Fast Parameter Estimation for MBHBs with Normalising Flows

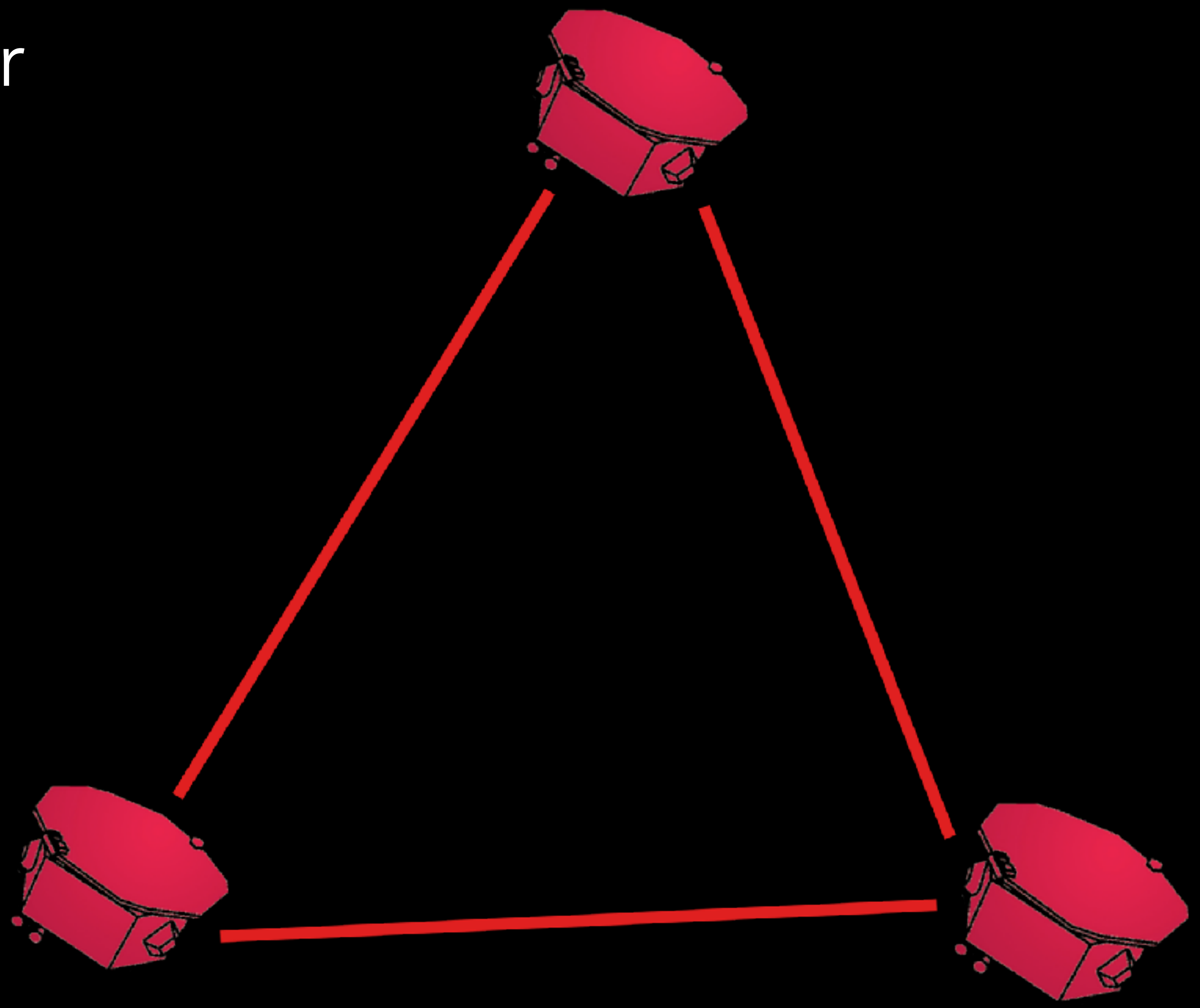
Natalia Korsakova

Laboratoire
AstroParticule & Cosmologie

APC, CNRS, FRANCE

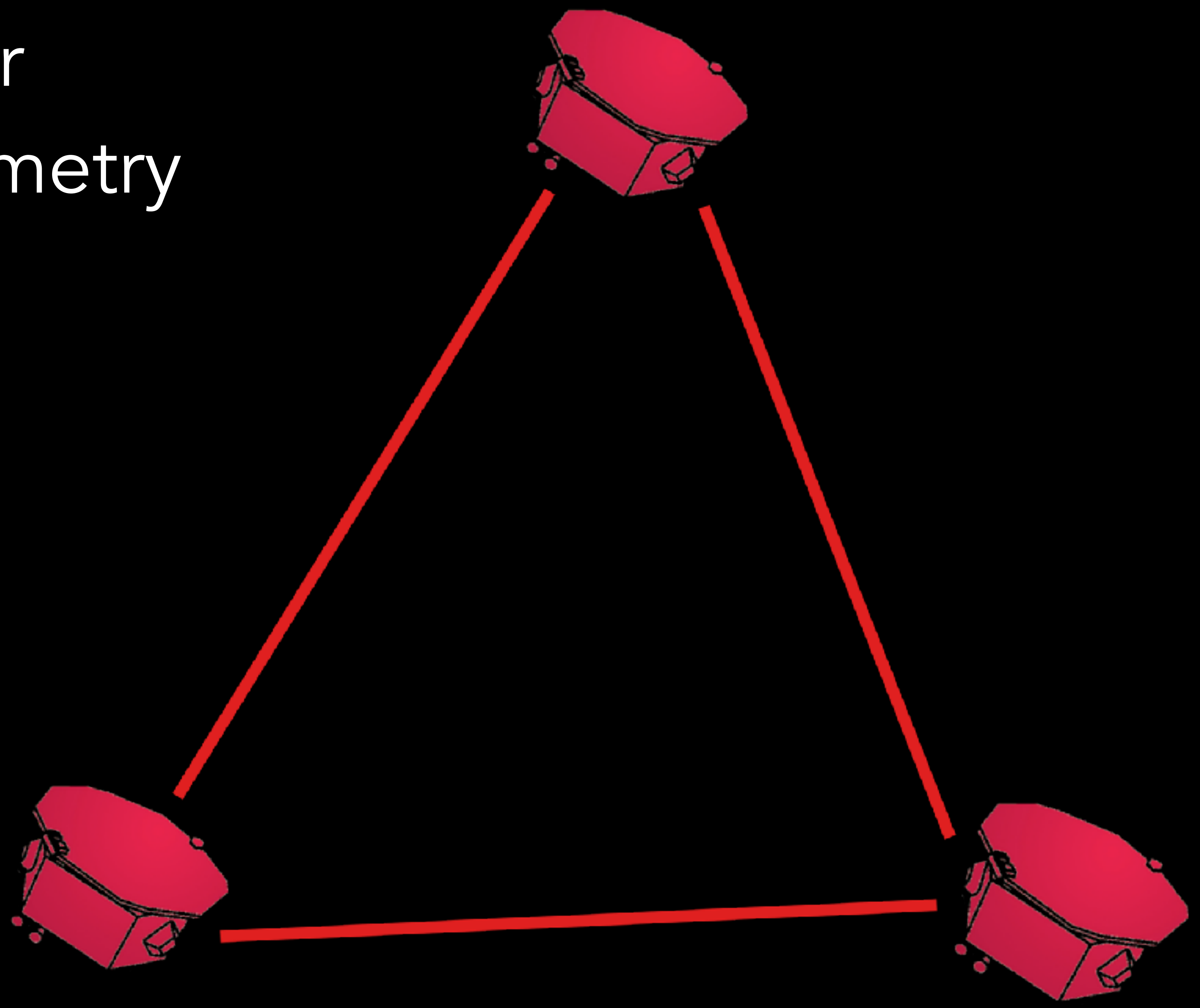
LASER INTERFEROMETER SPACE ANTENNA

- Space based gravitational wave detector



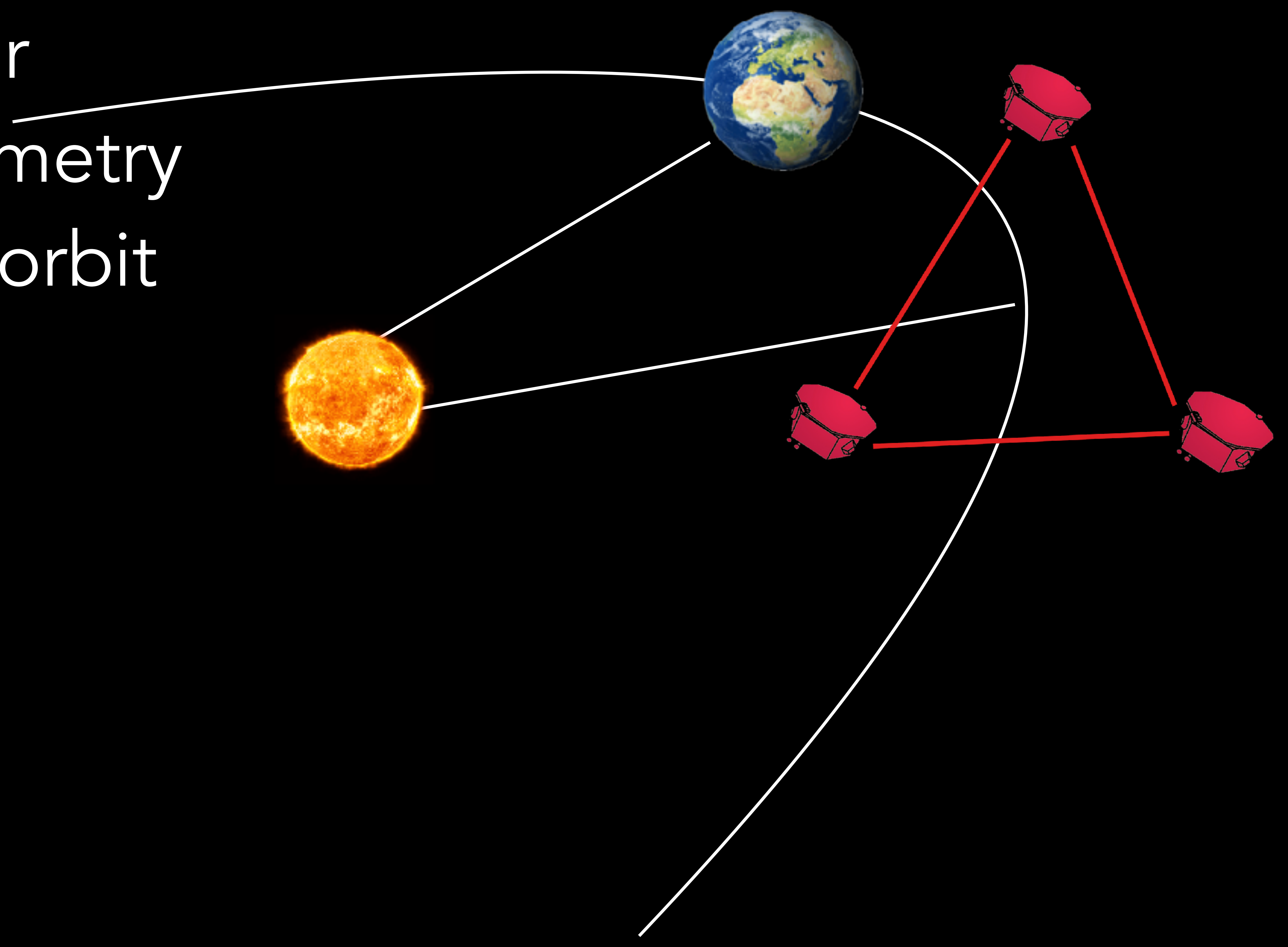
LASER INTERFEROMETER SPACE ANTENNA

- Space based gravitational wave detector
- Based on the principle of laser interferometry



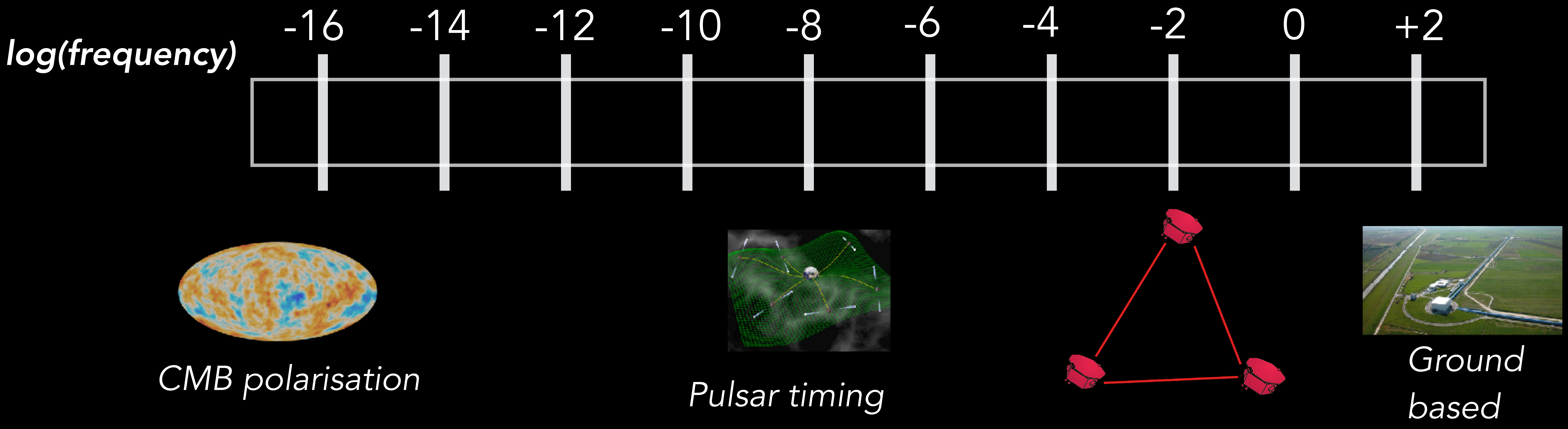
LASER INTERFEROMETER SPACE ANTENNA

- Space based gravitational wave detector
- Based on the principle of laser interferometry
- Following the Earth on the Heliocentric orbit



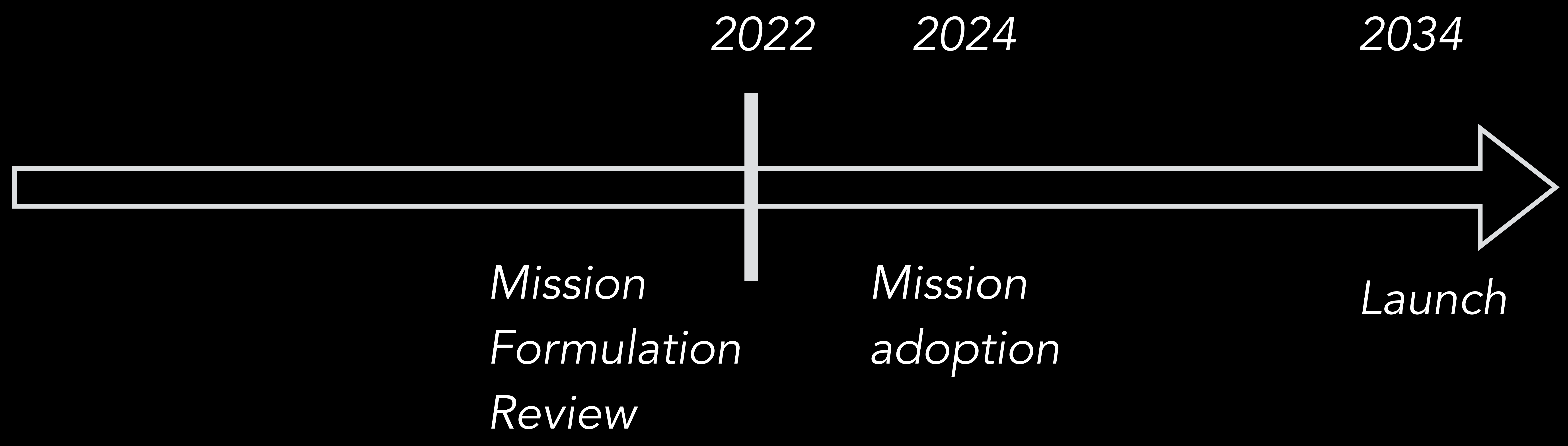
LASER INTERFEROMETER SPACE ANTENNA

- Space based gravitational wave detector
- Based on the principle of laser interferometry
- Following the Earth on the Heliocentric orbit
- Peak sensitivity at 20 mHz



LASER INTERFEROMETER SPACE ANTENNA

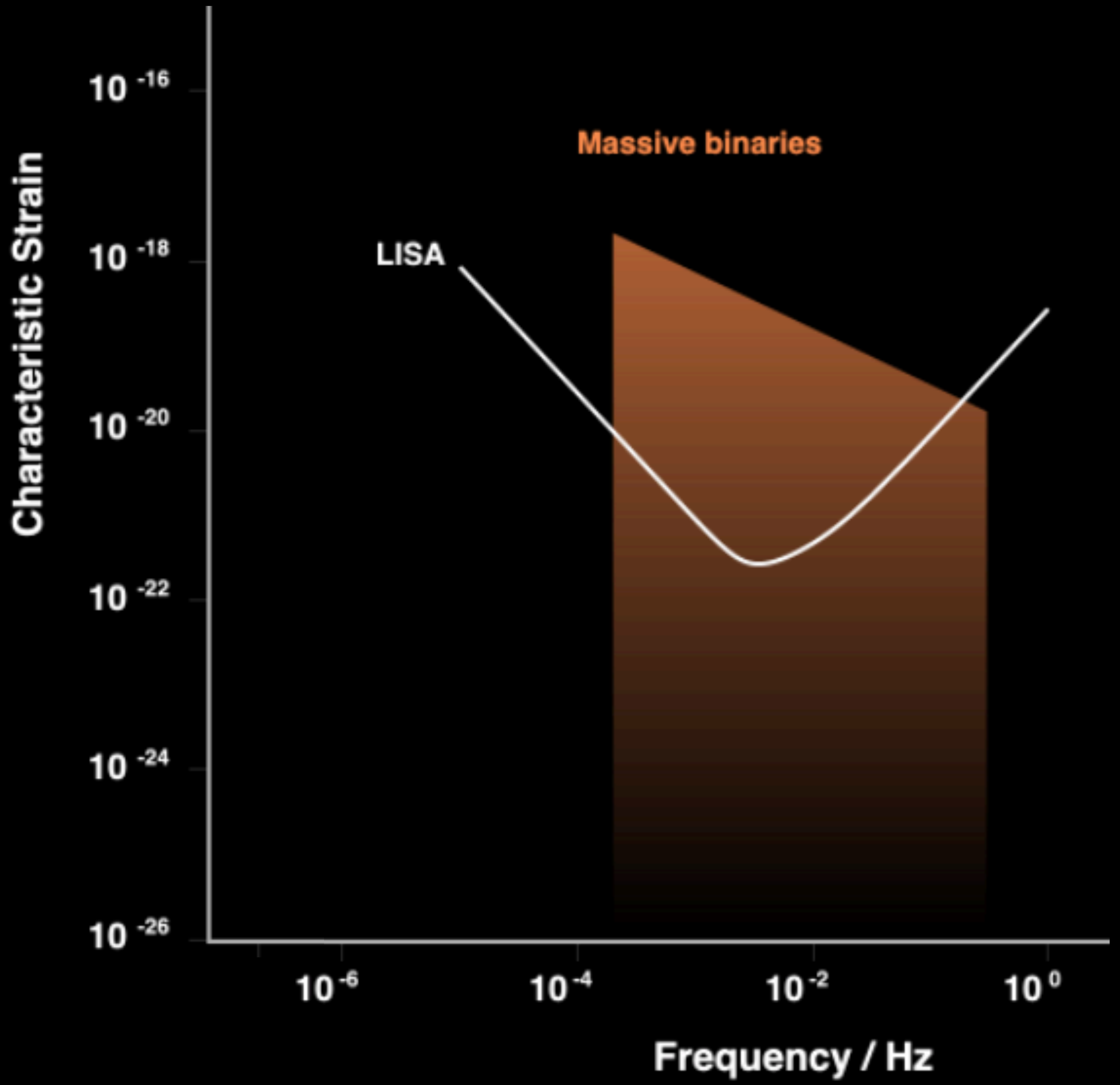
- Space based gravitational wave detector
- Based on the principle of laser interferometry
- Following the Earth on the Heliocentric orbit
- Peak sensitivity at 20 mHz
- Planned launch in 2034



MASSIVE BLACK HOLE BINARIES

image: gwplotter.com

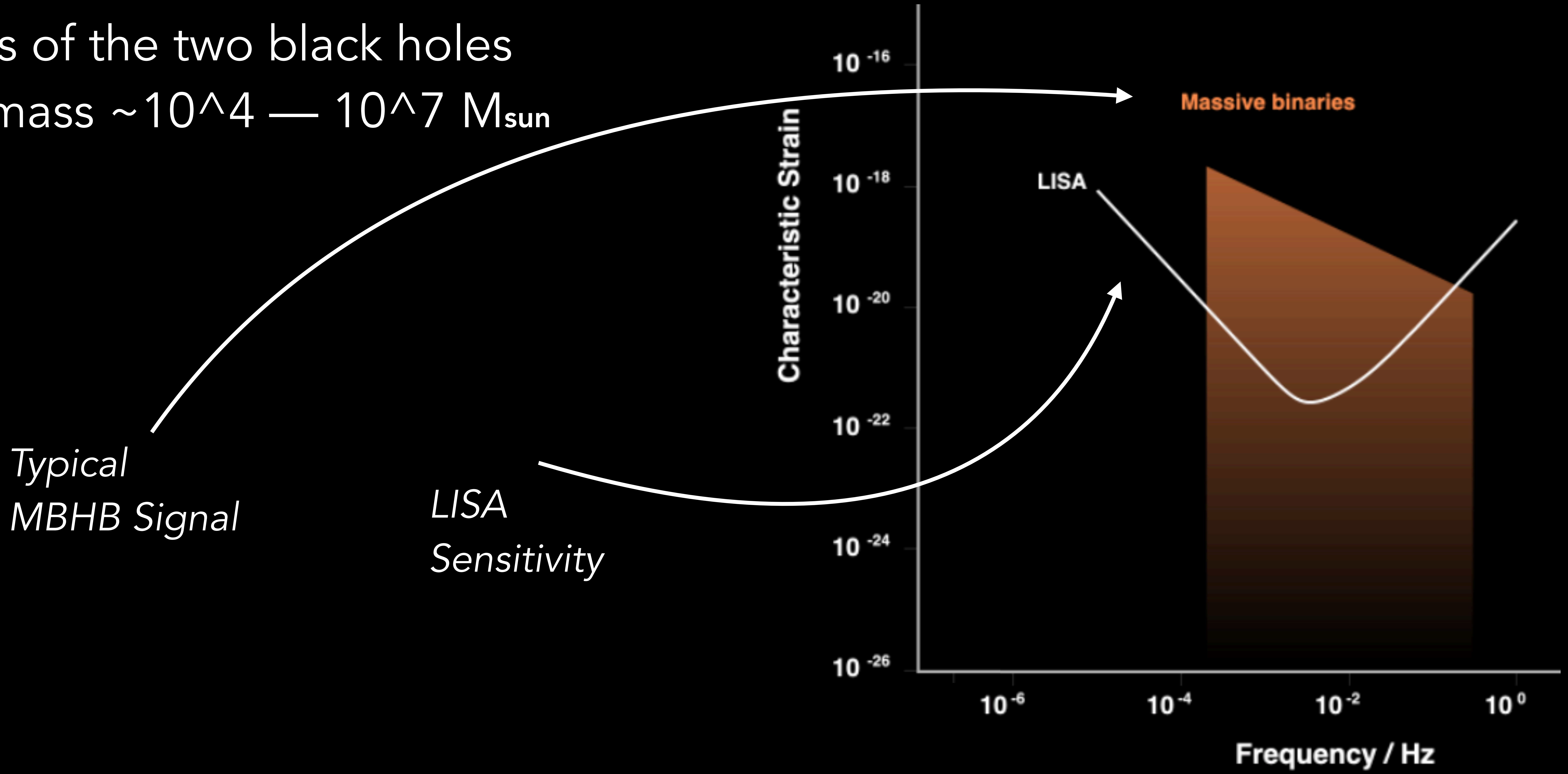
- Mergers of the two black holes of the mass ~10^4 — 10^7 M_{sun}



MASSIVE BLACK HOLE BINARIES

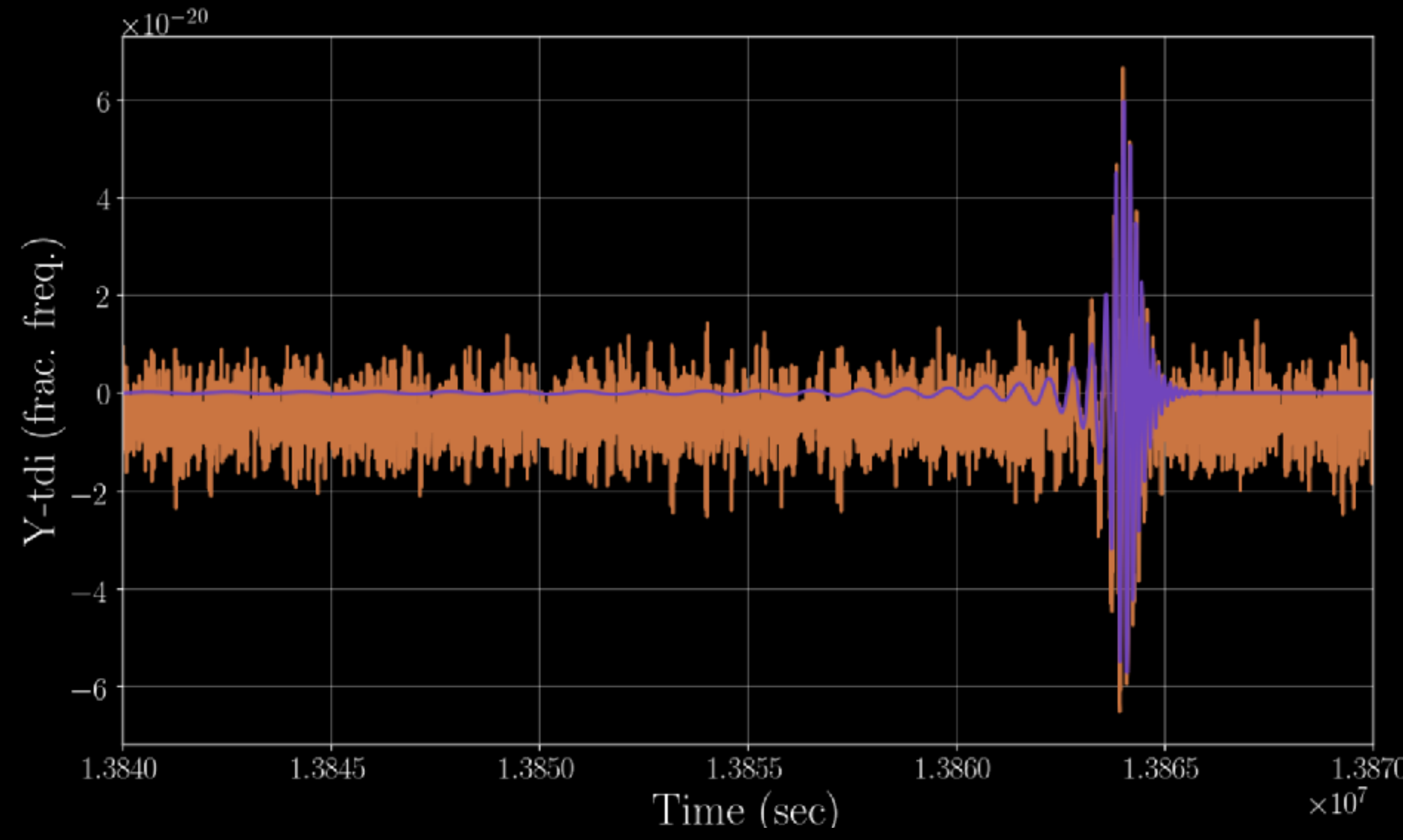
image: gwplotter.com

- Mergers of the two black holes of the mass ~10^4 — 10^7 M_{sun}



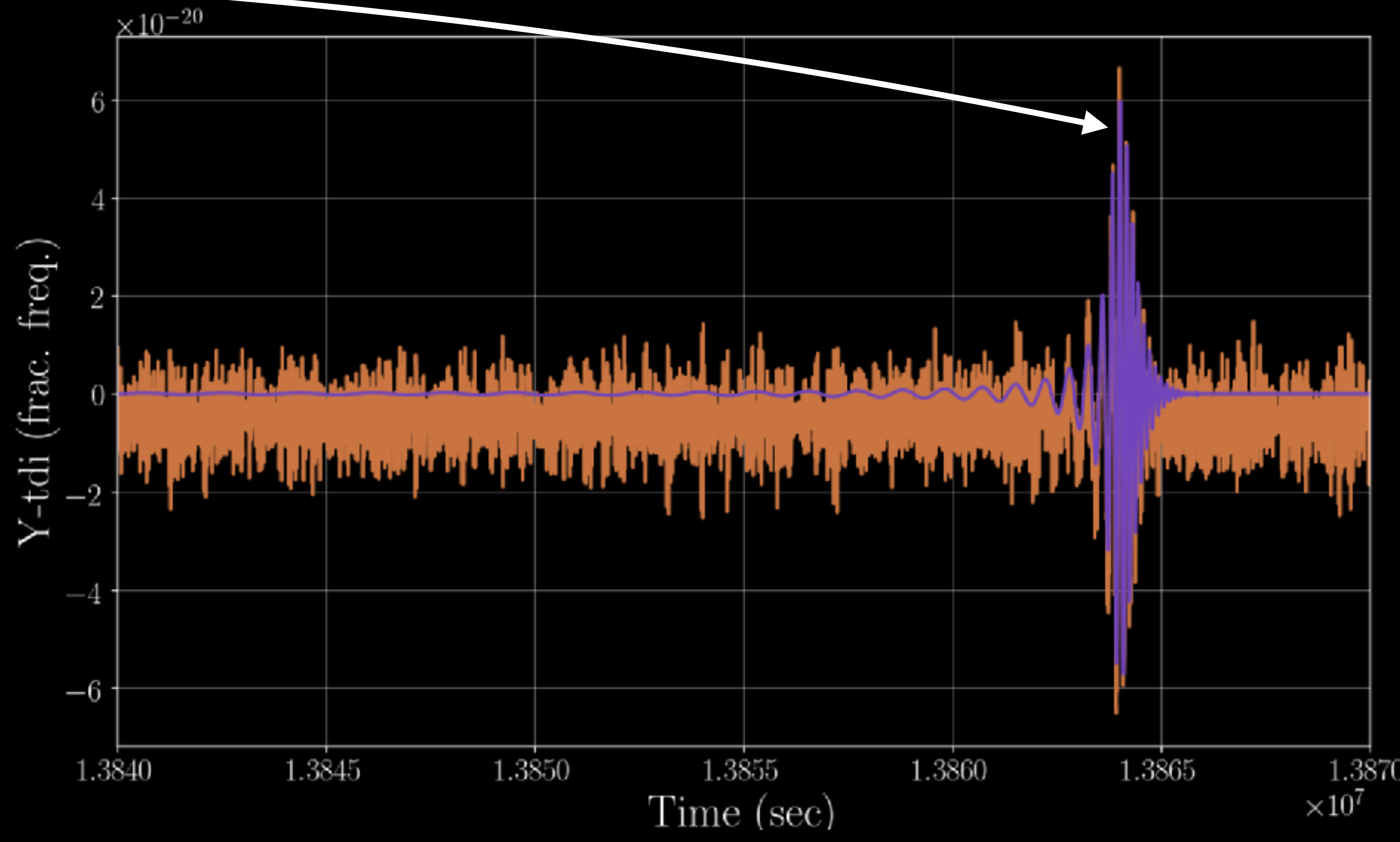
MASSIVE BLACK HOLE BINARIES

- Electromagnetic counterparts



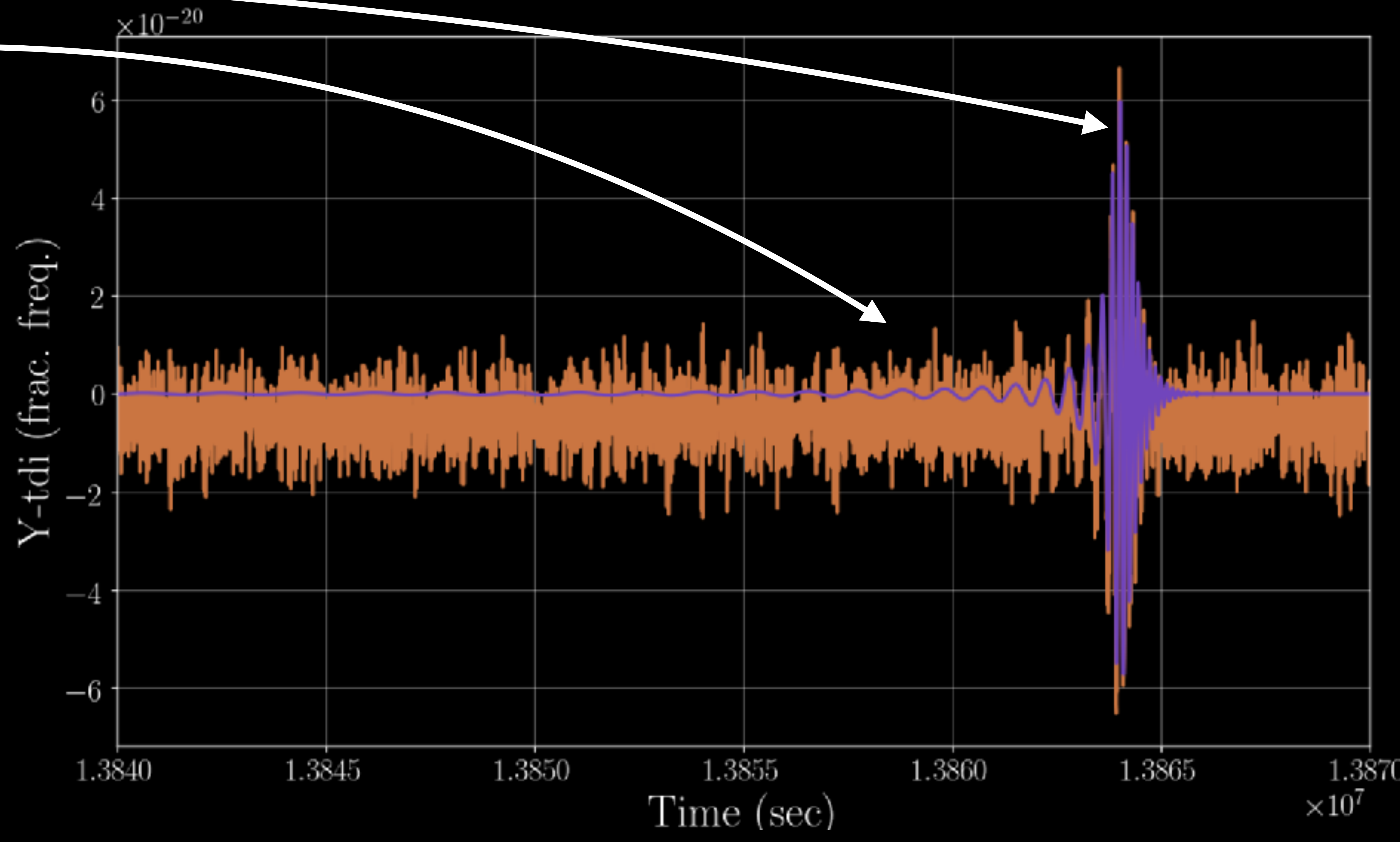
MASSIVE BLACK HOLE BINARIES

- Electromagnetic counterparts
- During merger



MASSIVE BLACK HOLE BINARIES

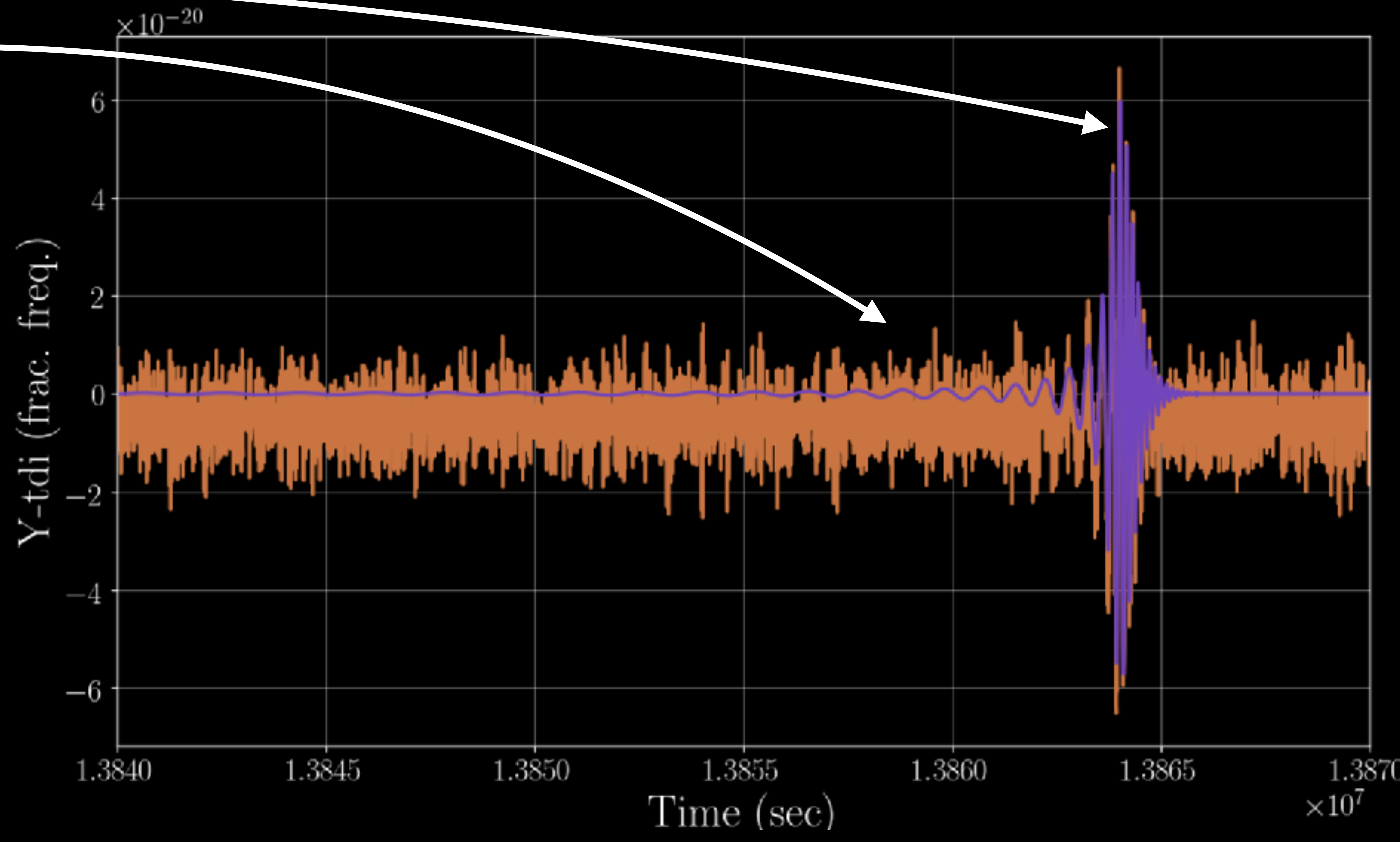
- Electromagnetic counterparts
- During merger
- or even during inspiral



MASSIVE BLACK HOLE BINARIES

- Electromagnetic counterparts
- During merger
- or even during inspiral

- EM counterparts can occur due to presence of
  - matter
  - magnetic fields



MASSIVE BLACK HOLE BINARIES

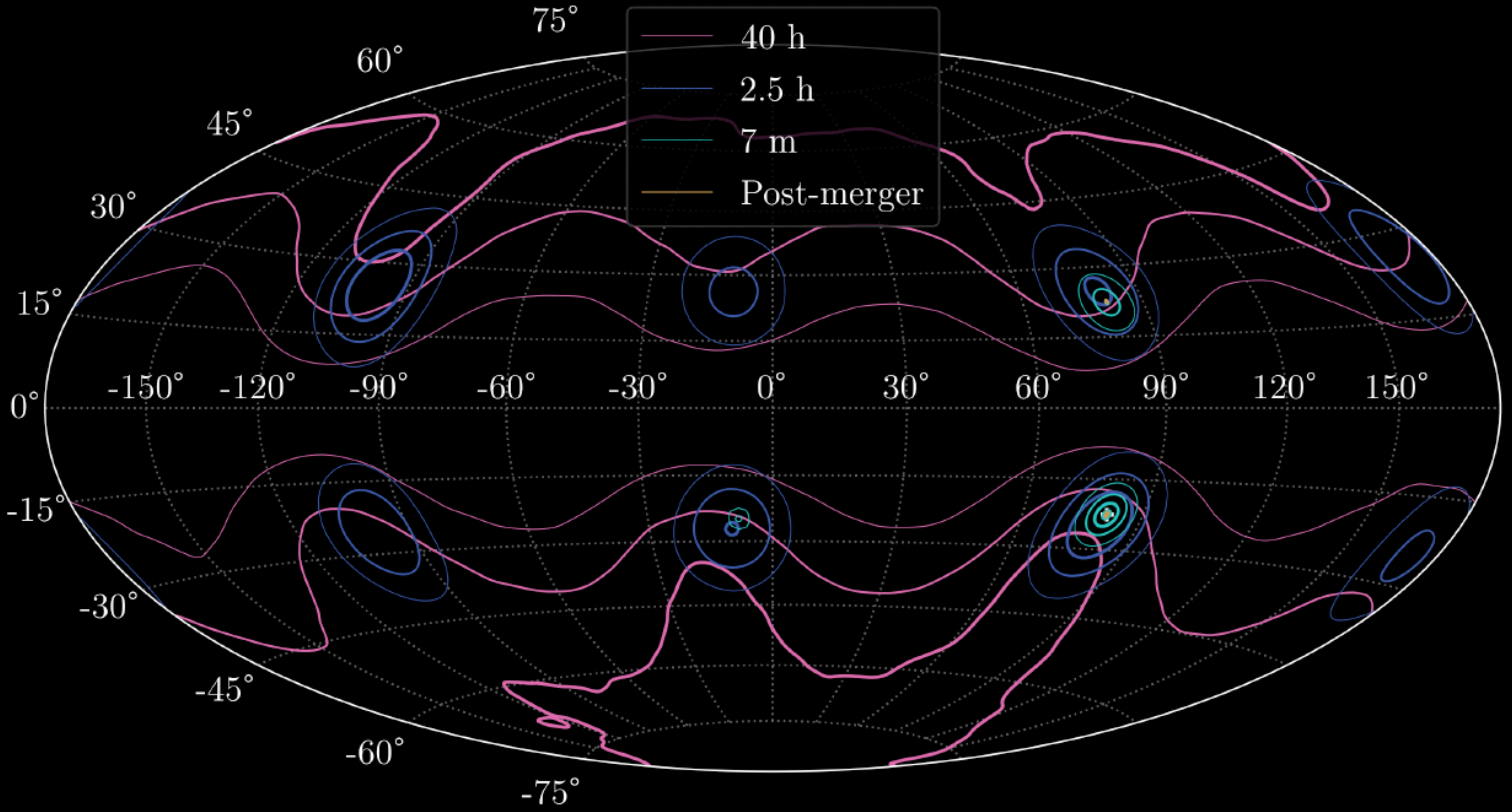


image: Marsat S. et al 2020 (arXiv:2003.00357)

INFERENCE

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

INFERENCE

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

- data model:

$$x = h(\theta) + n$$

« waveform template»

INFERENCE

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

- data model:

$$x = h(\theta) + n$$

« waveform template»

physical parameters

INFERENCE

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

- data model:

$$x = h(\theta) + n$$

« waveform template»

physical parameters

measurement noise

INFERENCE

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

<u>problem</u>:

marginal likelihood
has no exact solution

$$p(x) = \int p(x|\theta)p(\theta)\mathrm{d}\theta$$

INFERENCE

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

solutions:

- approximate inference:
  - MCMC/Nested sampling
    requires likelihood evaluation
    we can do it, but it is slow

INFERENCE

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

solutions:

- approximate inference:
  - MCMC/Nested sampling
    requires likelihood evaluation
    we can do it, but it is slow
  - Variational inference
    approximate the posterior distribution
    with a tractable distribution

INFERENCE

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

solutions:

- simplification to the model:
  - Gaussian mixture models

    too simple

INFERENCE

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

solutions:

- simplification to the model:
  - Gaussian mixture models

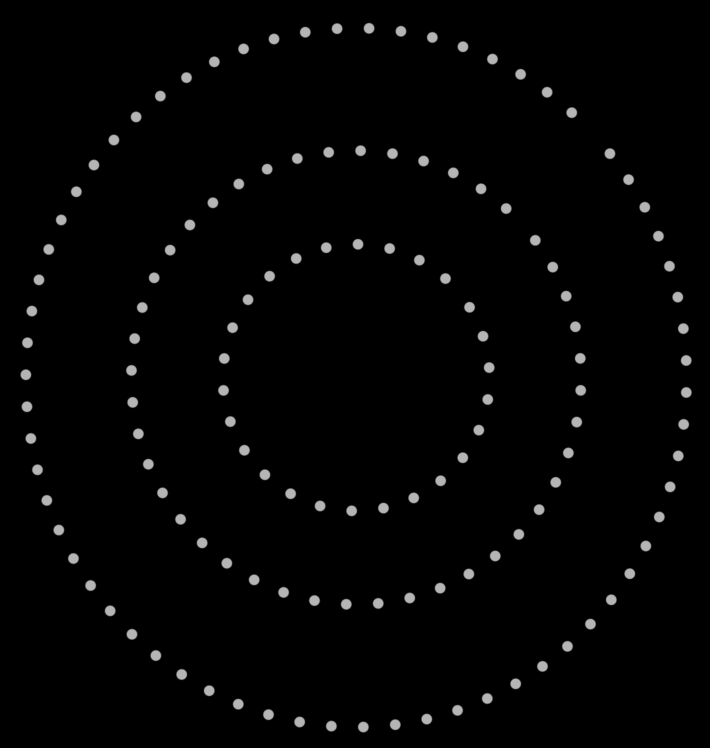
    too simple

  - Invertible models

    will talk about them today

NORMALISING FLOWS

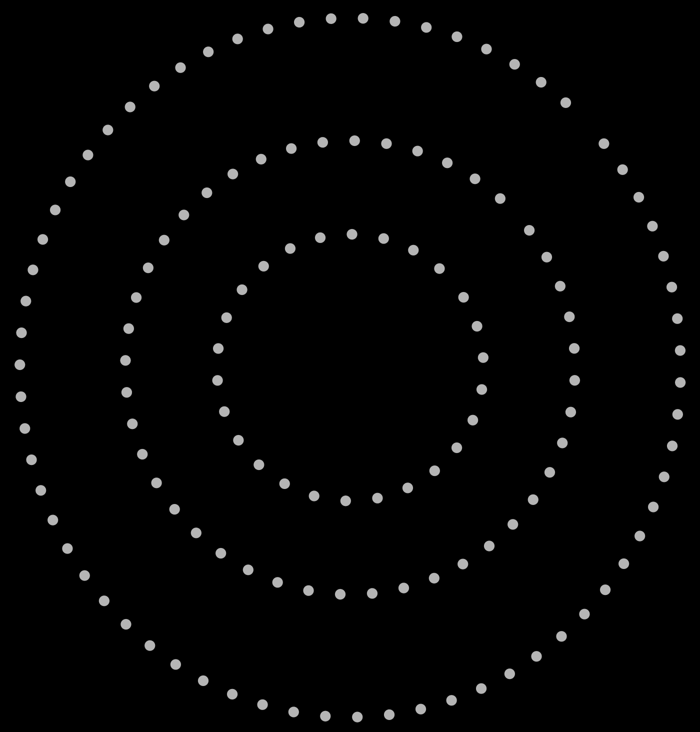
1. We have simple random generator



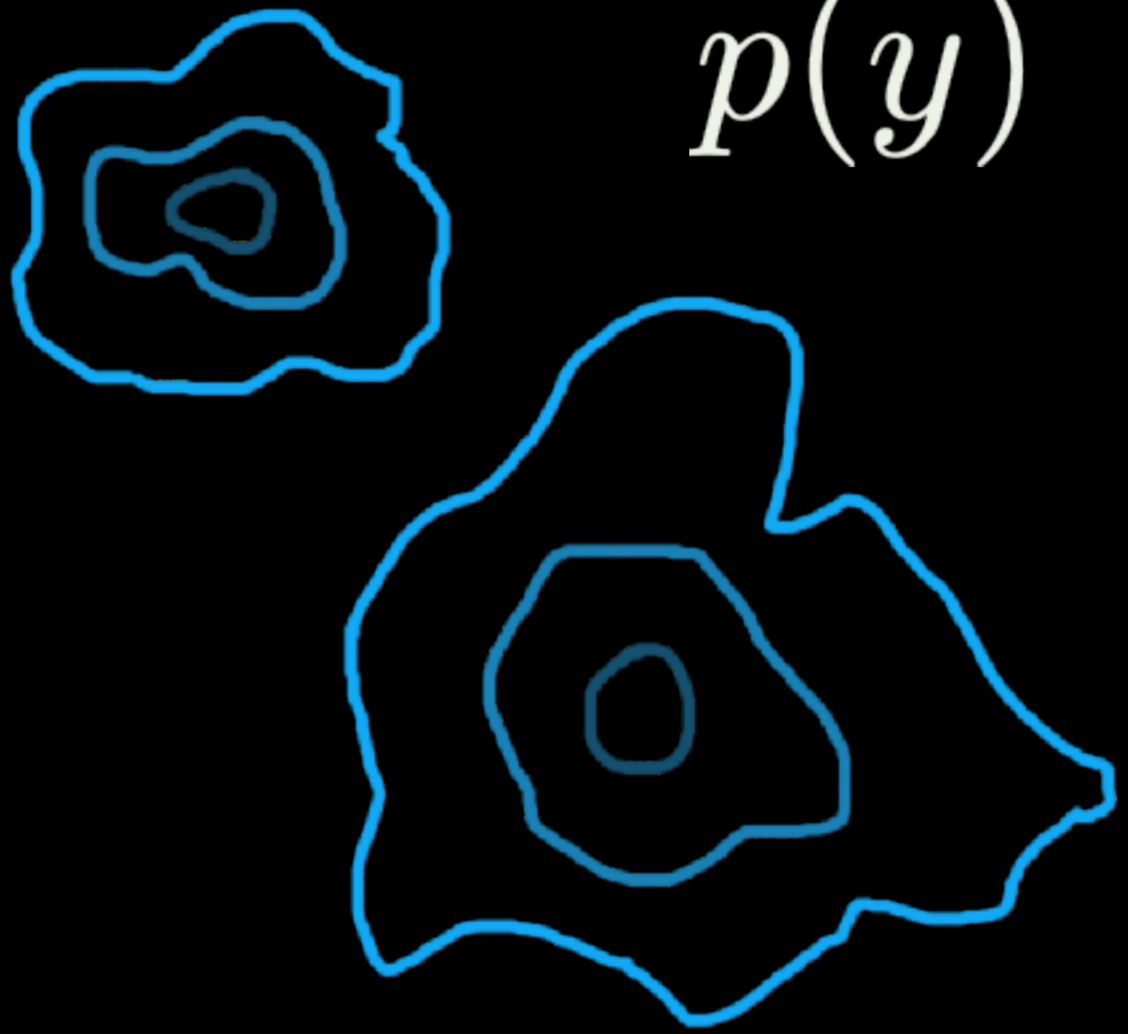
$$q(z) = \mathcal{N}(0, 1)$$

NORMALISING FLOWS

1. We have simple random generator
2. We want to sample from a more complex distribution



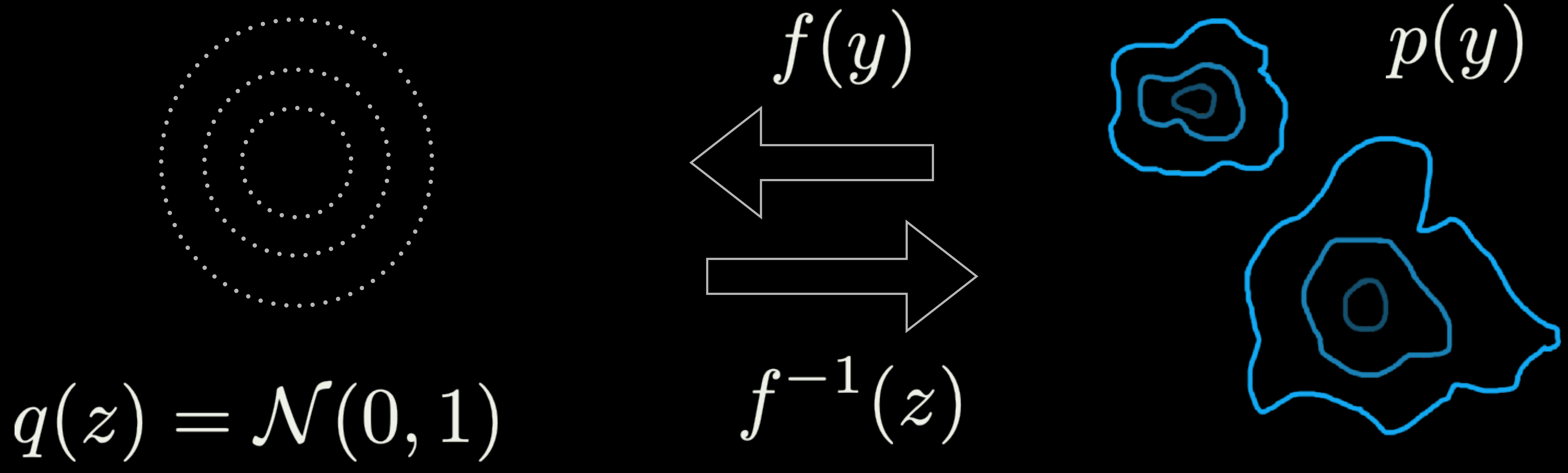
$$q(z) = \mathcal{N}(0, 1)$$



$$p(y)$$

NORMALISING FLOWS

1. We have simple random generator
2. We want to sample from a more complex distribution
3. We can estimate a bijective transformation which will allow us to do that



CHANGE OF VARIABLE EQUATION

$$p(y) = q(f(y))|\det(J_f(y))|$$

CHANGE OF VARIABLE EQUATION

$$p(y) = \boxed{q(f(y))} |\det(J_f(y))|$$

- f has to be a bijection

CHANGE OF VARIABLE EQUATION

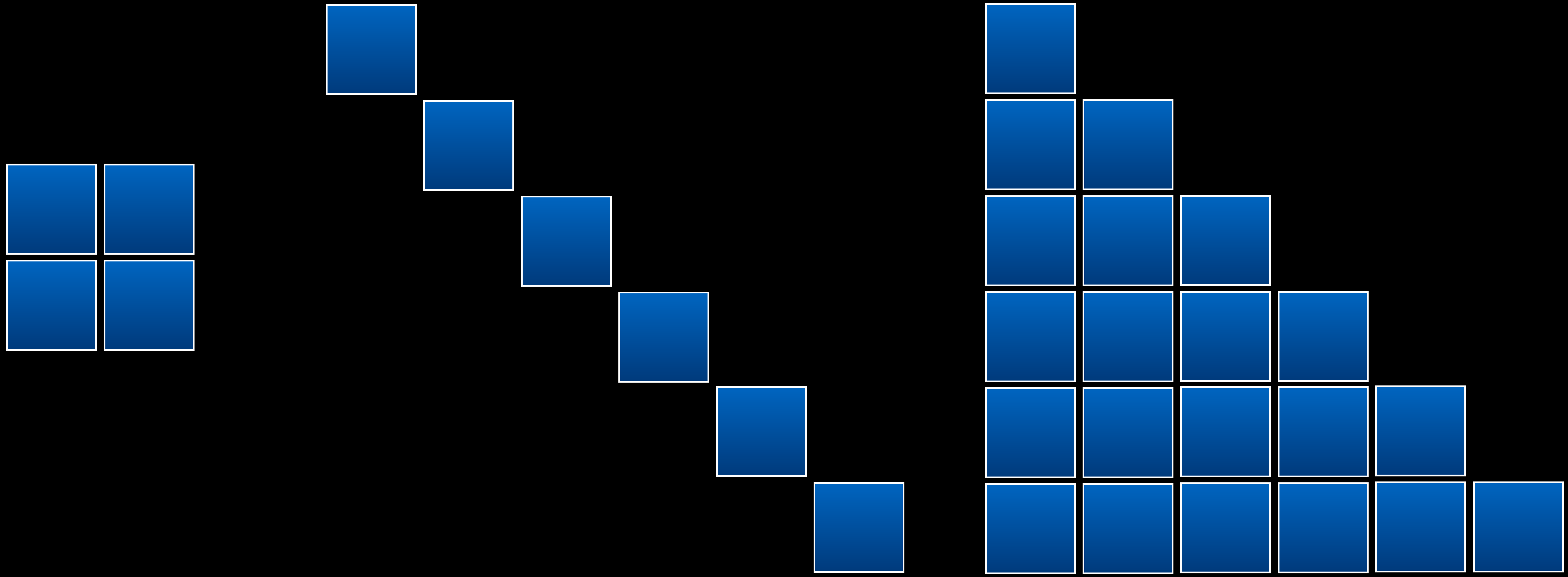
$$p(y) = q(f(y)) \left| \det(J_f(y)) \right|$$

- f has to be a bijection
- f and f^{-1} have to be differentiable
- Jacobian determinant has to be tractably invertable

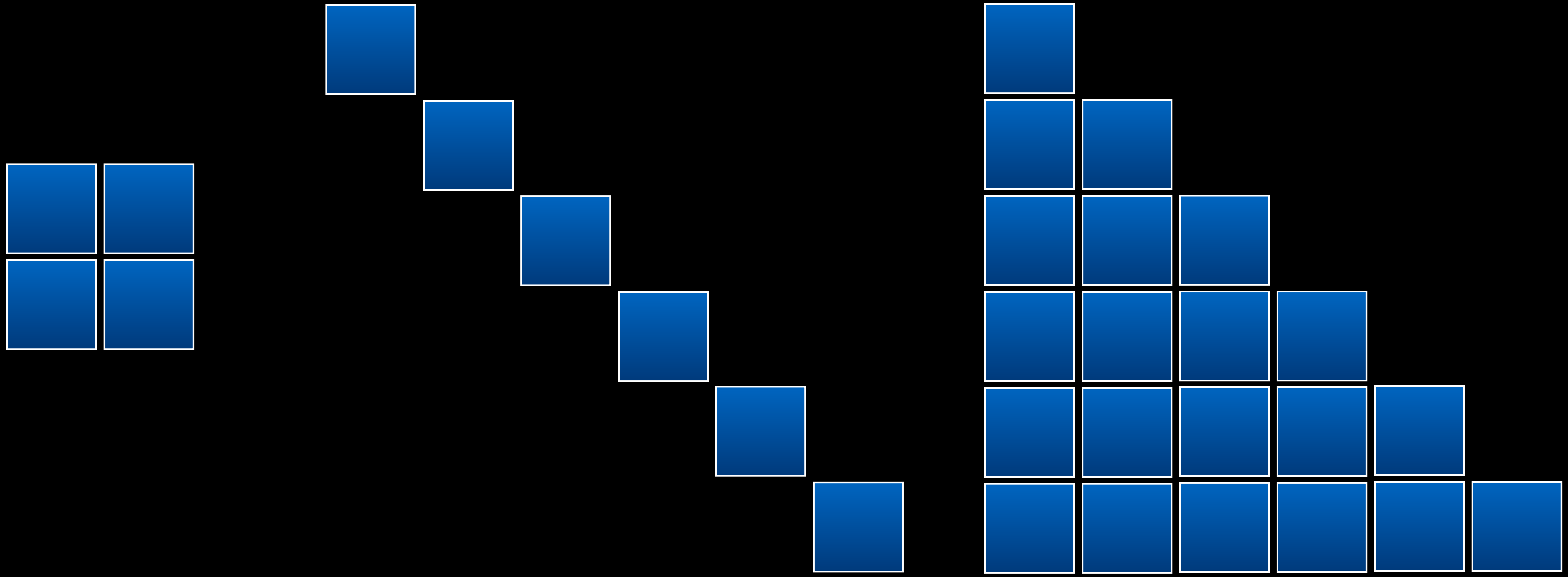
JACOBIAN

- The calculation of determinant Jacobian will take O(N^3)
- We need to speed it up
- For example, make Jacobian triangular matrix

JACOBIAN



JACOBIAN



Determinant of triangular matrix is a product of the elements on the diagonal

AFFINE TRANSFORM

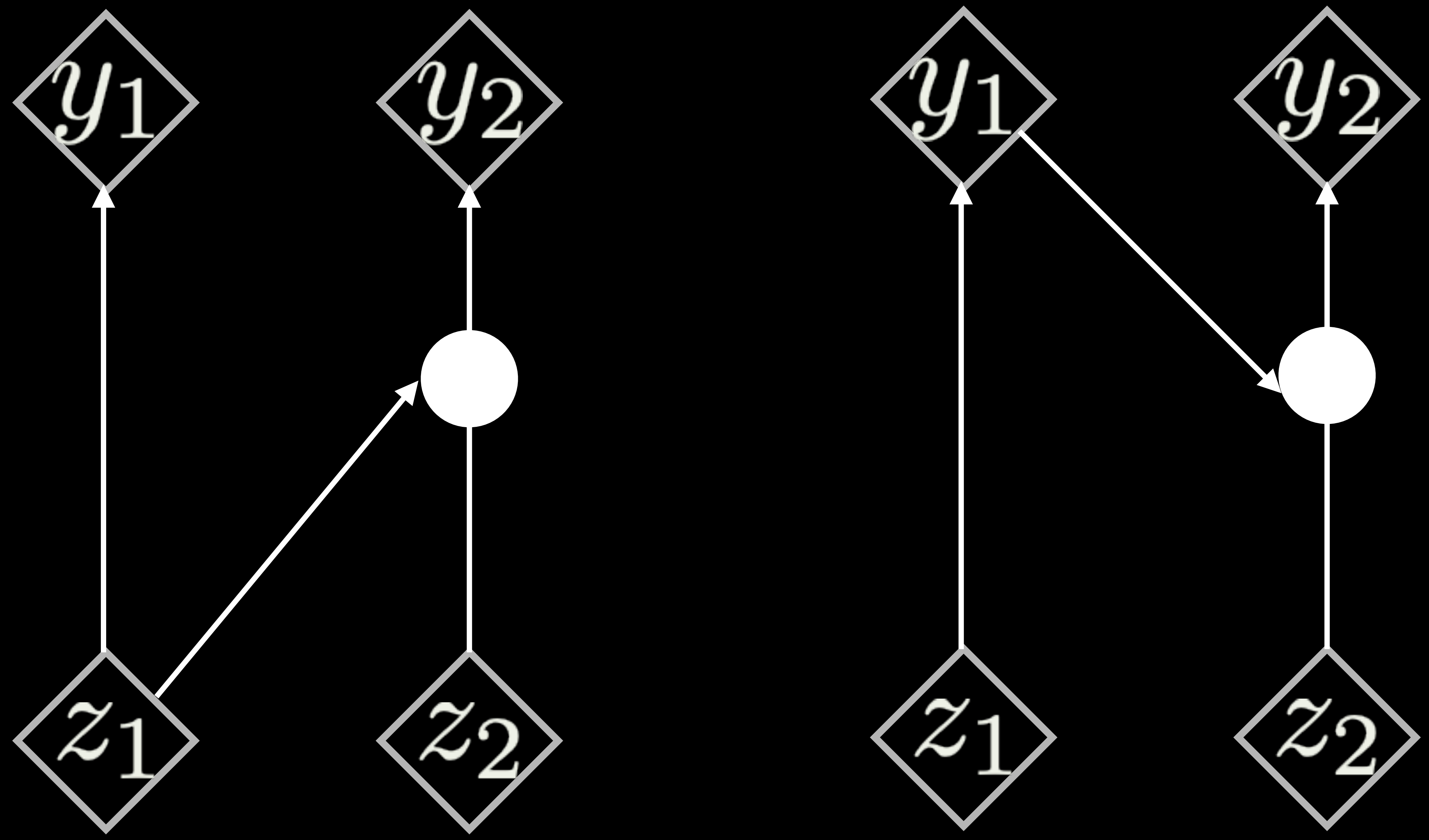
Location-scale transformation

$$\tau(z_i) = \alpha_i z_i + \beta_i$$

log-Jacobian becomes

$$\log |\det J_{g^{-1}}(z)| = \sum \log |\alpha_i|$$

COUPLING TRANSFORM



In each simple bijection,
part of the input vector
is updated using a function
which is simple to invert,
but which depends on the
remainder of the input vector
in a complex way.
The other part is left unchanged.

REAL NVP

Coupling transformation combined with affine transformation and its invention

$$\begin{cases} y_{1:d} &= x_{1:d} \\ y_{d+1:D} &= x_{d+1:D} \odot \exp\left(s(x_{1:d})\right) + t(x_{1:d}) \end{cases}$$

$$\Leftrightarrow \begin{cases} x_{1:d} &= y_{1:d} \\ x_{d+1:D} &= \left(y_{d+1:D} - t(y_{1:d})\right) \odot \exp\left(-s(y_{1:d})\right), \end{cases}$$

What is ***t*** and ***s***?

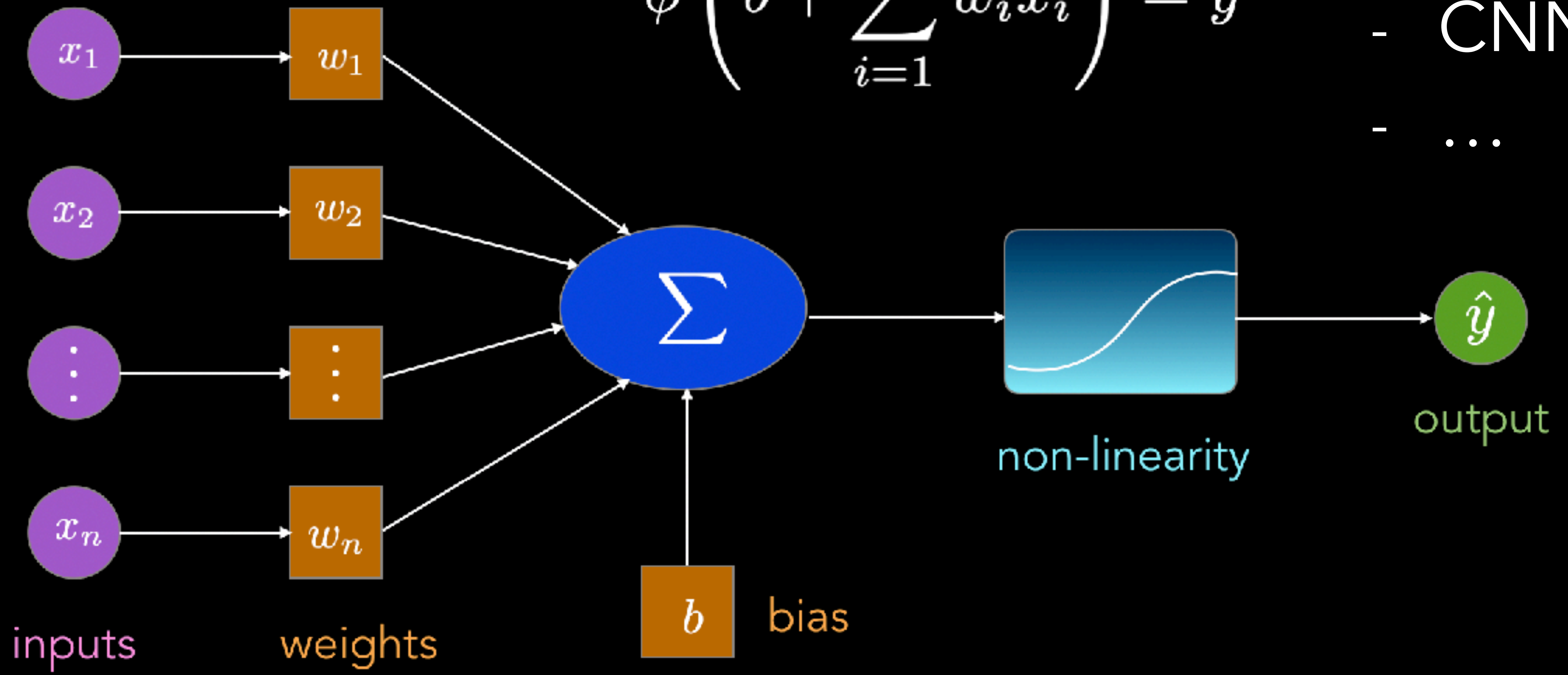
https://arxiv.org/abs/1605.08803

FUNCTION APPROXIMATION

can be parameterised by any NN:

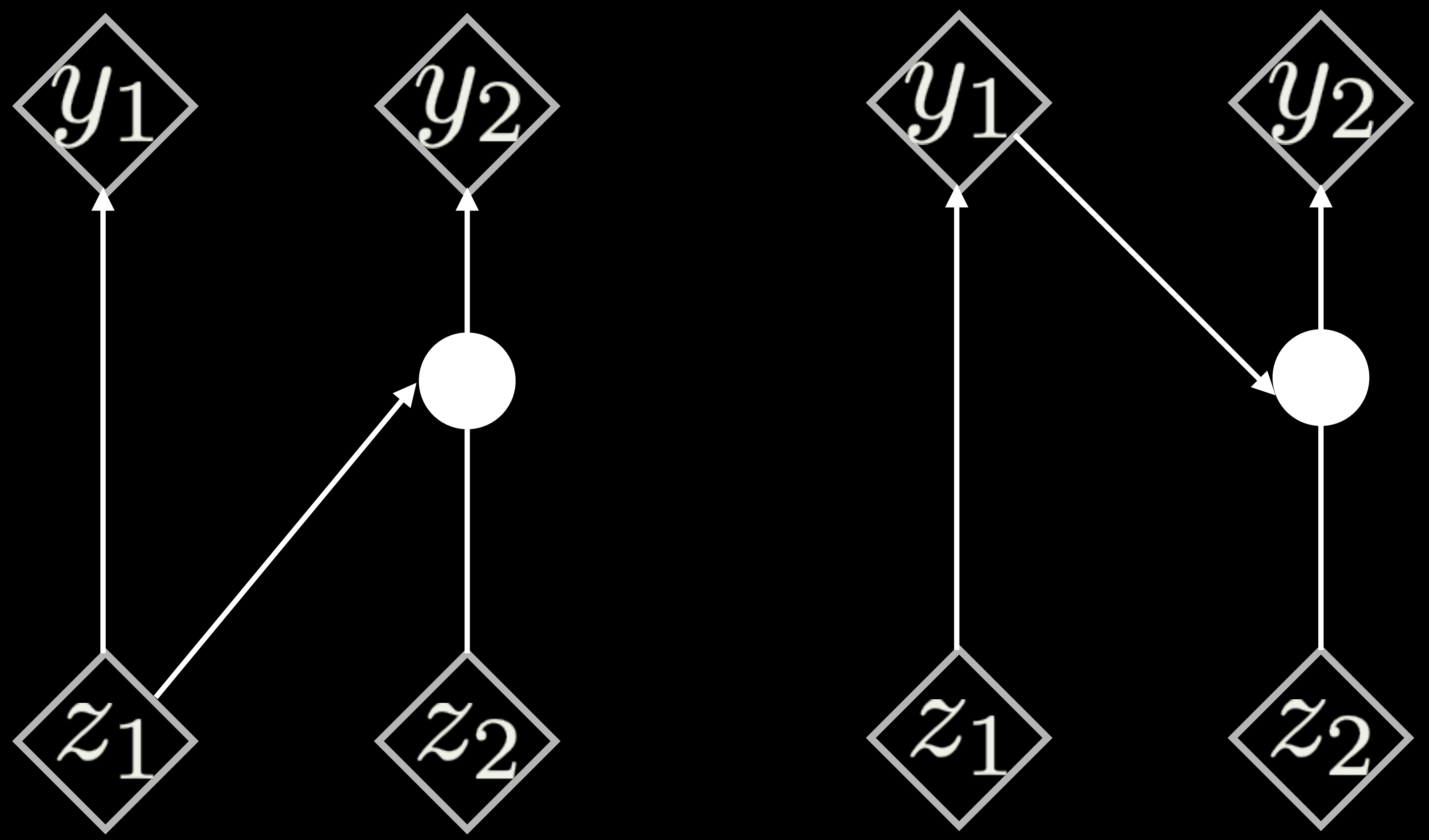
- Fully connected
- Residual
- CNN
- …

$$\phi\left(b + \sum_{i=1}^{n} w_i x_i\right) = \hat{y}$$



NEURAL SPLINE FLOWS

- Coupling transform



- Monotonic rational-quadratic spline transform

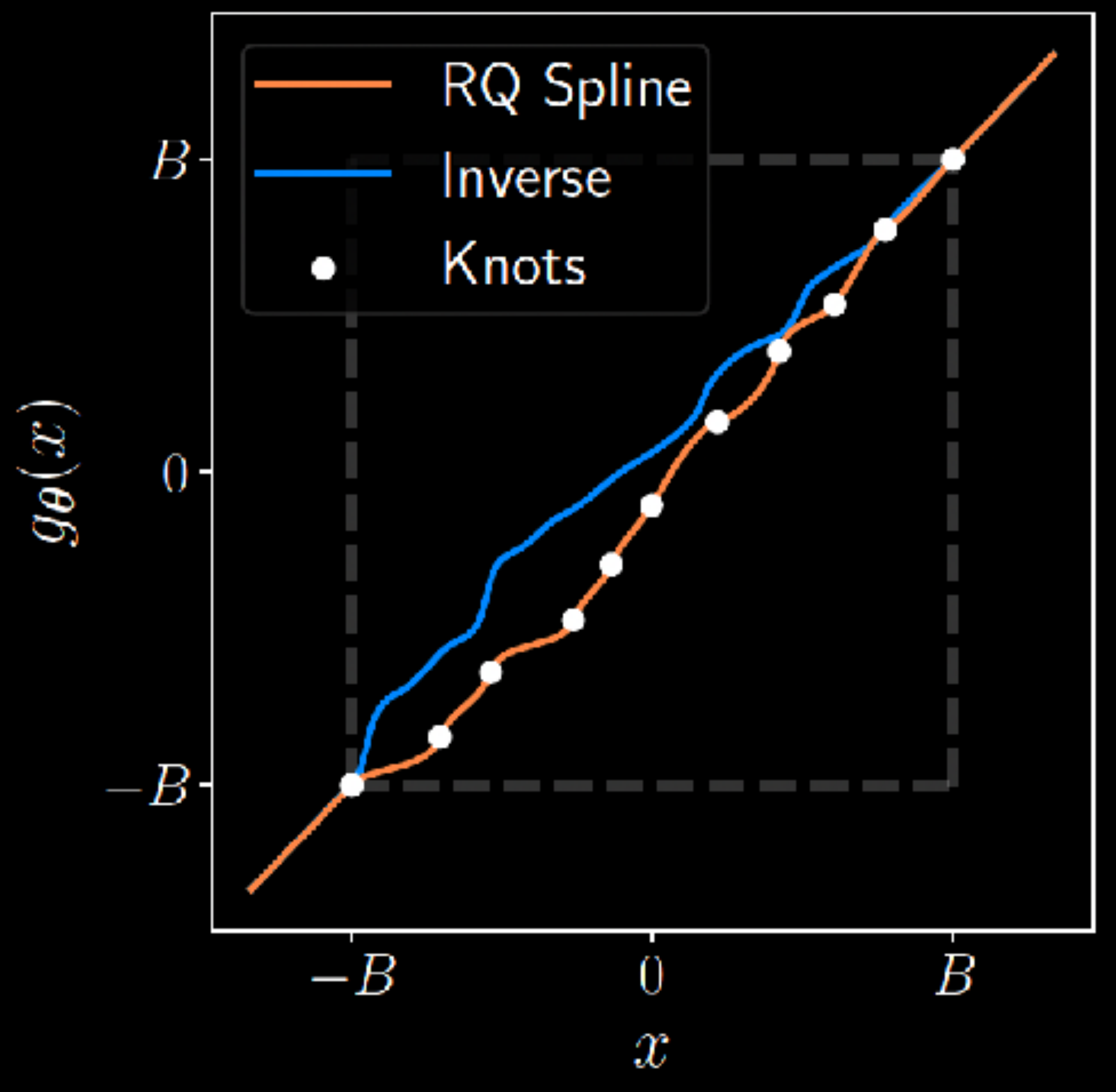
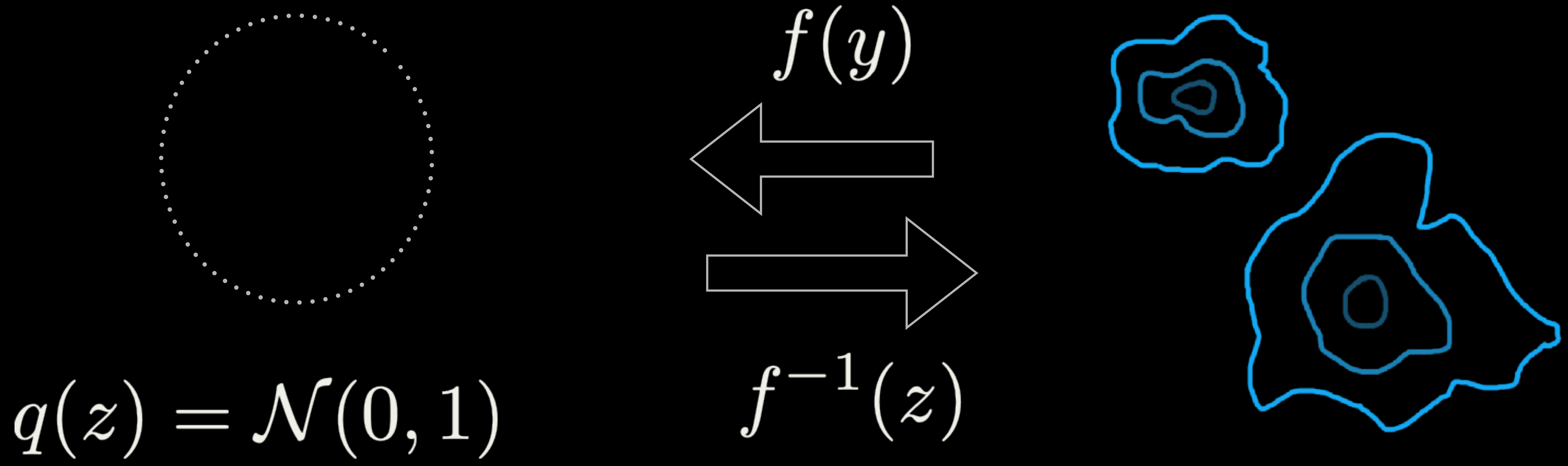


image: Duncan C. et al, Neural Spline Flows

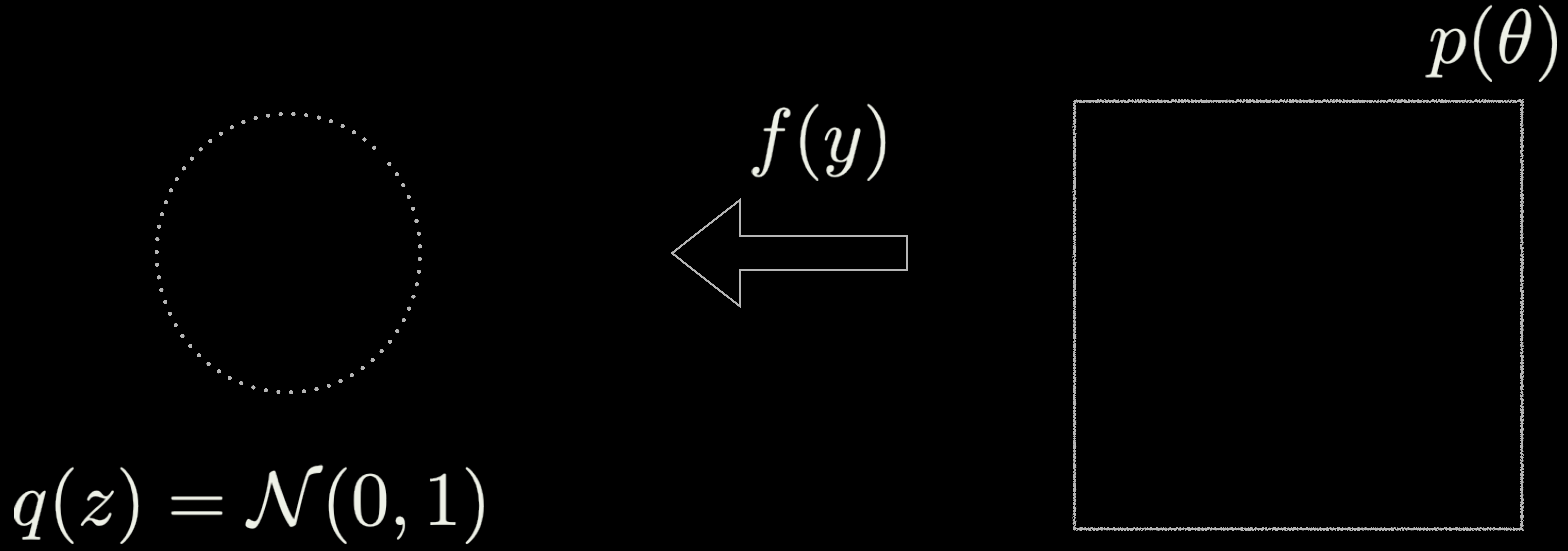
CONDITIONING

- Do not have access to samples from posterior



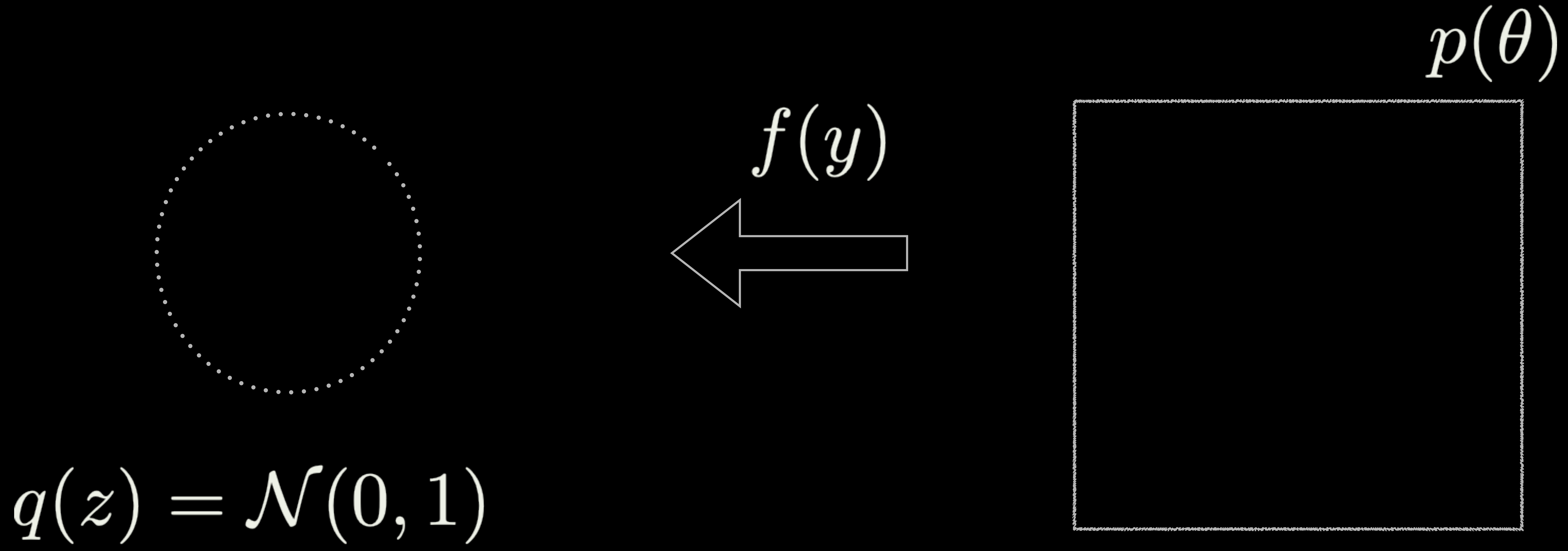
CONDITIONING

- Do not have access to samples from posterior
- Have access to samples from prior +



CONDITIONING

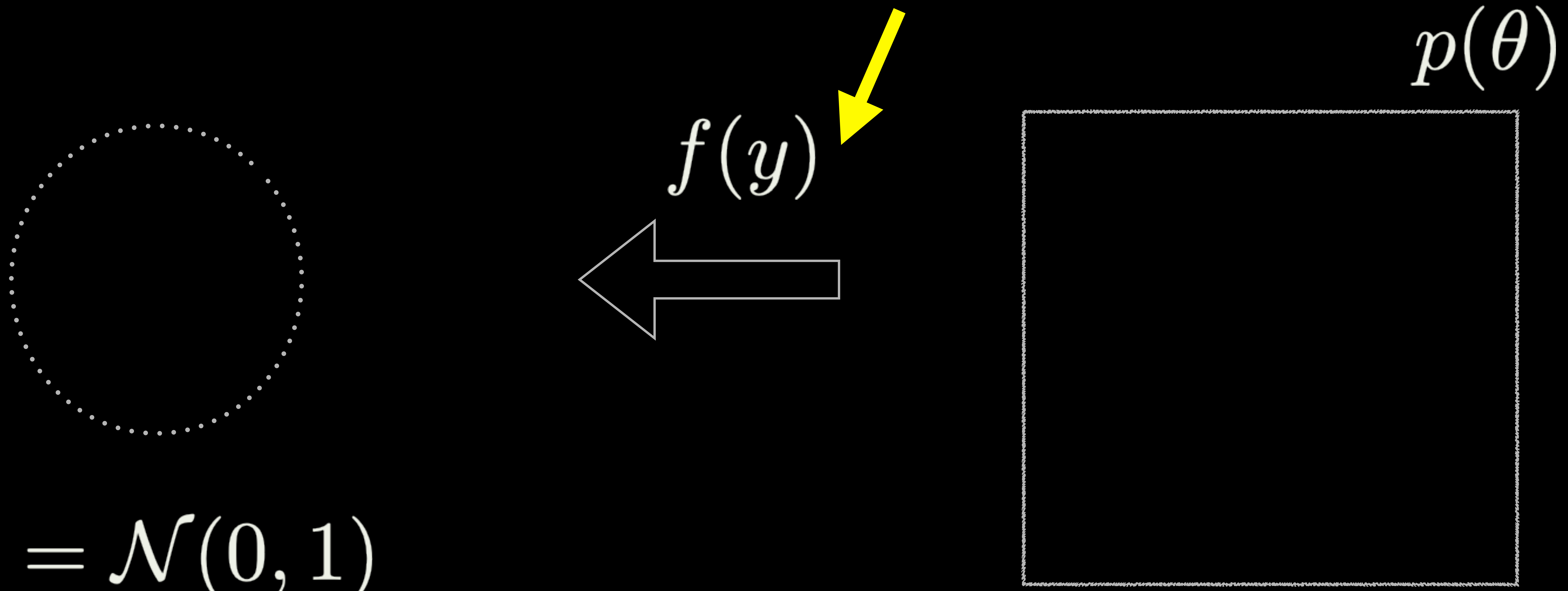
- Do not have access to samples from posterior
- Have access to samples from prior +
- Can generated simulated data $x = h(\theta) + n$



CONDITIONING

Condition map on simulated data

- Do not have access to samples from posterior
- Have access to samples from prior +
- Can generated simulated data $x = h(\theta) + n$



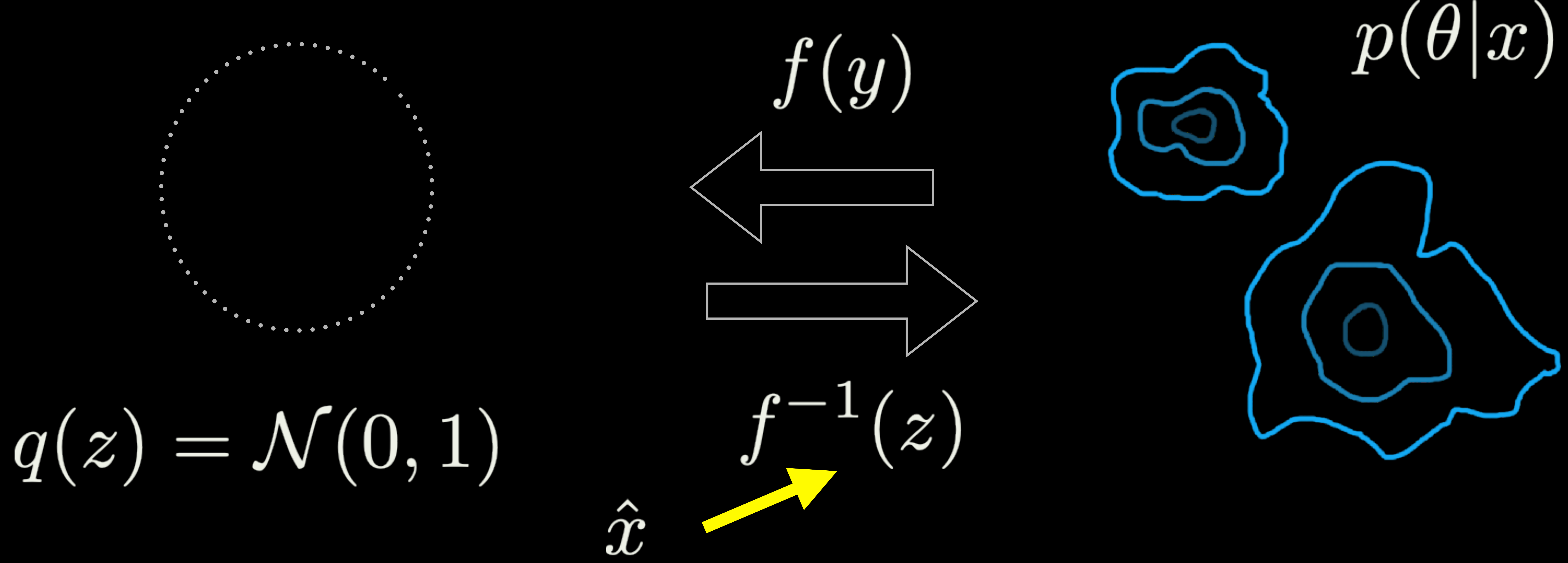
$$q(z) = \mathcal{N}(0, 1)$$

Therefore have access to the joint sample $p(x, \theta) = p(x|\theta)p(\theta)$

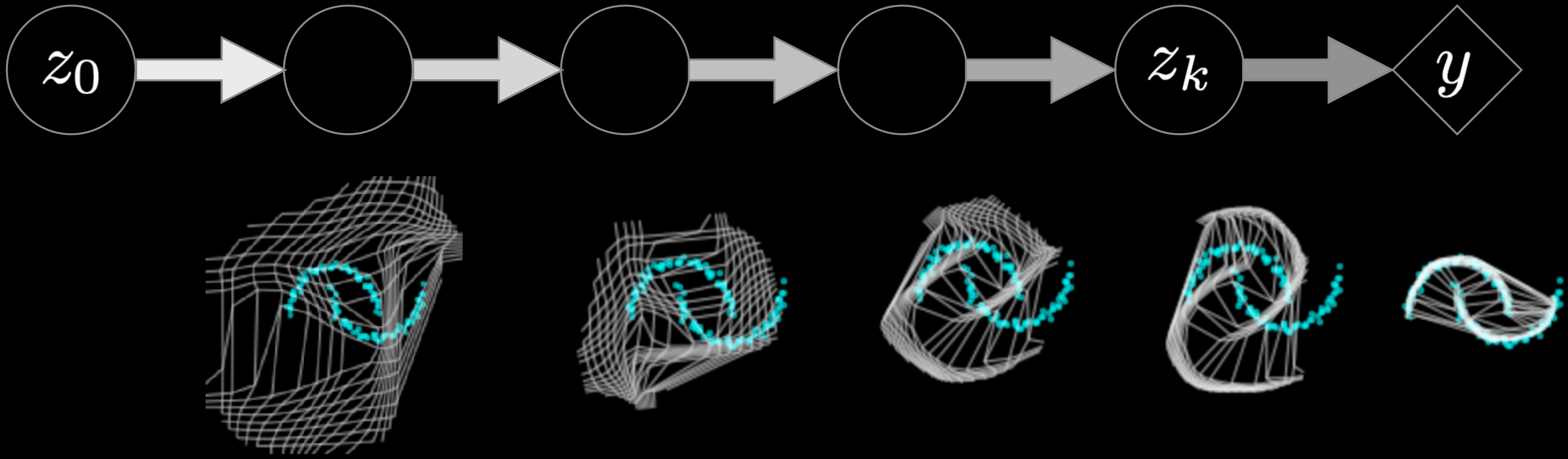
CONDITIONING

Condition inverted map on real data

- Do not have access to samples from posterior
- Have access to samples from prior +
- Can generated simulated data $x = h(\theta) + n$



COMPOSING FLOW



OPTIMISATION

- The flow is trained to maximise the total log likelihood of the data with respect to the parameters of the transform.

$$\log p(y|\lambda) = \sum_{i=1}^{N} \log \left[p(y\prime_i|\lambda)\right])$$

WAVEFORM EMBEDDING

- Low frequency sensitivity -> long waveforms
- Construct reduced orthogonal basis
- Use coefficients of the waveform projection on a new basis

WAVEFORM EMBEDDING

Decompose a matrix constructed of the set of waveforms

$$\mathbf{H} = \mathbf{V}\mathbf{\Sigma}\mathbf{U}^{\mathbf{T}}$$

WAVEFORM EMBEDDING

Decompose a matrix constructed of the set of waveforms

$$\mathbf{H} = \mathbf{V}\boldsymbol{\Sigma}\mathbf{U}^{\mathrm{T}}$$

Project sample simulated data on this basis

$$v'_{\alpha\mu} = \frac{1}{\sigma_\mu}\sum_{j=1}^{N} h_{\alpha j} u_{\mu j}$$

RESULTS

PRELIMINARY

CONCLUSIONS

- Alternative sampling method
- Can be used for low latency pipeline
- Can be used to approximate complex distributions
- Can use embedded data representations