Fast Parameter Estimation for Massive Black Hole Binaries with Normalising Flows

Natalia Korsakova
SYRTE/Paris Observatory

SYstèmes de Référence Temps-Espace

Laser

Interferometer

Space

Antenna

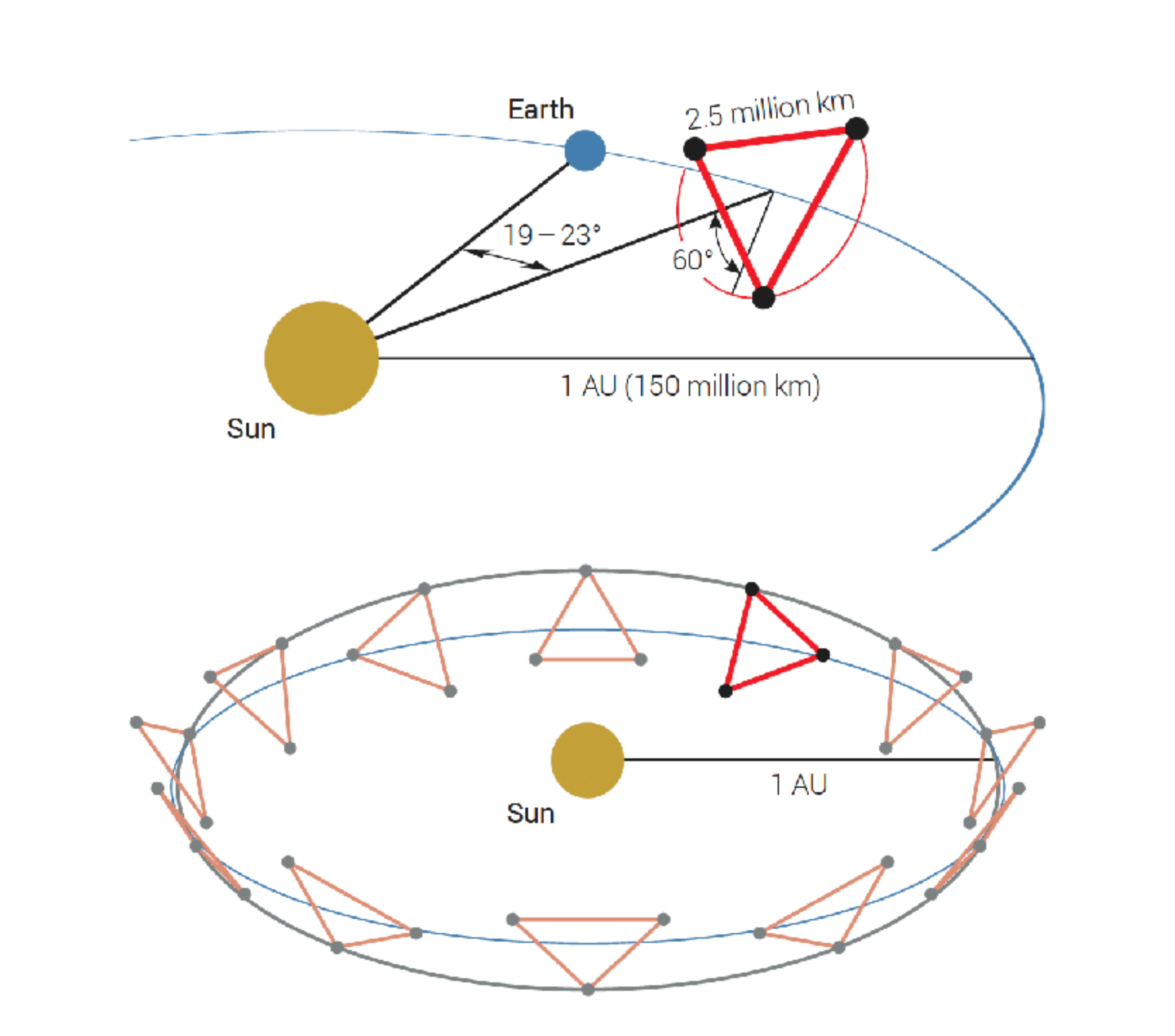


Image: LISA White paper

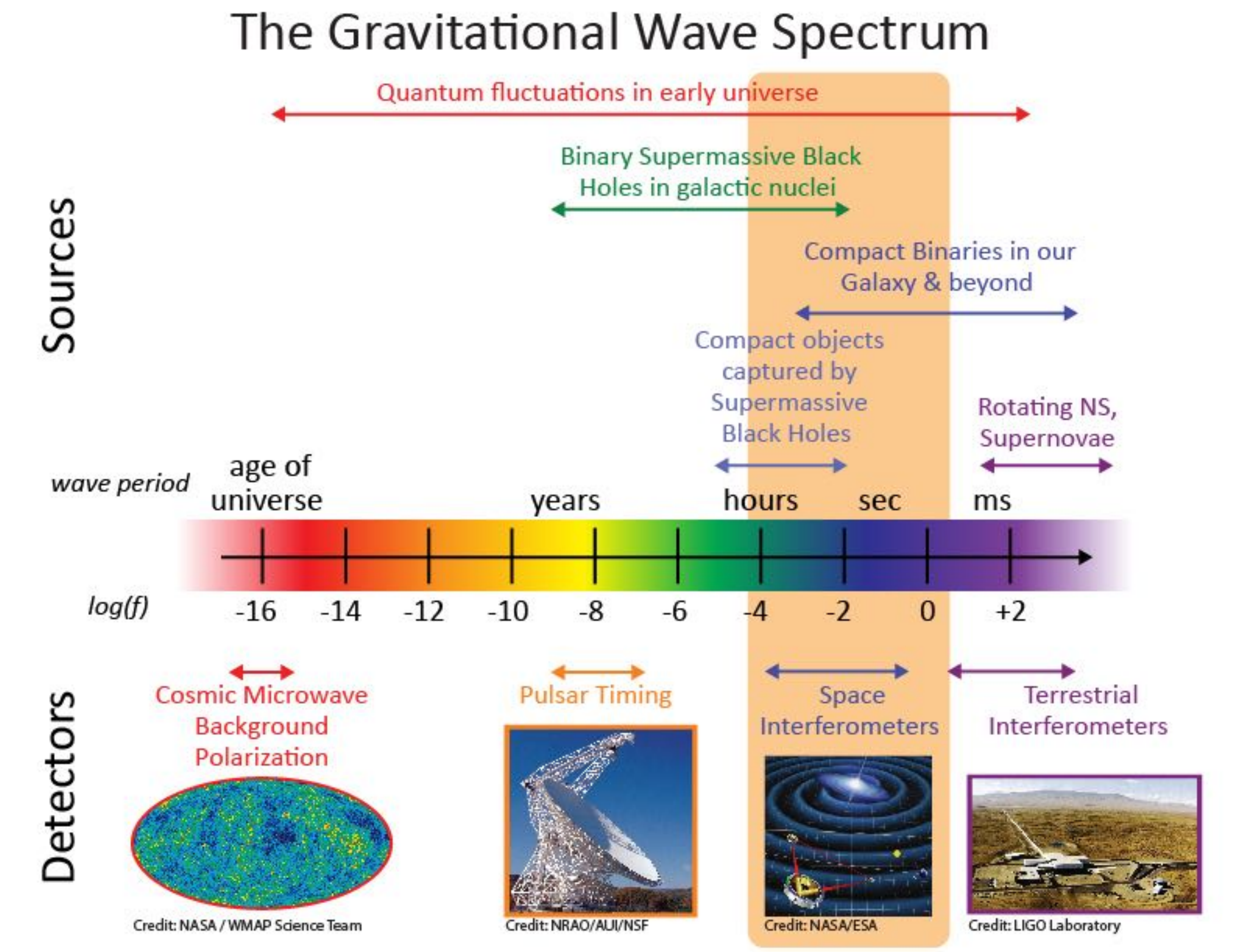
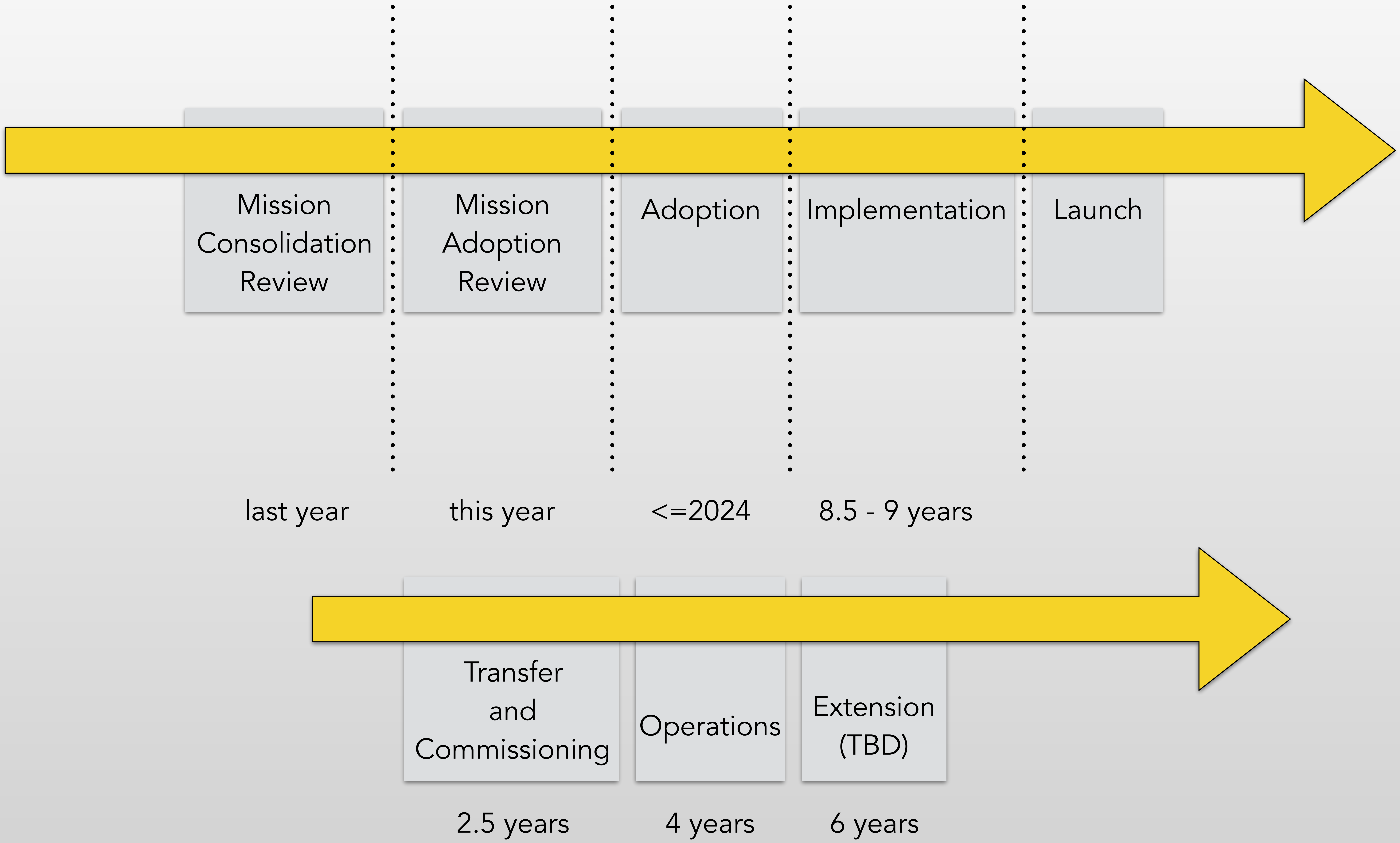

The Gravitational Wave Spectrum
Quantum fluctuations in early universe
Binary Supermassive Black Holes in galactic nuclei
Compact Binaries in our Galaxy & beyond
Compact objects captured by Supermassive Black Holes
Rotating NS, Supernovae
Sources
wave period
age of universe
years
hours
sec
ms
log(f)
-16
-14
-12
-10
-8
-6
-4
-2
0
+2
Detectors
Cosmic Microwave Background Polarization
Pulsar Timing
Space Interferometers
Terrestrial Interferometers
Credit: NASA / WMAP Science Team
Credit: NRAO/AUI/NSF
Credit: NASA/ESA
Credit: LIGO Laboratory

Image: NASA

TIMELINE



Mission Consolidation Review
Mission Adoption Review
Adoption
Implementation
Launch
last year
this year
<=2024
8.5 - 9 years
Transfer and Commissioning
Operations
Extension (TBD)
2.5 years
4 years
6 years

LISA NOISE AND SOURCES

- Massive Black Hole Binaries
- Galactic Binaries
- Extreme Mass Ratio Inspirals
- Stellar Origin Black Hole Binaries
- …

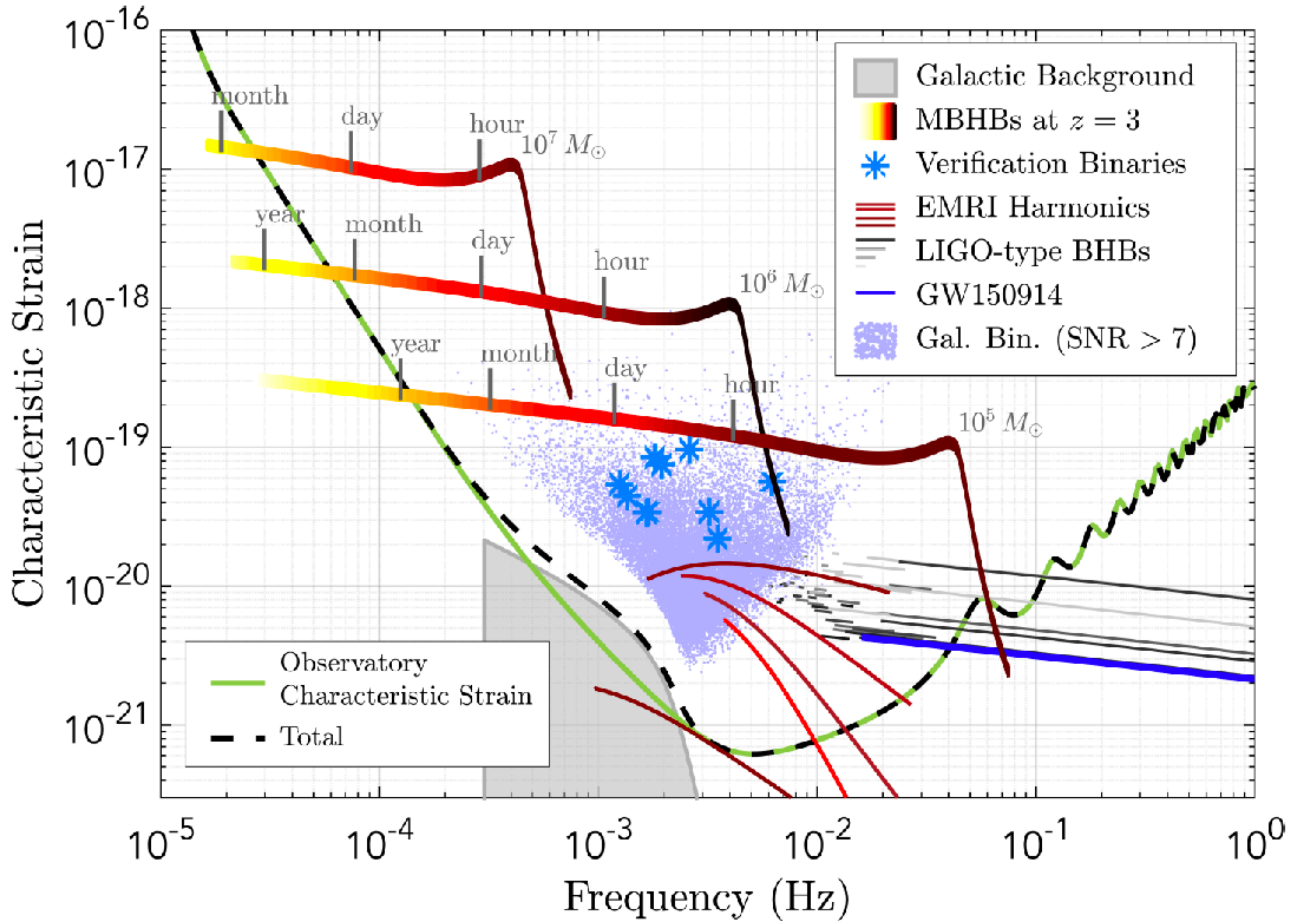


Image: LISA White paper

MASSIVE BLACK HOLE BINARIES

Signals from MBHB mergers observed by LISA depend on
- assumptions regarding MBH formation,
- the recipes employed for the black hole mass growth via
  merger and
  gas accretion.

We consider two main scenarios for black hole formation

- "light seed" scenario ($\simeq 10^2 M_\odot$)
  remnant of Population III stars formed in low metallicity environment at $z \sim 15$-20

- "heavy seed" scenario ($>= 10^4 M_\odot$)
  direct collapse of protogalactic disk

MBHB POPULATION

heavy seed scenario with efficient formation of
black hole seeds in a large fraction of high-redshift haloes
-> hundreds a year

1. seeds are light, and many coalescences do not fall into the LISA band,
2. seeds are massive, but rare

->tens a year

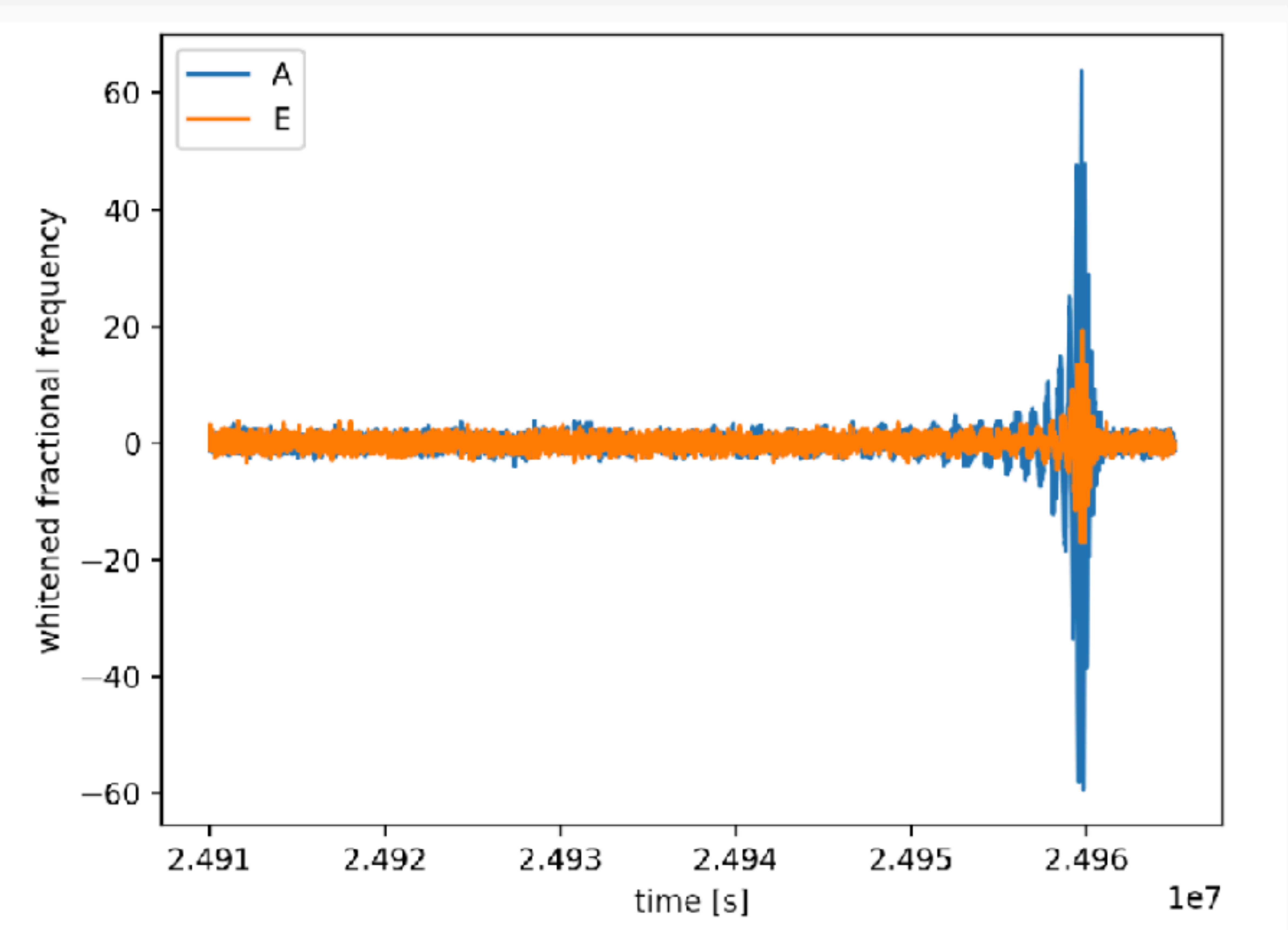
Massive Black Hole Binaries
— 10 to 100 sources / year

MASSIVE BLACK HOLE BINARIES EM COUNTERPARTS

Multiple authors suggest that
the electromagnetic counterparts will be observed
as a transient during merger or also during inspiral and merger.

Electromagnetic counterparts will occur
due to presence of
- matter or
- magnetic fields.

For example:
- Accretion during merger
- Jets produced by the external magnetic fields
- …



INFERENCE

We can estimate the posterior probability distribution of the parameters using Bayes' theorem

$$p(y|x) = \frac{p(x|y)\,p(y)}{p(x)}$$

Prior

Likelihood

Marginal likelihood

INFERENCE

The problem is that we have to compute marginal likelihood for the observation:

$$p(\mathbf{x}) = \int p(\mathbf{x}, \mathbf{z}) \mathrm{d}\mathbf{z}$$

That are the difference way to estimate marginal probability

INFERENCE

It is not possible to perform exact inference for the general problem.

We have to introduce some simplifications.

INFERENCE

It is not possible to perform exact inference for the general problem.
We have to introduce some simplifications.

We can use approximate inference:

- Sample from the exact posterior: MCMC or Nested sampling (slow)
- Variational Inference: approximate the posterior distribution with a tractable distribution

INFERENCE

It is not possible to perform exact inference for the general problem.
We have to introduce some simplifications.

We can use approximate inference:

- Sample from the exact posterior: MCMC or Nested sampling (slow)
- Variational Inference: approximate the posterior distribution with a tractable distribution

There are some exceptions for the models with some simplifications:

- Gaussian mixture models (Very simplified)
- **Invertible models**

INVERTIBLE TRANSFORM

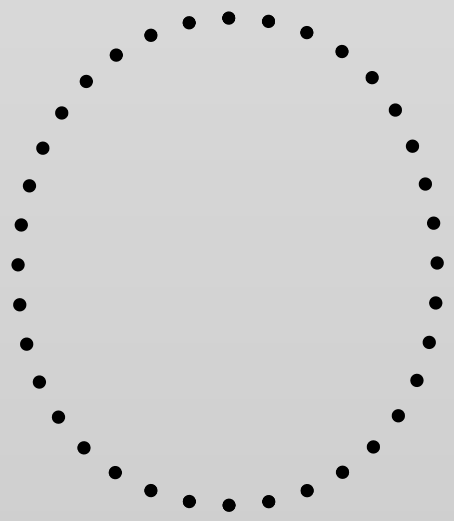
The basic idea:

INVERTIBLE TRANSFORM

The basic idea:

1. we have a simple random generator;

$$z \sim f_Z(z)$$



For example: $z \sim \mathcal{N}(0, 1)$

INVERTIBLE TRANSFORM

The basic idea:

1. we have a simple random generator;
2. we want want to transform it to be able to sample from a more complex distribution expression for which we do not know;

$z \sim f_Z(z)$ $\qquad\qquad$ $y \sim f_Y(y)$

For example: $z \sim \mathcal{N}(0, 1)$

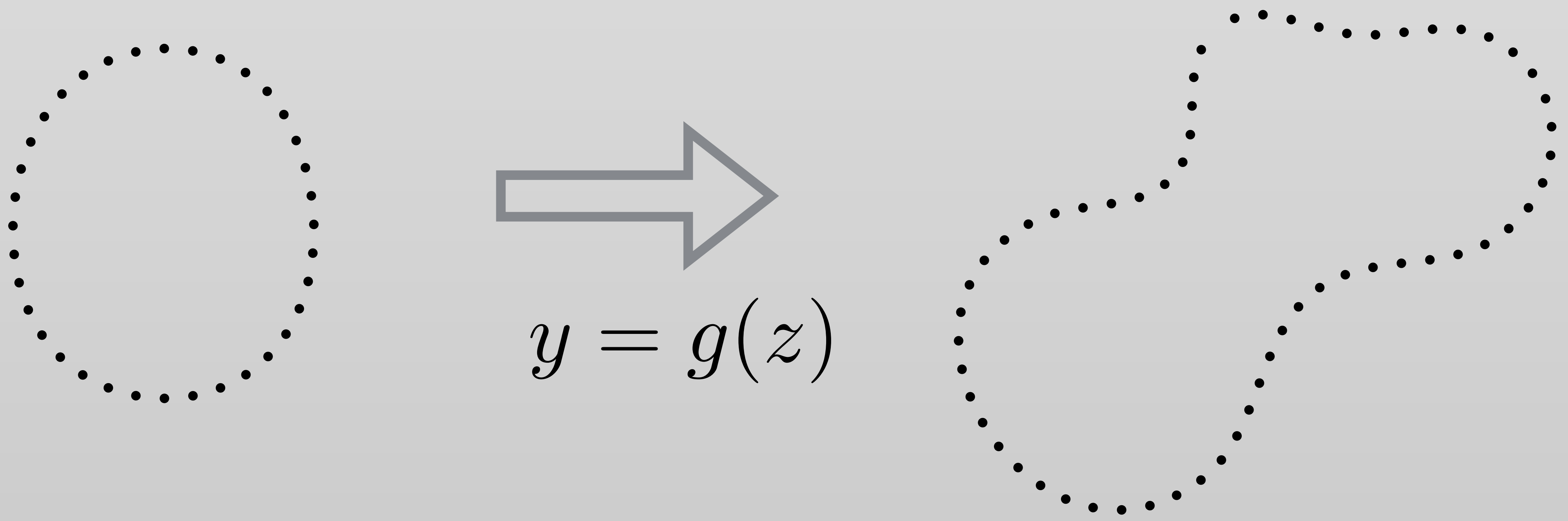
INVERTIBLE TRANSFORM

The basic idea:

1. we have a simple random generator;
2. we want want to transform it to be able to sample from a more complex distribution expression for which we do not know;
3. we pass it through a *bijective* transformation to produce a more complex variable.

$z \sim f_Z(z)$

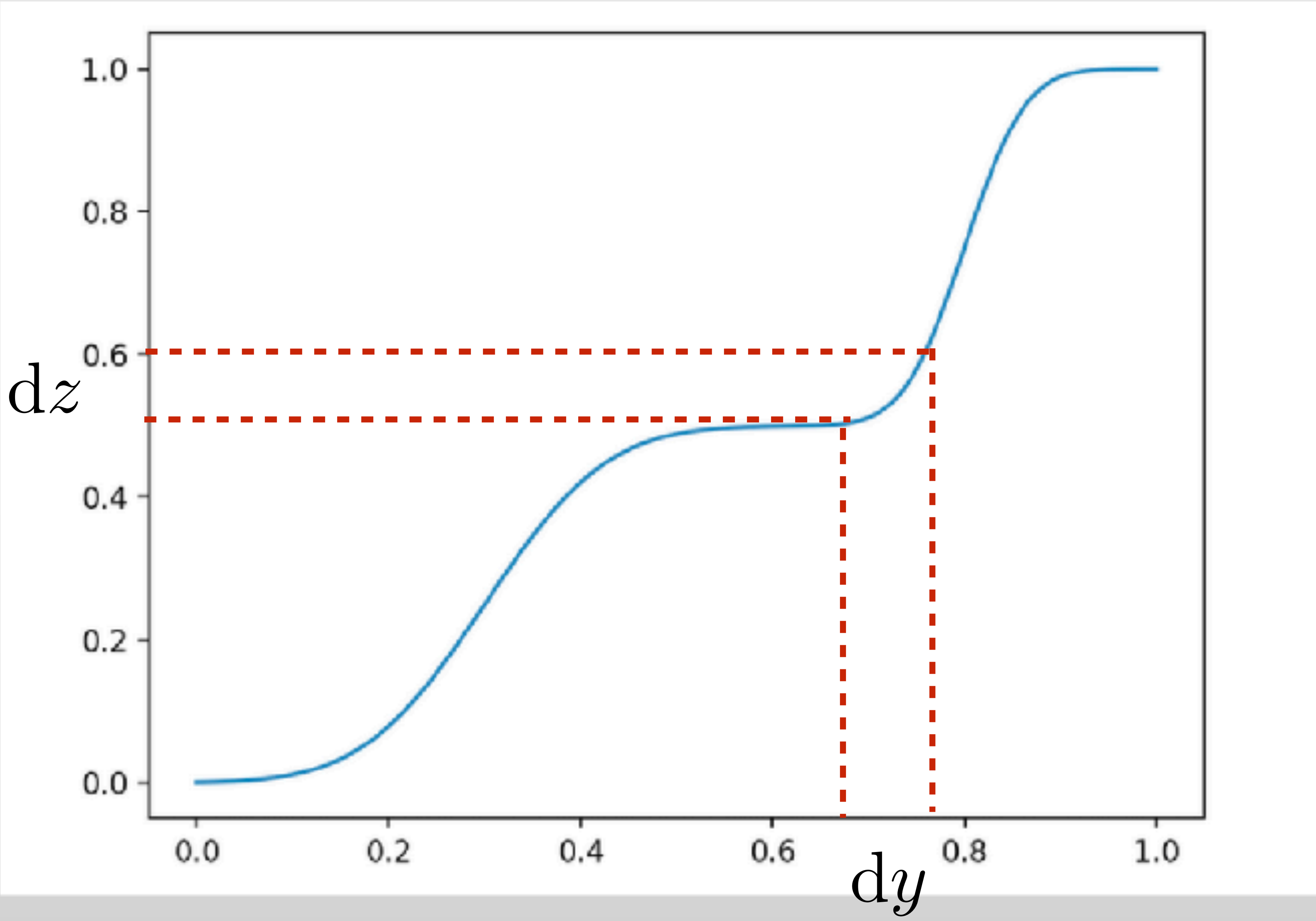
$y \sim f_Y(y)$



$y = g(z)$

For example: $z \sim \mathcal{N}(0, 1)$

CHANGE OF VARIABLE



$$f_Z(z)\mathrm{d}z = f_Y(y)\mathrm{d}y$$

$$f_Y(y) = f_Z(z)\left|\frac{\mathrm{d}z}{\mathrm{d}y}\right|$$

CHANGE OF VARIABLE

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y)$$

$$= \frac{\mathrm{d}}{\mathrm{d}y} F_Z(g^{-1}(y))$$

Chain rule

$$= f_Z(g^{-1}(y)) \left| \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) \right|$$

CHANGE OF VARIABLE

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y)$$

$$= \frac{\mathrm{d}}{\mathrm{d}y} F_Z(g^{-1}(y))$$

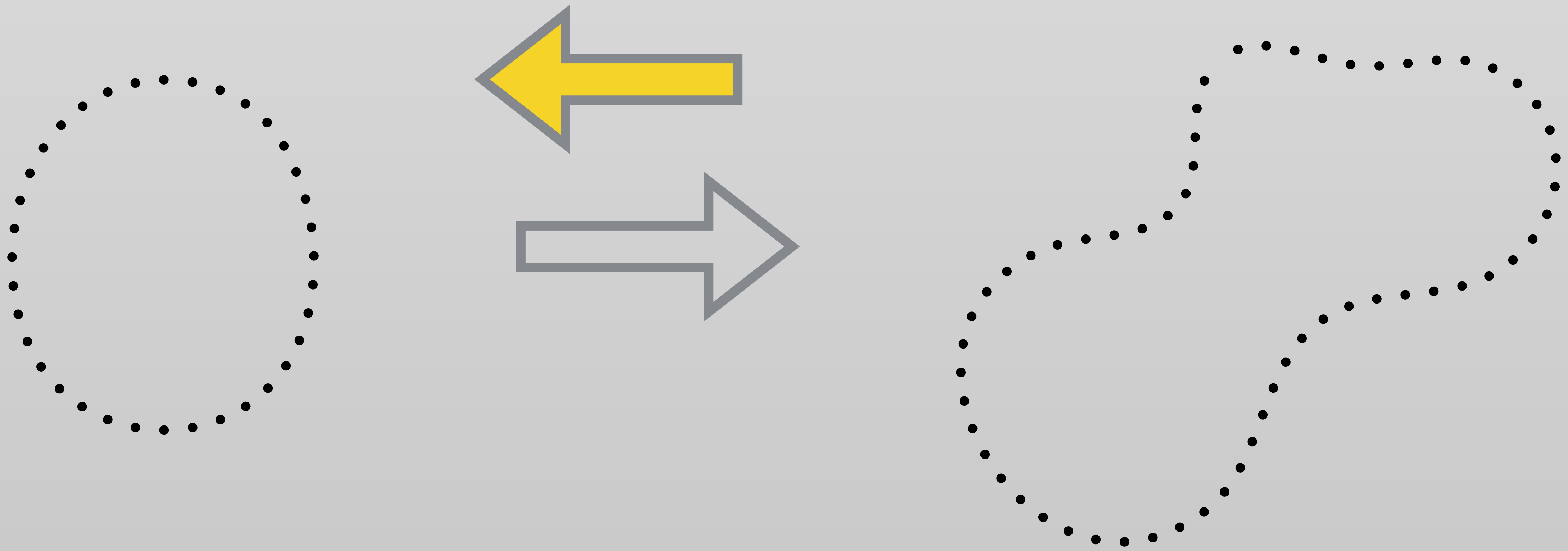
Chain rule

$$= f_Z(g^{-1}(y)) \left| \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) \right|$$

Multidimensional case

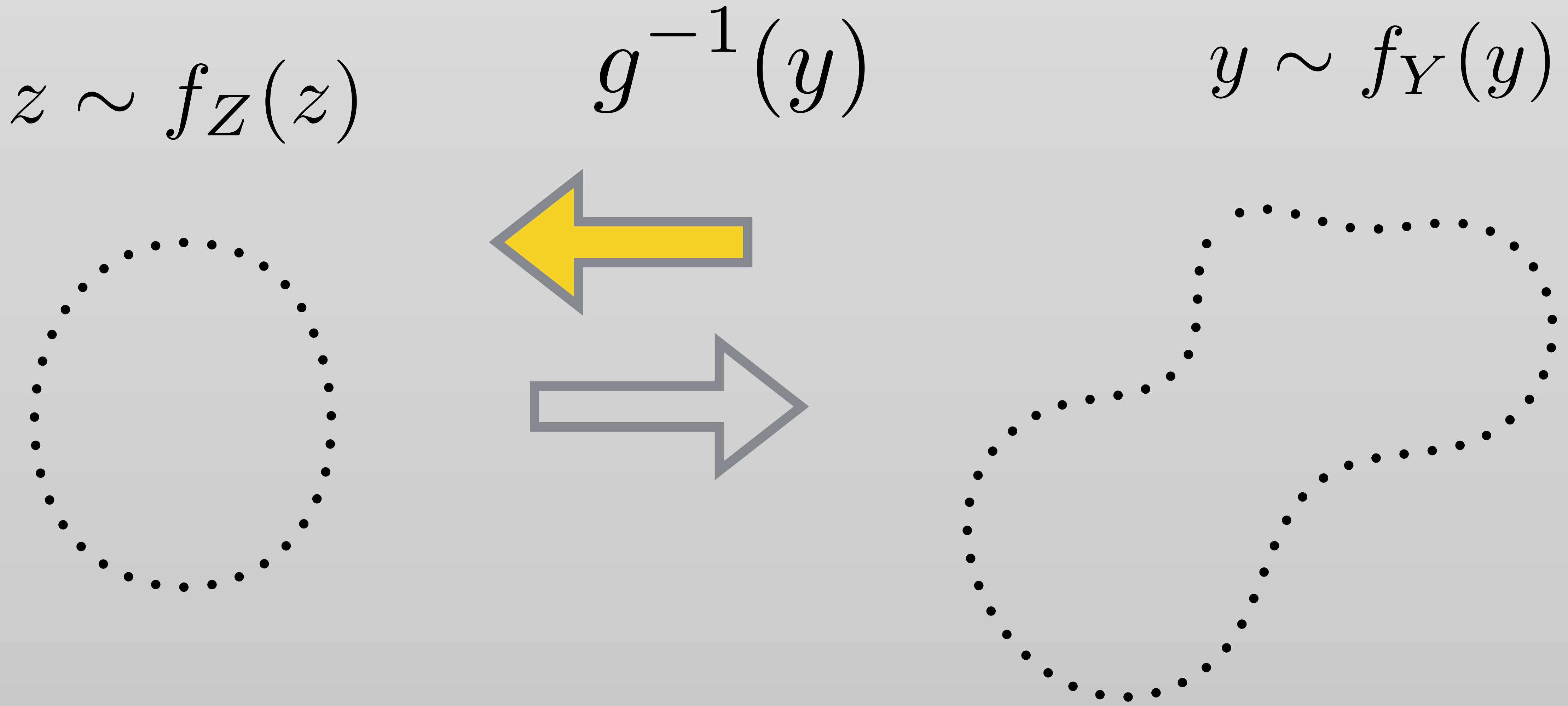
$$f_Y(y) = f_Z(g^{-1}(y)) \left| \det \frac{\partial g^{-1}(y)}{\partial y} \right|$$

$$g^{-1}(y)$$



CHANGE OF VARIABLE

$$\log[f_Y(y)] = \log[f_Z(g^{-1}(y))] + \log\left[\left|\det\frac{\partial g^{-1}(y)}{\partial y}\right|\right]$$

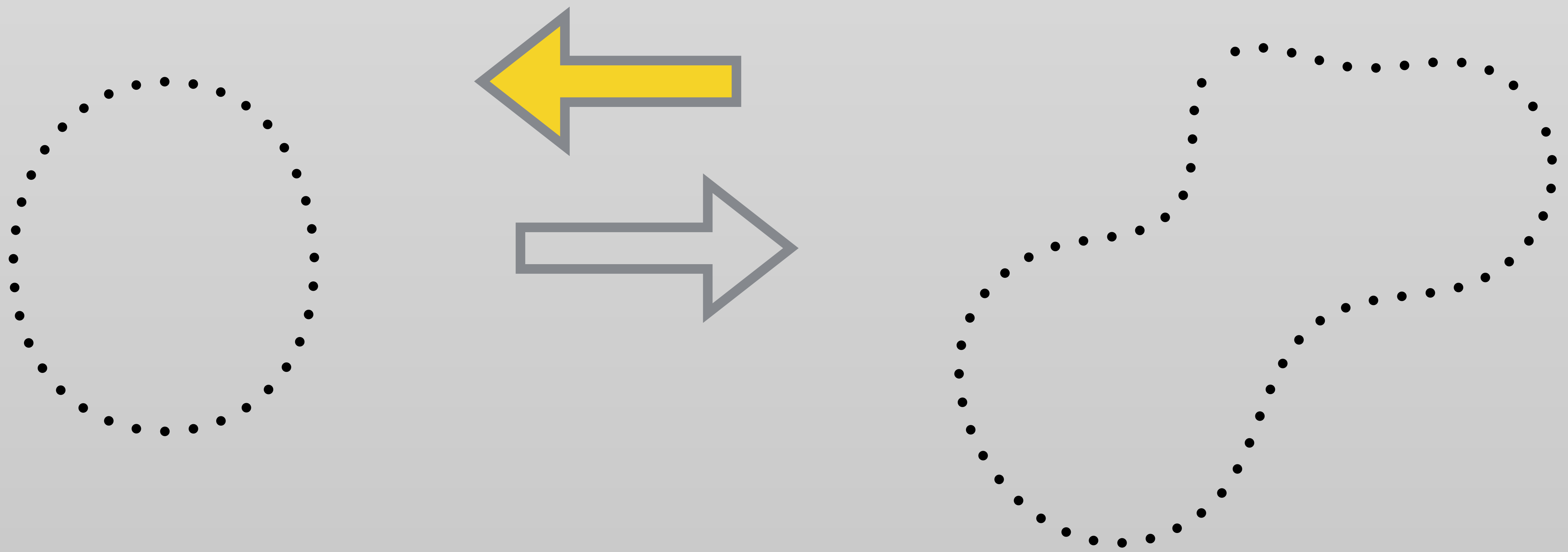


CHANGE OF VARIABLE

$$\log[f_Y(y)] = \log[f_Z(g^{-1}(y))] + \log\left[\left|\det\frac{\partial g^{-1}(y)}{\partial y}\right|\right]$$

1. $g(y)$ has to be a bijection

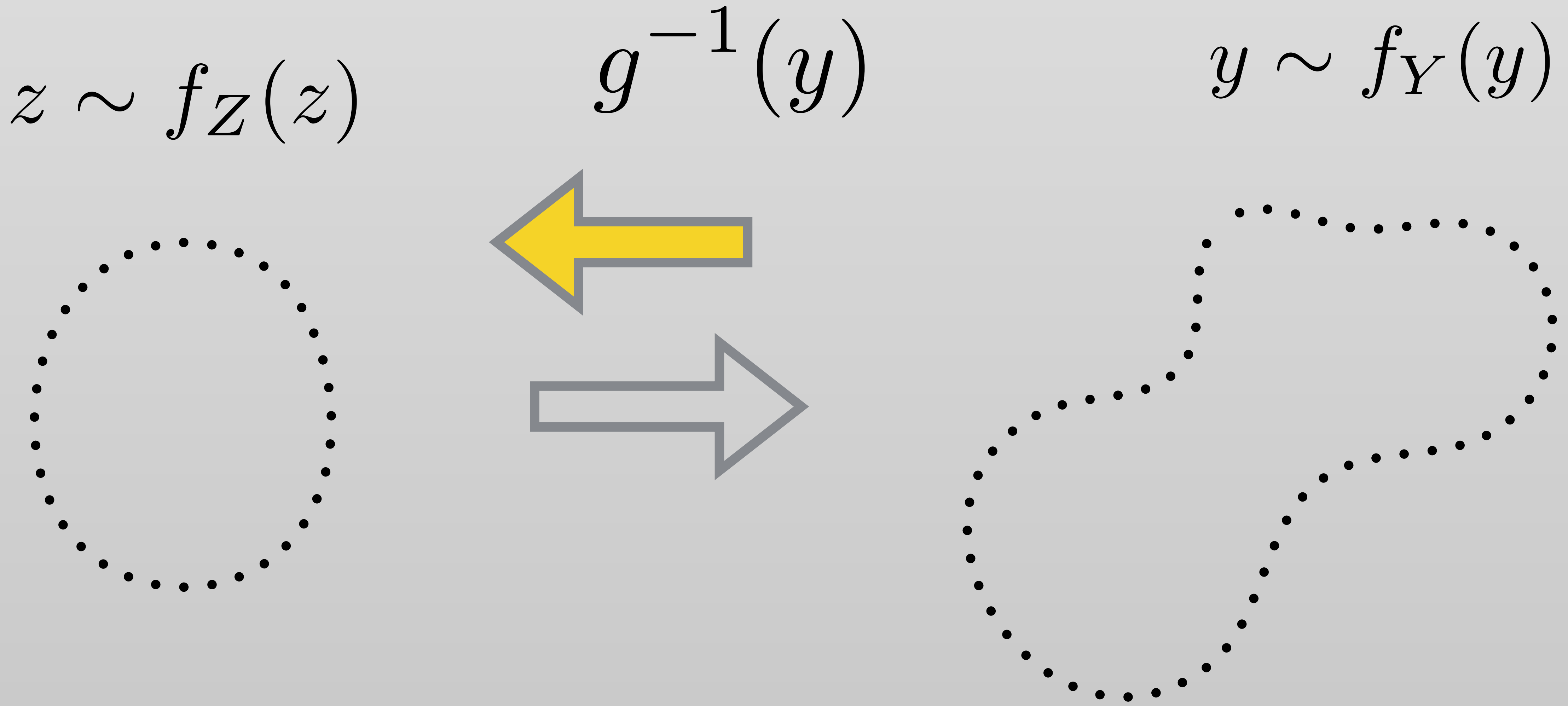
$z \sim f_Z(z)$ $\quad g^{-1}(y) \quad$ $y \sim f_Y(y)$



CHANGE OF VARIABLE

$$\log[f_Y(y)] = \log[f_Z(g^{-1}(y))] + \log\left[\left|\det \frac{\partial g^{-1}(y)}{\partial y}\right|\right]$$

1. $g(y)$ has to be a bijection
2. $g(y)$ and $g^{-1}(y)$ have to be differentiable
3. Jacobian determinant has to be tractably inverted



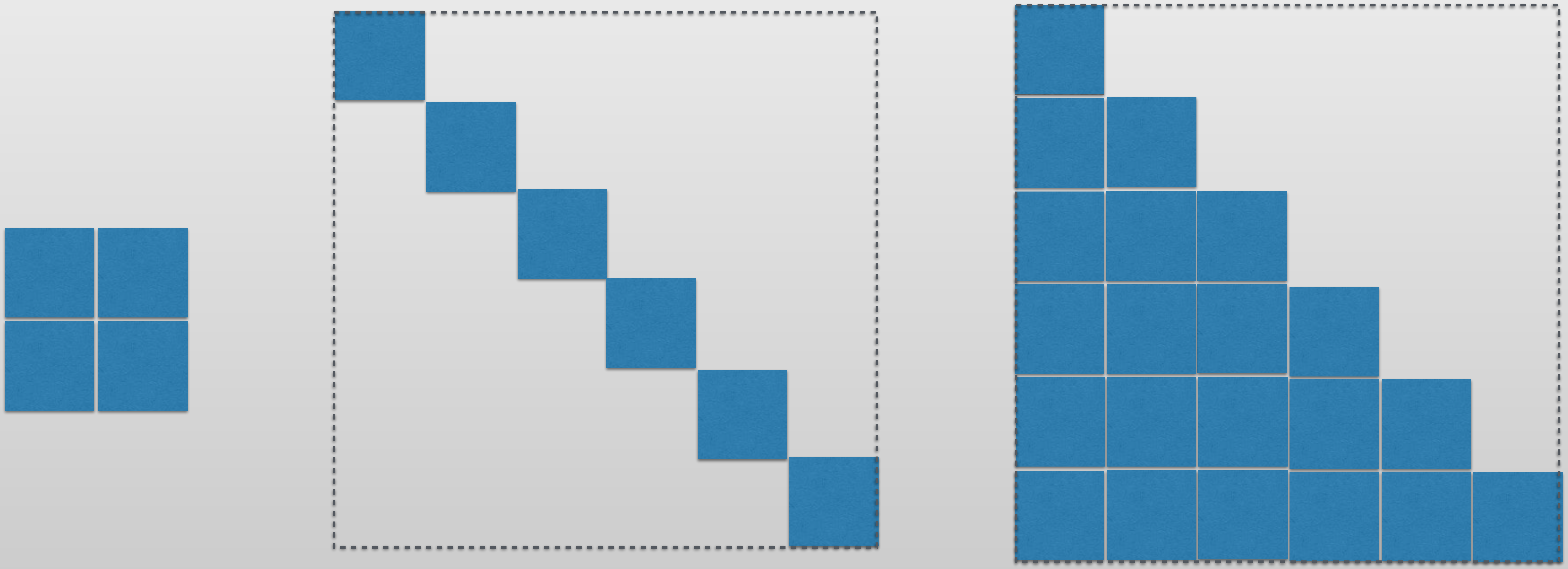
JACOBIAN

$$J_{g^{-1}}y = \begin{bmatrix} \frac{\partial g_1^{-1}}{\partial z_1} & \cdots & \frac{\partial g_1^{-1}}{\partial z_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_n^{-1}}{\partial z_1} & \cdots & \frac{\partial g_n^{-1}}{\partial z_n} \end{bmatrix}$$

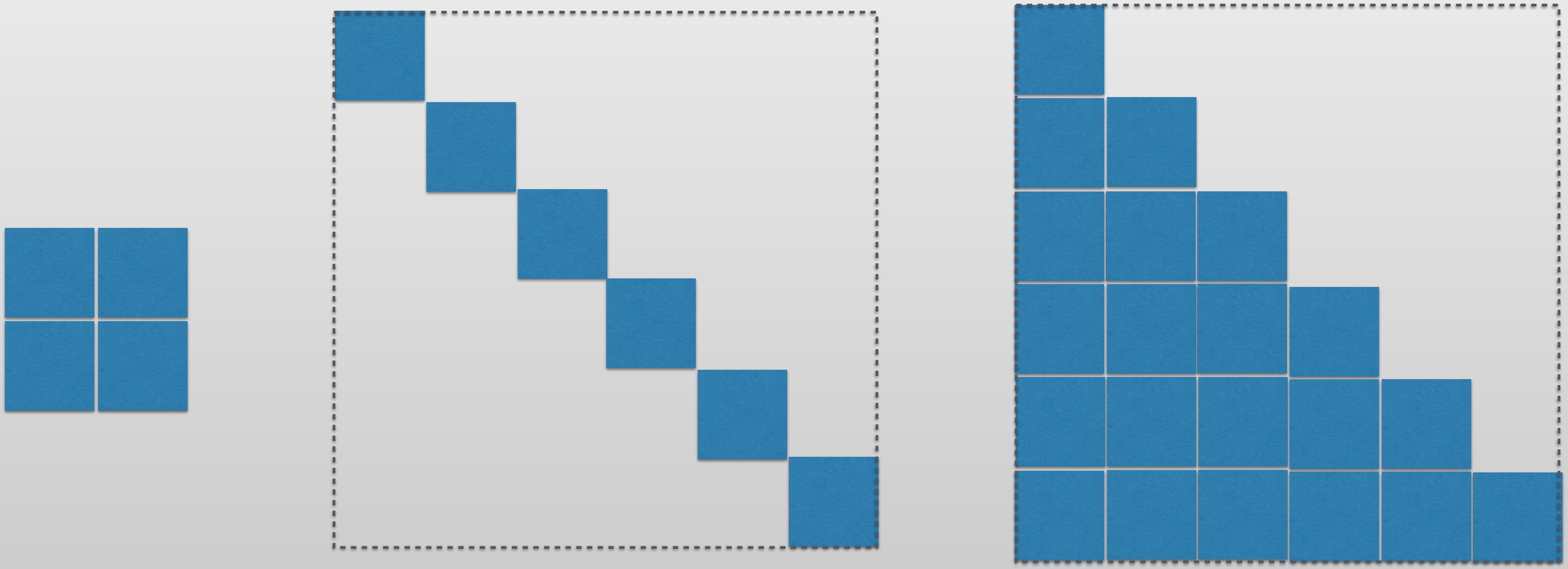
The calculation of determinant Jacobian will take $O(n^3)$

We have to find a way to make it faster

SIMPLIFYING JACOBIAN



SIMPLIFYING JACOBIAN



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

AFFINE TRANSFORMATIONS

location-scale transformation:

$$\tau(z_i; \mathbf{h}_i) = \alpha_i z_i + \beta_i \qquad \mathbf{h}_i = \{\alpha_i, \beta_i\}$$

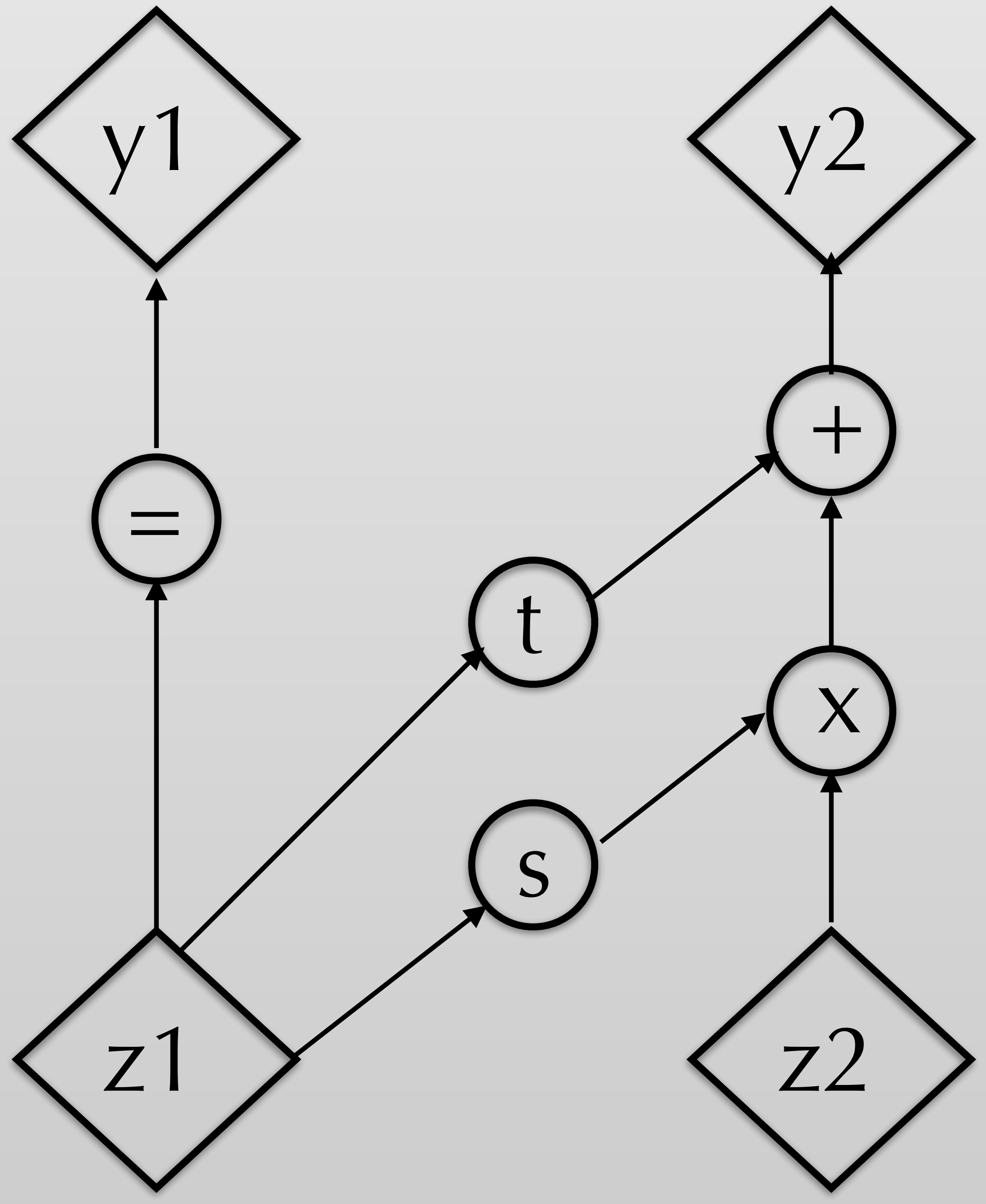
Invertibility for $\alpha_i \neq 0$

log-Jacobian becomes

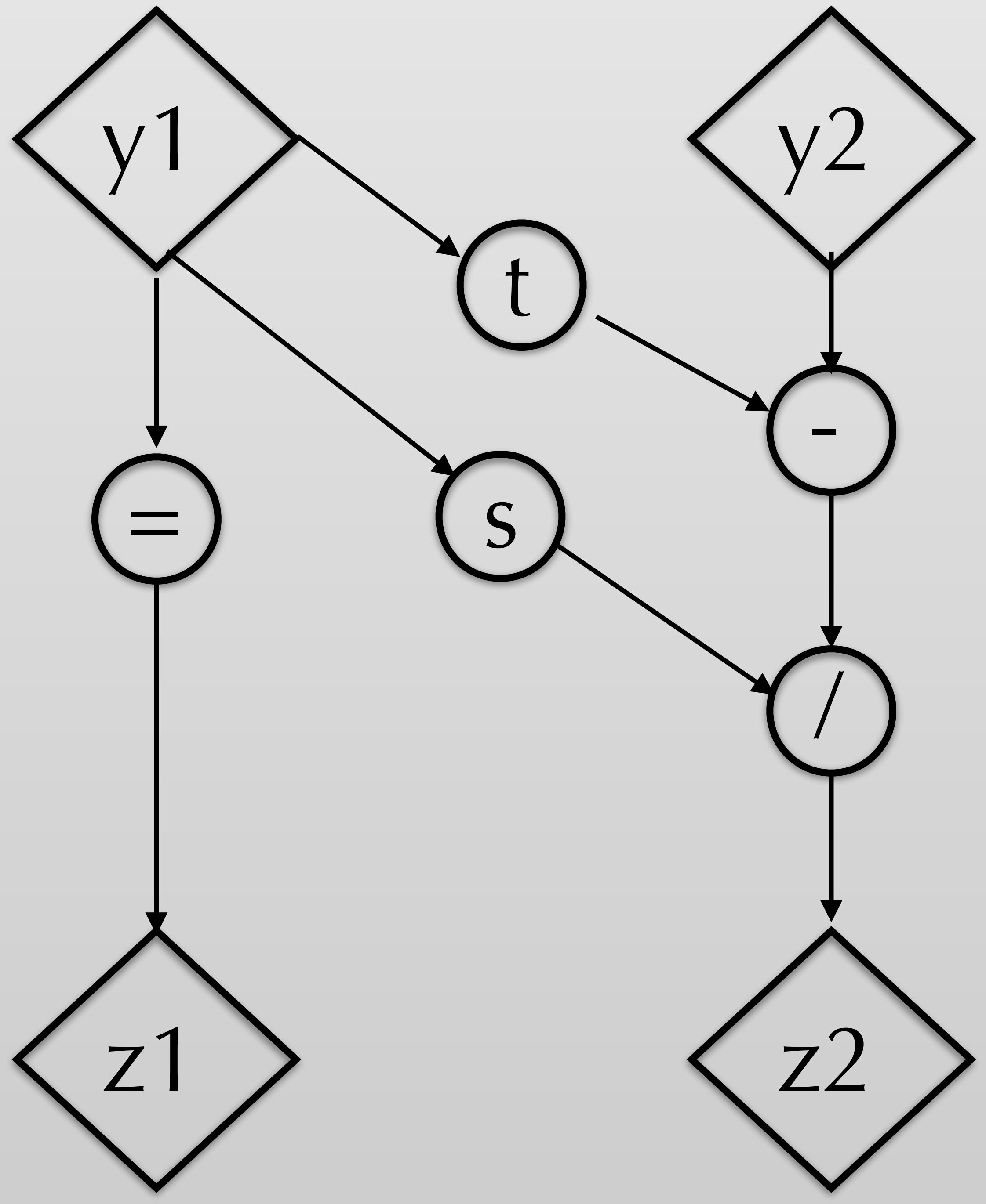
$$\log|\det J_{g^{-1}}(\mathbf{z})| = \sum_{i=1}^{N} \log|\alpha_i|$$

COUPLING TRANSFORM

Split input into two parts: z1 and z2



Forward propagation



Inverse propagation

REAL NVP

Coupling transform combined with affine transformation:

$$y_{1:d} = z_{1:d}$$
$$y_{d+1:D} = z_{d+1:D} \cdot \exp(s(z_{1:s})) + t(z_{1:d})$$

Jacobian of this transformation

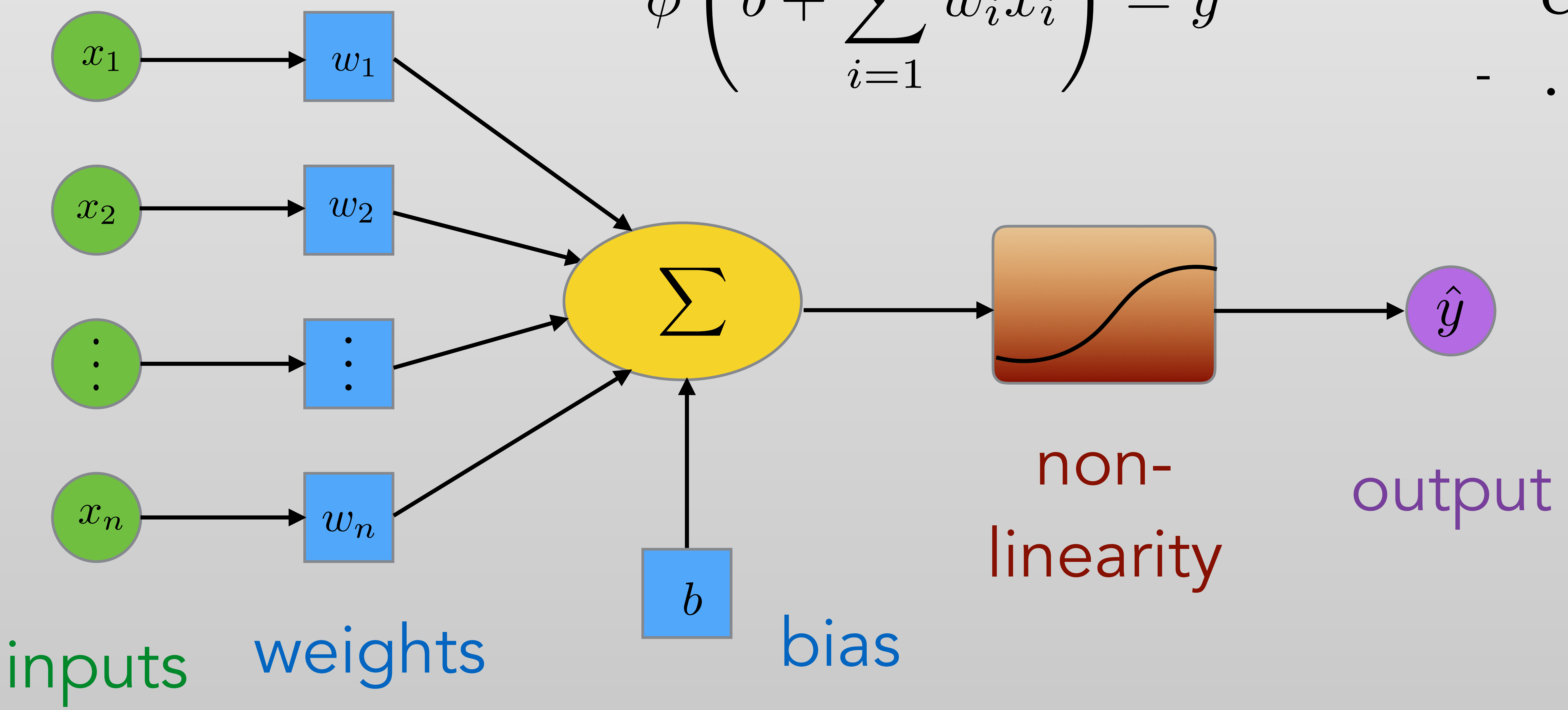
$$\frac{\partial y}{\partial z} = \begin{bmatrix} \mathbf{I}_d & 0 \\ \frac{\partial y_{d+1:D}}{\partial z_{1:d}} & \mathrm{diag}(\exp[s(z_{1:d})]) \end{bmatrix}$$

What is functions t and s?

PARAMETERISATION WITH THE NN

Neuron

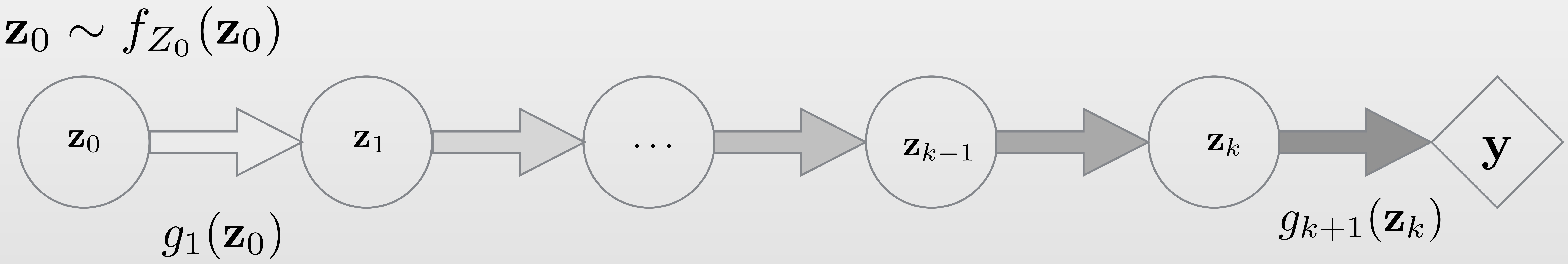
$$\phi\left(b + \sum_{i=1}^{n} w_i x_i\right) = \hat{y}$$



The architecture can be any:

- fully connected
- residual network
- convolutional network
- …

COMPOSING FLOWS



Function composition

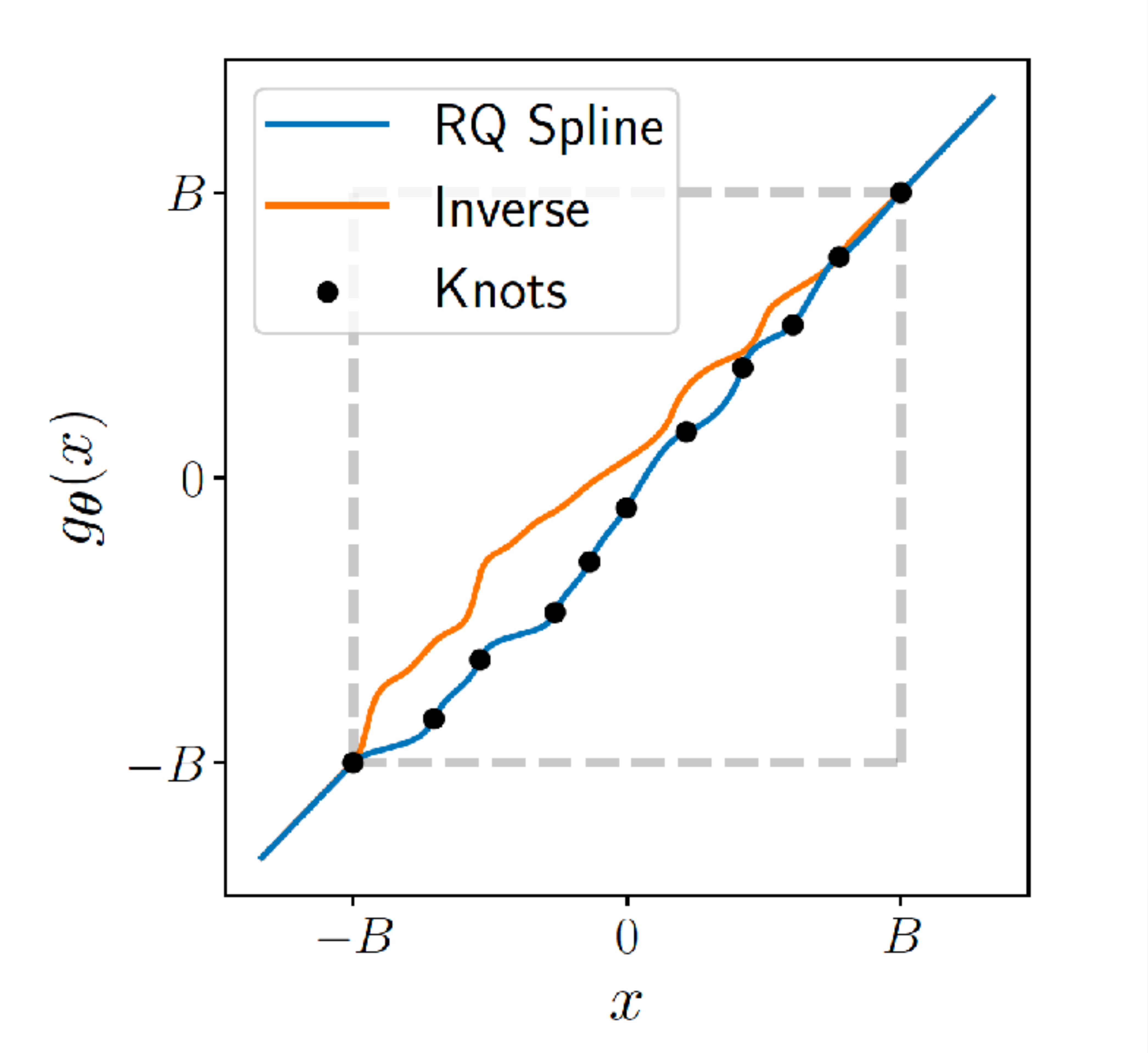
$$(g_1 \circ g_2)^{-1} = g_1^{-1} \circ g_2^{-1}$$

Jacobian composition

$$\det(J_1 \cdot J_2) = \det(J_1) \cdot \det(J_2)$$

SPLINE NEURAL FLOW

Replace affine transform
with tractable piecewise function.
For example,
Rational Quadratic Splines



Conor Durkan et al, Neural Spline Flows

OPTIMISATION

The flow is trained by maximising the total log likelihood of the data with respect to the parameters of the transformation:

$$\log p(\mathcal{D}|\theta) = \sum_{i=1}^{N} \log[f_Y(y_i|\theta)]$$

θ — parameters of the Neural Network with we use to parameterise our transform

OPTIMISATION

The flow is trained by maximising the total log likelihood of the data with respect to the parameters of the transformation:

$$\log p(\mathcal{D}|\theta) = \sum_{i=1}^{N} \log[f_Y(y_i|\theta)]$$

Use change of variable equation:

$$\log[f_Y(y)] = \log[f_Z(g^{-1}(y))] + \log\left[\left|\det \frac{\partial g^{-1}(y)}{\partial y}\right|\right]$$

CONDITIONING ON THE WAVEFORM

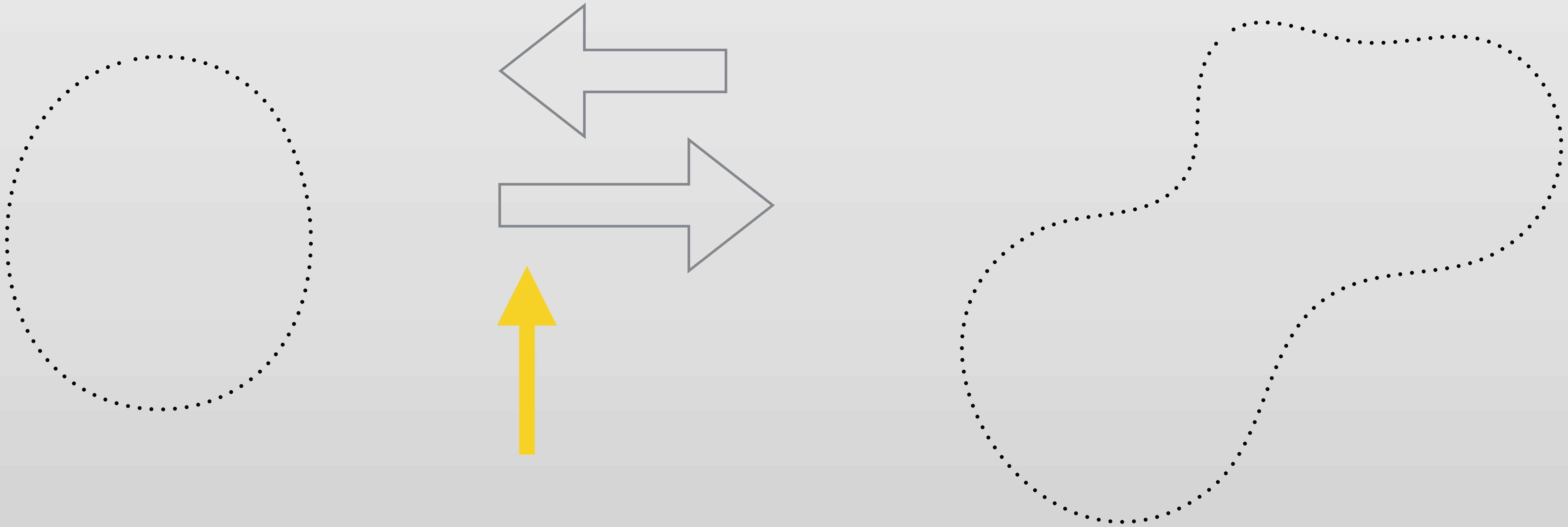
We do not have access to the samples form the posterior, as in the examples that we have just considered.

But we have access to the samples from the prior and the simulations of the data.

LIKELIHOOD FREE INFERENCE

Samples from a prior of a physical parameter

$$y \sim f_Y(y)$$



Condition map on the simulated data:

$$\mathbf{x} = h(\mathbf{y}) + \mathbf{n}$$

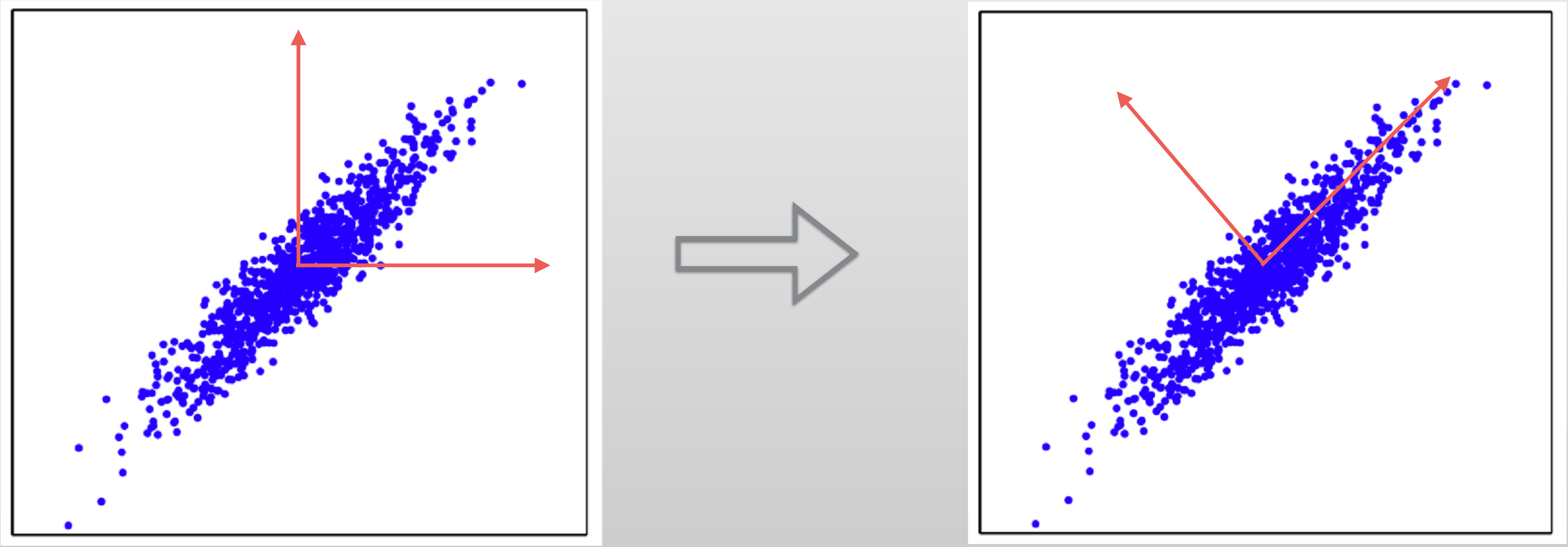
Therefore we have access to the joint sample: $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{x}|\mathbf{y})$

WAVEFORM EMBEDDING

- LISA observes signals in low frequency, therefor the waveforms are long.
- Conditioning does not work well with the long waveform, have to find a way to reduce in.
- It can be done, for example, by constructing new orthogonal basis

  which maximises variance in the space of the waveforms.
- And using the coefficients of the projection of the waveforms to the new basis.
- We implement it with Singular Value Decomposition.

WAVEFORM EMBEDDING



WAVEFORM EMBEDDING

Decompose a matrix constructed of the waveforms

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

matrix composed of reconstruction coefficients

matrix composed of basis vectors

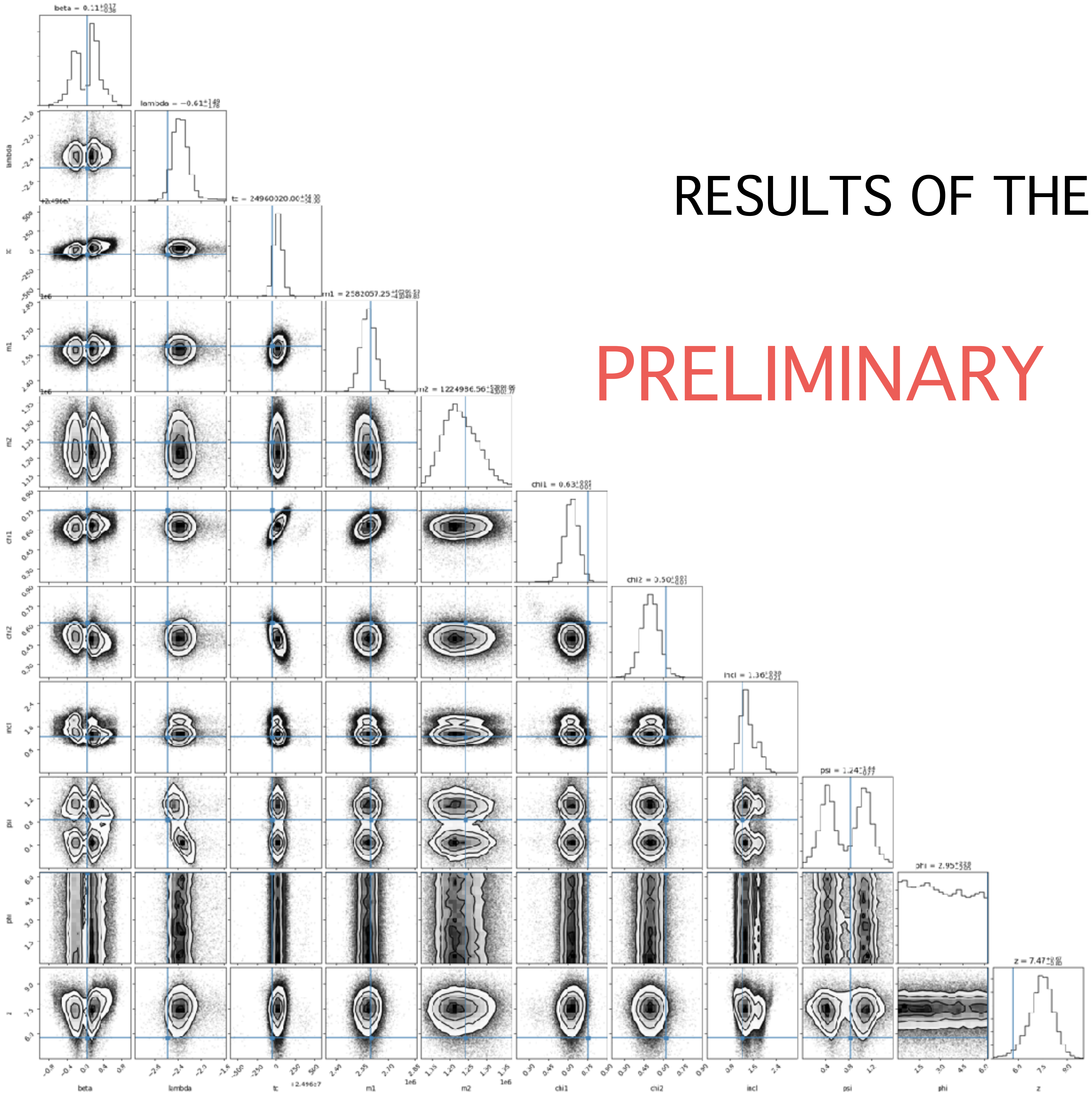
matrix containing the singular values

WAVEFORM EMBEDDING

Project the waveform onto the reduced basis in the following way:

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

RESULTS OF THE PARAMETER ESTIMATION



PRELIMINARY

Questions?