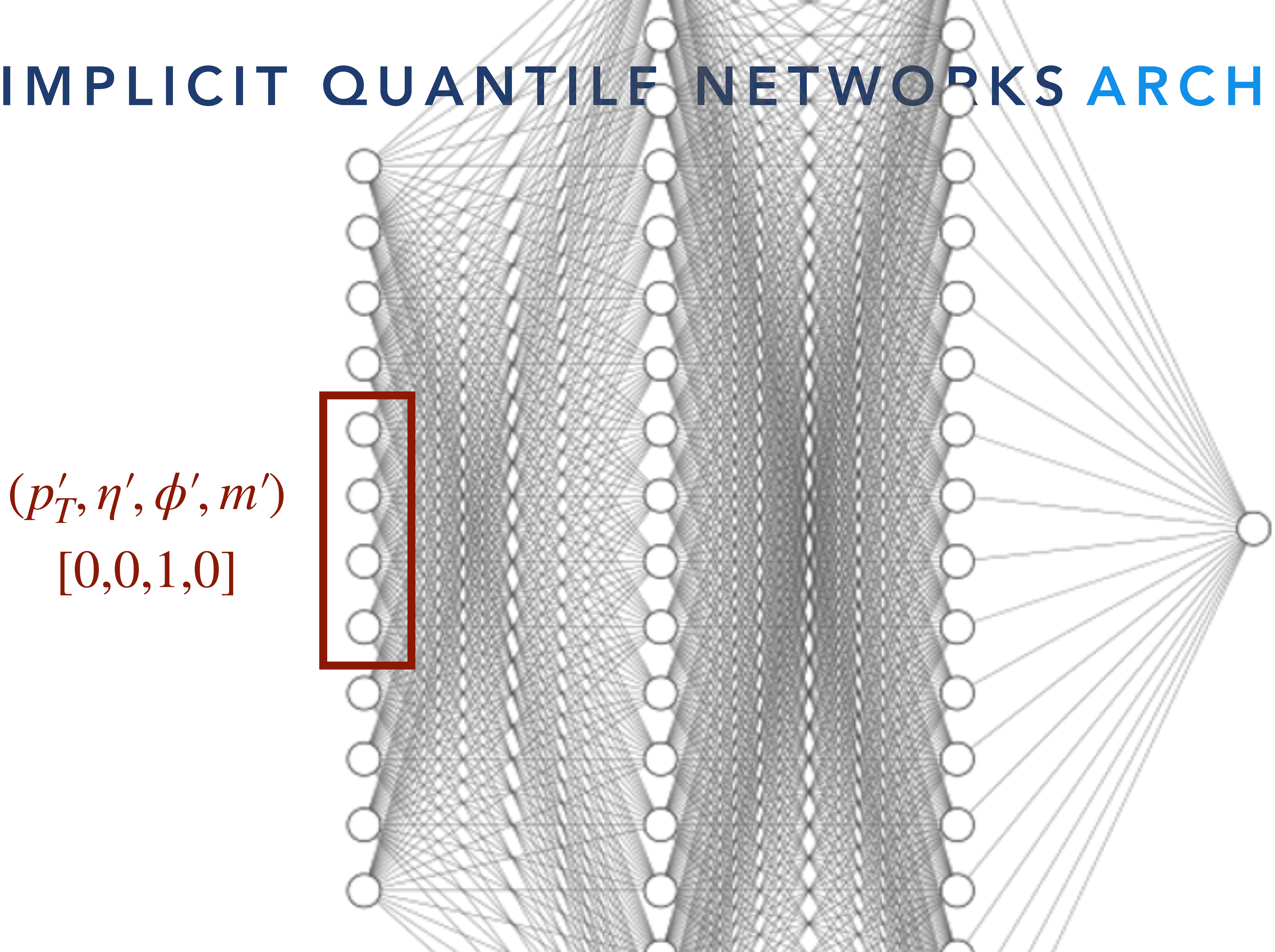
IMAGE CREDIT: HTTPS://ALEXLENAIL.ME/NN-SVG/

IMPLICIT QUANTILE NETWORKS ARCHITECTURE

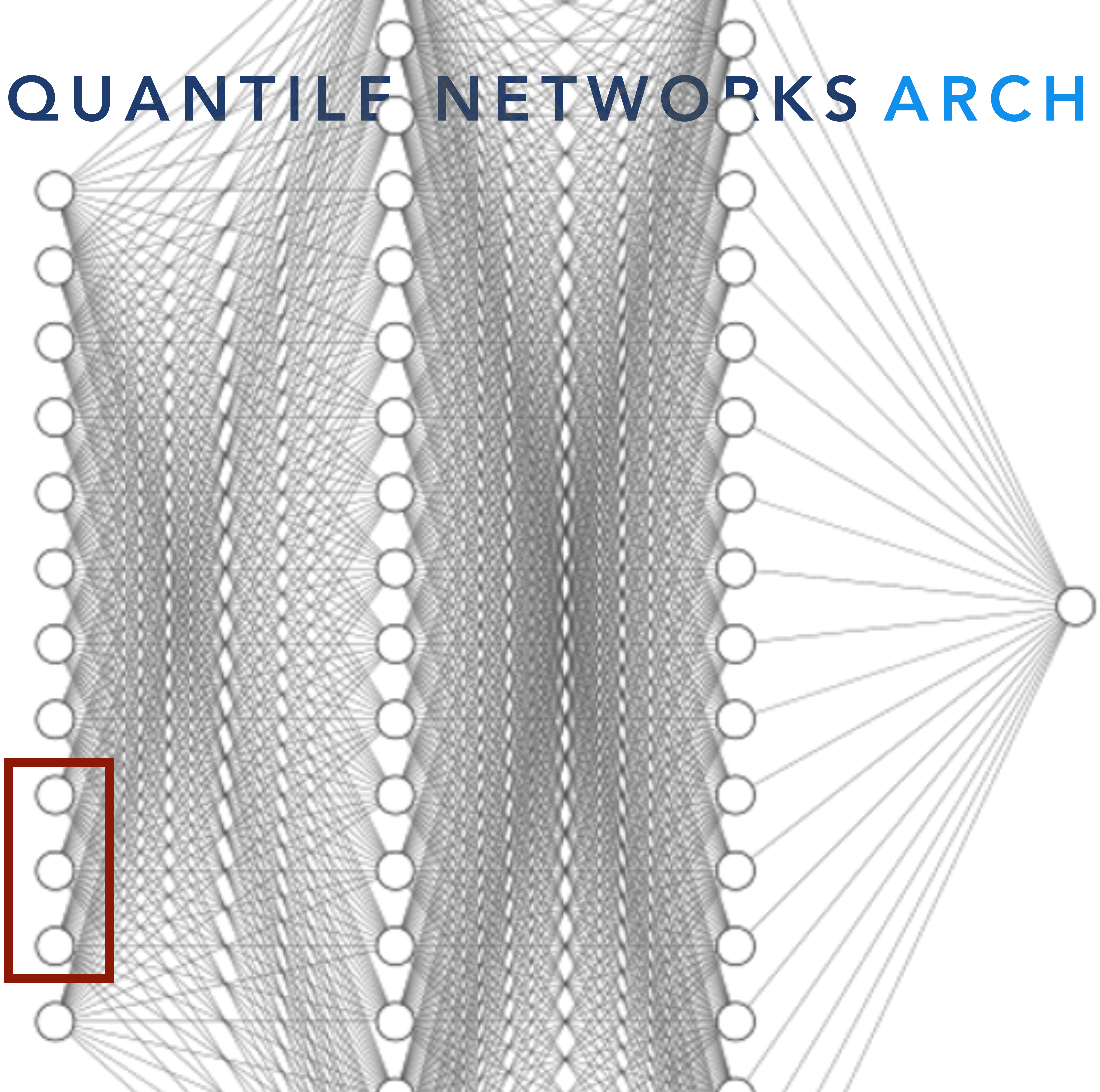
(p_T, η, ϕ, m)



IMPLICIT QUANTILE NETWORKS ARCHITECTURE

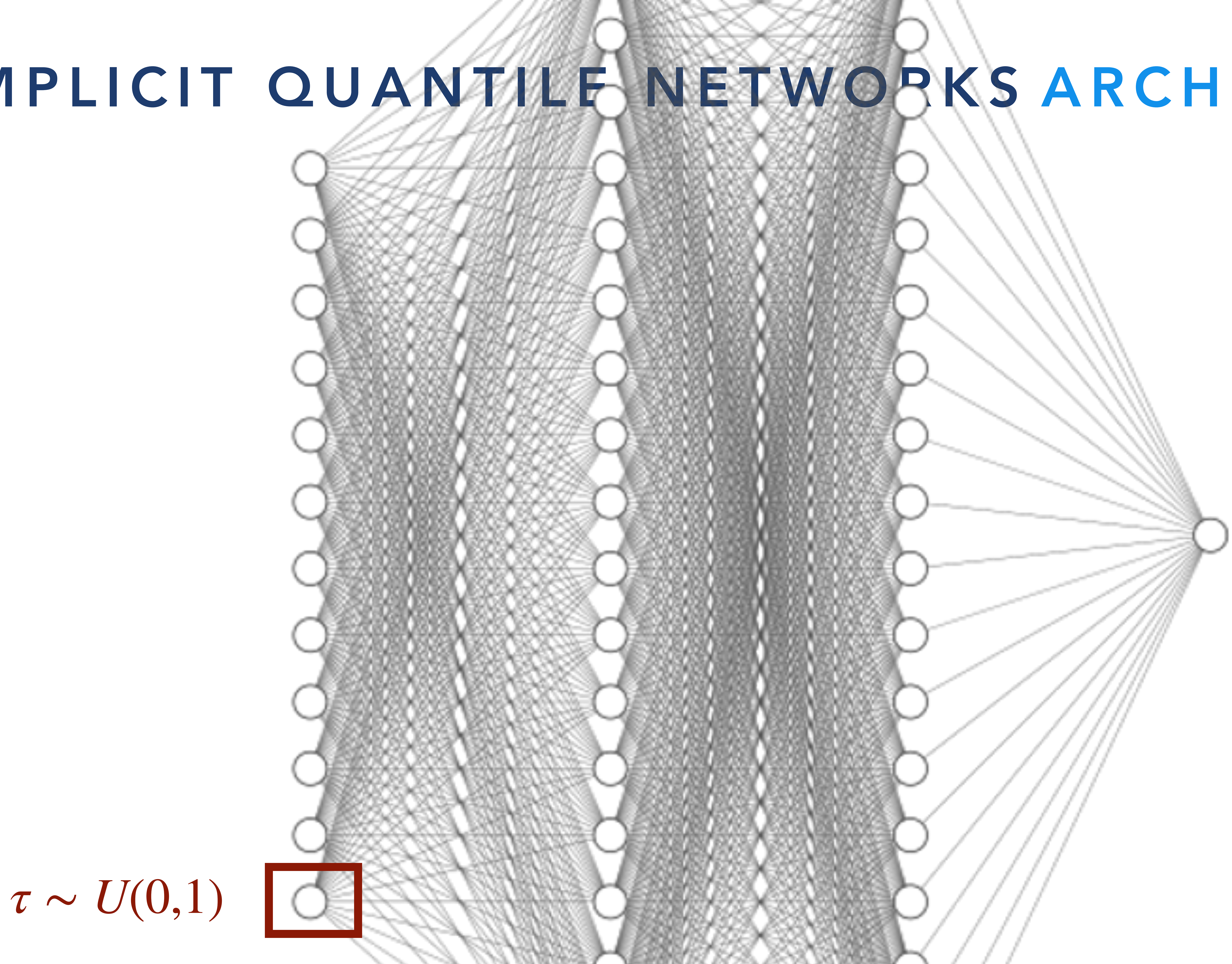


IMPLICIT QUANTILE NETWORKS ARCHITECTURE

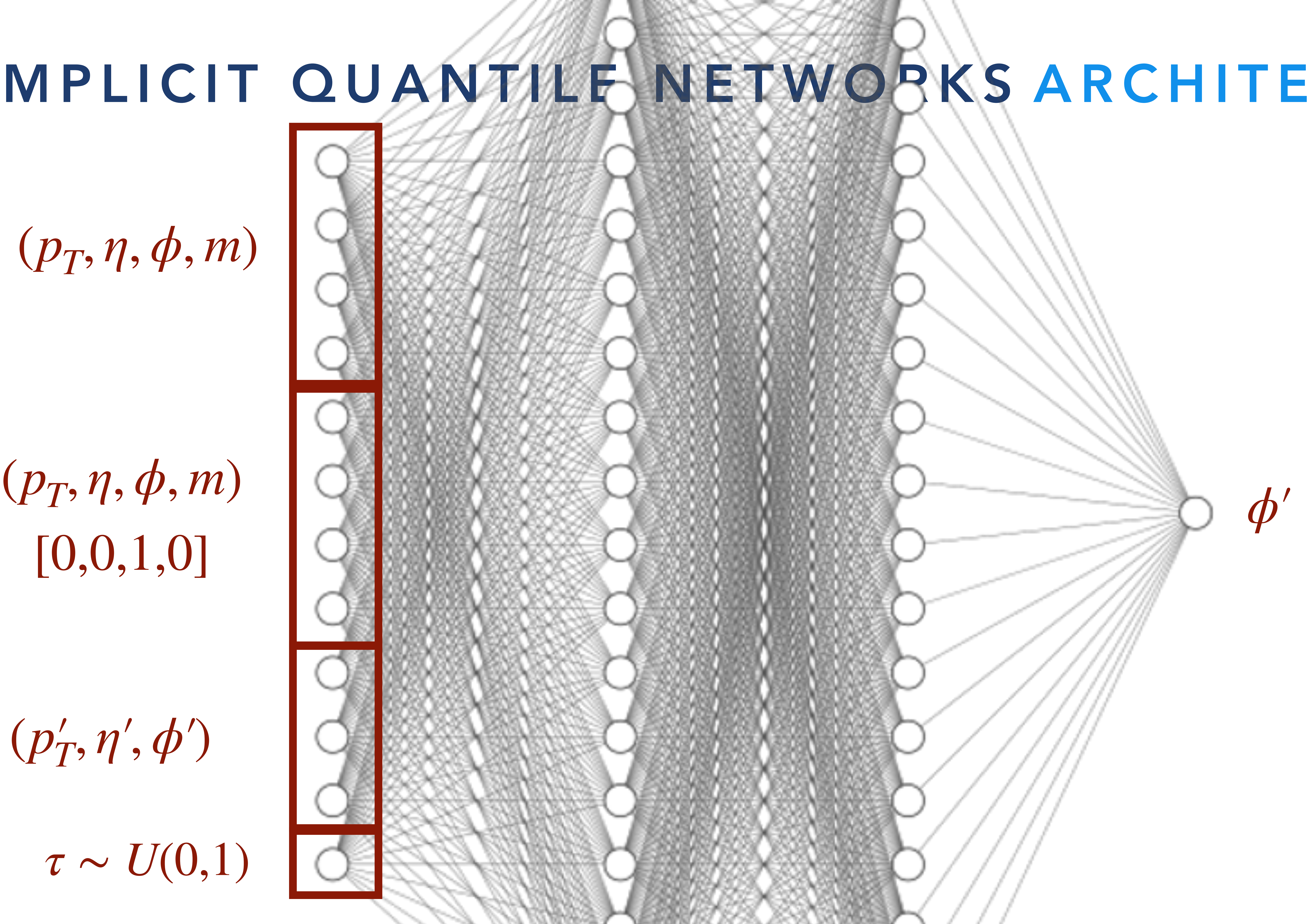


(p_T', η', ϕ')

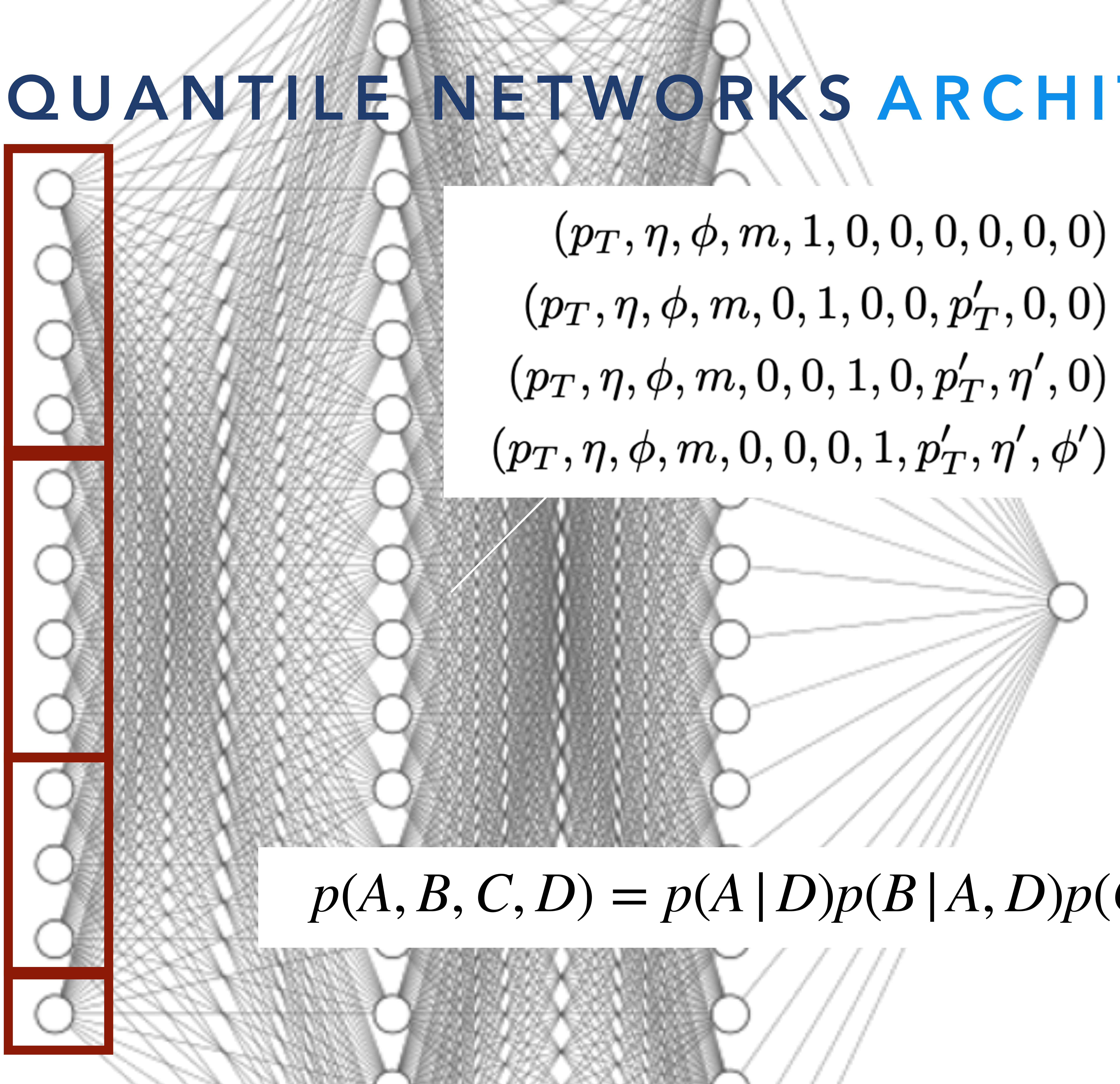
IMPLICIT QUANTILE NETWORKS ARCHITECTURE



IMPLICIT QUANTILE NETWORKS ARCHITECTURE



IMPLICIT QUANTILE NETWORKS ARCHITECTURE

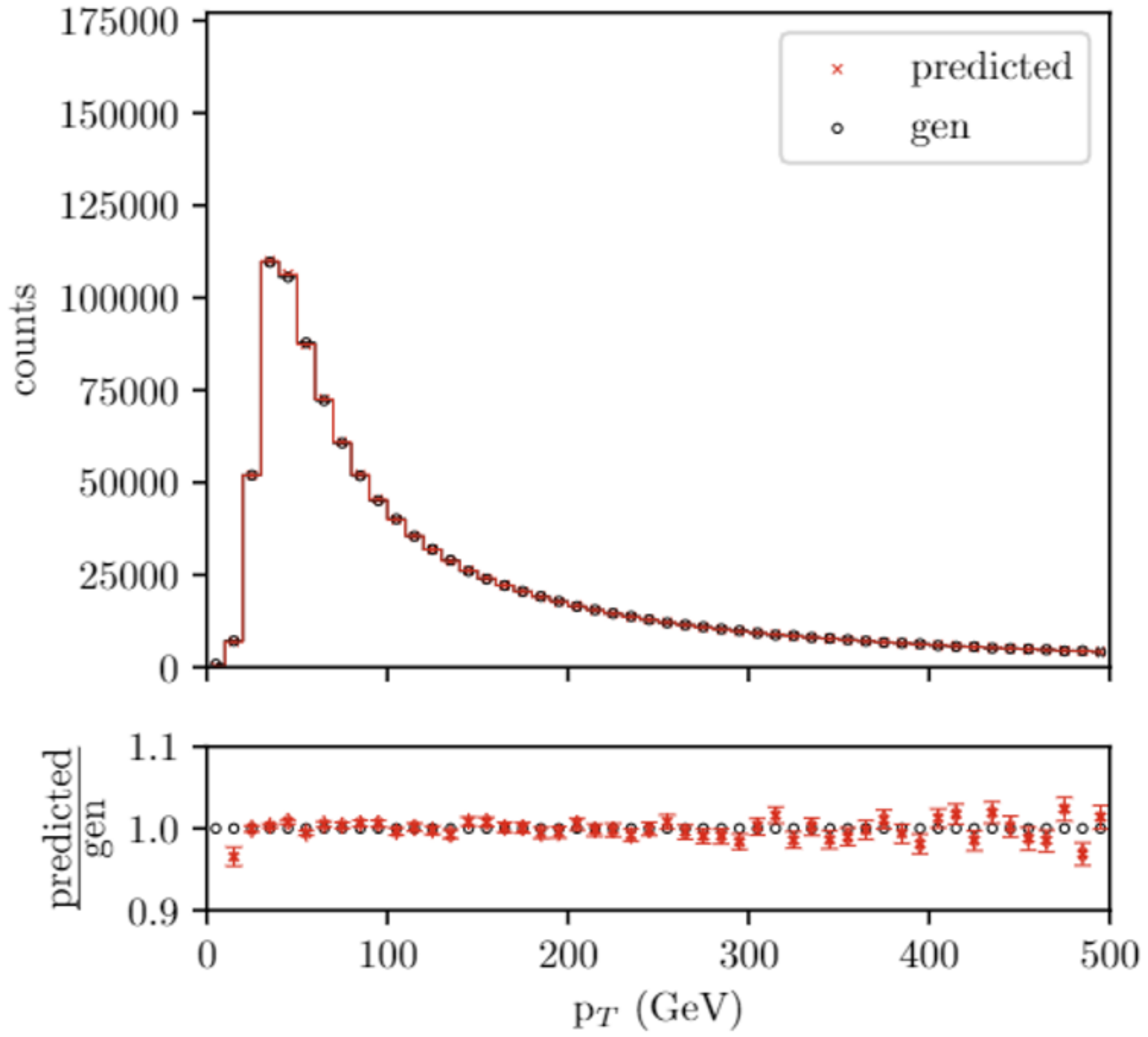


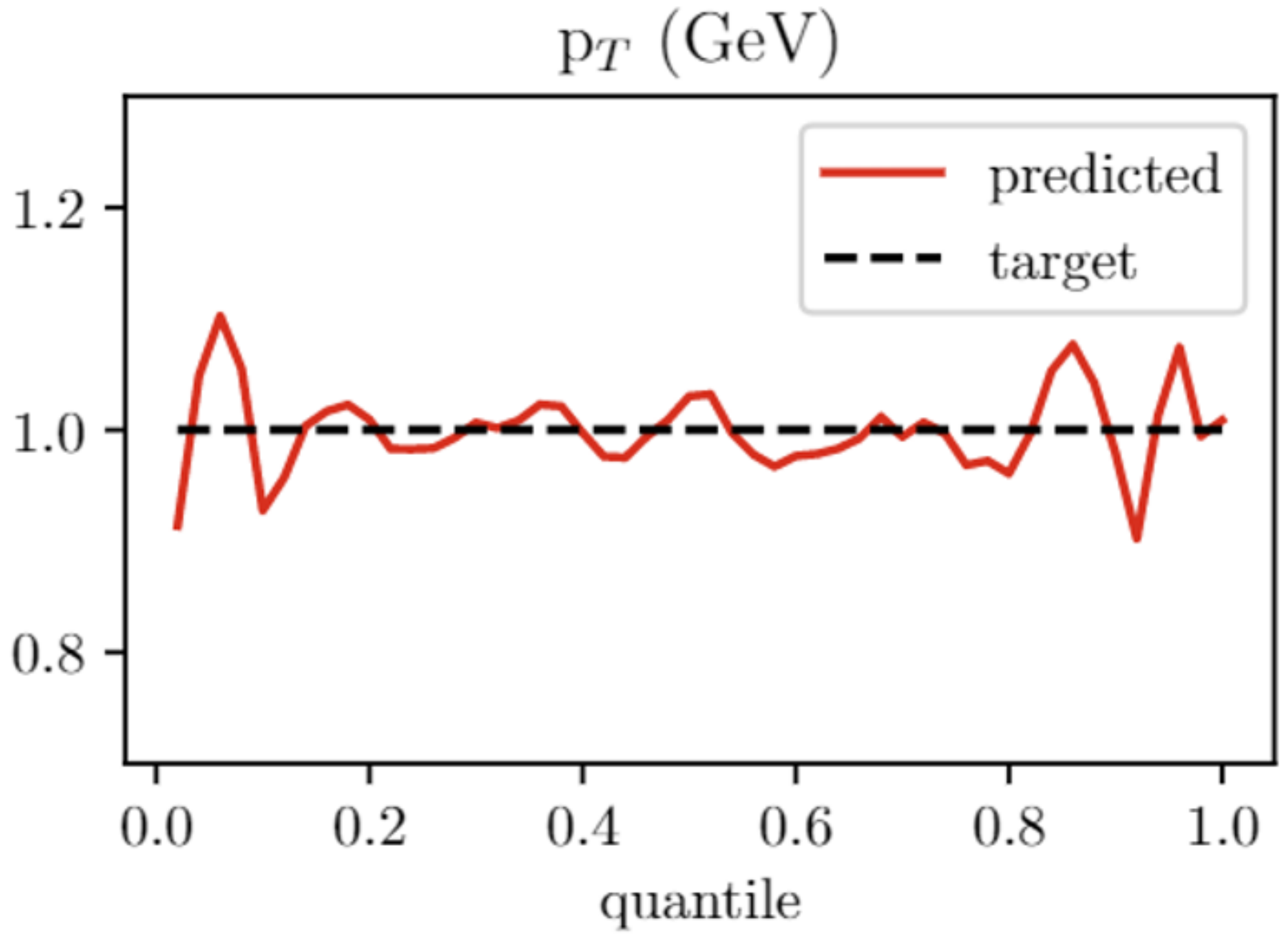
$$
\begin{aligned}
(p_T, \eta, \phi, m, 1, 0, 0, 0, 0, 0, 0) &\to (p'_T),\\
(p_T, \eta, \phi, m, 0, 1, 0, 0, p'_T, 0, 0) &\to (\eta'),\\
(p_T, \eta, \phi, m, 0, 0, 1, 0, p'_T, \eta', 0) &\to (\phi'),\\
(p_T, \eta, \phi, m, 0, 0, 0, 1, p'_T, \eta', \phi') &\to (m'),
\end{aligned}
$$

$$p(A, B, C, D) = p(A \mid D)\,p(B \mid A, D)\,p(C \mid A, B, D)$$

RESULTS JET CORRECTION

RESULTS JET CORRECTION





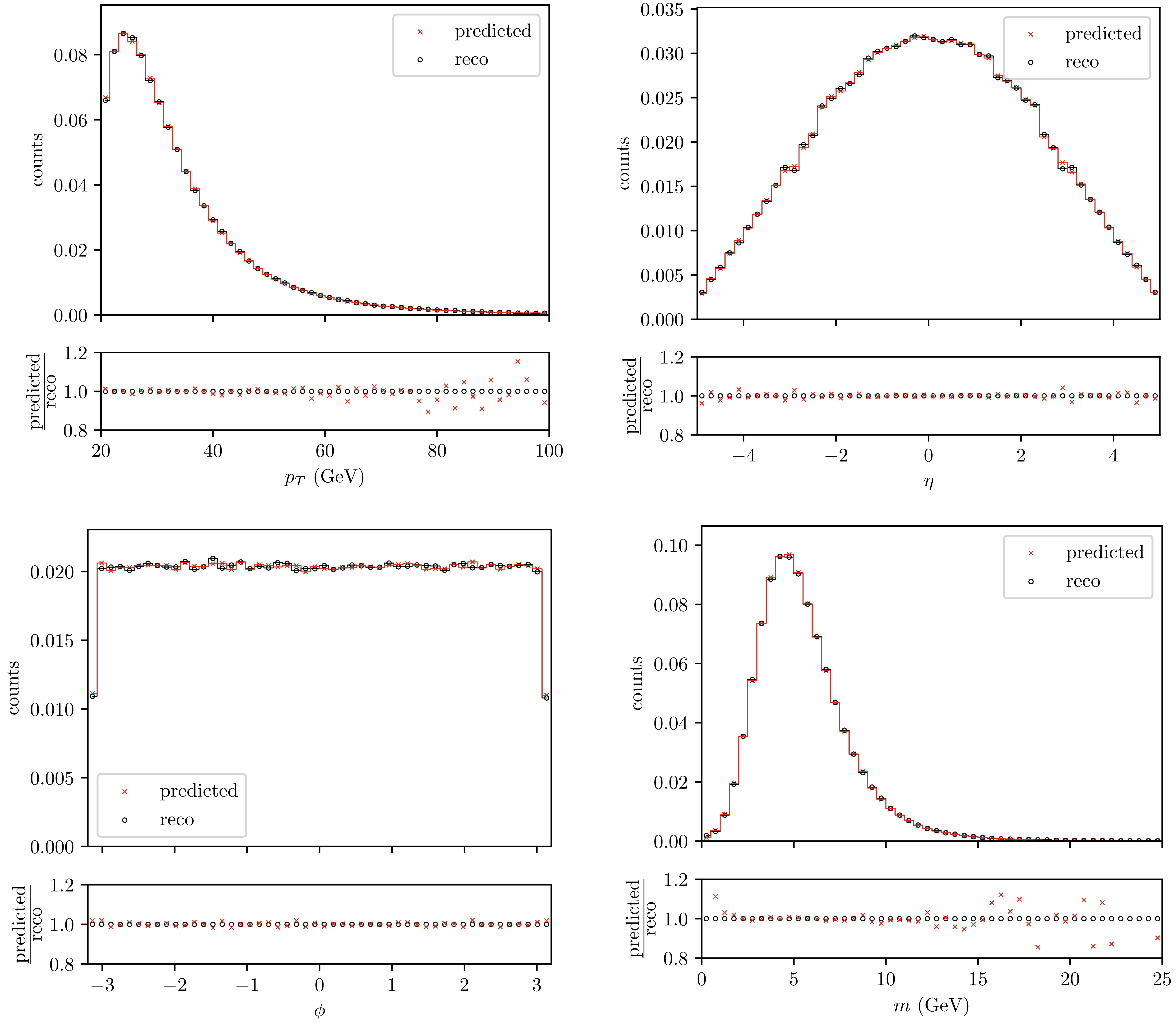
EV GENER

RECO JETS

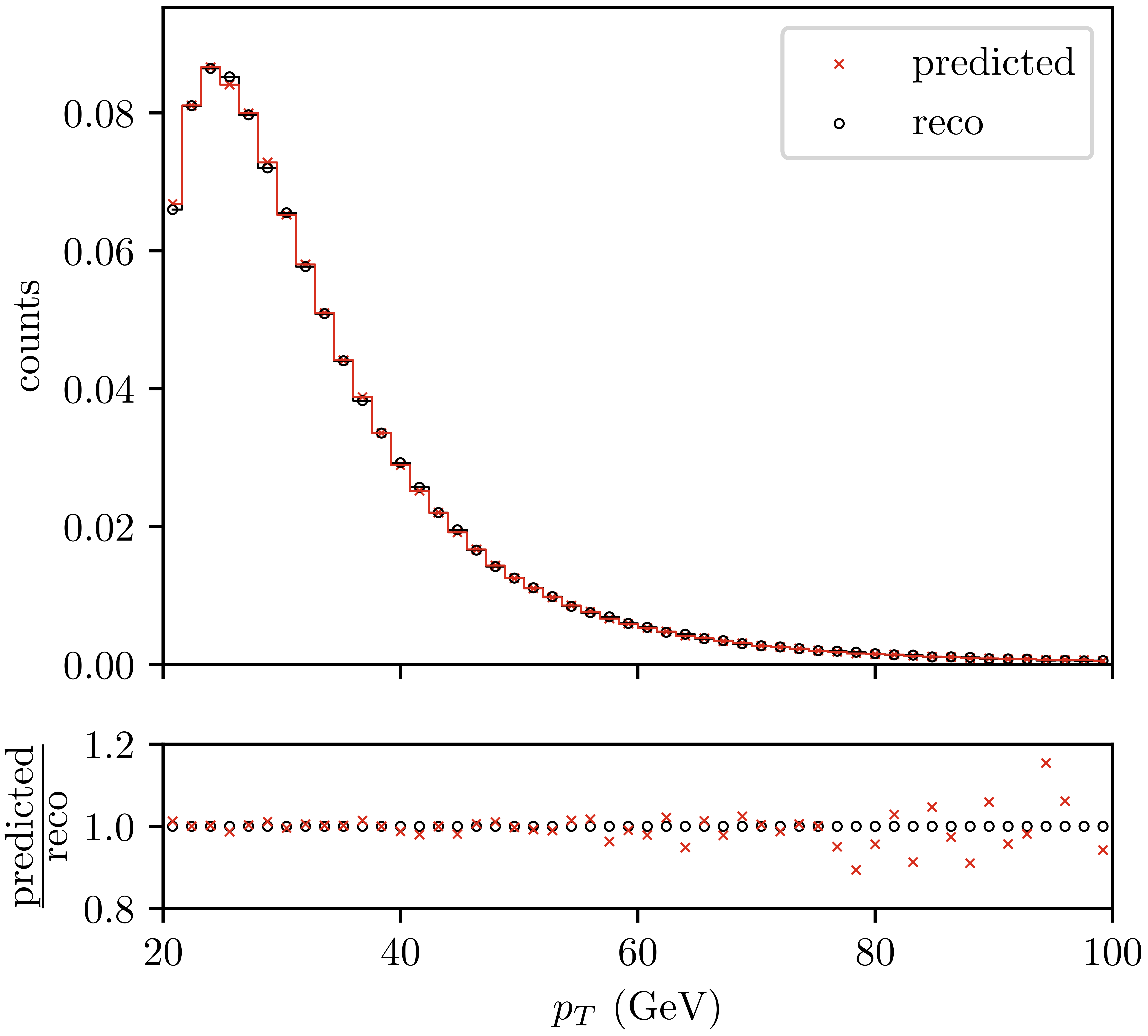
jet simulation

RESULTS JET SIMULATION

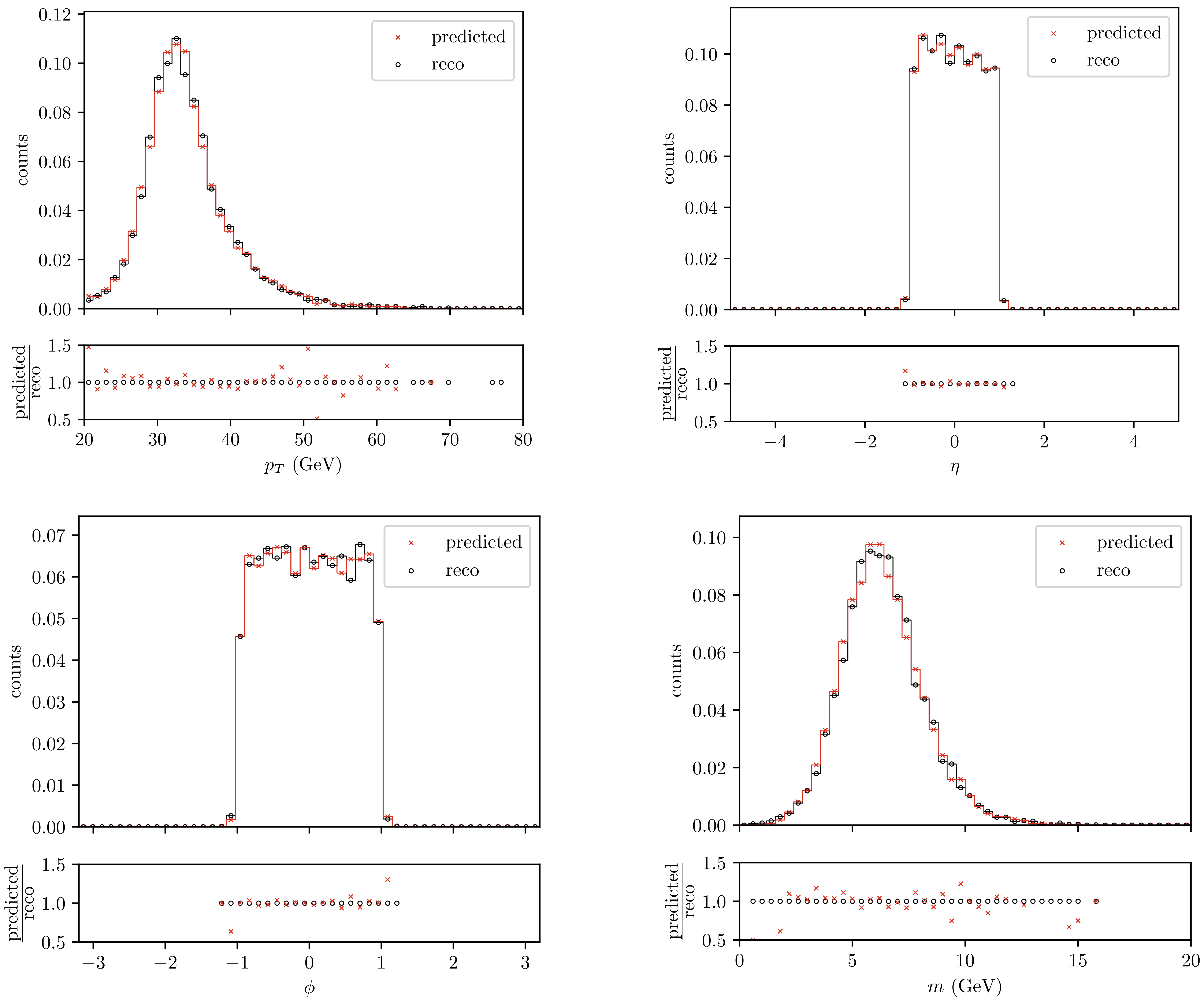
RESULTS JET SIMULATION



RESULTS JET SIMULATION



RESULTS JET SIMULATION: SUBSPACE



THANK YOU!

BRADEN KRONHEIM[1,2], **MICHELLE KUCHERA[1]**, HARRISON PROSPER[3], RAGHU RAMANUJAN[1], ALI AL KADHIM[3]

arXiv:2111.11415

NeurIPS 2021 – Thirty-fifth Workshop on Machine Learning and the Physical Sciences, Dec 2021, Vancouver, Canada

THIS WORK WAS SUPPORTED BY THE NATIONAL SCIENCE FOUNDATION UNDER COOPERATIVE AGREEMENT OAC-1836650.1 AND THE NATIONAL SCIENCE FOUNDATION UNDER GRANT NO. 2012865.

ALPhA
DAVIDSON COLLEGE

iris hep
Institute for Research & Innovation in Software for High Energy Physics

CERN

CMS
Compact Muon Solenoid

DAVIDSON