

データサイエンスが切り拓く天文学の未来

統計数理研究所

池田 忠朗

自己紹介

データ科学について

天文分野での実践例

EHTでは

Weak Lensing Mass Map

まとめ

自己紹介

1996.3. 東大 計数工学専攻 修士(工学)

～2001.3 理研 (PD)

～2003.1 九州工業大学

～現在 統計数理研究所

情報幾何, 統計学, 情報理論, 信号処理

2013年頃から天文データ解析,

EHTC メンバー, 天文台連携教員, Kavli IPMU 訪問研究員

統計科学（データ科学）を専門に行う研究所

- 創立 1944
- 2010年、港区南麻布から立川市に移転.
- 22の大学共同利用機関のひとつ.
- 教員50人弱, ビジター約80人/年, ポスドク約70人, 院生約30人.

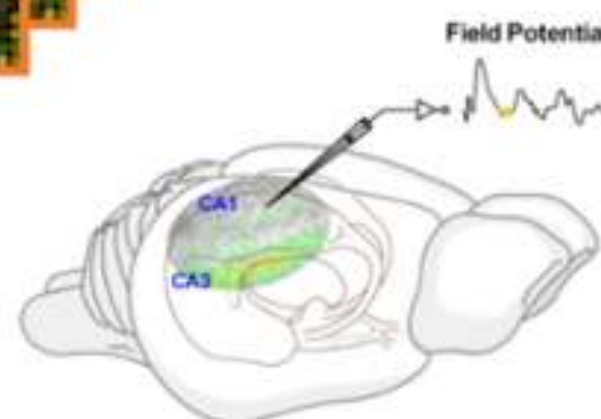
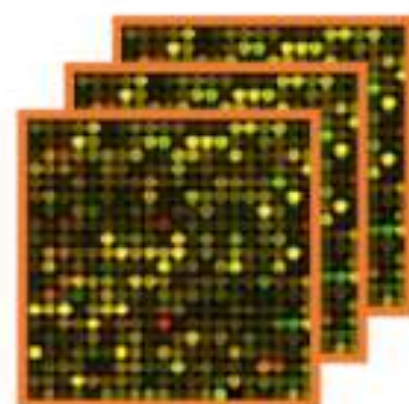
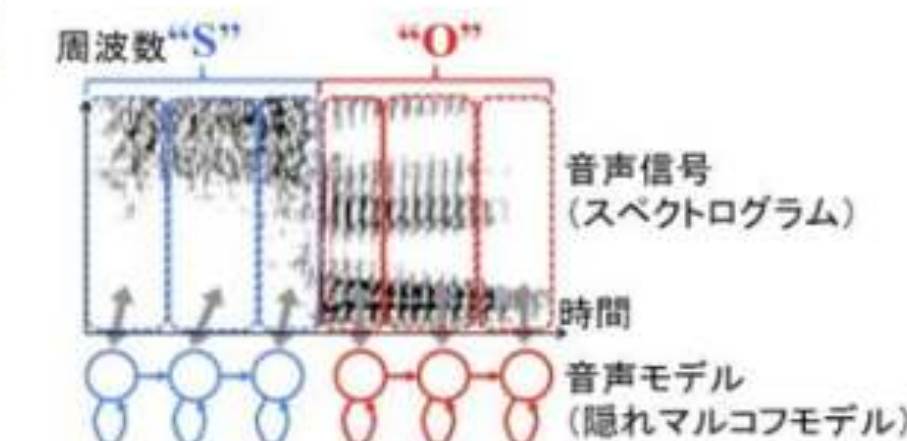
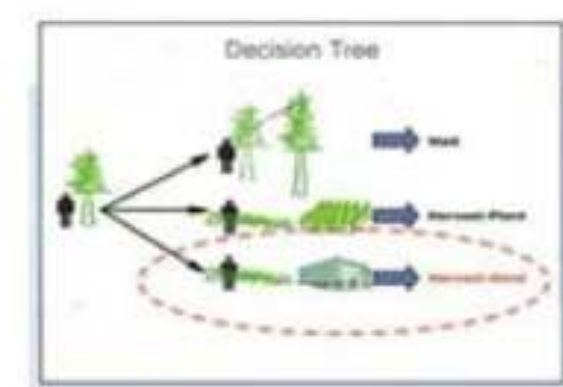
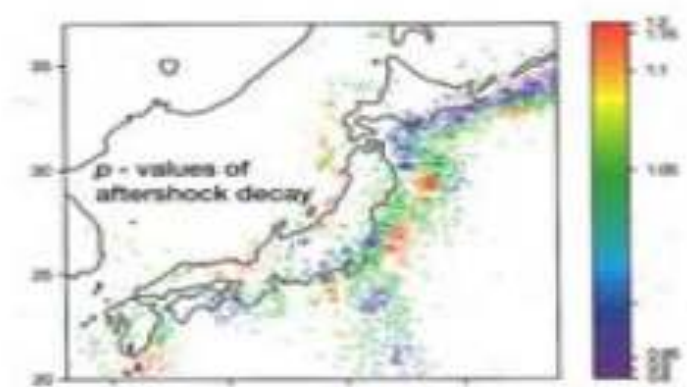
Theory and Methodology

- Information Criterion
- Time series analysis
- Survey methodology
- Quantification theory
- Multivariate analysis

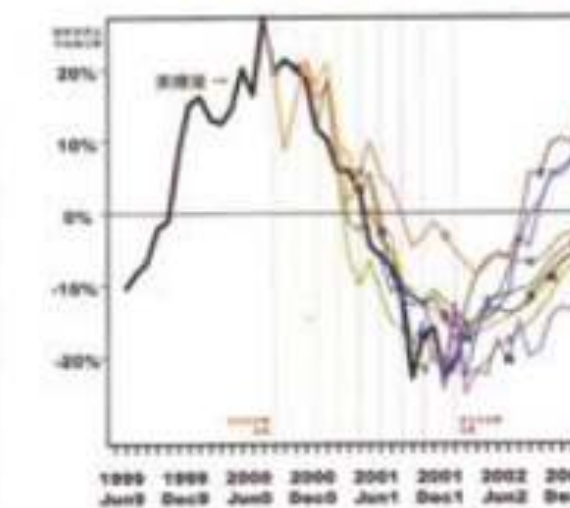
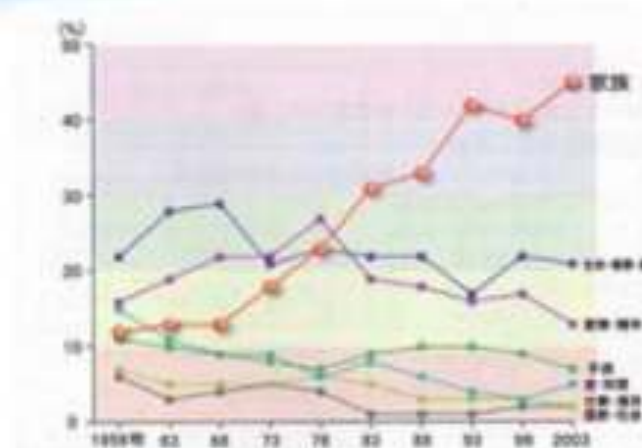
- MCMC, particle filter
- Ensemble machine
- Sparse modeling
- Kernel method
- Meta-analysis
- Bayesian modeling
- Data assimilation method



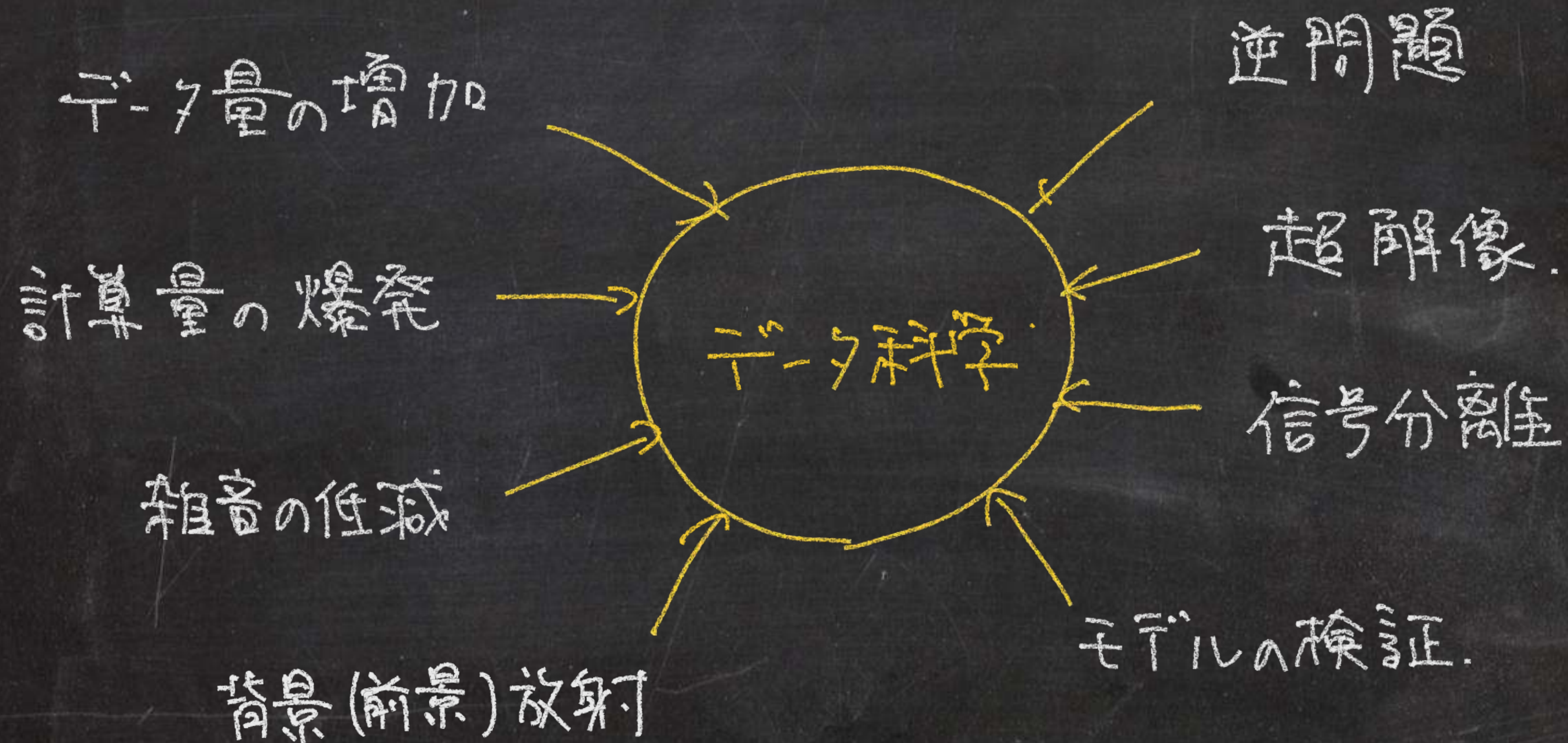
データ科学の研究と 普及・教育



ランダム化臨床試験のデザインの一例



天文学との関わり



自己紹介

データ科学について

天文分野での実践例

EHTでは

Weak Lensing Mass Map

まとめ

データ科学

自然科学

社会科学

データ科学

対象に関する知識を深め、広める

データの処理方法、解析方法に関する知識を深め、広める

データ科学

目的は統計学と変わらない。

Lies, damned lies, and statistics

確率とは何か。

実証可能性

モデルの妥当性

意思決定が必要 $\implies$ 統計学は重要
データ科学



Mark Twain, 1907

データ科学と天文学

観測 $\Longrightarrow$ データ $\Longrightarrow$ 自分のサイエンス

装置開発
(予算)

川原調にいけば

計算資源
データ科学

うまくいけば

誰かがやってくれる

予算?

データ科学と天文学

観測



データ



自分のサイエンス

装置開発
(予算)

計算資源
データ科学

全てを自分のサイエンスの一部として理解する。

自己紹介

データ科学について

天文分野での実践例

EHTでは

Weak Lensing Mass Map

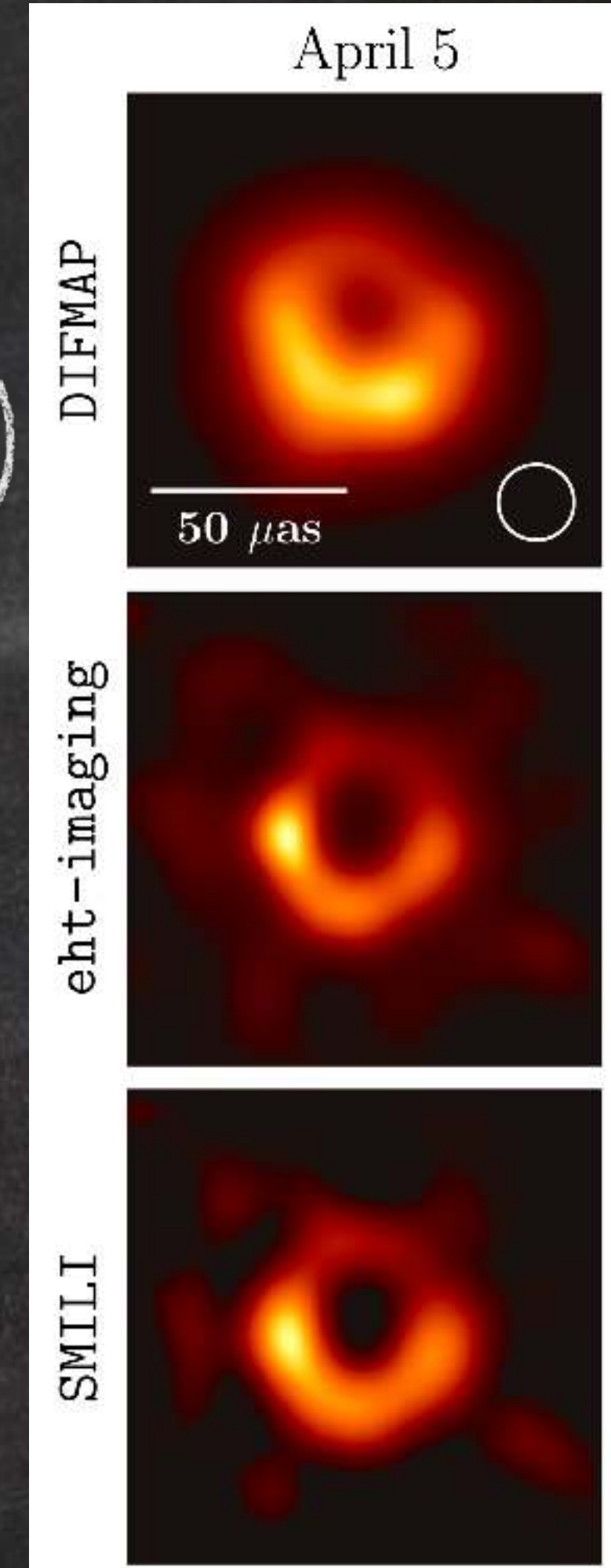
まとめ

EHT に対し M87 のイメージ (2019 年 paper IV)

CLEAN : 伝統的な方法

RML 法 (regularized Maximum Likelihood)

- eht-imaging
- SMILI



EHT による $Sgr A^*$ のイメージ (2022 年 paper III)

CLEAN : 伝統的な方法

RML 法 (regularized Maximum Likelihood)

- eht-imaging

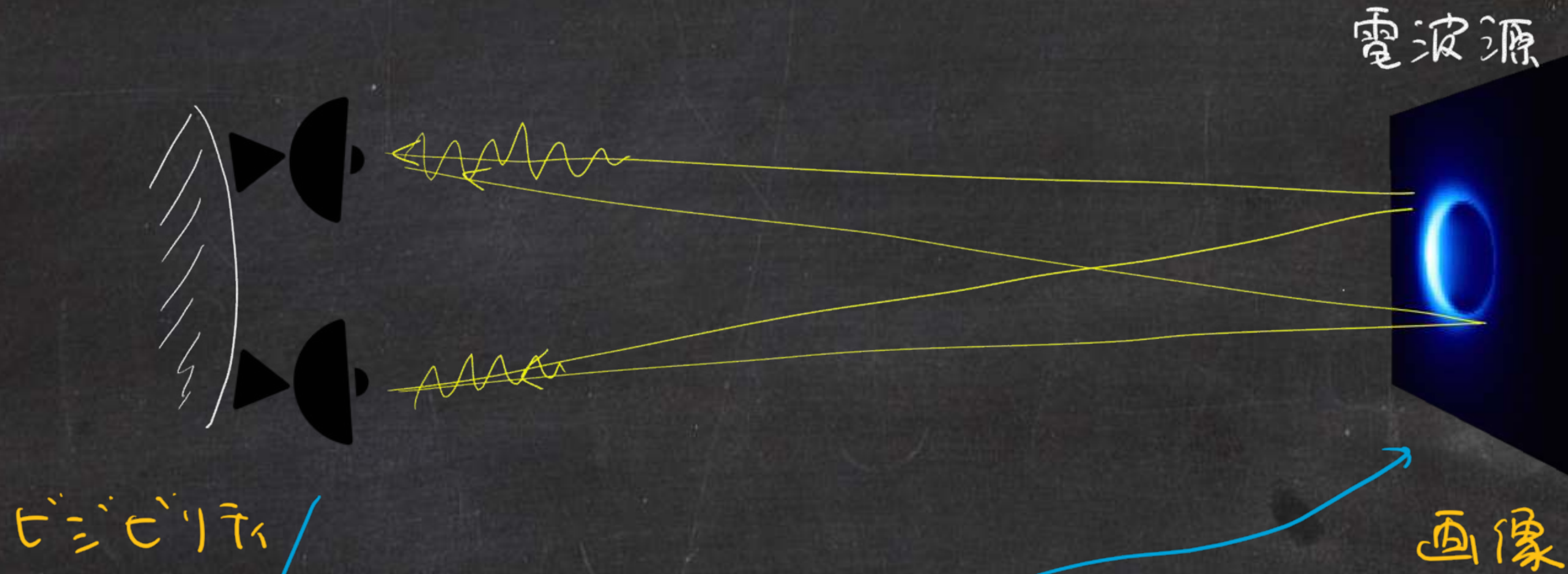
- SMILI

Bayesian 法

- Themis



電波干渉計



$$V(u, v) = \iint I(x, y) e^{2\pi i k (ux + vy)} dx dy$$

計算機上で干渉させ、フーリエ成分を計算する。



電波干渉計

V は I の フーリエ変換 に ノイズが加わったもの.

$p(\underline{V} | \underline{I})$: 観測モデル.

$p(\underline{I})$: 画像に関する仮定.

$p(\underline{I} | \underline{V}) \propto p(\underline{V} | \underline{I}) \cdot p(\underline{I})$ 事後分布.

電波干渉計の「ファージング」法は いづれも Bayes 的に理解できる.



電波干渉計

$$p(\underline{I}|\underline{V}) \propto \phi(\underline{V}|\underline{I}) \cdot p(I) \quad \text{事後分布.}$$

- V が与えられたとき, $p(\underline{I}|\underline{V})$ を最大にする $\underline{I}$ を求める.

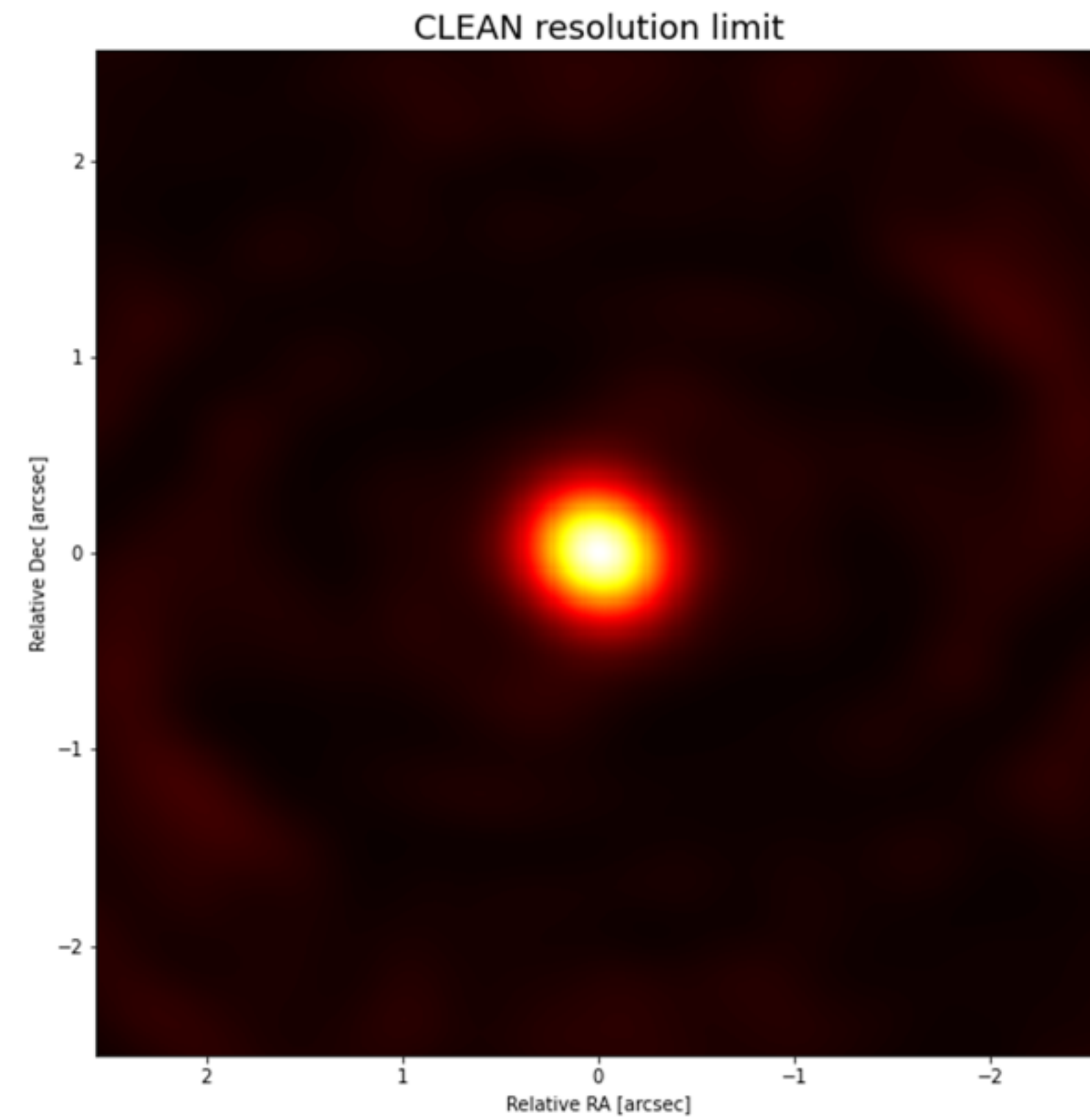
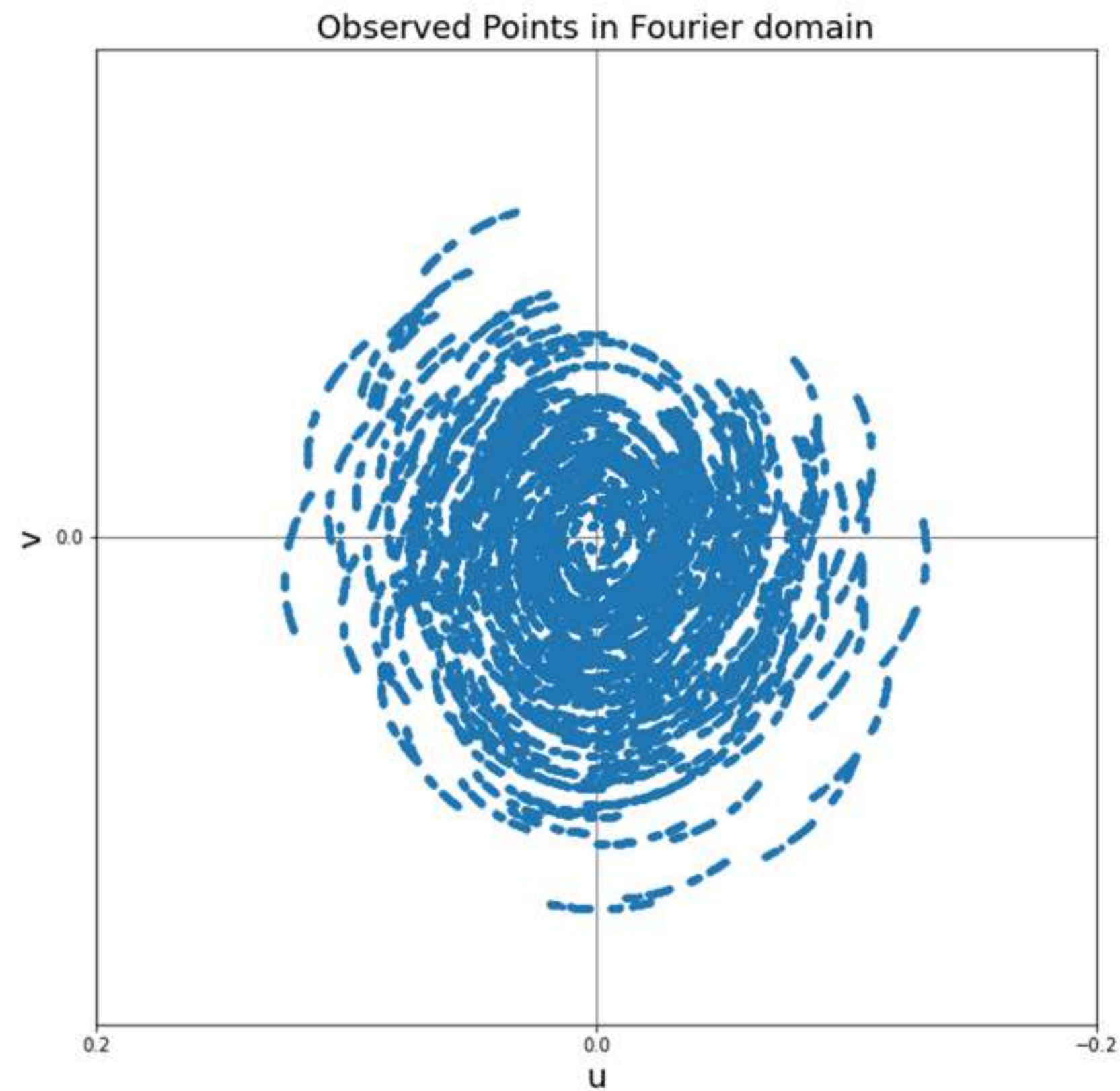
MAP推定 (事後確率最大), CLEAN, RML

- V が与えられたとき $p(\underline{I}|\underline{V})$ の分布自体を評価する.

Bayes 法

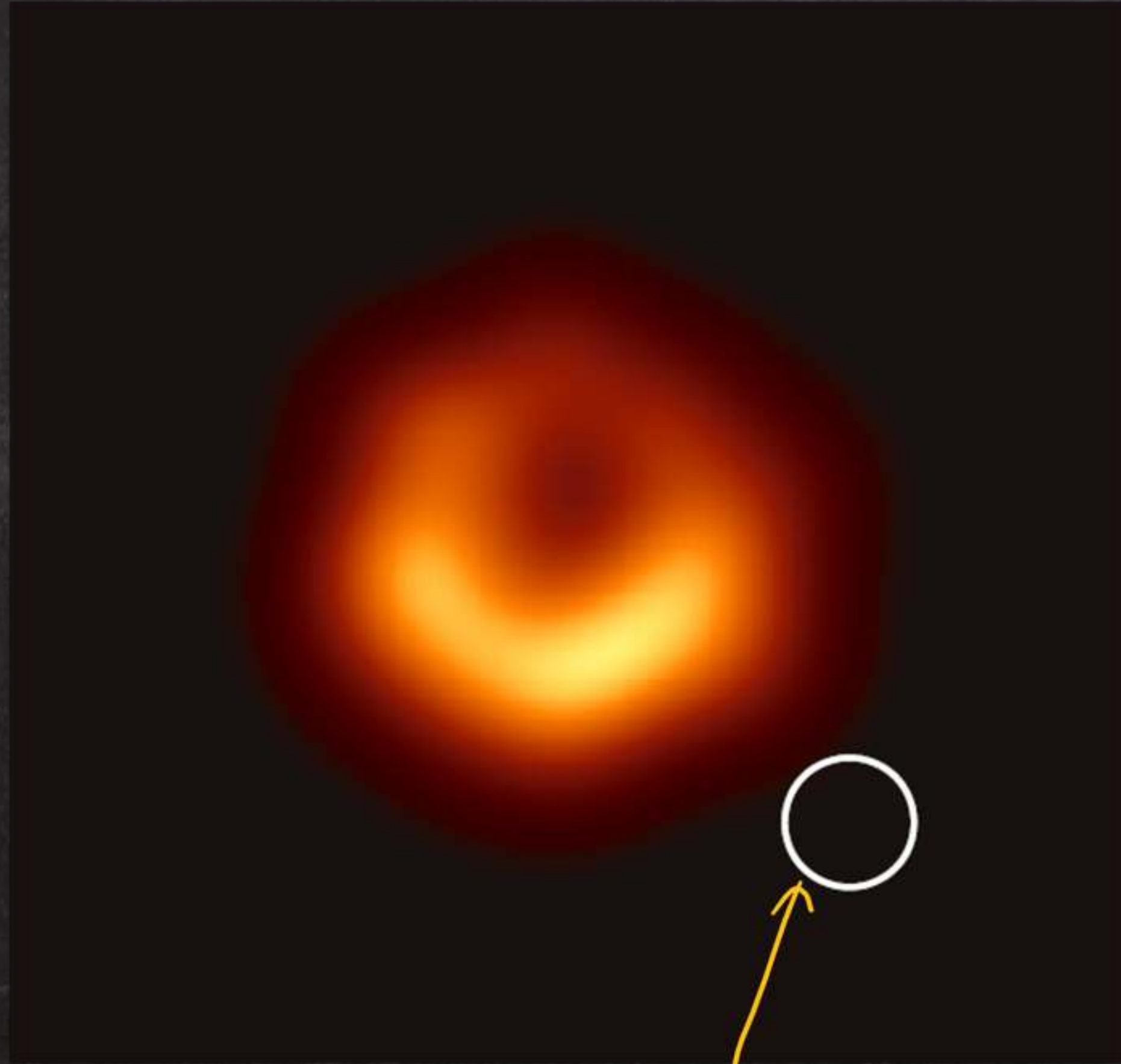


CLEAN法

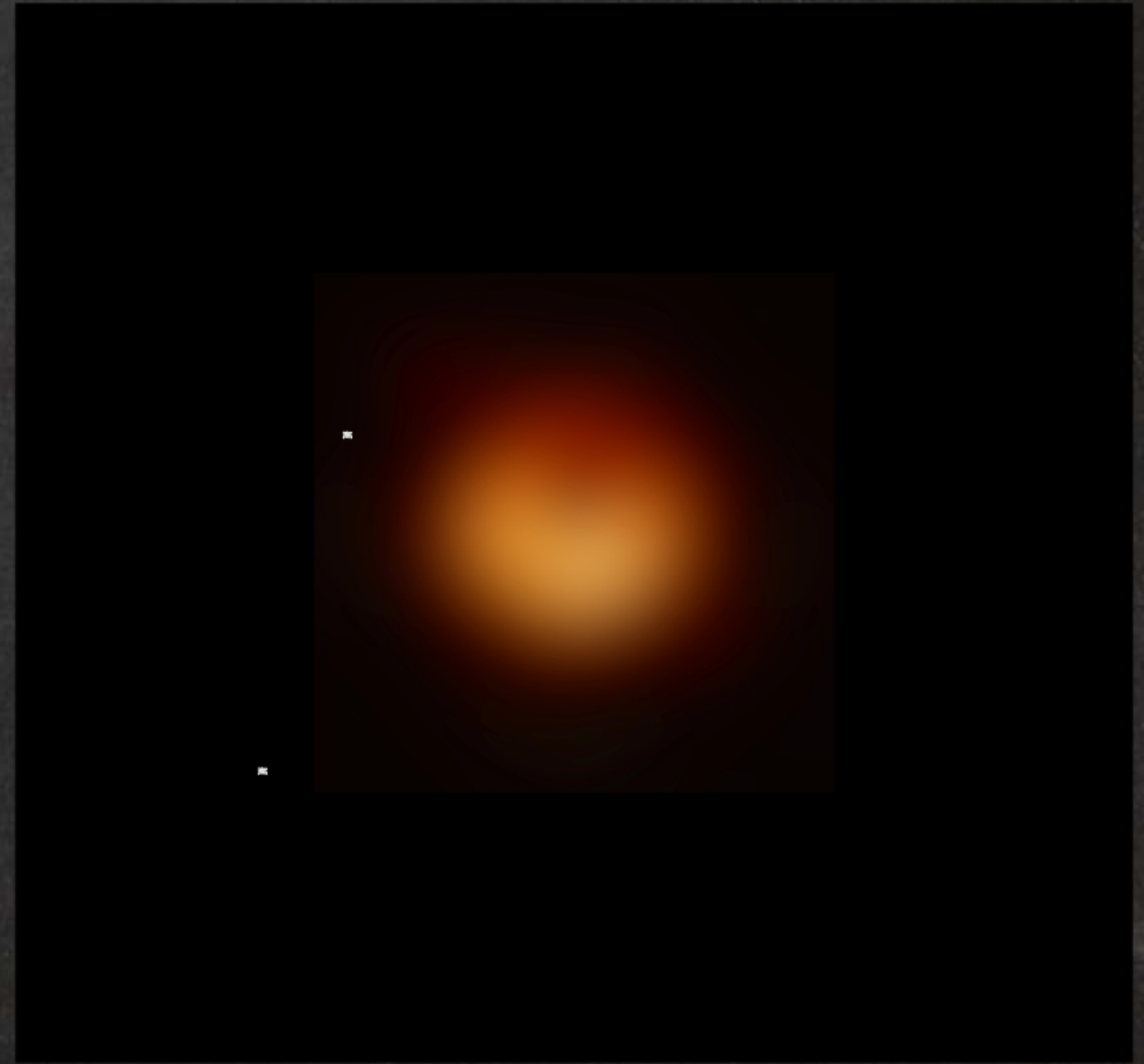


M87の予想したサイズ

6.5×10^9 SM



3.2×10^9 SM



Event Horizon Telescope

2023, 12/8, 仁科記念講演会

CLEAN + self calibration

$P(V|I)$: 観測モデル, 尤度

χ^2 誤差

$P(I)$: 事前分布, 制約条件

光源 = 点源の集合 ゲインの大きさ, 位相

アルゴリズム

Greedy

画像にゲインを合わせ?

データの理解

点源を CLEAN で重ね込む ... 1つの画像



RML (Regularized Maximum Likelihood) 法

$p(V|I)$: 観測モデル, 尤度

χ^2 誤差, Closure 量

$p(I)$: 事前分布, 制約条件

スパース性, 滑らかさ, Max Ent, Total flux

アルゴリズム

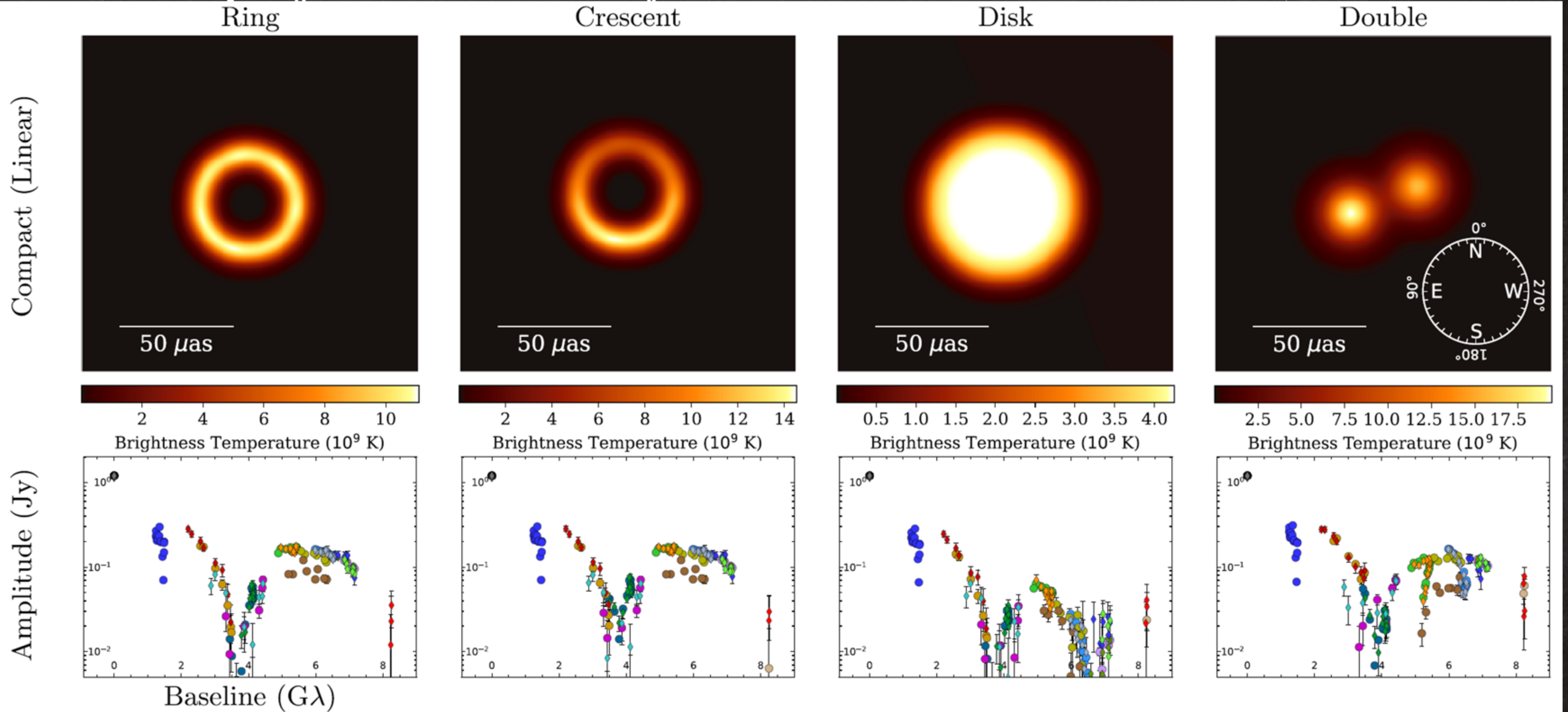
最適化法の導入

データの理解

= 画像化のためのパラメータの理解 $\Rightarrow$ 画像の決め方

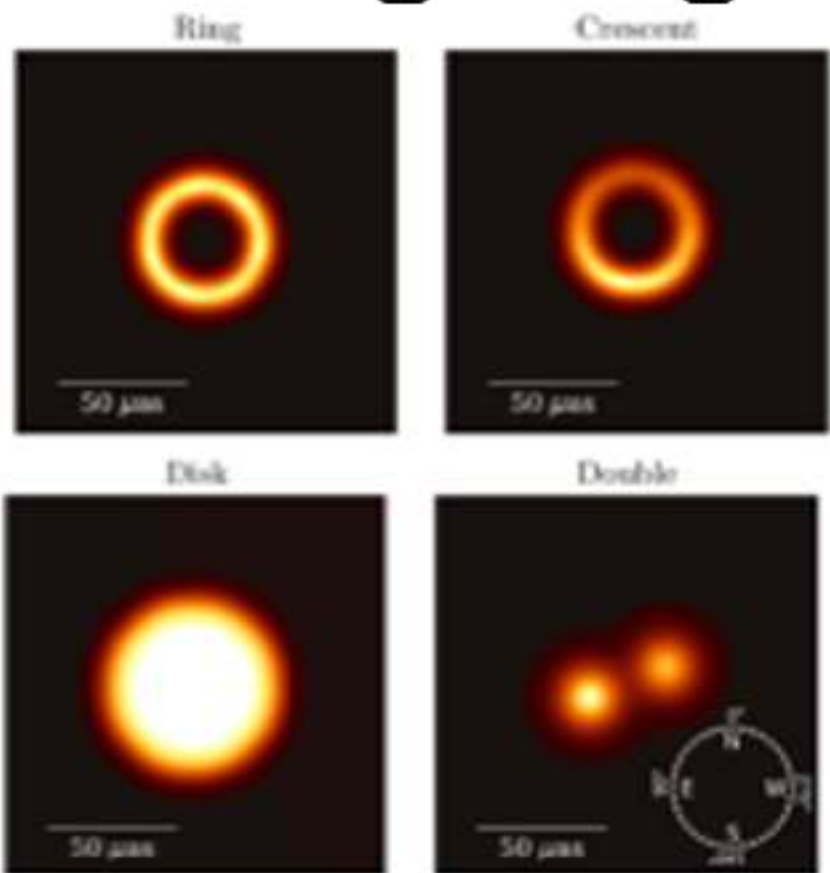


Can imaging methods distinguish these?



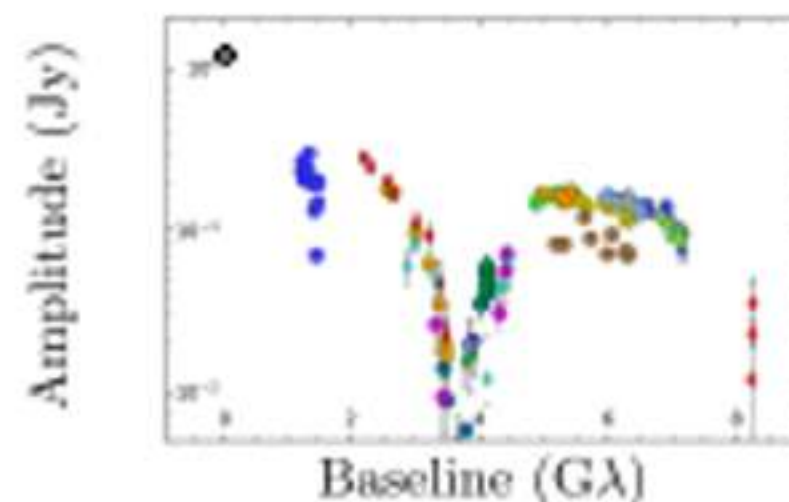
Imaging Process

Training Images



**Synthetic
Observations**

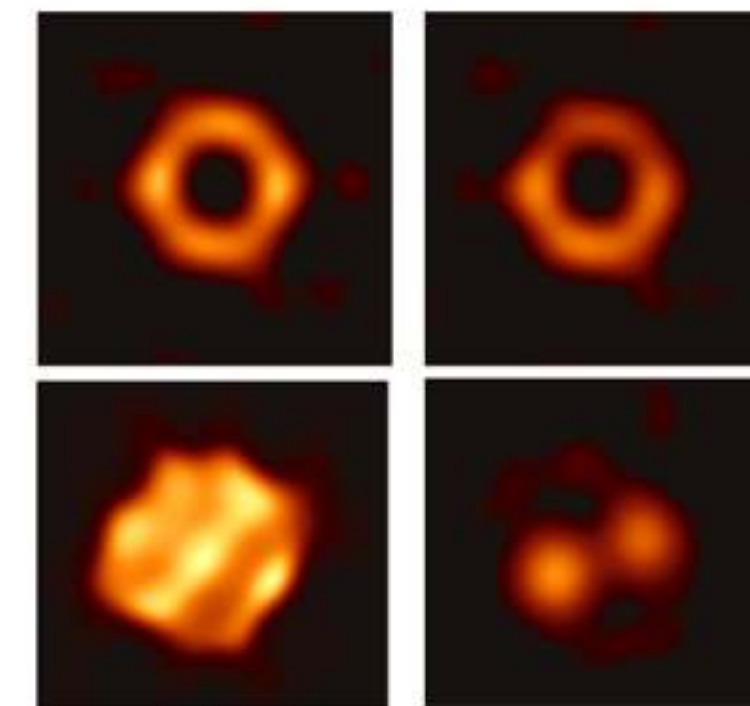
Synthetic Data Sets



**50 k
Parameter
sets**

**3 Imaging
Pipelines**

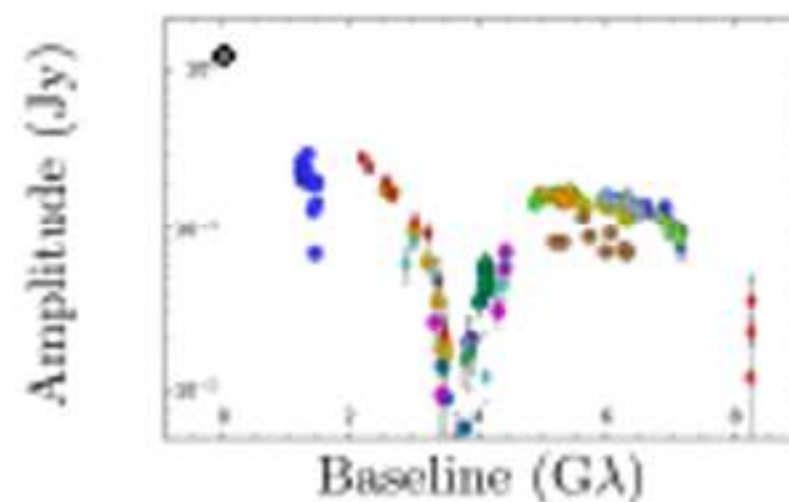
Reconstructed Images



**Fidelity
check**

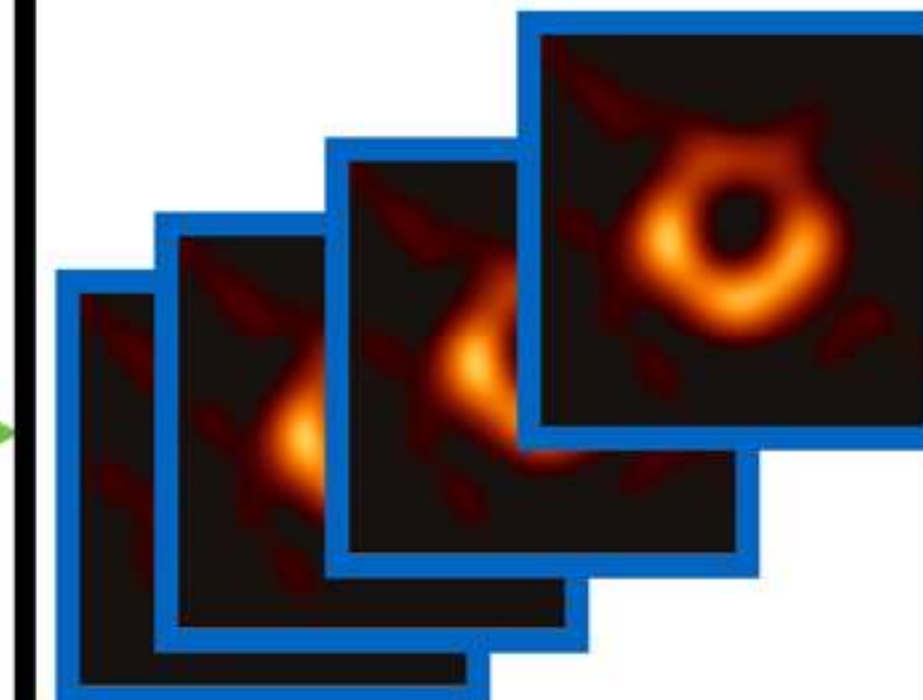
**Optimal
Parameters**

Real M87 Data



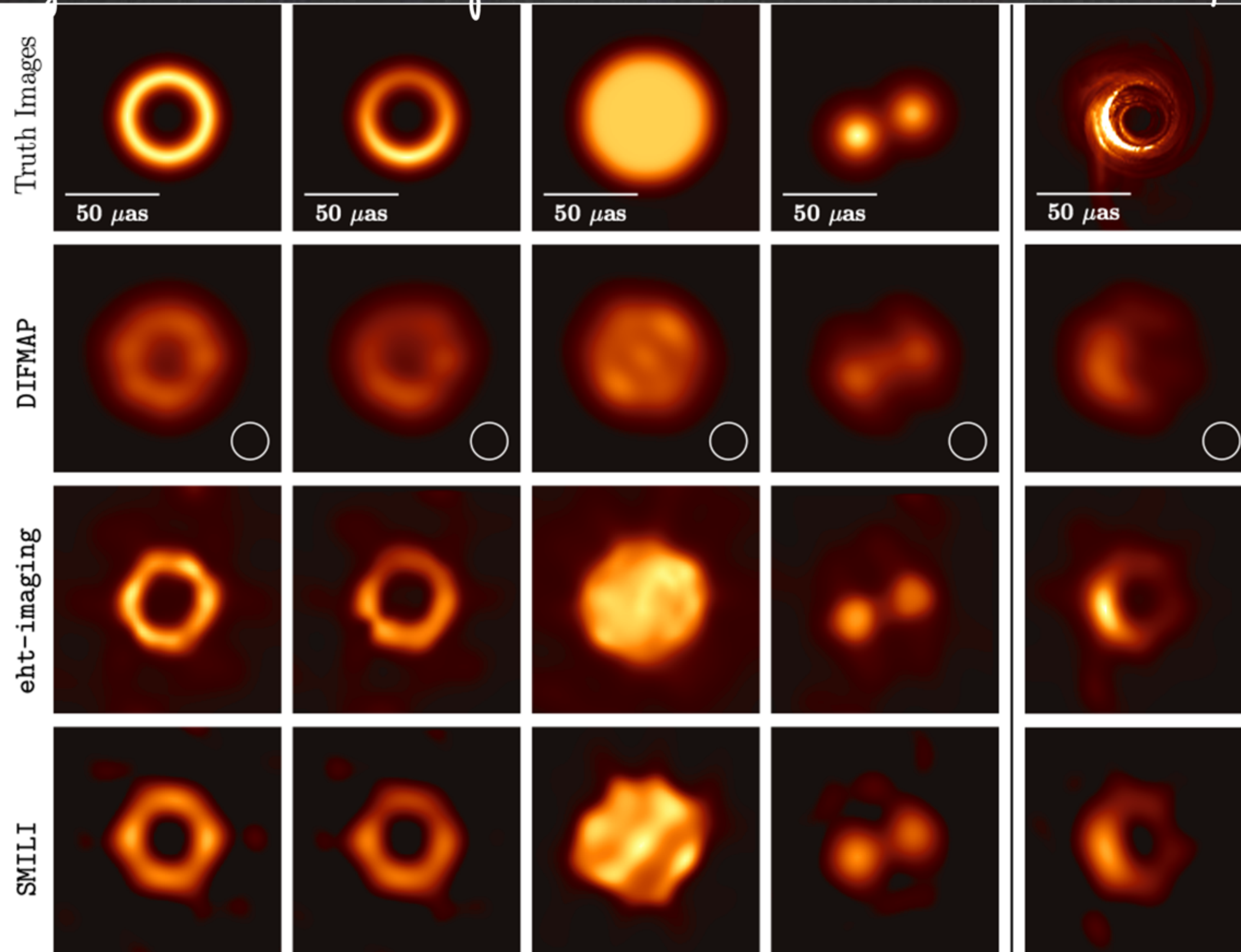
**3 Imaging
Pipelines
+ chisq cut off**

Fiducial M87 Images



Google Cloud

Can imaging methods distinguish these? Yes they can!



EHT M87 Paper IV



Event Horizon Telescope

2023, 12/8, 仁科記念講演会

未来の RML (Regularized Maximum Likelihood) 法

$p(V|I)$: 観測モデル, 尤度

χ^2 誤差, Closure 量

$p(I)$: 事前分布, 制約条件

NN などの利用

アルゴリズム

最適化法の導入 $\Rightarrow$ NN などの利用

データの理解

= 画像化のためのパラメータの理解 $\Rightarrow$ 画像の決め方



未来の Bayes 的アプローチ

$p(V|I)$: 観測モデル, 尤度

χ^2 誤差, Closure 量

$p(I)$: 事前分布, 制約条件

画像の事前分布 $\Rightarrow$ N.N. などの利用

アルゴリズム

MCMC

$\Rightarrow$ N.N. などの利用

データの理解

事後分布からのサンプリング (多数の画像)



自己紹介

データ科学について

天文分野での実践例

EHTでは

Weak Lensing Mass Map

まとめ

方法論の進歩

1990年代

Boosting, カーネル法

1990年代後半 ~ 2010年頃

スパース推定, ランダムフォレスト

2010年代 ~

深層学習

2020年代以降

??

深層学習の能.

関数近似の能力

$$y \sim f(x)$$

$$x \sim p(x) \longrightarrow y \sim p(y) \approx p(f(x))$$

確率分布の変換が可能となった.

CycleGAN

junyanz.github.io/CycleGAN/



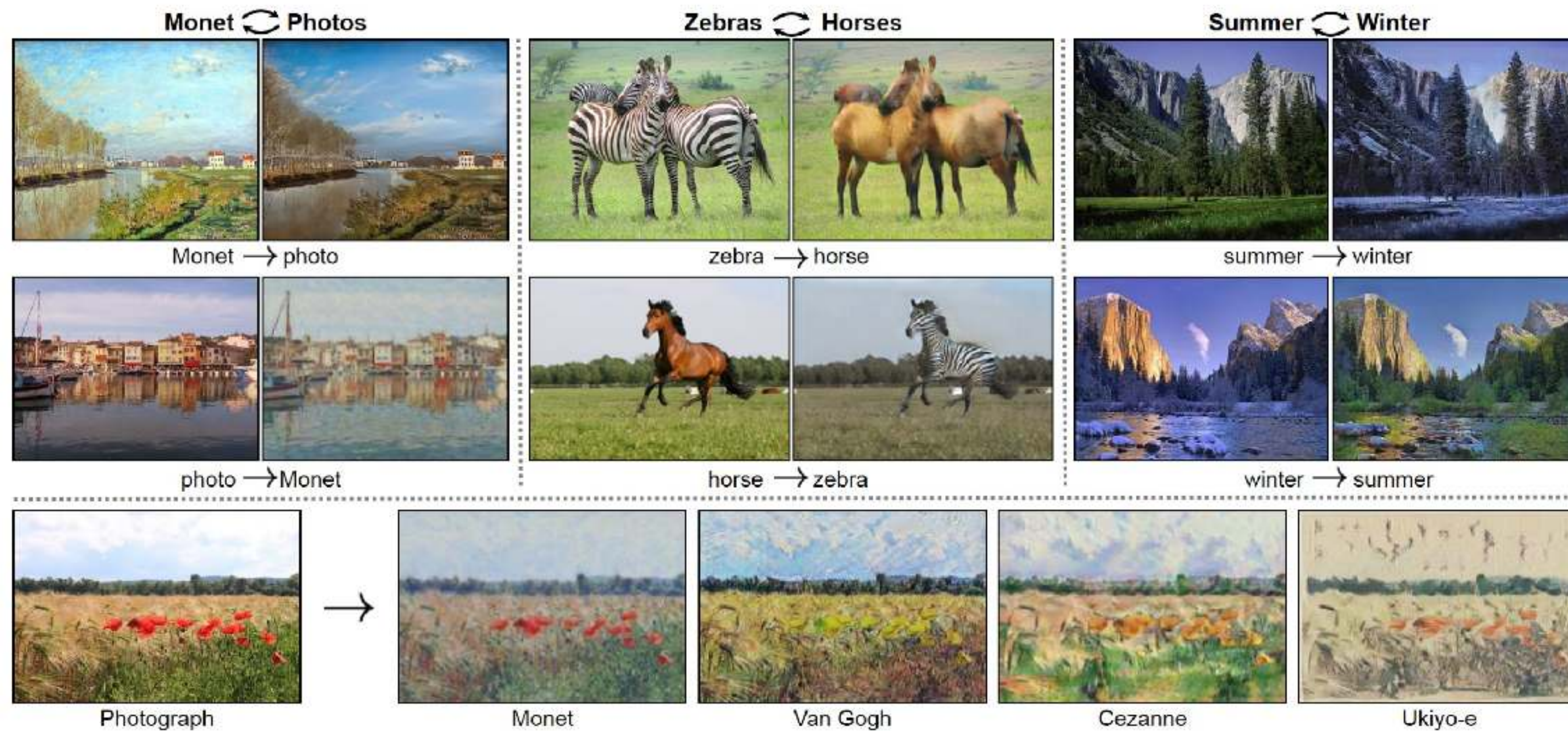
Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

Jun-Yan Zhu* **Taesung Park*** **Phillip Isola** **Alexei A. Efros**

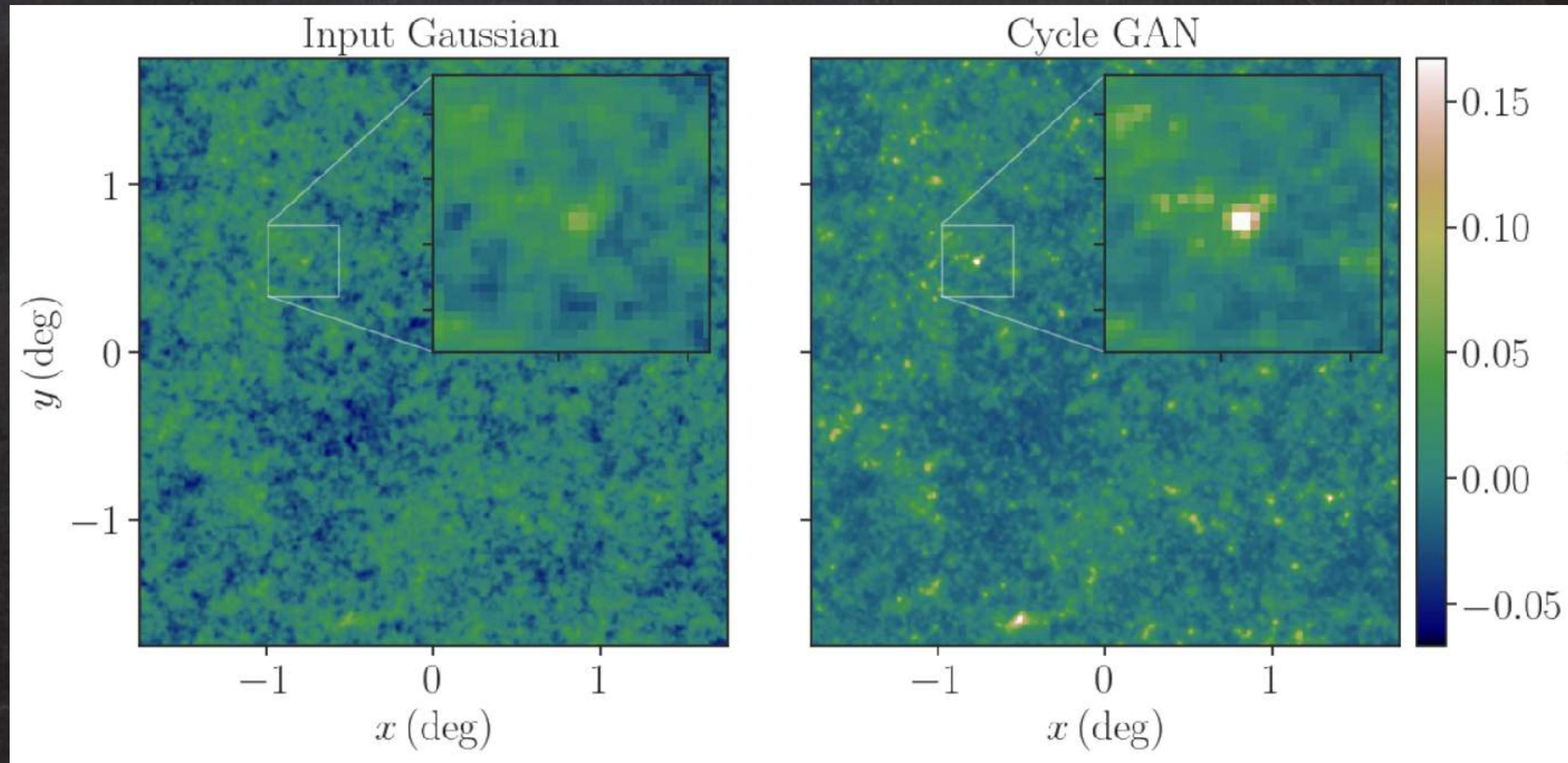
UC Berkeley

In ICCV 2017

Paper | **PyTorch code** | **Torch code**



Weak Lensing Mass Map



M. Shirasaki and S. Ikeda, "Neural style transfer of weak lensing mass maps," arXiv:2310.17141 [astro-ph.CO]

自己紹介

データ科学について

天文分野での実践例

EHTでは

Weak Lensing Mass Map

まとめ

天文学とデータ科学

現状

海外の方法の再現、改善、応用

国内のデータ科学 ... 周回遅れ

↳ ここから導入 $\Rightarrow$ 2週遅れ

将来

各プロジェクトにおいてデータ科学は必須

国内で立案、実行できる体制を

データ科学と天文学

観測



データ



自分のサイエンス

装置開発
(予算)

計算資源
データ科学

全てを自分のサイエンスの一部として理解する。

pathは通っているのか

誰がpathを通すのか

そのプロセスを理解しているのか