# Evaluating Summary Statistics with Mutual Information for Cosmological Inference

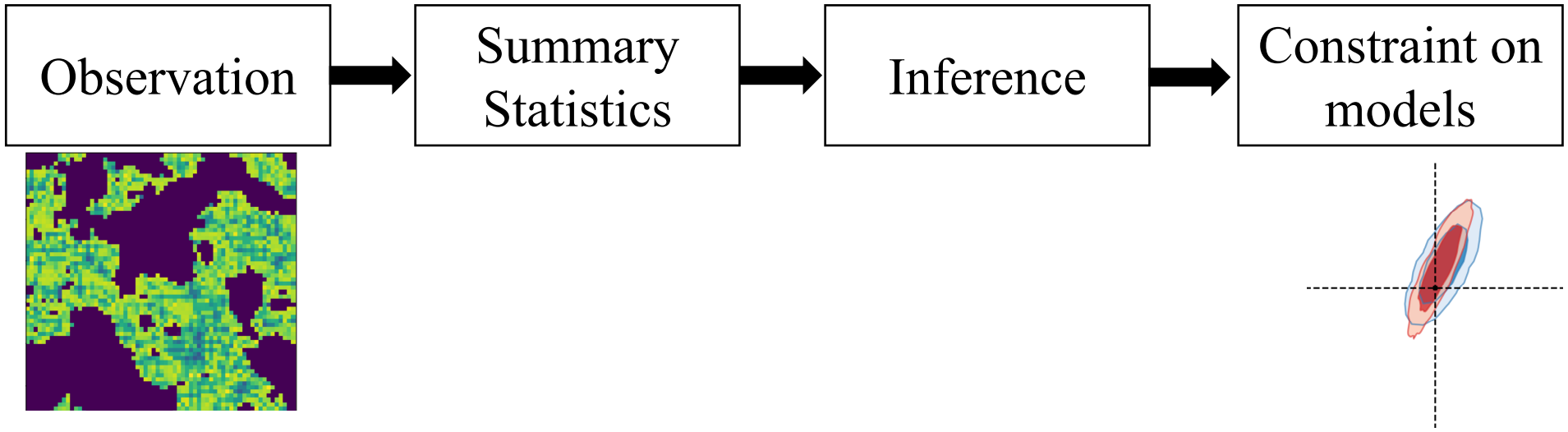
睦策　Ce Sui

In collaboration with

Xiaosheng Zhao, Tao Jing, Yi Mao

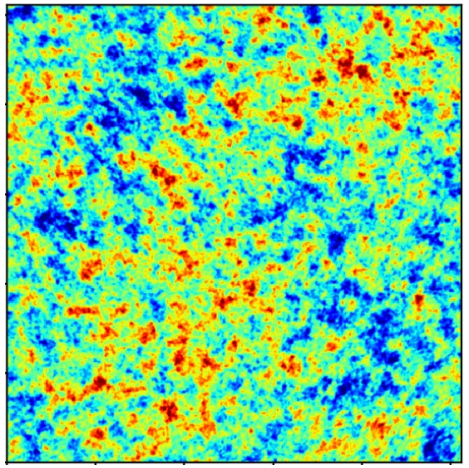
Tsinghua University

# Statistical Inference in Cosmology



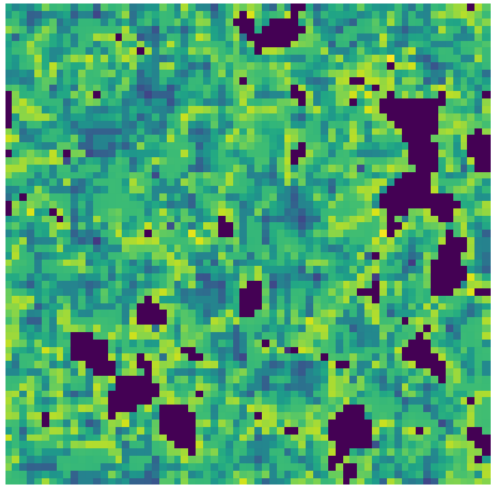
# How to select optimal statistics ?

CMB



→ Power Spectrum

21cm

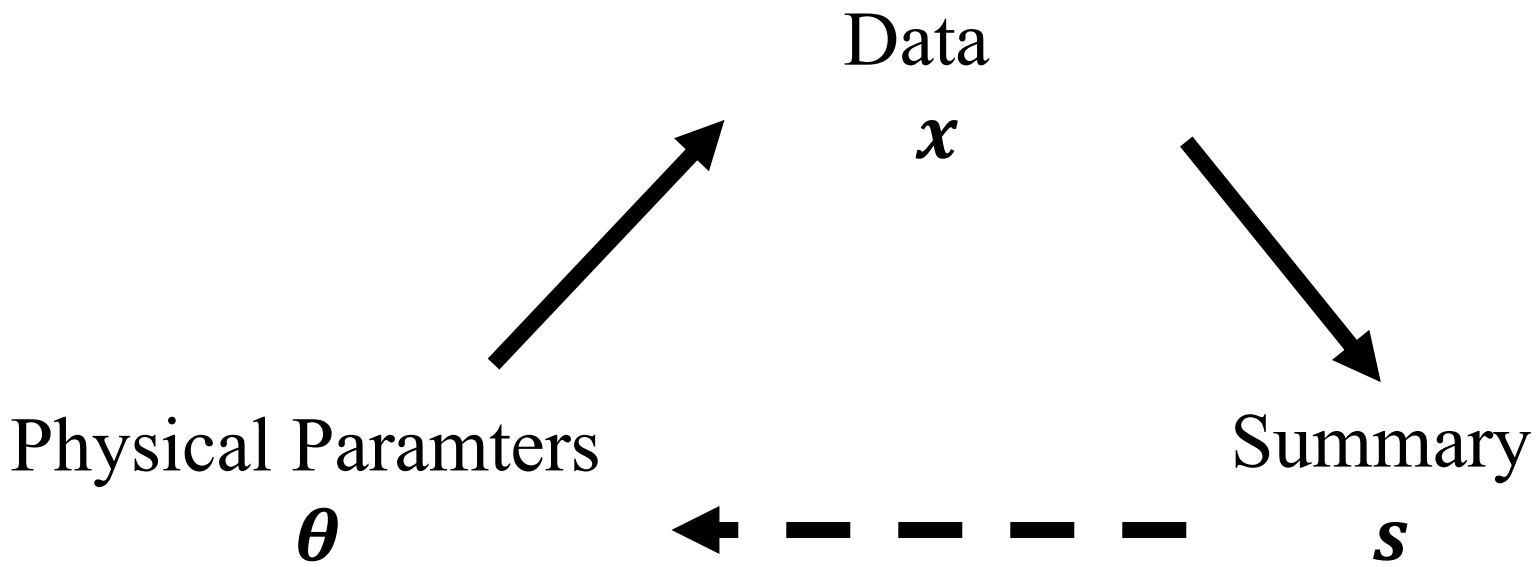


Spatial Correlation → Power Spectrum, Bispectrum

Morphological Properties → Minkowski Functionals / Bubble Size Distribution

Signal decomposition → Wavelet Transform

# How to select optimal statistics ?




1. Through theoretic interpretation

2. Fisher Analysis

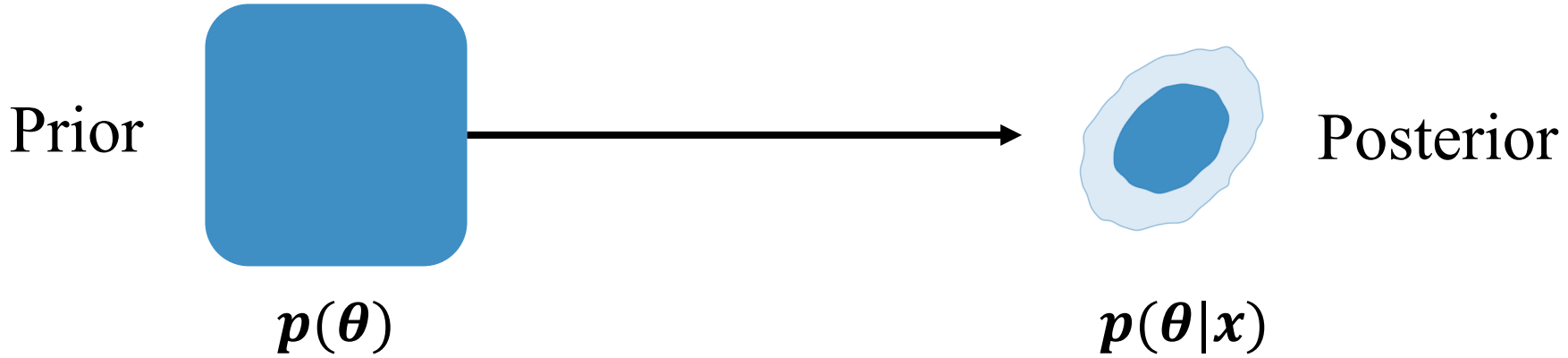
3. Running inferences with one mock observation

# Method: Mutual information as a probe of Sufficient Statistics

**Mutual information**

$$I(\theta; x) = \boldsymbol{D_{KL}}[p(\theta, x)||p(\theta)p(x)] = E_{p(\theta,x)}\left[\log\frac{p(\theta, x)}{p(\theta)p(x)}\right]$$

MI is a fundamental measure of Statistical Dependence




$$I(\theta; x) = E_{p(x)}[\boldsymbol{D_{KL}}(p(\theta|x)||p(\theta)]$$

It evaluates how much uncertainty is reduced

# Method: Mutual information as a probe of Sufficient Statistics

**Mutual information**

$$I(\theta; x) = \boldsymbol{D_{KL}}[p(\theta, x)||p(\theta)p(x)] = E_{p(\theta,x)}\left[\log\frac{p(\theta, x)}{p(\theta)p(x)}\right]$$

**Why mutual information?**

Sufficient Statistics $\longleftrightarrow$ $P(\theta|x, s) = P(\theta|s)$ $\longleftrightarrow$ $I(\theta; s) = I(\theta; x)$

MI reflects “proximity” to sufficiency

# Method: Mutual information as a probe of Sufficient Statistics

**Mutual information**

$$I(\theta; x) = \boldsymbol{D_{KL}}[p(\theta, x)||p(\theta)p(x)] = E_{p(\theta,x)}\left[\log\frac{p(\theta, x)}{p(\theta)p(x)}\right]$$

**Why mutual information?**

$$\text{Sufficient Statistics} \longleftrightarrow P(\theta|x, s) = P(\theta|s) \longleftrightarrow I(\theta; s) = I(\theta; x)$$

**How to estimate mutual information?**

$$I(\theta; x) \equiv E_{p(\theta,x)}\left[\log\frac{p(\theta|x)}{p(\theta)}\right] \approx E_{p(\theta,x)}\left[\log\frac{\boldsymbol{q(\theta|x)}}{p(\theta)}\right]$$

Variational Lower Bound

# Summary Statistics Considered in this work

**Power Spectrum** $\langle \delta(k)\delta(k') \rangle = (2\pi)^3 \delta^D(k+k')P(k)$

**Bipectrum** $\langle \delta(k_1)\delta(k_2)\delta(k_3) \rangle = (2\pi)^3 \delta^D(k_1+k_2+k_3)B(k_1,k_2,k_3)$

**Scattering Transform** $S\delta[\lambda_1,\dots,\lambda_k] = \left|\psi_{\lambda_k} \star \cdots |\psi_1 \star \delta|\right|$

Use designed wavelets to convolve the field

(Anden & Mallat 2014, Eickenberg et al. 2017,2018)

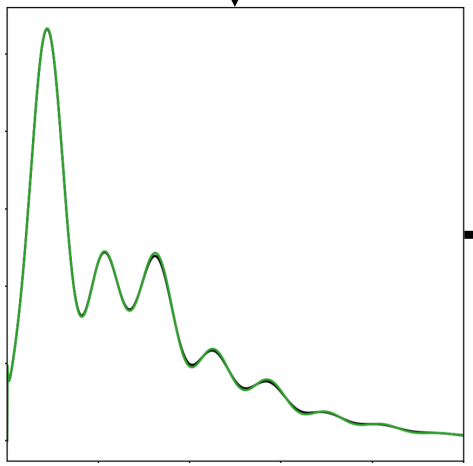
# Experiments I:

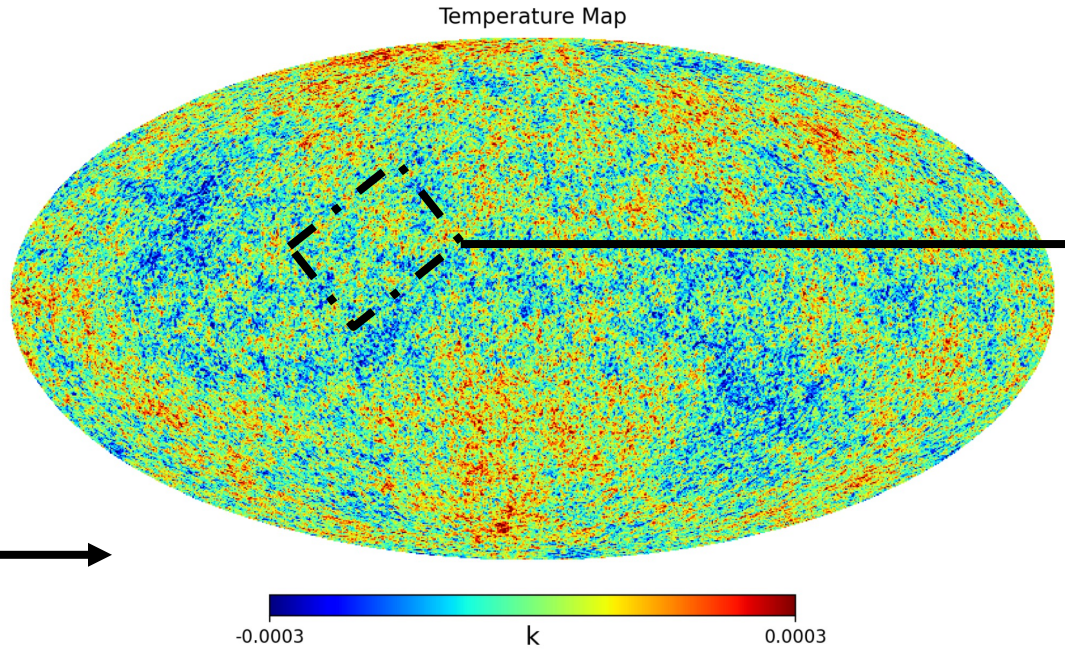
Validate the method in CMB-like Gaussian random fields(GRF)

$\boldsymbol{\Omega_c \in [0.21, 0.31]}$

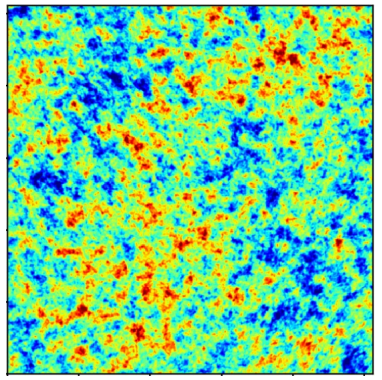
$\boldsymbol{h \in [0.64, 0.76]}$

Use ***CAMB*** to generate Power Spectral





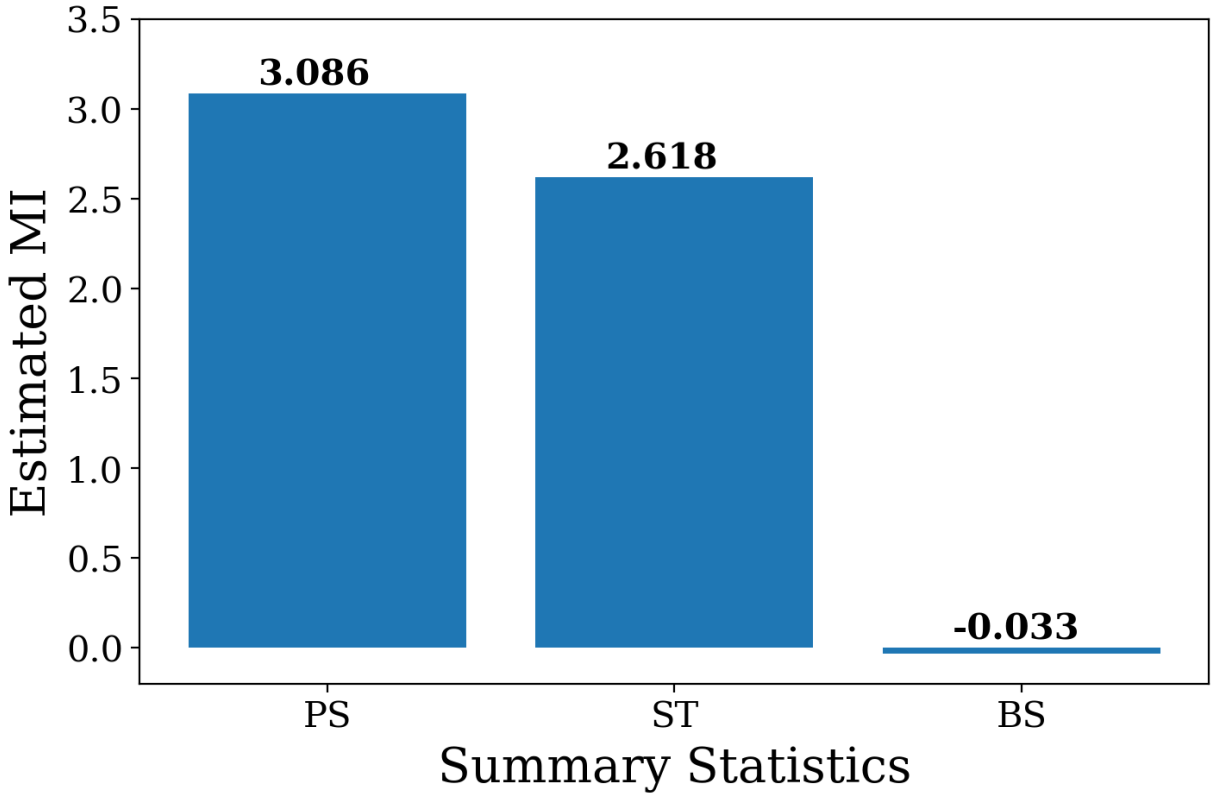

Use ***Healpy*** to generate CMB-like GRF



Patch for experiments

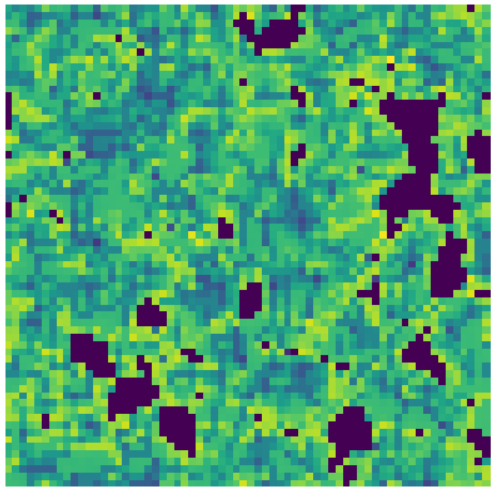
# Experiments I:

Validate the method in CMB-like Gaussian random fields(GRF)



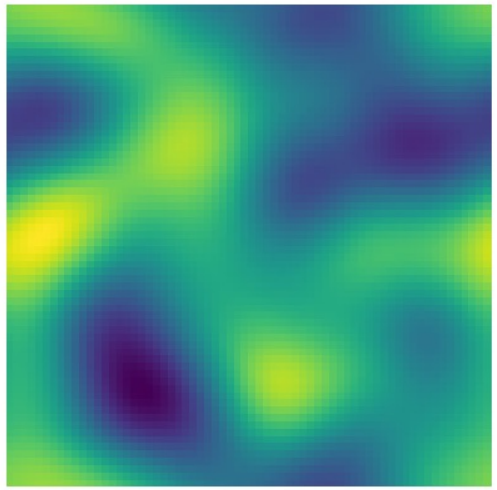

Results are consistent with theoretical interpretations

# Experiments II:
# Evaluate Statistics in an EoR inference Task

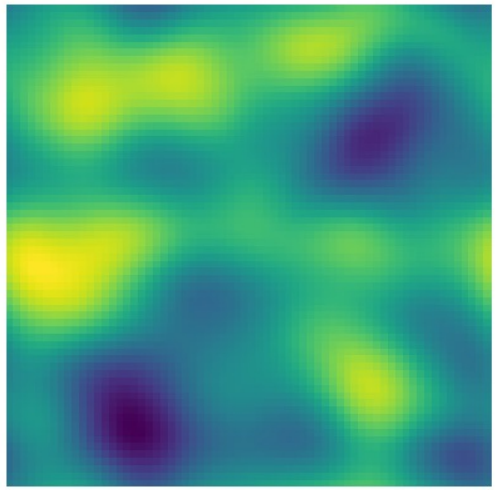


**Pure 21cm signal (simulated by 21cmFAST)**

Mesinger & Furlanetto (2007)



**+Thermal Noise (SKA1-low configuration)**



**+Residual Foreground (GSM model +SVD)**

Zheng et al. (2017)

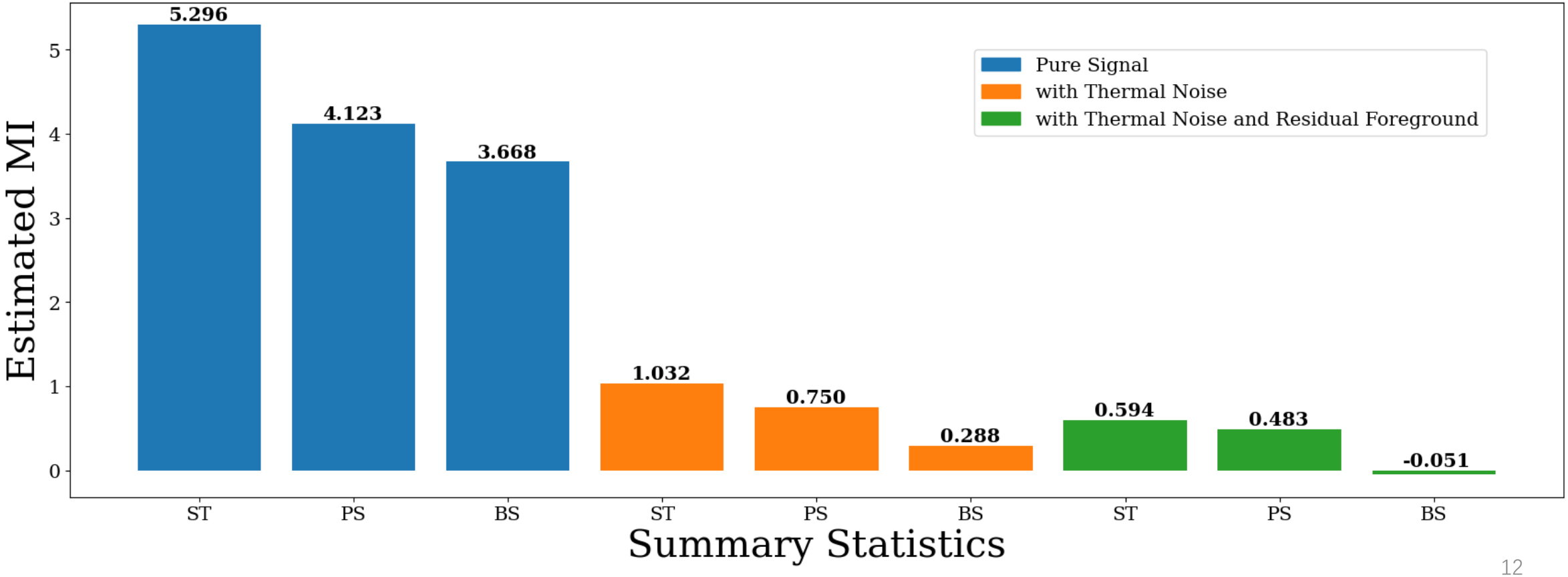
**Reionization Parameters**

$\zeta$ the ionizing efficiency

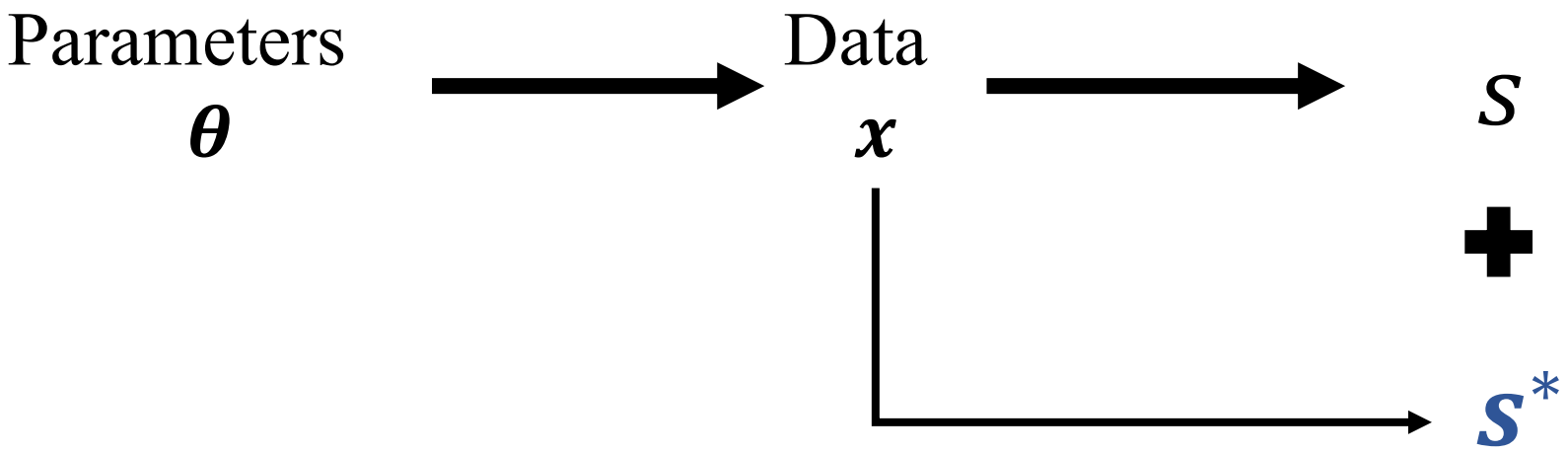
$T_{vir}$ the minimum virial temperature of halos that host ionizing sources

# Experiments: MI-based comparison of statistics in an EoR Inference task

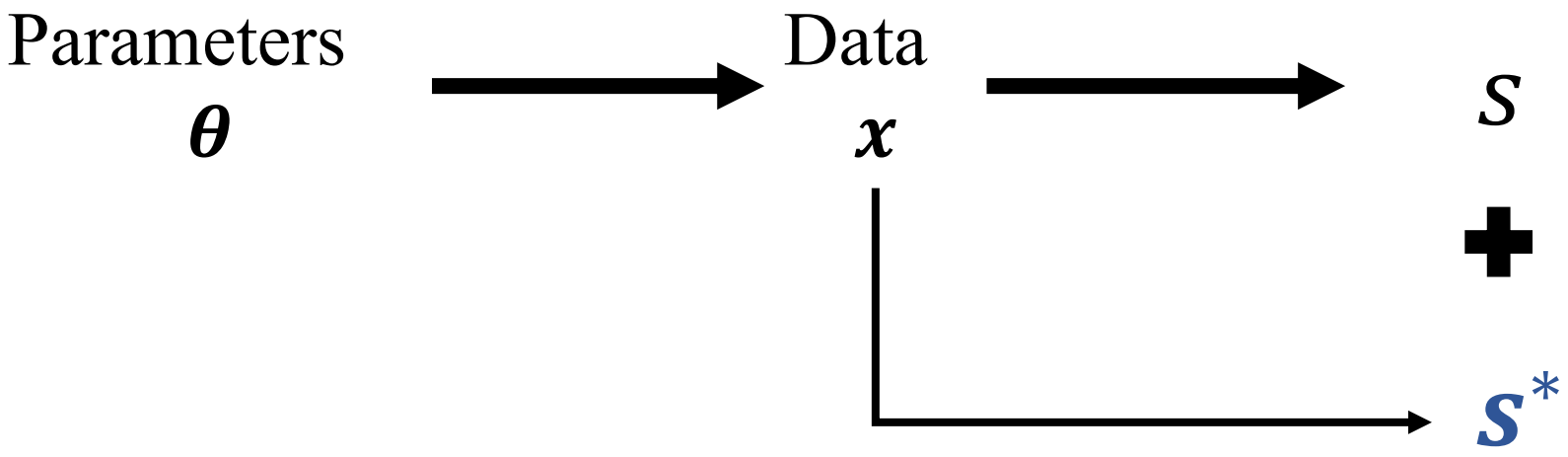
**ST:** Wavelet Transform **PS:** Power spectrum **BS:** Bispectrum



# Mutual Information for evaluating Complementary Summary

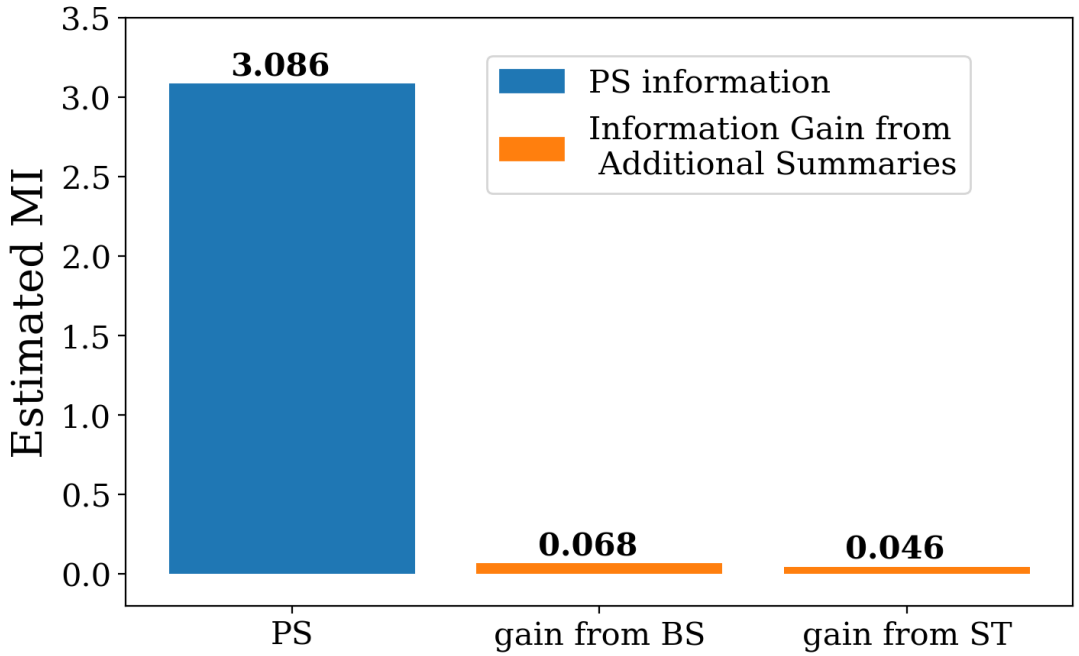


# Mutual Information for evaluating Complementary Summary

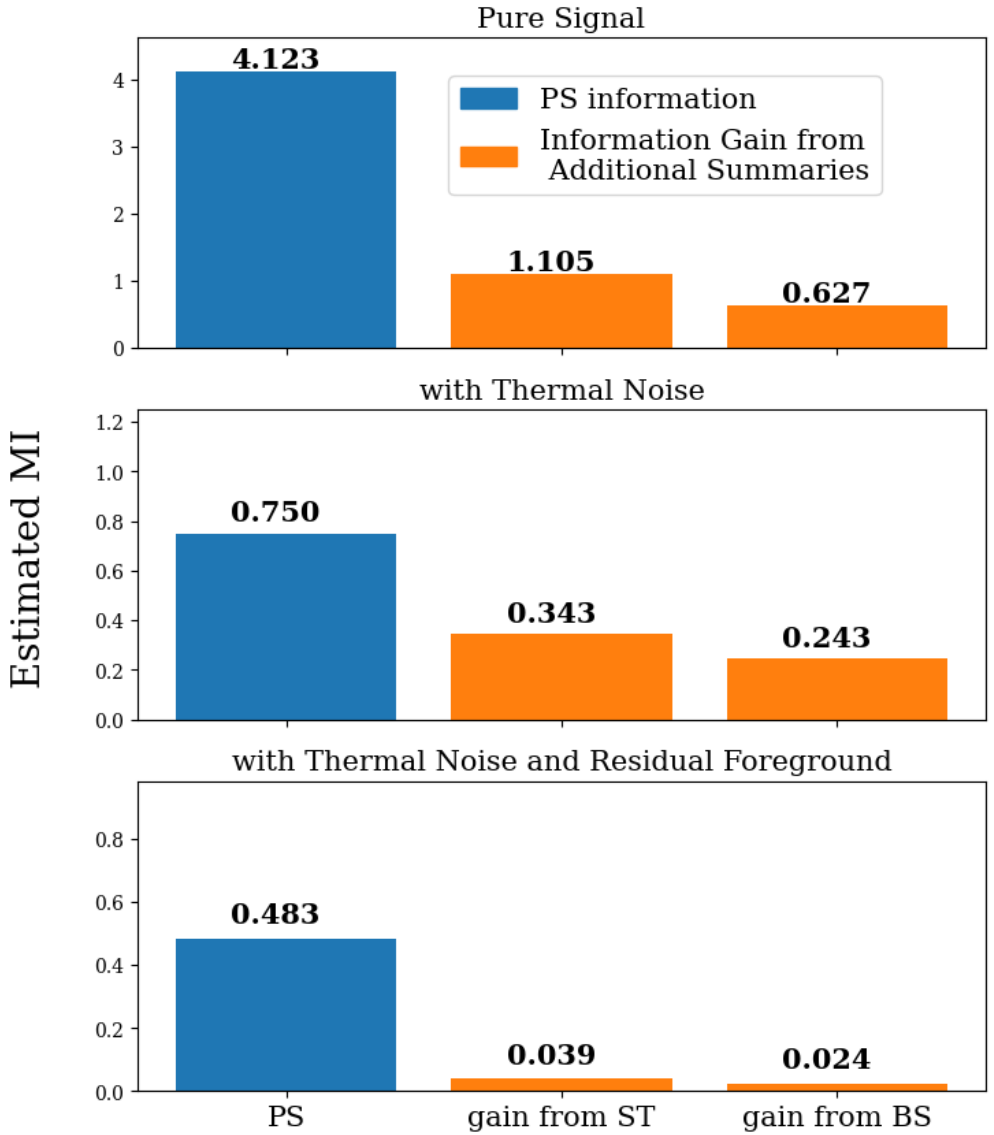



Conditional Mutual Information: $I(\theta; S^* | S)$

# Mutual Information for evaluating Complementary Summary




**CMB Experiment**




**21cm Experiment**

# Mutual Information for Learning Summaries

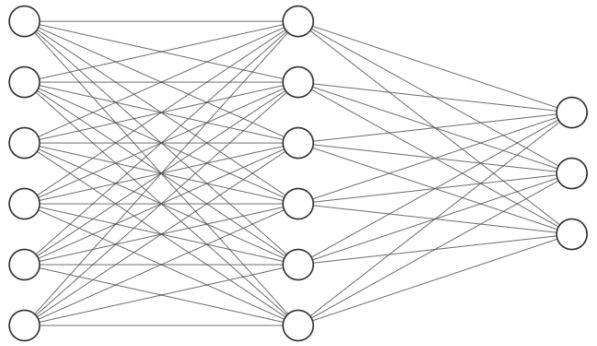
1. Learn a new optimal summary by maximizing mutual information

$$\max_{S} I(\theta; S(x))$$

Parameters
$\boldsymbol{\theta}$



Data
$\boldsymbol{x}$



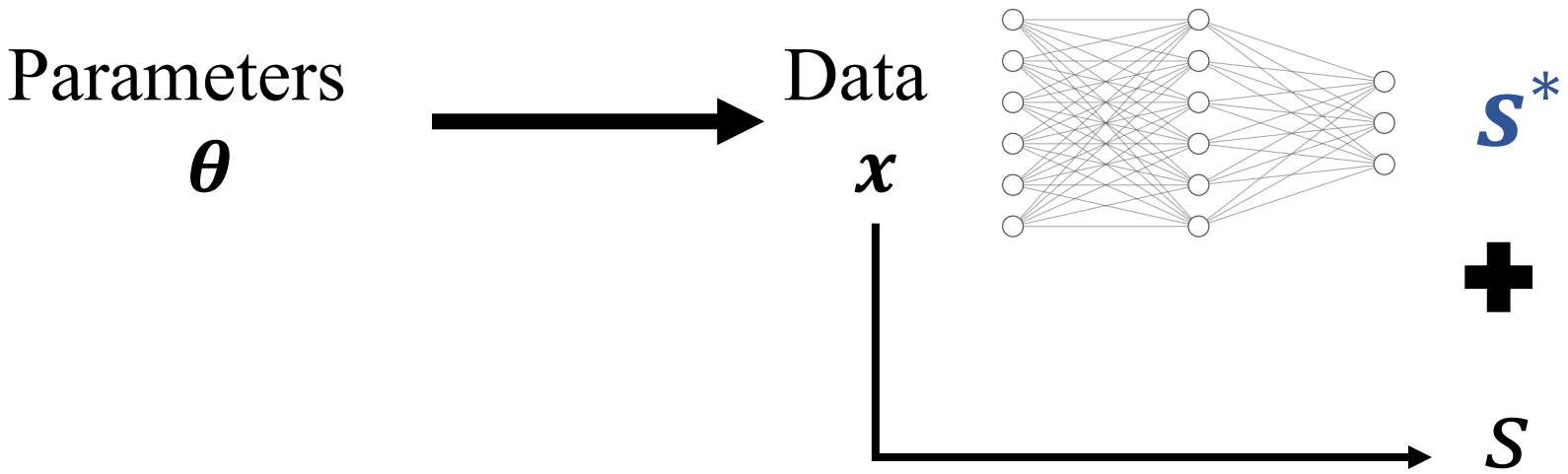
Summary
$\boldsymbol{s}$

**Compressor Neural Network**
**S**

# Mutual Information for Learning Summaries

2. Learn a complementary summary by maximizing conditional mutual information

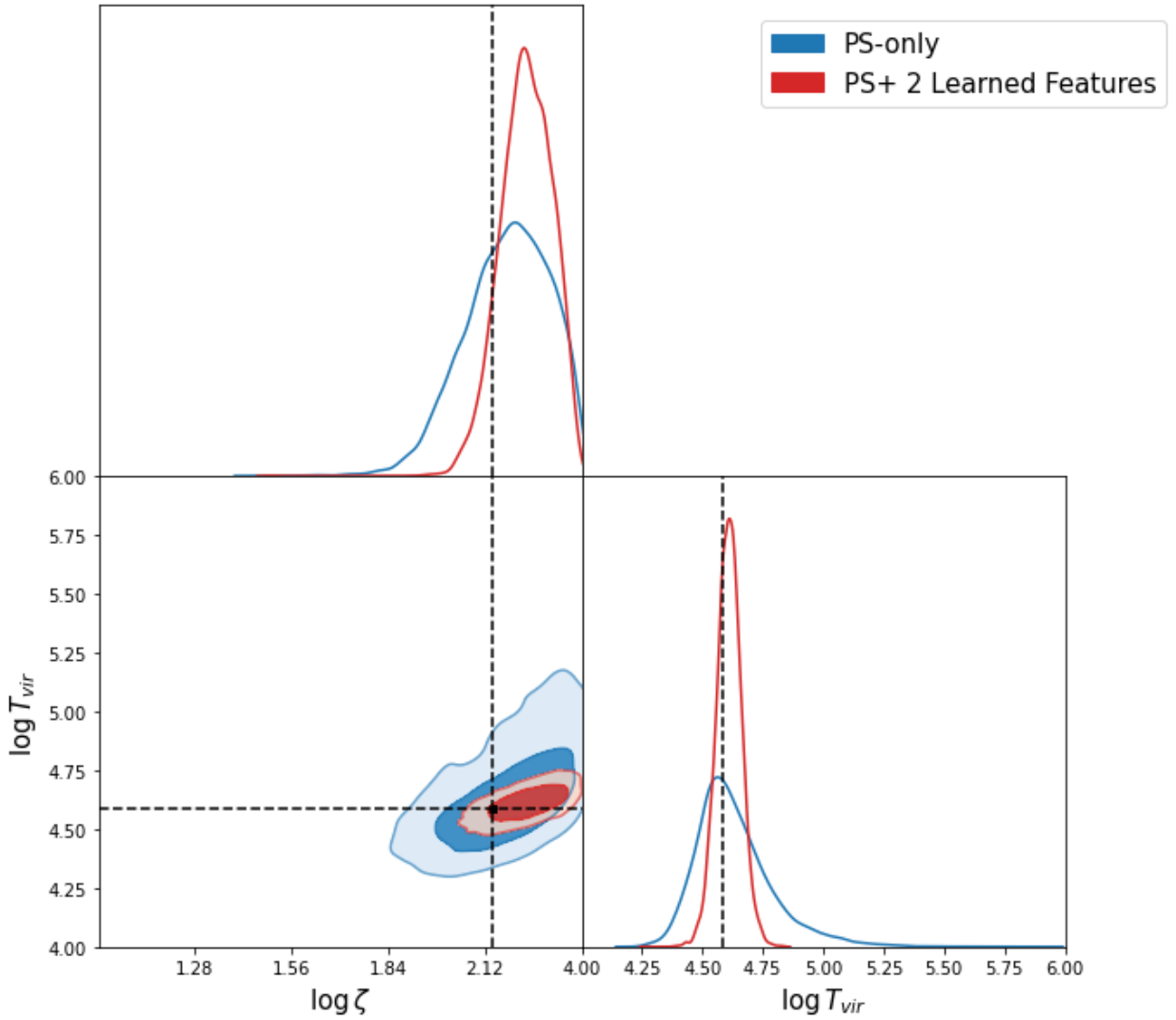
$$\max_{S} I(\theta; S^*(x) | S)$$



2. Learn a complementary summary by maximizing conditional mutual information

**Learn two complementary features for Power Spectrum**

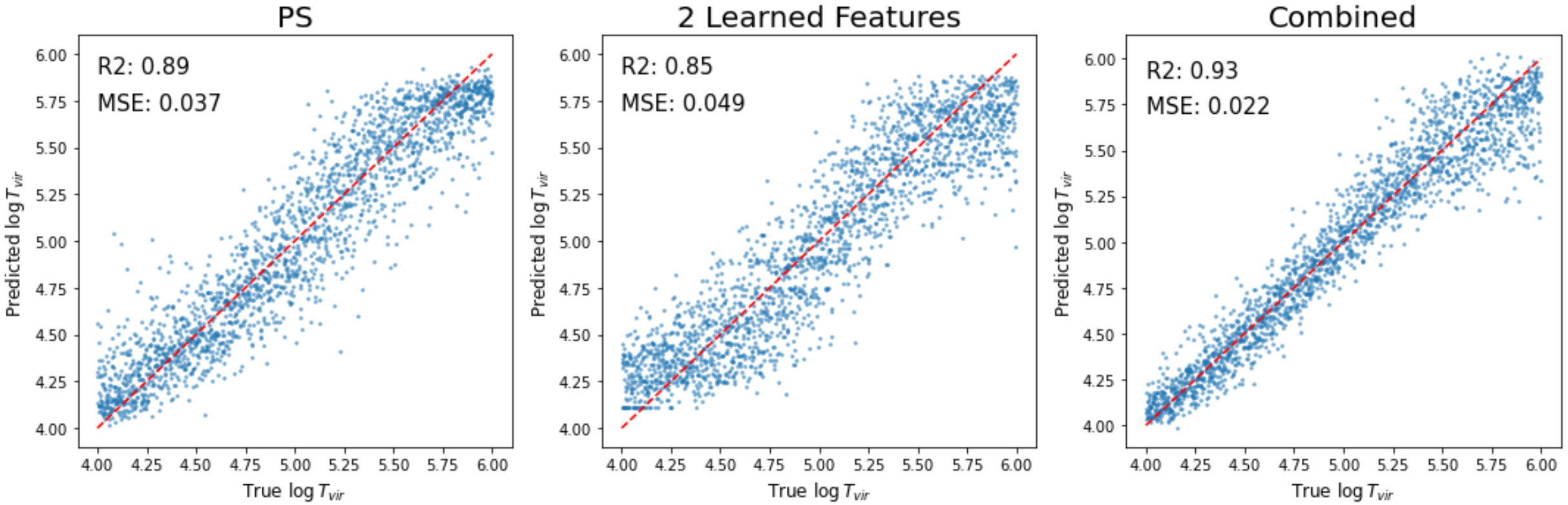
**Experiment:**
Infer Reionization Parameter from 21cm images.




**Preliminary results, Paper in Prep**

# 2. Learn a complementary summary by maximizing conditional mutual information

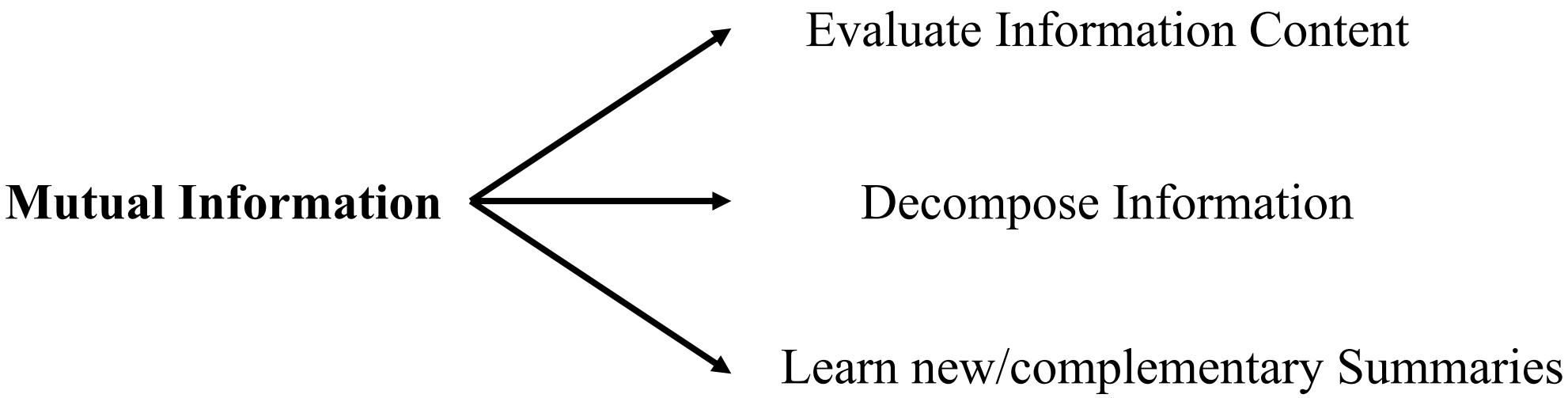
**Preliminary results, Paper in Prep**




## Regression Performance

The learned representations are indeed complementary to power spectrum

# Summary




**More Details:**

Previous conference paper: https://arxiv.org/abs/2307.04994

Codes and experiments: https://github.com/suicee/MI4StatsEval

Journal Paper in prep