Challenges in LISA data analysis: Disentanglement and fast parameter estimation

Natalia Korsakova

01.09.2021

Laser

Interferometer

Space

Antenna

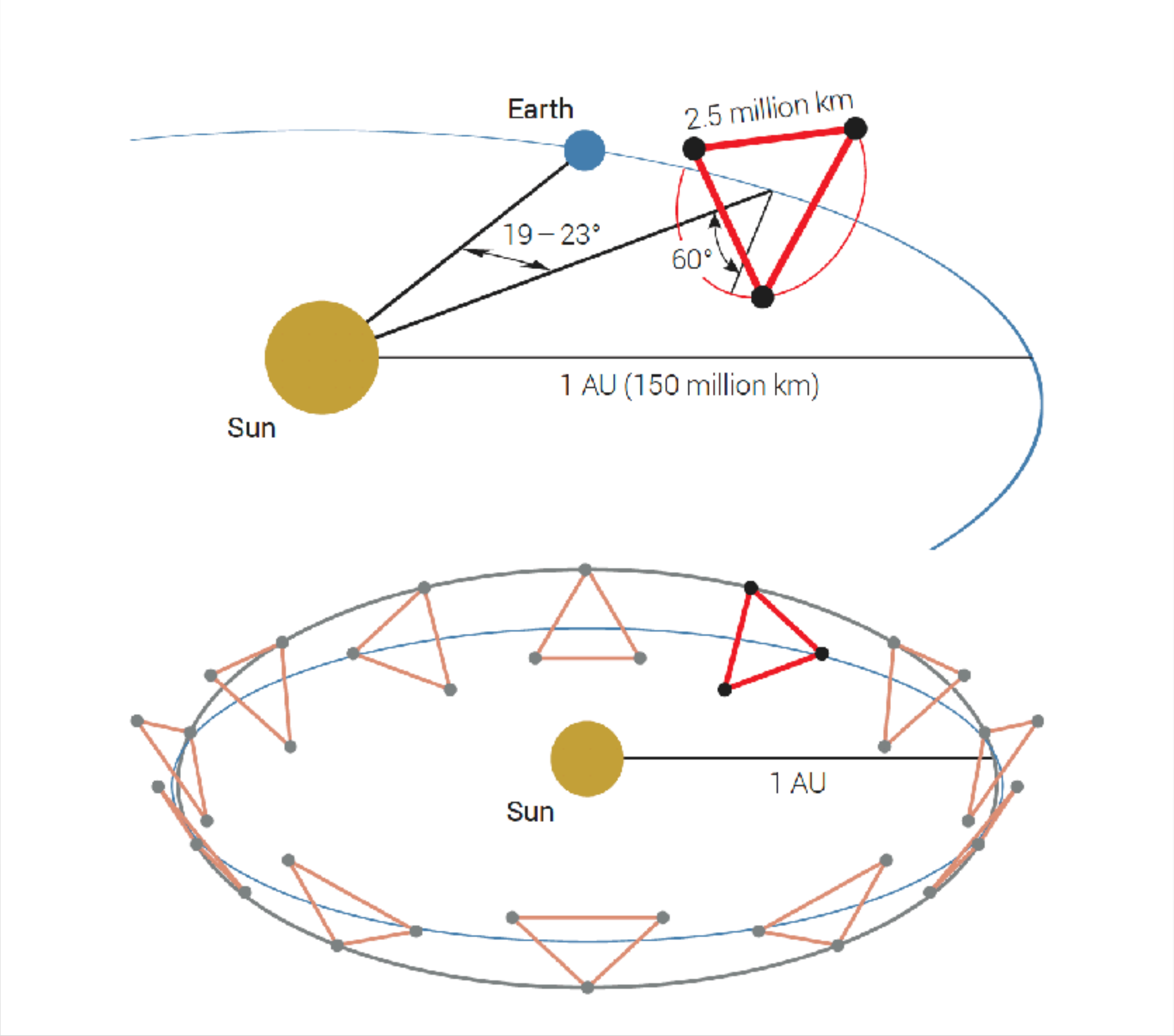


Image: LISA White paper

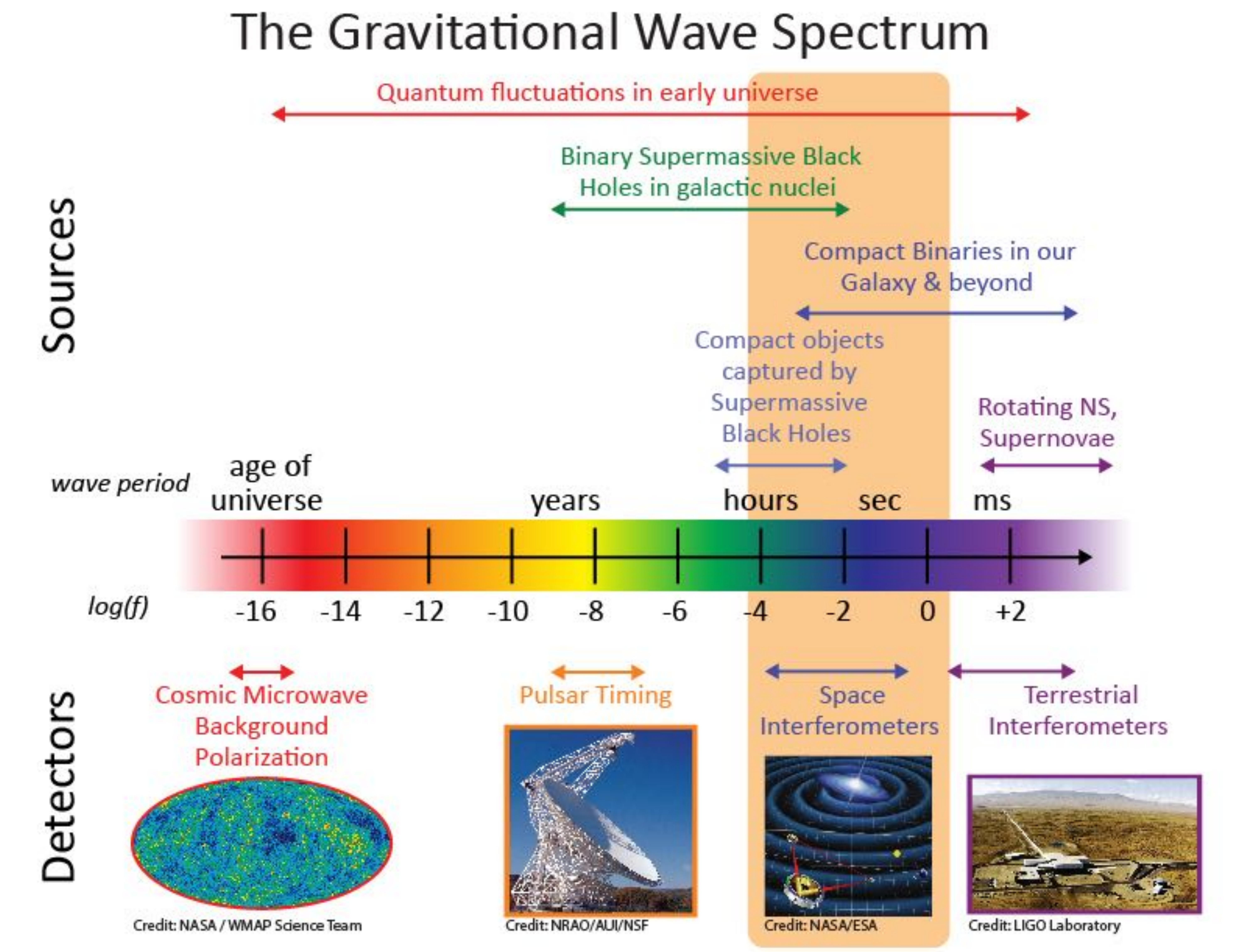
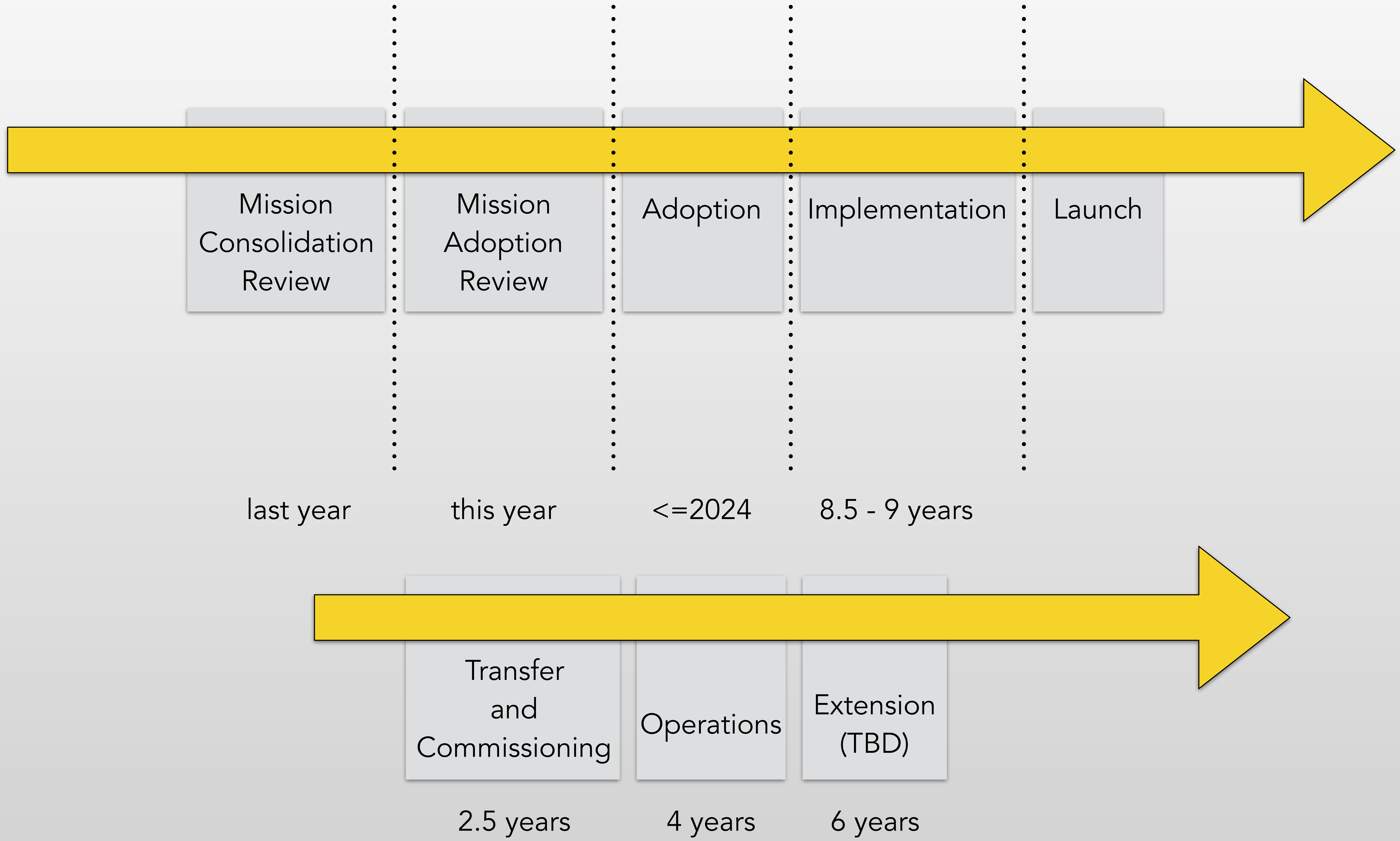

The Gravitational Wave Spectrum
Sources
Quantum fluctuations in early universe
Binary Supermassive Black Holes in galactic nuclei
Compact Binaries in our Galaxy & beyond
Compact objects captured by Supermassive Black Holes
Rotating NS, Supernovae
wave period
age of universe
years
hours
sec
ms
log(f)
-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
- …

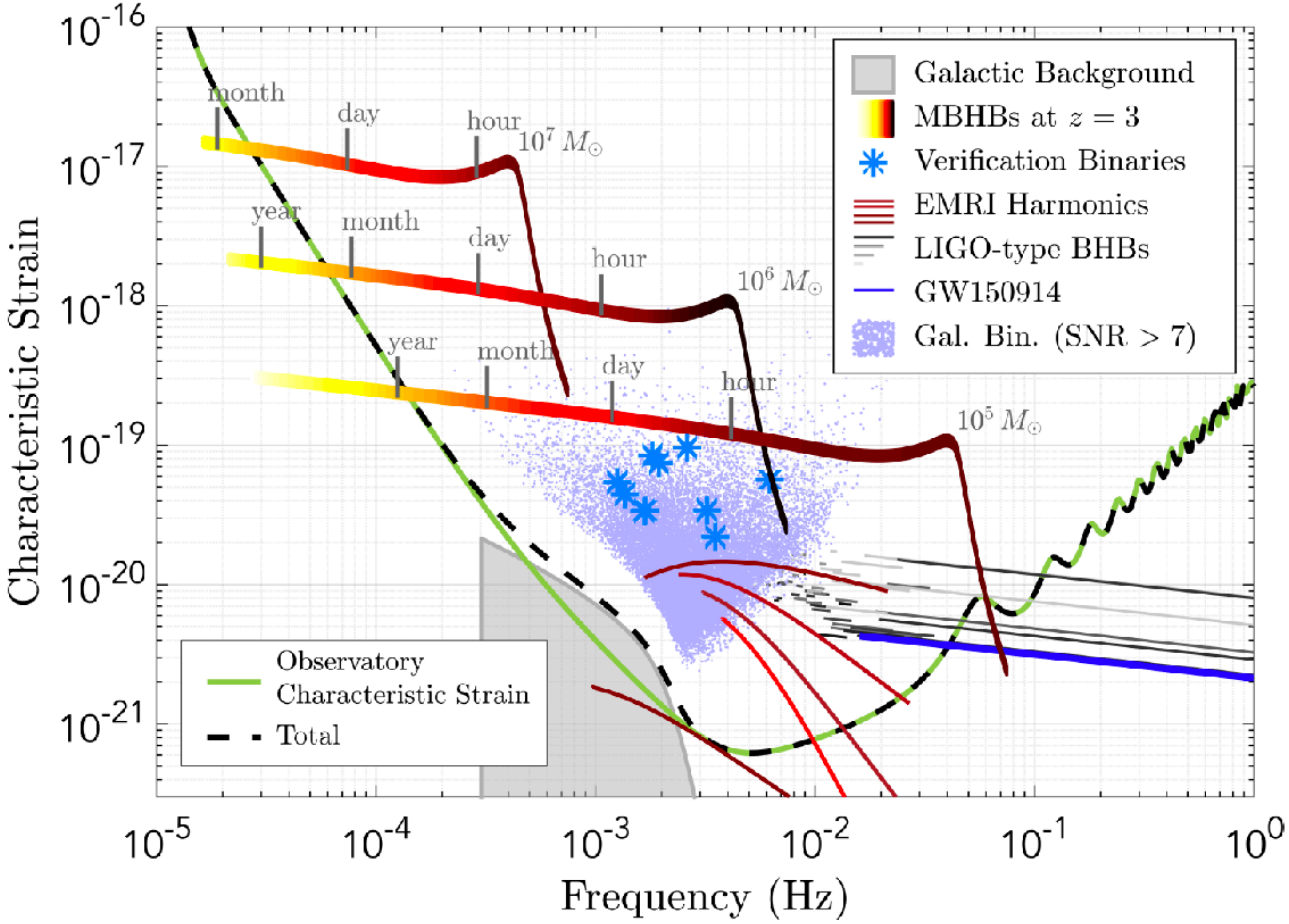


Image: LISA White paper

MASSIVE BLACK HOLE BINARIES

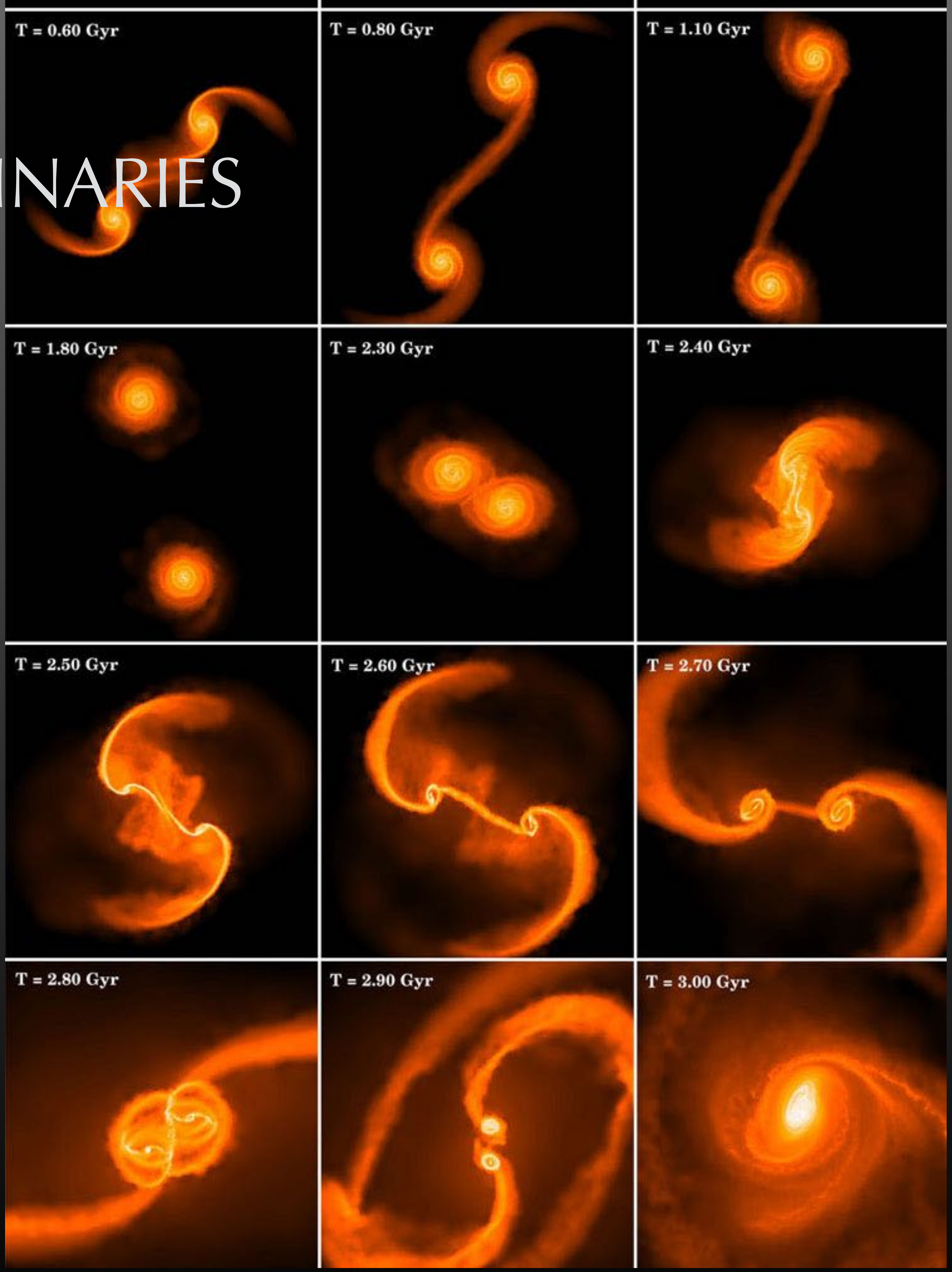


Image credit:
SLAC
National Accelerator Laboratory

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 ~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

MBHB SIGNAL

Example of a time series of the MBHB around coalescence

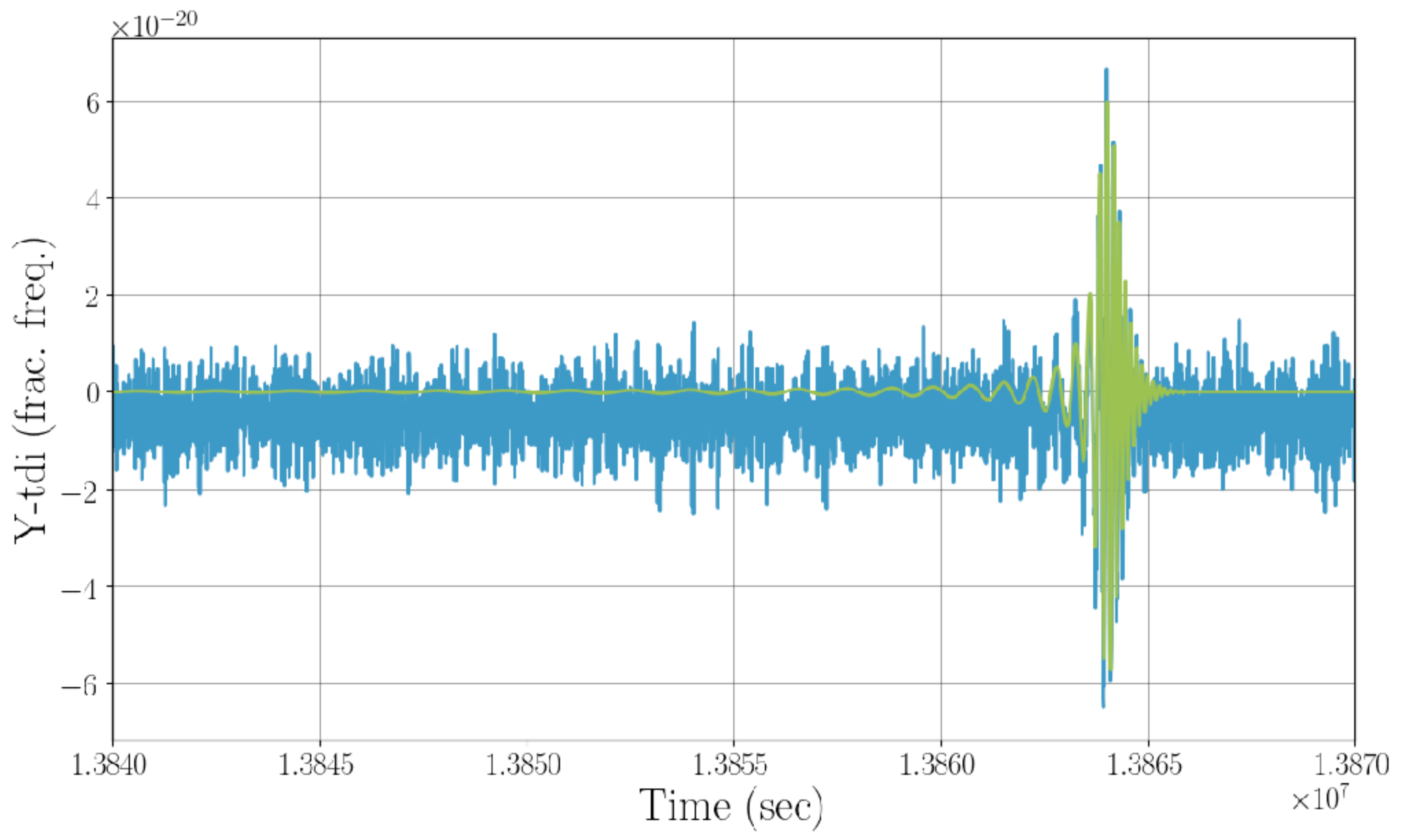


Image: LDC

GALACTIC BINARIES

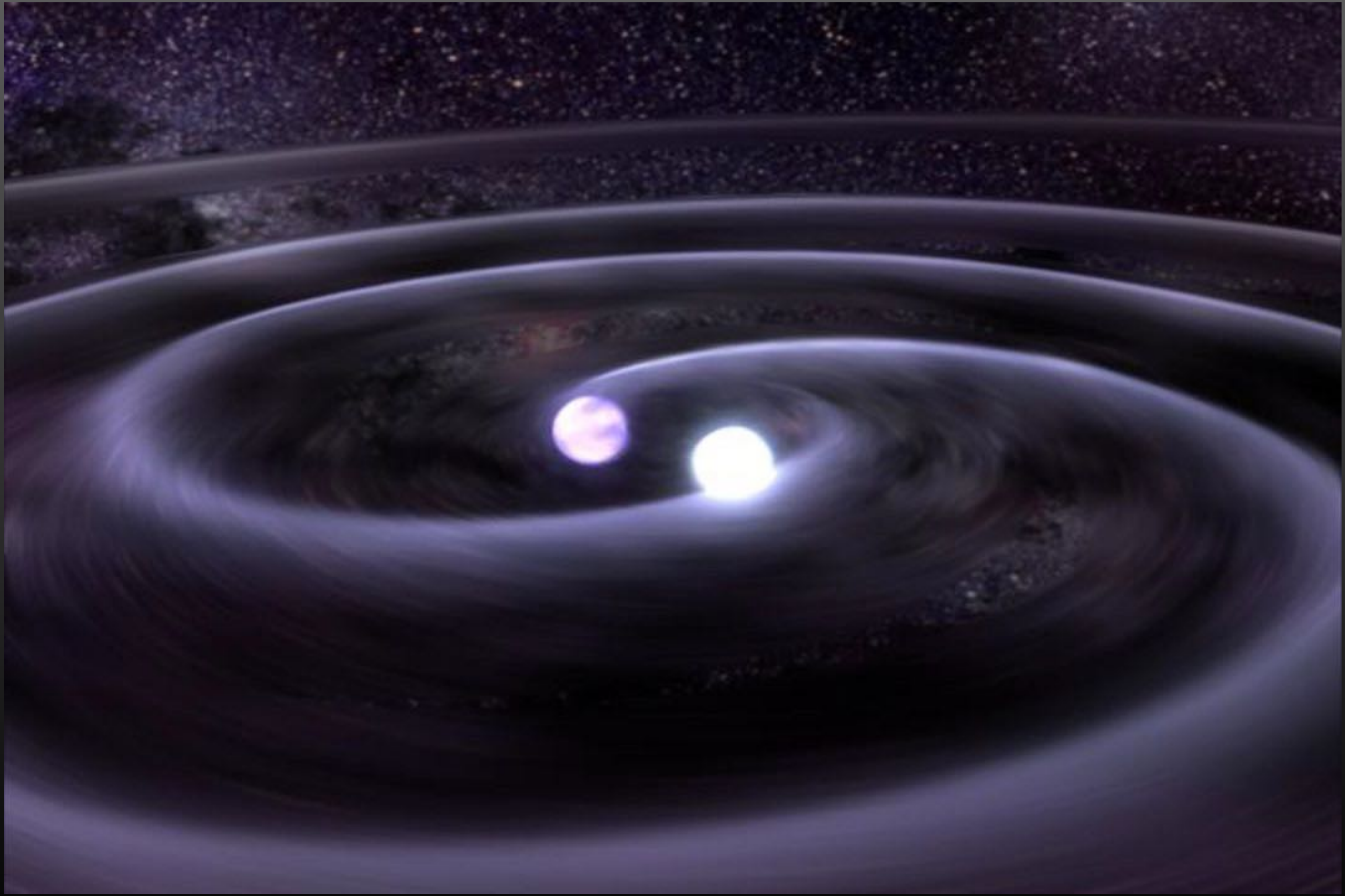


Image credit: NASA

GALACTIC BINARIES

- *Resolvable (~25000) and Confusion background (~10^7 all together)*
- *Verification binaires*

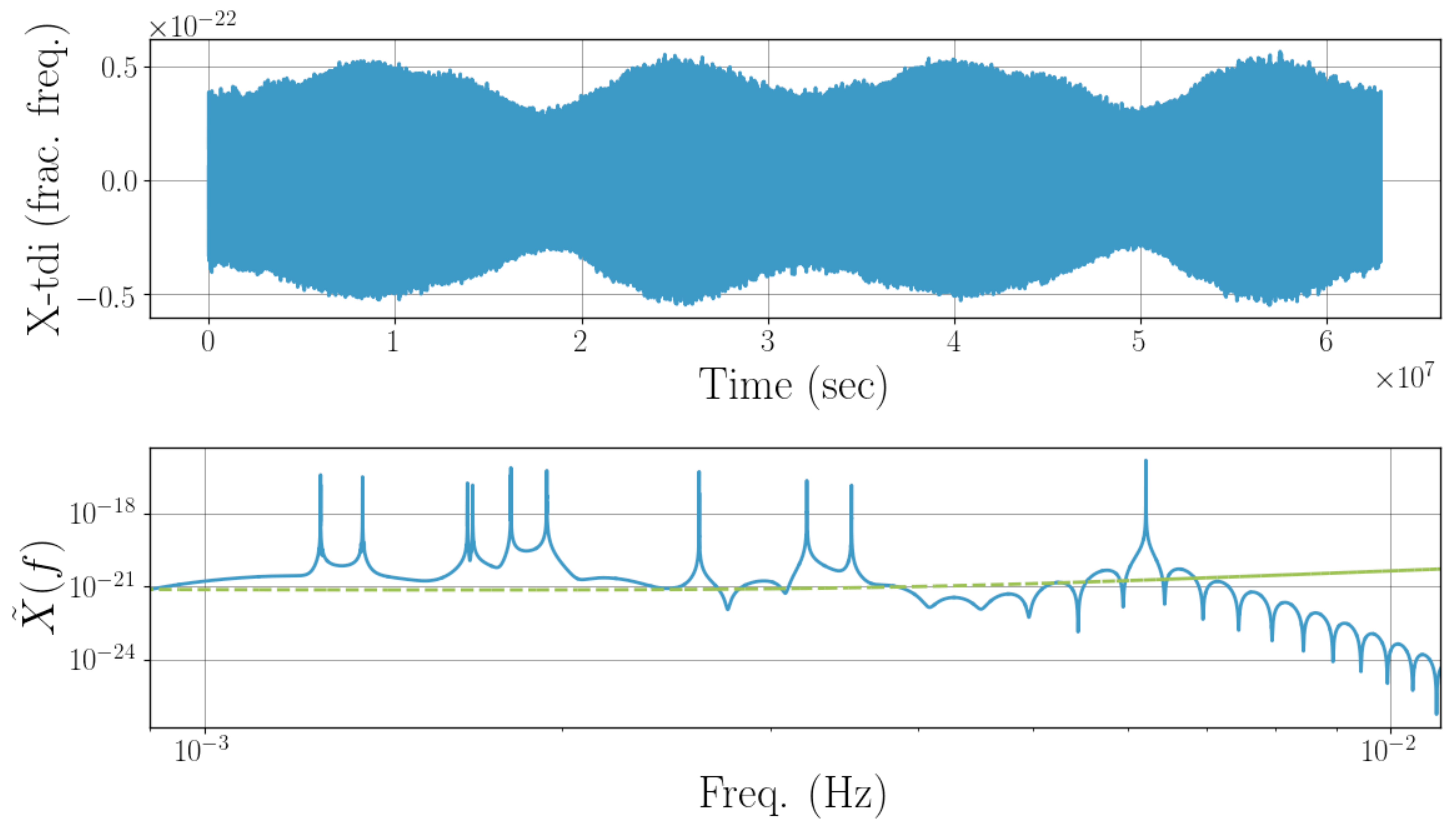


Image: LDC

EXTREME MASS RATIO INSPIRALS

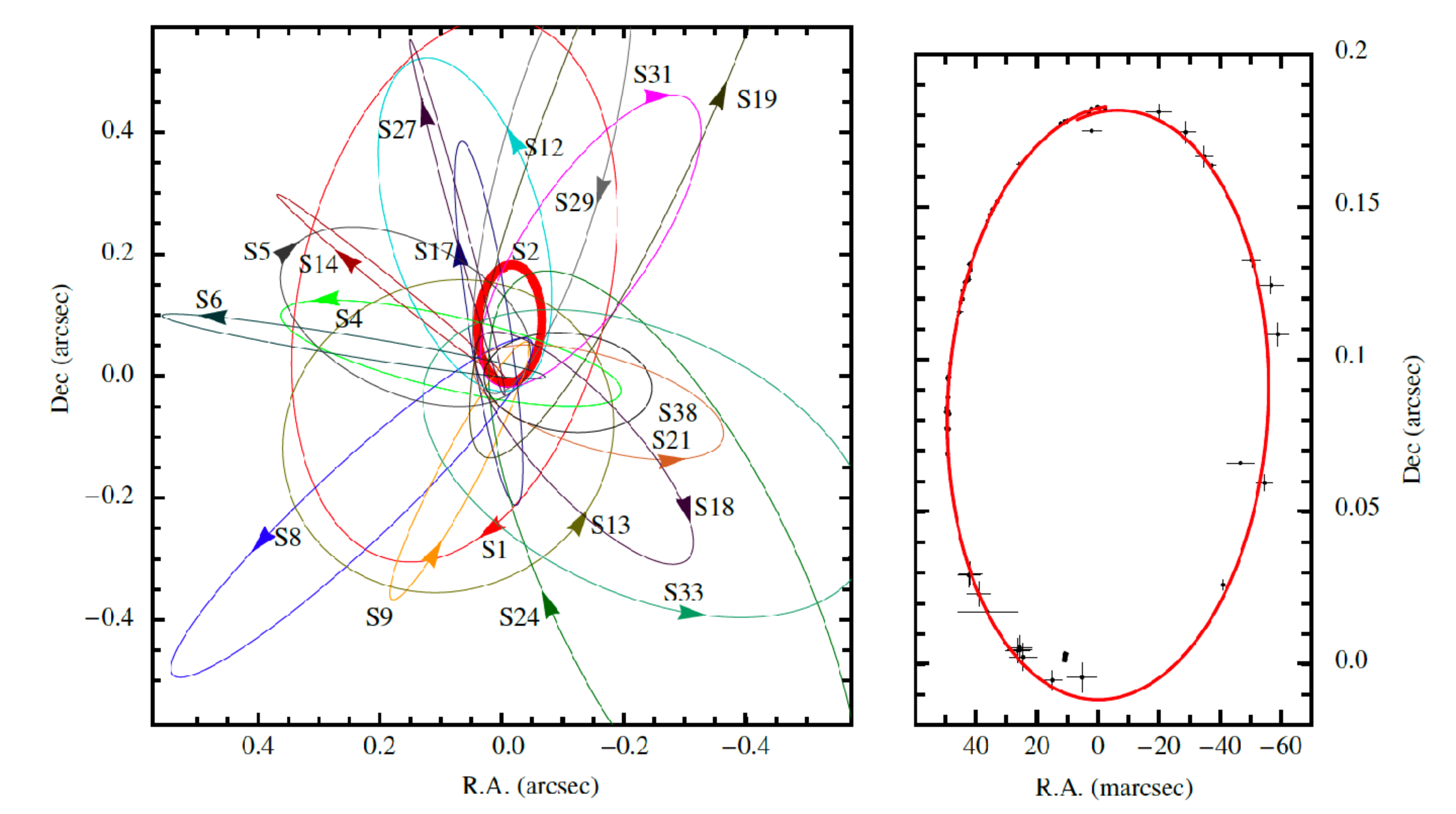


Image: Gillessen et al. (2009)

EXTREME MASS RATIO INSPIRALS

- Extreme mass ratio inspiral is produced by the compact object captured by MBH. The object gradually falls for the 10^4 - 10^6 cycles in the strong gravity.
- The waveforms are determined by three characteristic frequencies: the orbital frequency, the perihelion precession frequency and the frequency of precession of the orbital plane
- Usually have significant eccentricity

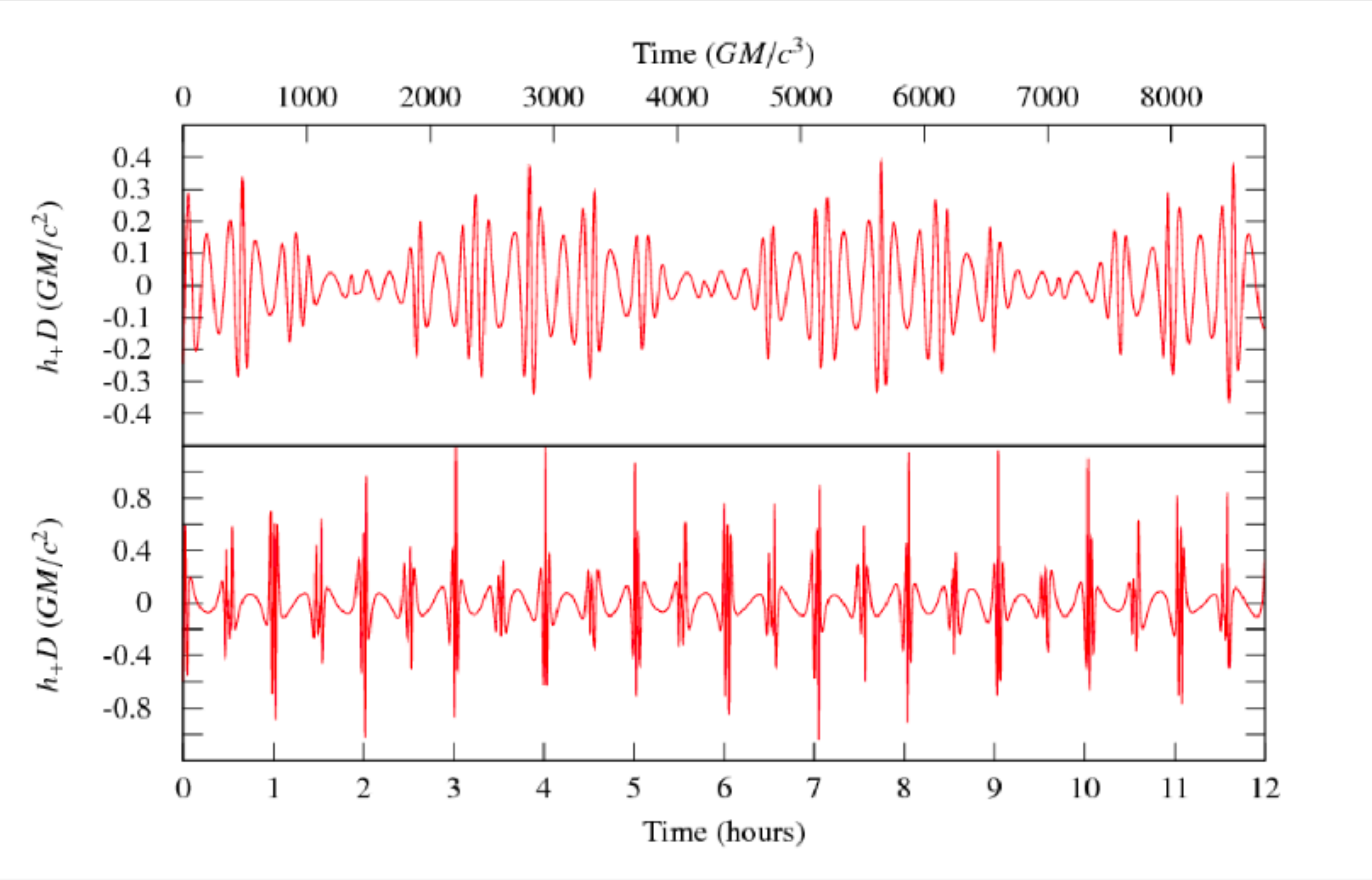


Image: Drasco and Hughes (2006)

EXTREME MASS RATIO INSPIRALS

Astrophysical population model:

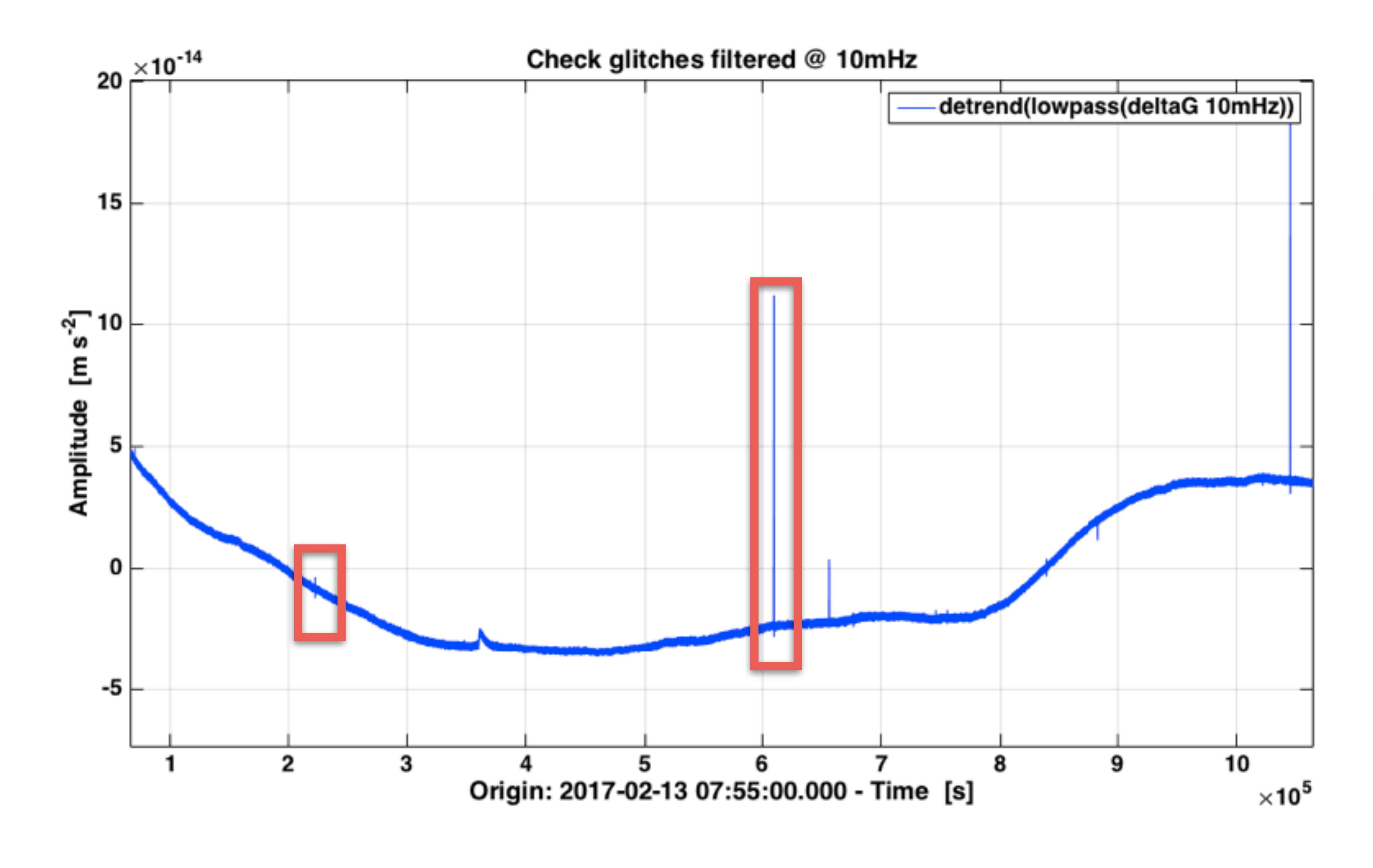
- mass distribution

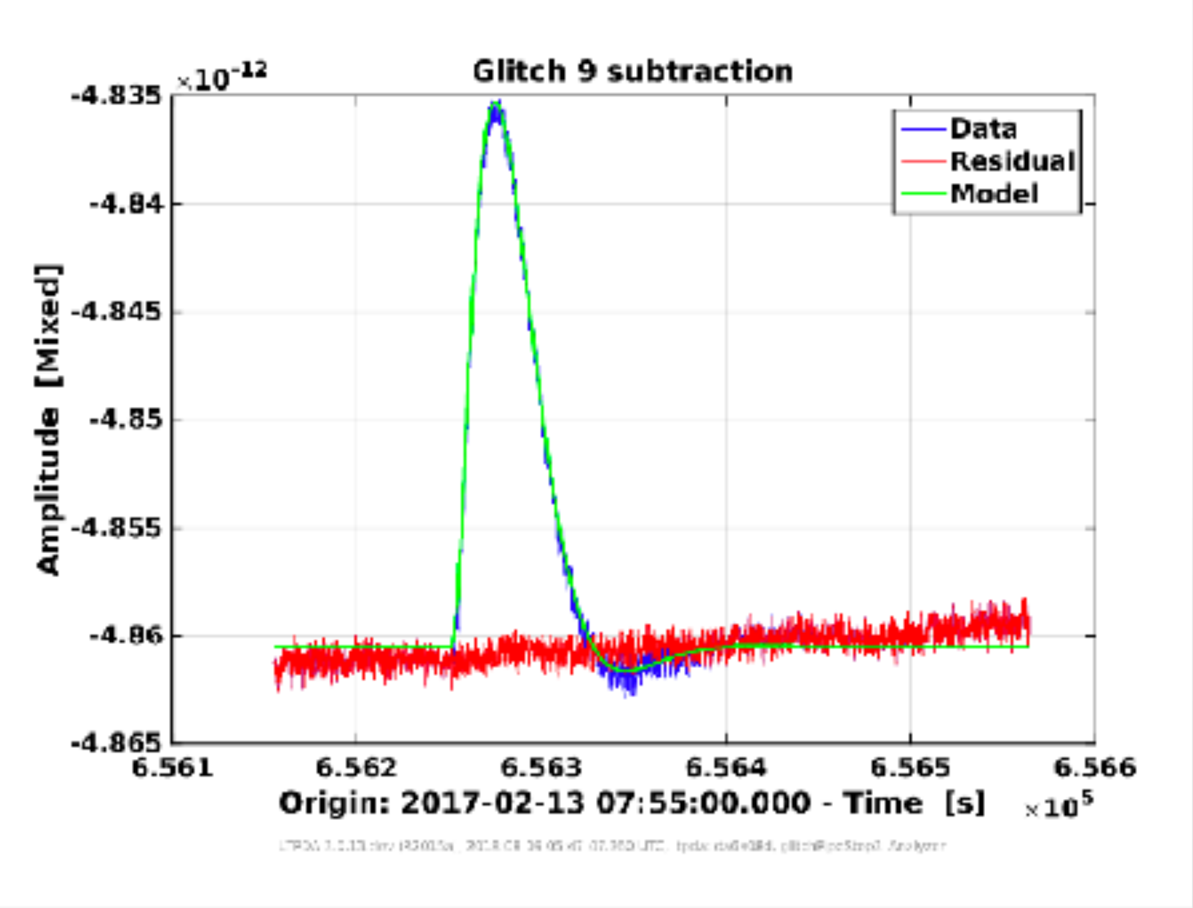
$$\frac{\mathrm{d}n}{\mathrm{d}\log M} = A\left(\frac{M}{3\times 10^6 M_\odot}\right)^{\alpha} \mathrm{Mpc}^{-3}$$

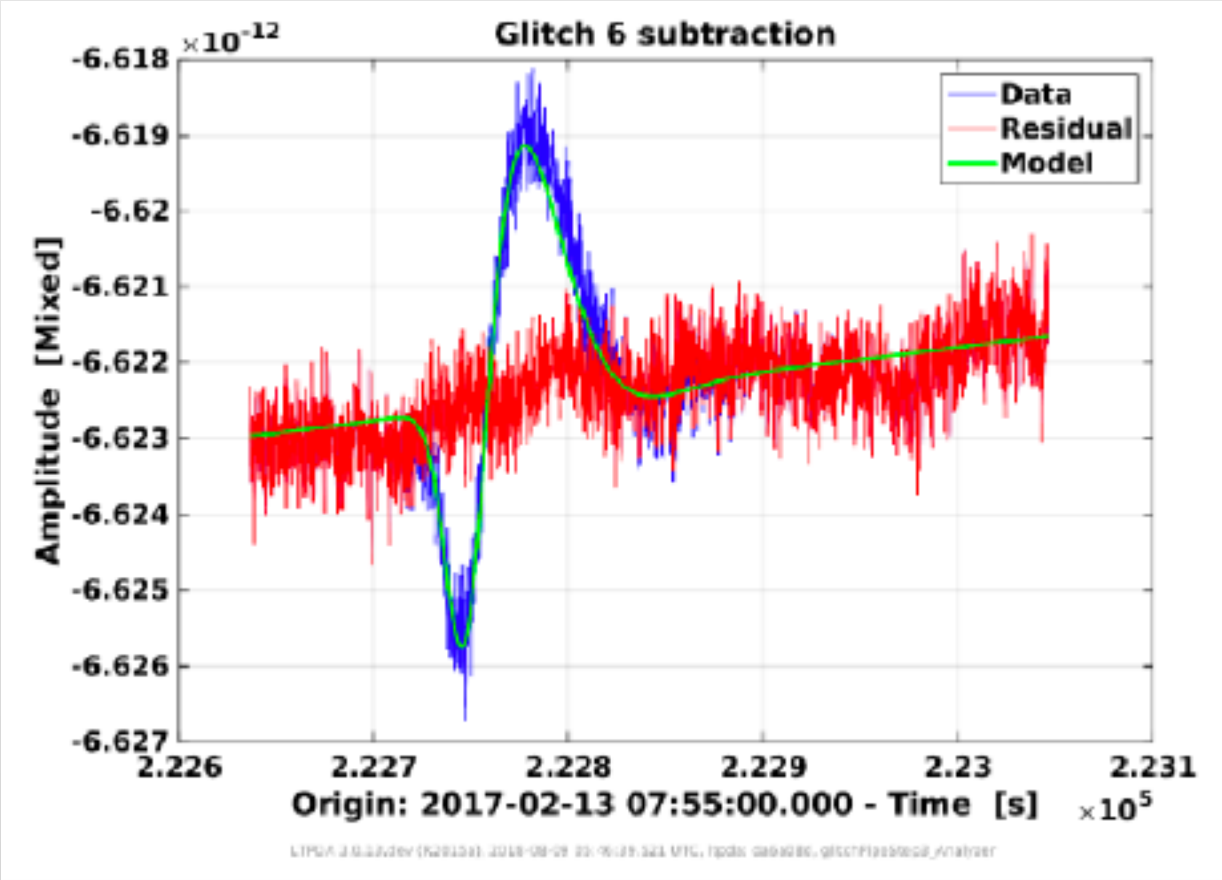
- spin distribution
  typically close to the maximum limit of 0.98
- EMRI rate per MBH
- M-sigma relation
- Properties of the compact object — typically black hole

From 1 to 10000 per year

NOISE ARTEFACTS



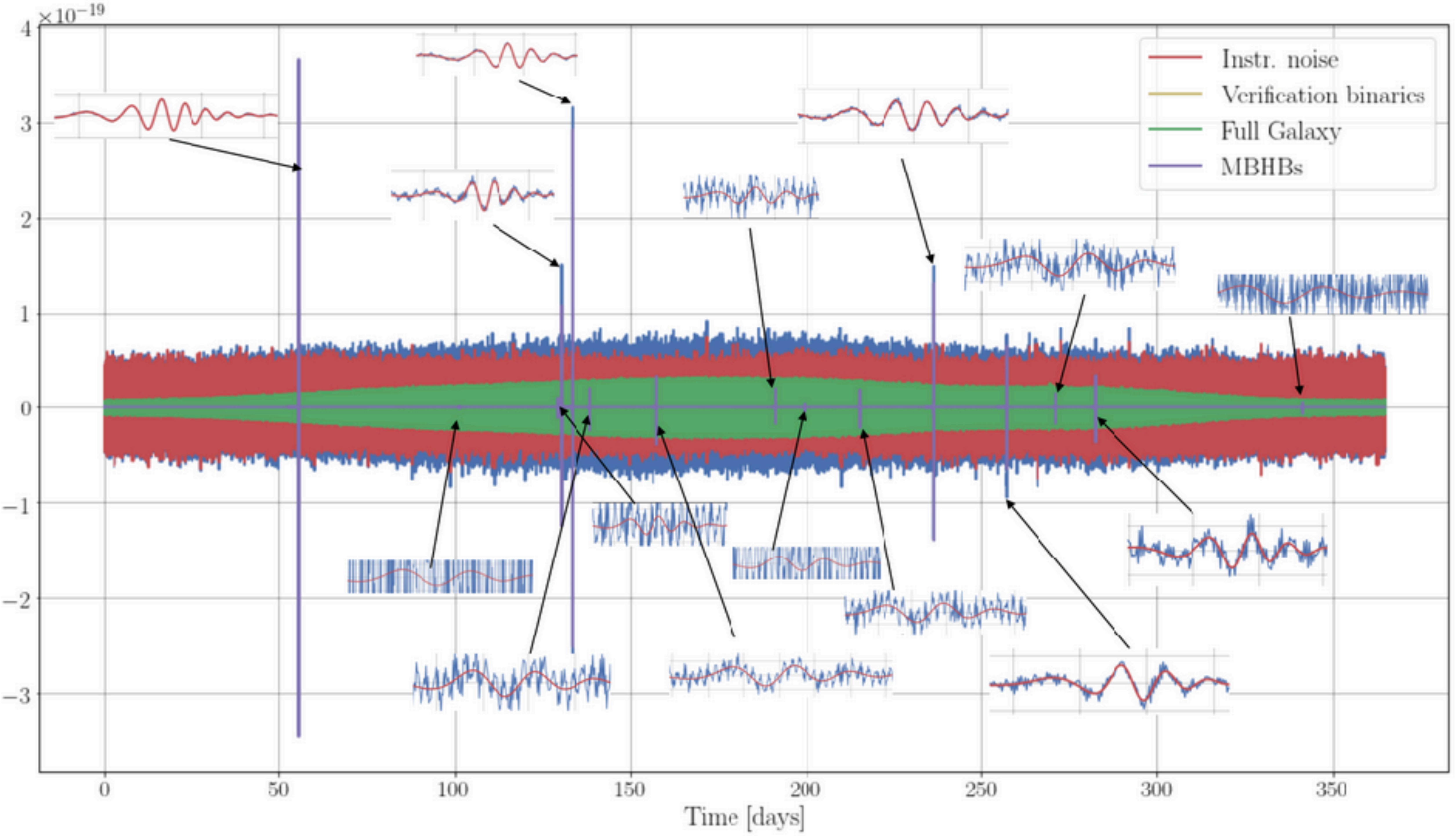




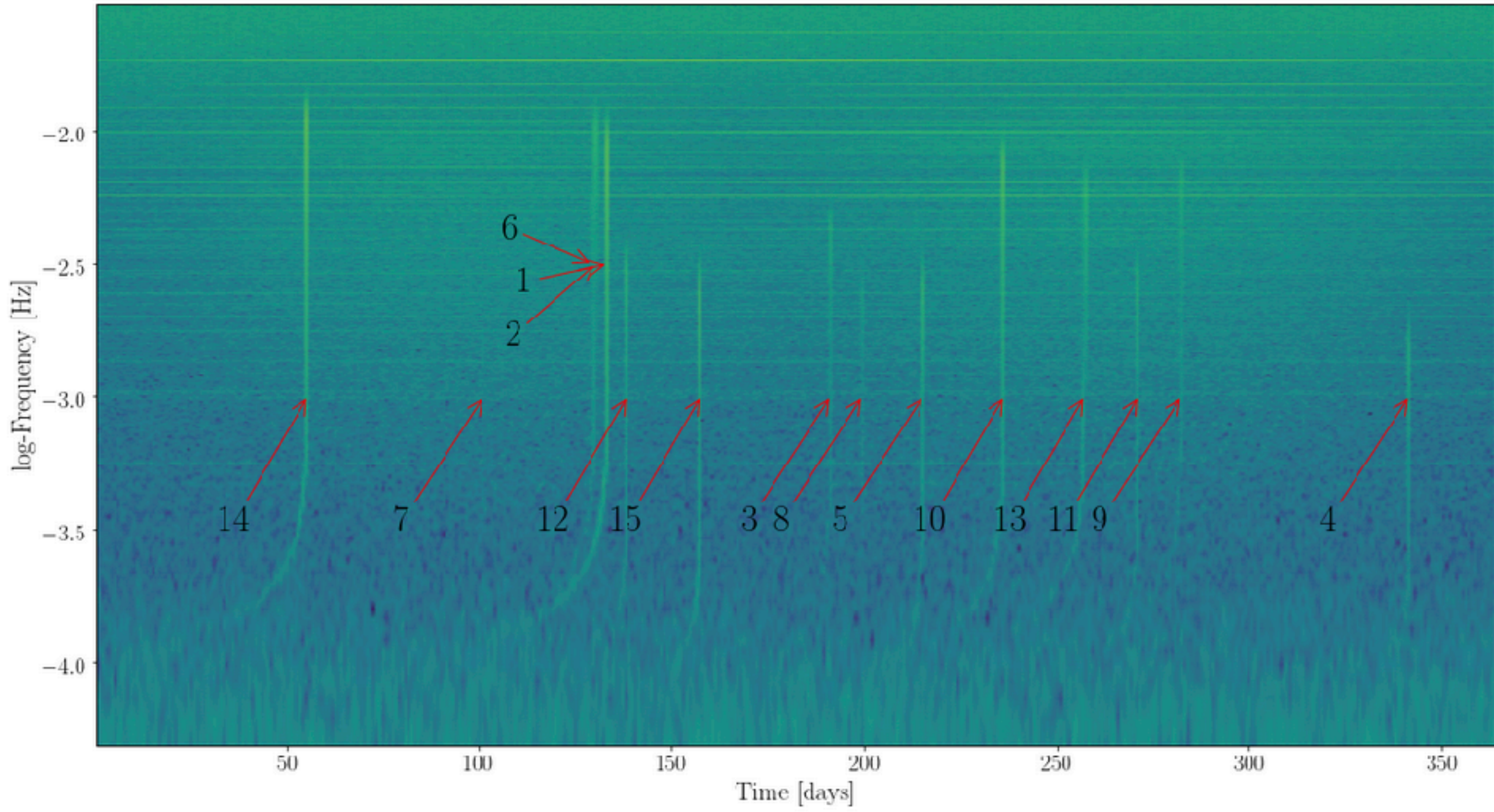
Can be modelled as Poisson distribution with

$$\lambda \approx 1^{d}10^{h}30^{m}$$

LDC WITH MIXED SIGNALS: SANGRIA

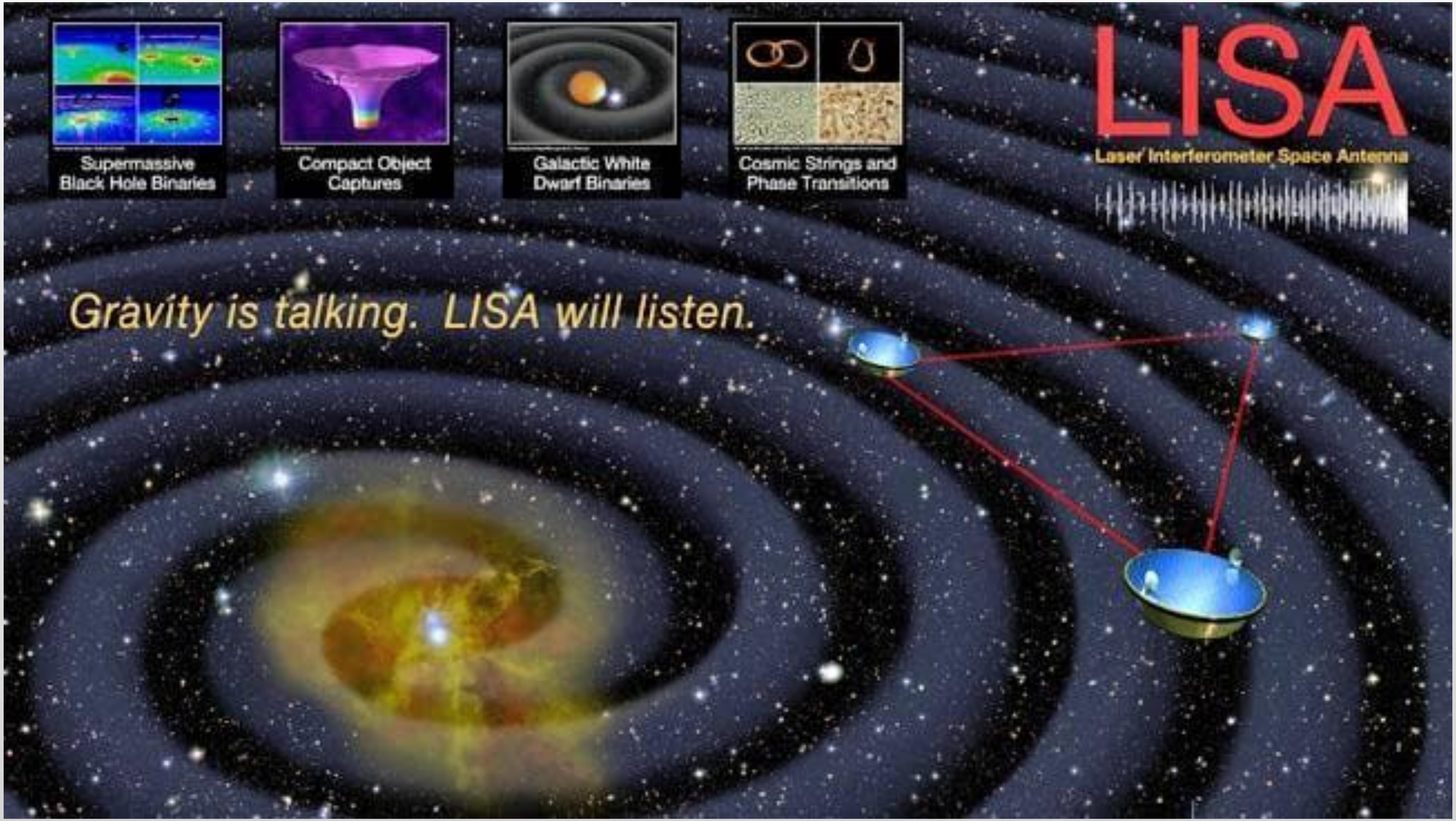


Sangria in TD



Sangria in T-FD

SOURCE SEPARATION PROBLEM



SIGNAL MIXTURE PROBLEMS ANALOGY

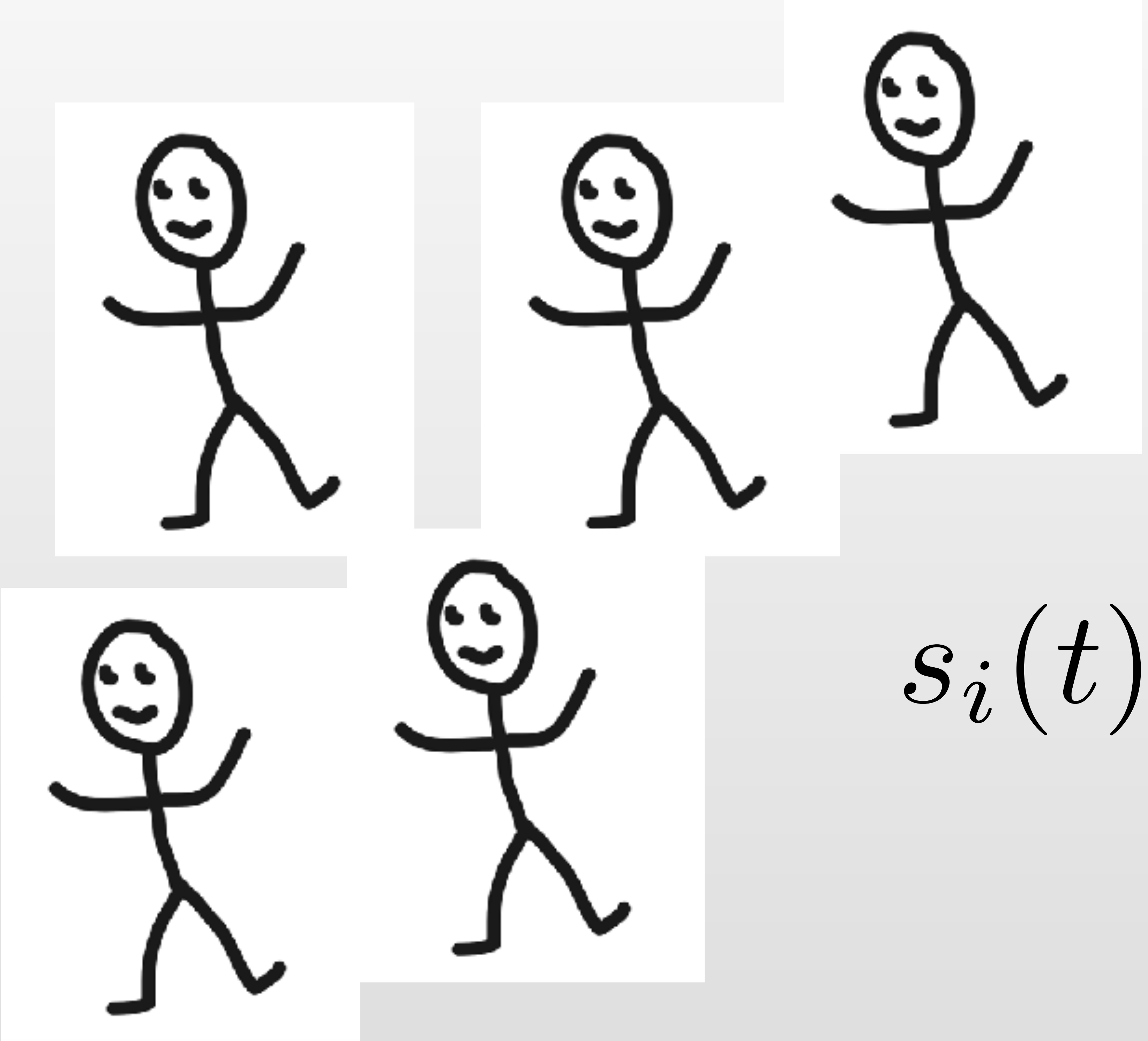
Gravitational wave signal measured by detector is

$$x(t) = \mathbf{D}(\hat{n}, f) : \mathbf{h}(f, \xi)$$

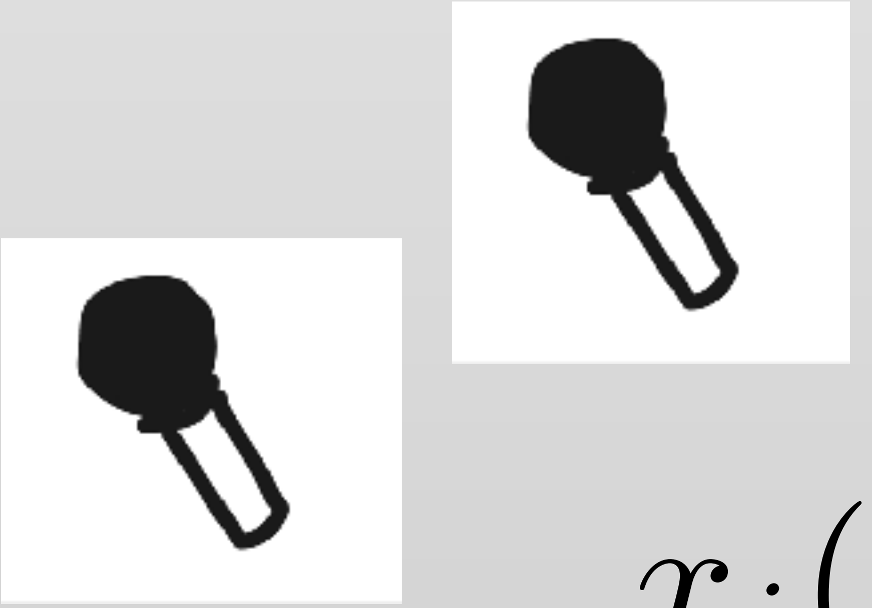
*where **h** is gravitational wave strain that is produced by each astrophysical object,*

***D** is the response of the detector.*

COCKTAIL PARTY PROBLEM



$s_i(t)$



$x_i(t)$

BLIND SOURCE SEPARATION

$$\begin{cases} x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) \\ x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t) \end{cases}$$

observe

want to recover

mixing coefficients

In general this is ill posed problem because we deal with underdetermined source separation problem.

We can generalise it as $\mathbf{x} = \mathbf{A}\mathbf{s}$

MISIC SEPARATION

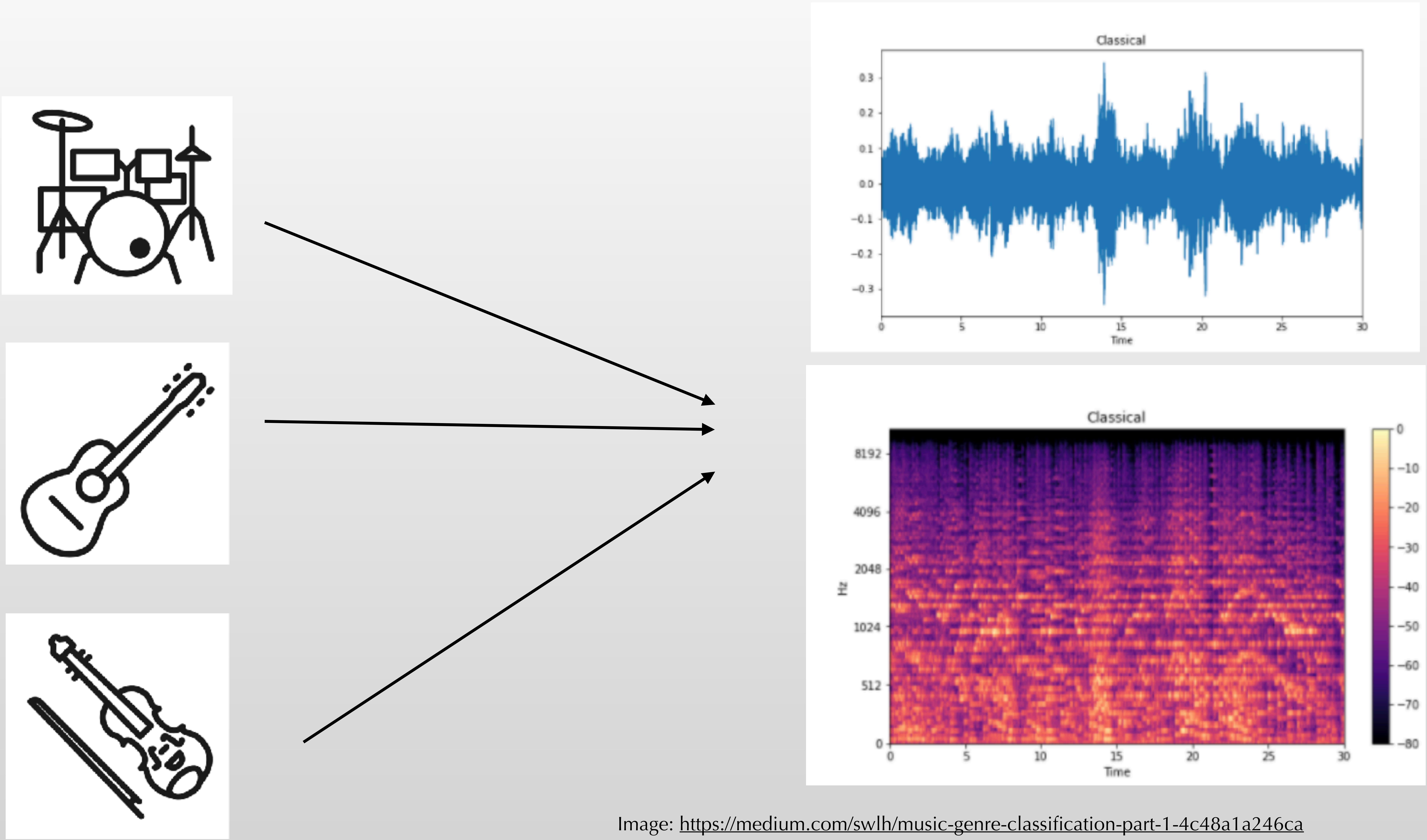


Image: https://medium.com/swlh/music-genre-classification-part-1-4c48a1a246ca

BLIND SOURCE SEPARATION

Traditional way to solve this problem was to find the independent components in the data.

COVARIENCE

Central moments — *covariance* matrix

$$\mathbf{C_x} = \mathrm{E}\{(\mathbf{x} - \mathbf{m_x})(\mathbf{x} - \mathbf{m_x})^T\}$$

Cross-covariance matrix

$$\mathbf{C_{xy}} = \mathrm{E}\{(\mathbf{x} - \mathbf{m_x})(\mathbf{y} - \mathbf{m_y})^T\}$$

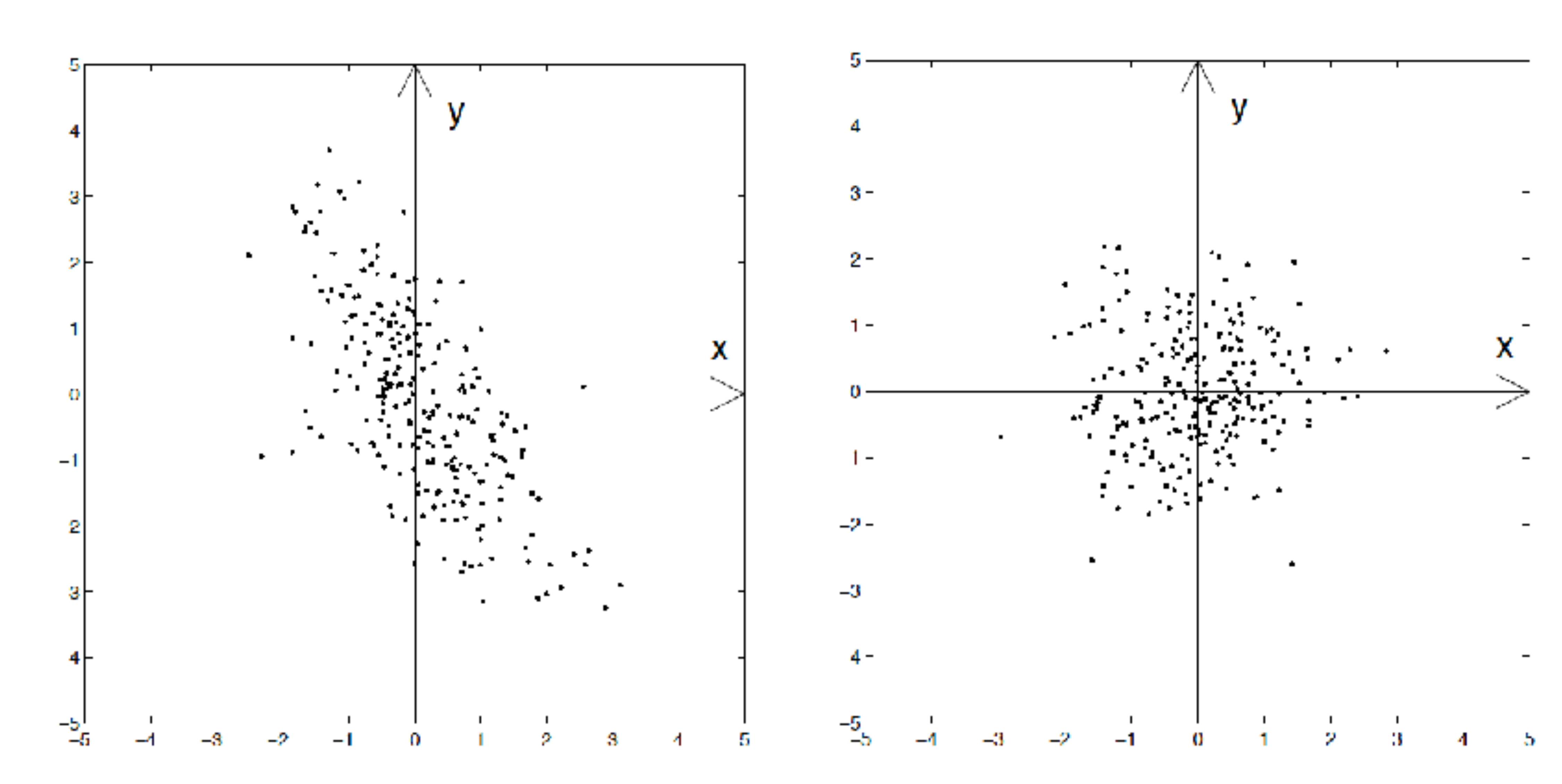
Two vectors are uncorrelated if

$$\mathbf{C_{xy}} = \mathrm{E}\{(\mathbf{x} - \mathbf{m_x})(\mathbf{y} - \mathbf{m_y})^T\} = \mathbf{0}$$

For one vector, similar condition, different components mutually uncorrelated:

$$\mathbf{C_x} = \mathrm{E}\{(\mathbf{x} - \mathbf{m_x})(\mathbf{x} - \mathbf{m_x})^T\} = \mathbf{D}$$

where $\mathbf{D} = \mathrm{diag}(\sigma_{x_1}^2, \sigma_{x_2}^2, \ldots, \sigma_{x_n}^2)$



Negative covariance Zero covariance

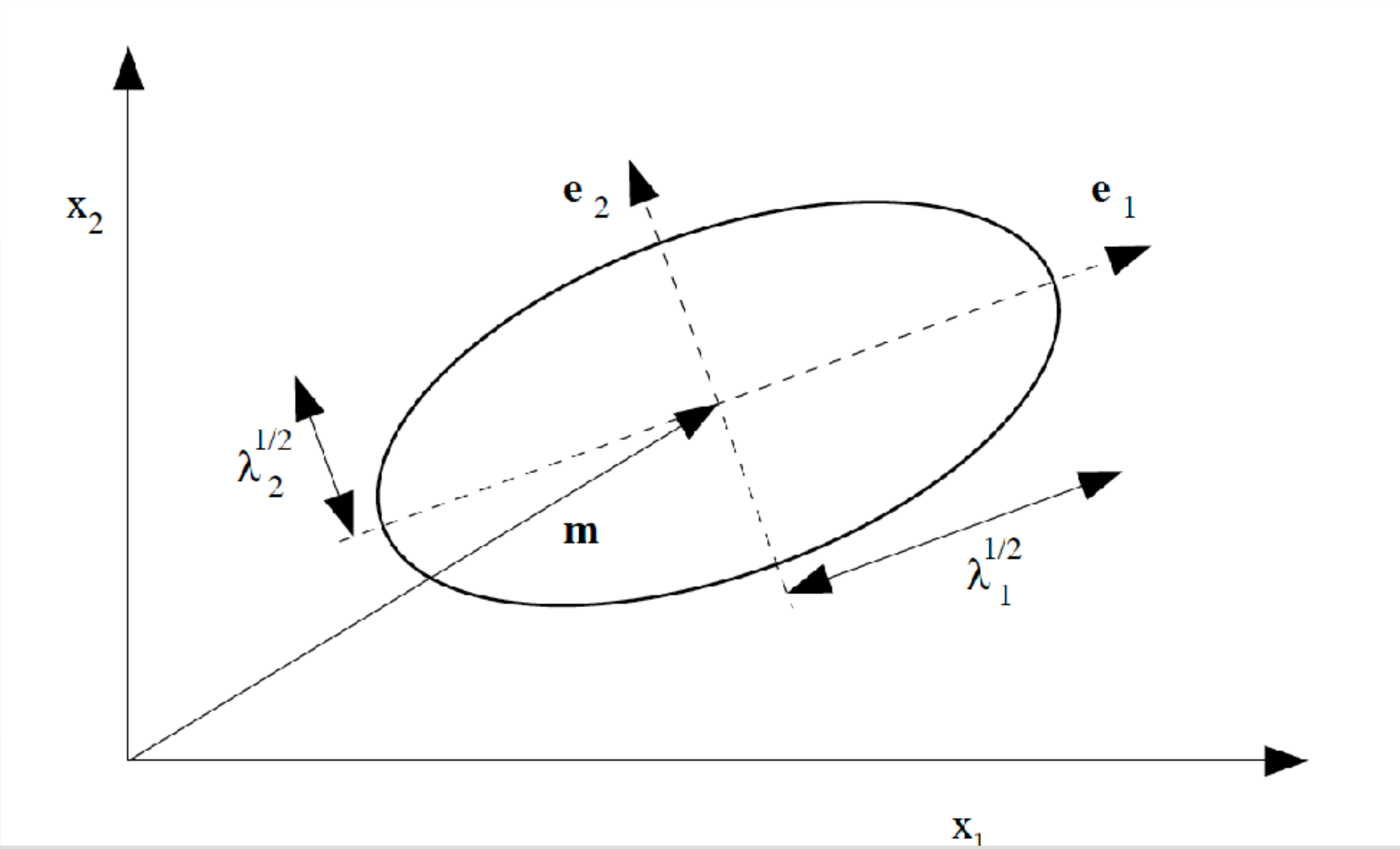
Image: Aapo Hyvaerinen, Juha Karhunen, and Erkki Oja, Independent Component Analysis

UNCORRELATEDNESS

$$\mathbf{C_x} = \mathbf{EDE}^T = \sum_{i=1}^{n} \lambda_i \mathbf{e}_i \mathbf{e}_i^T$$

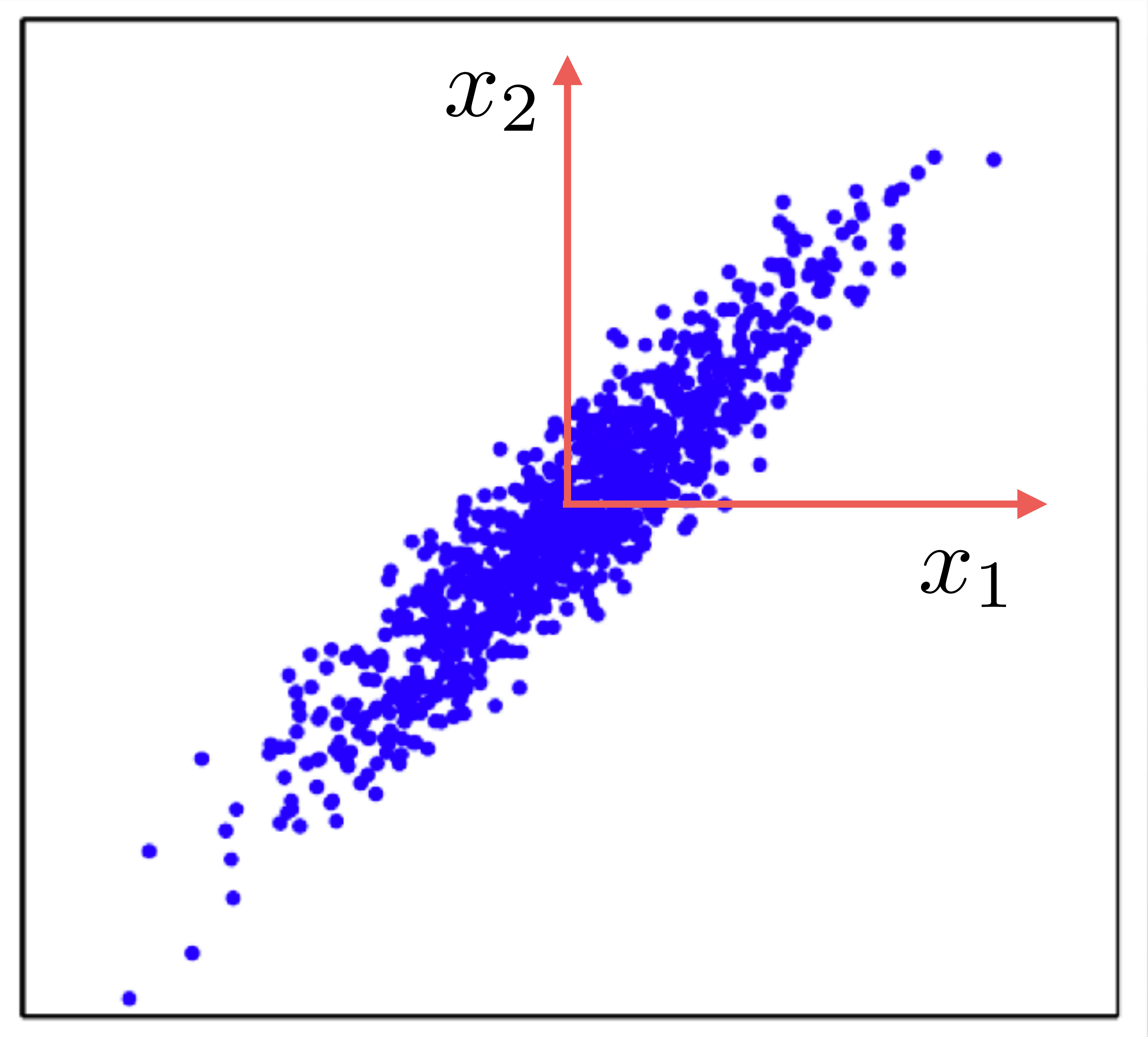
where $\mathbf{E}$ is an orthogonal matrix, i.e. rotation, having as its columns eigenvectors of covariance matrix.

And $\mathbf{D} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is diagonal matrix containing respective eigenvalues.



Applying this rotation $\mathbf{u} = \mathbf{E}^T(\mathbf{x} - \mathbf{m_x})$ to $\mathbf{x}$ will make components of $\mathbf{u}$ uncorrelated.

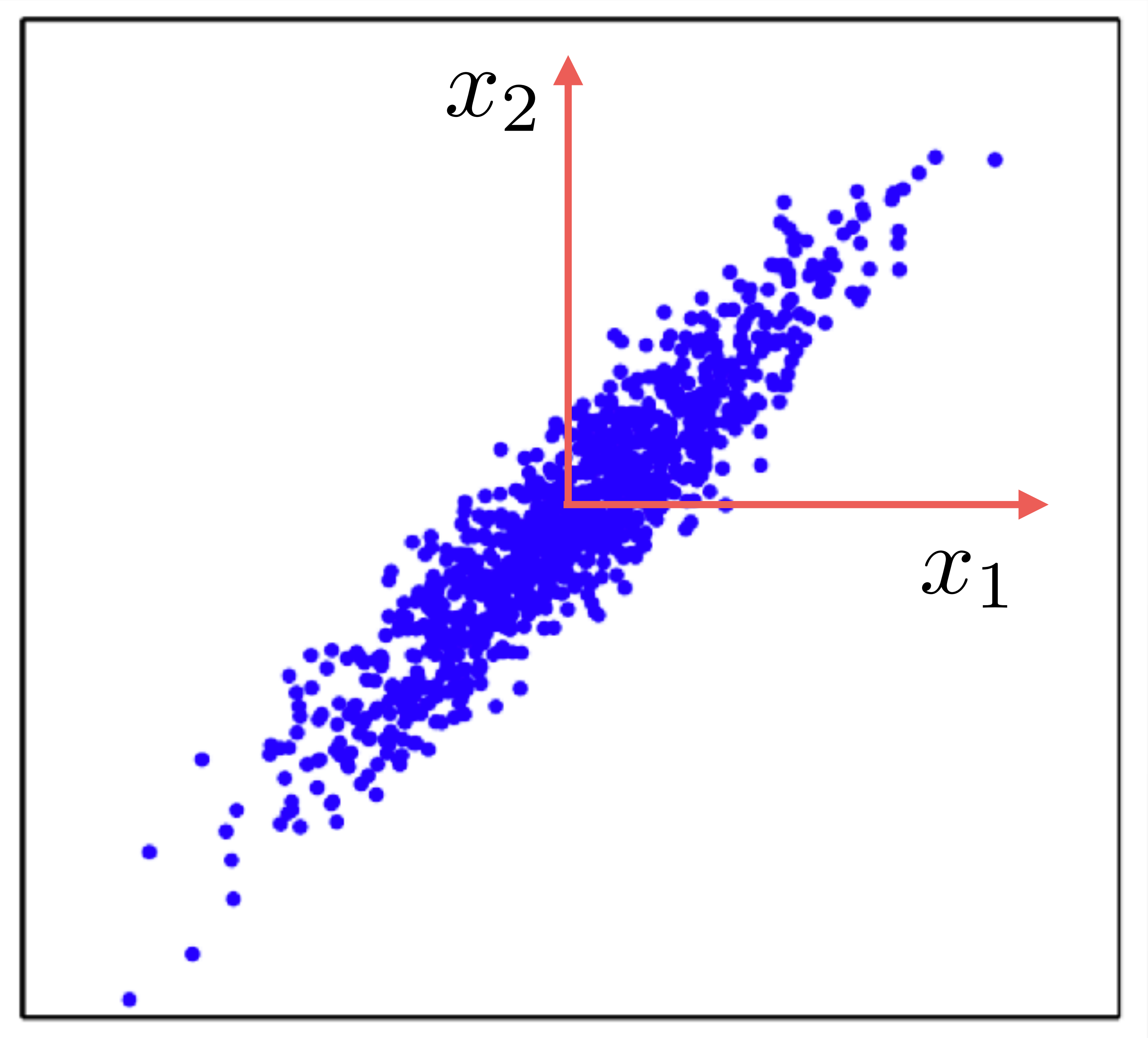
PRINCIPLE COMPONENT ANALYSIS



PCA maps original data
into a new coordinate system
which maximises variance of the data

$$y_1 = \sum_{k=1}^{n} w_{k1} x_k$$

PRINCIPLE COMPONENT ANALYSIS



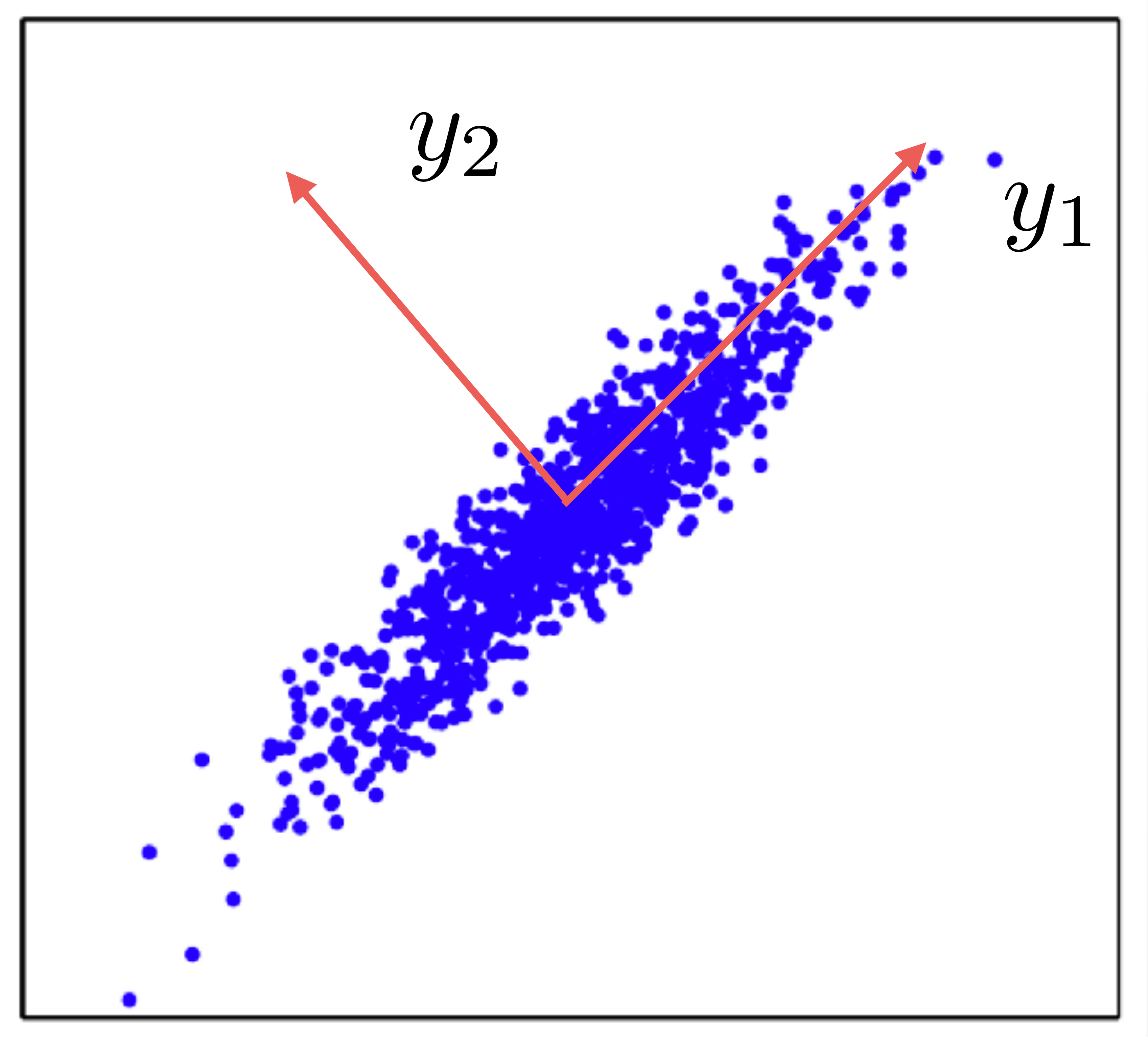
The mapping to the new basis can be expressed using the eigenvectors of the Covariance matrix

$$C = E\{\mathbf{x}\mathbf{x}^T\}$$

Eigenvalue decomposition

$$\mathbf{C} = \mathbf{U}\mathbf{D}\mathbf{U}^T$$

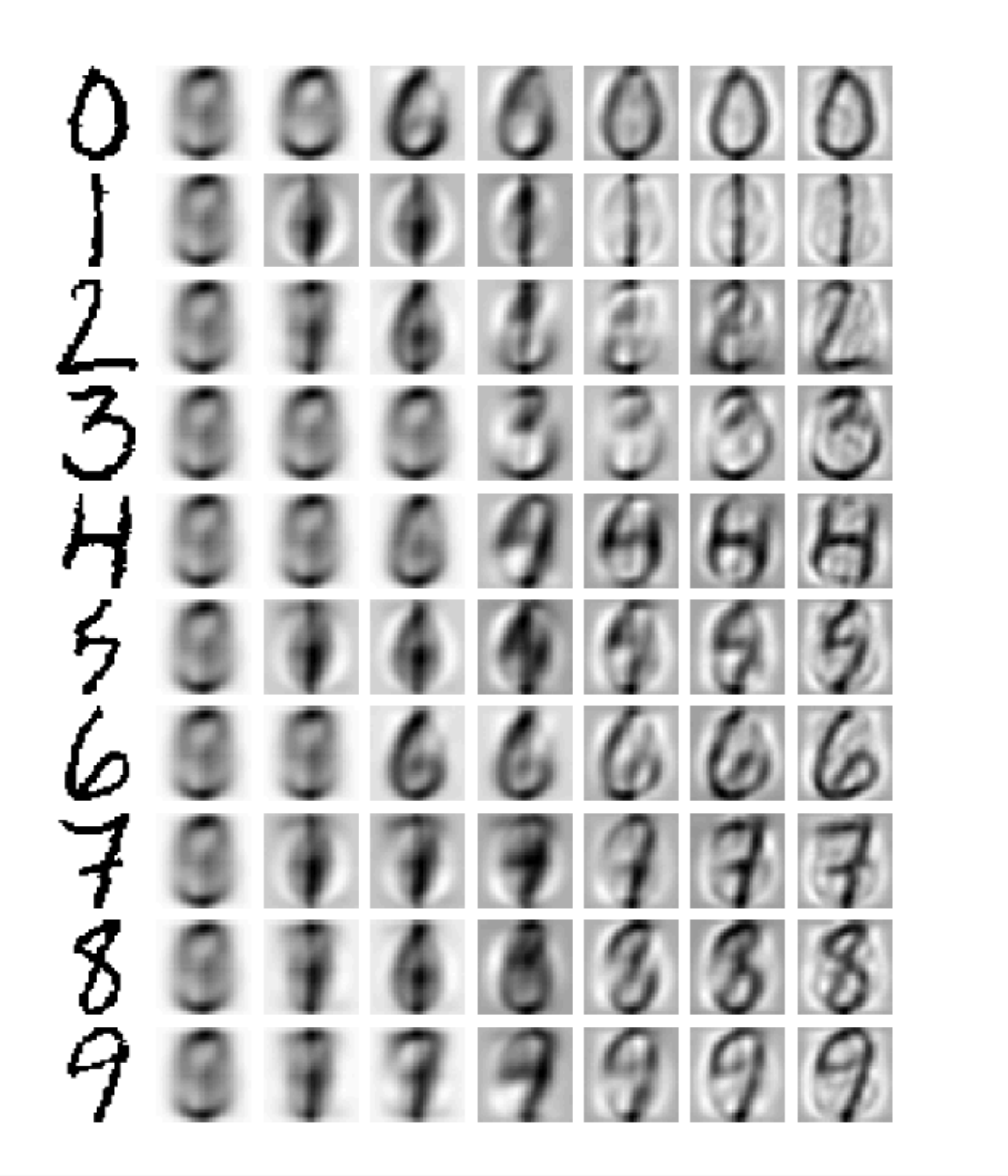
PRINCIPLE COMPONENT ANALYSIS



The vector of principle components will be

$$\mathbf{y} = \mathbf{U}^T \mathbf{x}$$

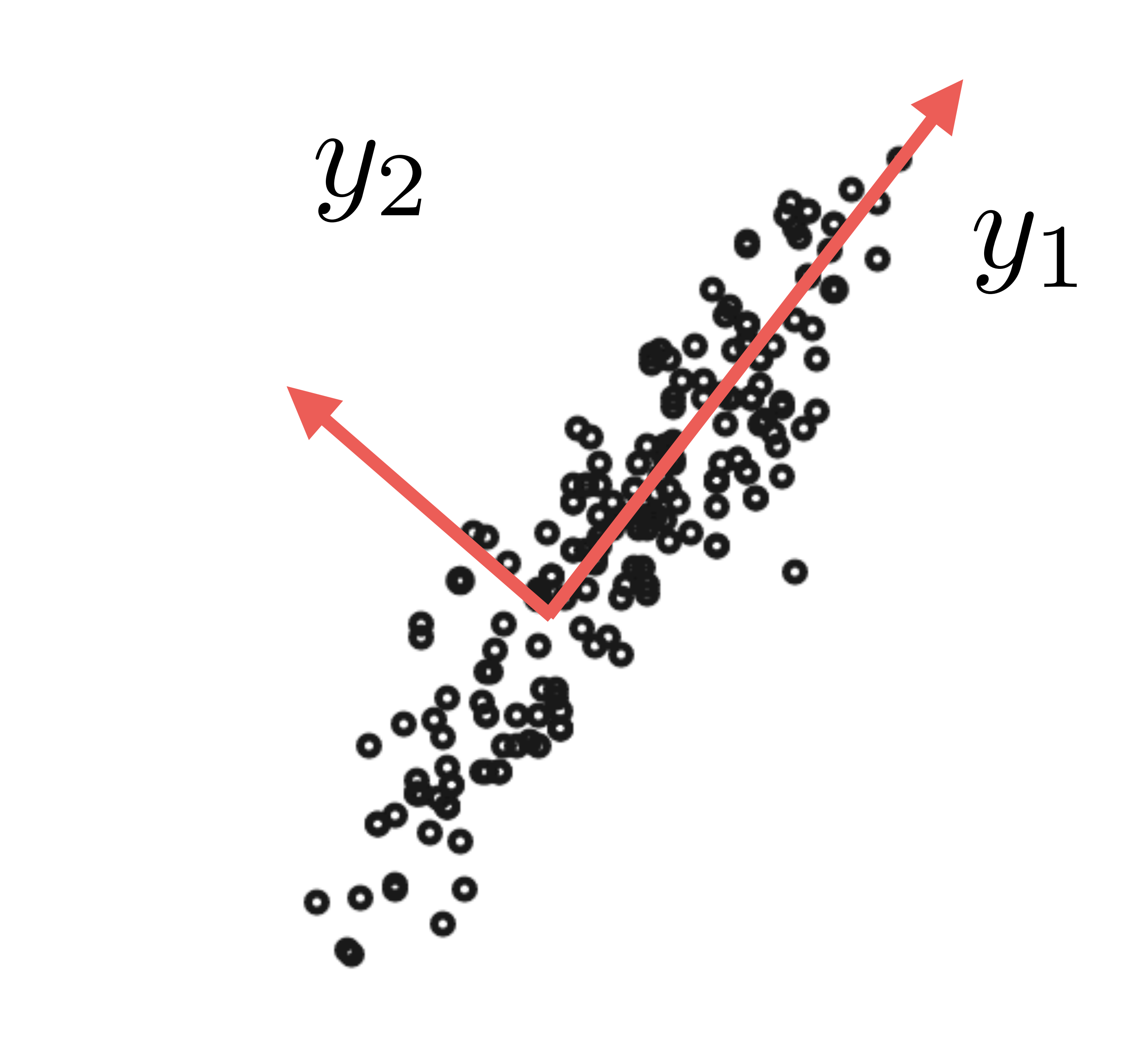
DATA COMPRESSION WITH PCA

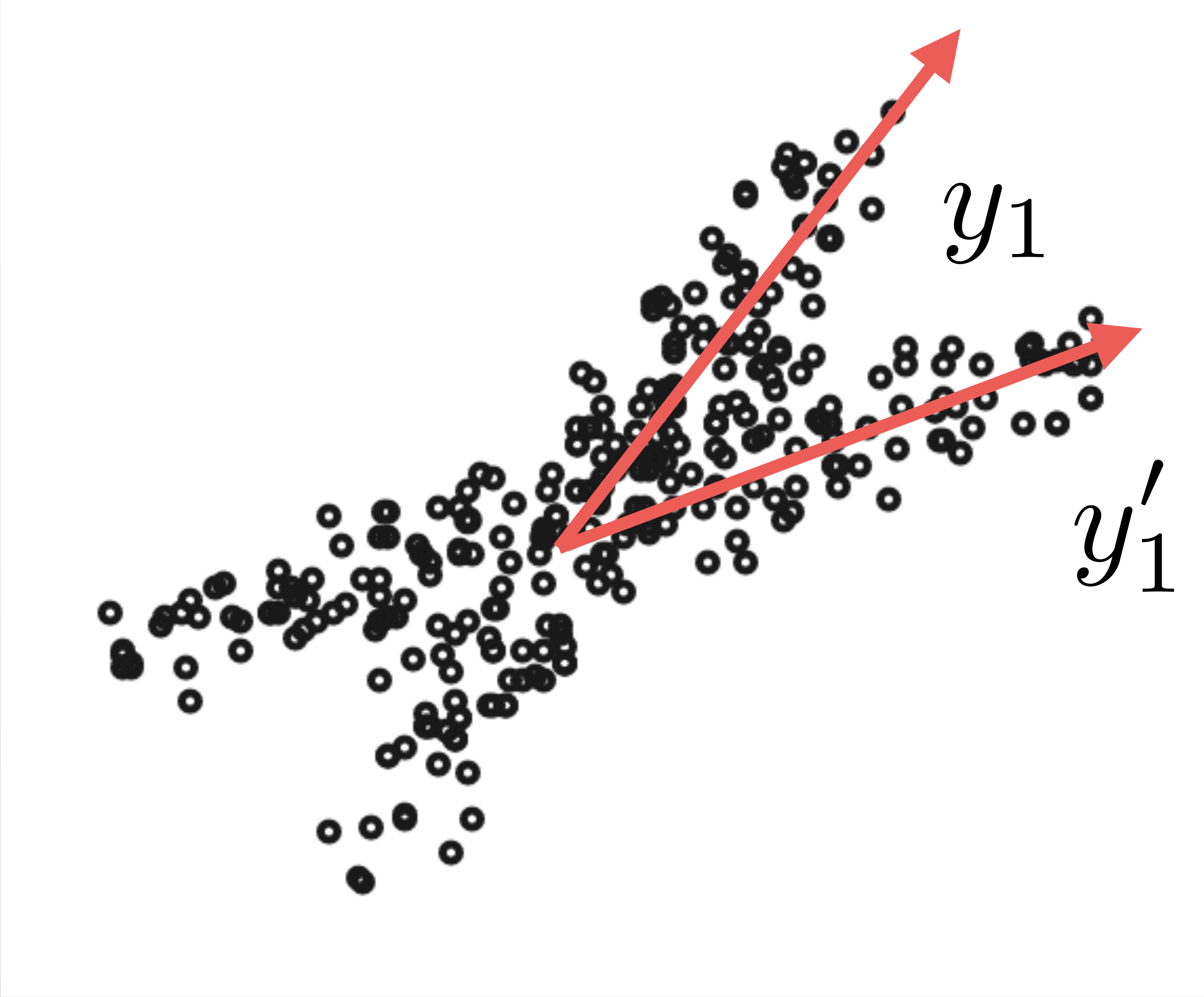


$$\hat{\mathbf{x}} = \sum_{i=1}^{m} y_i \mathbf{e}_i$$

Data compression

PRINCIPLE COMPONENT ANALYSIS





STATISTICAL INDEPENDENCE

The random variable x is independent y,
if knowing y does not give any additional information on x

$$p_{x,y}(x,y) = p_x(x)p_y(y)$$

<— joint density factorized into a product of marginal densities

Statistical independence is much stronger property than uncorrelatedness

$$\mathrm{E}\{g(x)h(y)\} = \mathrm{E}\{g(x)\}\mathrm{E}\{h(y)\}$$

If random variables are Gaussian,
independence and uncorrelatedness
become the same thing.

$$p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y}) = p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})p_{\mathbf{x}}(\mathbf{x}) = p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y})p_{\mathbf{y}}(\mathbf{y})$$

Recall

$$\mathbf{C}_{\mathbf{xy}} = \mathrm{E}\{(\mathbf{x}-\mathbf{m}_{\mathbf{x}})(\mathbf{y}-\mathbf{m}_{\mathbf{y}})^T\} = \mathbf{0}$$

which is equivalent to

$$\mathbf{R}_{\mathbf{xy}} = \mathrm{E}\{\mathbf{x}\mathbf{y}^T\} = \mathrm{E}\{\mathbf{x}\}\mathrm{E}\{\mathbf{y}^T\} = \mathbf{m}_{\mathbf{x}}\mathbf{m}_{\mathbf{y}}^T$$

STATISTICAL INDEPENDENCE

For dependent variables

$$p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y}) = p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})p_{\mathbf{x}}(\mathbf{x}) = p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y})p_{\mathbf{y}}(\mathbf{y})$$

Bayes' rule

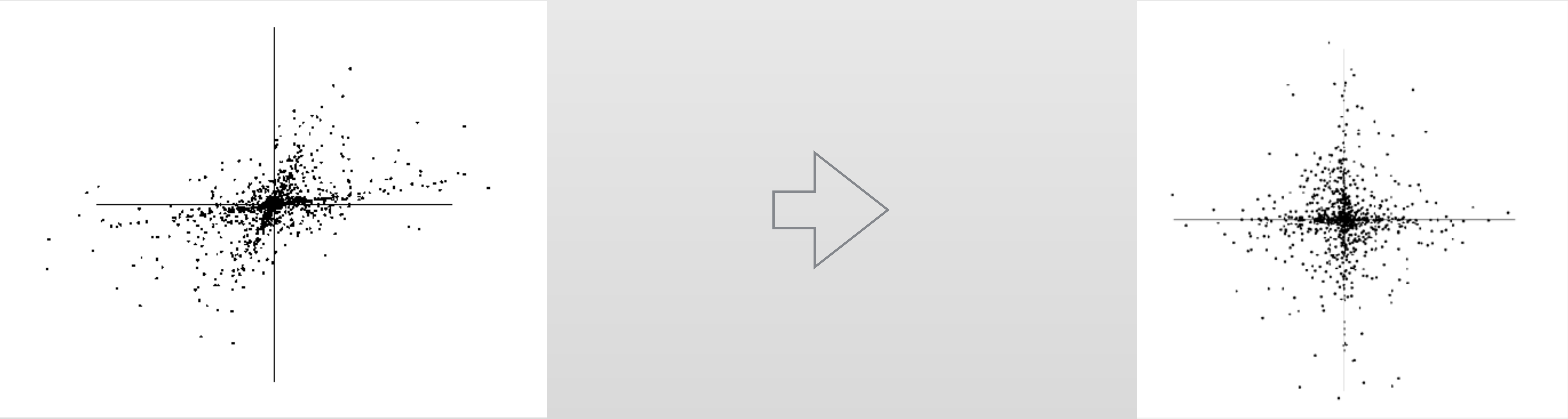
$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) = \frac{p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y})p_{\mathbf{y}}(\mathbf{y})}{p_{\mathbf{x}}(\mathbf{x})}$$

where the dominator is

$$p_{\mathbf{x}}(\mathbf{x}) = \int_{-\infty}^{\infty} p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\boldsymbol{\eta})p_{\mathbf{y}}(\boldsymbol{\eta})d\boldsymbol{\eta}$$

INDEPENDENT COMPONENT ANALYSIS

We want to extract independent components by making assumption of their non-gaussianity



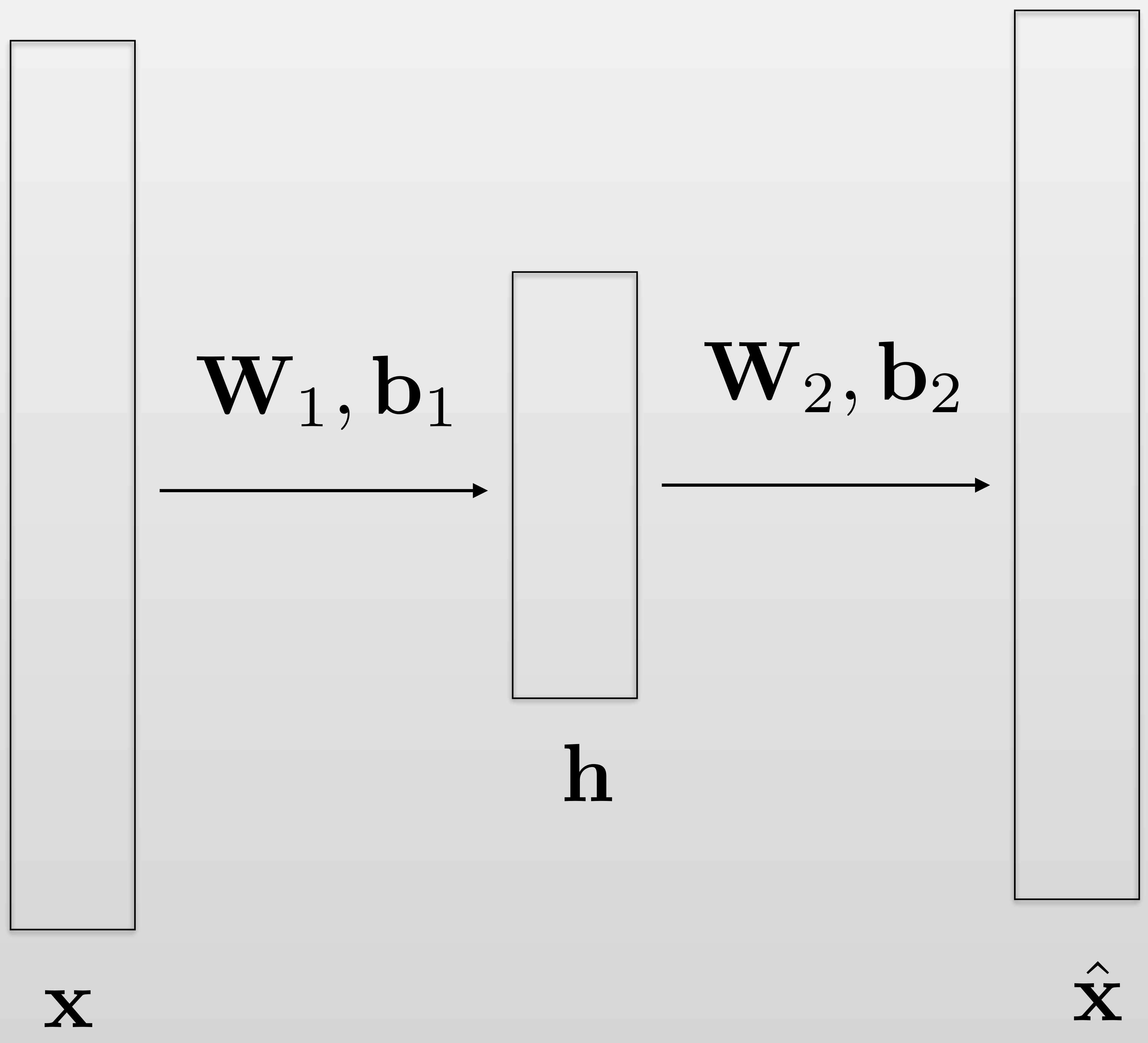
PRINCIPLE COMPONENT ANALYSIS

It has been shown that it is possible to formulate PCA in terms of Neural Networks

$$\hat{\mathbf{x}} = \mathbf{W}\mathbf{W}^T\mathbf{x}$$

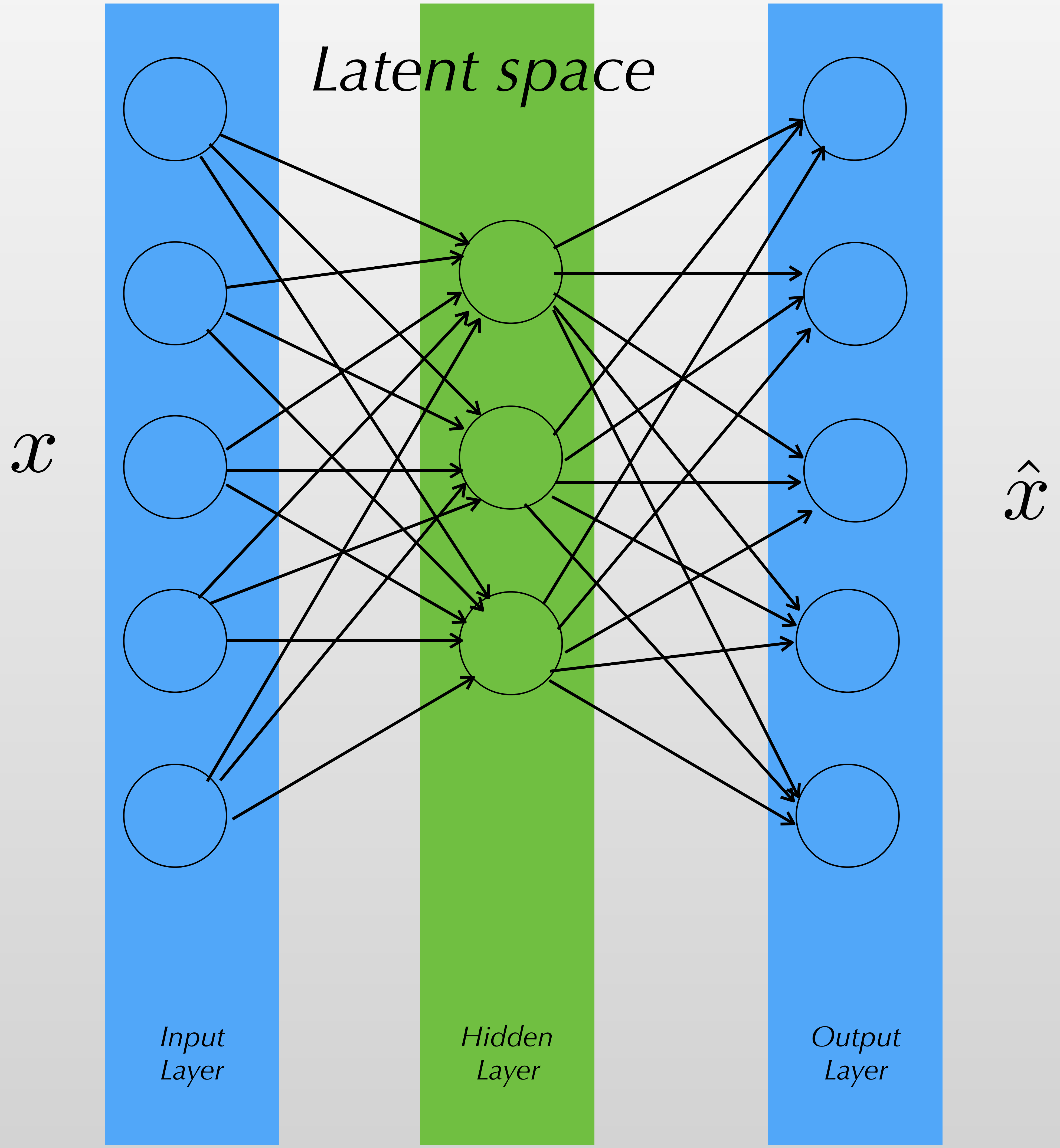
$$J_{MSE} = \frac{1}{T}\sum_{j=1}^{T} ||\hat{\mathbf{x}}(j) - \mathbf{W}\mathbf{W}^T\mathbf{x}(j)||^2$$

PRINCIPLE COMPONENT ANALYSIS



$$\mathbf{y} = \mathbf{W}_2\mathbf{h} + \mathbf{b}_2$$

AUTOENCODER

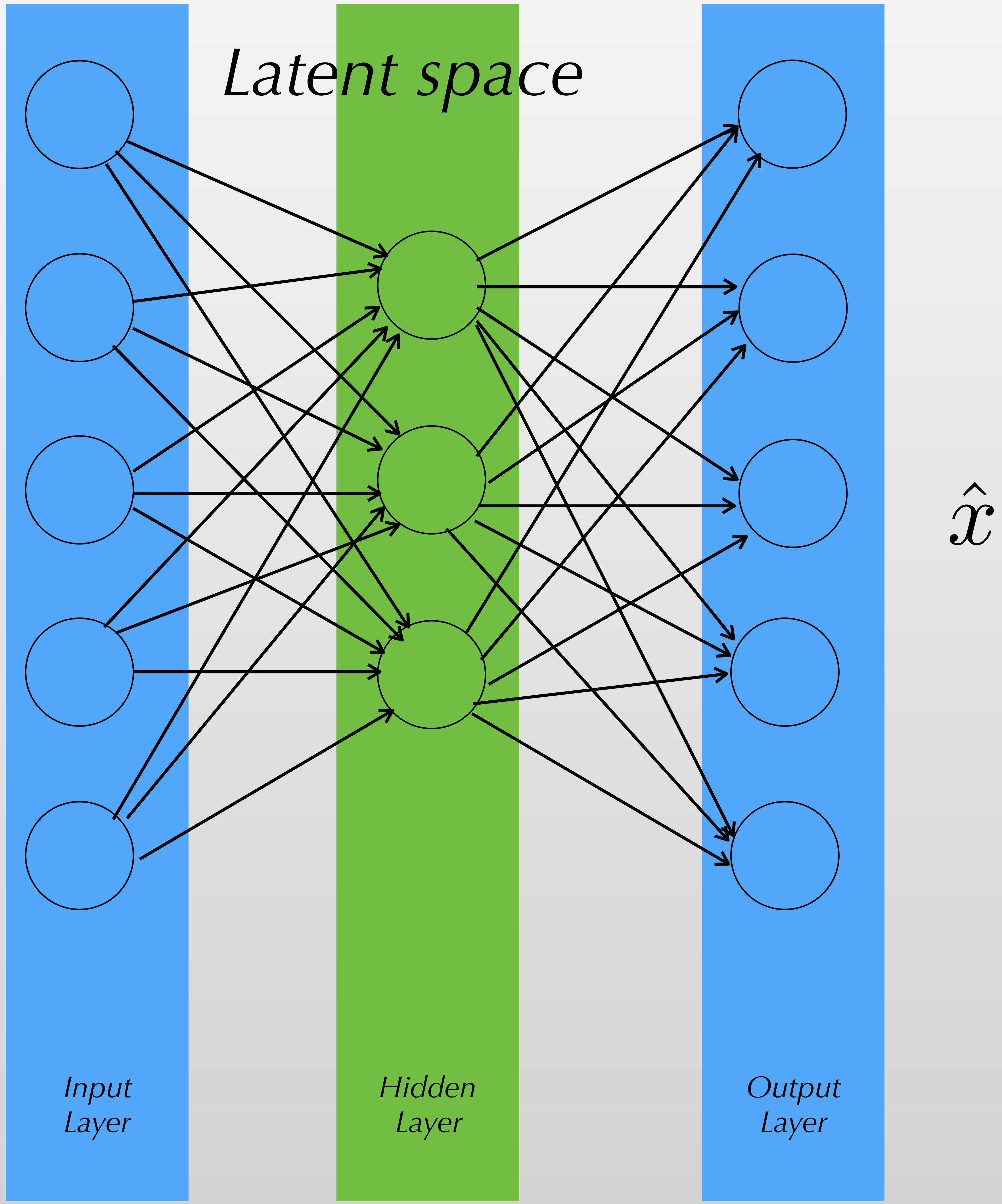


Autoencoders is unsupervised learning technique, which solves the task of representational learning.

Learning is done by comparing reconstruction to original input.

$$\mathcal{L}\left(x, \hat{x}\right)$$

AUTOENCODER



Variations:
- Denoising autoencoders
- Contractive auto encoders
- <u>Undercomplete autoencoders</u>

LINEAR VS NONLINEAR DIMENSIONALITY REDUCTION

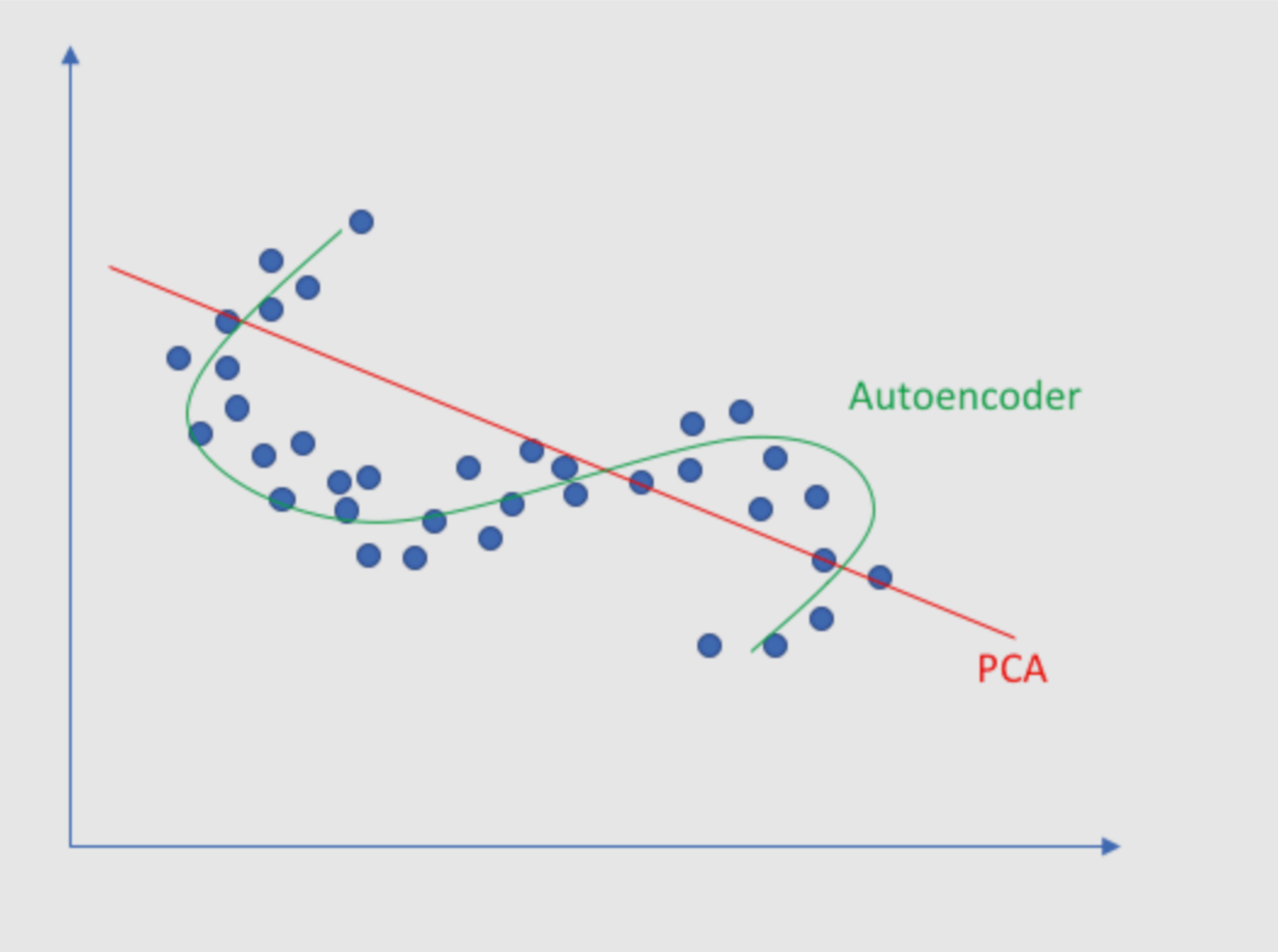
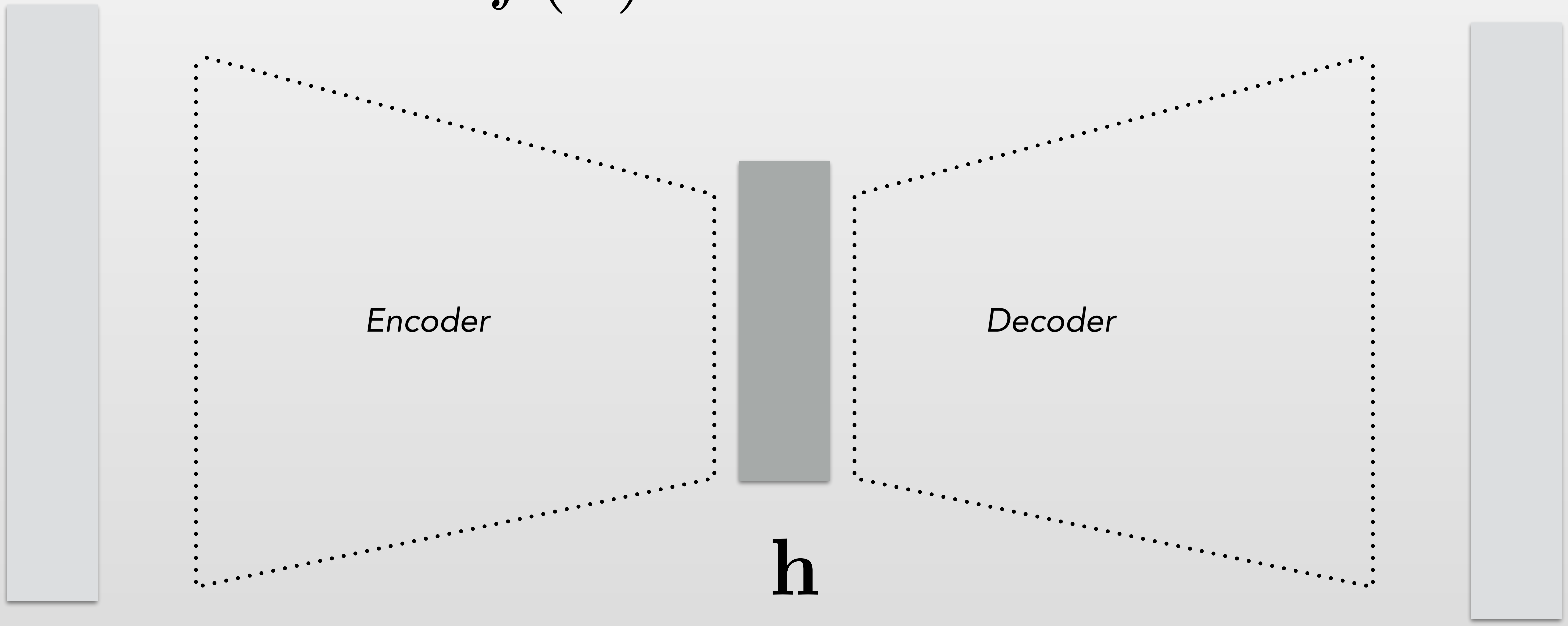


Image: https://www.jeremyjordan.me/autoencoders/

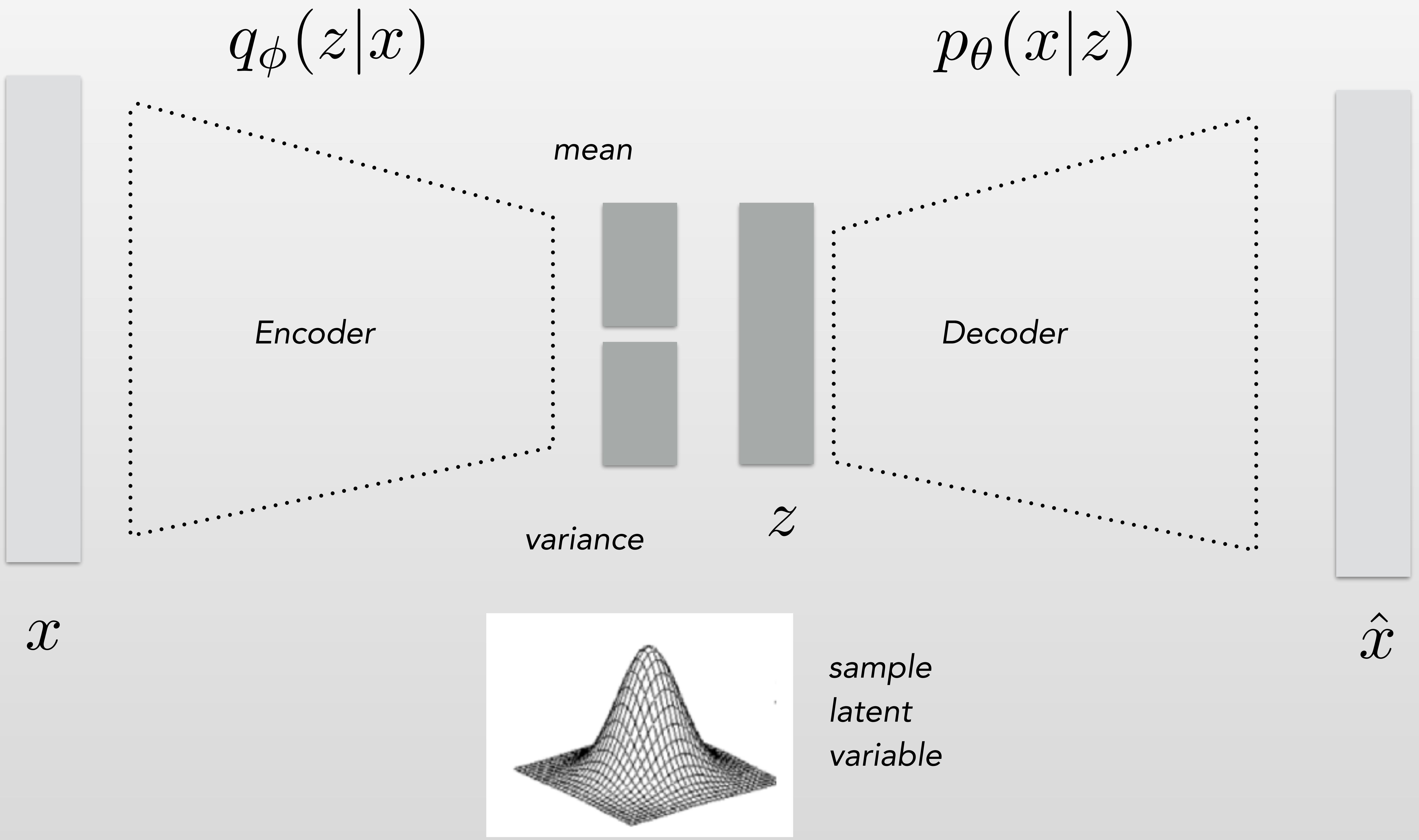
AUTOENCODER

$$\mathbf{h} = f(\mathbf{x}) \qquad \hat{\mathbf{x}} = g(\mathbf{h})$$



$\mathbf{x}$ $\mathbf{h}$ $\hat{\mathbf{x}}$

VARIATIONAL AUTOENCODER



VARIATIONAL AUTOENCODER

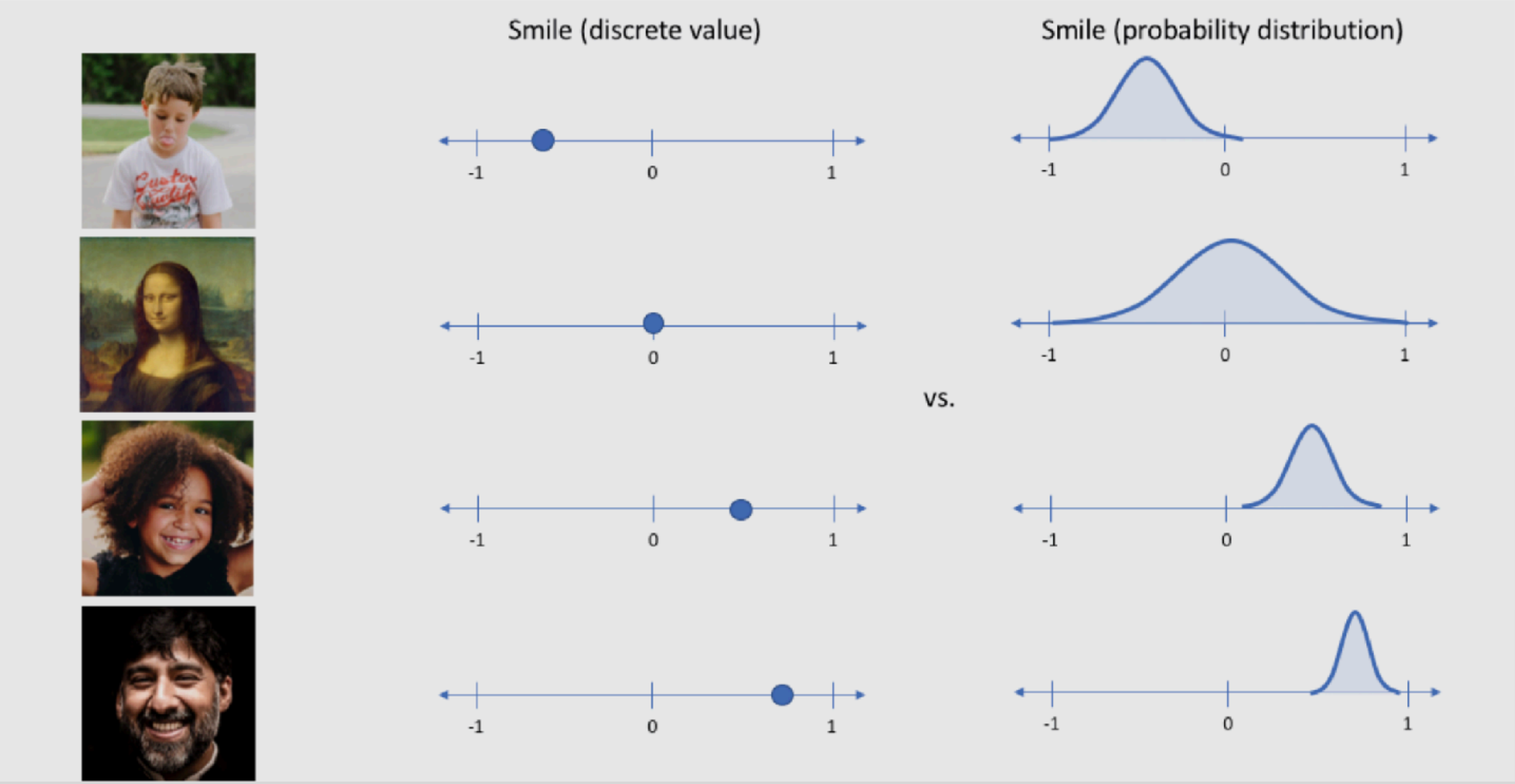
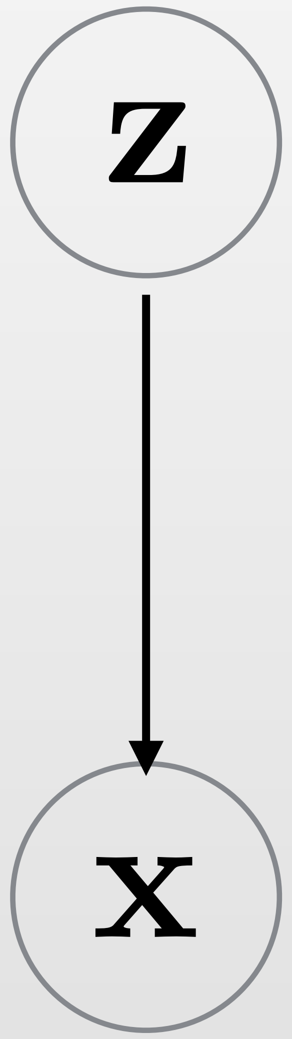


Image: https://www.jeremyjordan.me/variational-autoencoders/

VARIATIONAL AUTOENCODER



observe

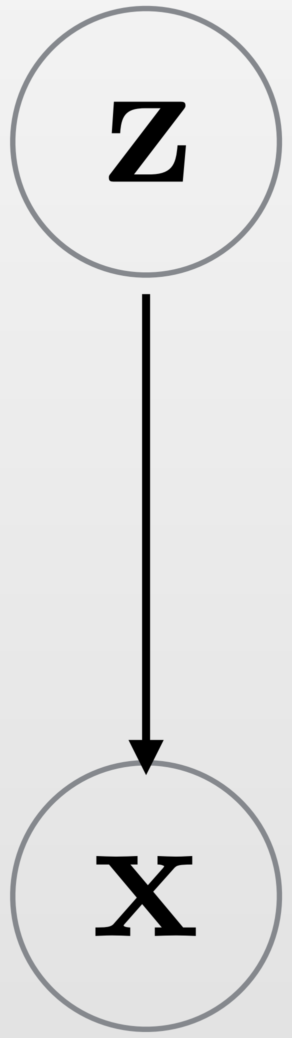
$$p(z|x) = \frac{p(x|z)p(z)}{p(x)}$$

<— *we want to estimate the latent variables given the data*

$$p(x) = \int p(x|z)p(z)dz = E_{p(z)}\left[p(x|z)\right]$$

<— *this is intractable*

VARIATIONAL AUTOENCODER



observe

Lets approximate $p(z|x)$ *with* $q(z|x)$

such that we set a condition that they are close to each other as possible.

We can enforce this condition by minimising Kullback–Leibler divergence

KL DIVERGENCE

Information

$$I = -\log p(x)$$

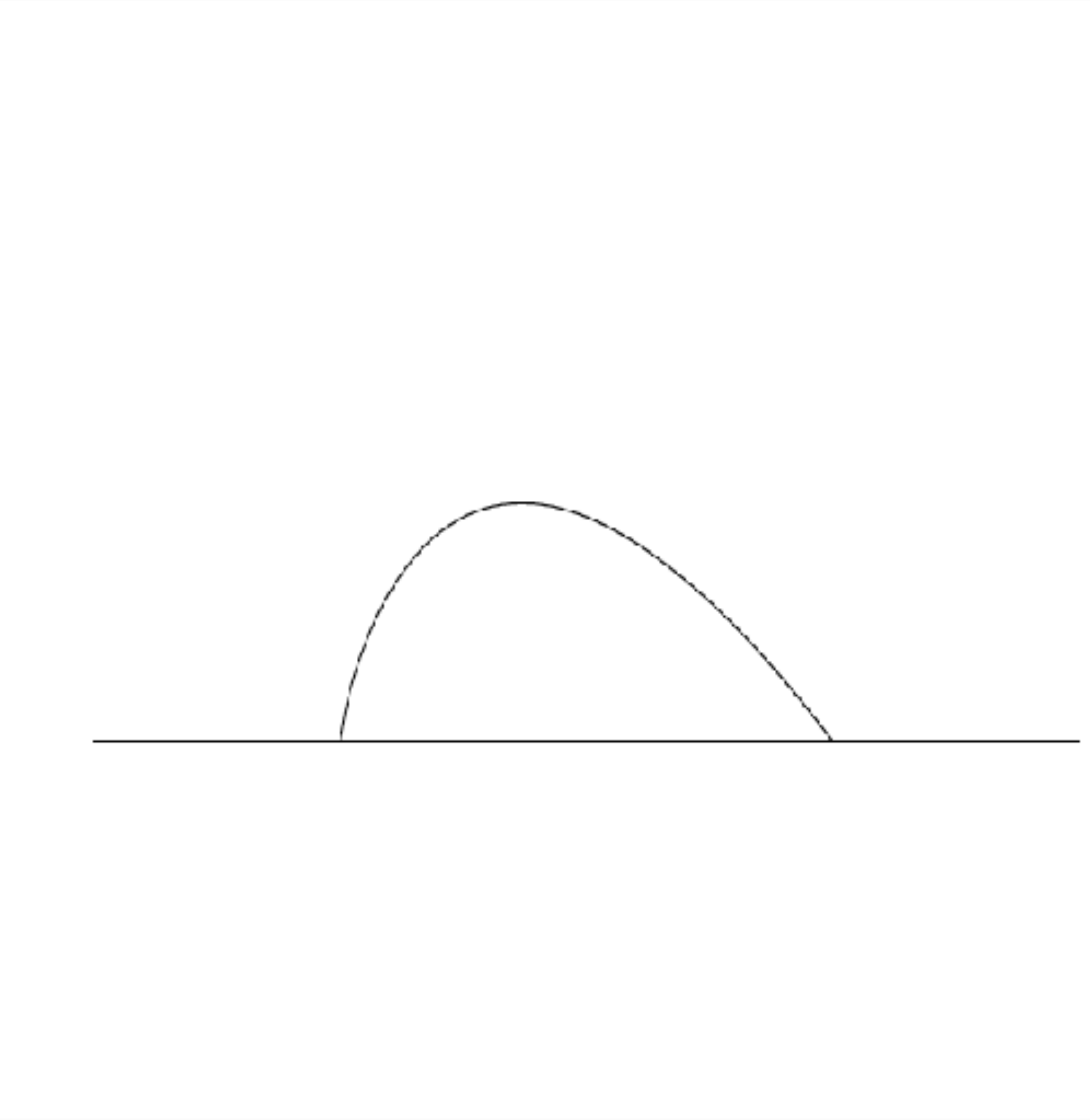
Entropy

$$H = -\sum p(x) \log p(x)$$

KL - divergence

$$D_{\mathrm{KL}}(P \parallel Q) = -\sum_{x \in \mathcal{X}} p(x) \log q(x) + \sum_{x \in \mathcal{X}} p(x) \log p(x)$$
$$= \mathrm{H}(P, Q) - \mathrm{H}(P)$$

$$D_{\mathrm{KL}}(P \parallel Q) = -\sum_{x \in \mathcal{X}} P(x) \log \left(\frac{Q(x)}{P(x)} \right)$$



Mutual information can be measured as KL divergence

VARIATIONAL AUTOENCODER

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

Introduce tractable $q(z|x)$

$$= \log \int p(x,z)\frac{q(z|x)}{q(z|x)}\mathrm{d}z \qquad \geq \mathbb{E}_{q(z|x)}\log\frac{p(x,z)}{q(z|x)}$$

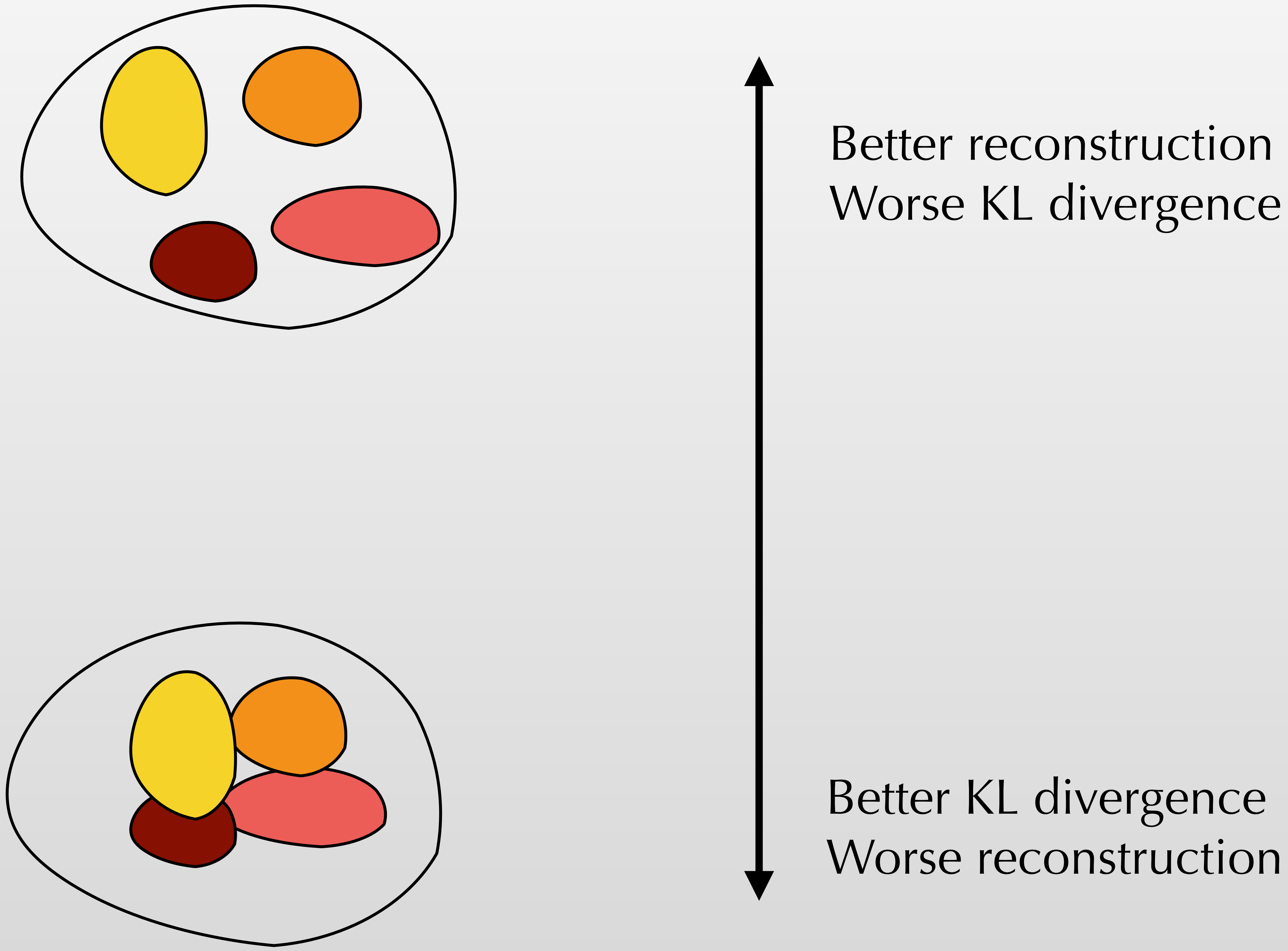
Jensen inequality

$$= \mathbb{E}_{q(z|x)}\log\frac{p(x|z)p(z)}{q(z|x)} \qquad = \mathbb{E}_{q(z|x)}\log p(x|z) + \mathbb{E}_{q(z|x)}\log\frac{p(z)}{q(z|x)}$$

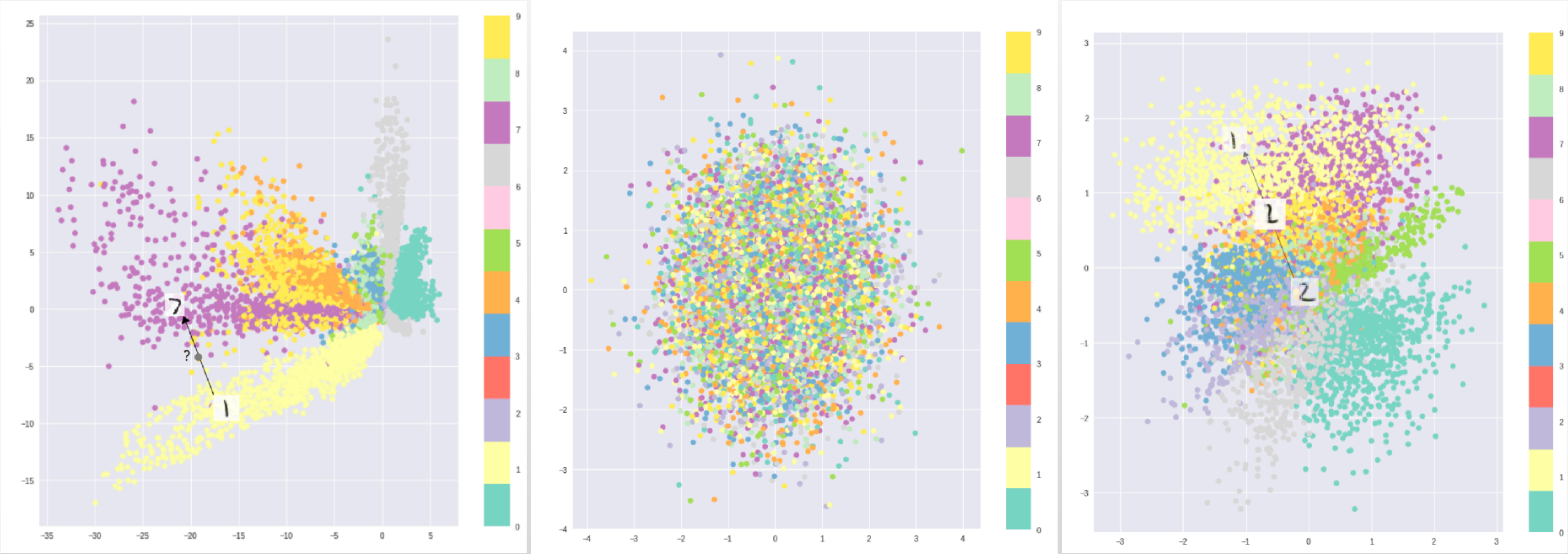
$$= \text{likelihood} - D_{KL}[q(z|x)||p(z)]$$

ELBO — evidence lower bound

VARIATIONAL AUTOENCODER



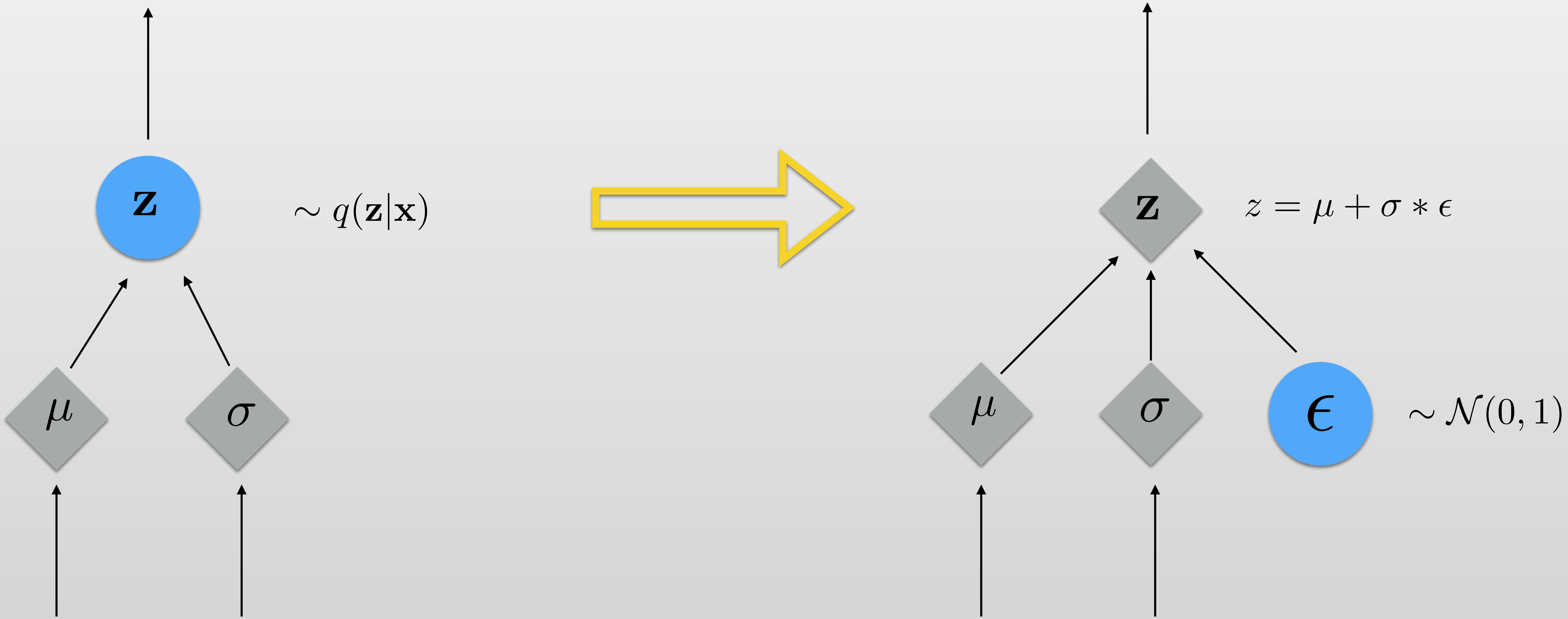
VARIATIONAL AUTOENCODER



https://towardsdatascience.com/intuitively-understanding-variational-autoencoders-1bfe67eb5daf

VARIATIONAL AUTOENCODER

Reparemeterization trick: used to propagate back the error

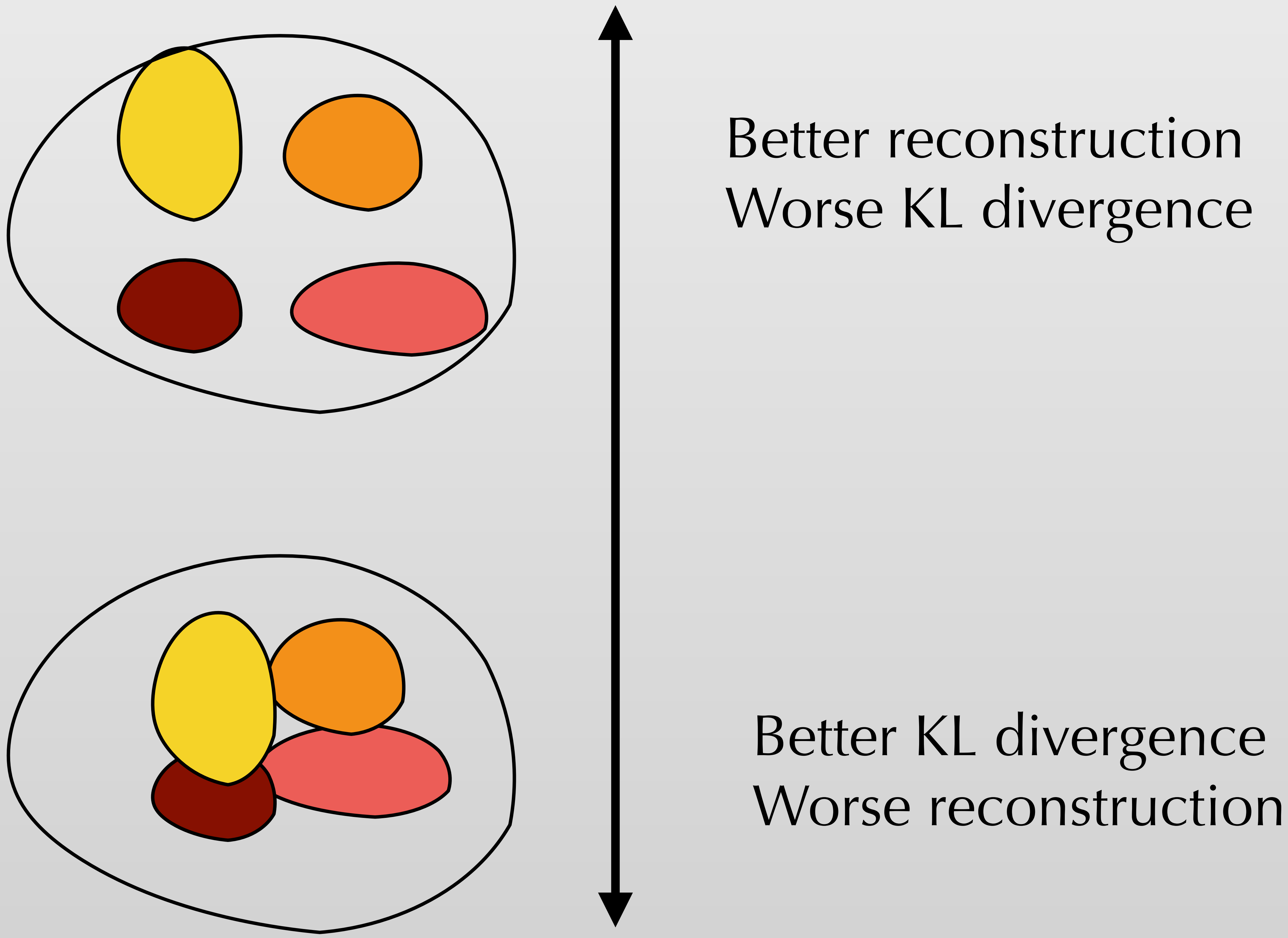


β - VAE

Introduce parameterisation into the learning criterion

$$\mathcal{L}(\theta, \phi; \mathbf{x}, \mathbf{z}, \beta) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] - \beta\, D_{KL}(q_\phi(\mathbf{z}|\mathbf{x})||p(\mathbf{z}))$$

Which allows to control this:



DENOISING AUTOENCODER

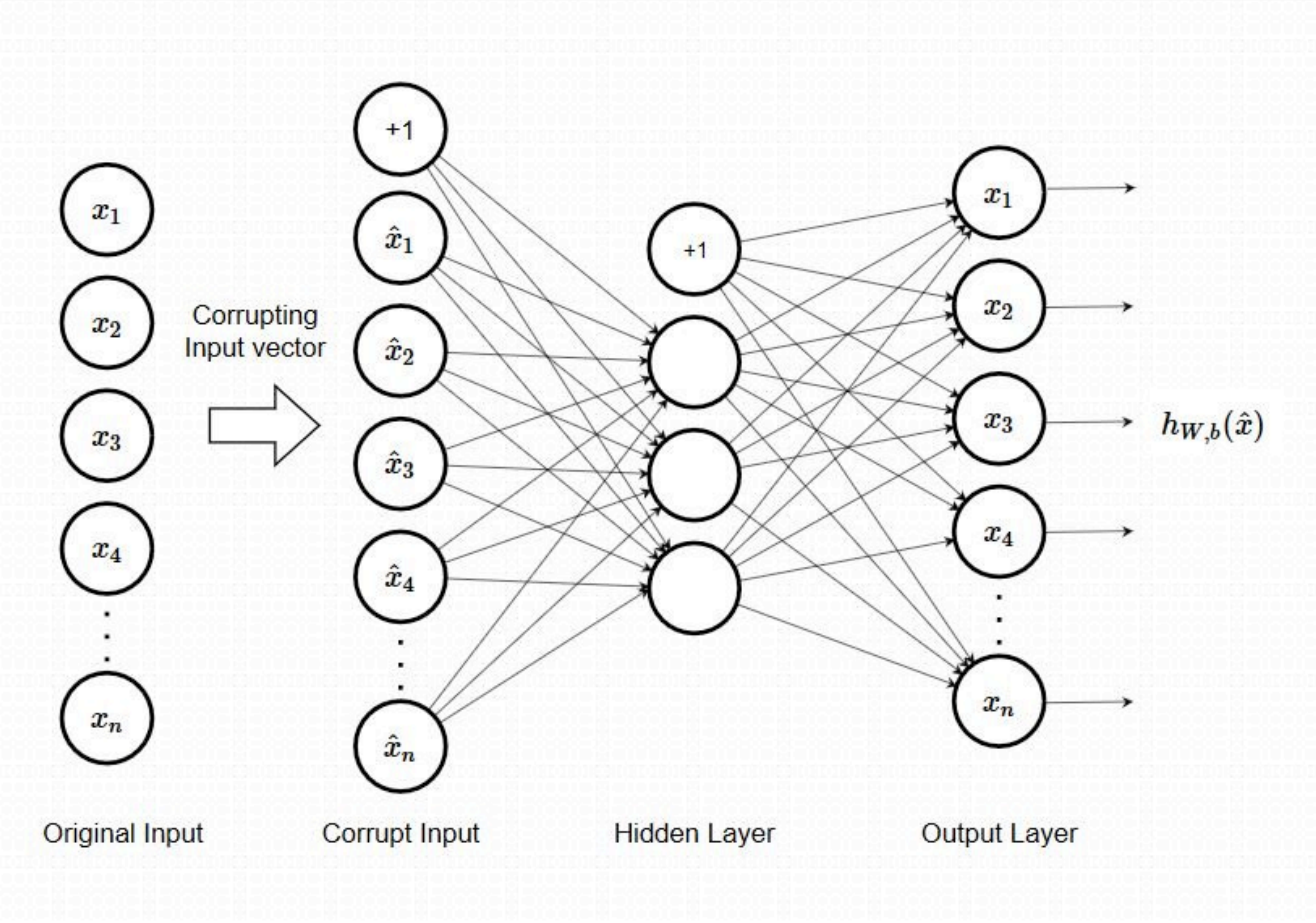
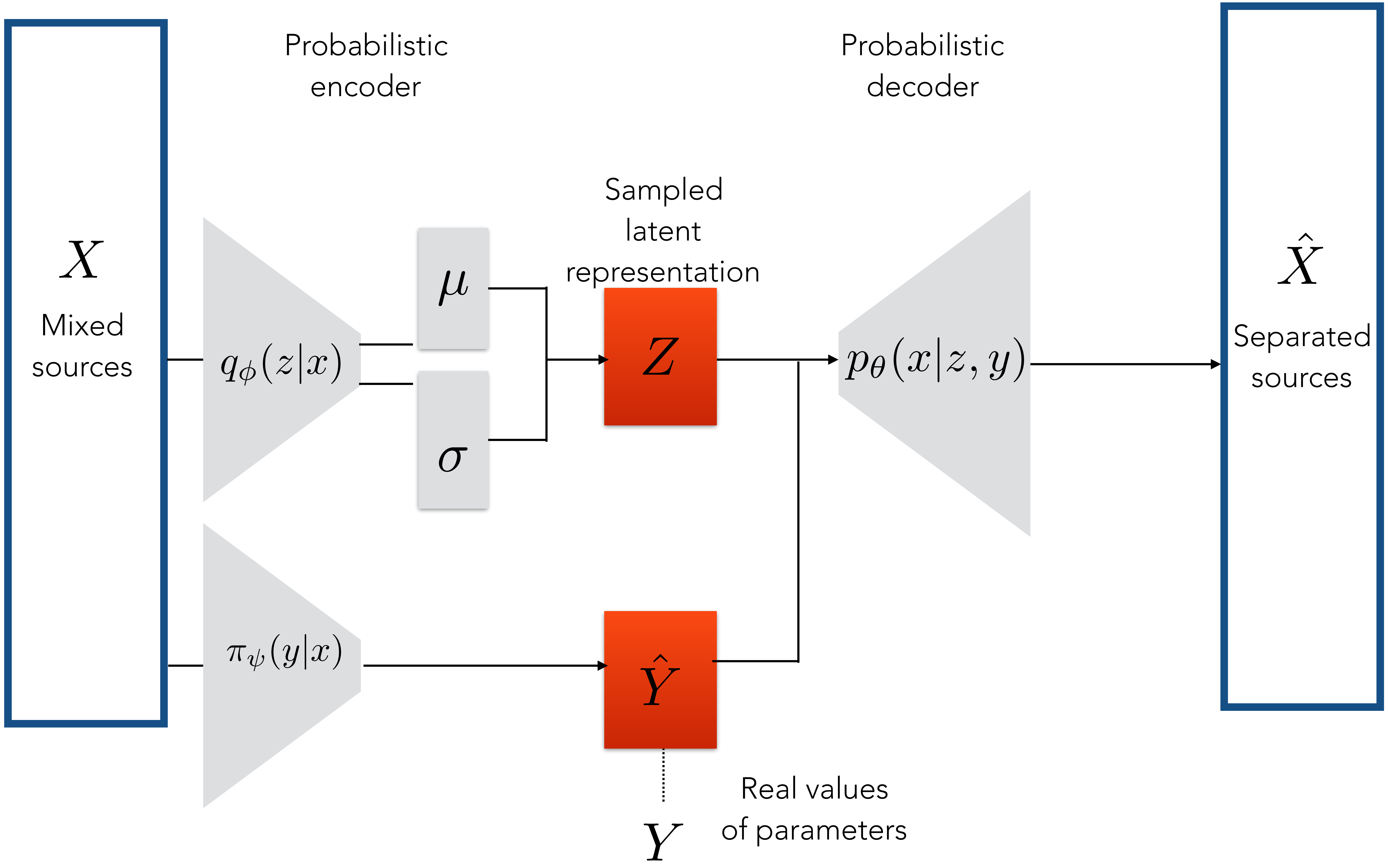


Image: Static hand gesture recognition using stacked Denoising Sparse Autoencoders, Kumar, Nandi, Kala

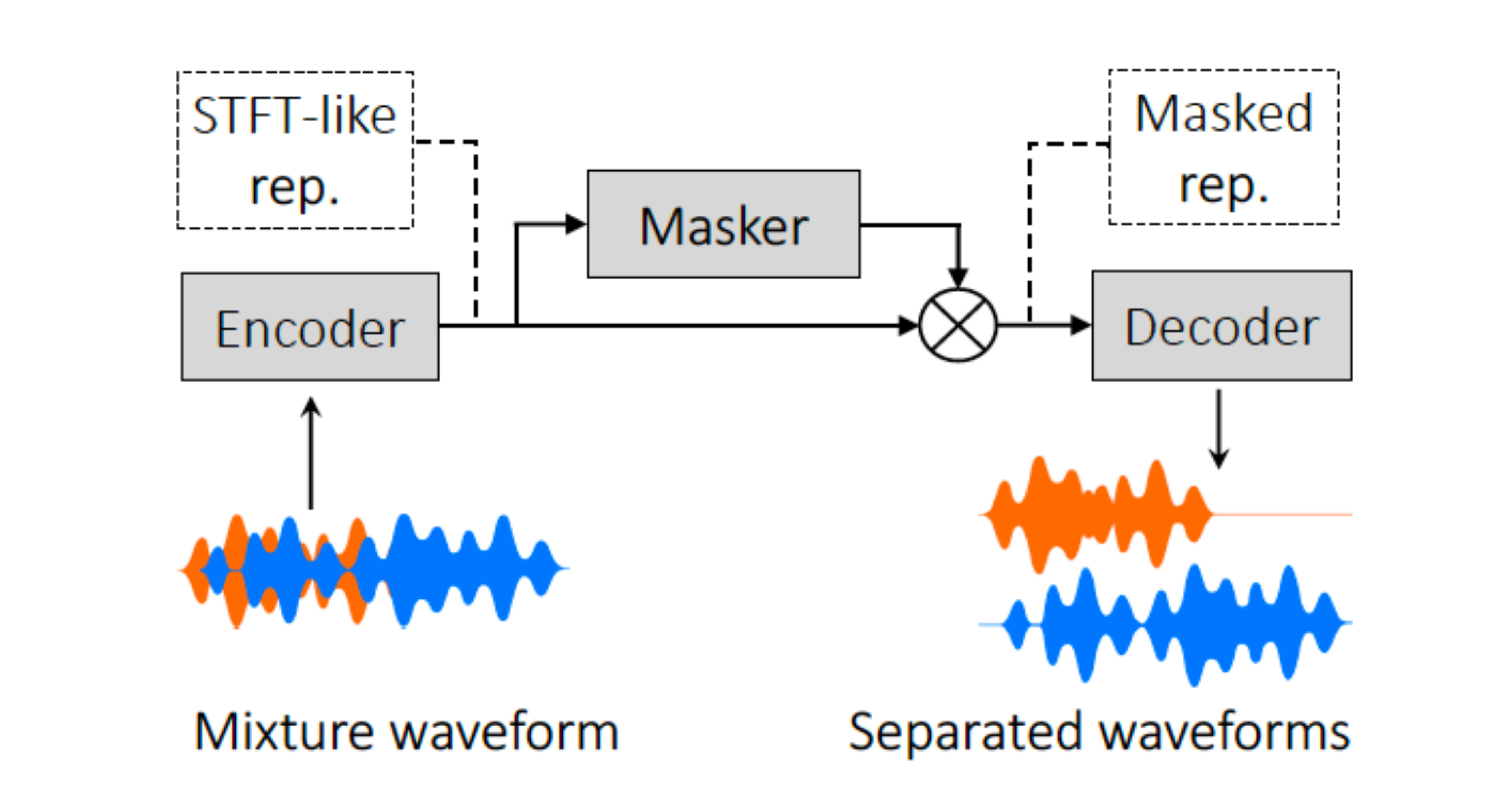
SEMI-SUPERVISED LEARNING



MODERN APPROACHES TO SOURCE SEPARATION

Deep Learning allowed for a huge jump in the performance of the algorithms for source separation.

Typical architecture:



Frameworks to that implement popular approaches and provide training datasets. For example, Asteroid.

Image: Asteroid: the PyTorch-based audio source separation toolkit for researchers, Pariente et al

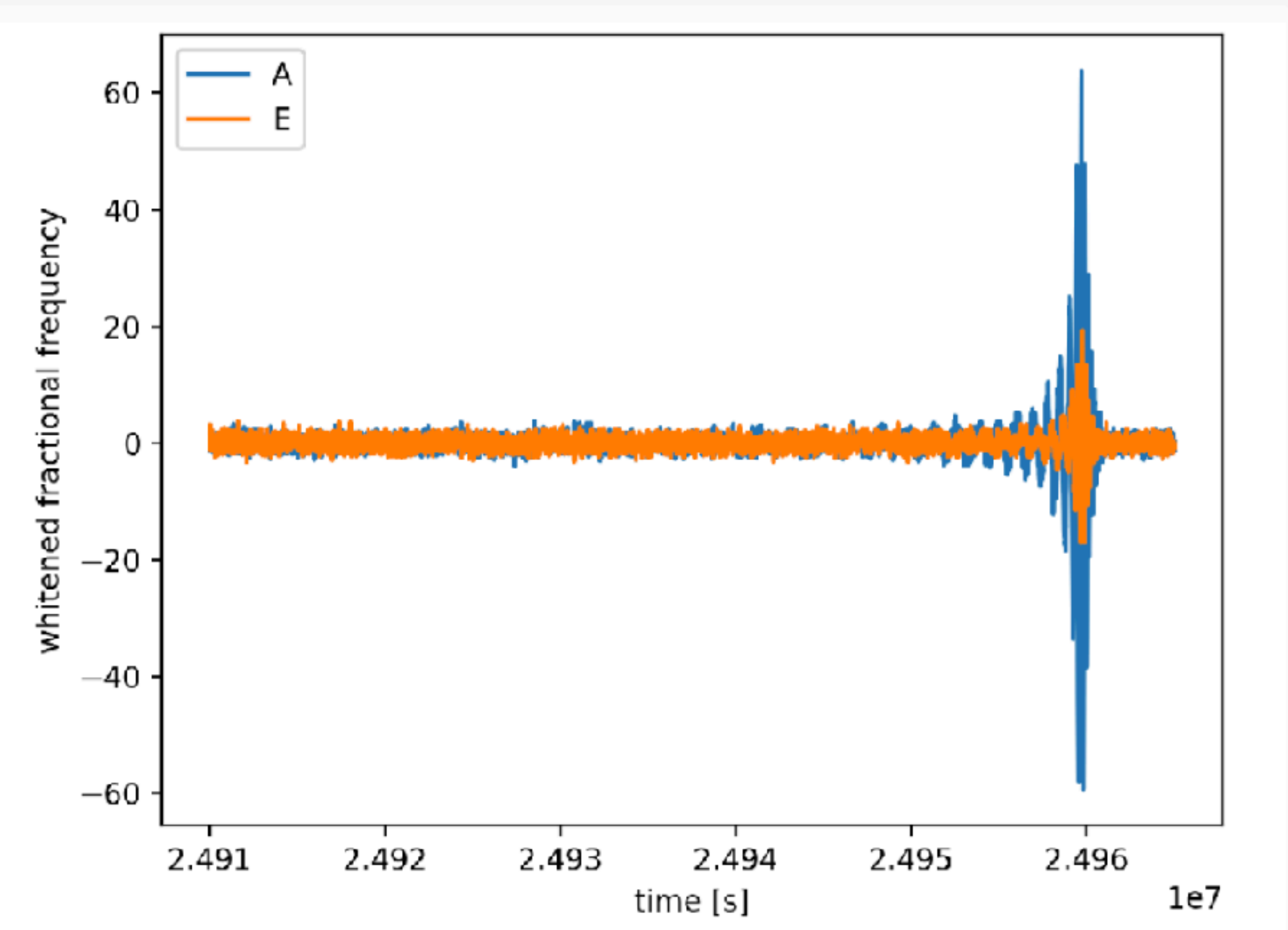
FAST PARAMETER ESTIMATION

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
- …



MASSIVE BLACK HOLE BINARIES EM COUNTERPARTS

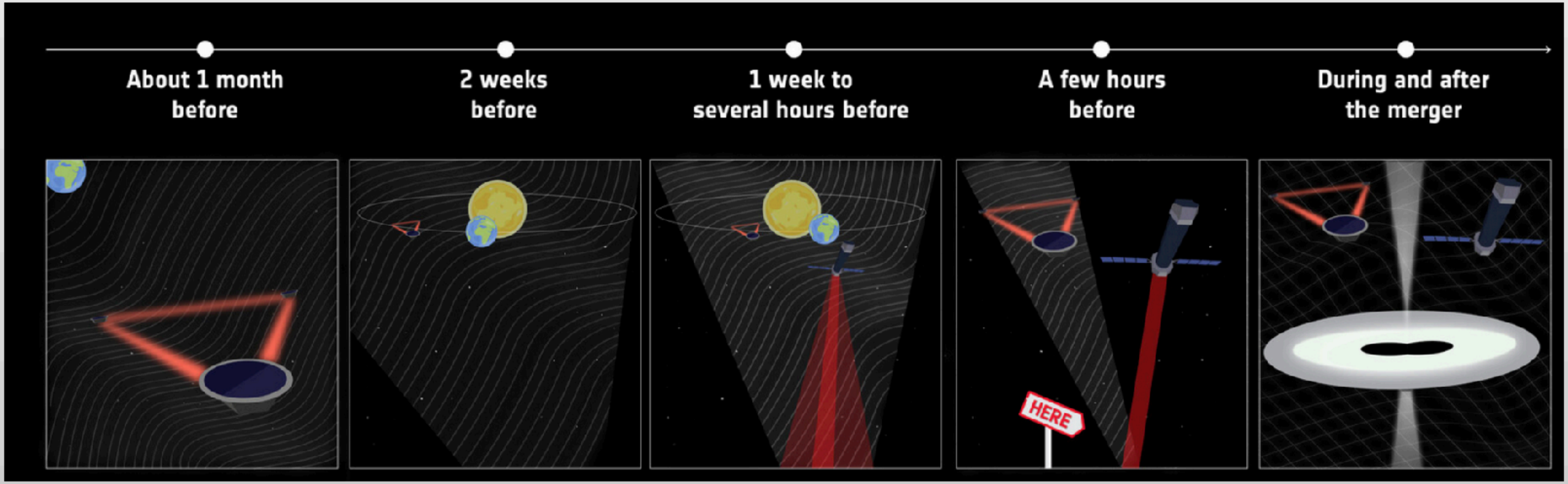


Image: ESA

LISA — Athena synergy heavily relies on the estimation of the sky-localisation all the way along the observation of the gravitational wave for MBHBs

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

Marginal likelihood

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:

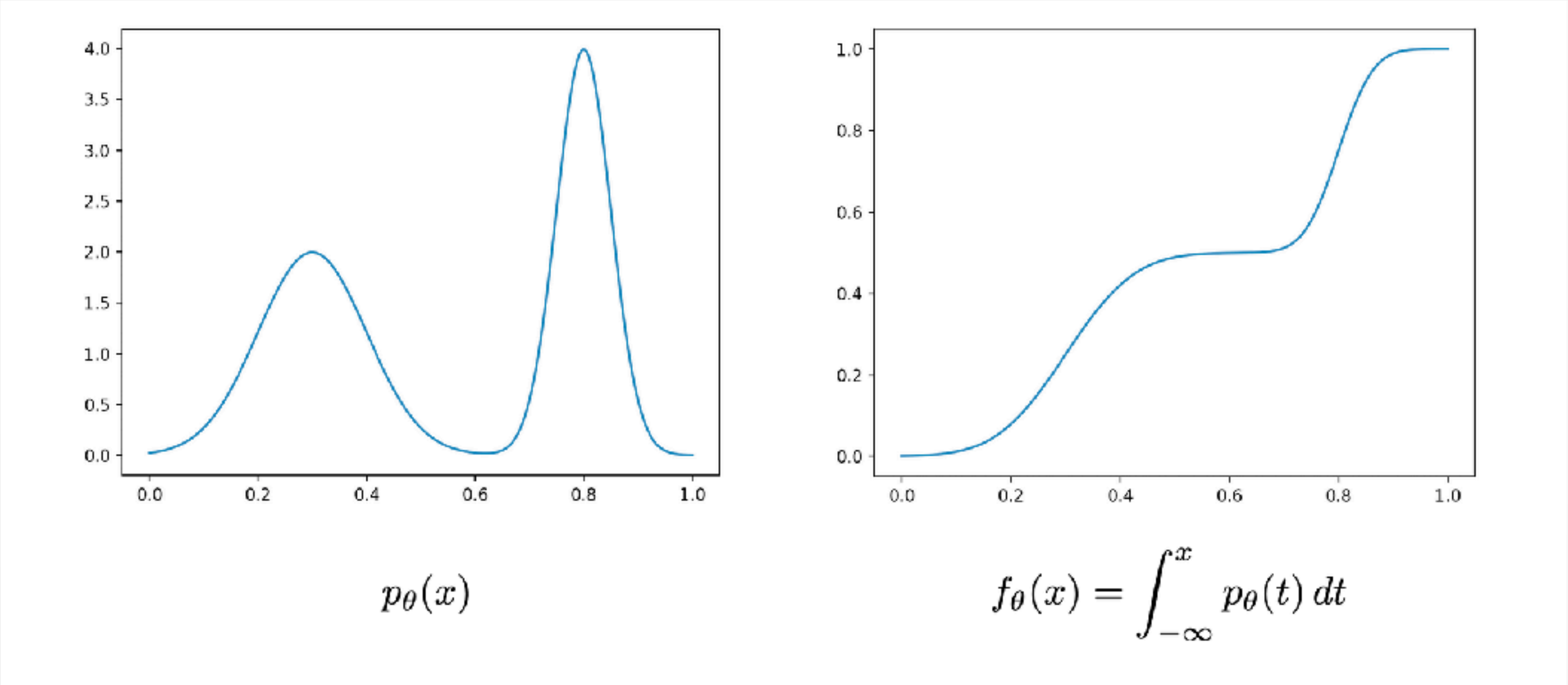
- Sample from the exact posterior: MCMC or Nested sampling (slow)
- Variational Inference: approximate the posterior distribution with a tractable distribution

There are some exceptions for the models with some simplifications:

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

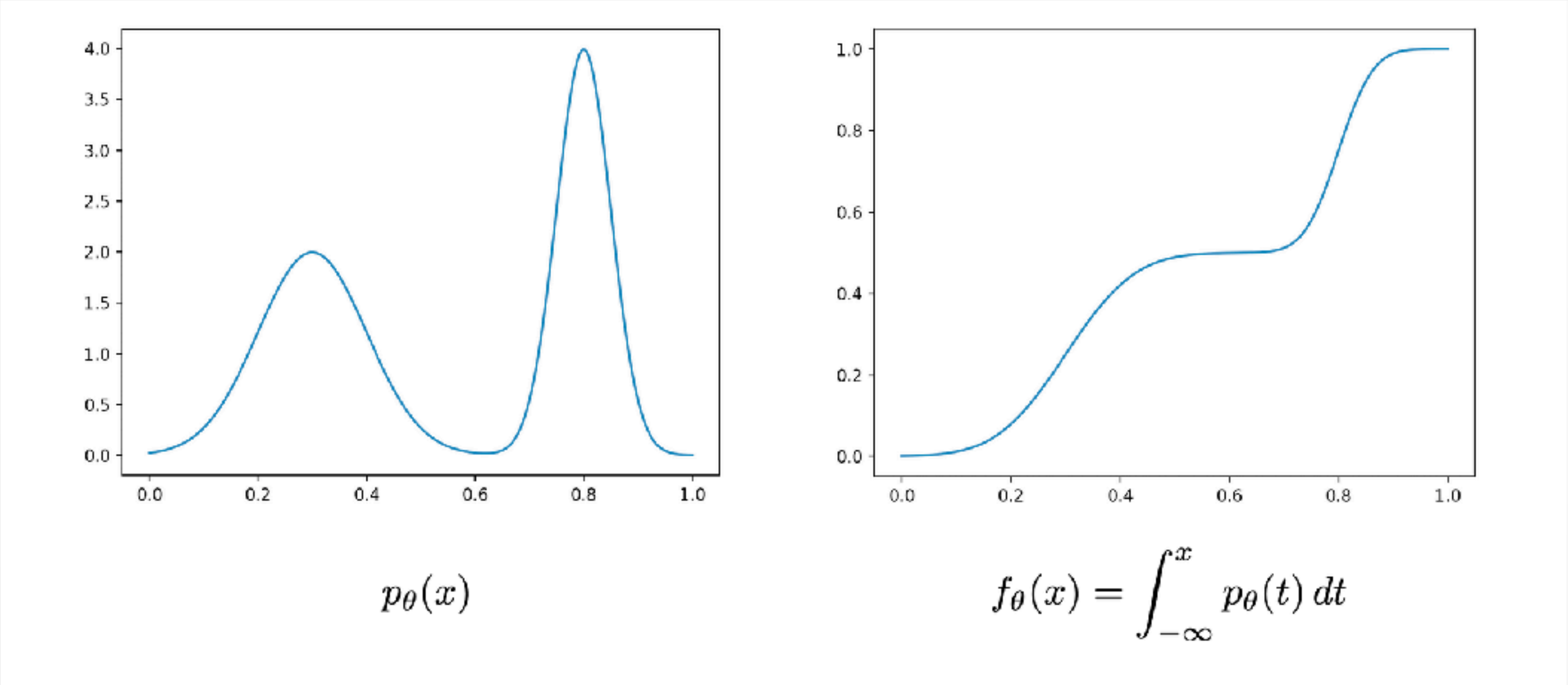
INVERTIBLE TRANSFORM

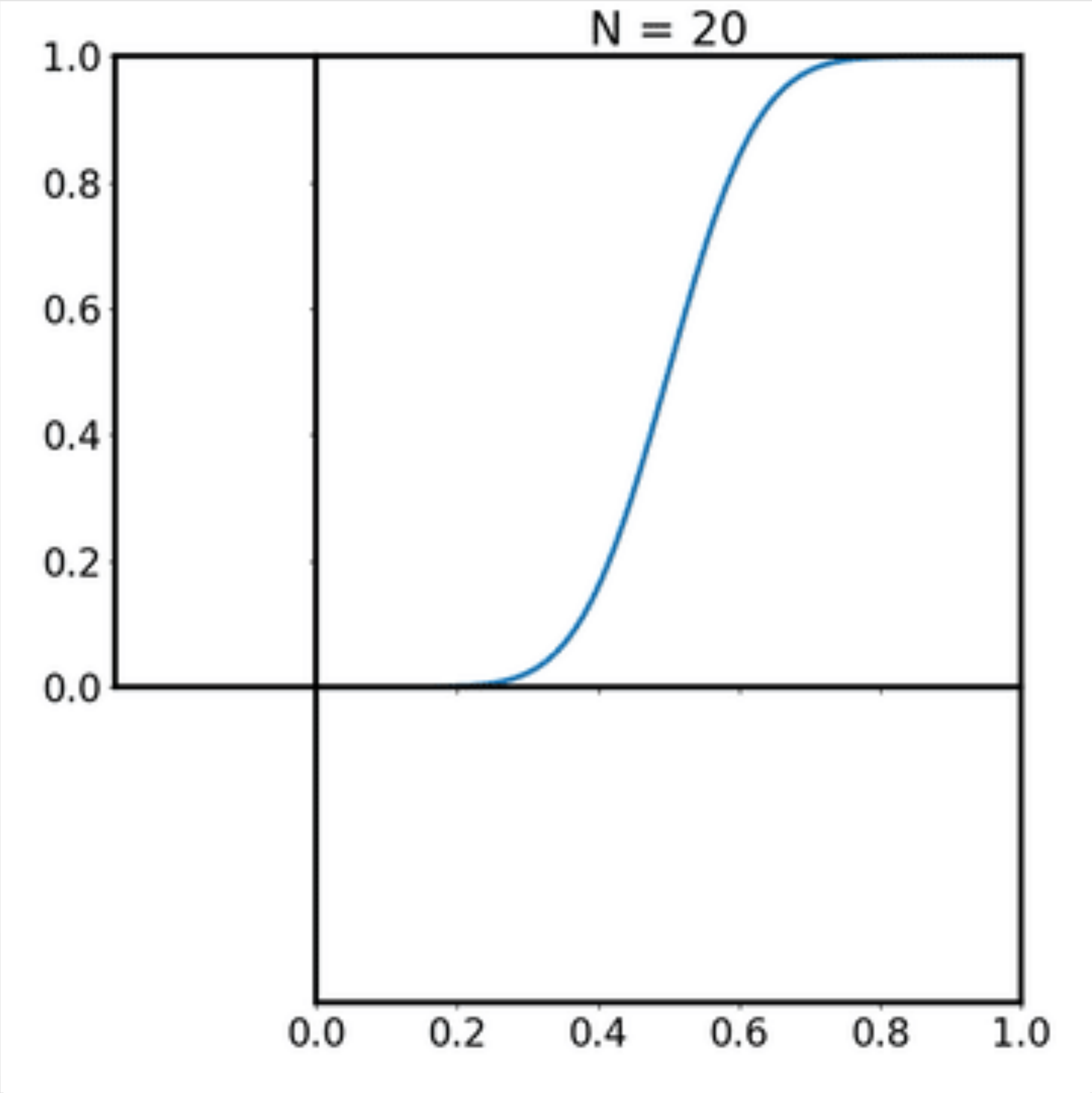
If x is a continuous random variable with CDF f(x),
then the random variable y = f(x) has a uniform distribution on [0, 1].



INVERTIBLE TRANSFORM

If x is a continuous random variable with CDF f(x),
then the random variable y = f(x) has a uniform distribution on [0, 1].





INVERTIBLE TRANSFORM

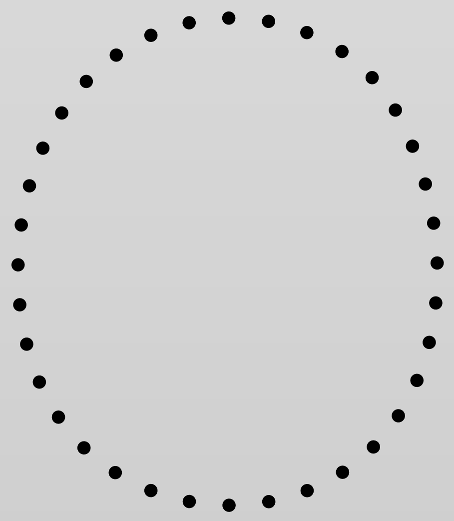
The basic idea:

INVERTIBLE TRANSFORM

The basic idea:

1. we have a simple random generator;

$$z \sim f_Z(z)$$



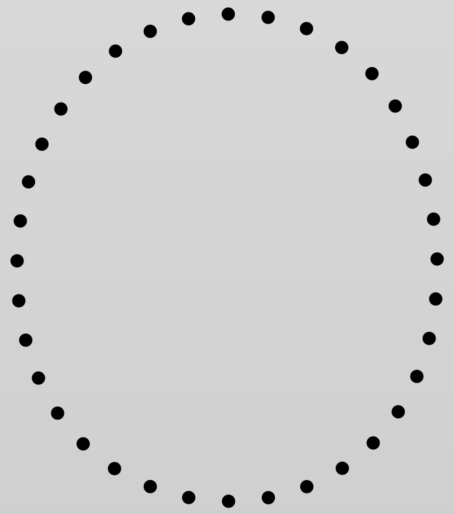
For example: $z \sim \mathcal{N}(0, 1)$

INVERTIBLE TRANSFORM

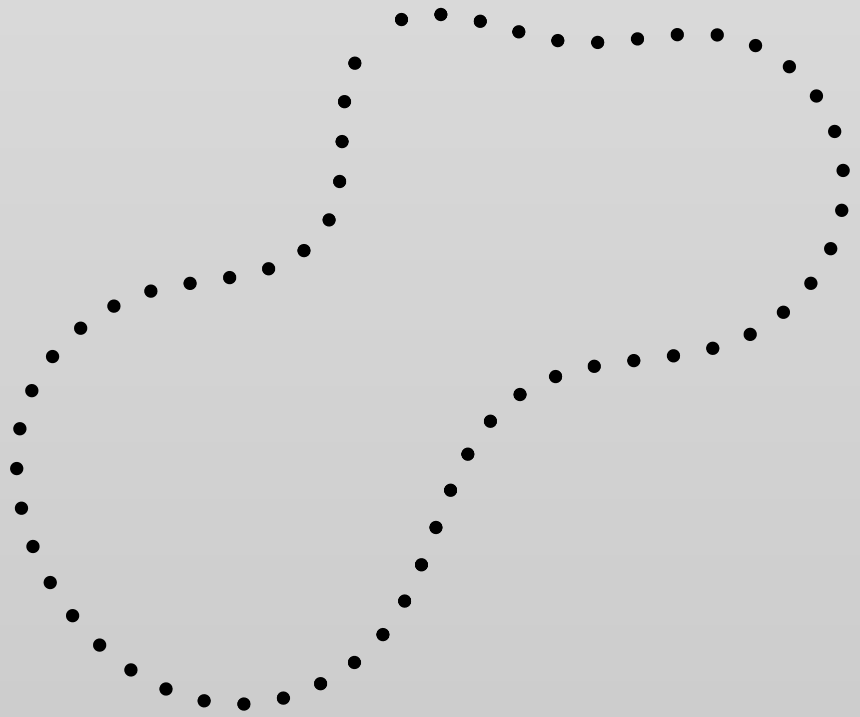
The basic idea:

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;

$z \sim f_Z(z)$



$y \sim f_Y(y)$

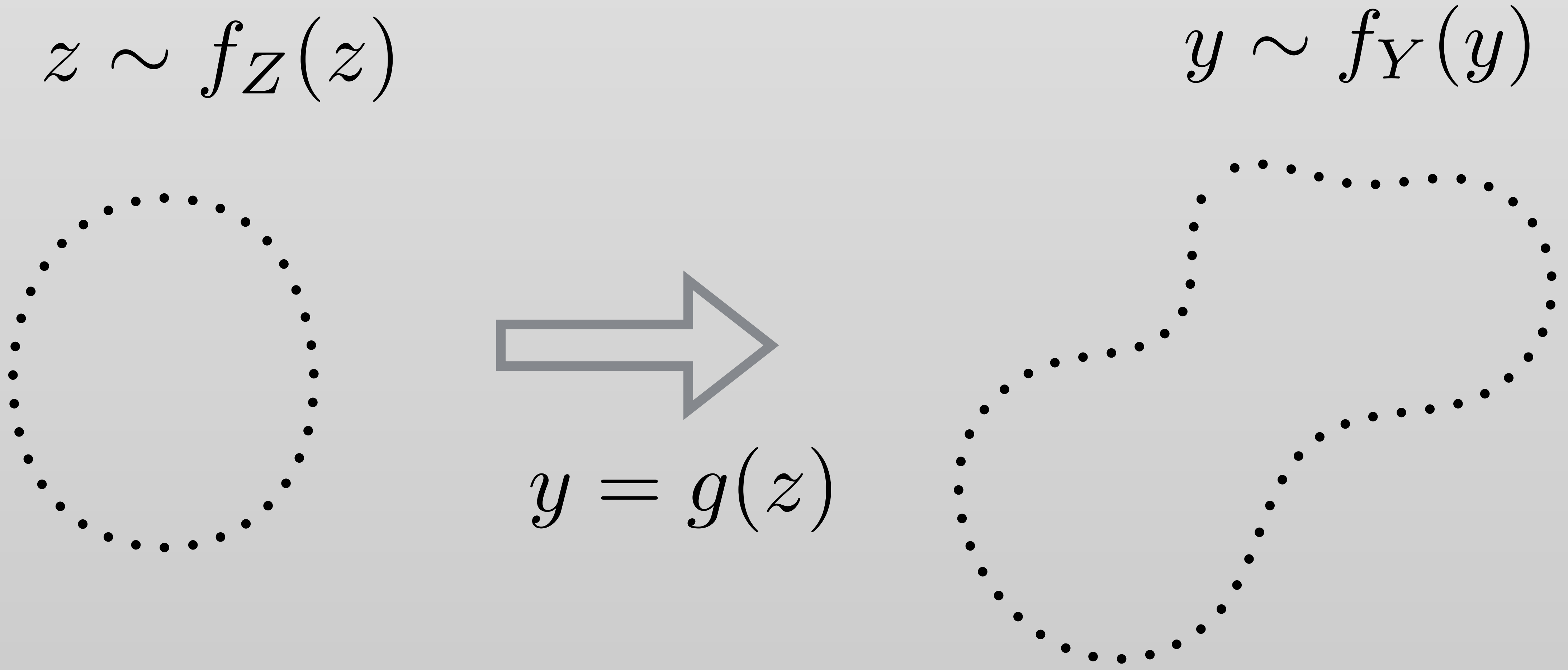


For example: $z \sim \mathcal{N}(0, 1)$

INVERTIBLE TRANSFORM

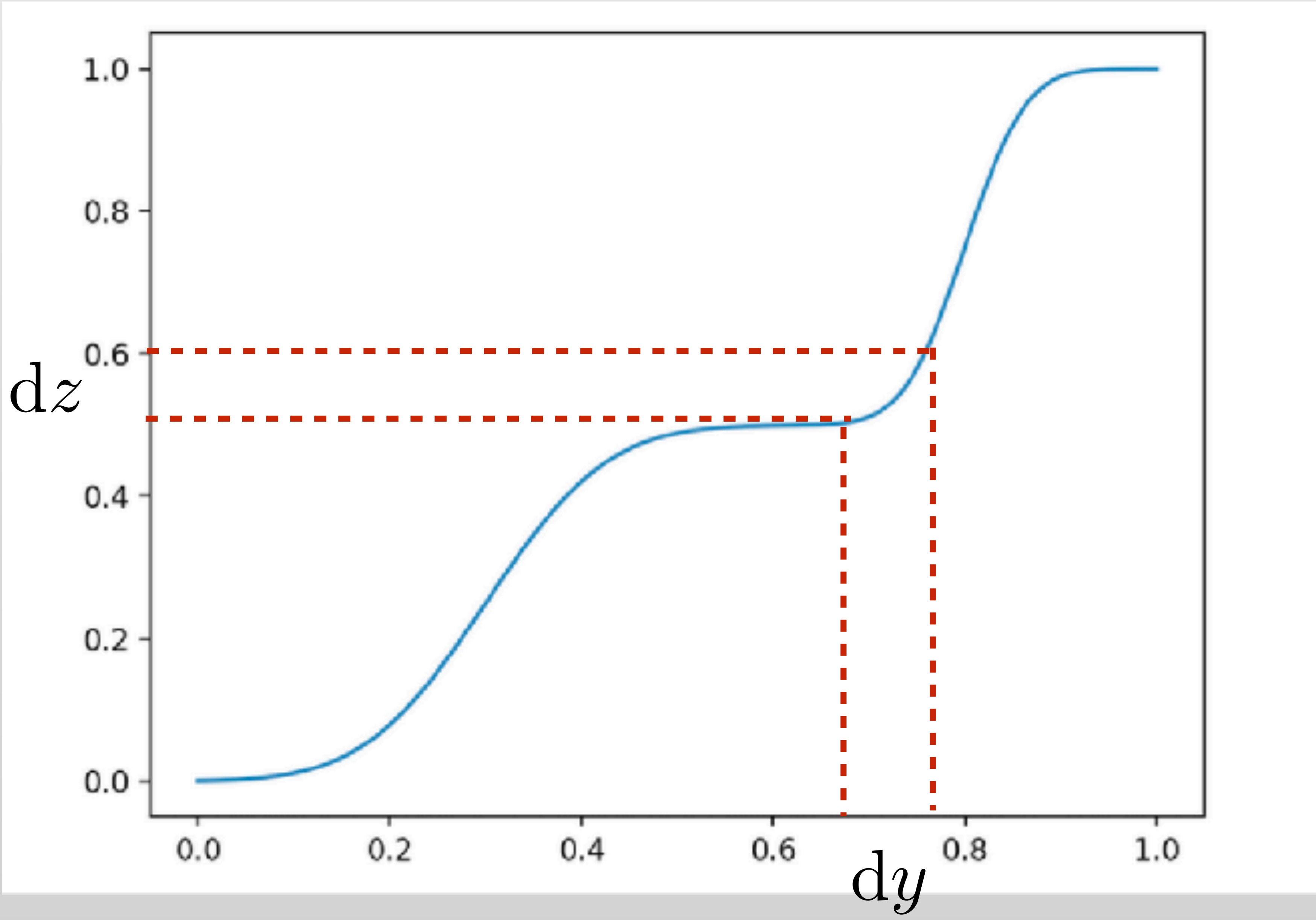
The basic idea:

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.



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

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

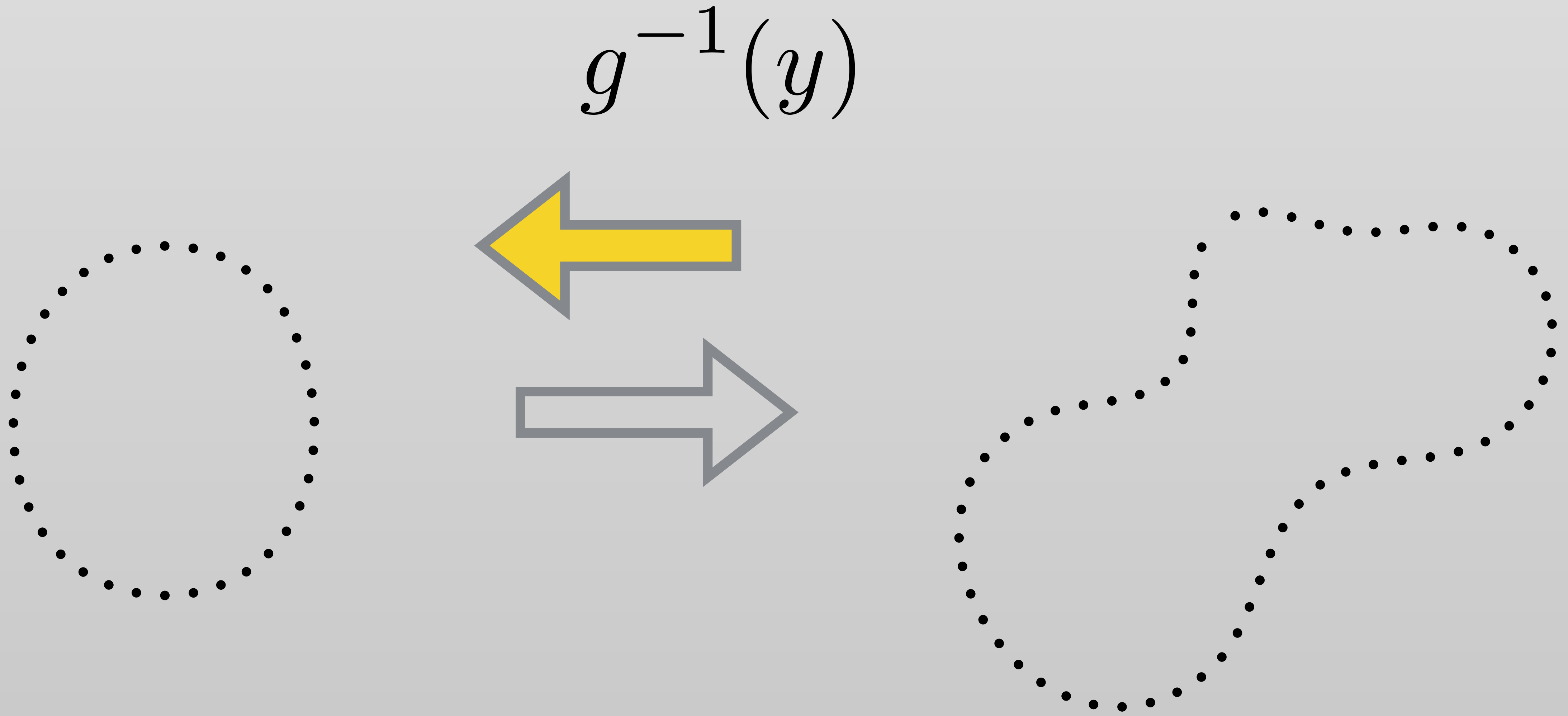
$$= \frac{\mathrm{d}}{\mathrm{d}y} F_Z(g^{-1}(y))$$

Chain rule

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

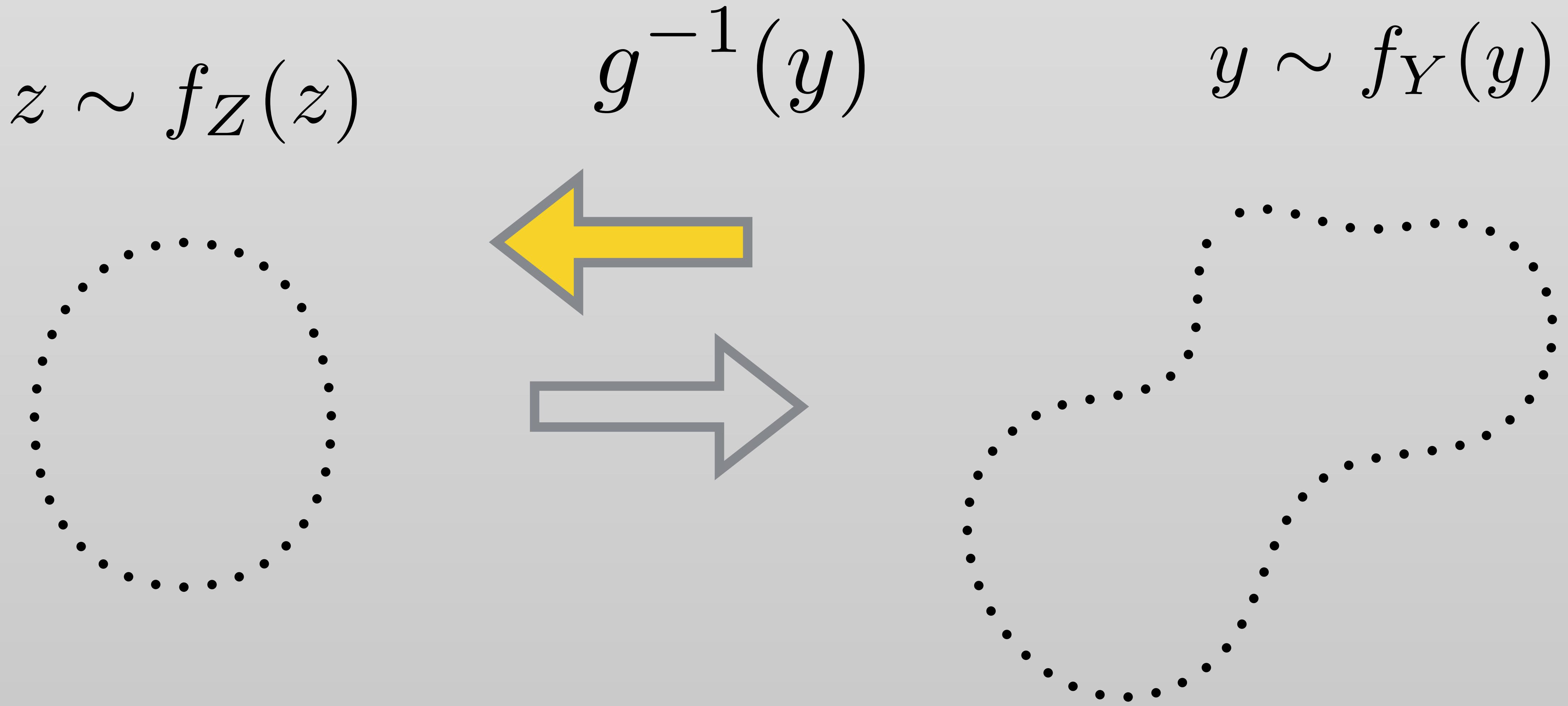
Multidimensional case

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



CHANGE OF VARIABLE

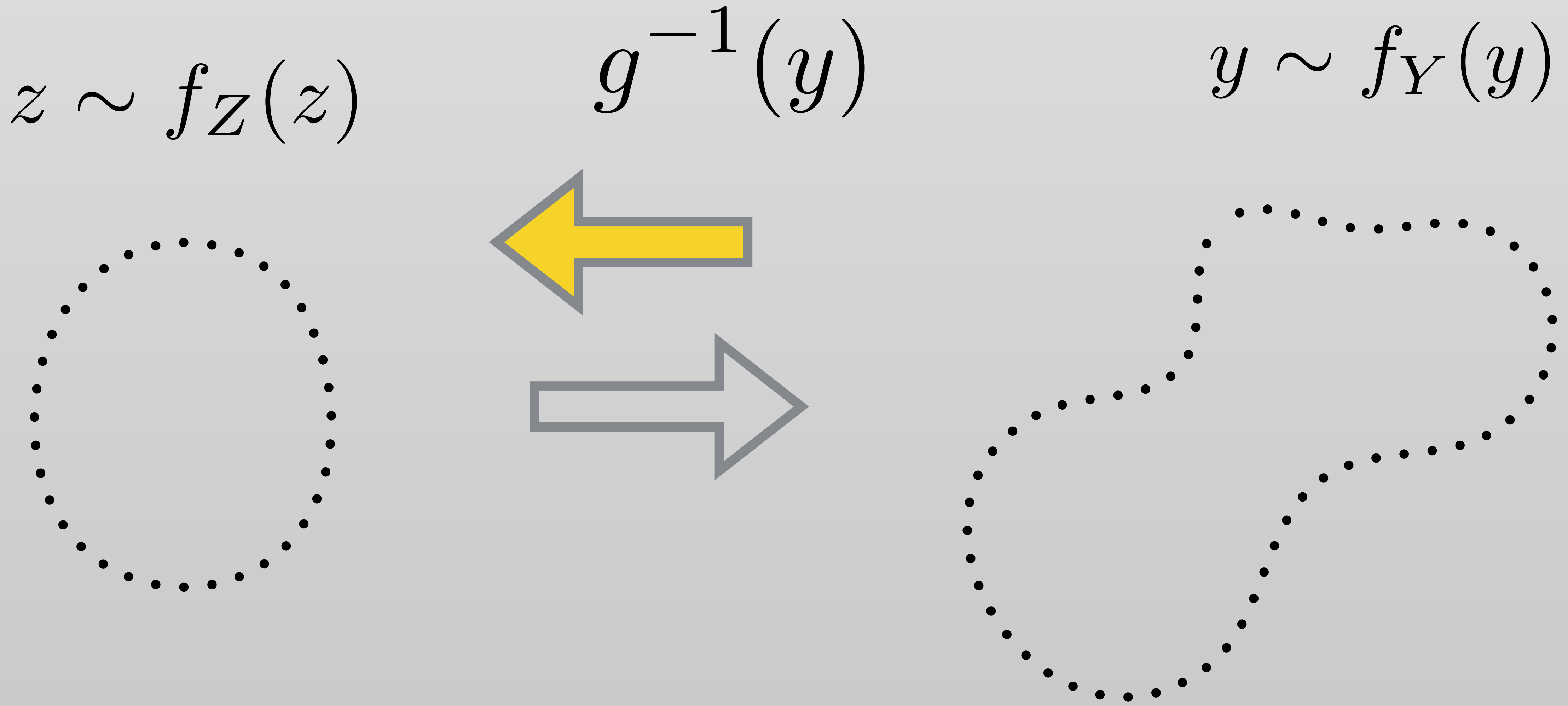
$$\log[f_Y(y)] = \log[f_Z(g^{-1}(y))] + \log\left[\left|\det\frac{\partial g^{-1}(y)}{\partial y}\right|\right]$$



CHANGE OF VARIABLE

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

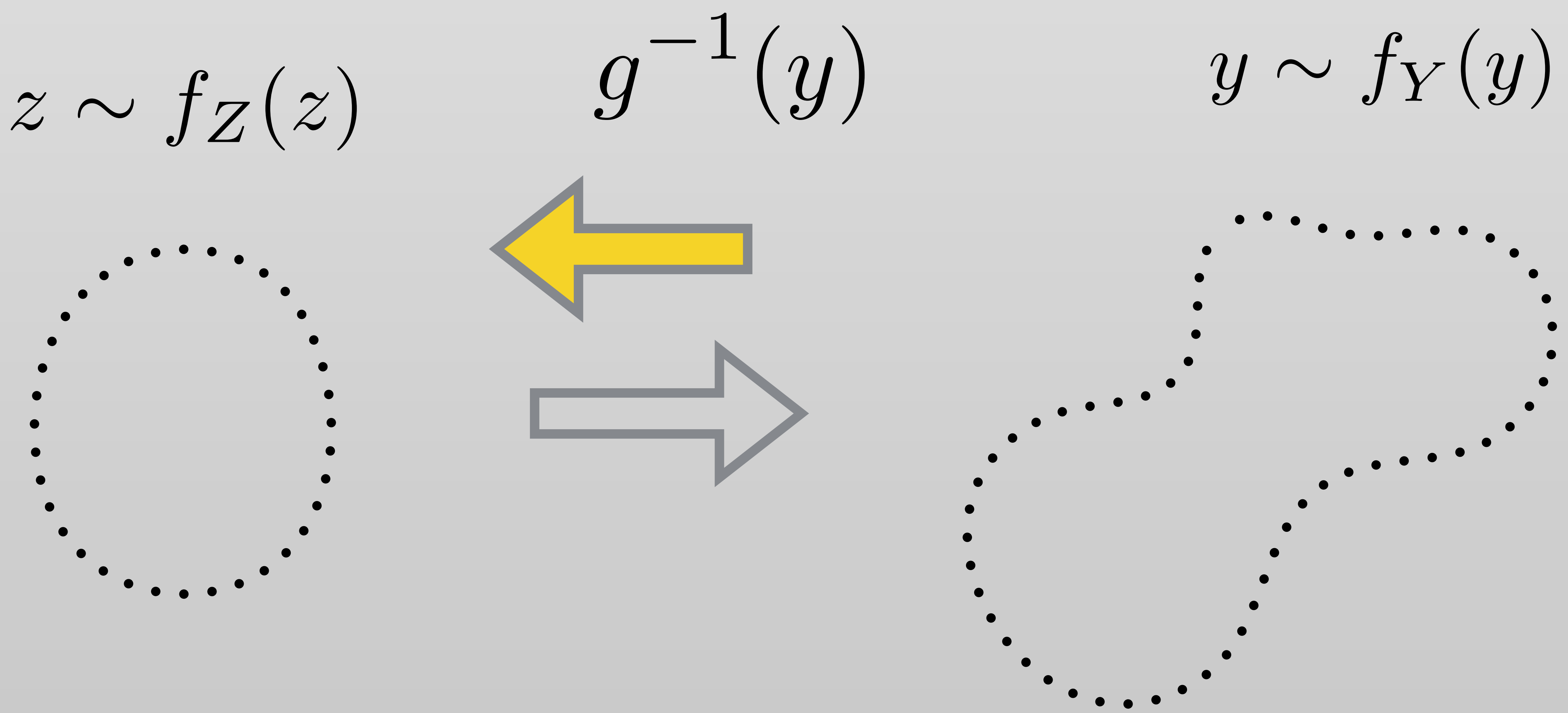
1. $g(y)$ has to be a bijection



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



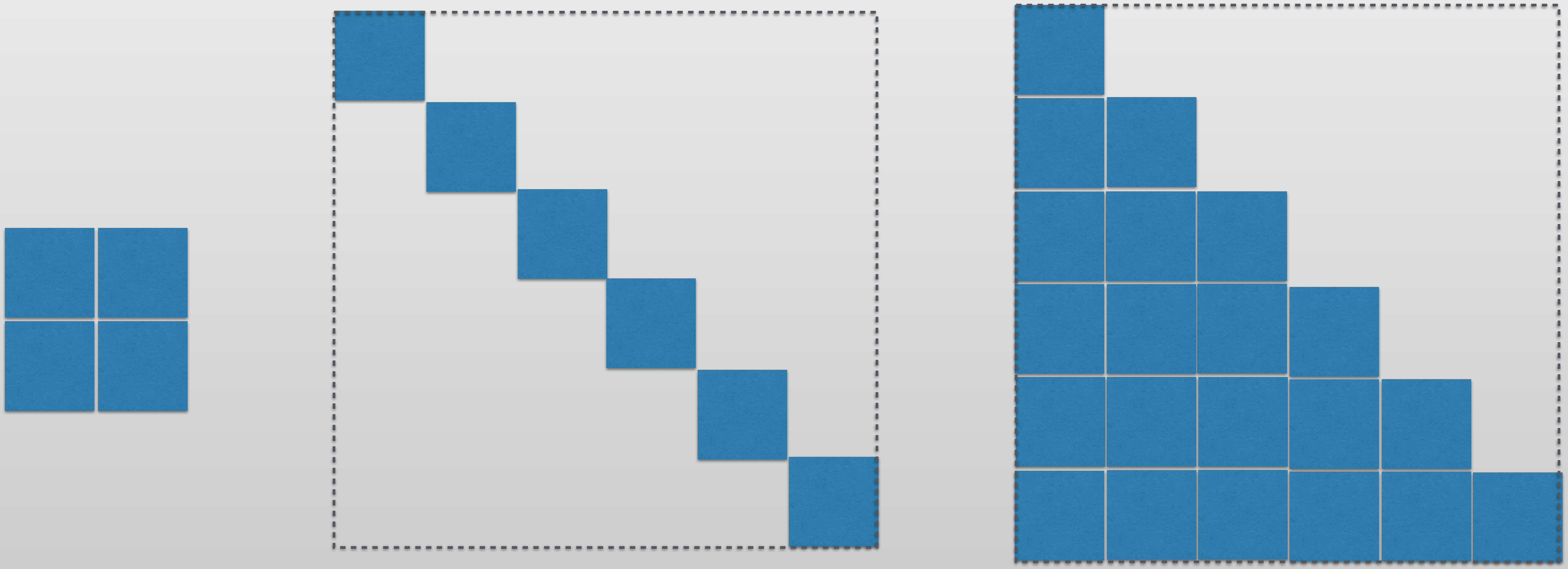
JACOBIAN

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

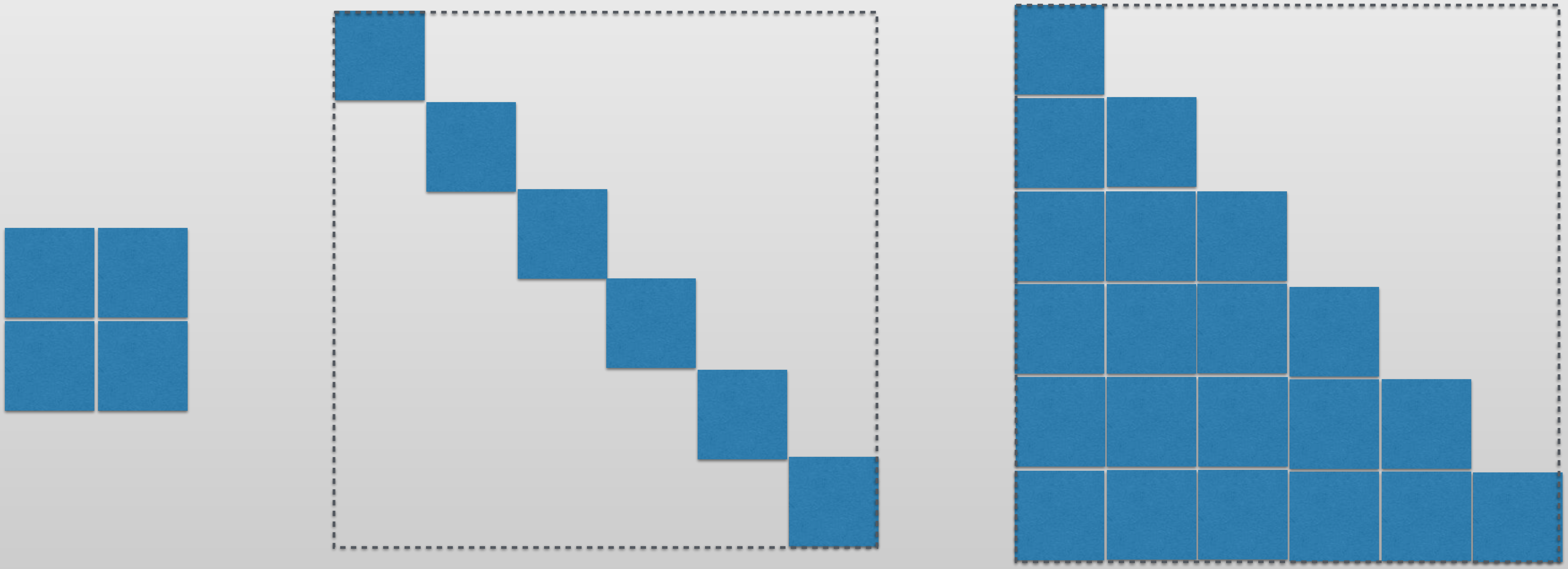
The calculation of determinant Jacobian will take $O(n^3)$

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:

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

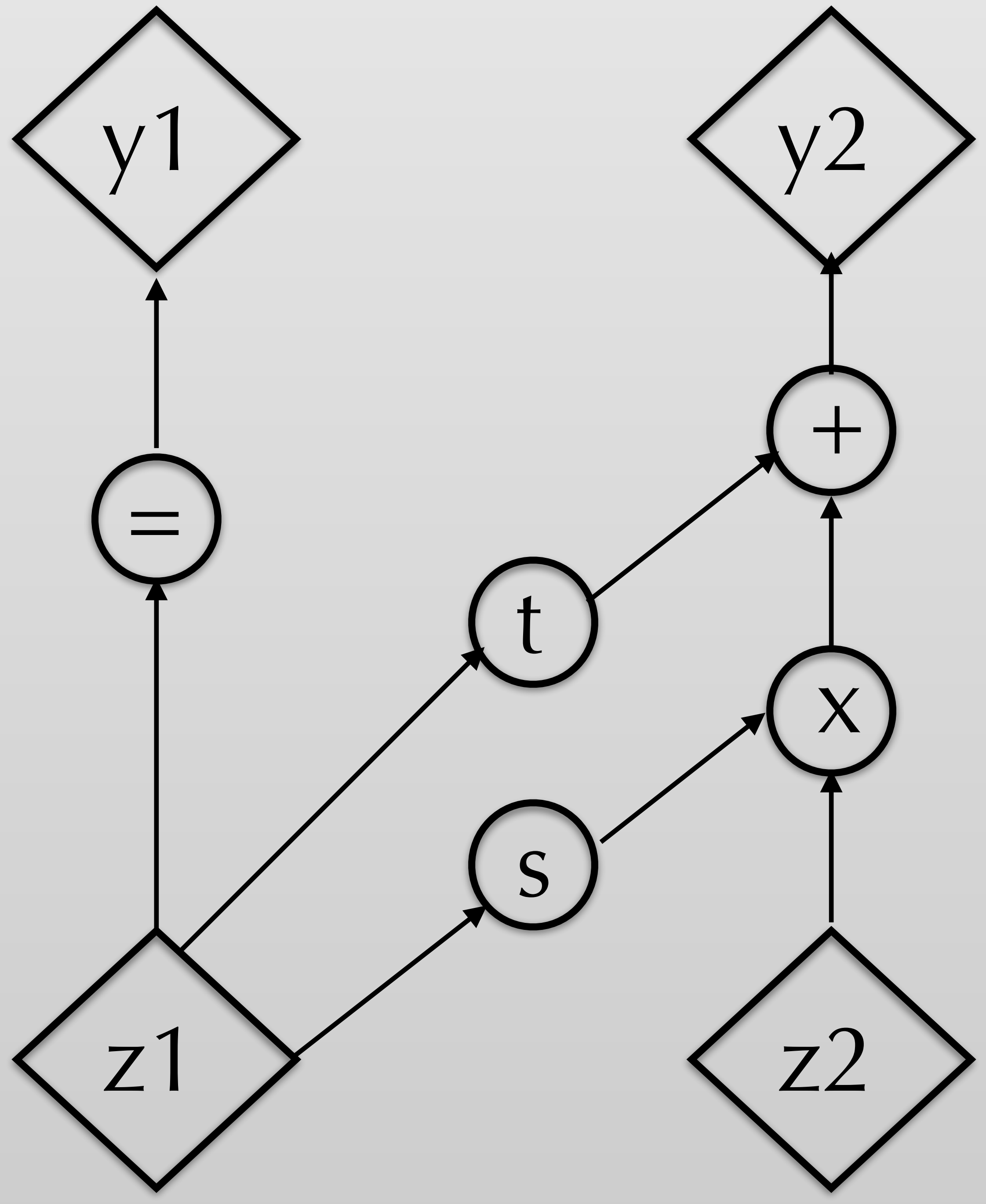
Invertibility for $\alpha_i \neq 0$

log-Jacobian becomes

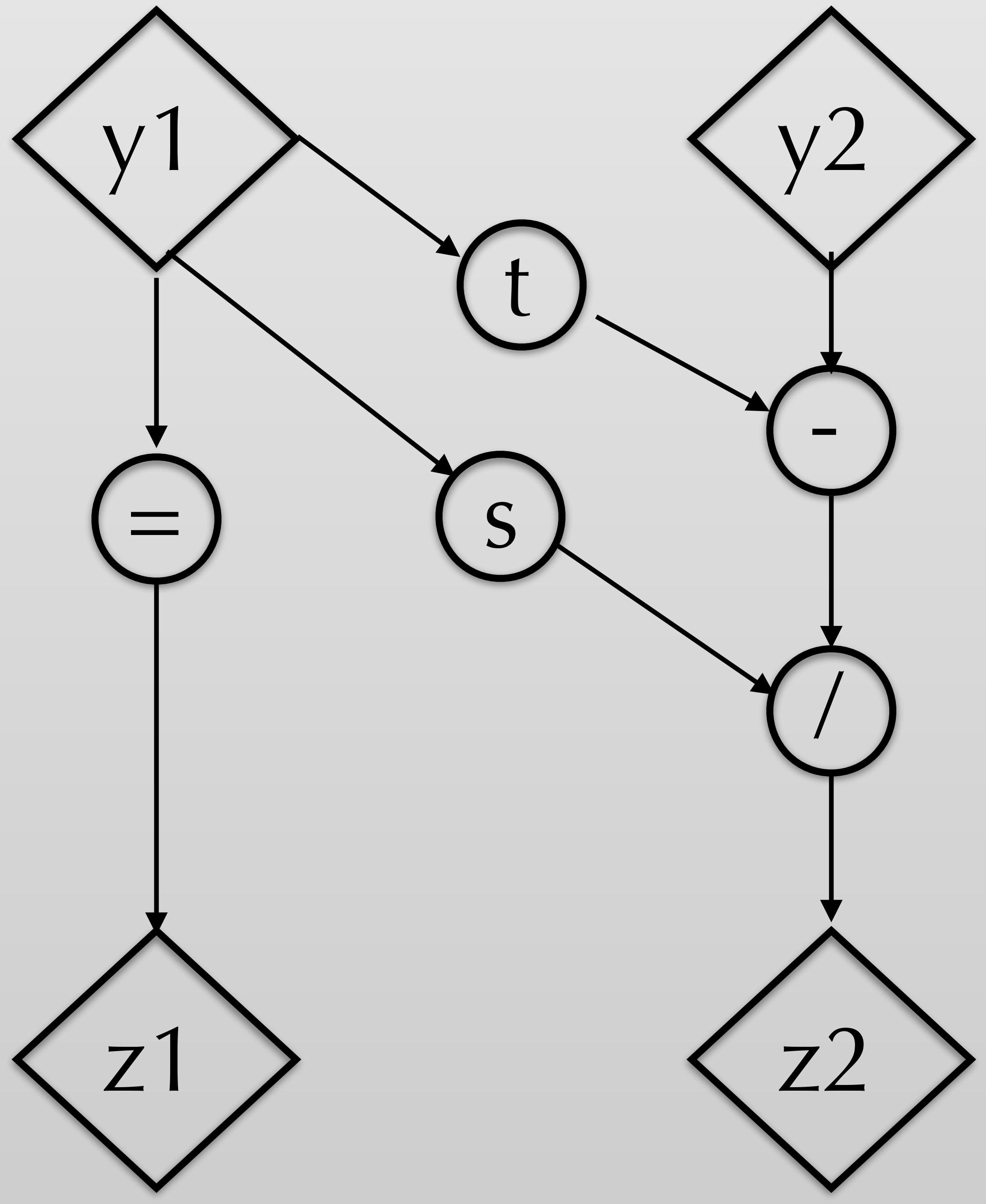
$$\log|\det J_{g^{-1}}(\mathbf{z})| = \sum_{i=1}^{N} \log|\alpha_i|$$

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

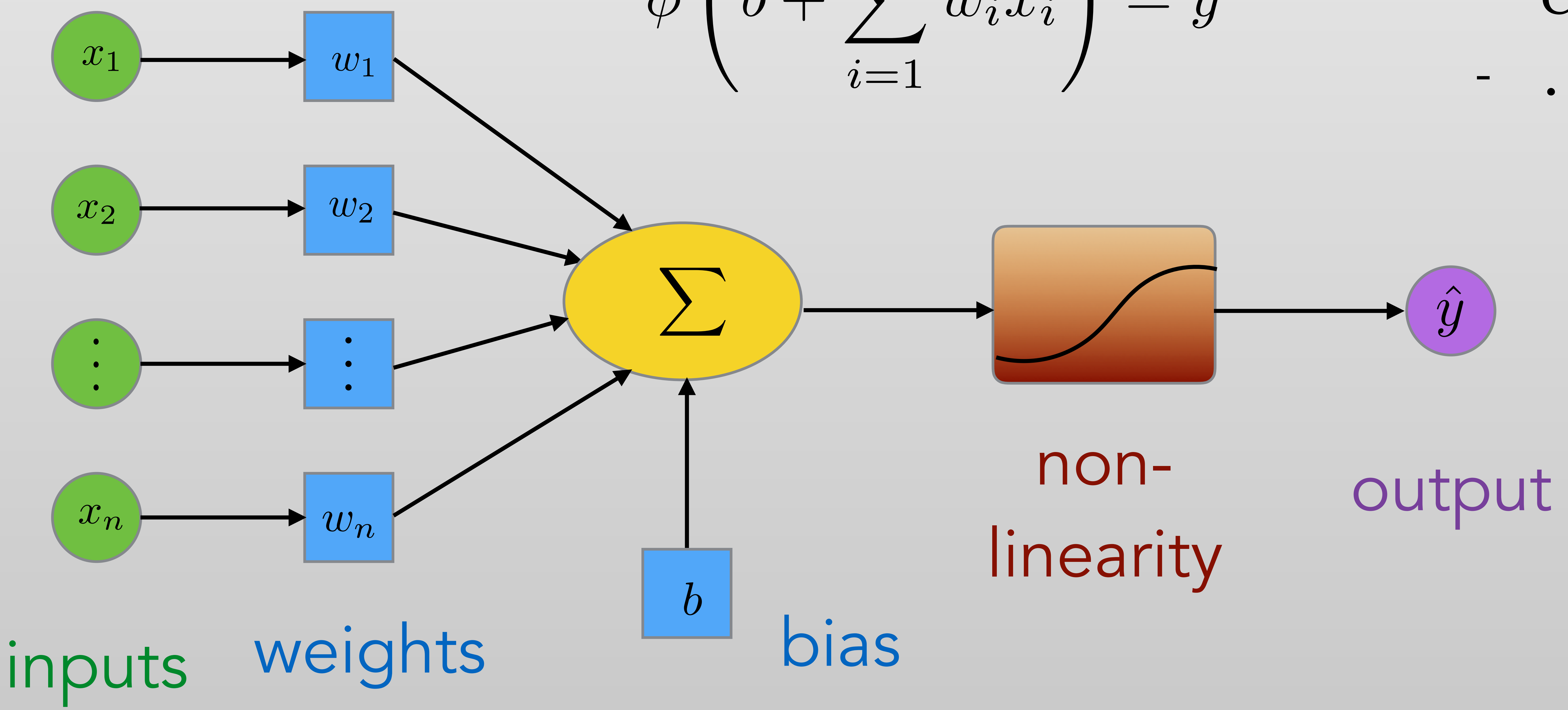
$$\frac{\partial y}{\partial z} = \begin{bmatrix} \mathbf{I}_d & 0 \\ \frac{\partial y_{d+1:D}}{\partial z_{1:d}} & \text{diag}(\exp[s(z_{1:d})]) \end{bmatrix}$$

What is functions t and s?

PARAMETERISATION WITH THE NN

Neuron

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

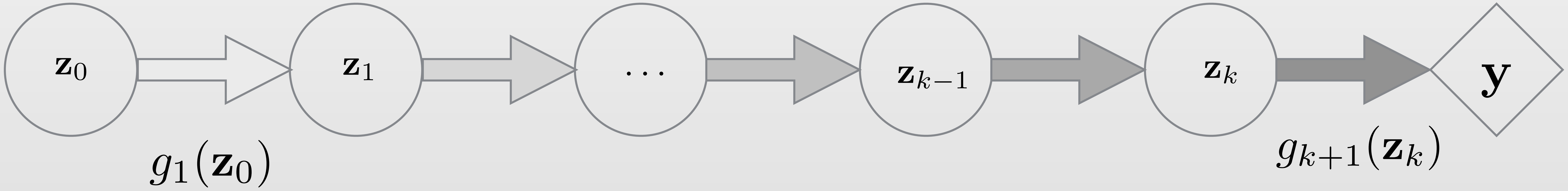


The architecture can be any:

- fully connected
- residual network
- convolutional network
- …

COMPOSING FLOWS

$\mathbf{z}_0 \sim f_{Z_0}(\mathbf{z}_0)$



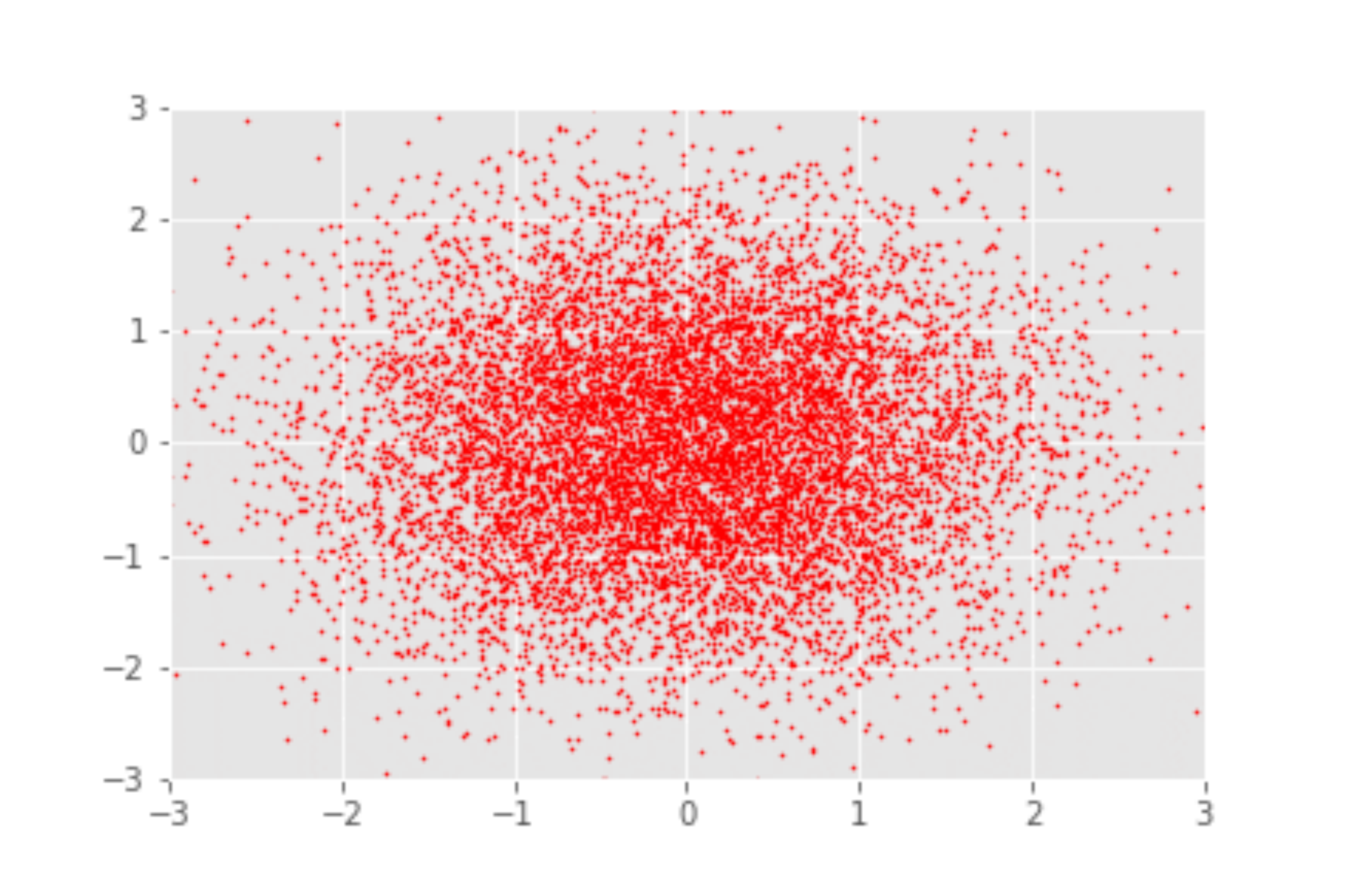
Function composition

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

Jacobian composition

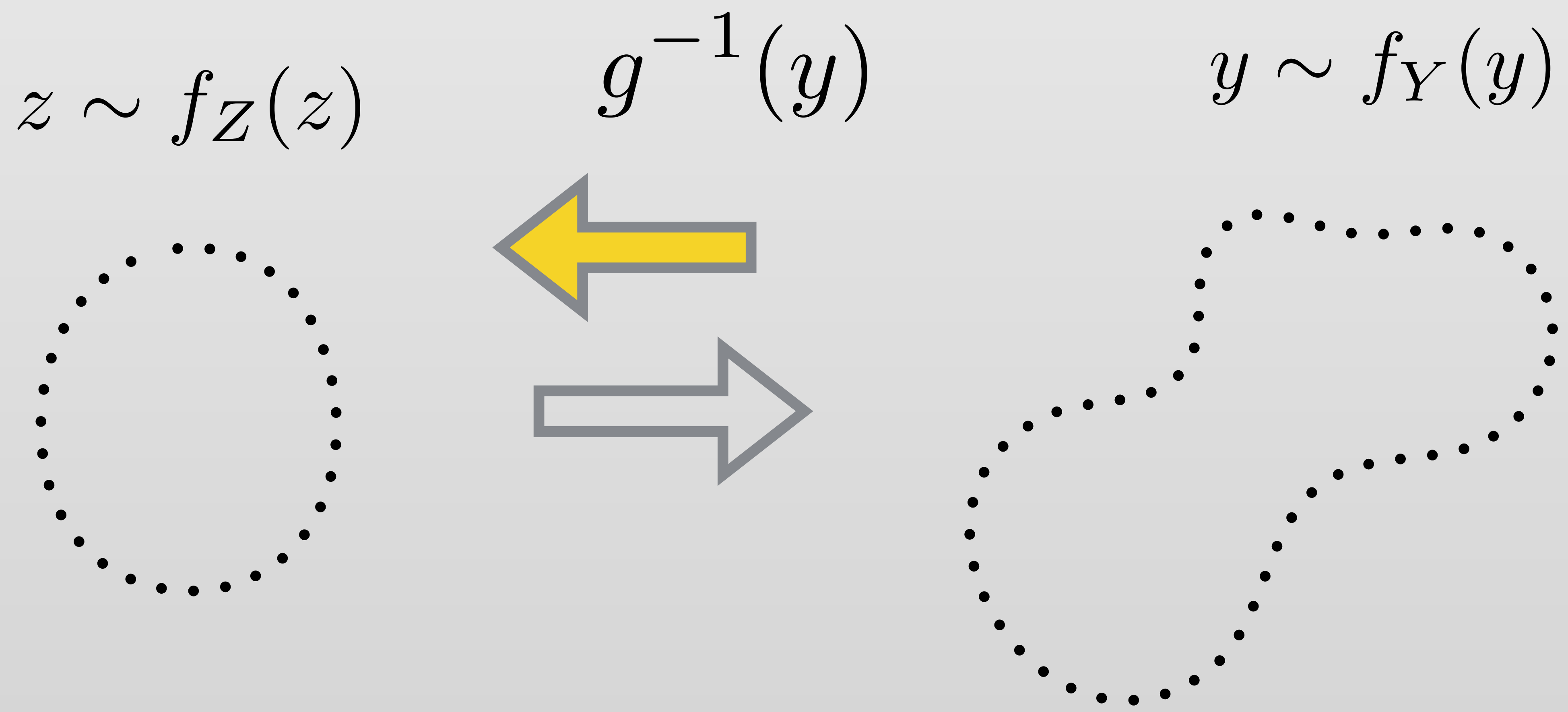
$$\det(J_1 \cdot J_2) = \det(J_1) \cdot \det(J_2)$$

FLOW



github.com/papercup-open-source/
tutorials

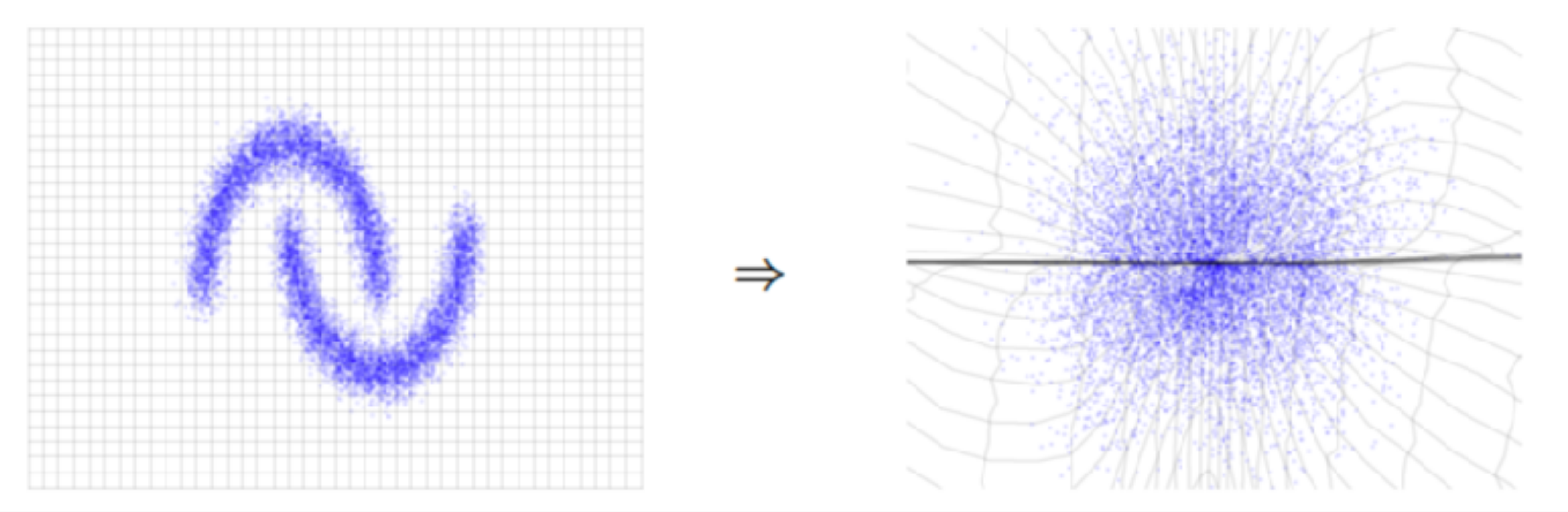
SAMPLE GENERATION



SAMPLE GENERATION

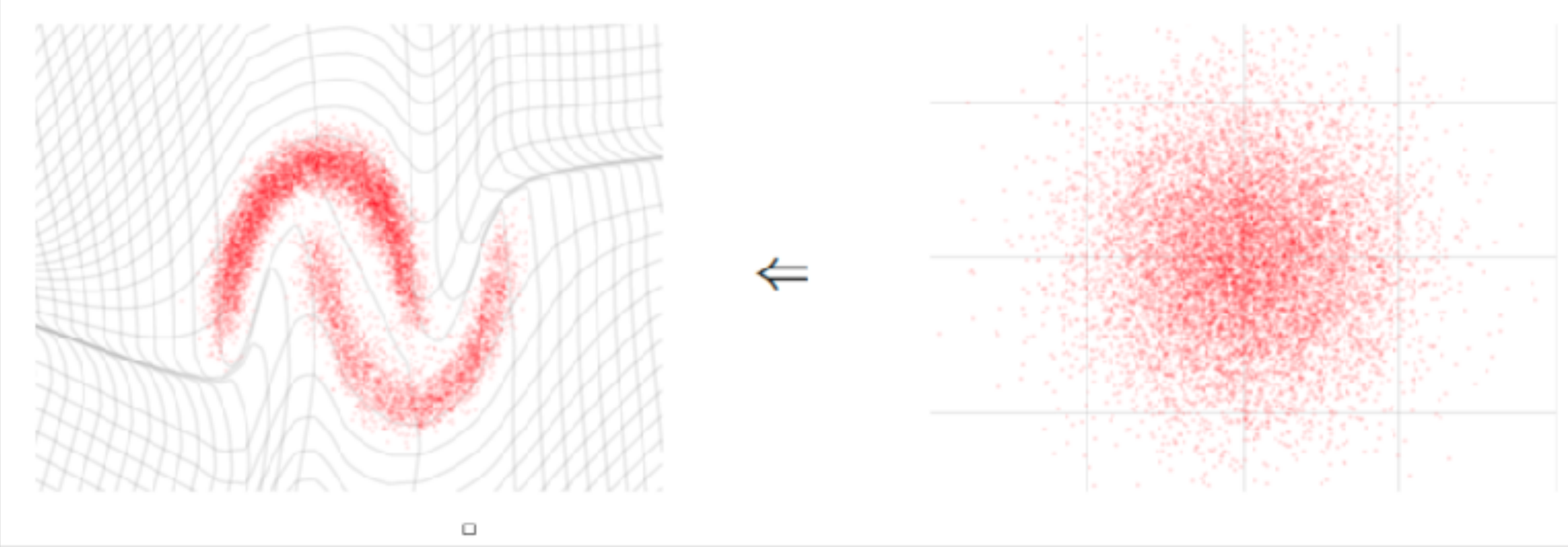
Data space   Latent space

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



Sample generation

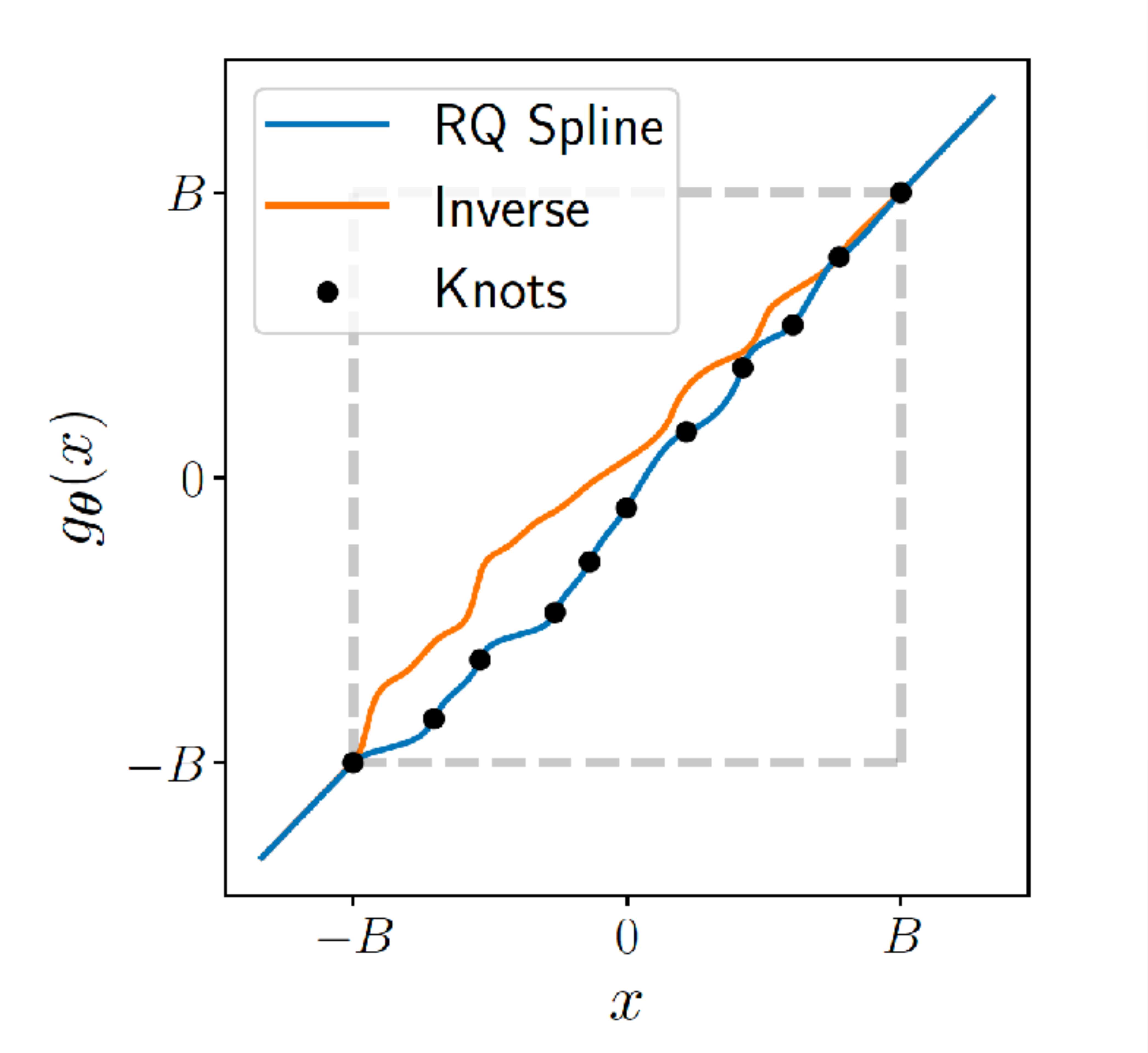
$$z \sim f_Z(z)$$
$$y = g(z)$$



Laurent Dinh et al, Density estimation using realNVP

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

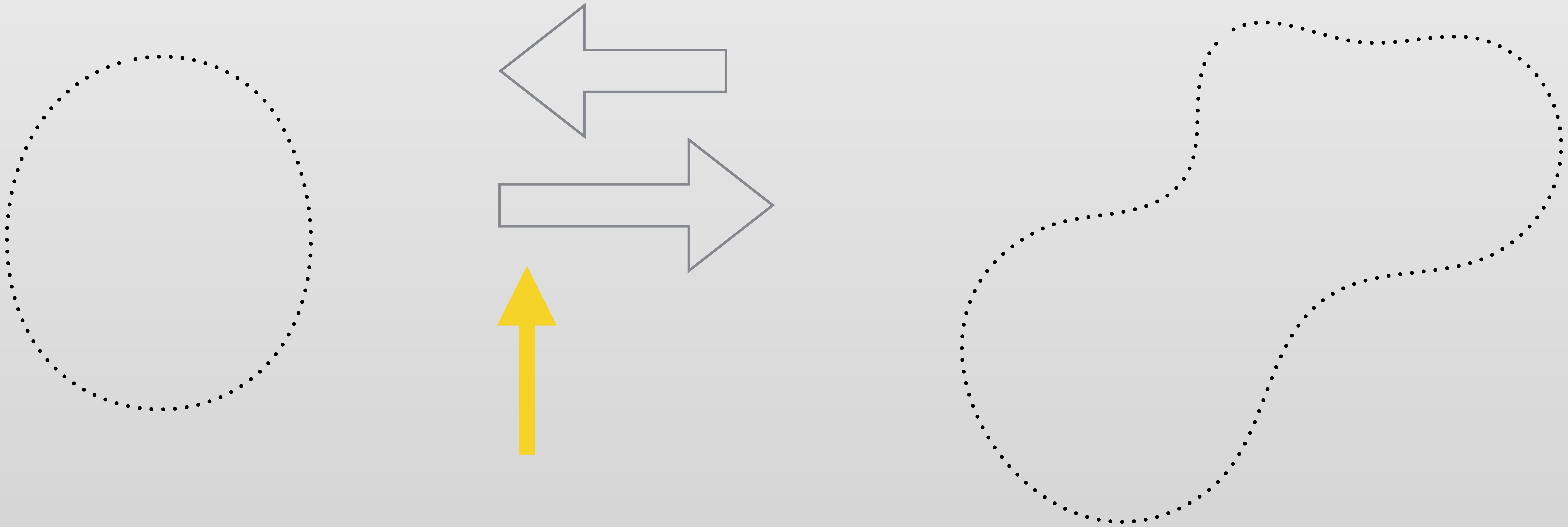
We do not have access to the samples form the posterior,
as in the examples that we have just considered.

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:

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

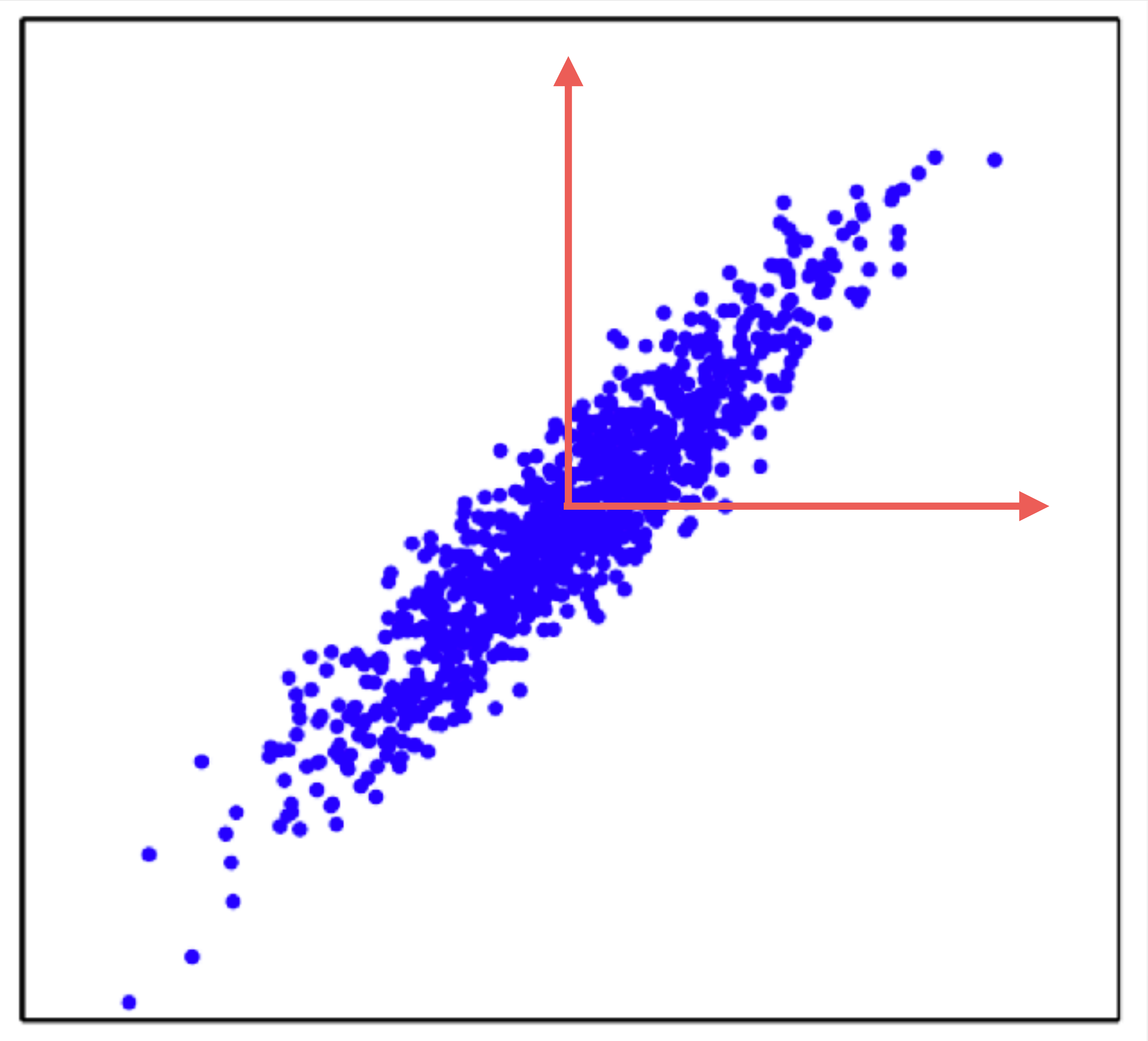
Therefore we have access to the joint sample: $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{x}|\mathbf{y})$

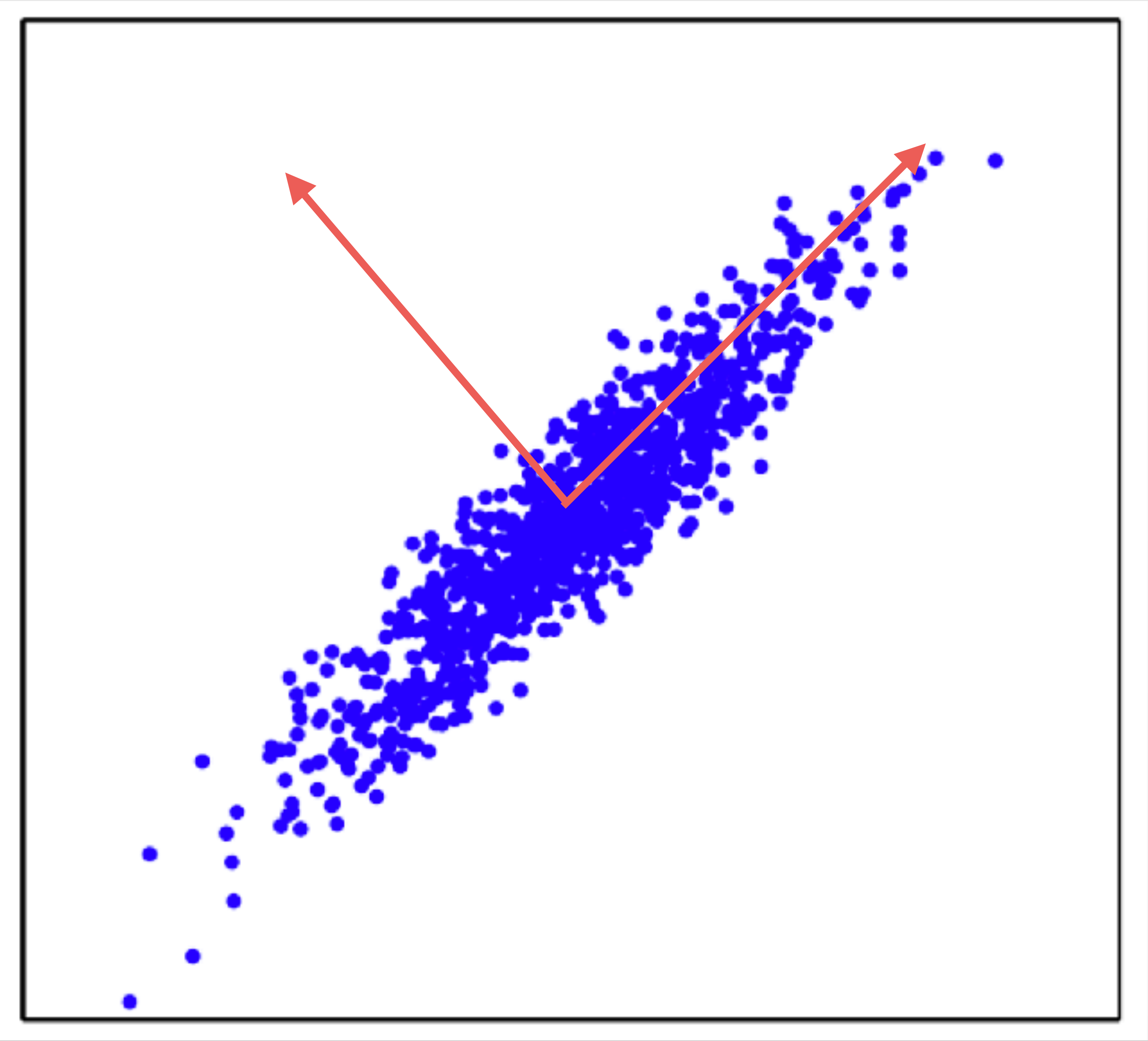
WAVEFORM EMBEDDING

- 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:

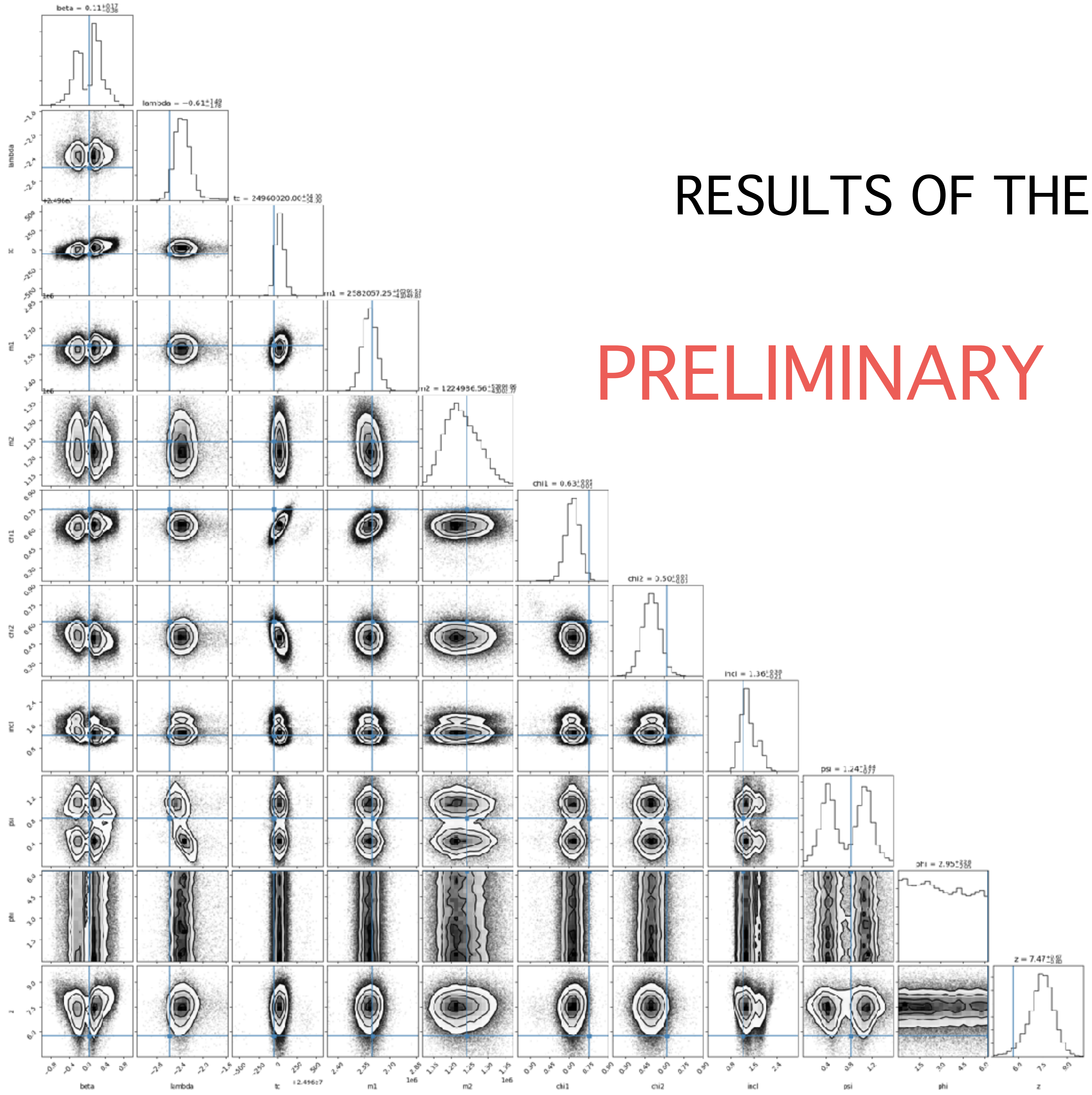
$$v'_{\alpha\mu} = \frac{1}{\sigma_\mu} \sum_{j=1}^{N} h_{\alpha j} u_{\mu j}$$

Exercises:

1. Separation of the signals on the toy spectrograms

2. Simple implementation of the Real NVP flow

RESULTS OF THE PARAMETER ESTIMATION



PRELIMINARY