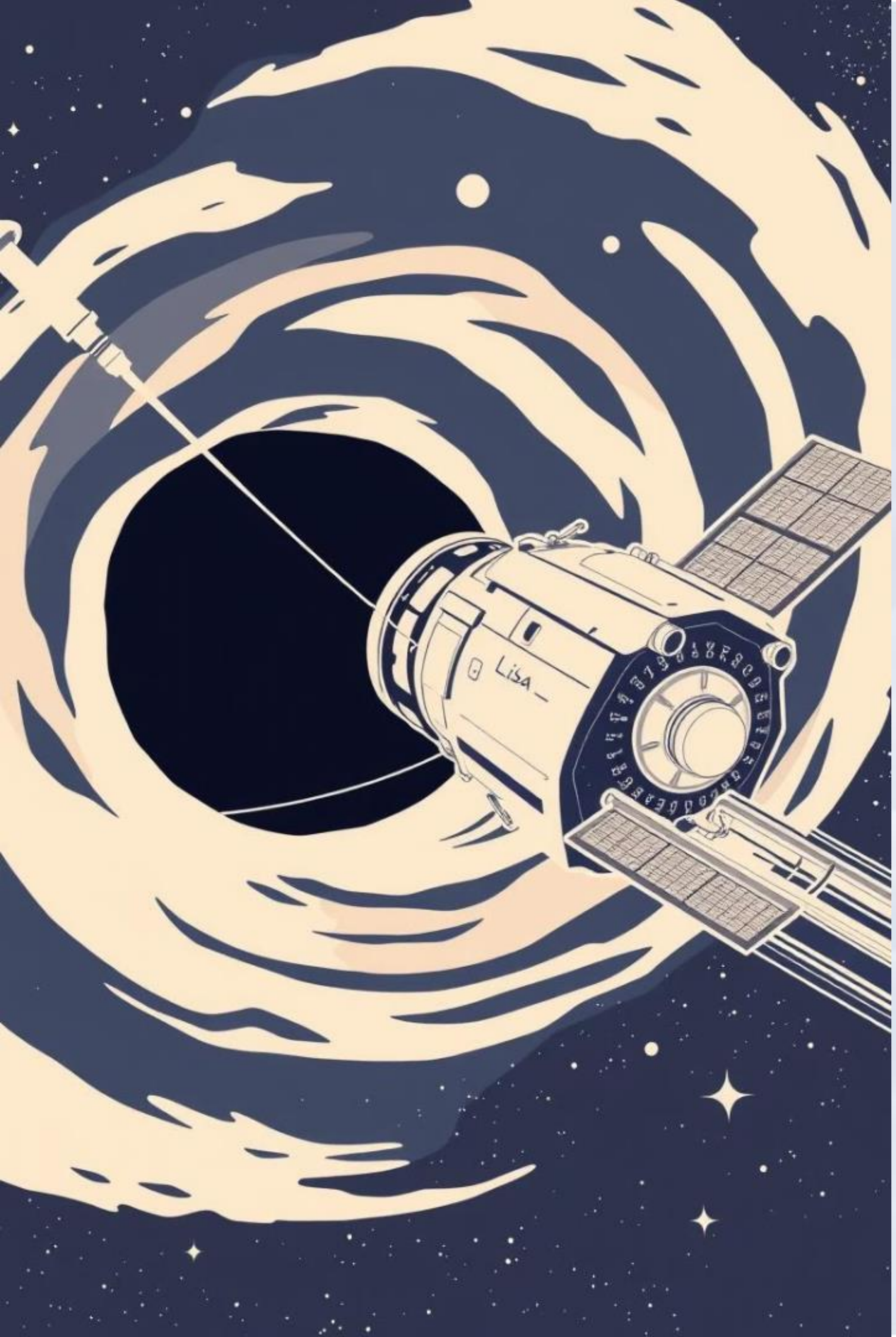


Maxent 2025 in Auckland

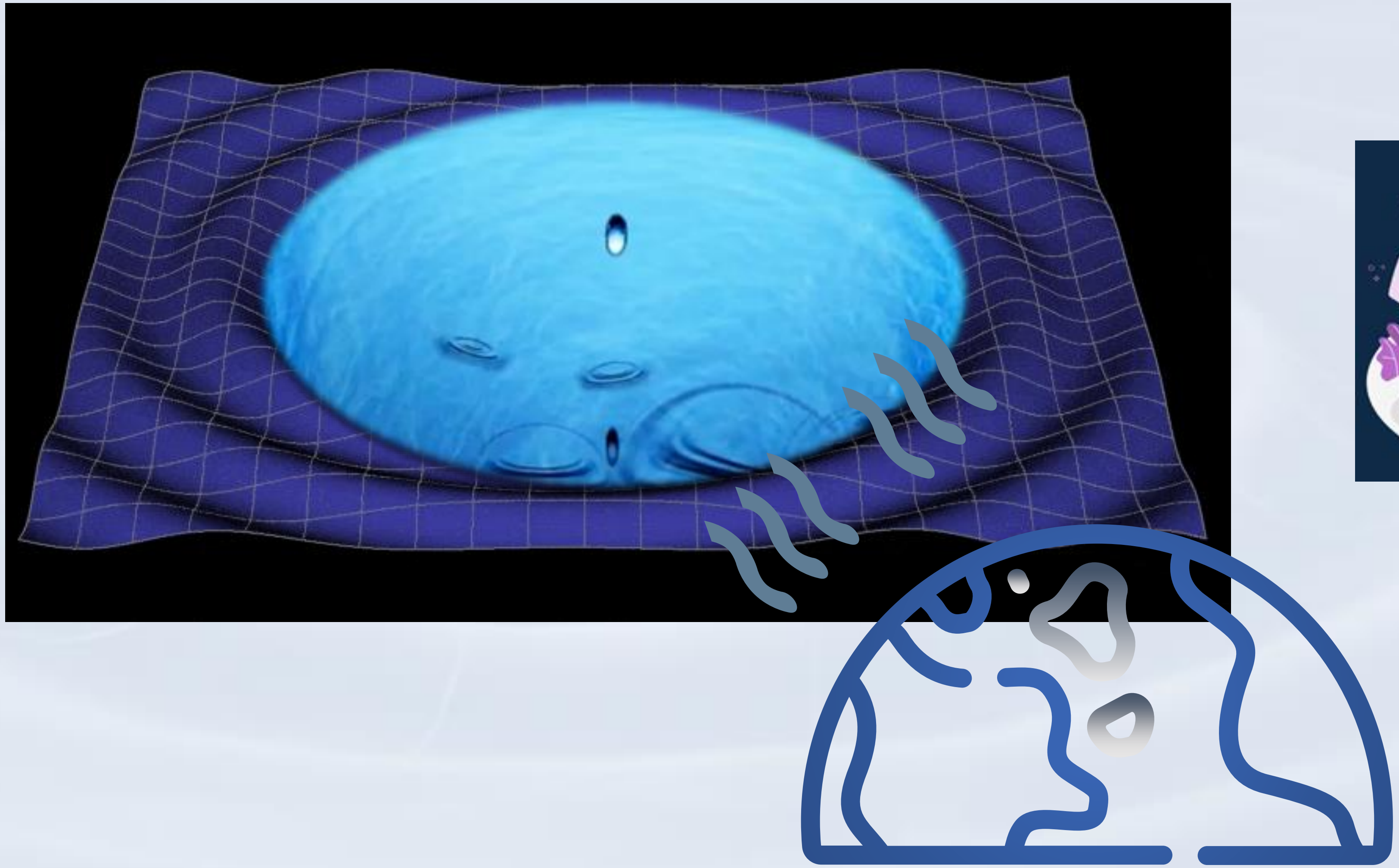
Inference from Imperfection:

Rapid Gravitational Wave Parameter Estimation with Data Gaps in LISA using conditional Flow Matching

Ruiting(Mica) Mao
with Kate Lee and Matt Edwards
University of Auckland

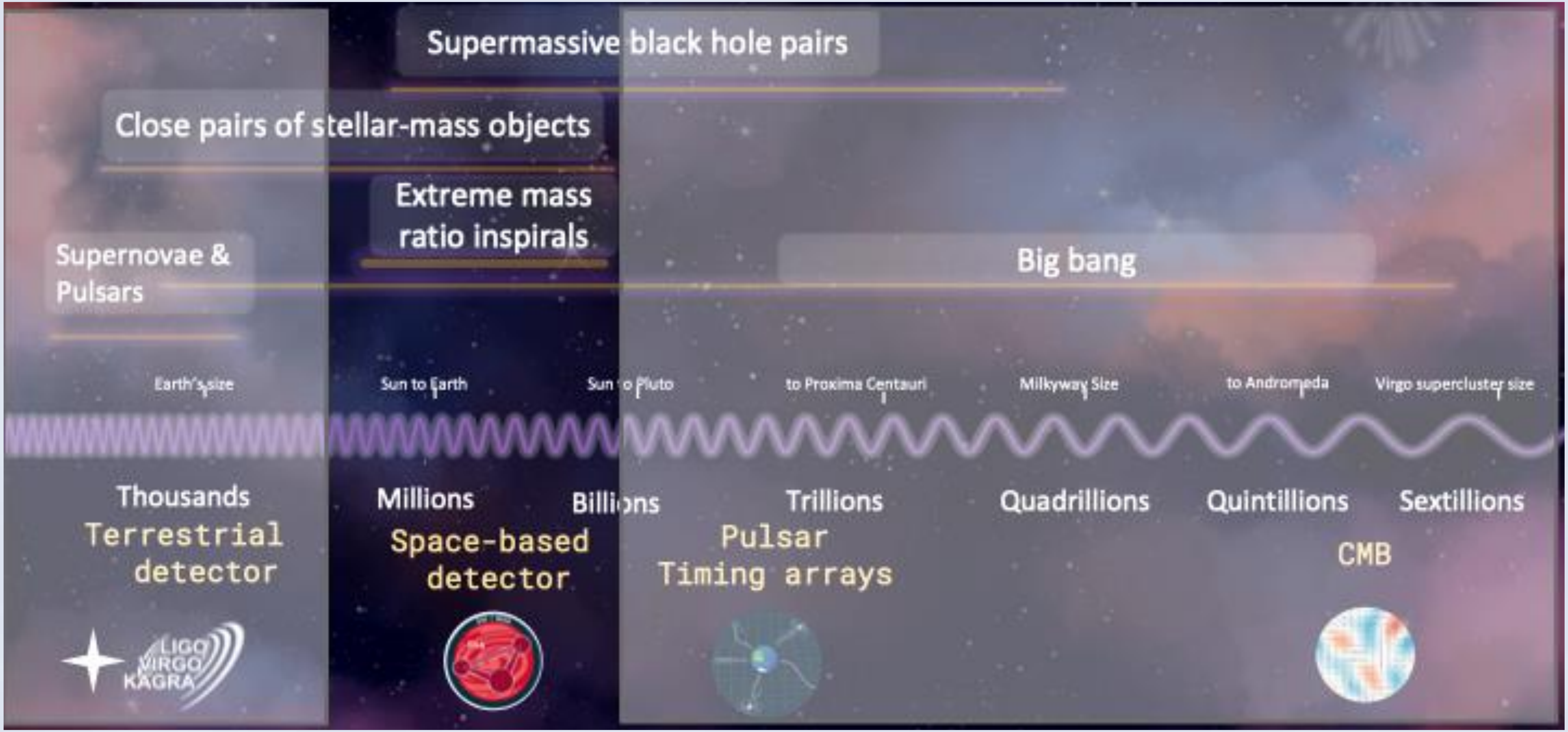
Dec 2025

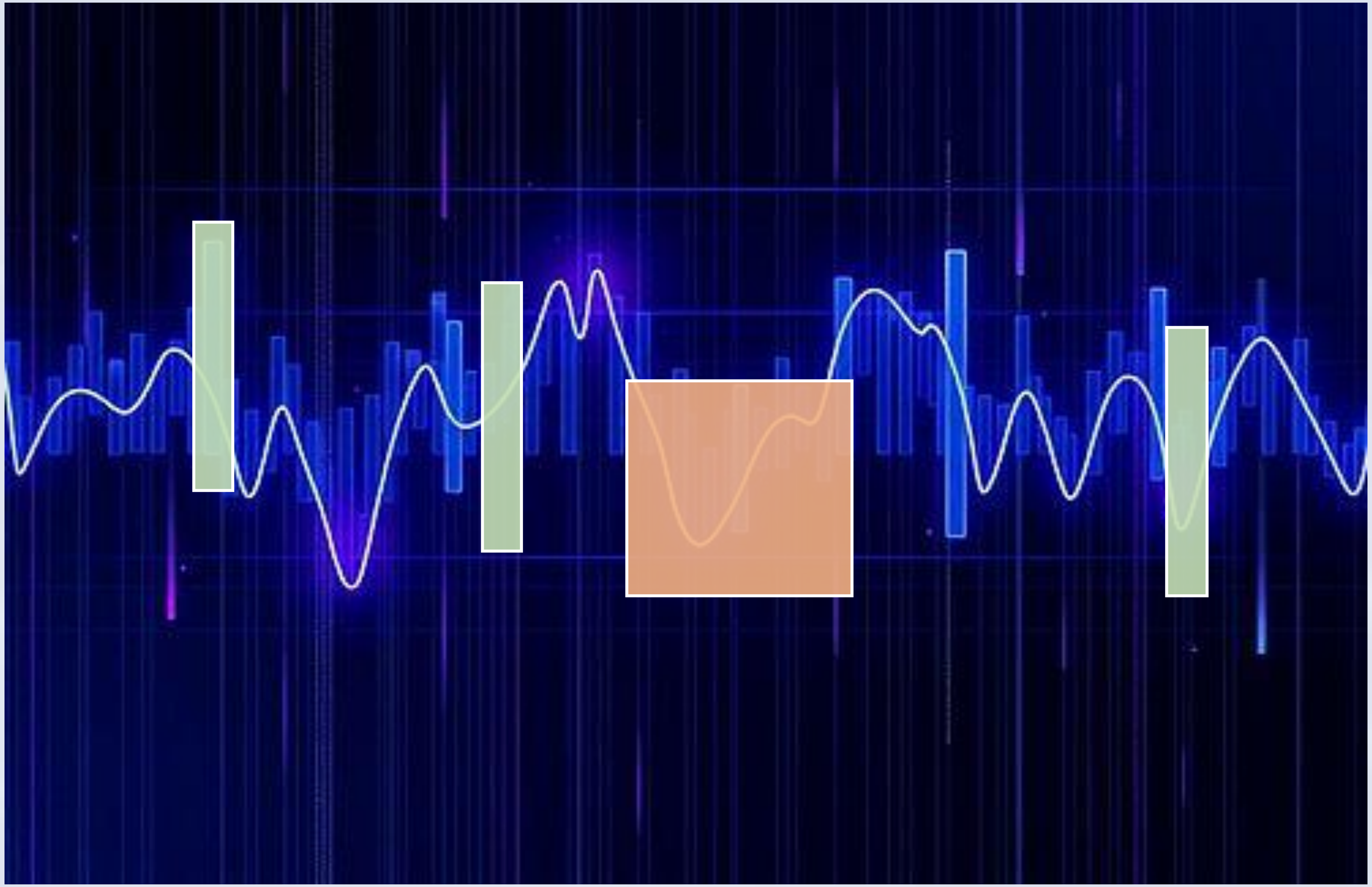
GW : Ripples in Space-Time



Being in space: LISA

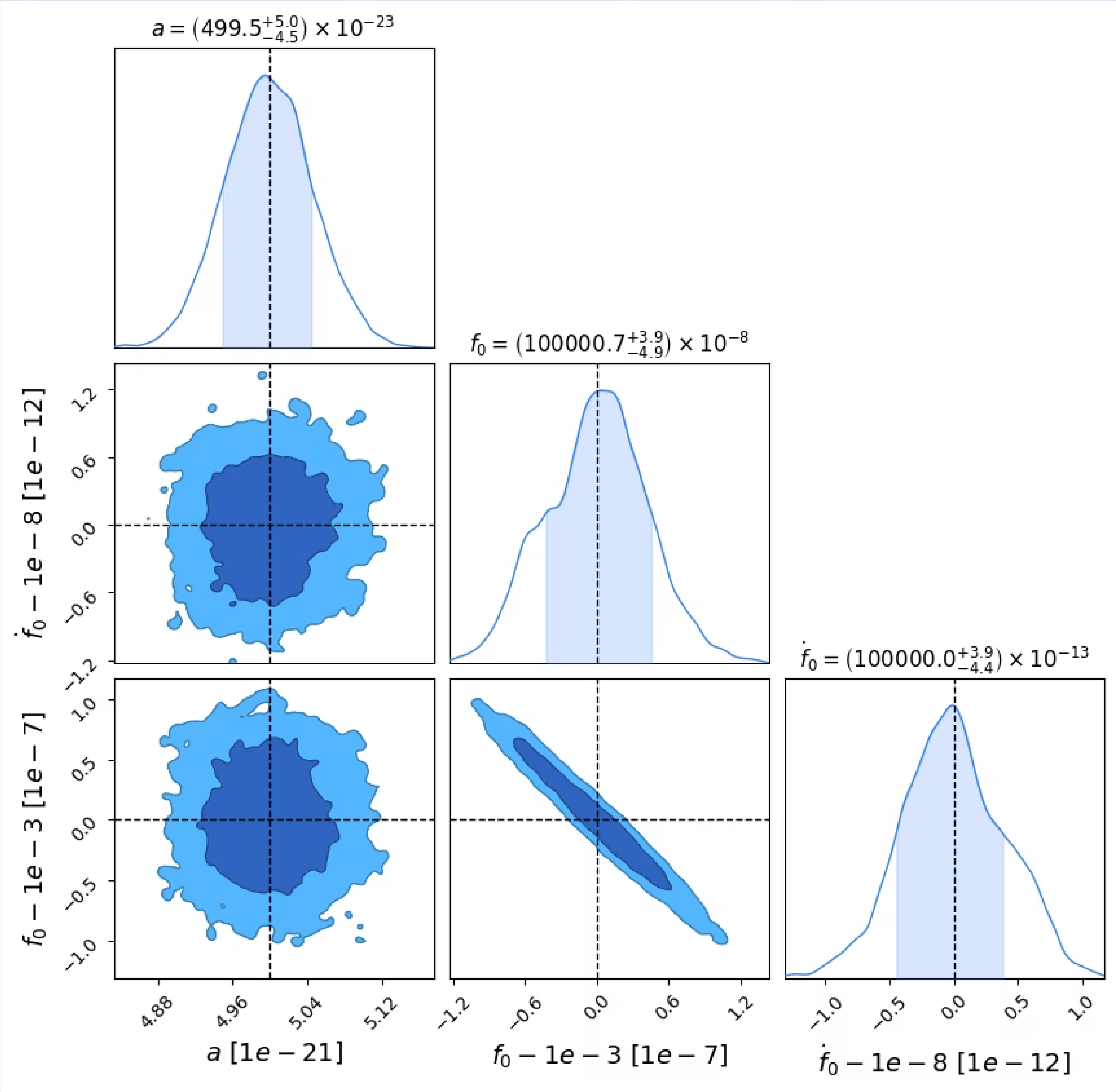
LISA: millihertz frequency band





The Core Problem:

Biased Inference



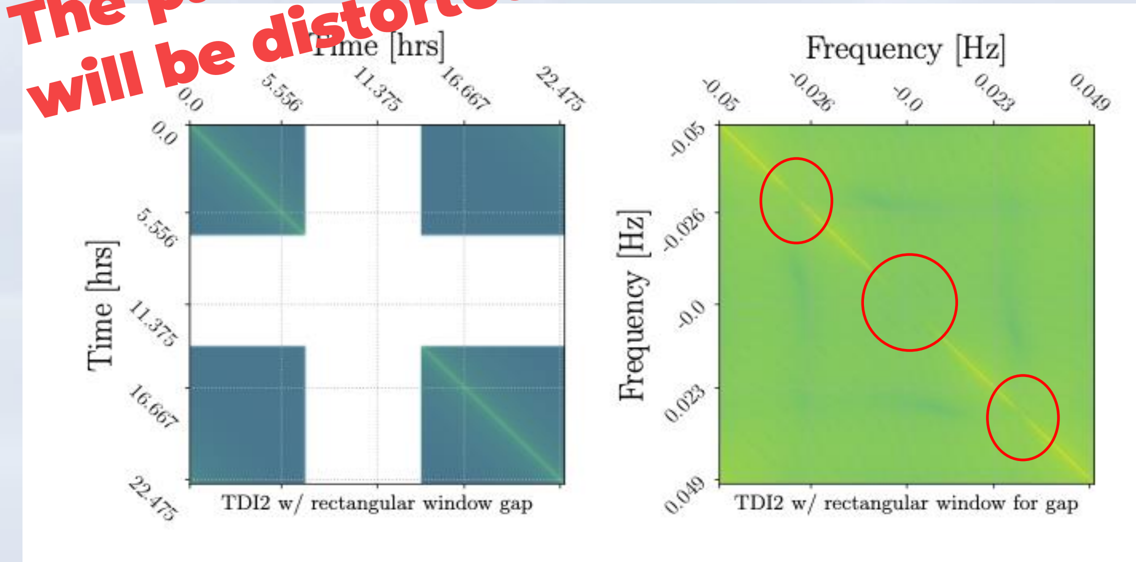
Bayesian inference with Whittle likelihood :

$$p(d|\boldsymbol{\theta}) = -\frac{1}{2}\sum(d-h_\mathrm{m}|d-h_\mathrm{m})$$

Assuming stationary Gaussian noise :

$$\mathrm{Re}(\hat{n}(f_i)), \mathrm{Im}(\hat{n}(f_i)) \sim N\left(0, \frac{S_n(f_i)}{4\Delta f}\right)$$

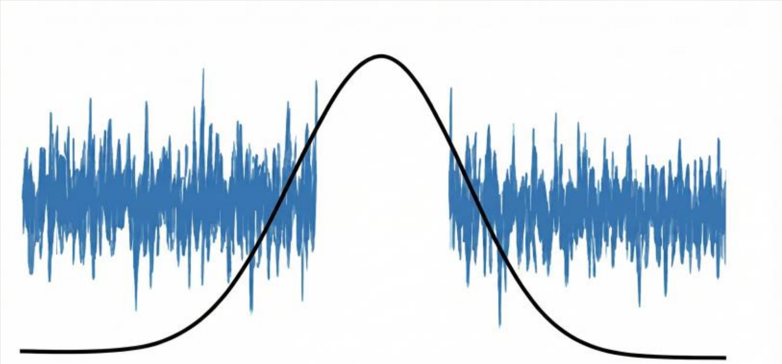
The parameter estimation with MCMC will be distorted!



Previous Research

Standard & Apodization Techniques

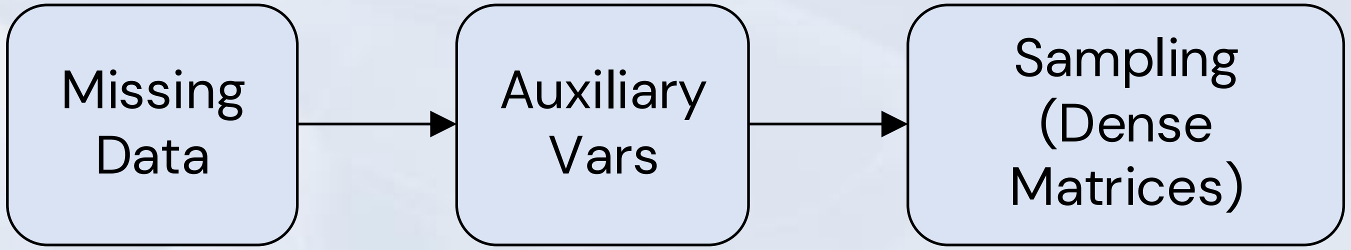
Windowing[2014]: forced to throw away data, with loss of information


Apodization (Windowing)

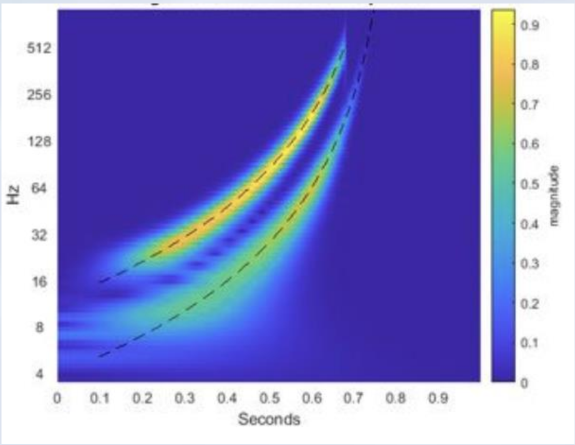
Critique: Alters noise stochasticity, violates stationarity, biases estimation [Burke et al., 2025]

Reconstruction: Augmentation

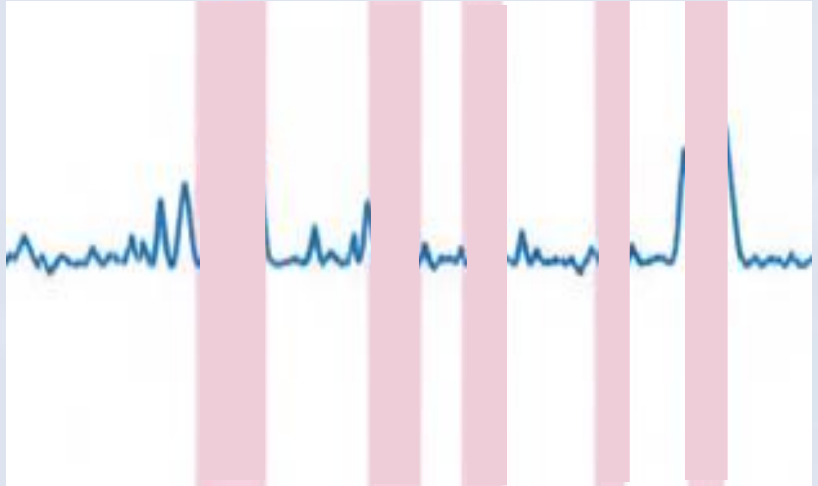
- **Bayesian Data Augmentation[2019]**: Computationally prohibitive for long signals

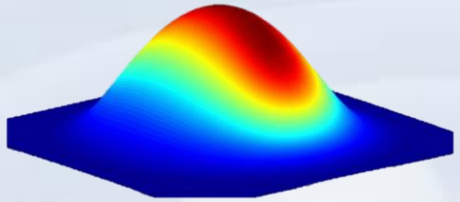


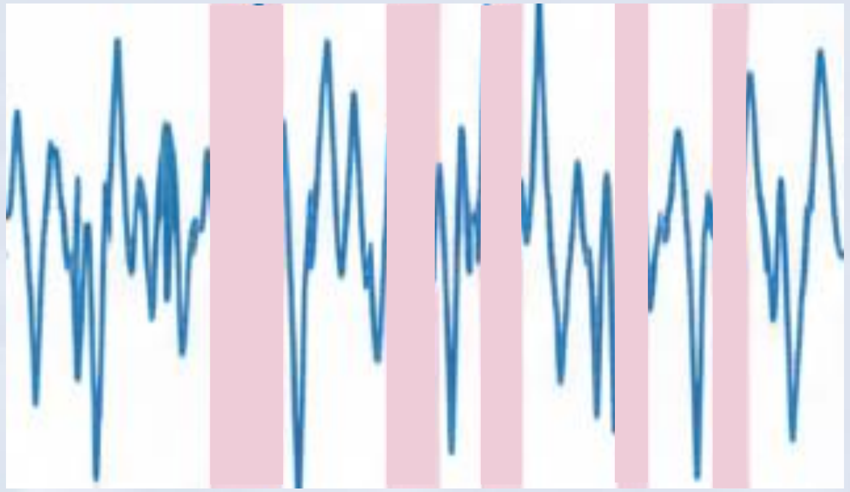
- **Wavelet Domain Augmentation [2025]:** Relies on strict local stationarity assumption.



From Imputation to Direct Inference







Previous Work:
Imputation: BiGRU-CAE

Denoising is needed when considering real signal

Target: Direct parameter Inference from “gapped “signal

A Robust, Scalable SBI Framework

Architecture: Embed & Flow

The Summarizer

This network compresses the high-dimensional input *d(t)* into a low-dimensional, dense summary statistic vector, *s*.

The Flow Matching Engine

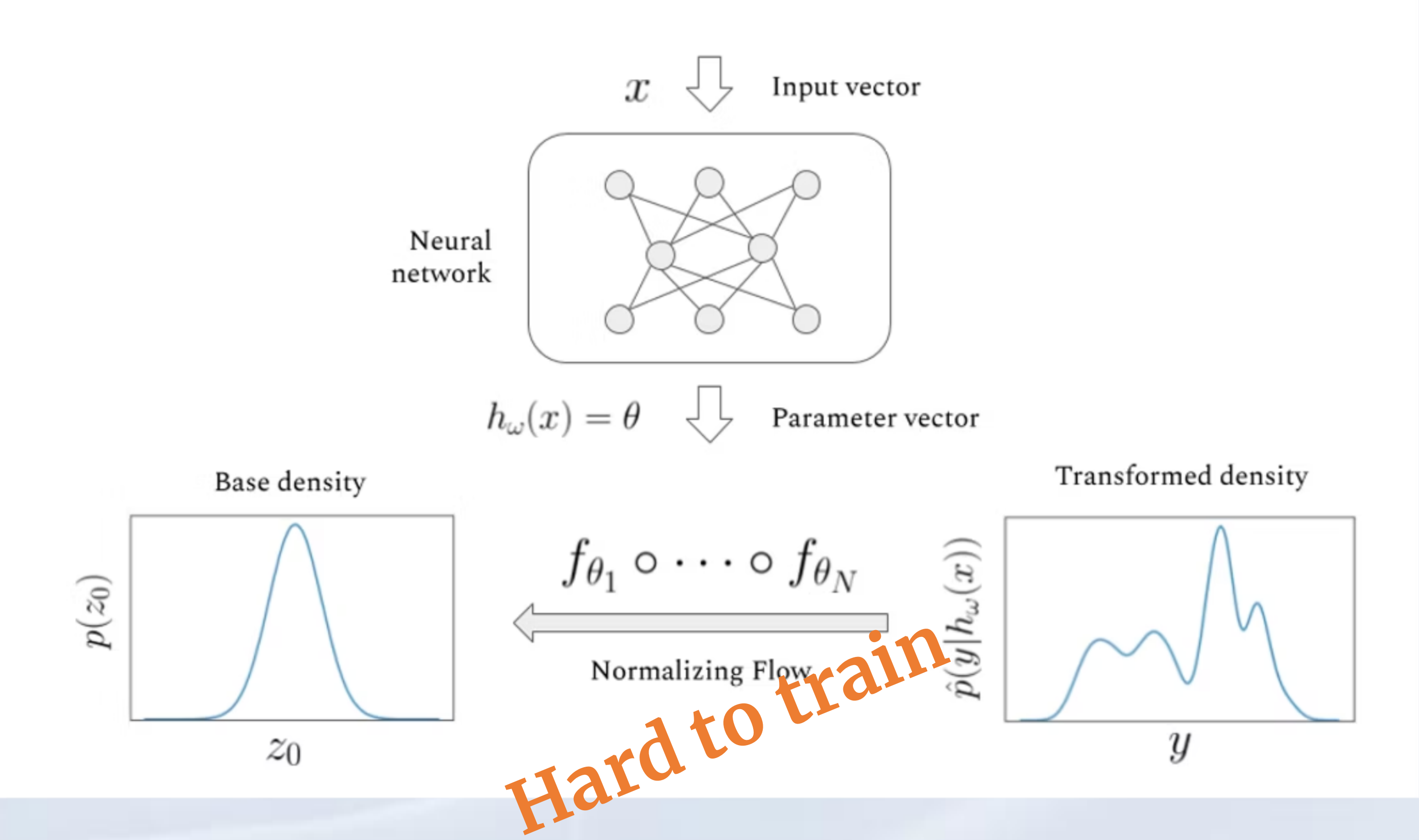
Takes *s* as input, which conditions the vector vector field *v_t(\theta|s)*, transforming the base base distribution into the final posterior *p(\theta|d)*.

Joint Training is Key

The Summarizer learns features that are specifically **optimized for parameter estimation**, ensuring the minimal amount of information is lost.

Inference Engine: Why Flow Matching?

Normalizing Flows: The Foundation



Simple Prior

Invertible Transform

Complex Posterior

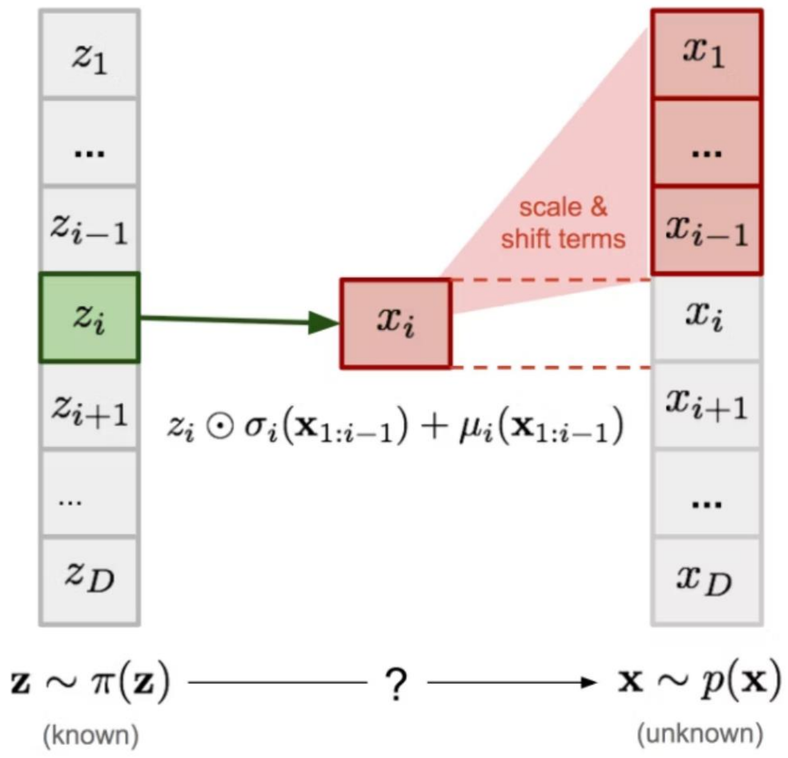
MAF: Forcing Invertibility

The Constraint Challenge

Autoregressive NN for transformation layers
Restrictive architectural structure ensures invertibility

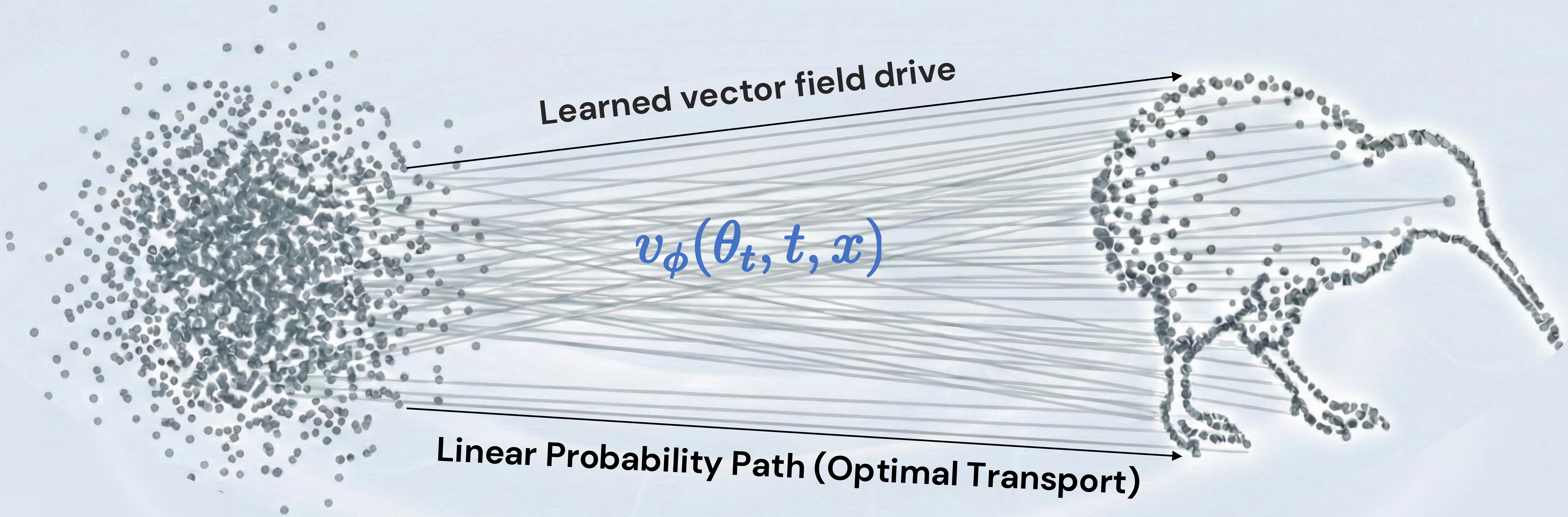
Simplified Math
Tractable Jacobian determinant calculations

Might be unstable



Masked Autoregressive Flow (MAF)

Flow Matching: Breaking Free



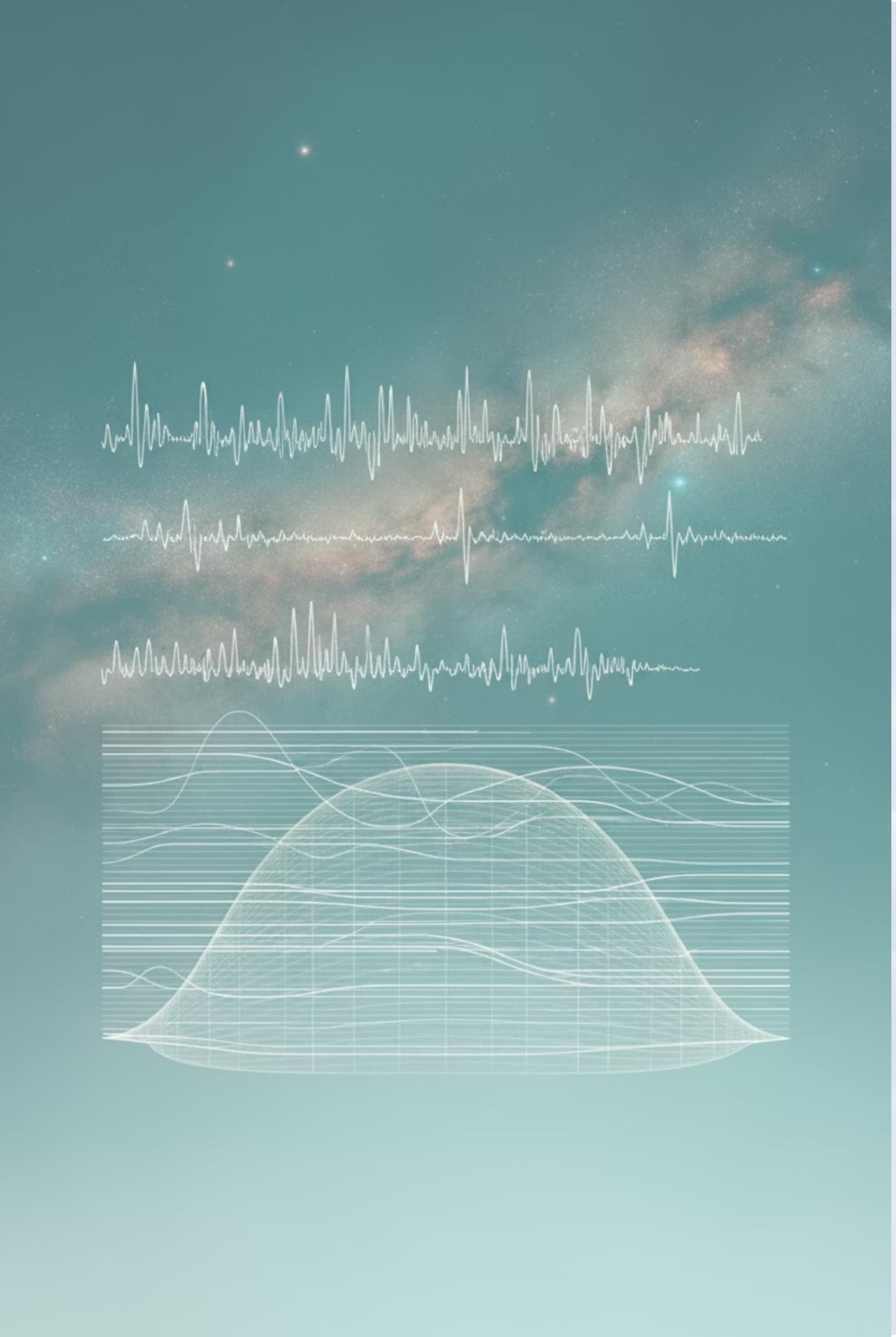
Base distribution
t=0: $z \sim \pi$

$$u_t(z,\theta) = \frac{d}{dt}\psi_t(z,\theta) = \theta - z$$

$$(1-t)z + t\theta$$

Target distribution
t=1: $\theta \sim p(\theta|x)$

$$\mathcal{L}(\phi) = \mathbb{E}_{t\sim\mathcal{U}(0,1),\, z\sim\pi(z),\, \theta\sim p(\theta|x)}\left[\left\|v_\phi(t,\psi_t(z,\theta),x) - (\theta - z)\right\|^2\right]$$

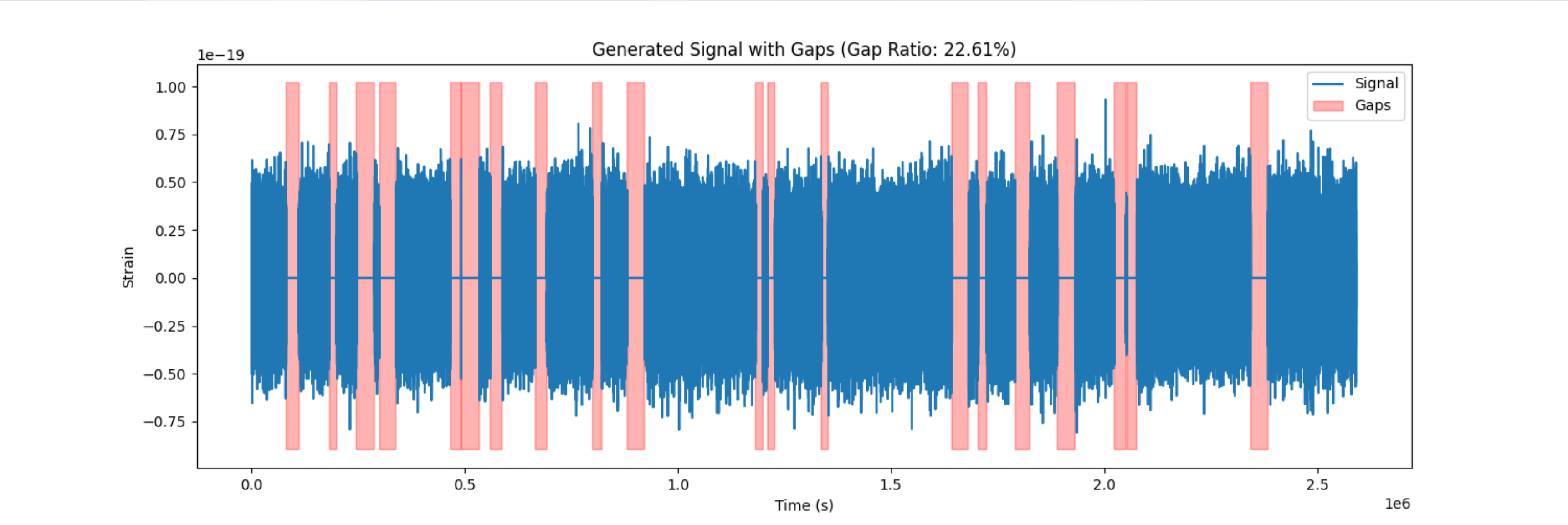


FM VS MAF

on 30-day GB-like signal

Input: signals with gaps in time domain

No worries about the spectrual leakage during FFT



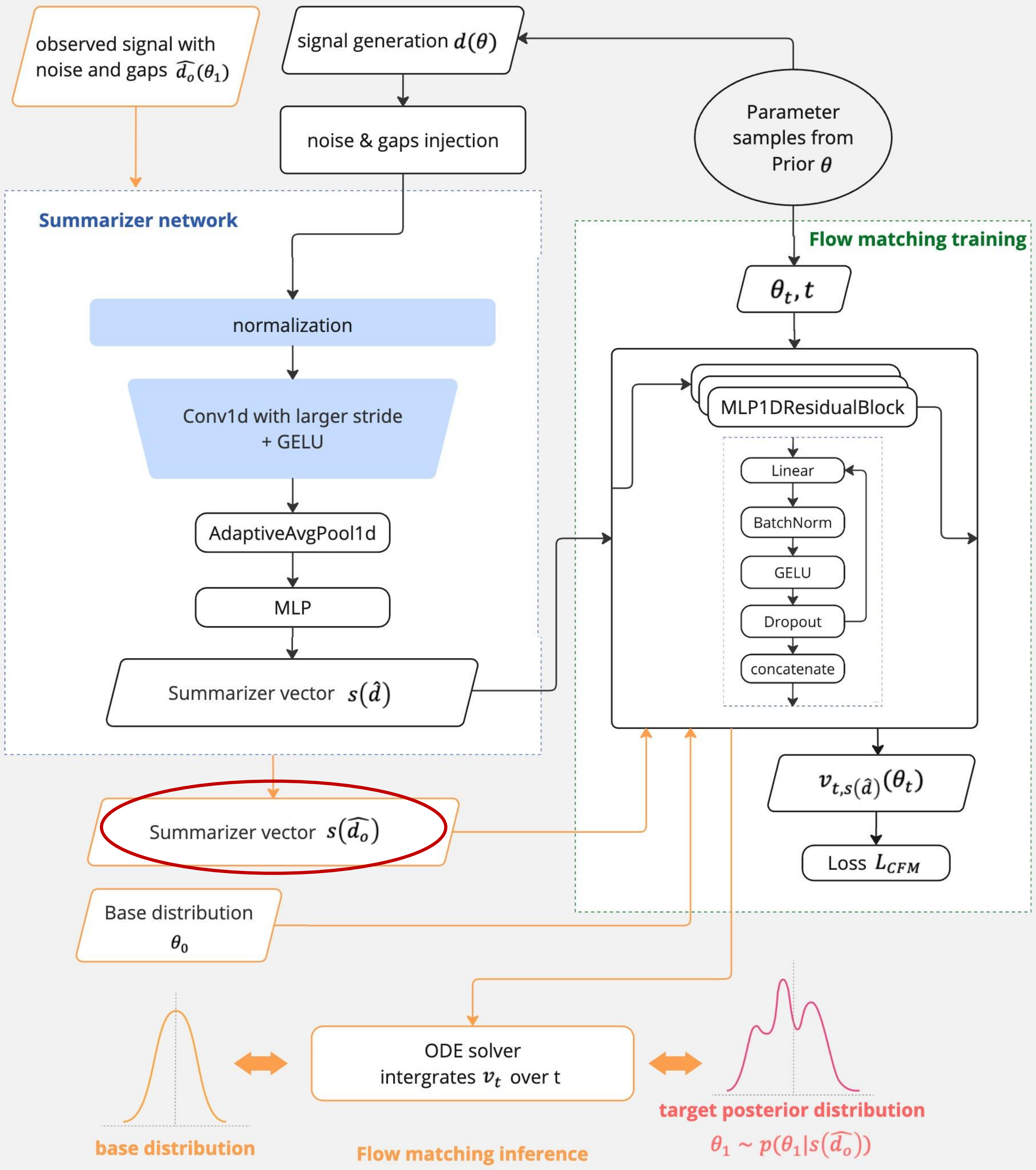
Date generated by GPU-accelerated *fastlisaresponse* package

Summarizer discussion

High dimension: 50K
Only think about the first layer if MLP applied…
50K*512…… computational impossible

Conv1d with large stride layers are applied

curriculum training by adding noise gradually
randomly adding gaps



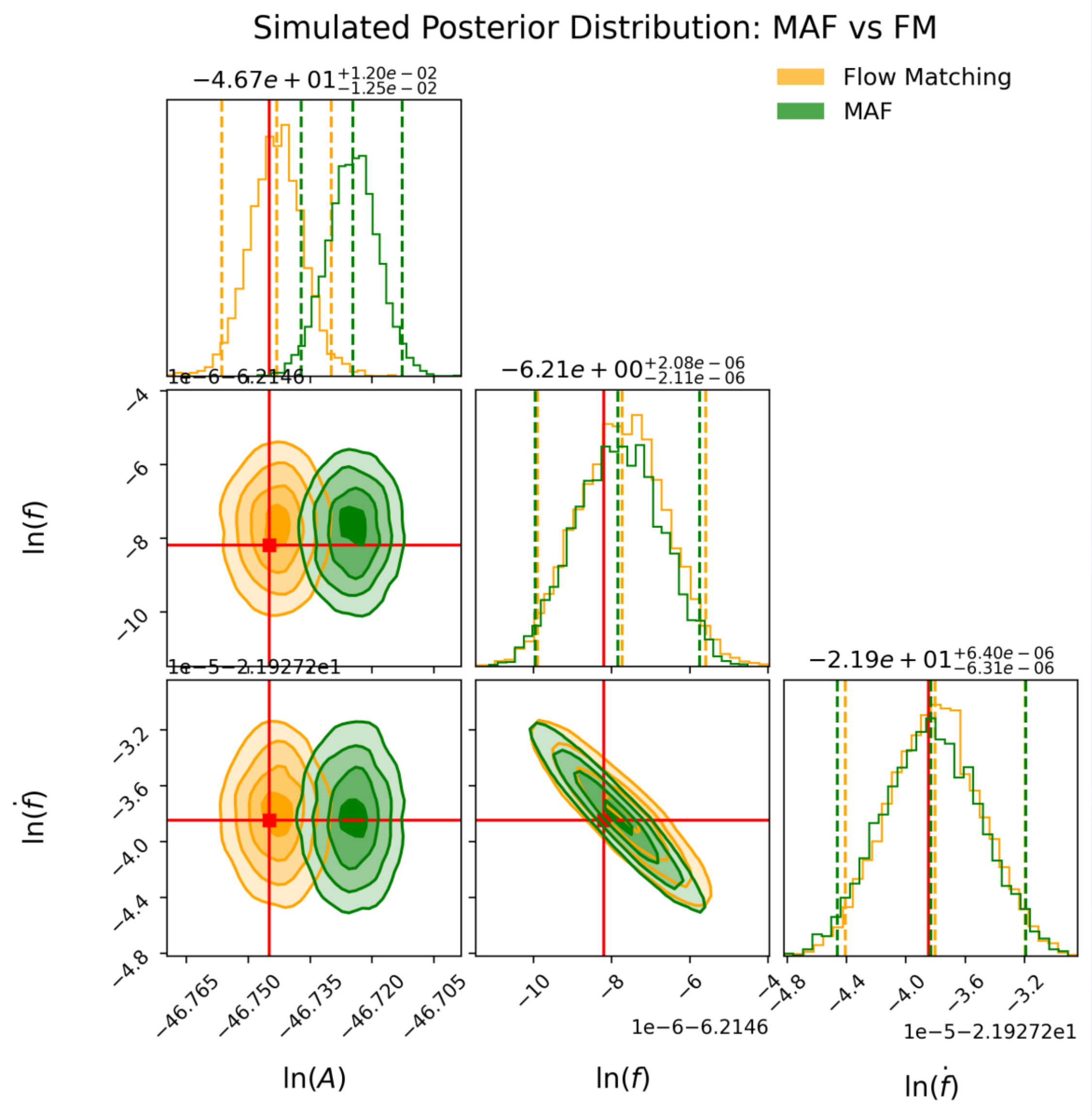
30-Day Signal

FM vs. MAF Comparison

robust

FM Model (orange) accurate estimation

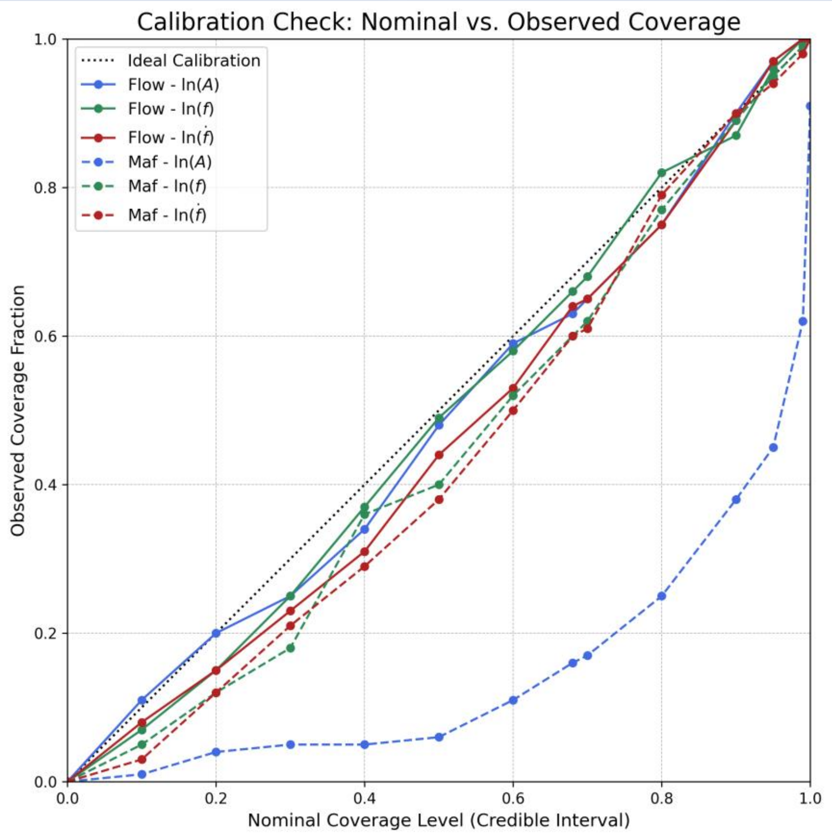
MAF Model (Green) fails at amplitude

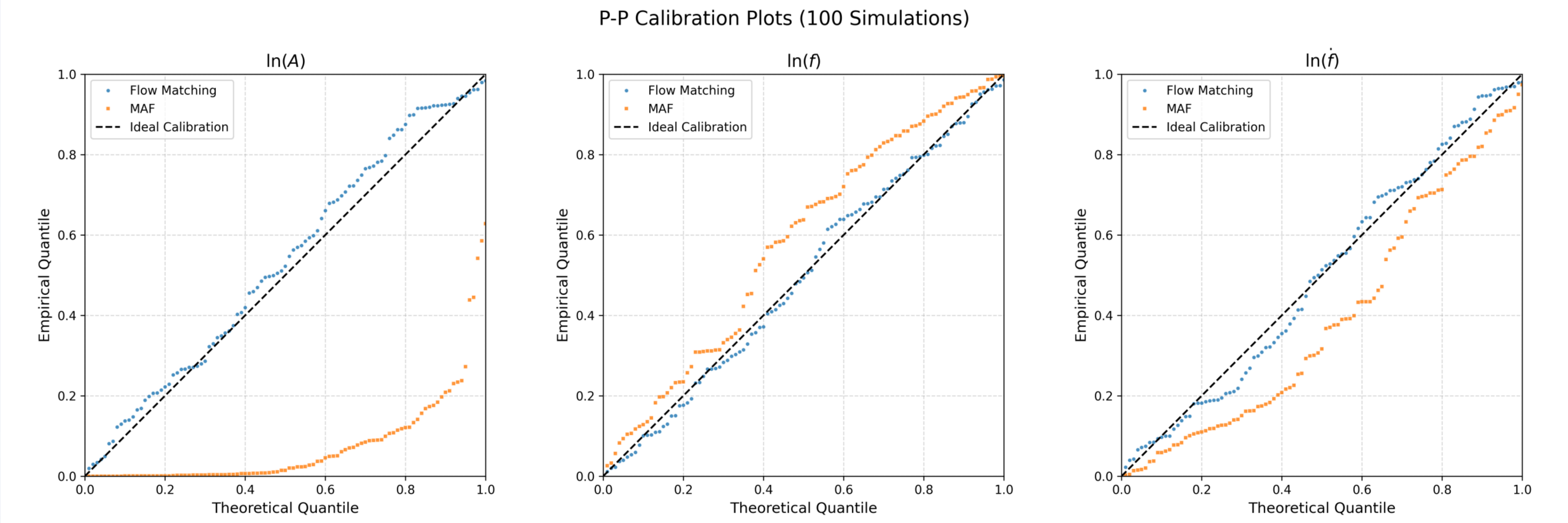


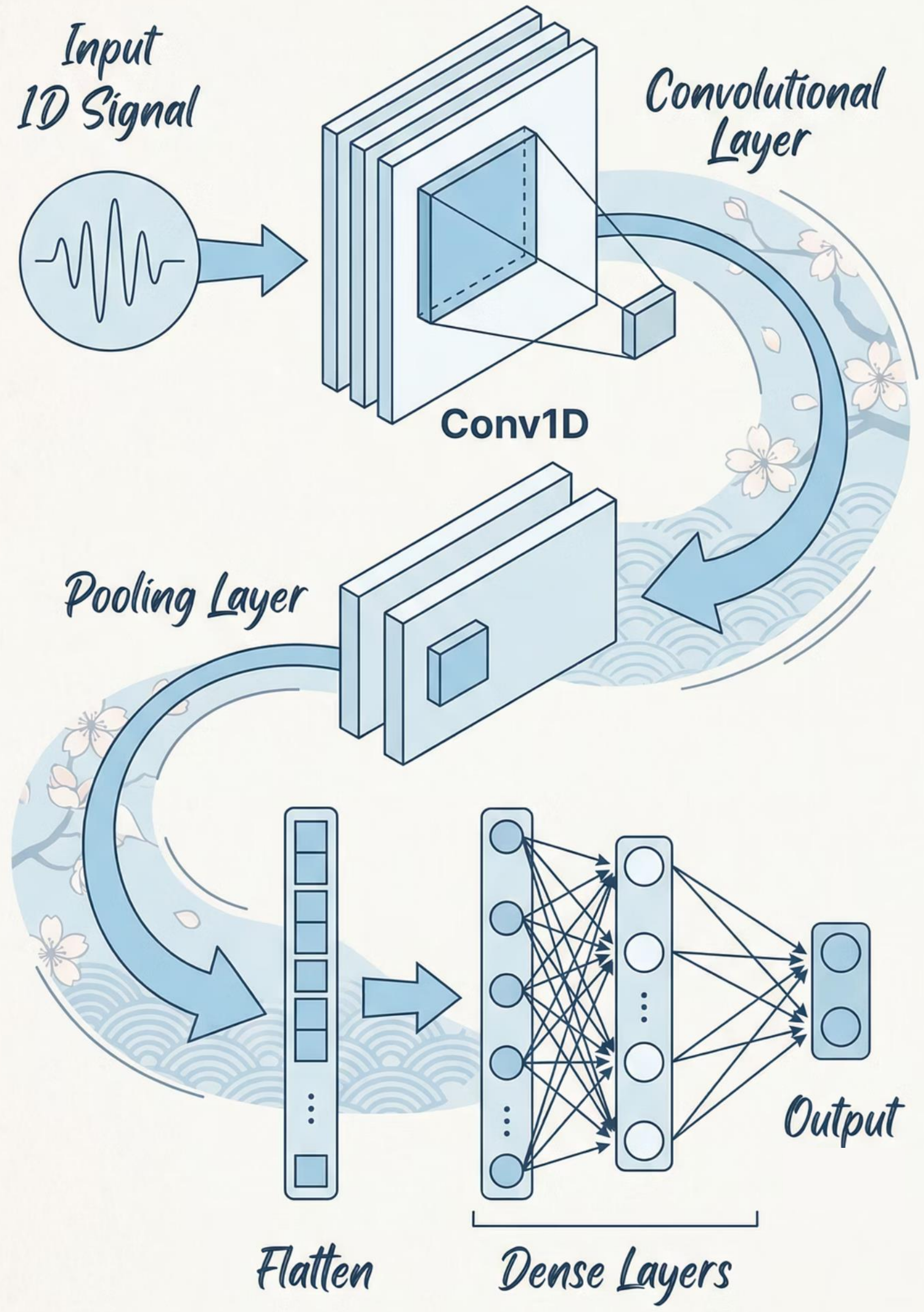
Statistical reliability check

More stable

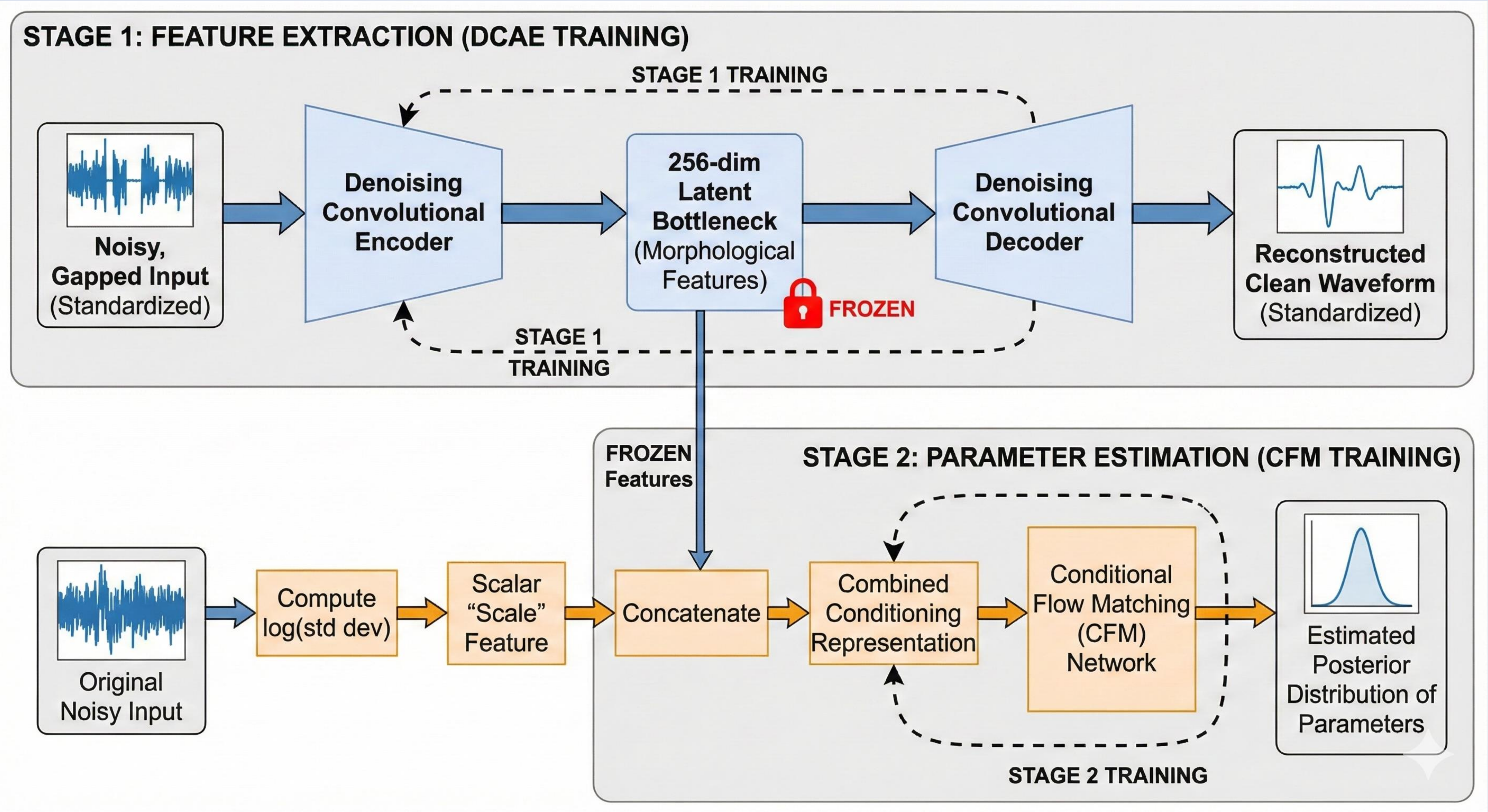
Well-Calibrated



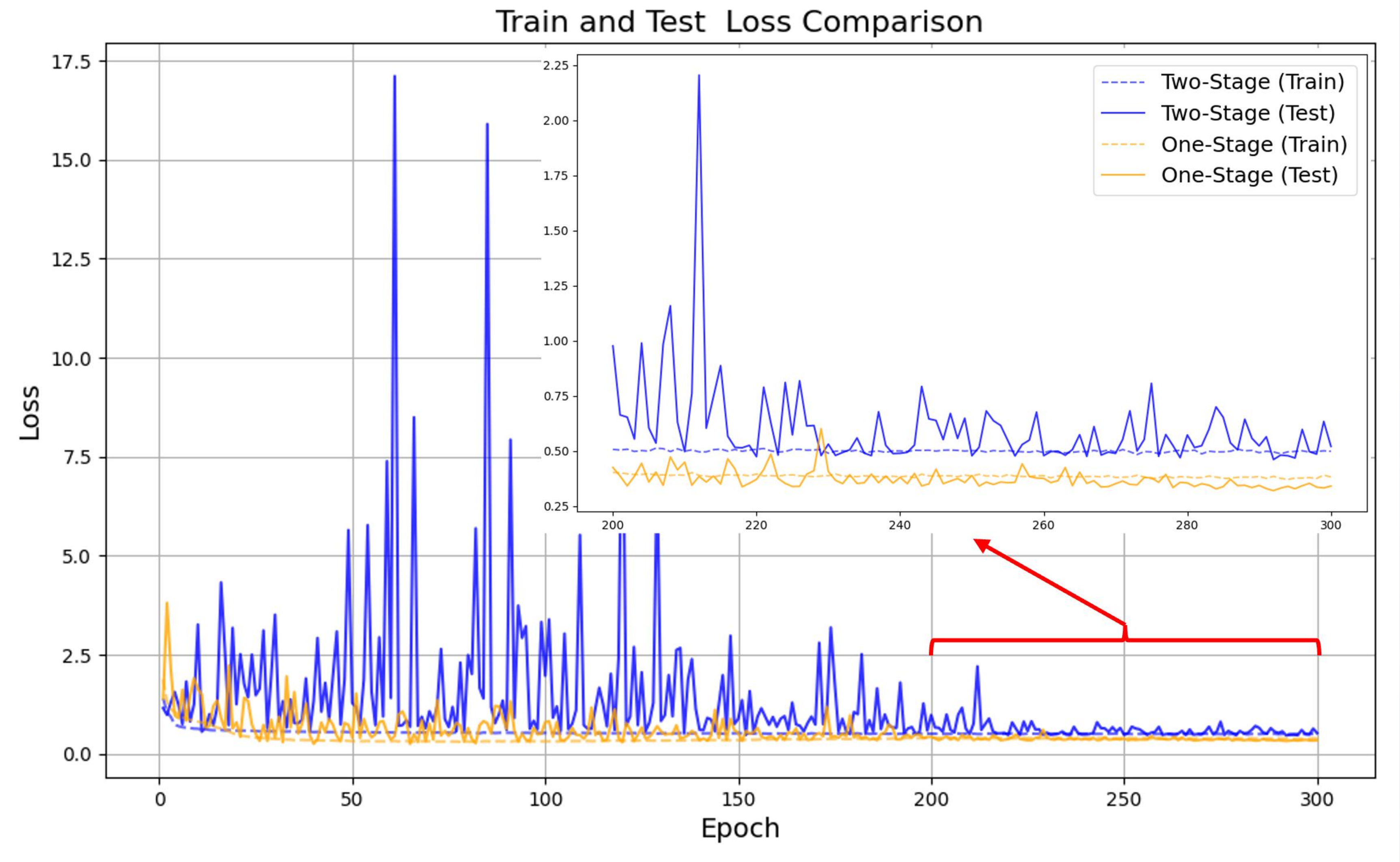




What if decoupling the summarizer during training?


STAGE 1: FEATURE EXTRACTION (DCAE TRAINING)
STAGE 1 TRAINING
Noisy, Gapped Input (Standardized)
Denoising Convolutional Encoder
256-dim Latent Bottleneck (Morphological Features)
FROZEN
Denoising Convolutional Decoder
Reconstructed Clean Waveform (Standardized)
STAGE 1 TRAINING
FROZEN Features
STAGE 2: PARAMETER ESTIMATION (CFM TRAINING)
Original Noisy Input
Compute log(std dev)
Scalar “Scale” Feature
Concatenate
Combined Conditioning Representation
Conditional Flow Matching (CFM) Network
Estimated Posterior Distribution of Parameters
STAGE 2 TRAINING

Joint vs decoupled training



Joint Training

Validation Loss

0.38

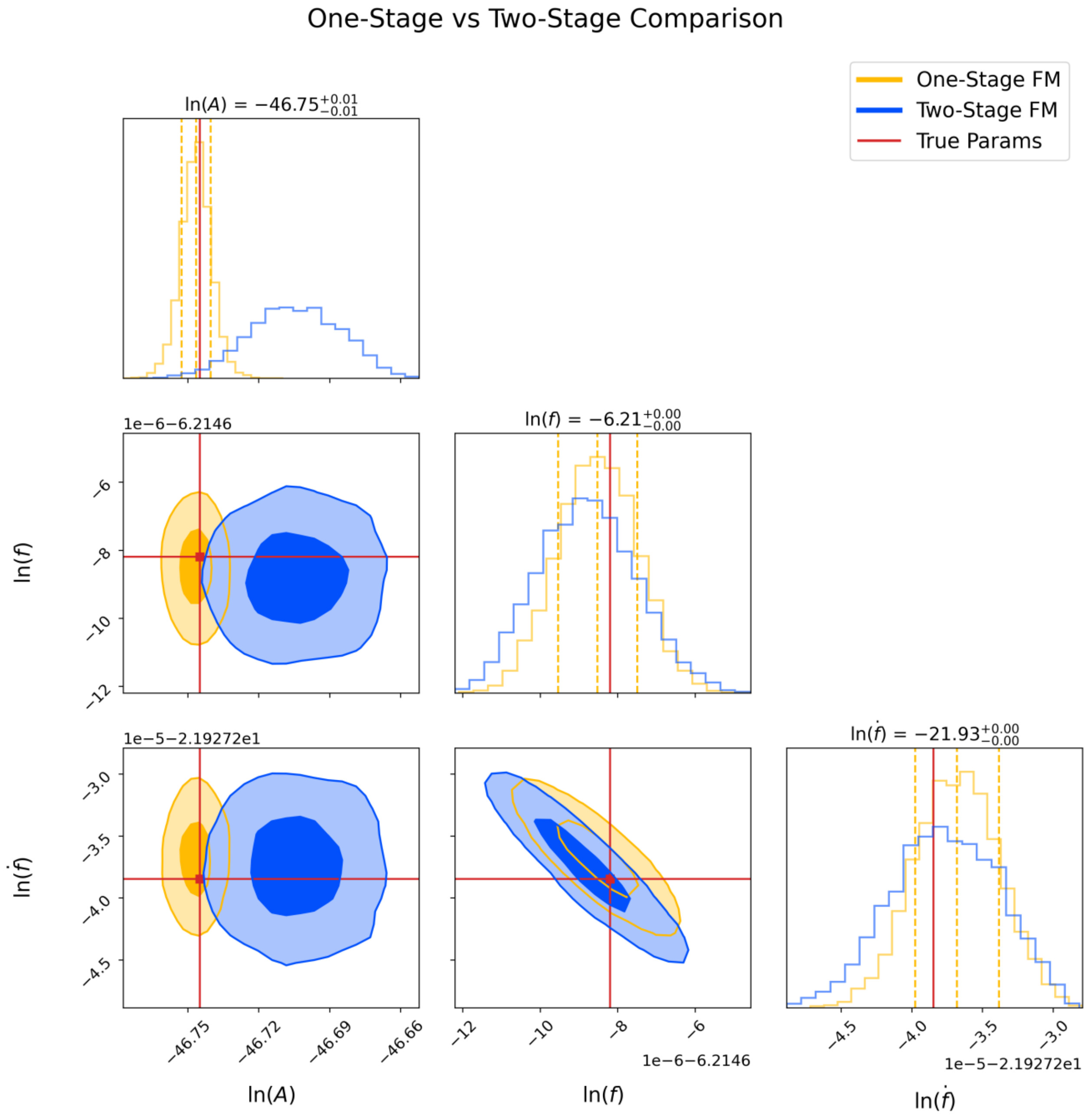
Two-Stage

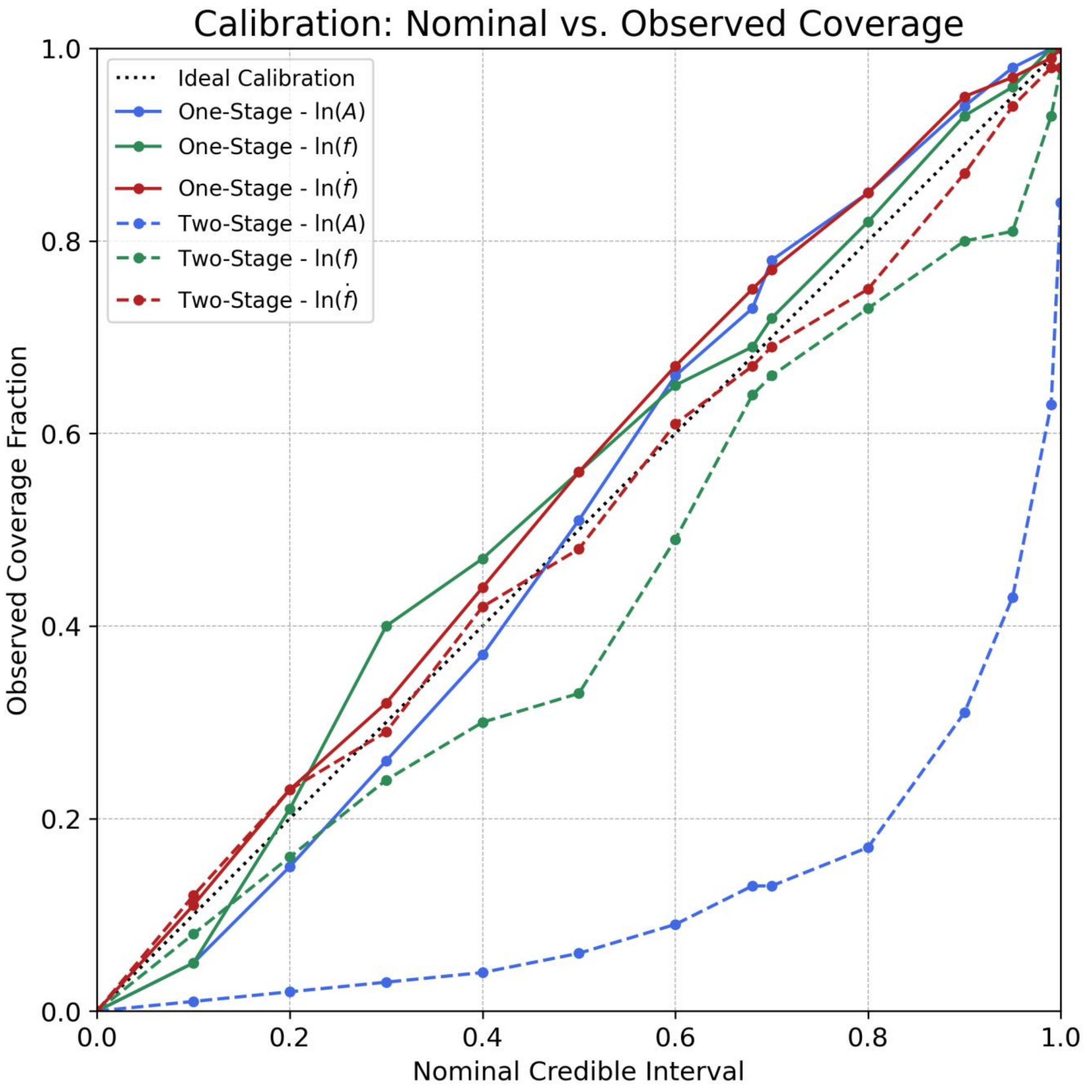
Validation Loss

0.50

Irreducible information bottleneck

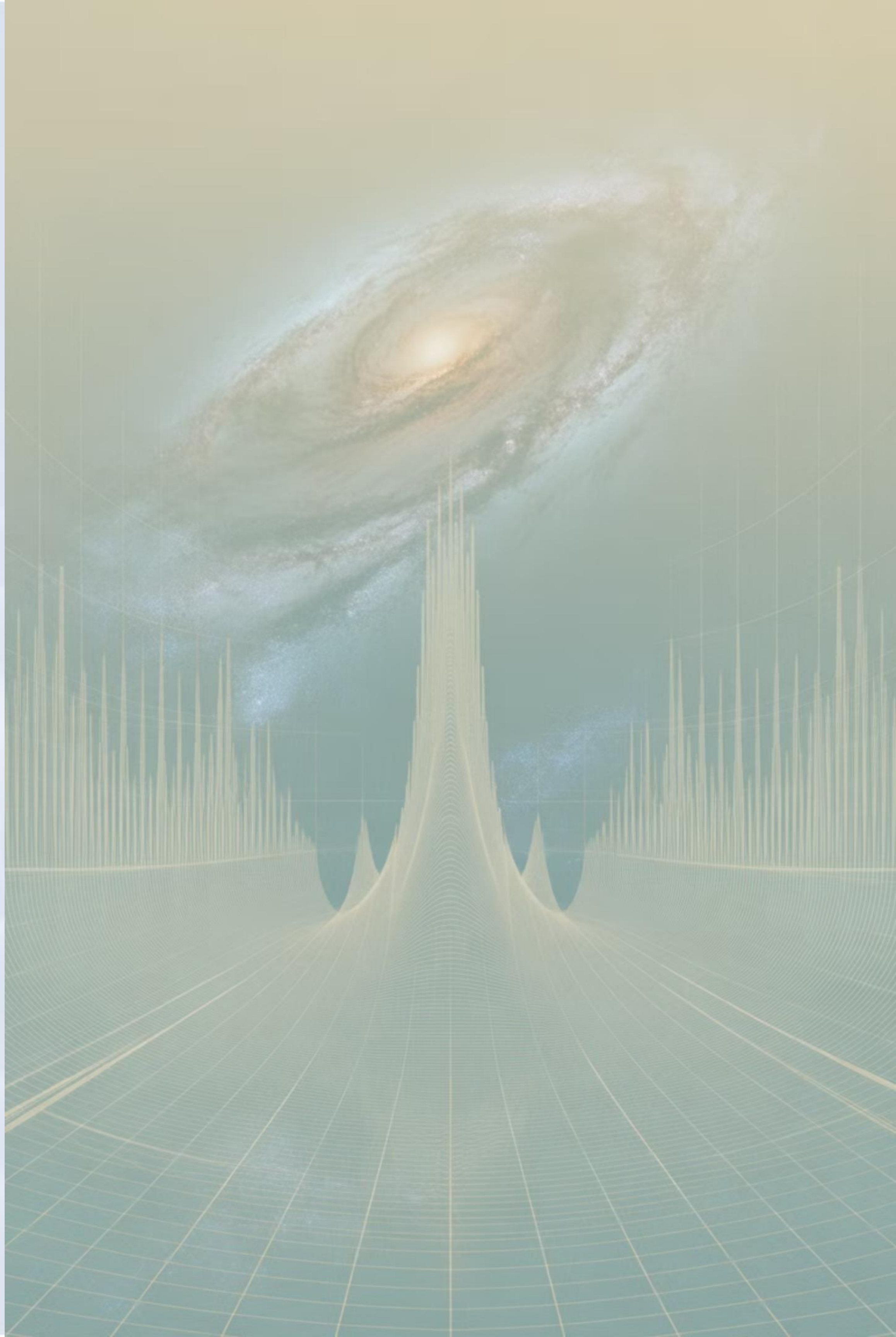
Decoupled training: biased result & wider uncertainty





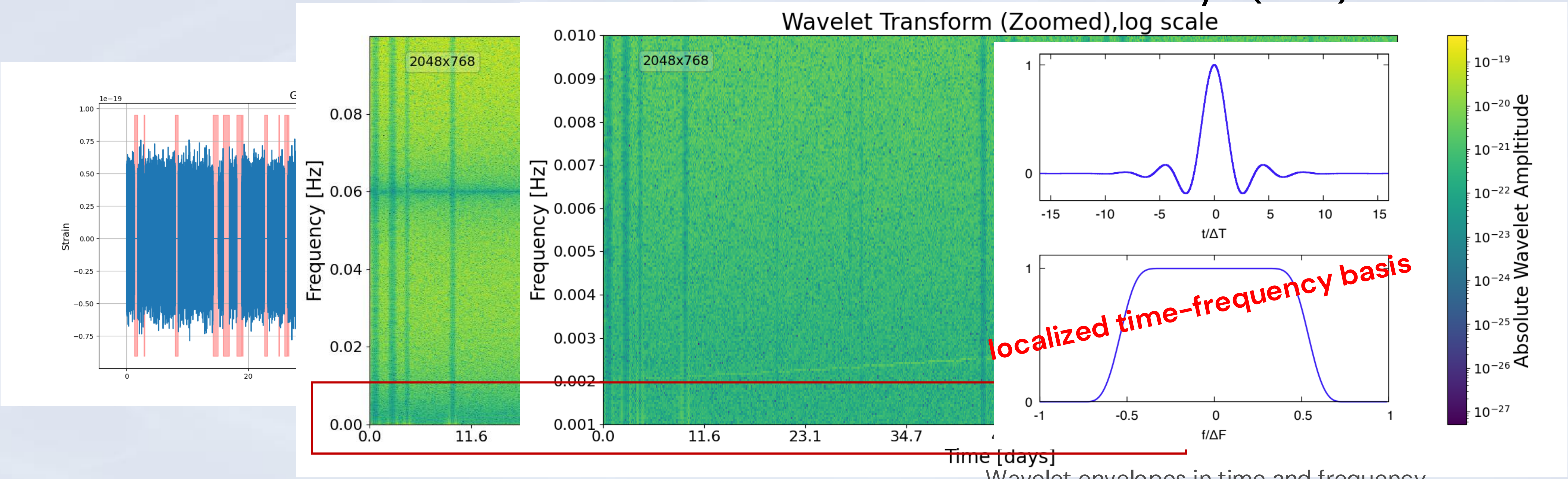
What if for a longer signal

Conv1d cannot handle ……



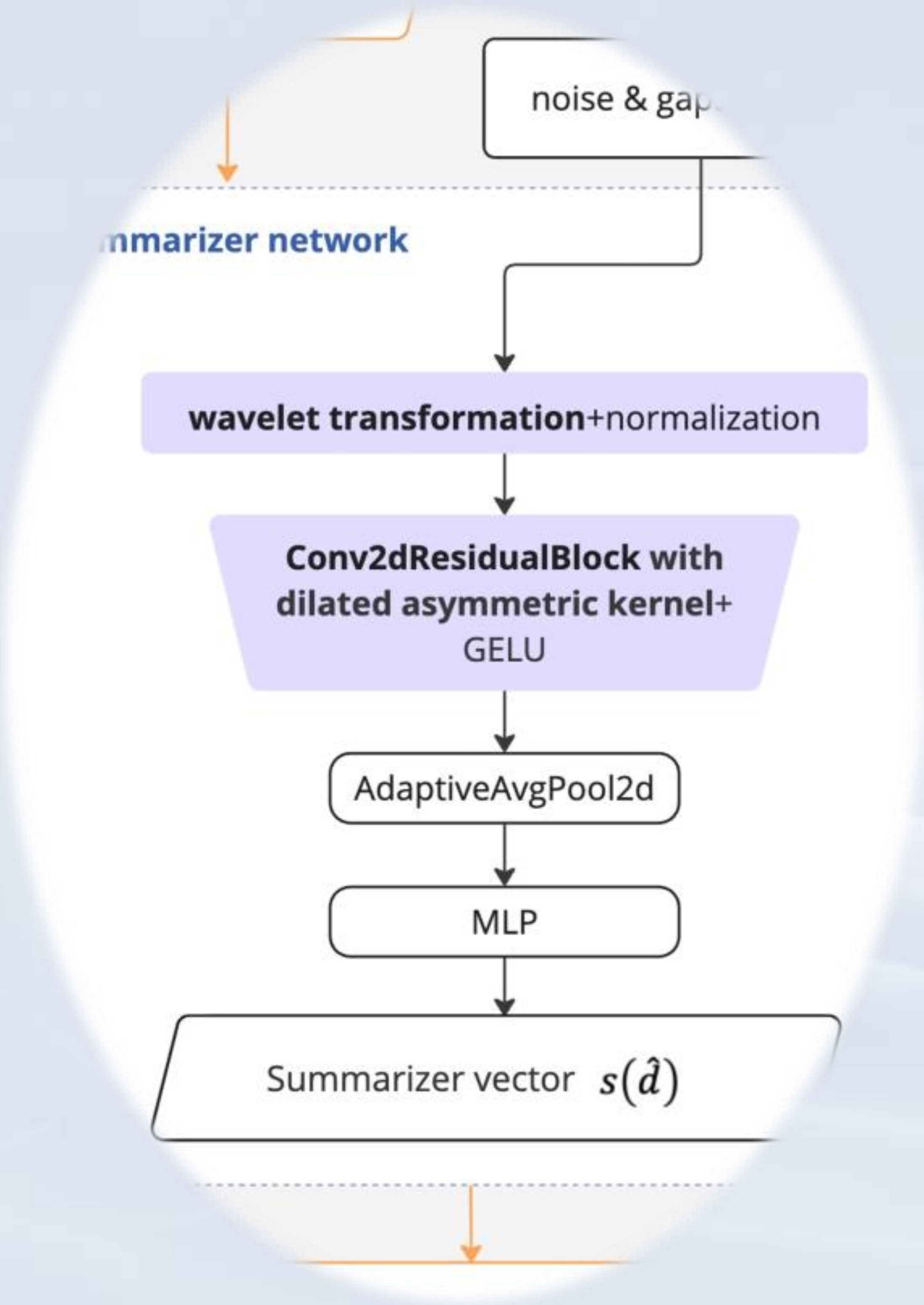
Wavelet: time – frequency domain representation

Wilson-Daubechies-Meyer (WDM) Wavelets



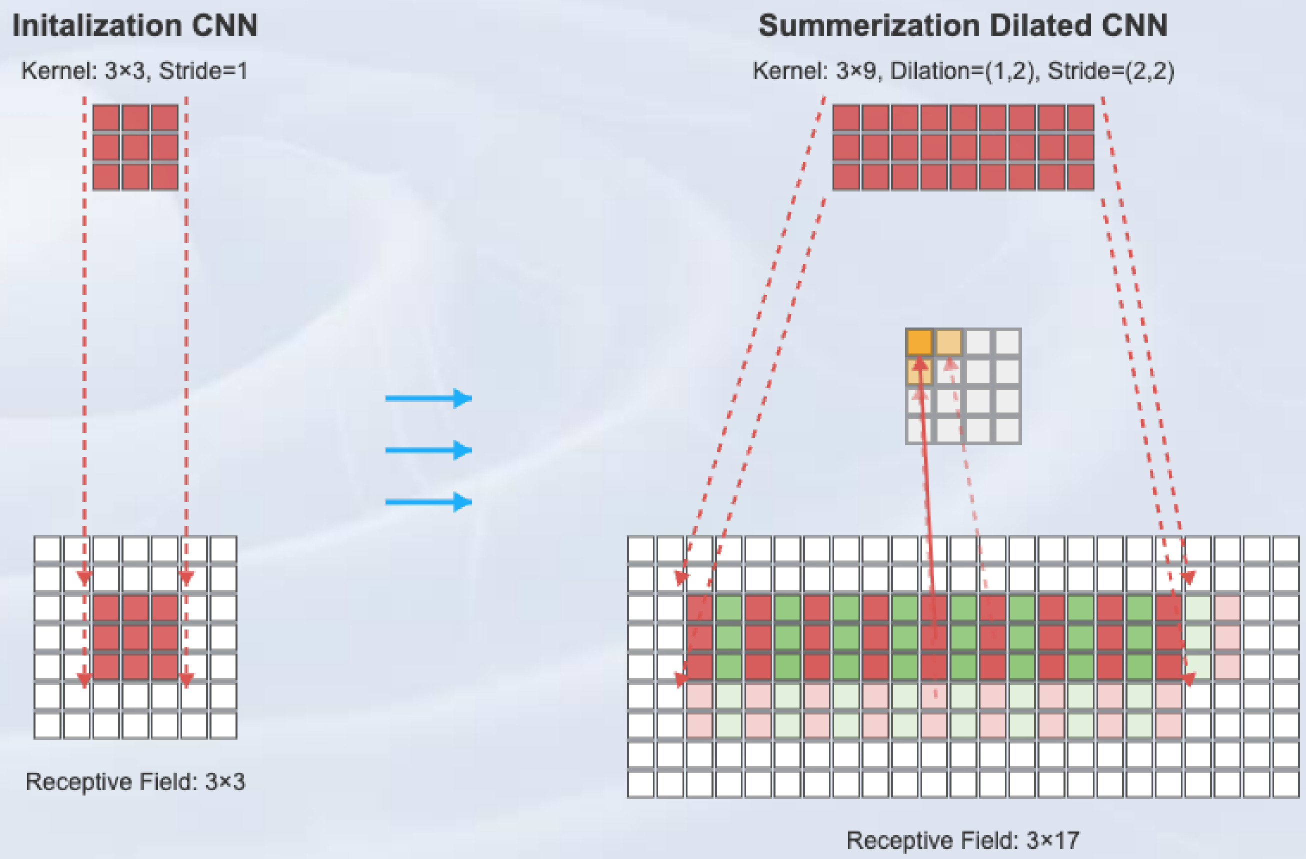
Wavelet envelopes in time and frequency

Cornish (2020): arXiv:2009.00043

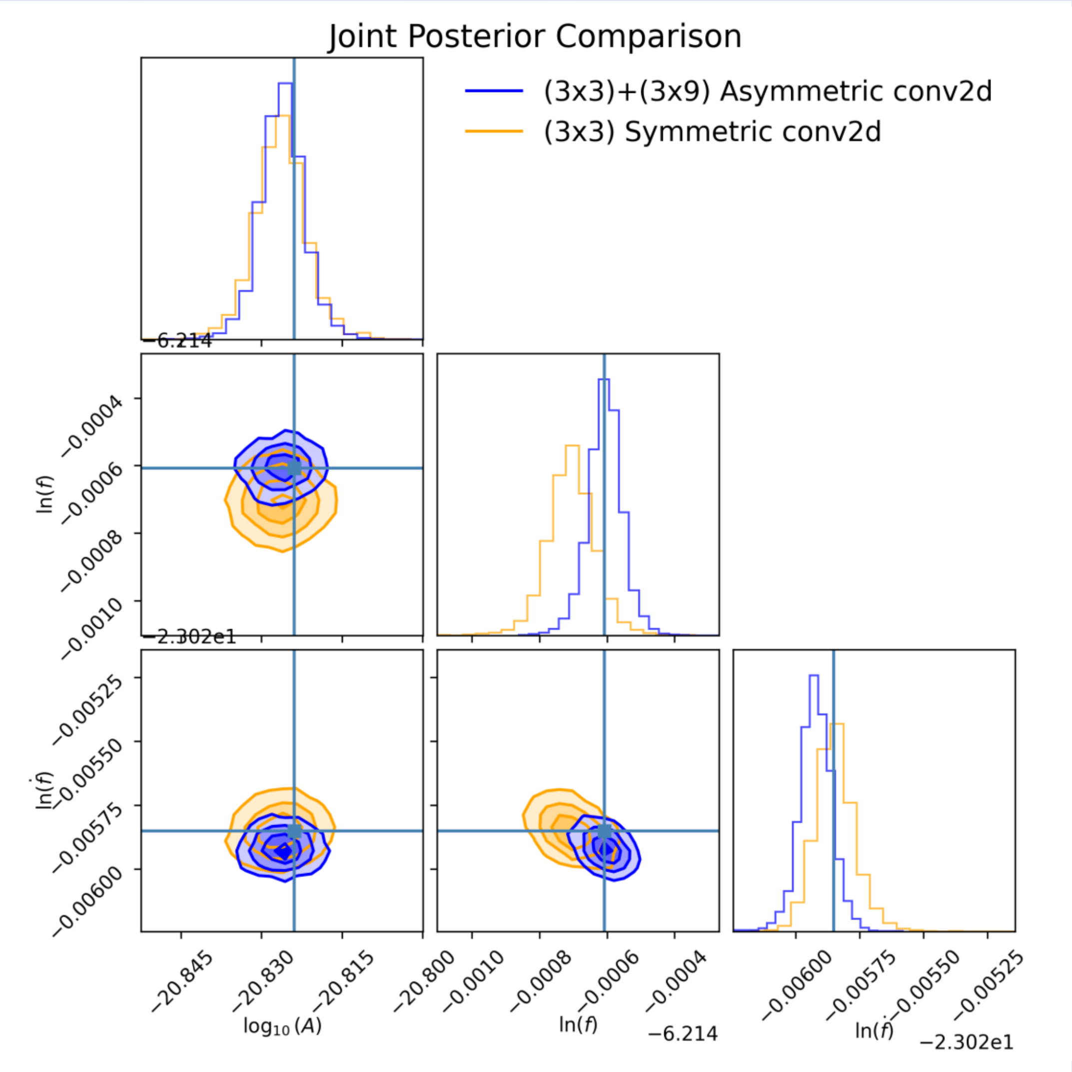


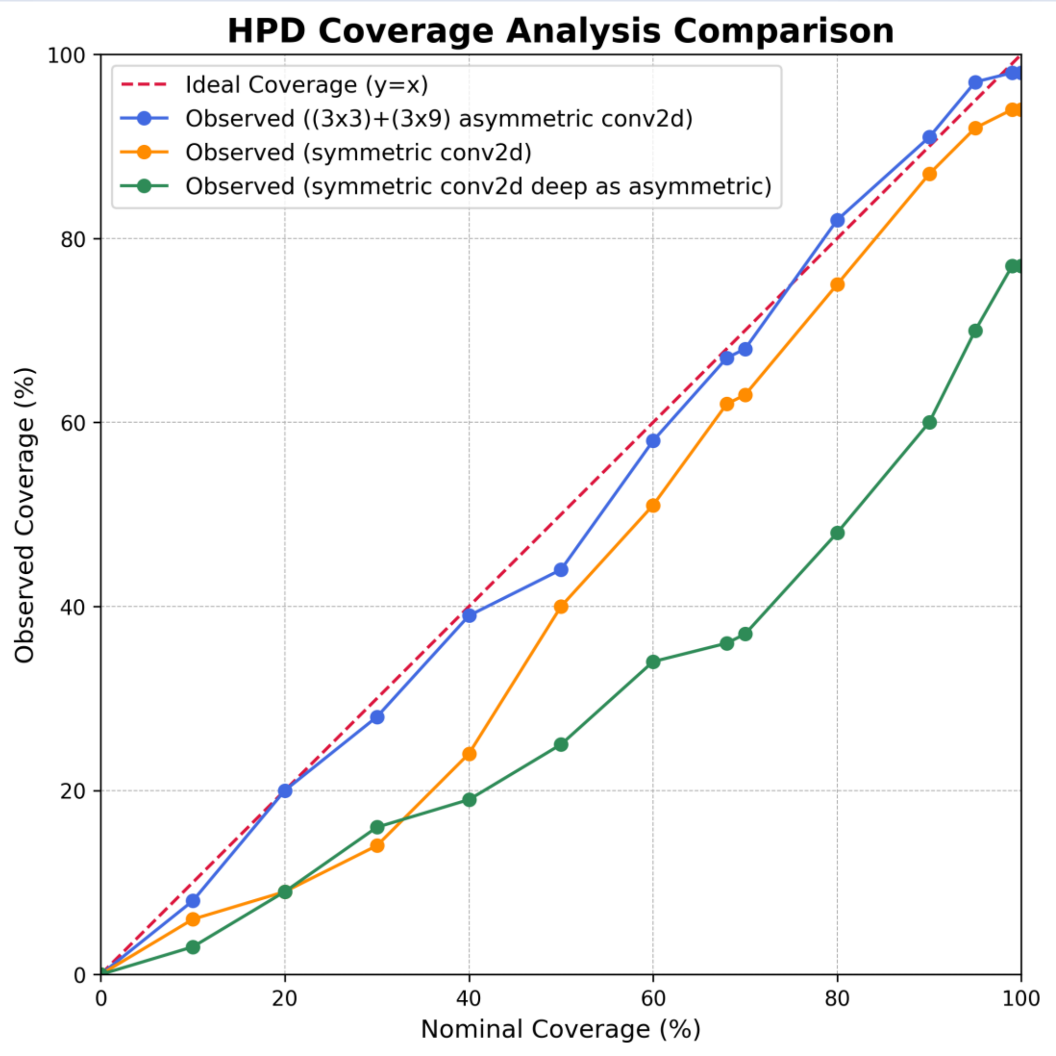
Summarizer for 2D spectrum

Isotropic issue for wavelet representation
Asymmetric Kernels is applied



Asymmetric Kernel better choice





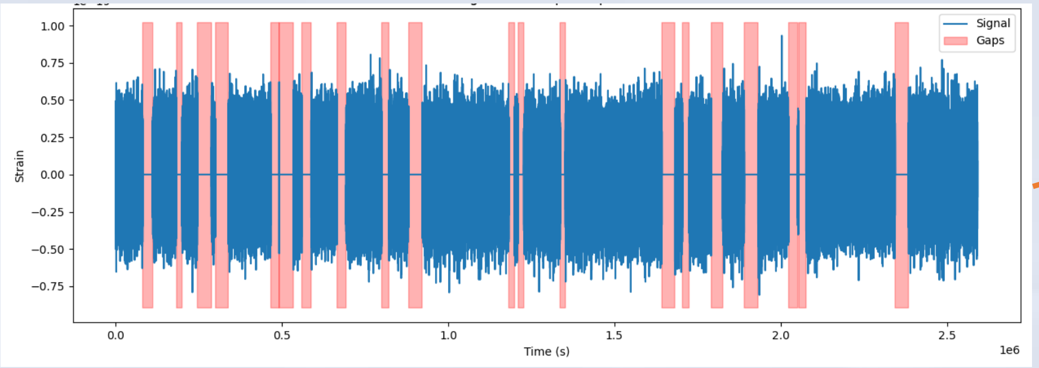
✓ **Asymmetric Kernel (3×9)**

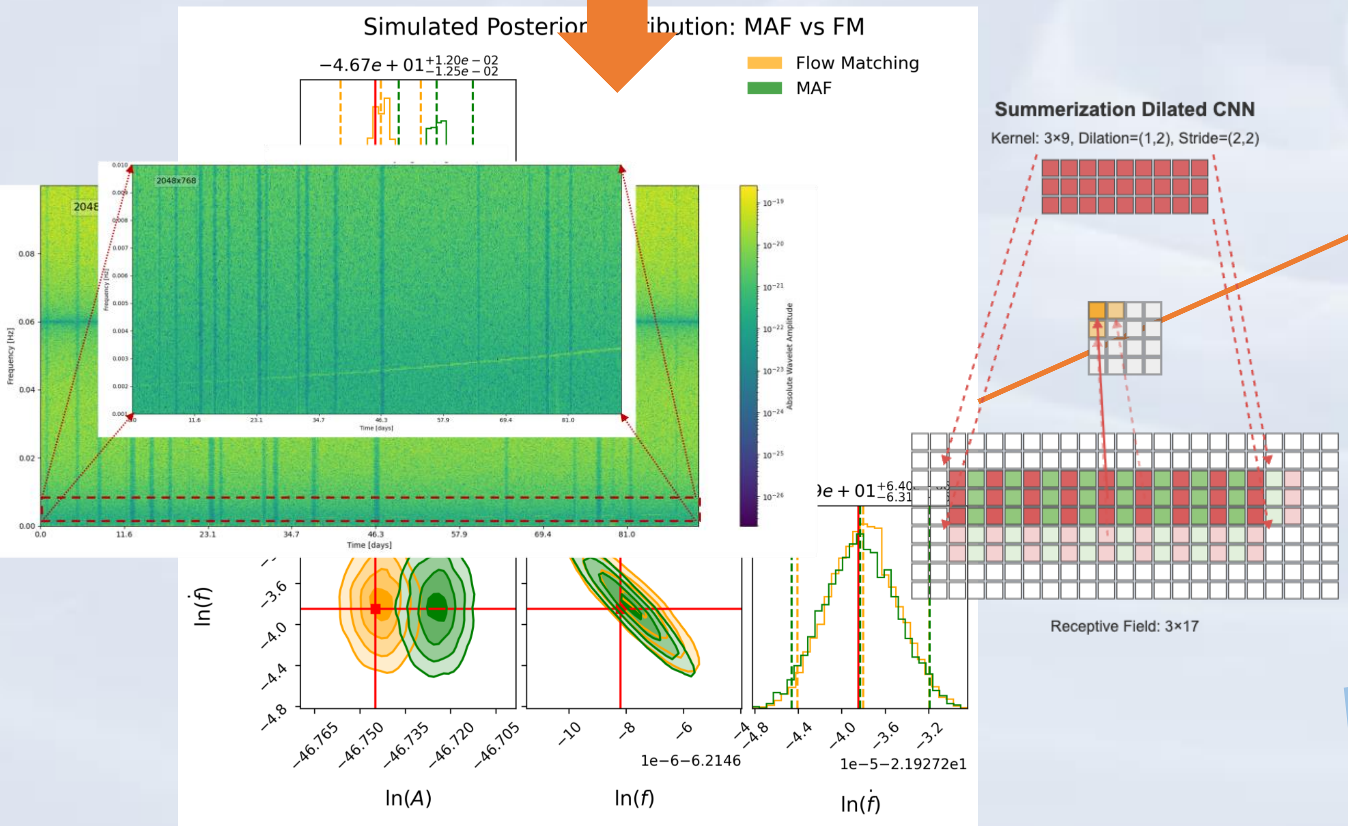
Well-calibrated, on the diagonal. Blue line in posterior, & PP-plot.

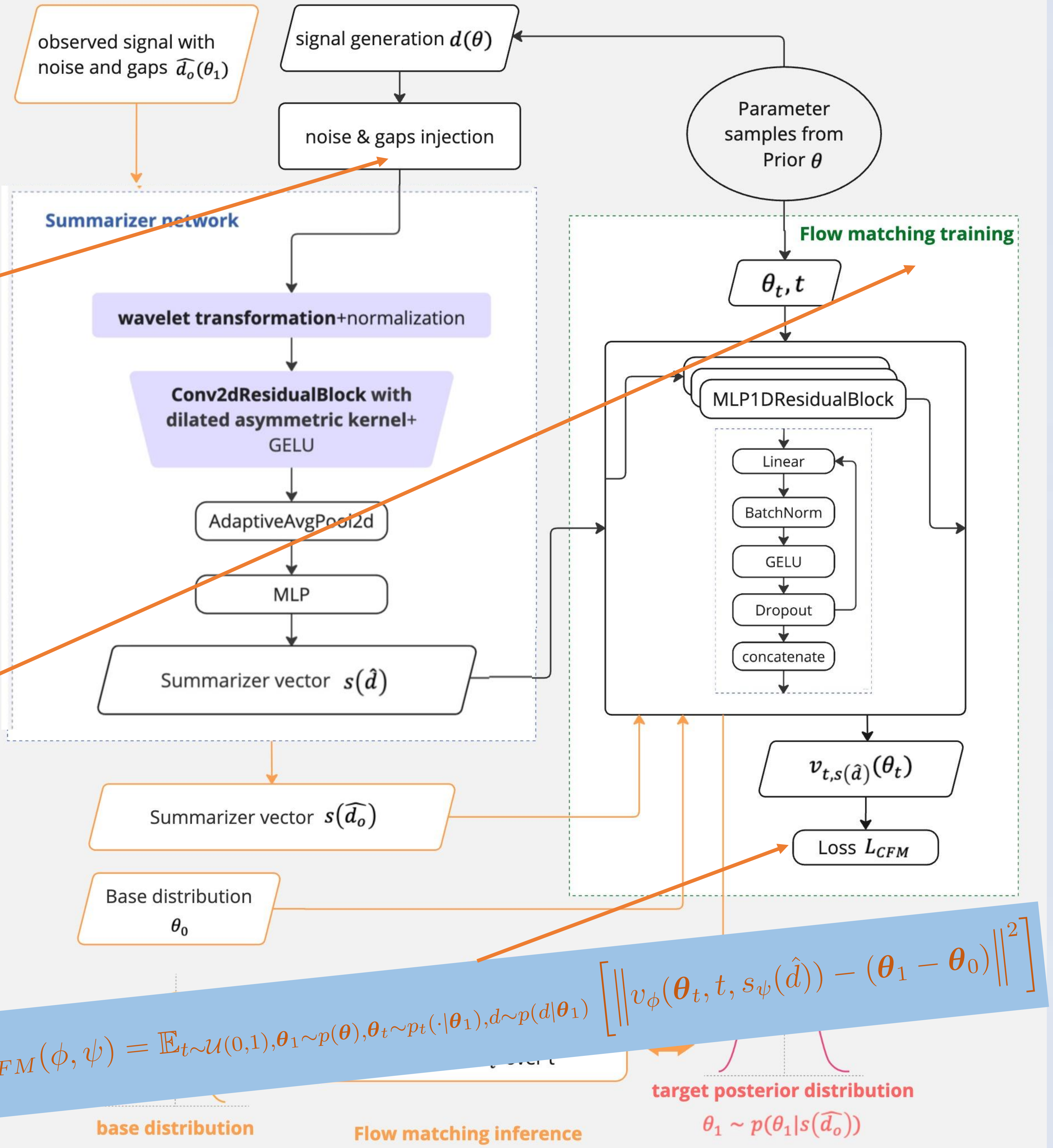
✗ **Symmetric Models (3×3)**

Both symmetric models (green, orange) are not calibrated.

Here is our model:







$$\mathcal{L}_{CFM}(\phi,\psi) = \mathbb{E}_{t\sim\mathcal{U}(0,1),\boldsymbol{\theta}_1\sim p(\boldsymbol{\theta}),\boldsymbol{\theta}_t\sim p_t(\cdot|\boldsymbol{\theta}_1),d\sim p(d|\boldsymbol{\theta}_1)}\left[\left\|v_\phi(\boldsymbol{\theta}_t,t,s_\psi(\hat{d}))-(\boldsymbol{\theta}_1-\boldsymbol{\theta}_0)\right\|^2\right]$$

On the Way...

01

Sufficient summarizer

How to measure the sufficiency of the summarizer

02

Multiple GB signals

Incorporating multiple overlapping GB signals.

03

Training efficiency

Package for flow matching based on wavelet transformation in Jax

Thanks for listening!

[FM vs MAF experiment results]

[One vs Two stage experiment results]