UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

ML4Jets, 07.11.2023

Diffusion Models

for LHC event generation

SPONSORED BY THE

Federal Ministry of Education and Research

Anja Butter, Nathan Huetsch, Sofia Palacios Schweitzer, Tilman Plehn, Peter Sorrenson, Jonas Spinner
arXiv: 2305.10475

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

ML4Jets, 07.11.2023

Uncertainty-Aware Diffusion Models

for LHC event generation

SPONSORED BY THE
Federal Ministry of Education and Research

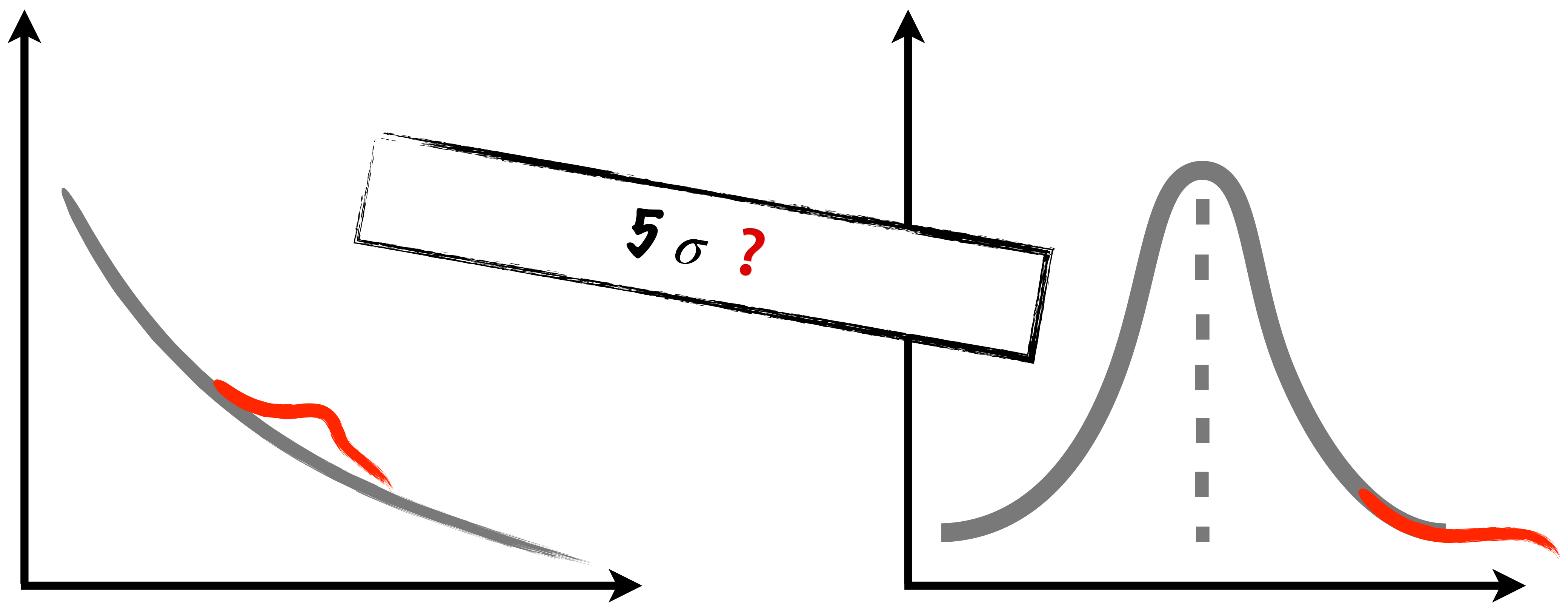
Anja Butter, Nathan Huetsch, Sofia Palacios Schweitzer, Tilman Plehn, Peter Sorrenson, Jonas Spinner
arXiv: 2305.10475

Why Event Generation?

Vast amount of data collected by collider experiments

Standard Model is probed

Theoretical predictions (simulation) need to match experimental statistics



Why ML Event Generation?

After high luminosity runs → ~ 20 times as much data

Theoretical predictions need to be even more precise (include higher correction terms)

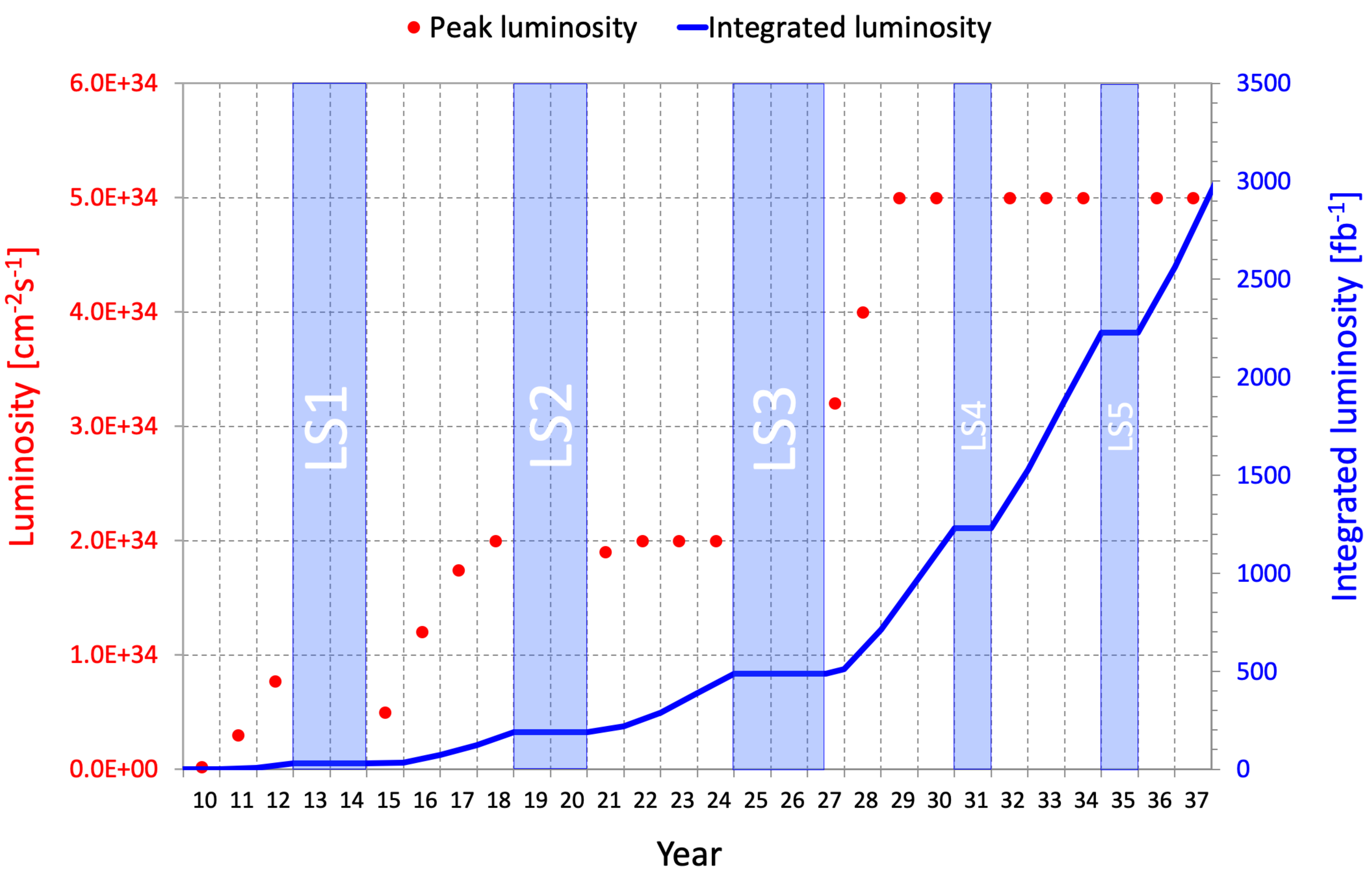


Figure from https://web.archive.org/web/20220706170326/https://lhc-commissioning.web.cern.ch/schedule/images/LHC-nominal-lumi-projection.png

Why ML Event Generation?

After high luminosity runs → ~ 20 times as much data

Theoretical predictions needs to be even more precise (include higher correction terms)

But: Currently computationally expensive

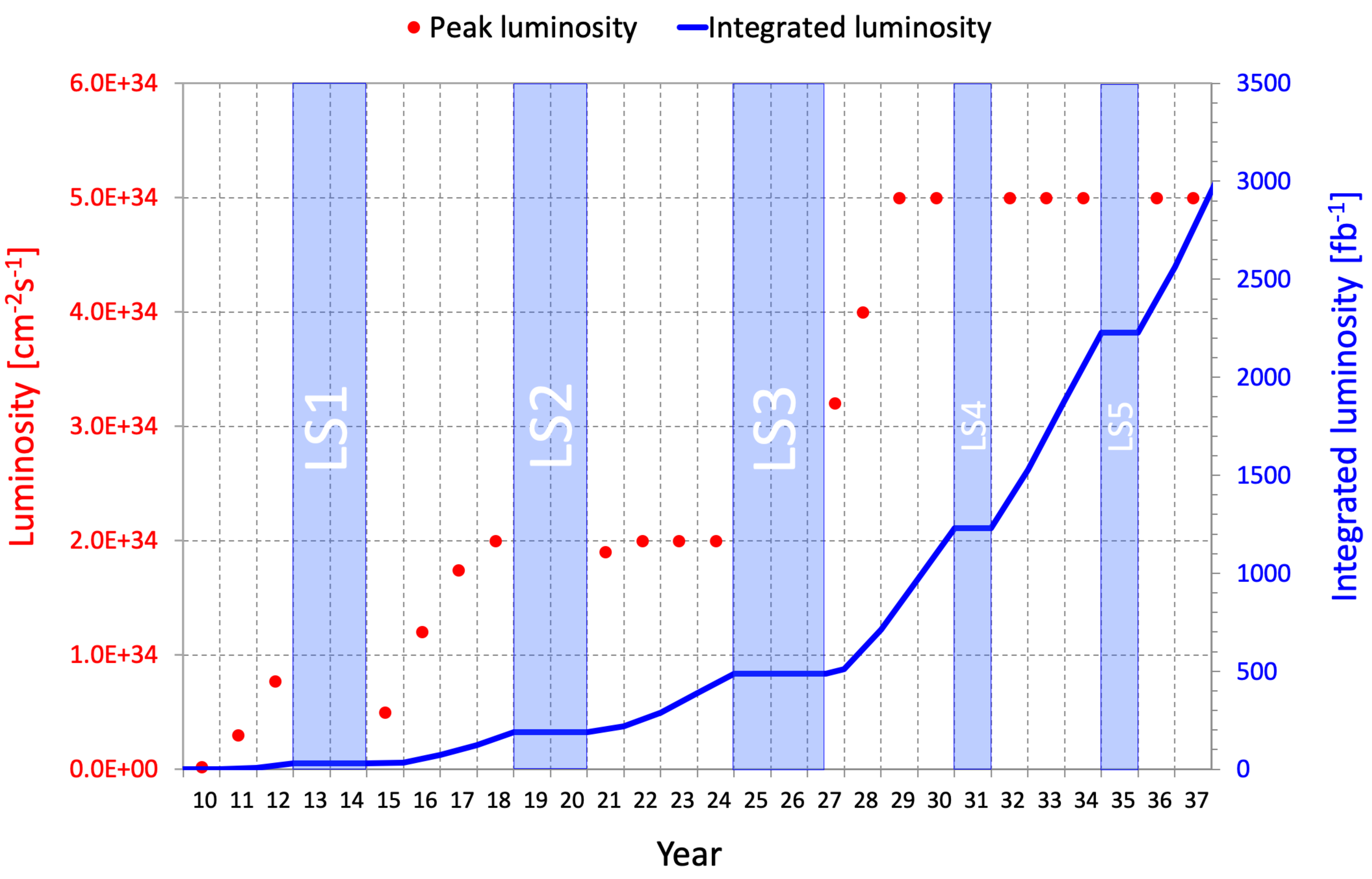
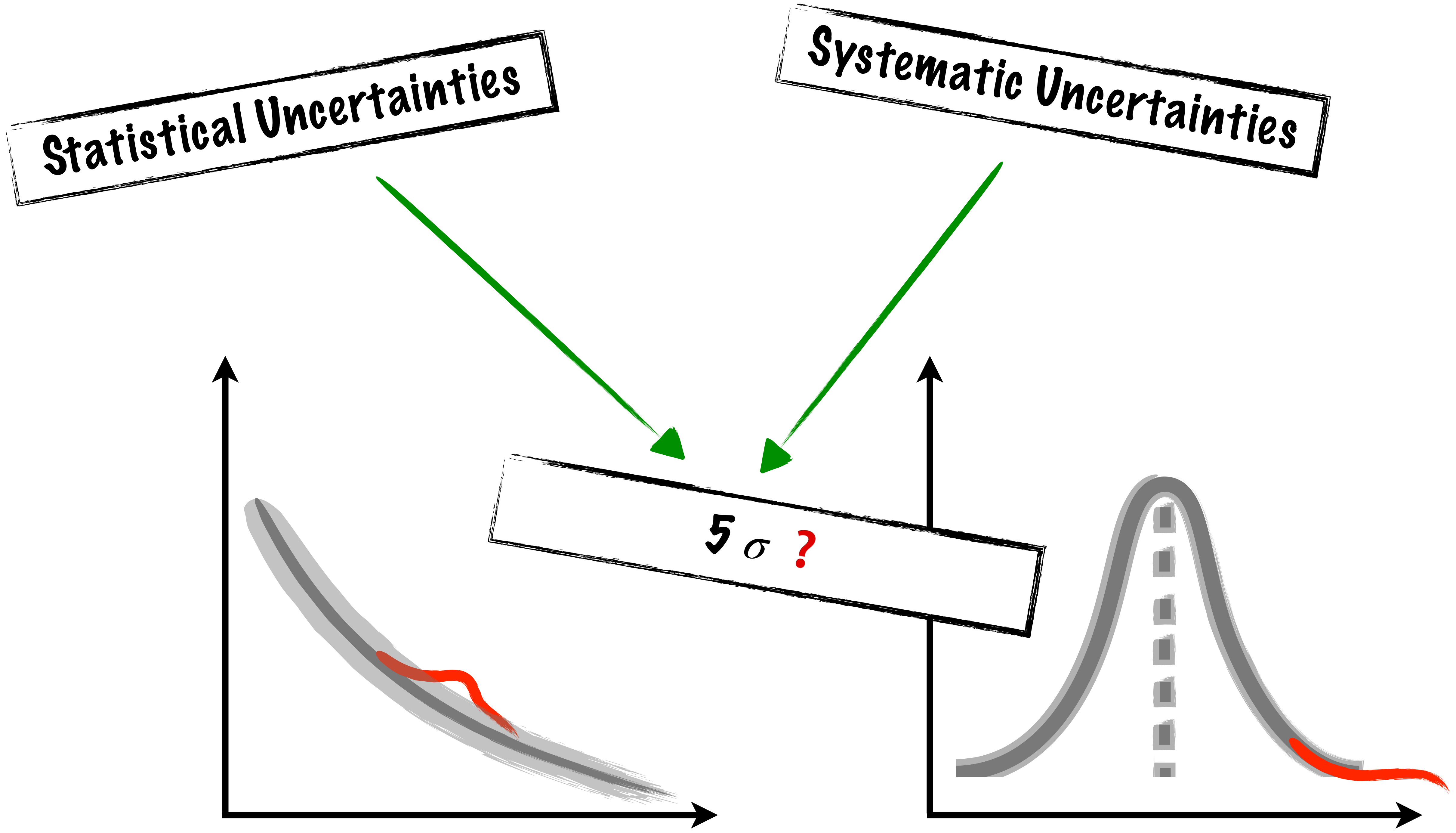


Figure from https://web.archive.org/web/20220706170326/https://lhc-commissioning.web.cern.ch/schedule/images/LHC-nominal-lumi-projection.png

Raising Awareness for Uncertainties

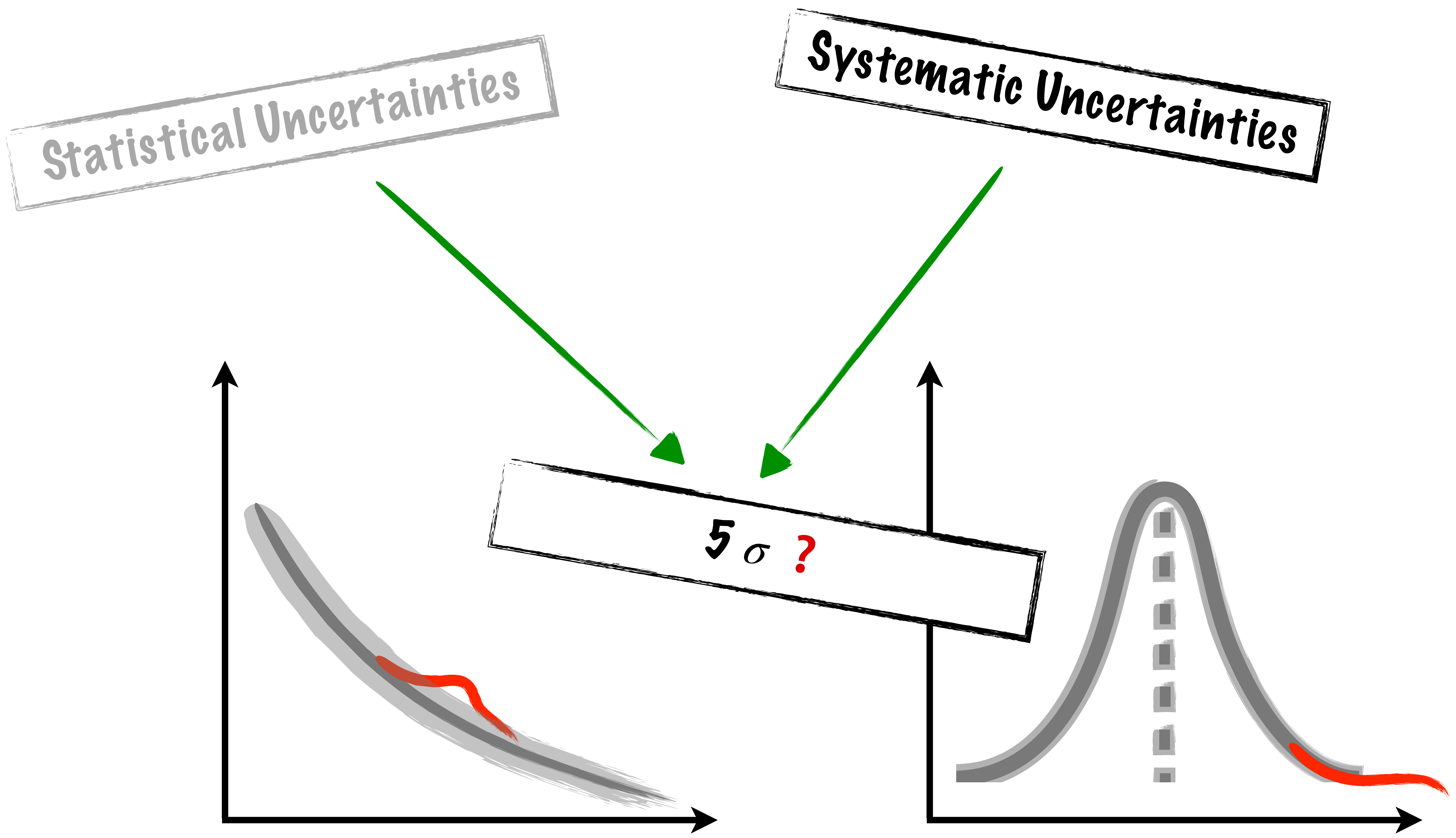
Being precise = estimating uncertainties



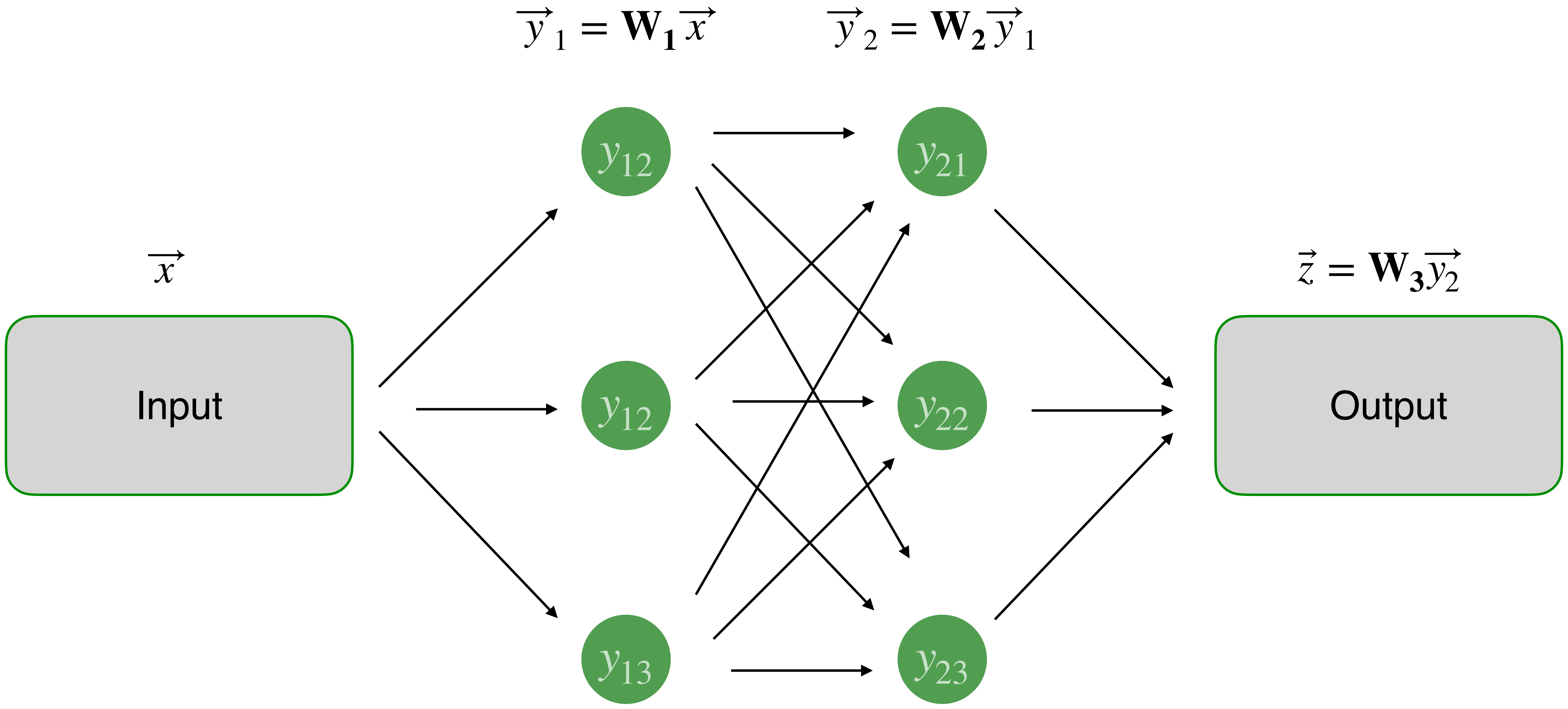
Raising Awareness for Uncertainties

Being precise = estimating uncertainties

How can we account for network uncertainties?

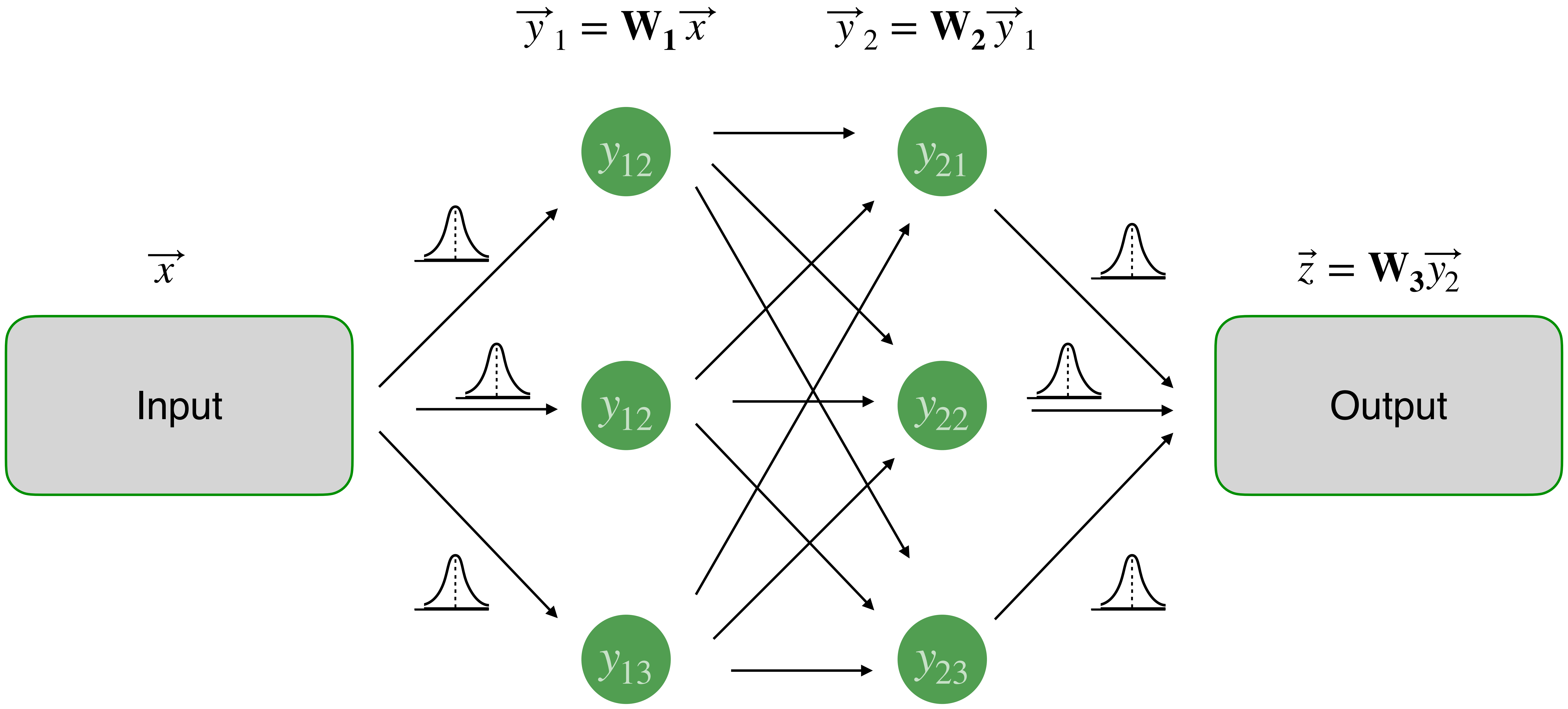


What about uncertainties?



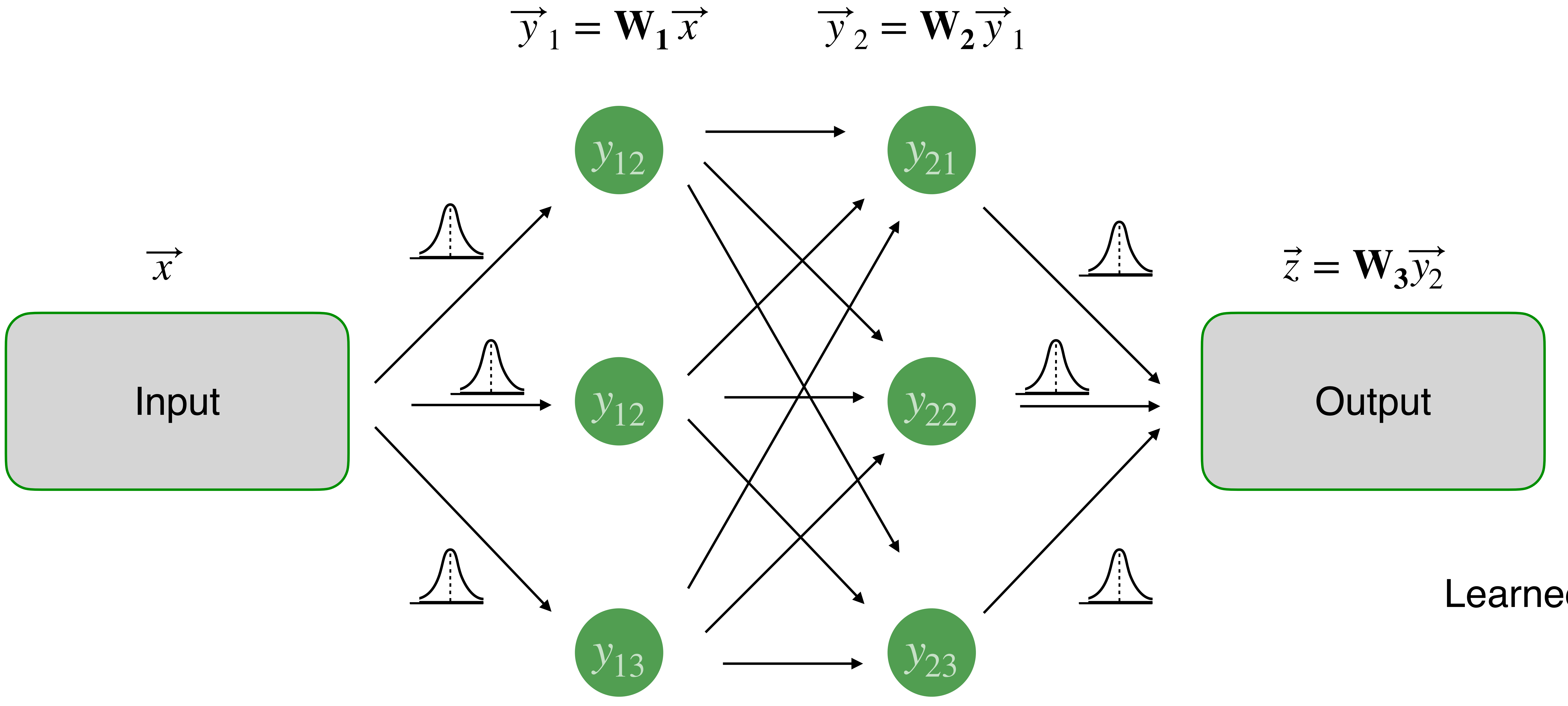
Once training is done: $\mathbf{W_1}, \mathbf{W_2}, \mathbf{W_3}$ fixed (*"Network output is deterministic"*)

What about uncertainties?



Bayesianization: We draw each entry from $\mathbf{W_1}, \mathbf{W_2}, \mathbf{W_3}$ from distribution $q(\theta \,|\, \mu_\phi, \sigma_\phi)$

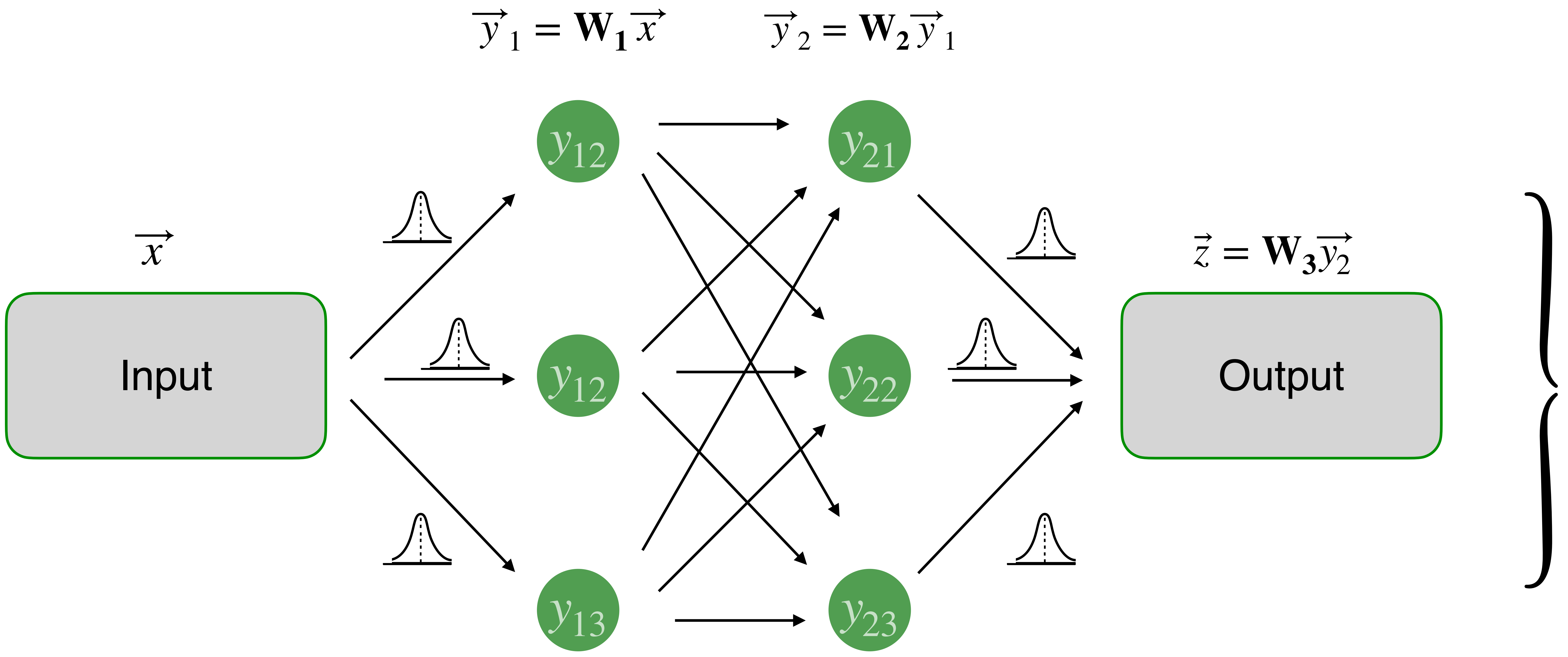
What about uncertainties?



Learned by network

Bayesianization: We draw each entry from $\mathbf{W_1}, \mathbf{W_2}, \mathbf{W_3}$ from distribution $q(\theta \,|\, \mu_\phi, \sigma_\phi)$

What about uncertainties?



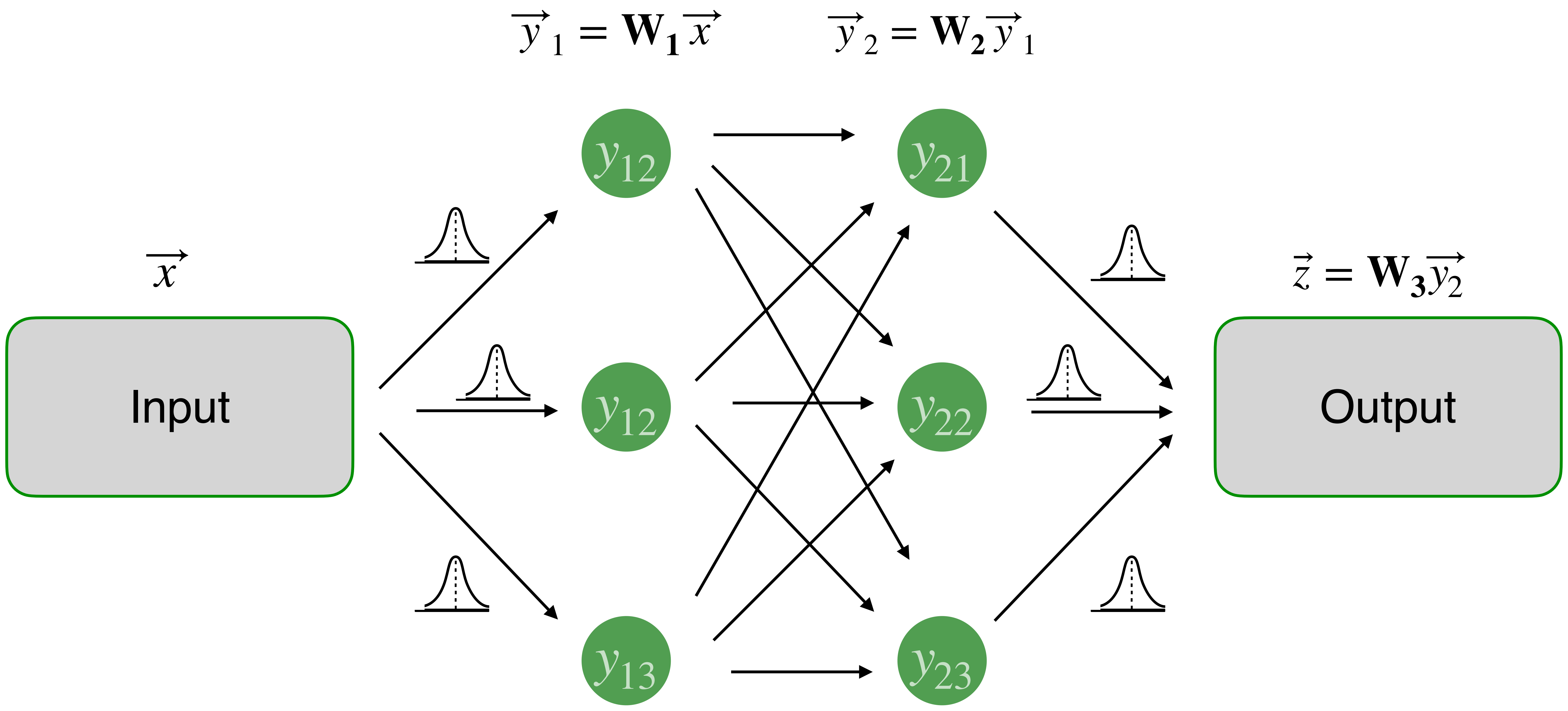
$$\langle \vec{z} \rangle = \frac{1}{N} \sum_i \vec{z}_i$$

$$\sigma^2_{pred} = \frac{1}{N} \sum_{i=1}^{N} \left(\langle \vec{z} \rangle - \vec{z}_i \right)^2$$

How to Bayesianize

1. Replace each linear layer with a *Bayesian* layer

2. Add additional regularisation term to likelihood loss



Diffusion Models (DDPM)

Phase Space

$x_0 \sim p(x_0)$

Latent Space

$x_T \sim p(x_T)(= \mathcal{N}(0,1))$

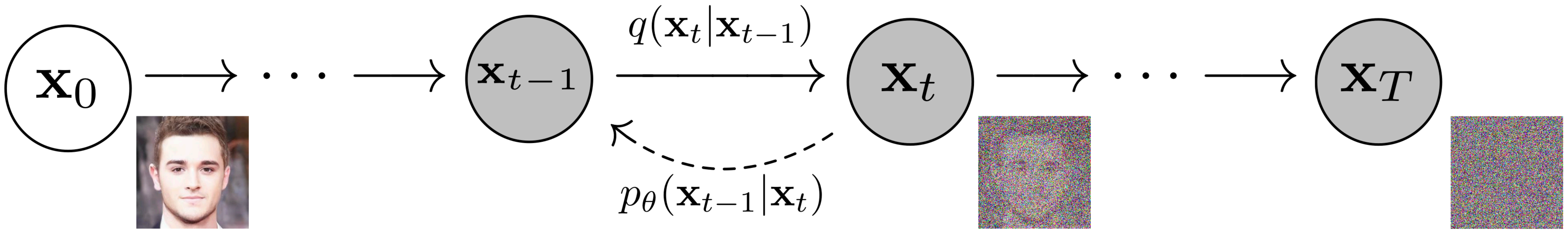
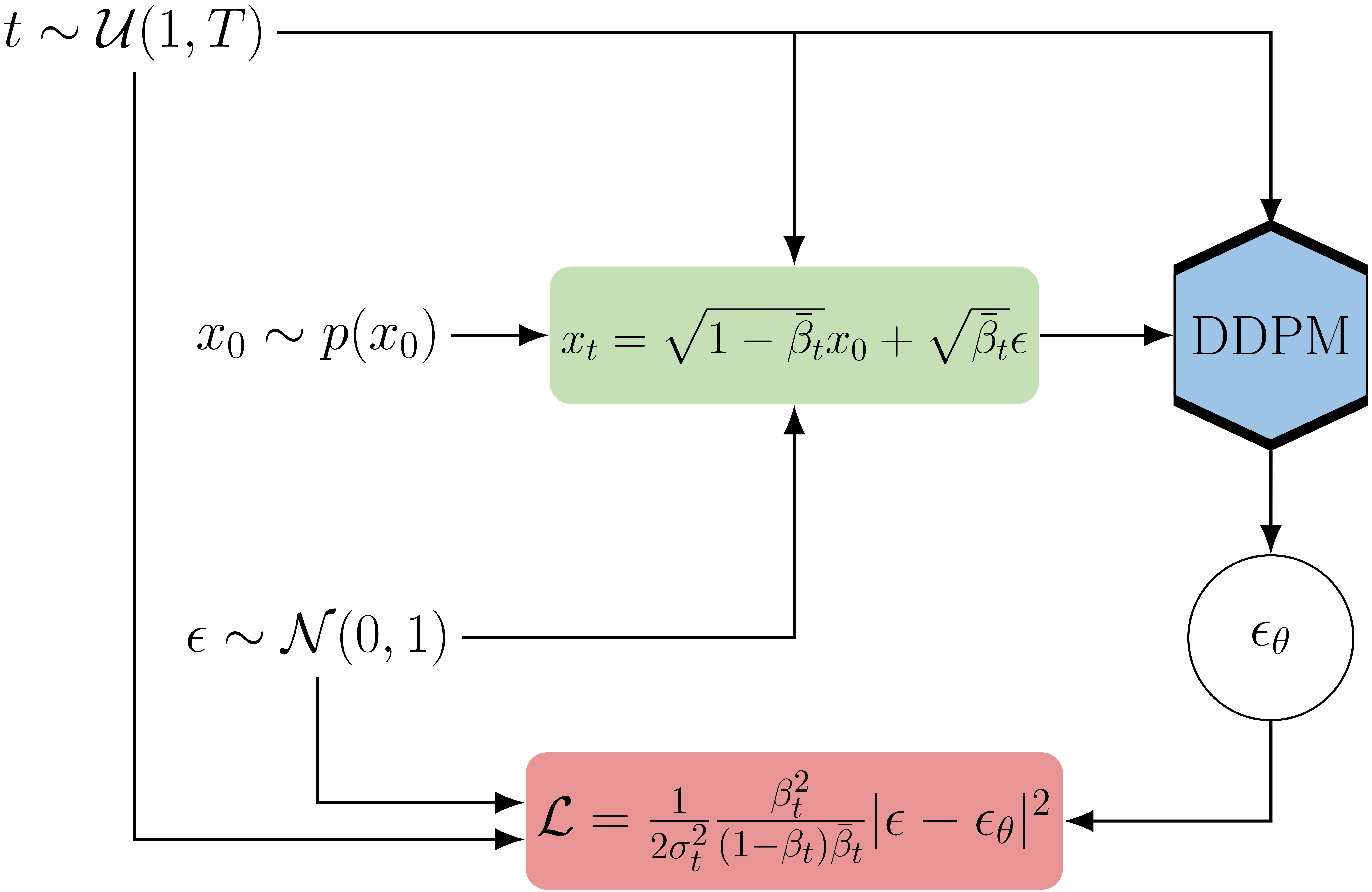
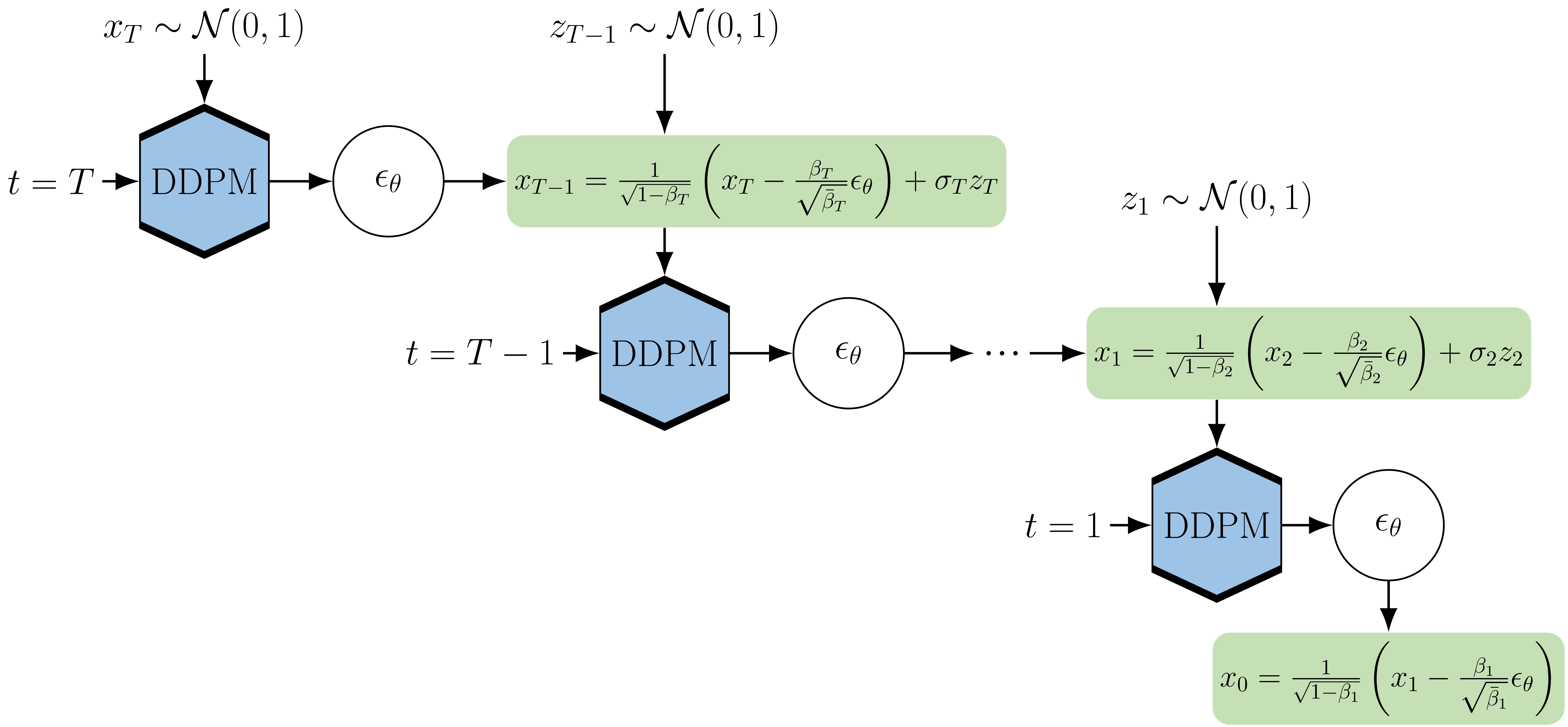


Figure from J.Ho et al.: arXiv:2006.11239

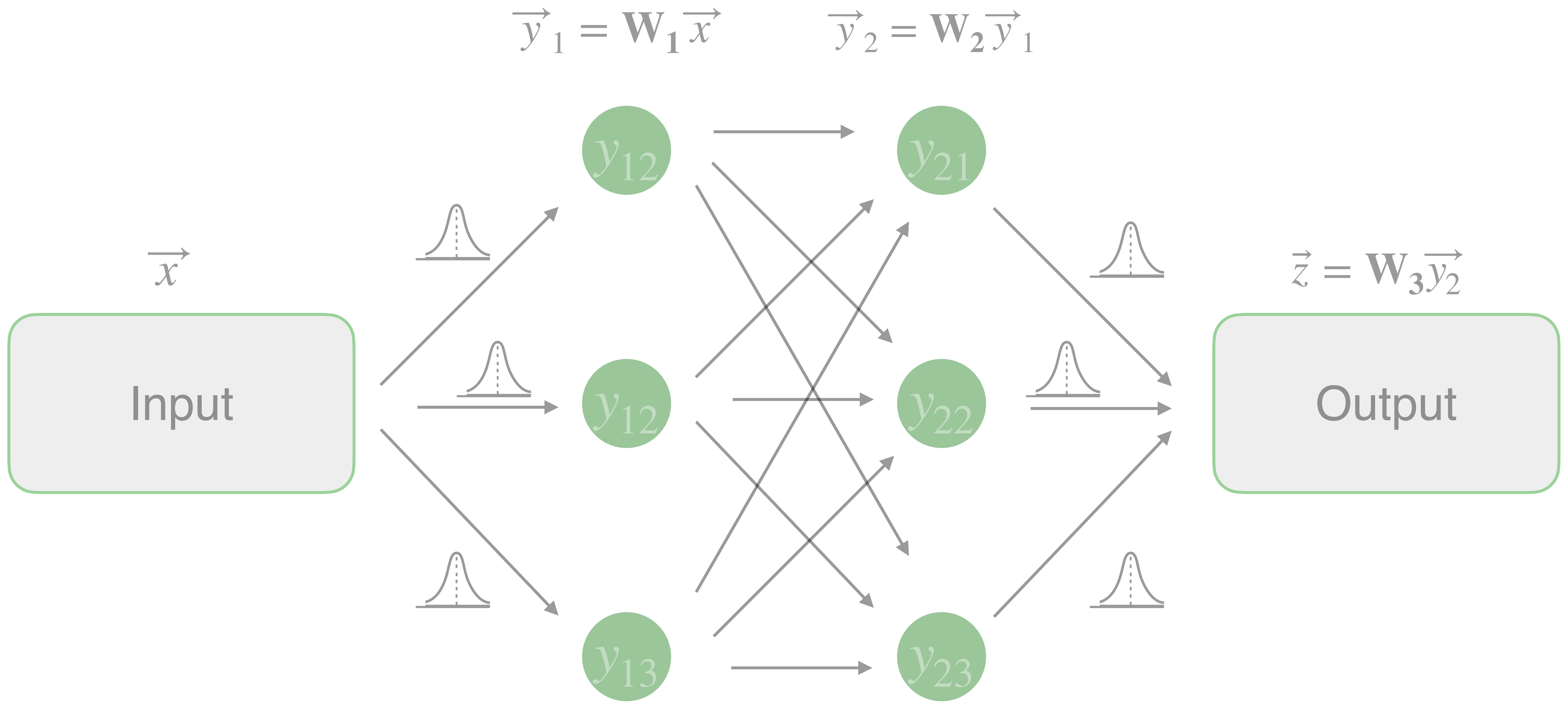
Diffusion Models (DDPM)



Diffusion Models (DDPM)



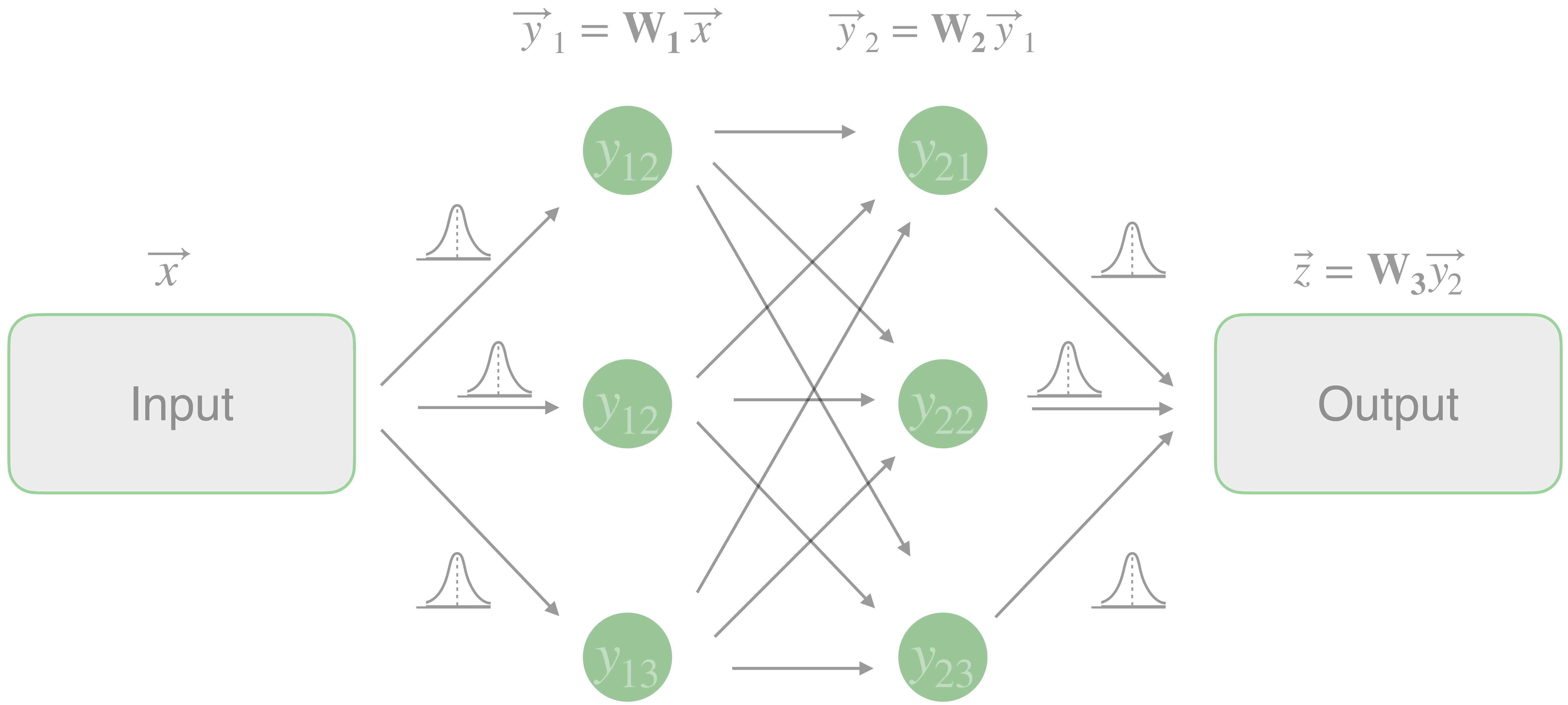
How to Bayesianize - DDPM



Likelihood Loss

$$\mathscr{L}_{DDPM} = -\log p_\theta \approx \frac{1}{2\sigma_t^2}\frac{\beta_t^2}{(1-\beta_t)\bar{\beta}_t}|\epsilon(t) - \epsilon_\theta(t)|^2$$

How to Bayesianize - DDPM



Likelihood Loss

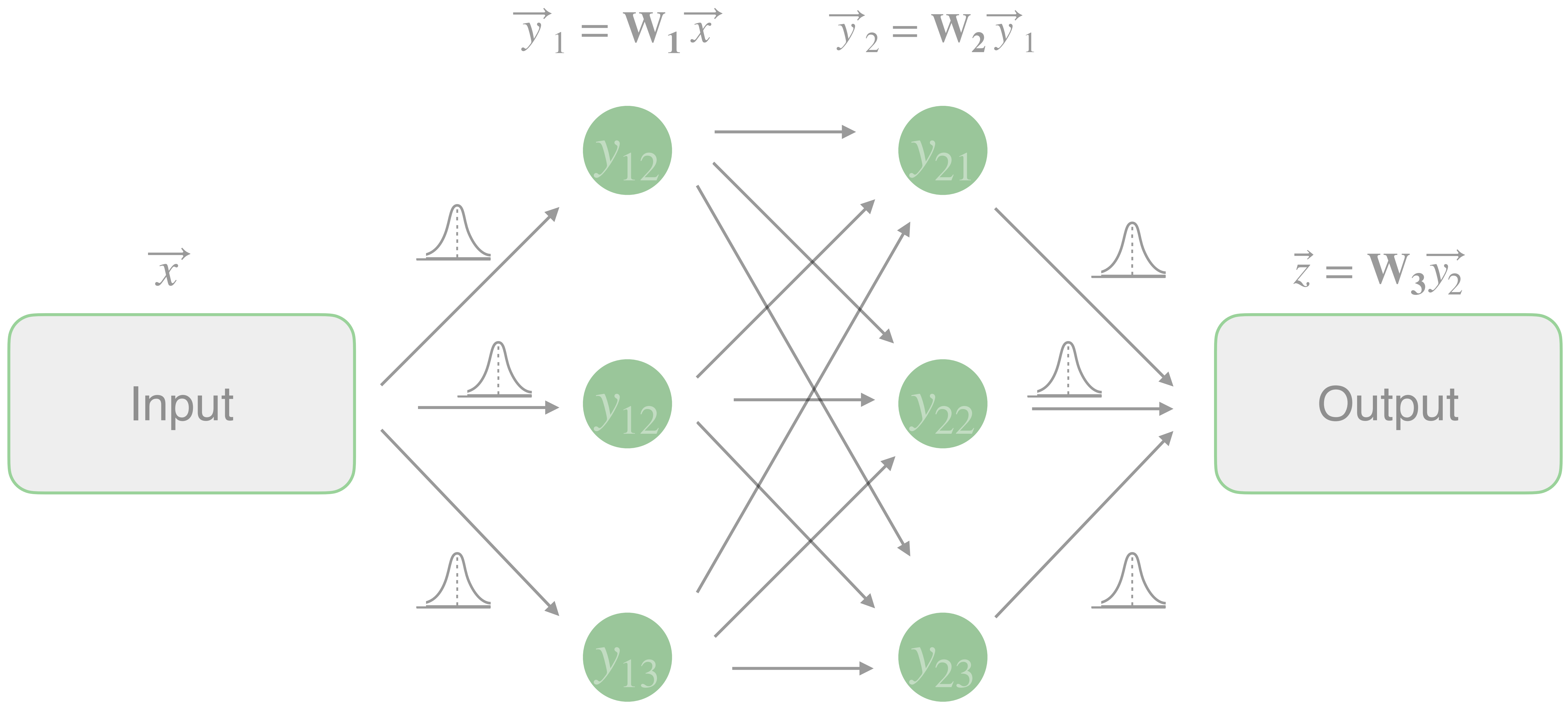
$$\mathscr{L}_{B-DDPM} = \left\langle \frac{1}{2\sigma_t^2} \frac{\beta_t^2}{(1-\beta_t)\bar{\beta}_t} |\epsilon(t) - \epsilon_\theta(t)|^2 \right\rangle_{\theta \sim q(\theta)}$$

Prior weight distribution

$$+ \quad \mathscr{D}_{KL}(q(\theta), p(\theta))$$

Learned weight distribution

How to Bayesianize - DDPM



Likelihood Loss

$$\mathcal{L}_{B-DDPM} = T \cdot \left\langle \frac{1}{2\sigma_t^2} \frac{\beta_t^2}{(1-\beta_t)\bar{\beta}_t} |\epsilon(t) - \epsilon_\theta(t)|^2 \right\rangle_{\theta \sim q(\theta), t \sim \mathcal{U}(0,T)}$$

Prior weight distribution

$$+ \quad \mathcal{D}_{KL}(q(\theta), p(\theta))$$

Learned weight distribution

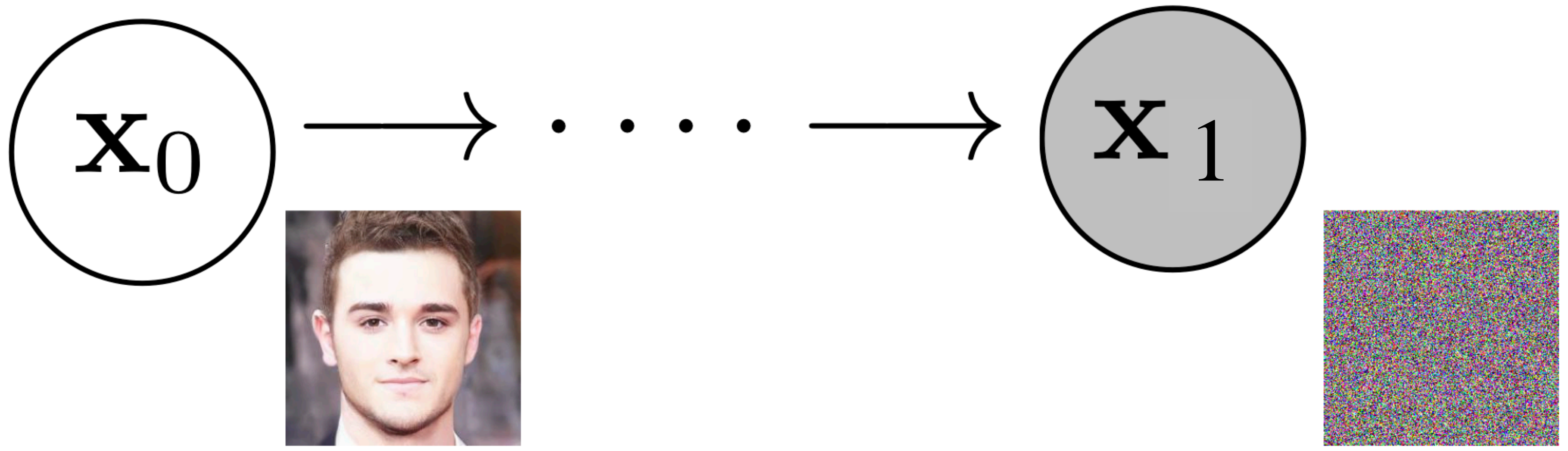
Diffusion Models (CFM)

Phase Space

$x_0 \sim p(x_0)$

Latent Space

$x_1 \sim p(x_1)(= \mathcal{N}(0,1))$

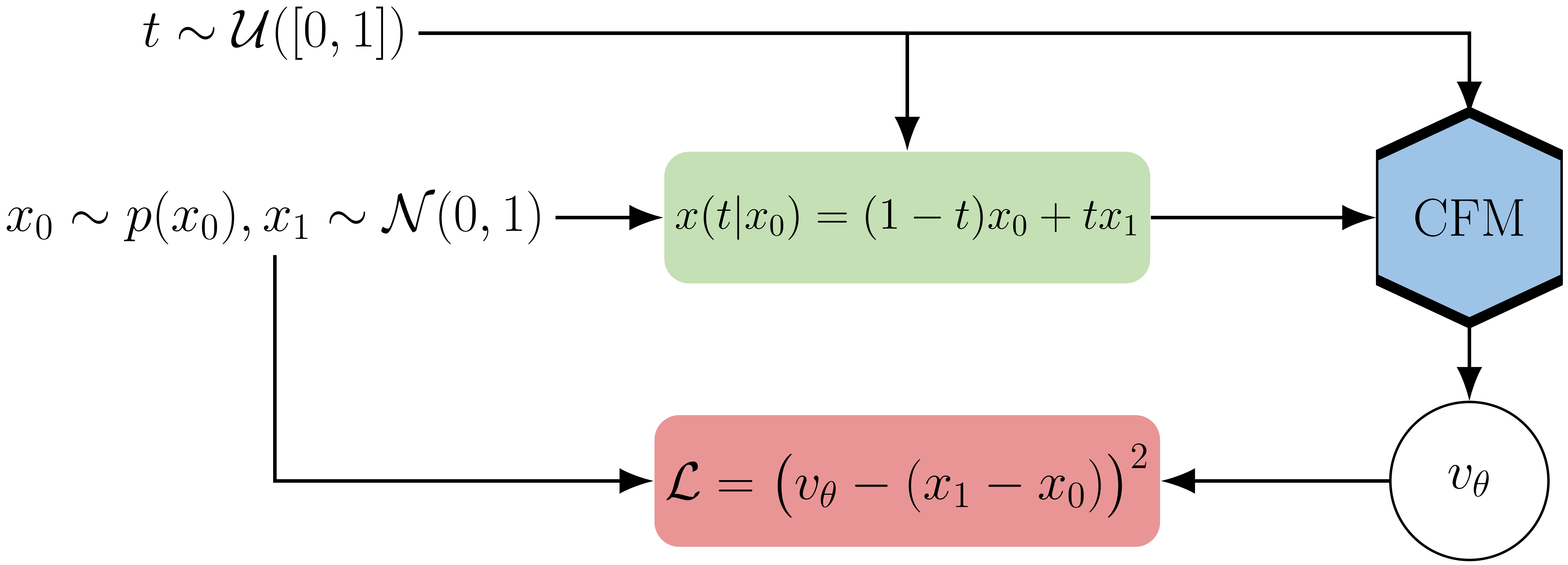


Continuous time evolution **t**

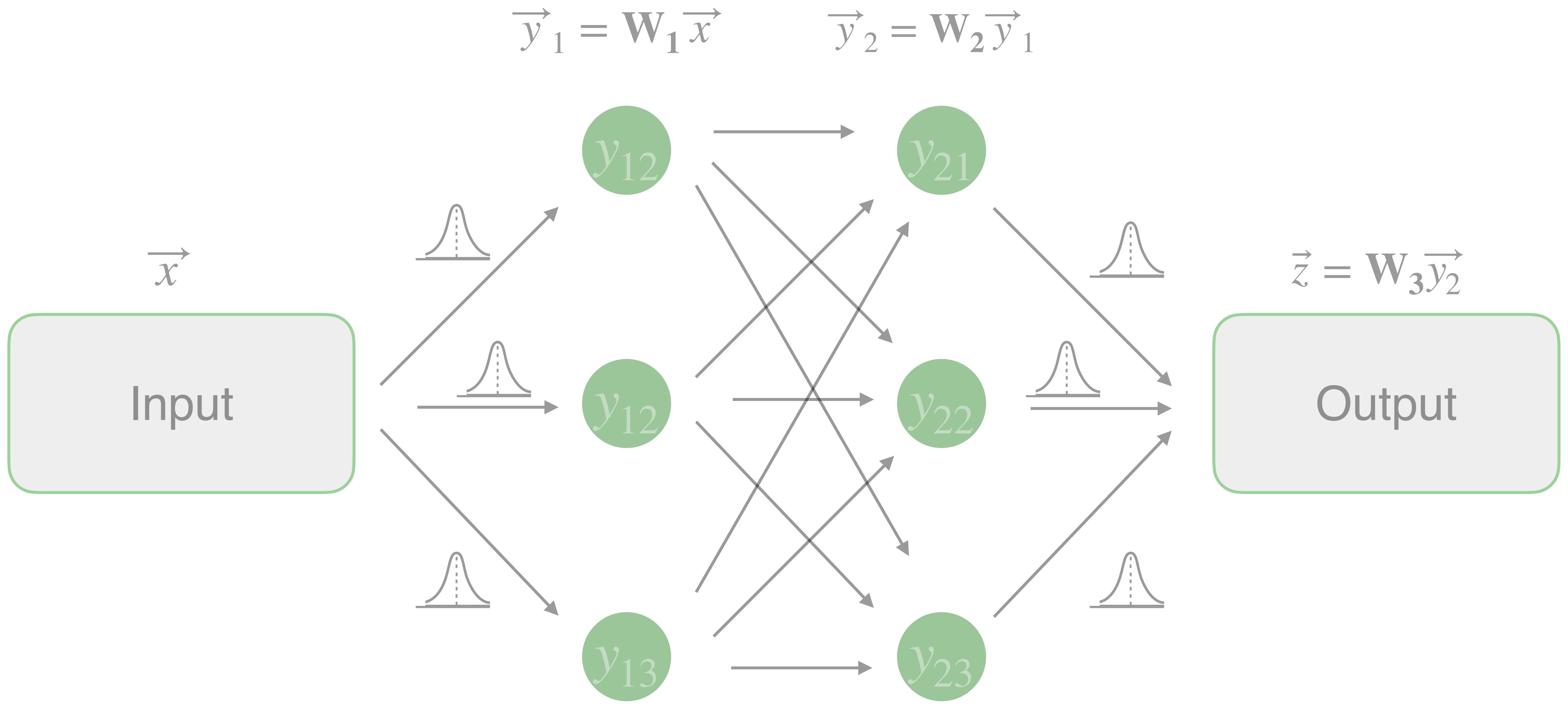
Evolution governed by $v \equiv \frac{dx}{dt}$

Figure from J.Ho et al.: arXiv:2006.11239

Diffusion Models (CFM)



How to Bayesianize - CFM

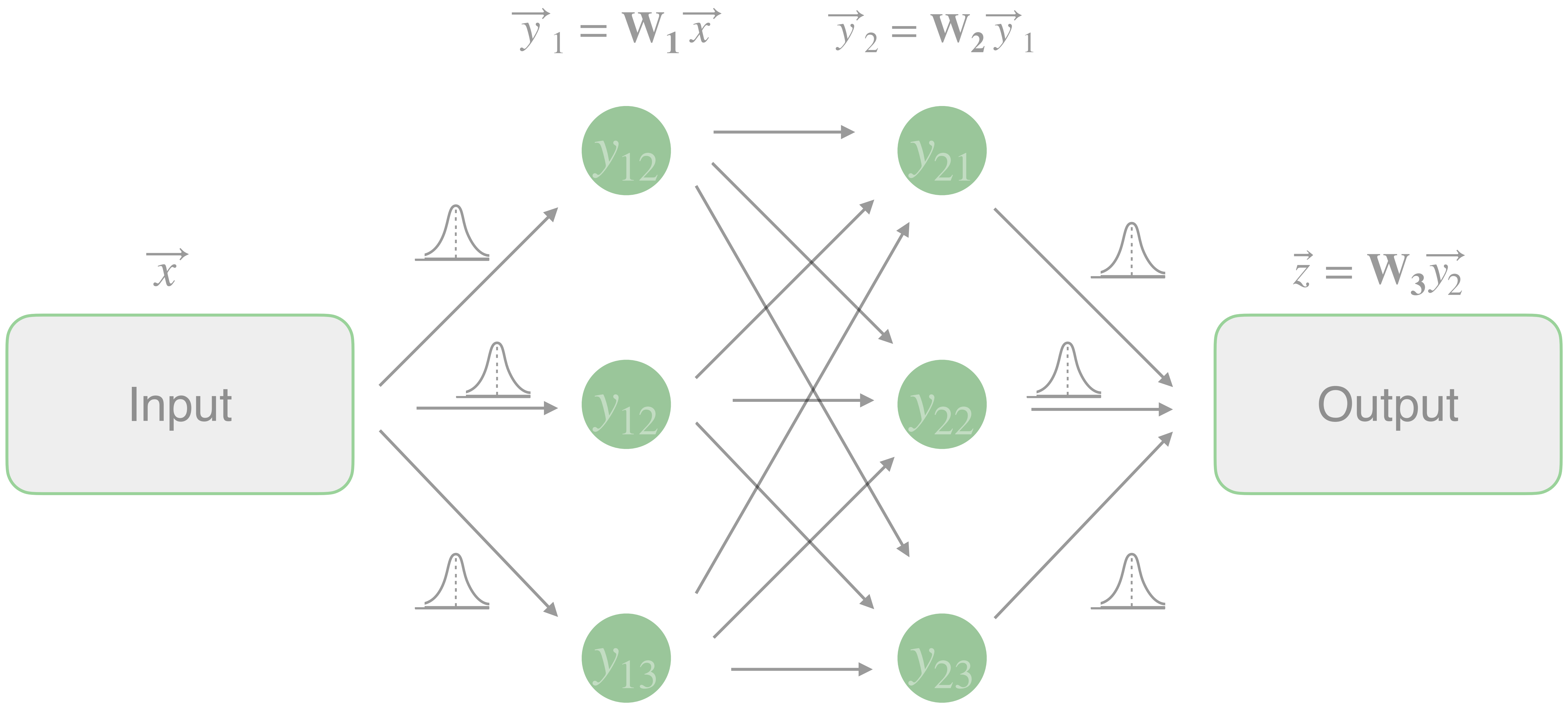


No Likelihood Loss

$$\mathcal{L}_{CFM} = \left(v_\theta - (x_1 - x_0)\right)^2$$

+ ? ? ?

How to Bayesianize - CFM



No Likelihood Loss

$$\mathscr{L}_{CFM} = \left\langle \left(v_\theta - (x_1 - x_0)\right)^2 \right\rangle_{\theta \sim q(\theta)}$$

Hyperparameter

$$+ \quad c \cdot \mathscr{D}_{KL}(q(\theta), p(\theta))$$

Choosing Test Phase-Space

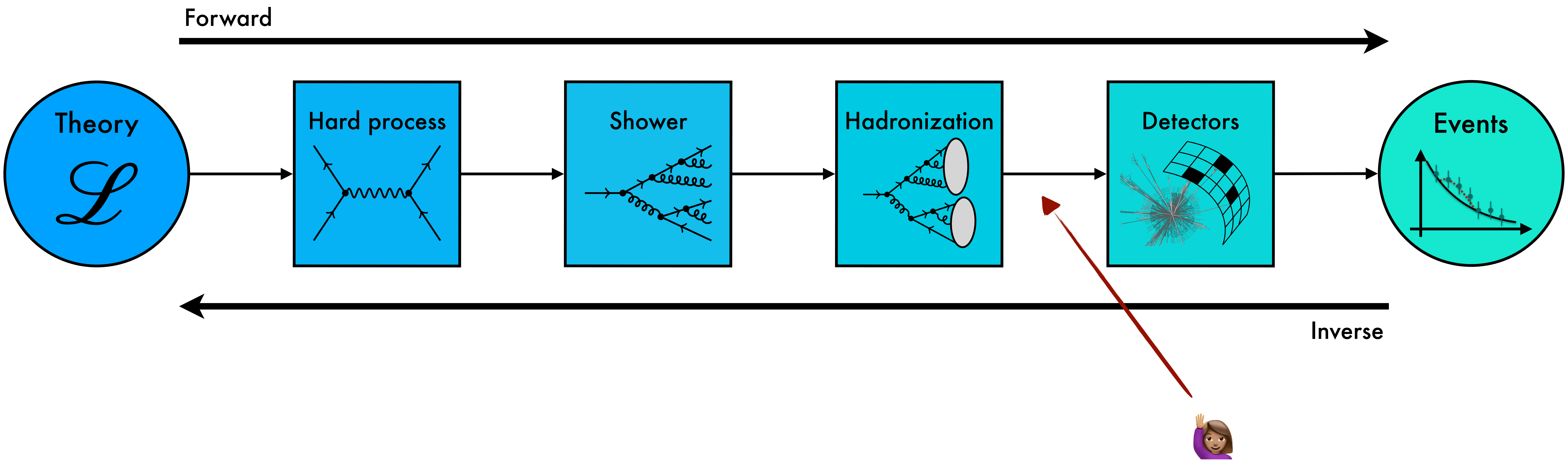
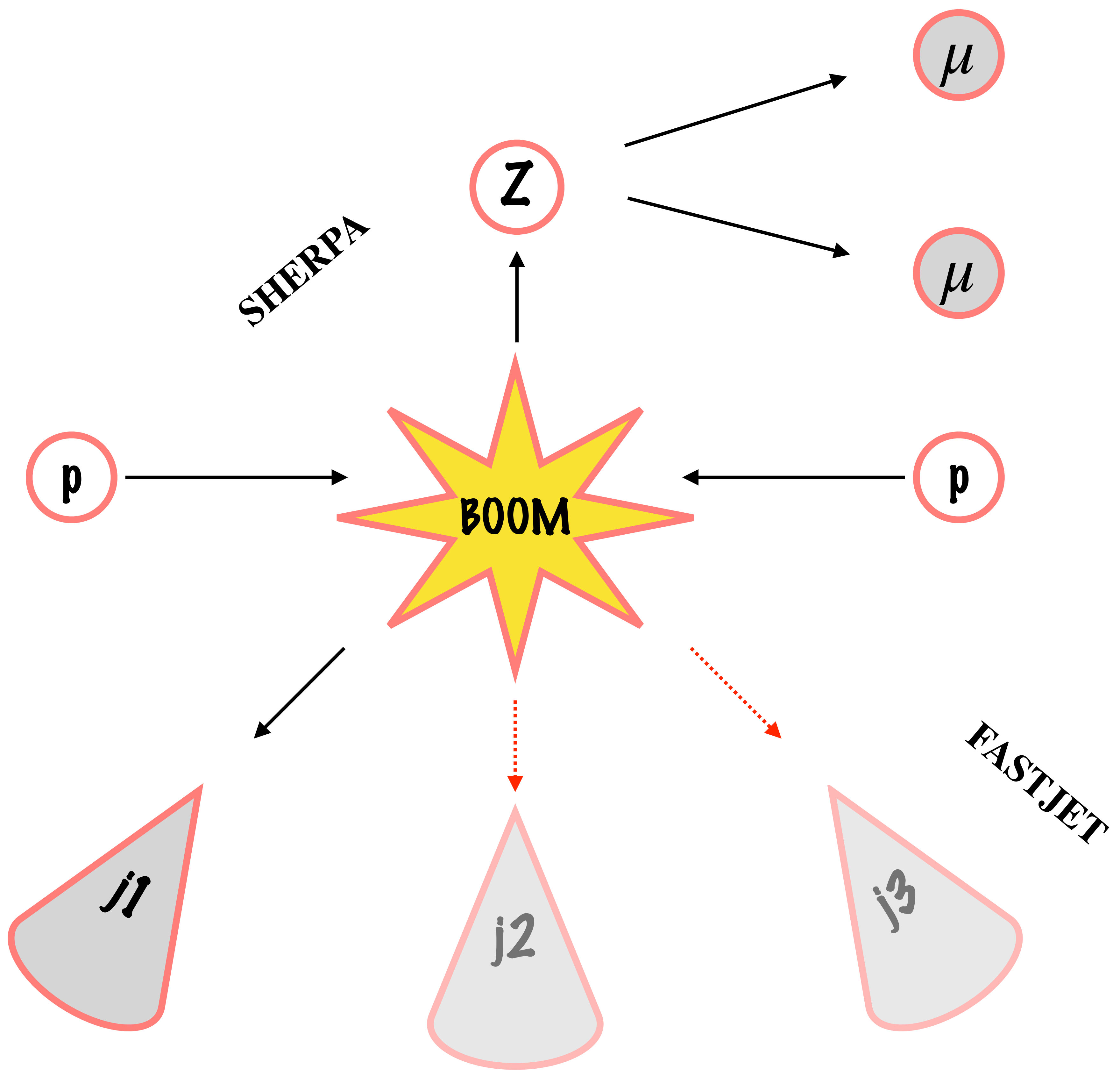


Figure from A. Butter et al.: arXiv:2203.07460, R. Winterhalder

Concrete Application — LHC



Z (→ μμ) + jets:

3 - 5 final state particles (including jets)

12 - 20 dimensional phase space

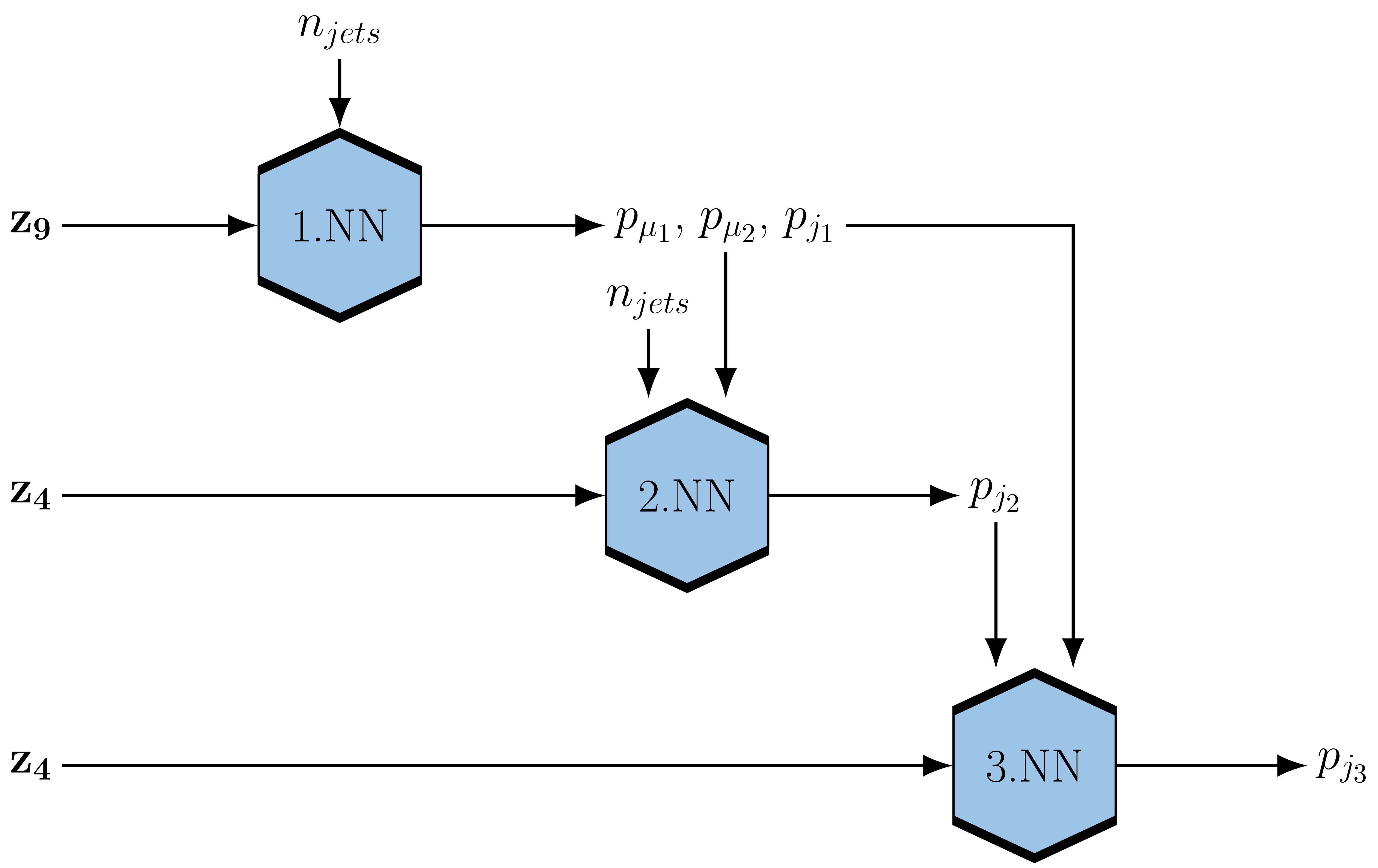
Smart preprocessing:

Global Phase Shift

Drop muon masses

→ reduces phase space to 9 - 17 dimensions

Concrete Application — LHC



Z ($\rightarrow \mu\mu$) + jets:

3 - 5 final state particles (including jets)

12 - 20 dimensional phase space

Smart preprocessing:

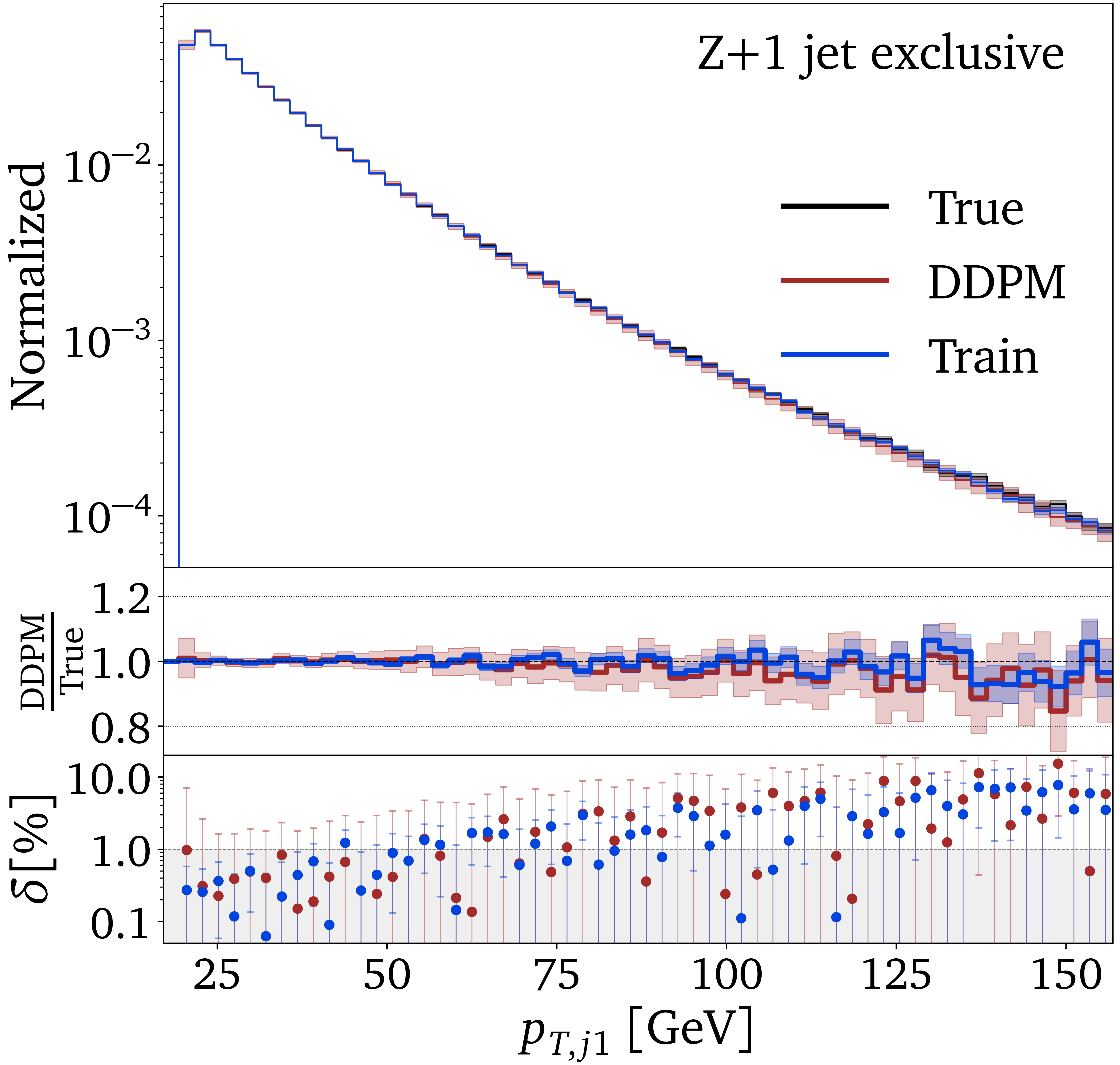
Global Phase Shift

Drop muon masses

$\rightarrow$ reduces phase space to 9 - 17 dimensions

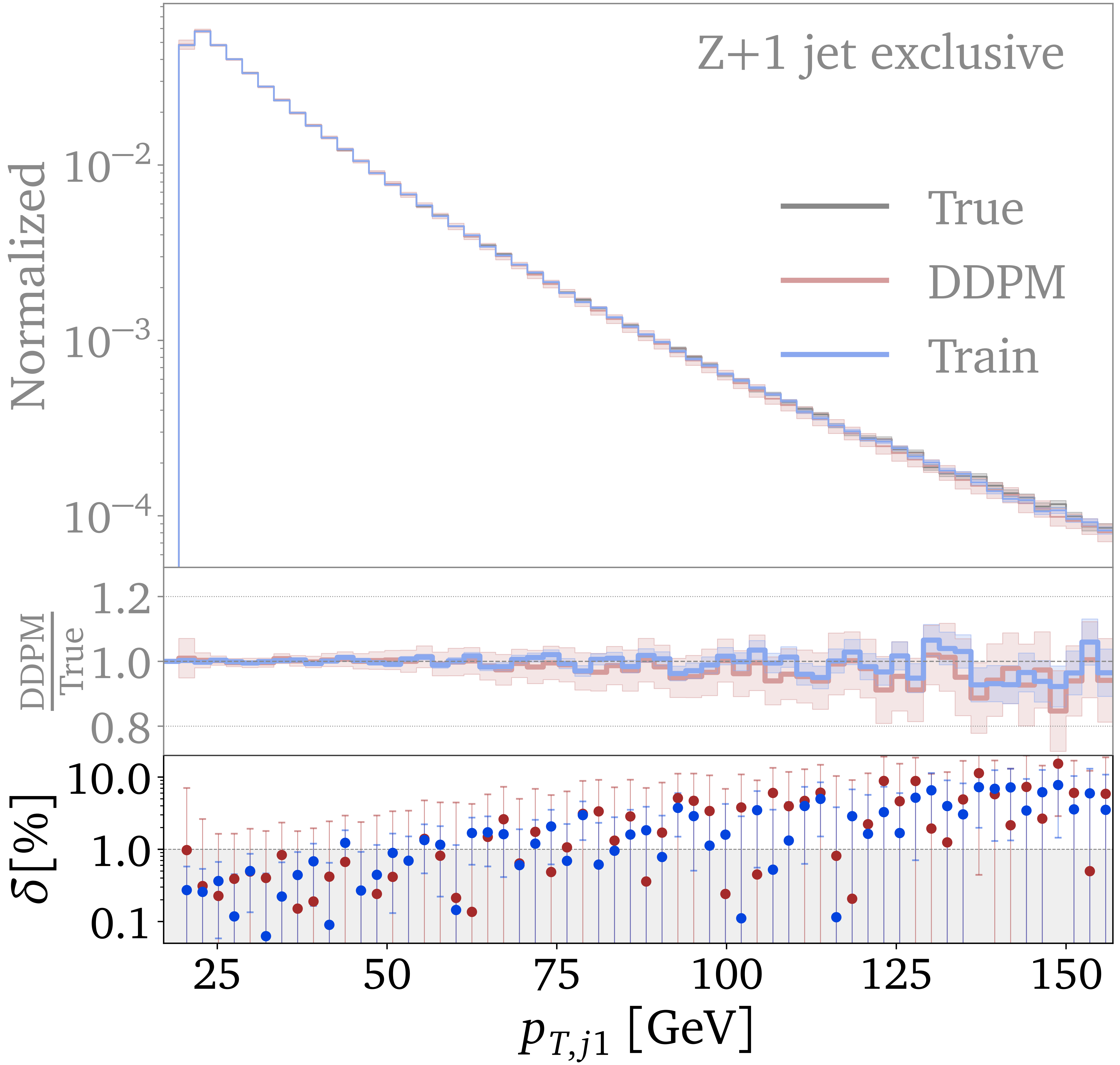
To be precise

Percent level precision (comparable to statistical uncertainty)



To be precise

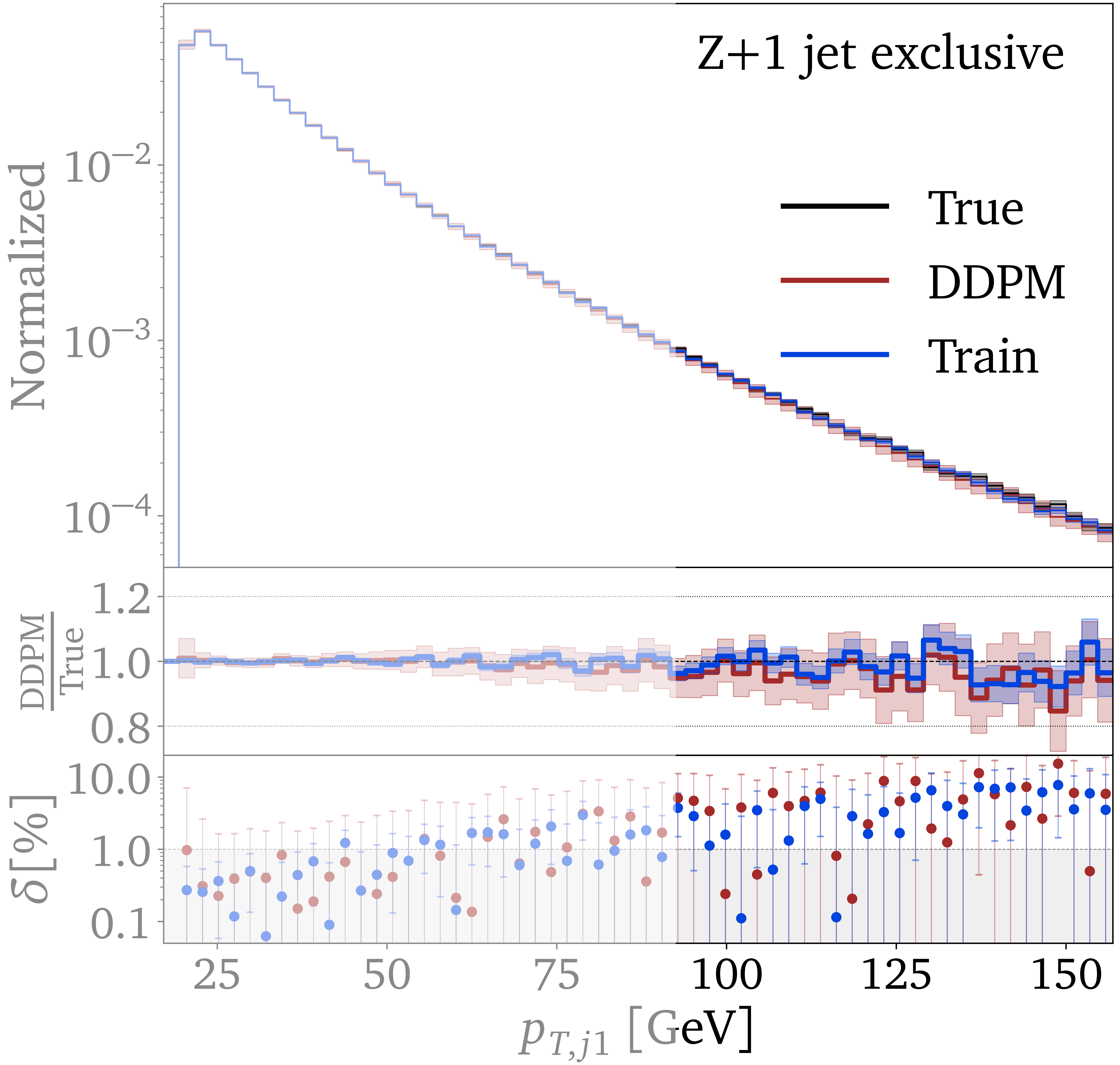
Percent level precision (comparable to statistical uncertainty)



To be precise

Percent level precision (comparable to statistical uncertainty)

Uncertainty well-defined

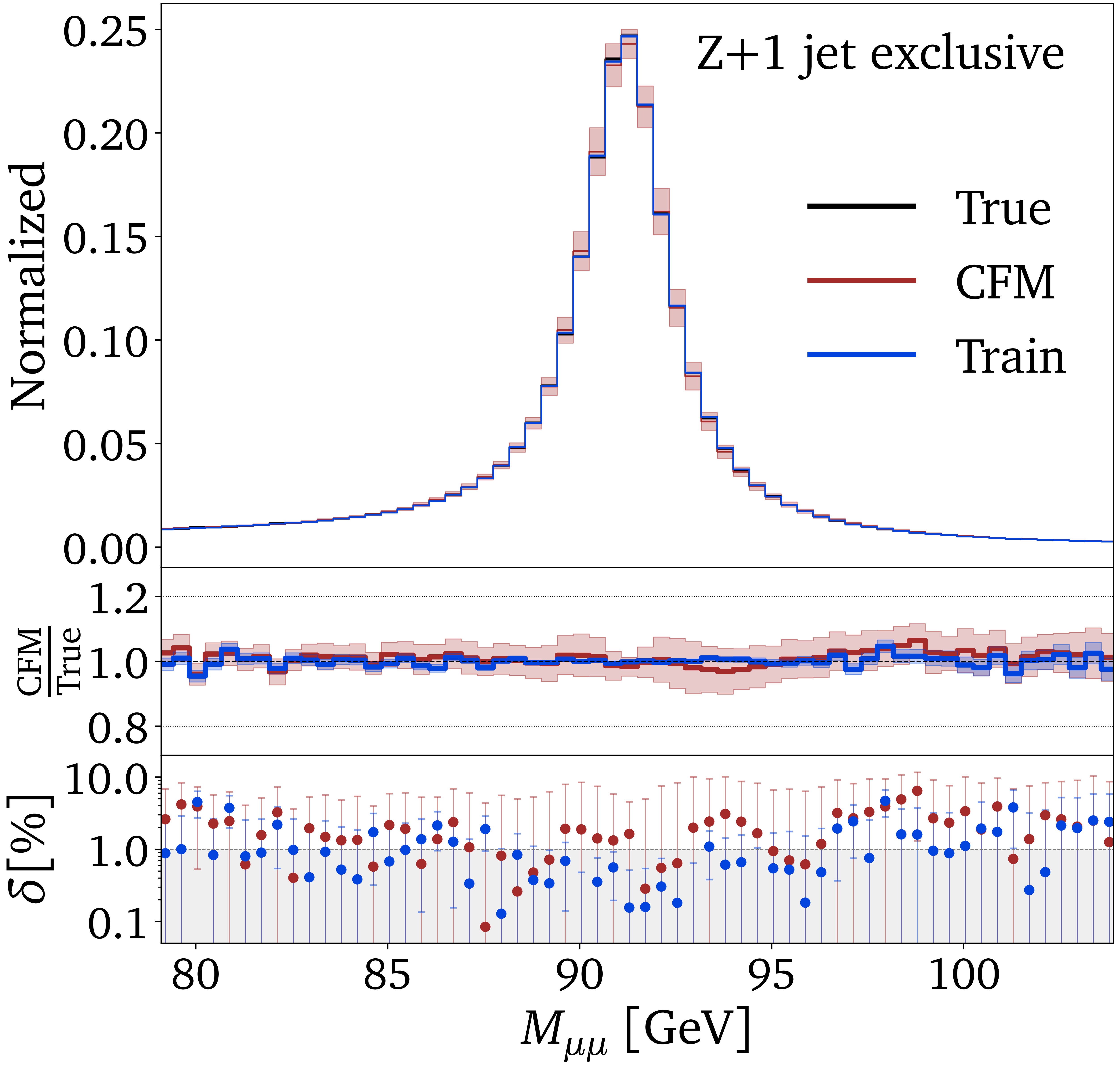


To be precise

Percent level precision (comparable to statistical uncertainty)

Uncertainty well-defined

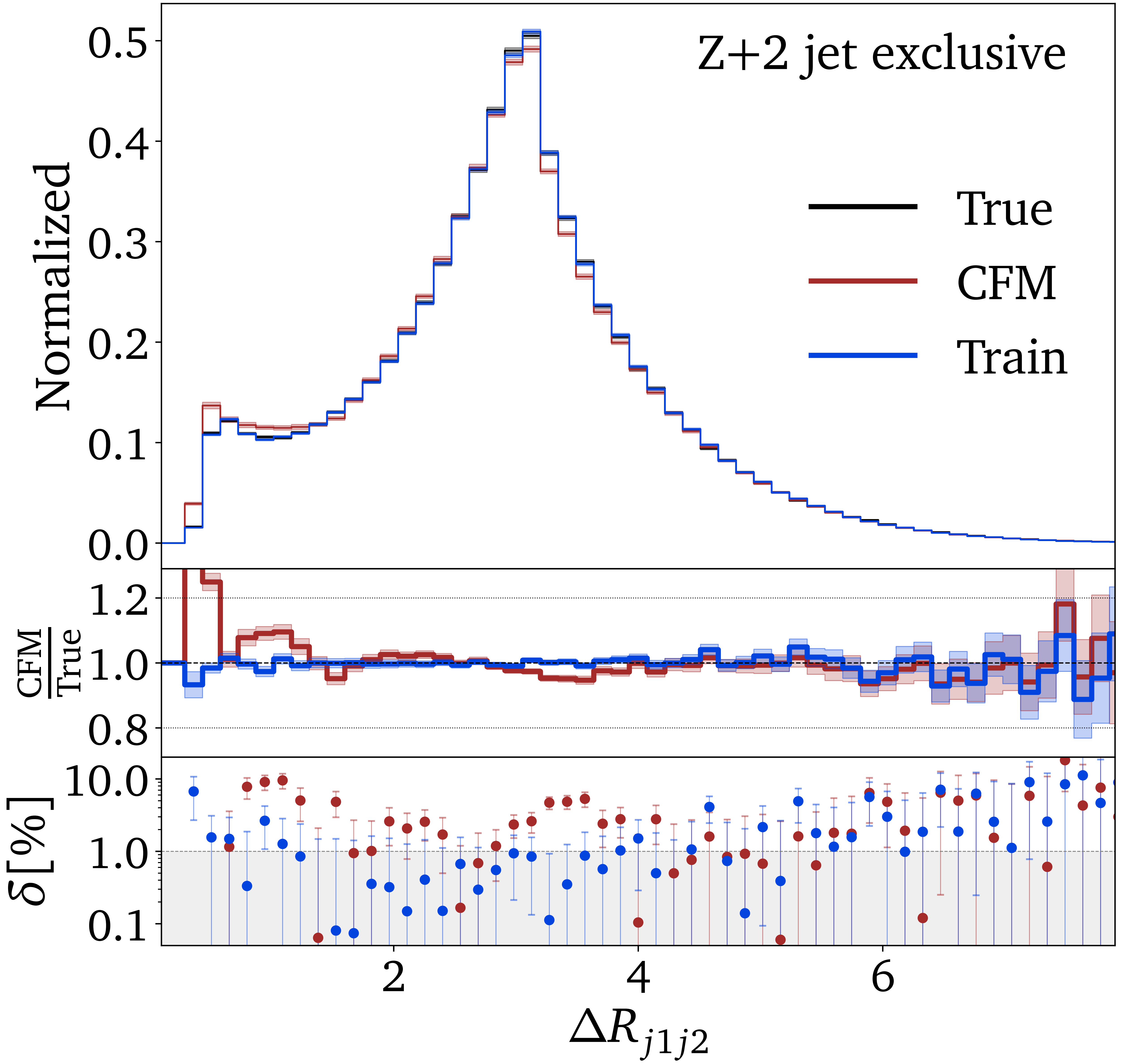
Surpasses INN precision (A. Butter et al.: arXiv:2110.13632)



To be precise

By construction:

Systematic failure modes of network not covered by Bayesian uncertainties



And now what?

Diffusion Models can compete with current benchmark (precision-wise)

A lot of on-going research (generation speed up, precision, etc.)

Bayesian Versions seem to estimate training uncertainty correctly

→ (Bayesian) Diffusion models show potential to be applied to particle physics tasks

Backup

Diffusion Models (DDPM)

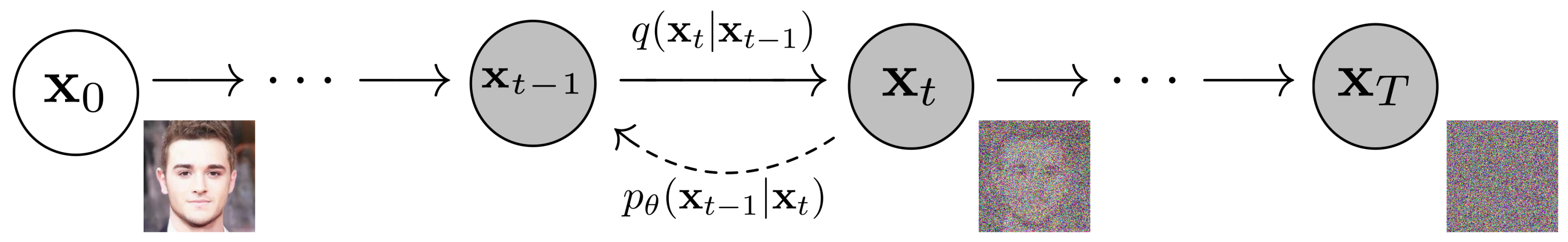
Phase Space

$x_0 \sim p(x_0)$

Forwards

Latent Space

$x_T \sim p(x_T)(= \mathcal{N}(0,1))$



$$q(x_1, \ldots, x_T | x_0) = \prod_{t=1}^{T} q(x_t | x_{t-1})$$

$$q(x_t | x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t} x_{t-1}, \beta_t)$$

where β_t follows noise scheduler

Figure from J.Ho et al.: arXiv:2006.11239

Diffusion Models (DDPM)

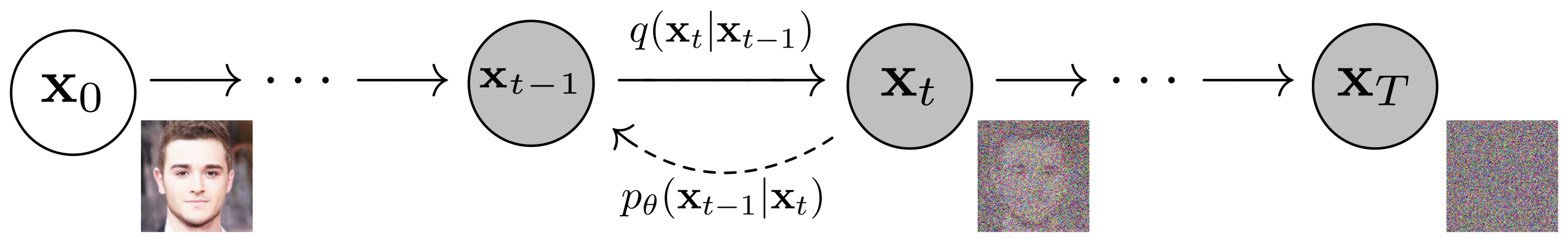
Backwards

Phase Space

$x_0 \sim p(x_0)$

Latent Space

$x_T \sim p(x_T)(= \mathcal{N}(0,1))$



To reverse: $q(x_{t-1} | x_t) = \dfrac{q(x_t | x_{t-1})q(x_{t-1})}{q(x_t)}$

…but we don’t know $q(x_t)$ & $q(x_{t-1})$

Figure from J.Ho et al.: arXiv:2006.11239

Diffusion Models (DDPM)

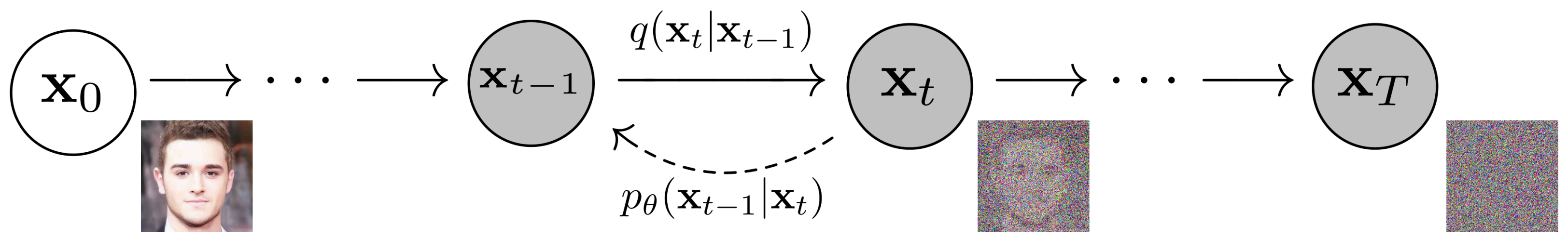
Backwards

Phase Space

$x_0 \sim p(x_0)$

Latent Space

$x_T \sim p(x_T)(= \mathcal{N}(0,1))$



Instead, learn: $p_\theta(x_{t-1}\,|\,x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \sigma^2_\theta(x_t, t))$

$$p(x_0, \ldots, x_T\,|\,\theta) = p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}\,|\,x_t)$$

Figure from J.Ho et al.: arXiv:2006.11239

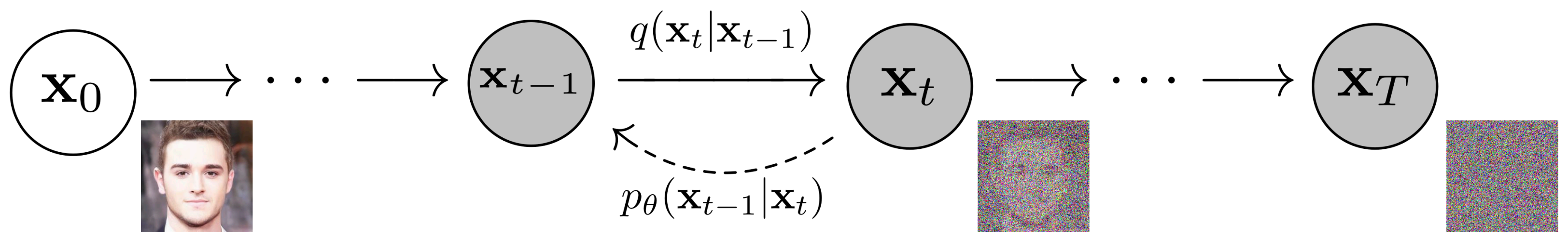
Diffusion Models (DDPM)

Phase Space

$x_0 \sim p(x_0)$

Latent Space

$x_T \sim p(x_T)(= \mathcal{N}(0,1))$



$$\mathscr{L}_{DDPM} = -\log p_\theta(x_0) \approx \frac{1}{2\sigma_t^2}\frac{\beta_t^2}{(1-\beta_t)\bar{\beta}_t}|\epsilon(t) - \epsilon_\theta(t)|^2$$

Figure from J.Ho et al.: arXiv:2006.11239

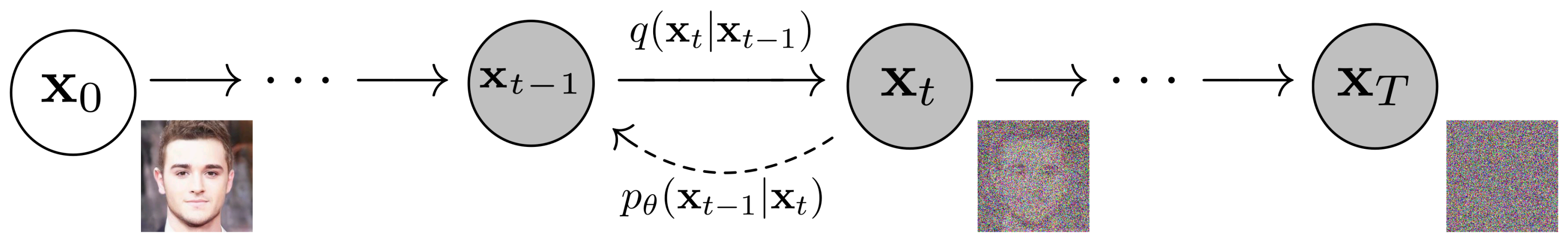
Diffusion Models (DDPM)

Phase Space

$x_0 \sim p(x_0)$

Latent Space

$x_T \sim p(x_T)(= \mathcal{N}(0,1))$



$$\mathscr{L}_{DDPM} = -\log p_\theta(x_0) \approx \frac{1}{2\sigma_t^2}\frac{\beta_t^2}{(1-\beta_t)\bar{\beta}_t}\,|\,\epsilon(t) - \epsilon_\theta(t)\,|^2$$

:
- Reparametrization
- Estimation
- …

Predicted and actual noise added at time t

Figure from J.Ho et al.: arXiv:2006.11239

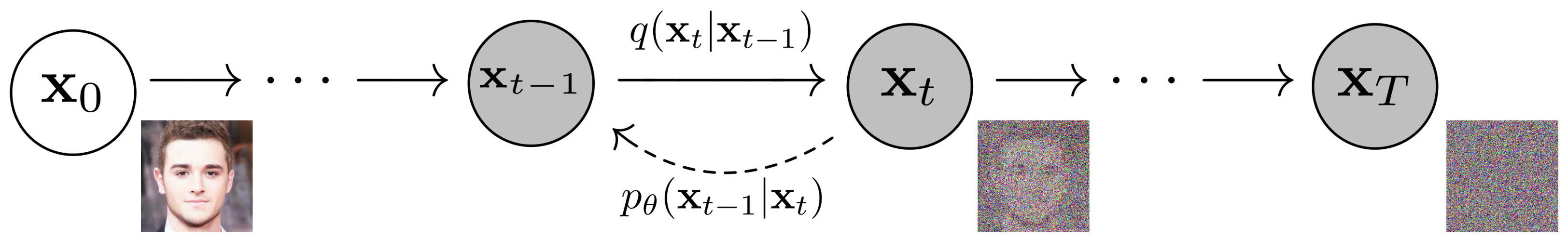
Diffusion Models (DDPM)

Phase Space

$x_0 \sim p(x_0)$

Latent Space

$x_T \sim p(x_T)(= \mathcal{N}(0,1))$



$$\mathscr{L}_{DDPM} = -\log p_\theta(x_0) \approx \frac{1}{2\sigma_t^2}\frac{\beta_t^2}{(1-\beta_t)\bar{\beta}_t}|\epsilon(t) - \epsilon_\theta(t)|^2$$

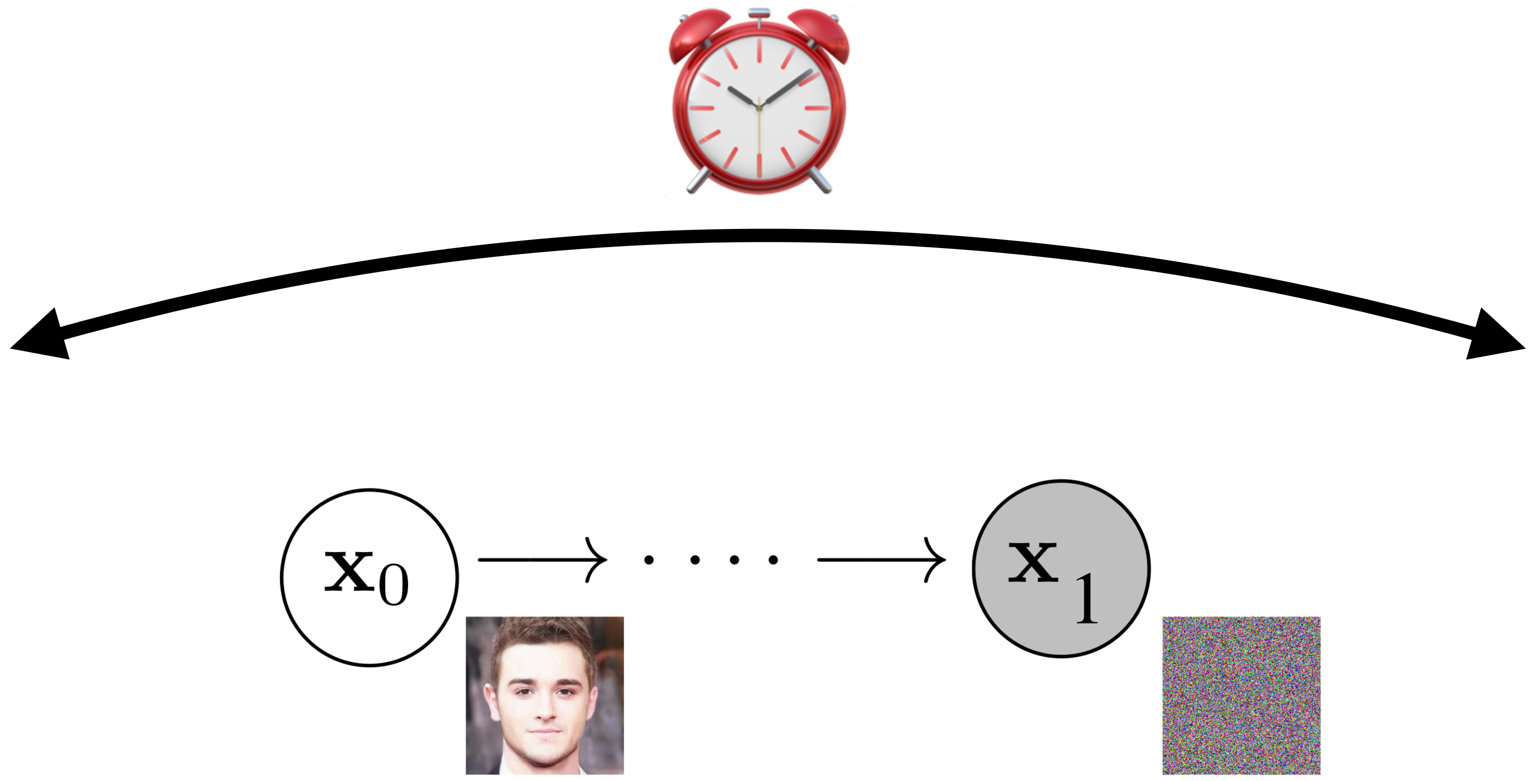
Denoising

Figure from J.Ho et al.: arXiv:2006.11239

Diffusion Models (CFM)

Phase Space

$x_0 \sim p(x_0)$



Latent Space

$x_1 \sim p(x_1)(= \mathcal{N}(0,1))$

1. Connect x_0 with x_1 by a linear trajectory: $x(t\,|\,x_0) = (1 - t)x_0 + tx_1$

2. Derive: $\frac{d}{dt}x(t\,|\,x_0) = (x_1 - x_0) \equiv v_\theta(x_1, x_0)$

3. Solve for x_0 (ODE): $x_0 = x_t - \int_0^1 v_\theta(x(t), t)dt$

Figure from J.Ho et al.: arXiv:2006.11239

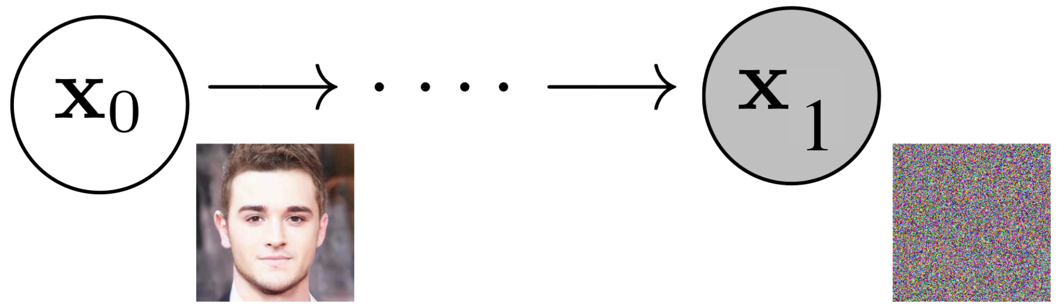
Diffusion Models (CFM)

Phase Space

$x_0 \sim p(x_0)$

Latent Space

$x_1 \sim p(x_1)(= \mathcal{N}(0,1))$



1. Connect x_0 with x_1 by a linear trajectory: $x(t\,|\,x_0) = (1-t)x_0 + tx_1$

2. Derive: $\frac{d}{dt}x(t\,|\,x_0) = (x_1 - x_0) \equiv v_\theta(x_1, x_0)$

3. Solve for x_0 (ODE): $x_0 = x_t - \int_0^1 v_\theta(x(t), t)\,dt$

Figure from J.Ho et al.: arXiv:2006.11239