REMOVAL AND MARGINALIZATION OF WELL-KNOWN CMB INSTRUMENTAL SYSTEMATICS

Adri Duivenvoorden
Max Planck Institute for Astrophysics

MAX-PLANCK-INSTITUT
FÜR ASTROPHYSIK

CosmoForward meeting
09-02-2026

CMB INSTRUMENTAL SYSTEMATICS

CMB INSTRUMENTAL SYSTEMATICS

- Often quite specific to a particular experiment

CMB INSTRUMENTAL SYSTEMATICS

- Often quite specific to a particular experiment
  - For example, the non-linearity in the analog-to-digital converters for *Planck*, discontinuities in the time-ordered data of SPIDER caused by cosmic rays

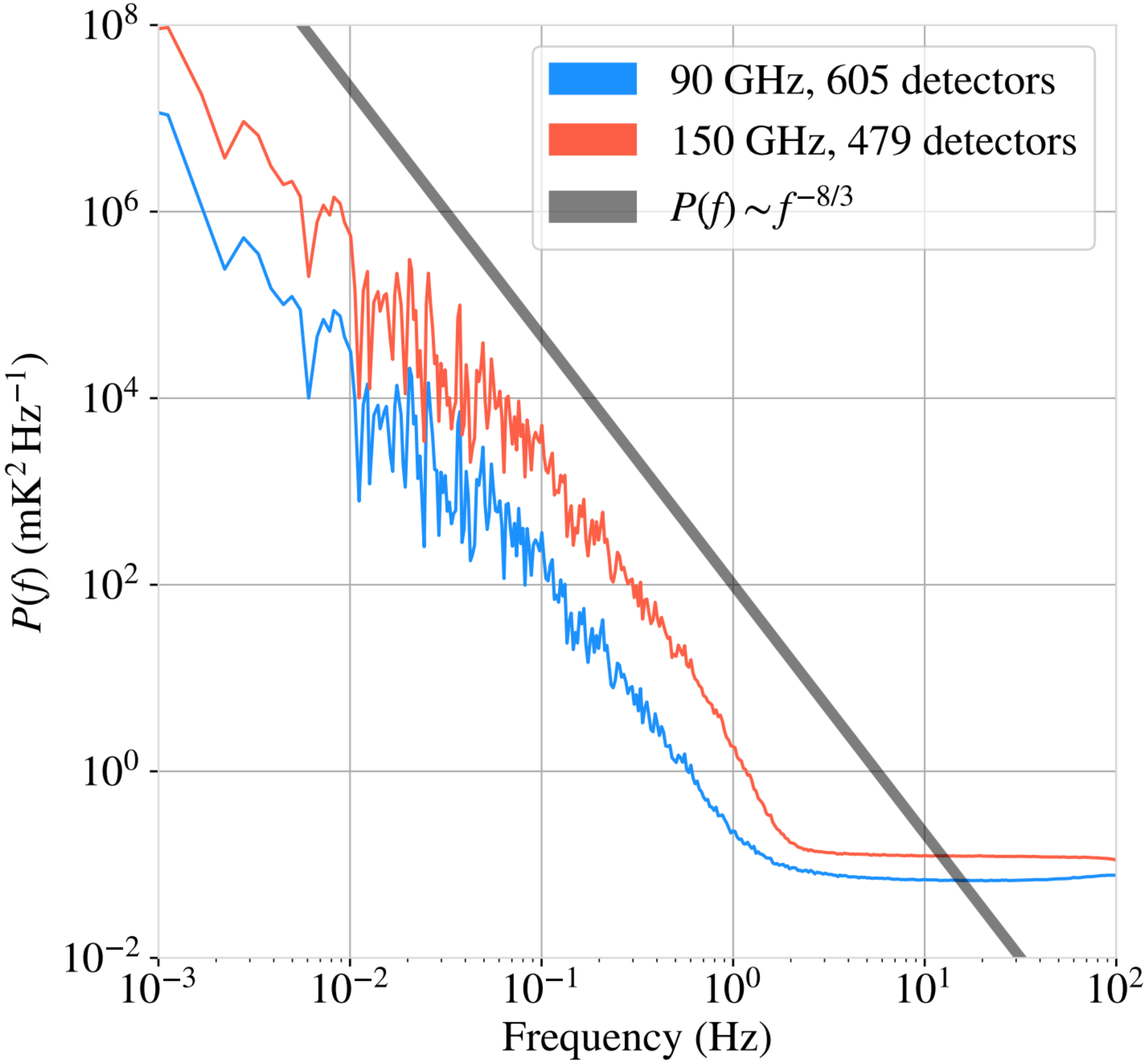
CMB INSTRUMENTAL SYSTEMATICS

- Often quite specific to a particular experiment
  - For example, the non-linearity in the analog-to-digital converters for *Planck*, discontinuities in the time-ordered data of SPIDER caused by cosmic rays
- Here, I will try to focus on more generic instrumental systematics

CMB INSTRUMENTAL SYSTEMATICS

- Often quite specific to a particular experiment
  - For example, the non-linearity in the analog-to-digital converters for *Planck*, discontinuities in the time-ordered data of SPIDER caused by cosmic rays
- Here, I will try to focus on more generic instrumental systematics
  - Although the topics will be skewed towards my own experience with the Atacama Cosmology Telescope

ATMOSPHERIC NOISE



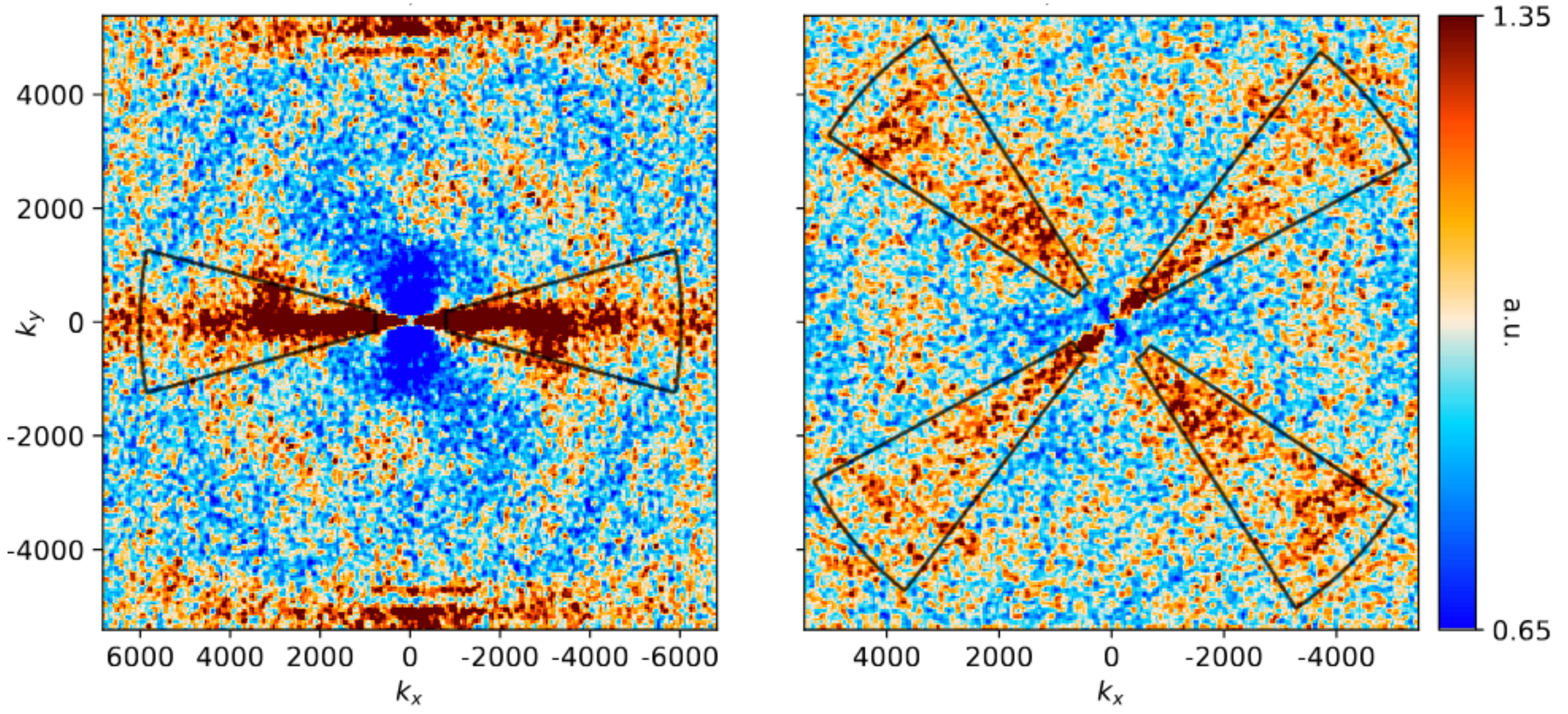
33 minutes of stare data for PA6

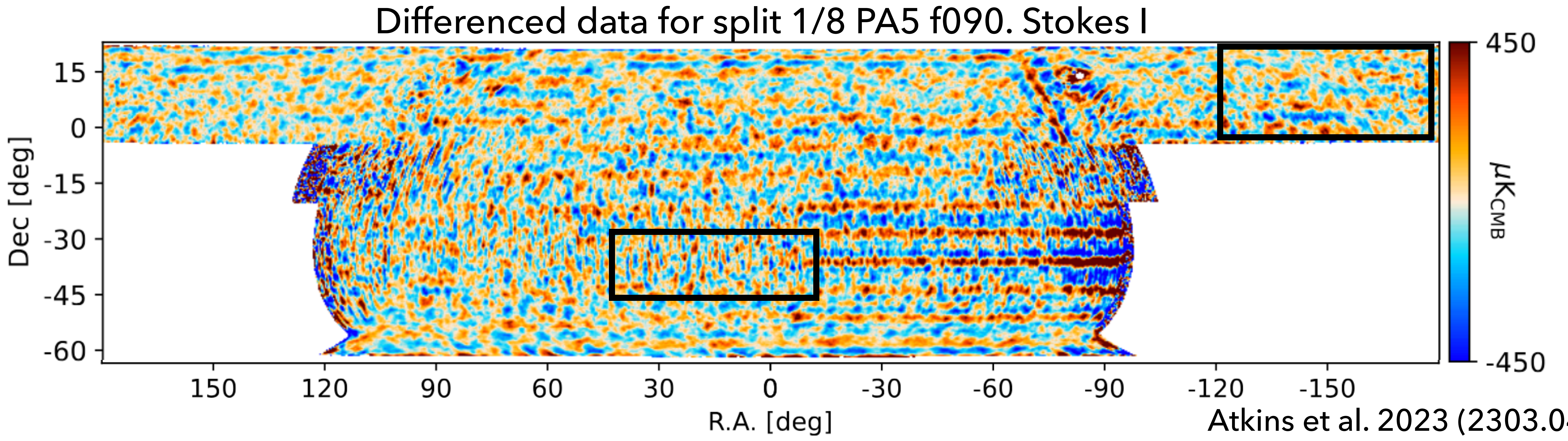
Morris et al. (2022), 2111.01319

NOISE PROPERTIES

Correlated atmospheric noise modulated by scan strategy results in complicated noise properties

- Difficult to simulate the atmosphere



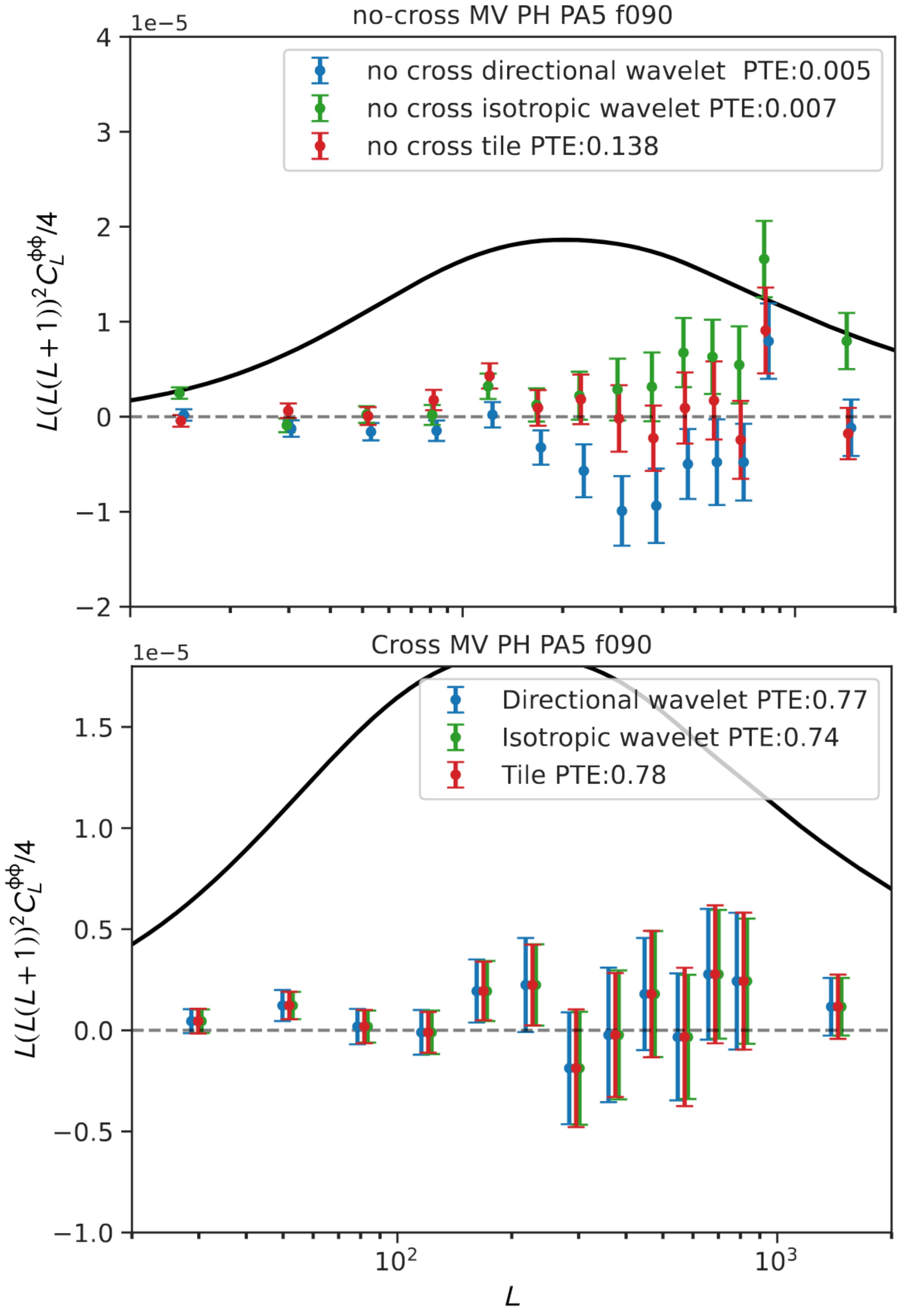


Atkins et al. 2023 (2303.04180)

ROBUST ESTIMATORS

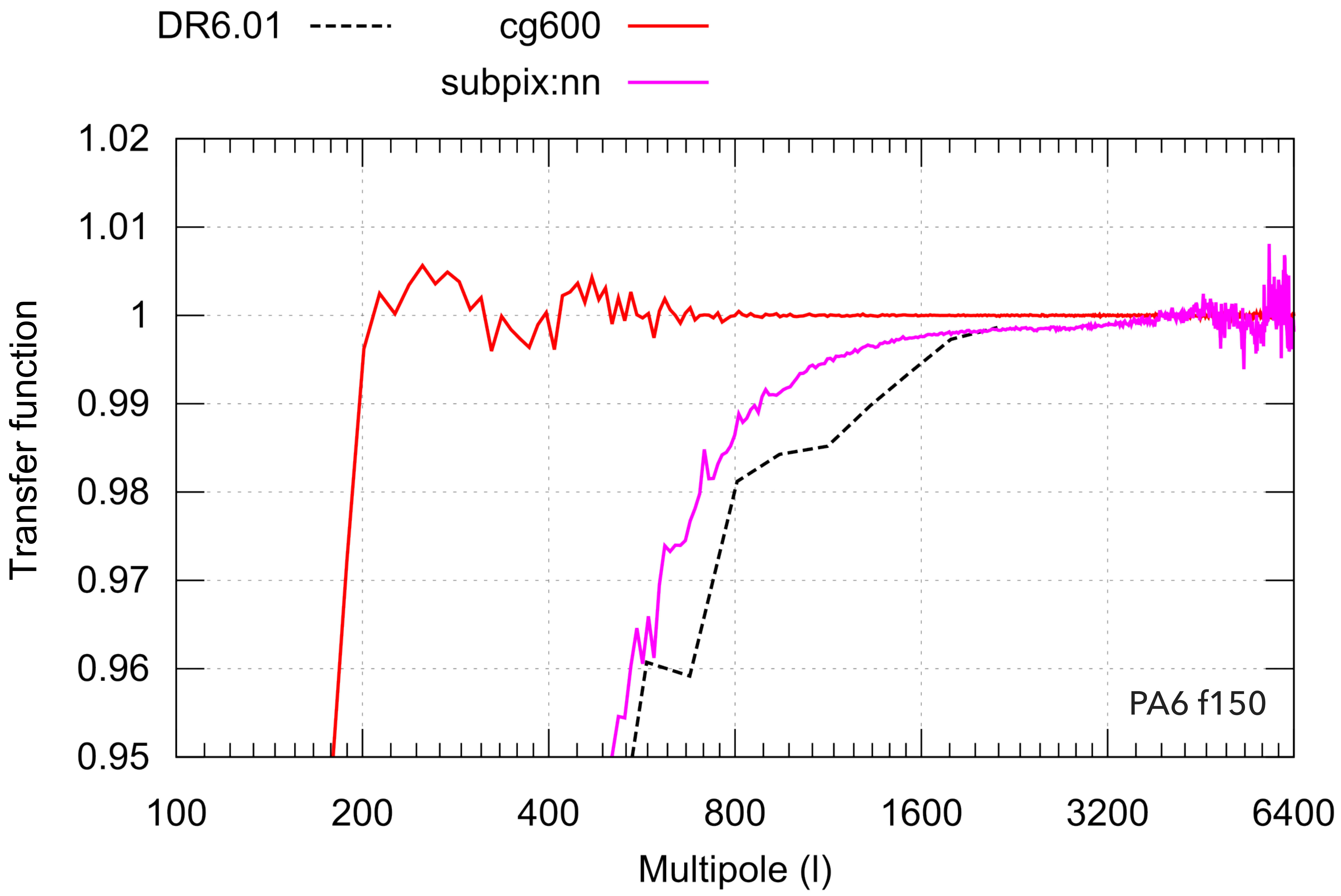
Construct estimators that rely on the cross-correlation between observations with independent noise

- Already the norm for power-spectrum estimation
- ACT DR6 demonstrated the effectiveness of this same idea for lensing estimation, using the estimator from Madhavacheril et al. (2020), 2011.02475



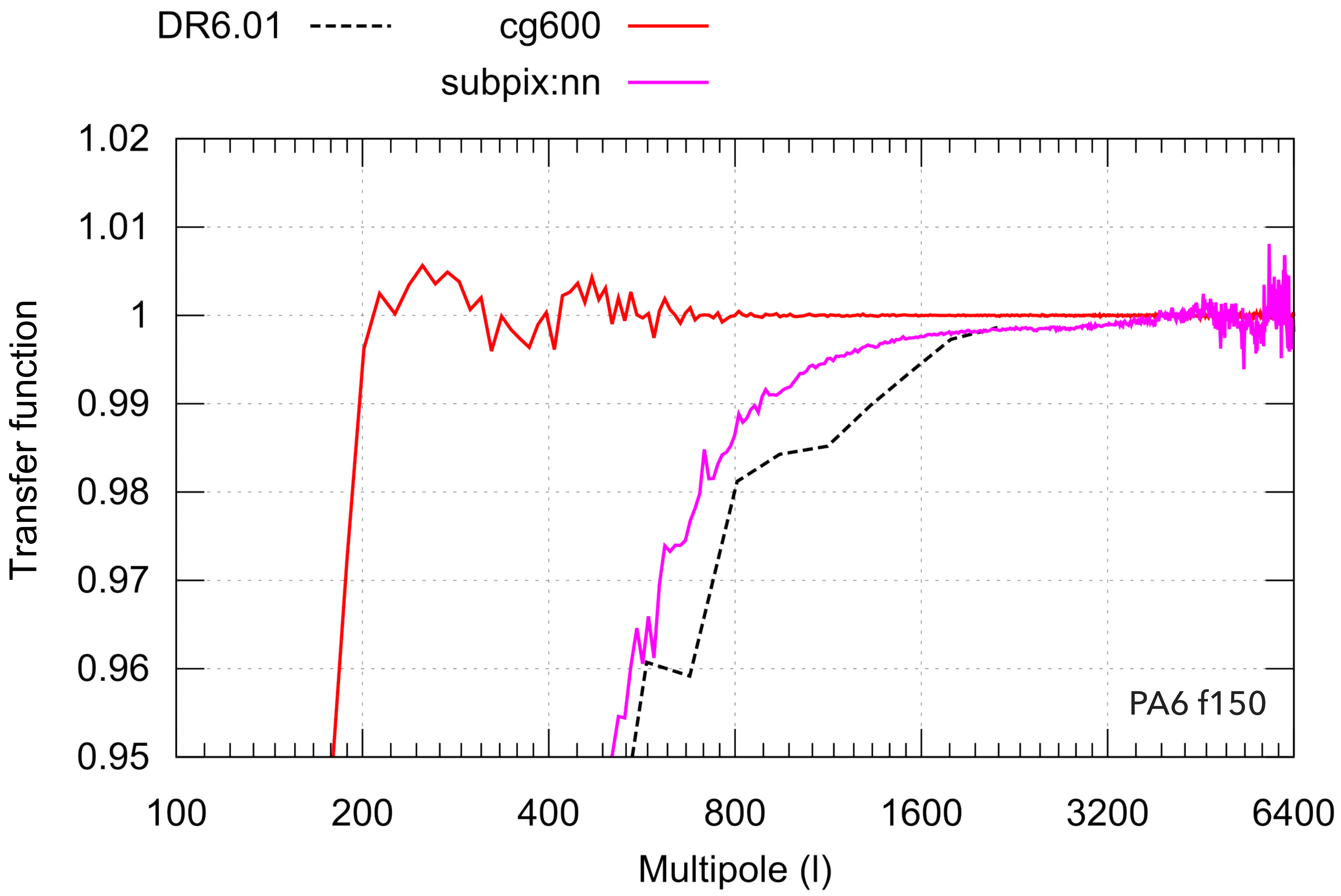
Qu et al, 2024 (2304.05202)

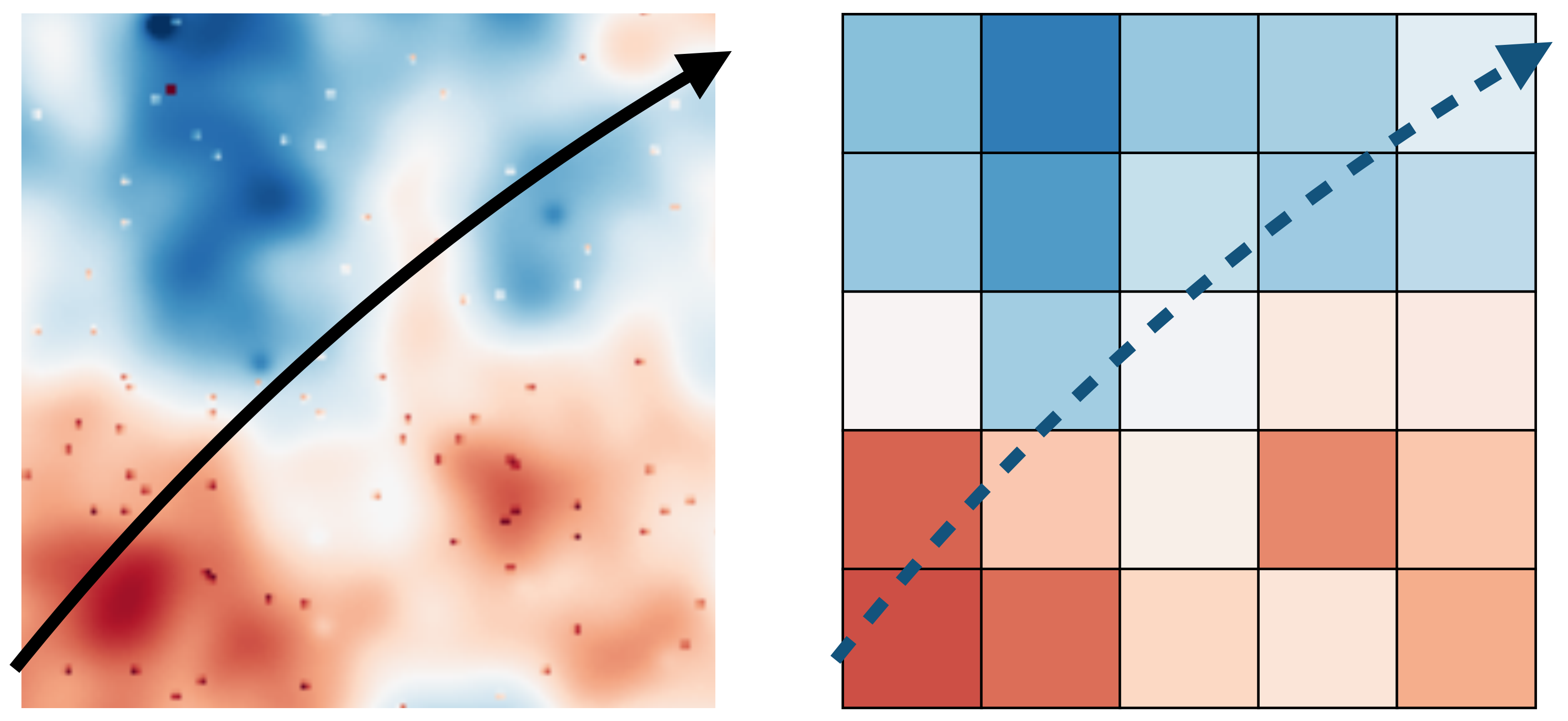
MAPMAKING SYSTEMATICS

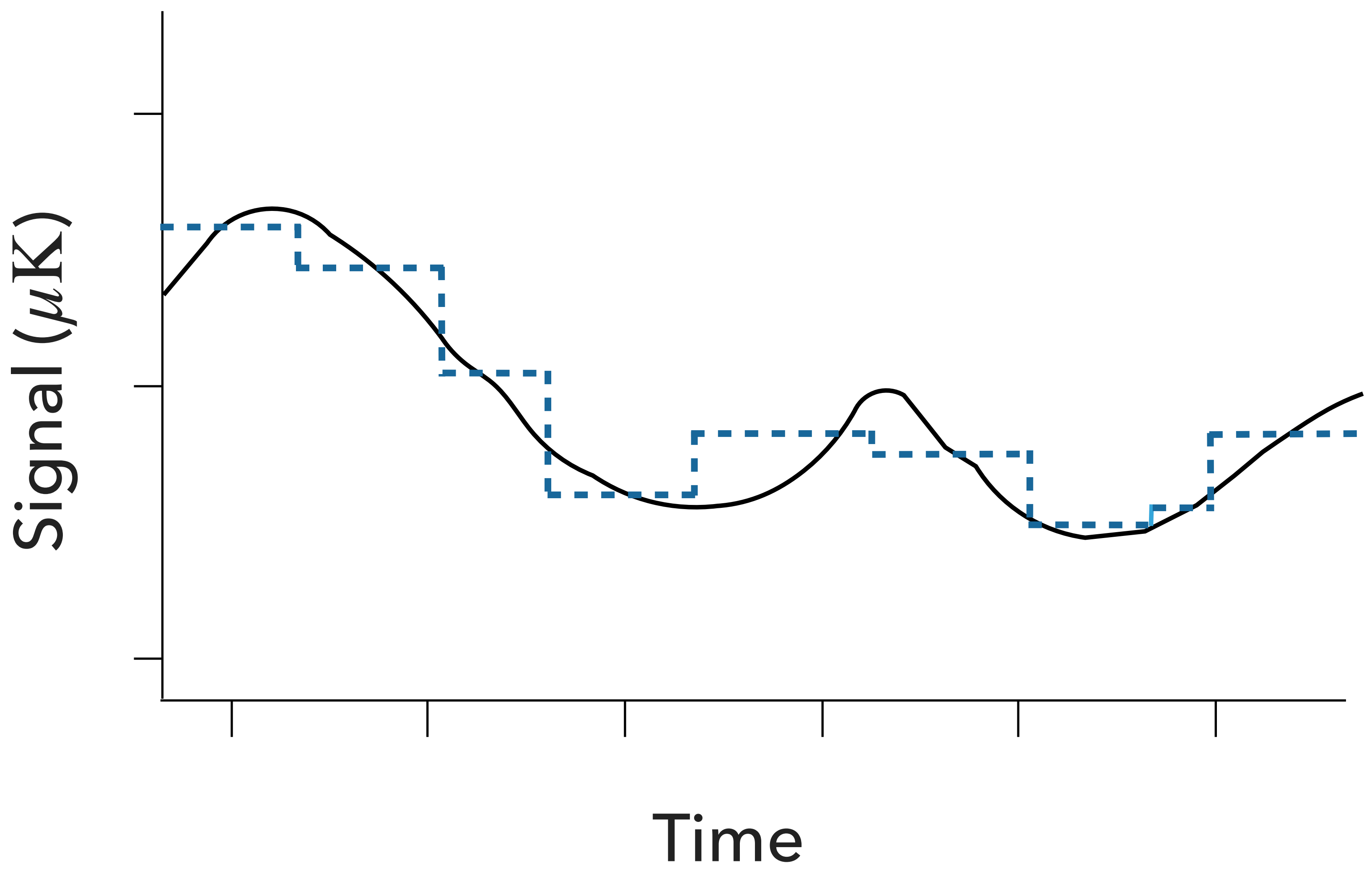


Naess et al, 2025 (2503.14451)

MAPMAKING SYSTEMATICS

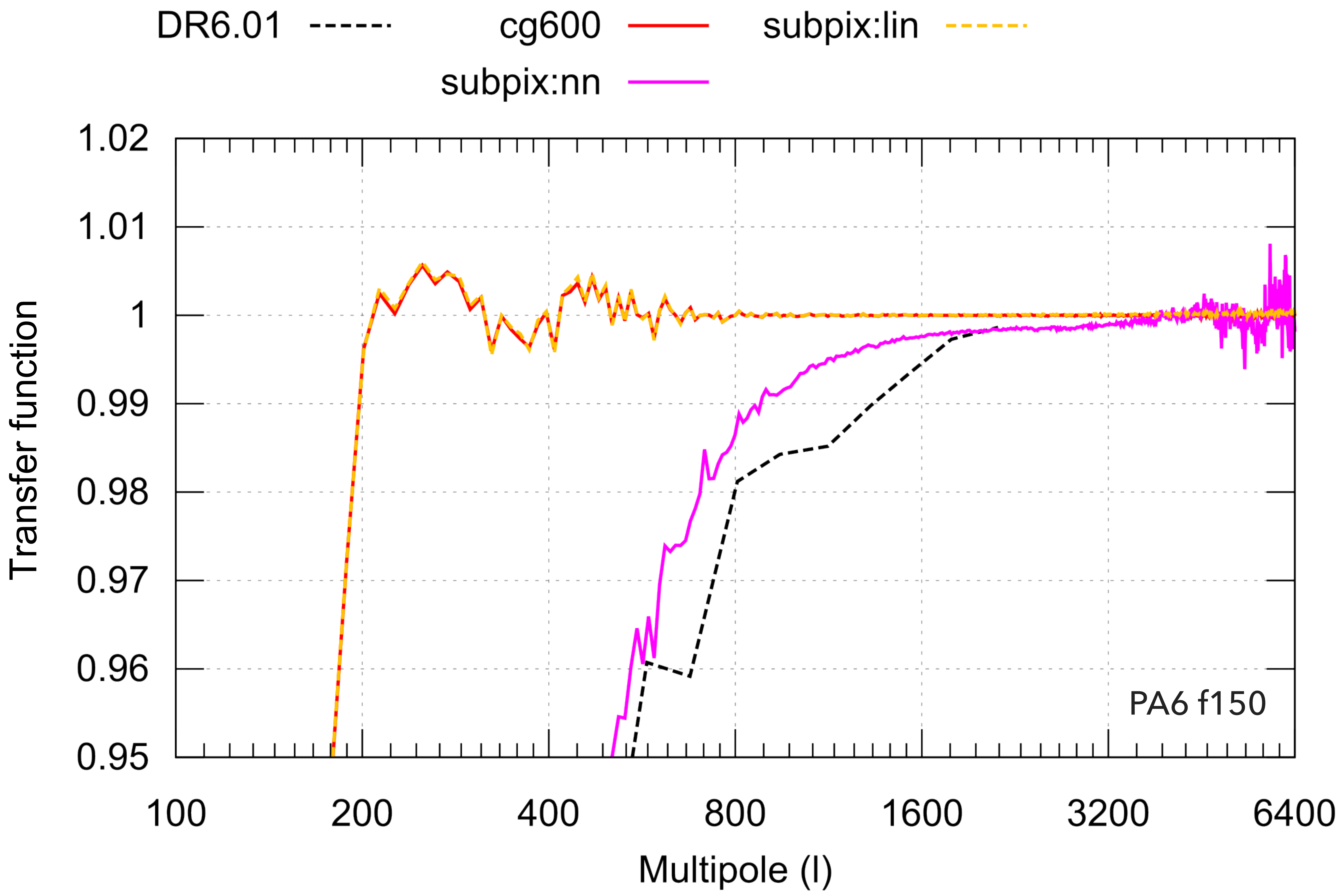






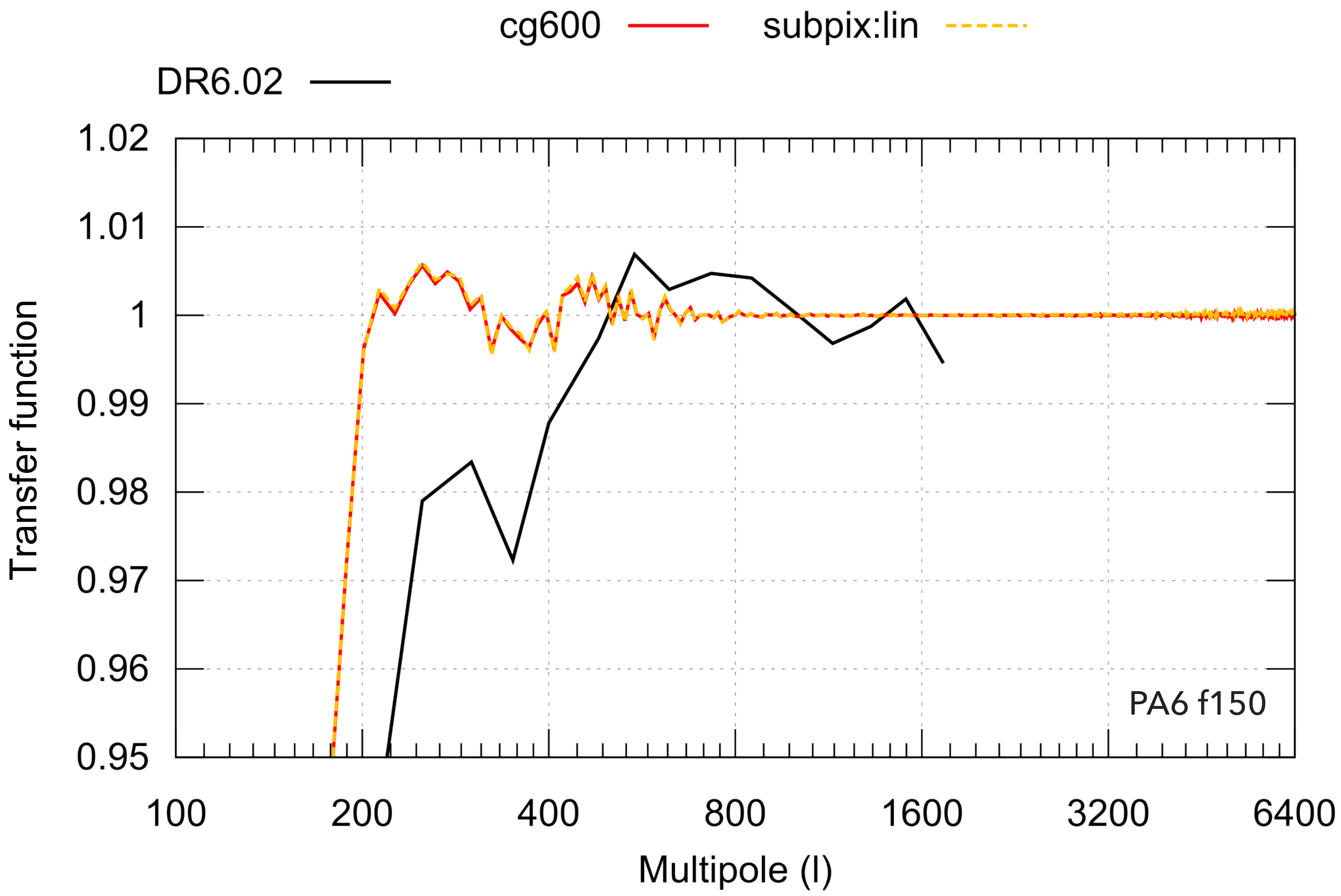
Naess et al, 2025 (2503.14451)

MAPMAKING SYSTEMATICS



Naess et al, 2025 (2503.14451)

MAPMAKING SYSTEMATICS



Naess et al, 2025 (2503.14451)

LARGE-SCALE POWER LOSS FROM GAIN ERRORS

Naess, Louis, 2023 (2210.02243)

LARGE-SCALE POWER LOSS FROM GAIN ERRORS

$$d = Ps + n$$

Naess, Louis, 2023 (2210.02243)

LARGE-SCALE POWER LOSS FROM GAIN ERRORS

$$d = Ps + n$$

$$\hat{s} = (P^\top N^{-1} P)^{-1} P^\top N^{-1} d$$

Naess, Louis, 2023 (2210.02243)

LARGE-SCALE POWER LOSS FROM GAIN ERRORS

$$d = Ps + n$$

$$\hat{s} = (P^\top N^{-1} P)^{-1} P^\top N^{-1} d$$

$$\langle \hat{s} \rangle = (P^\top N^{-1} P)^{-1} P^\top N^{-1} P s$$

Naess, Louis, 2023 (2210.02243)

LARGE-SCALE POWER LOSS FROM GAIN ERRORS

$$d = Ps + n$$

$$\hat{s} = (P^\top N^{-1} P)^{-1} P^\top N^{-1} d$$

$$\langle \hat{s} \rangle = (P^\top N^{-1} P)^{-1} P^\top N^{-1} P s$$

$$\langle \hat{s} \rangle = s$$

Naess, Louis, 2023 (2210.02243)

LARGE-SCALE POWER LOSS FROM GAIN ERRORS

$$d = Ps + n$$

$$d = GPs + Gn \qquad \text{(true)}$$

$$\hat{s} = (P^\top N^{-1} P)^{-1} P^\top N^{-1} d$$

$$\langle \hat{s} \rangle = (P^\top N^{-1} P)^{-1} P^\top N^{-1} P s$$

$$\langle \hat{s} \rangle = s$$

Naess, Louis, 2023 (2210.02243)

LARGE-SCALE POWER LOSS FROM GAIN ERRORS

$$d = Ps + n$$

$$\hat{s} = (P^\top N^{-1} P)^{-1} P^\top N^{-1} d$$

$$\langle \hat{s} \rangle = (P^\top N^{-1} P)^{-1} P^\top N^{-1} P s$$

$$\langle \hat{s} \rangle = s$$

$d = GPs + Gn$ (true)

$d = Ps + Gn$ (assumed)

Naess, Louis, 2023 (2210.02243)

LARGE-SCALE POWER LOSS FROM GAIN ERRORS

$$d = Ps + n$$

$$\hat{s} = (P^\top N^{-1} P)^{-1} P^\top N^{-1} d$$

$$\langle \hat{s} \rangle = (P^\top N^{-1} P)^{-1} P^\top N^{-1} P s$$

$$\langle \hat{s} \rangle = s$$

$d = GPs + Gn$ (true)

$d = Ps + Gn$ (assumed)

- G encodes per-detector gain errors

Naess, Louis, 2023 (2210.02243)

LARGE-SCALE POWER LOSS FROM GAIN ERRORS

$$d = Ps + n$$

$$\hat{s} = (P^\top N^{-1} P)^{-1} P^\top N^{-1} d$$

$$\langle \hat{s} \rangle = (P^\top N^{-1} P)^{-1} P^\top N^{-1} P s$$

$$\langle \hat{s} \rangle = s$$

$d = GPs + Gn$ (true)

$d = Ps + Gn$ (assumed)

- G encodes per-detector gain errors

$$\hat{s} = (P^\top G^{-1} N^{-1} G^{-1} P)^{-1} P^\top G^{-1} N^{-1} G^{-1} d$$

Naess, Louis, 2023 (2210.02243)

LARGE-SCALE POWER LOSS FROM GAIN ERRORS

$$d = Ps + n$$

$$\hat{s} = (P^\top N^{-1} P)^{-1} P^\top N^{-1} d$$

$$\langle \hat{s} \rangle = (P^\top N^{-1} P)^{-1} P^\top N^{-1} P s$$

$$\langle \hat{s} \rangle = s$$

$d = GPs + Gn$ (true)

$d = Ps + Gn$ (assumed)

- G encodes per-detector gain errors

$$\hat{s} = (P^\top G^{-1} N^{-1} G^{-1} P)^{-1} P^\top G^{-1} N^{-1} G^{-1} d$$

$$\langle \hat{s} \rangle = (P^\top G^{-1} N^{-1} G^{-1} P)^{-1} P^\top G^{-1} N^{-1} G^{-1} G P s$$

Naess, Louis, 2023 (2210.02243)

TRANSFER FUNCTION FROM GAIN ERRORS

- $\langle \hat{s} \rangle = (P^\top G^{-1} N^{-1} G^{-1} P)^{-1} P^\top G^{-1} N^{-1} G^{-1} G P s$

Naess, Louis, 2023 (2210.02243)

TRANSFER FUNCTION FROM GAIN ERRORS

- $\langle \hat{s} \rangle = (P^\top G^{-1} N^{-1} G^{-1} P)^{-1} P^\top G^{-1} N^{-1} G^{-1} G P s$

- 1D toy model with 2 detectors
  - $N_f = A_f \begin{pmatrix} 1 & \alpha_f \\ \alpha_f & 1 \end{pmatrix}, \; G = \begin{pmatrix} g_1 & 0 \\ 0 & g_2 \end{pmatrix}$

Naess, Louis, 2023 (2210.02243)

TRANSFER FUNCTION FROM GAIN ERRORS

- $\langle \hat{s} \rangle = (P^\top G^{-1} N^{-1} G^{-1} P)^{-1} P^\top G^{-1} N^{-1} G^{-1} G P s$

- 1D toy model with 2 detectors
  - $N_f = A_f \begin{pmatrix} 1 & \alpha_f \\ \alpha_f & 1 \end{pmatrix}, \; G = \begin{pmatrix} g_1 & 0 \\ 0 & g_2 \end{pmatrix}$

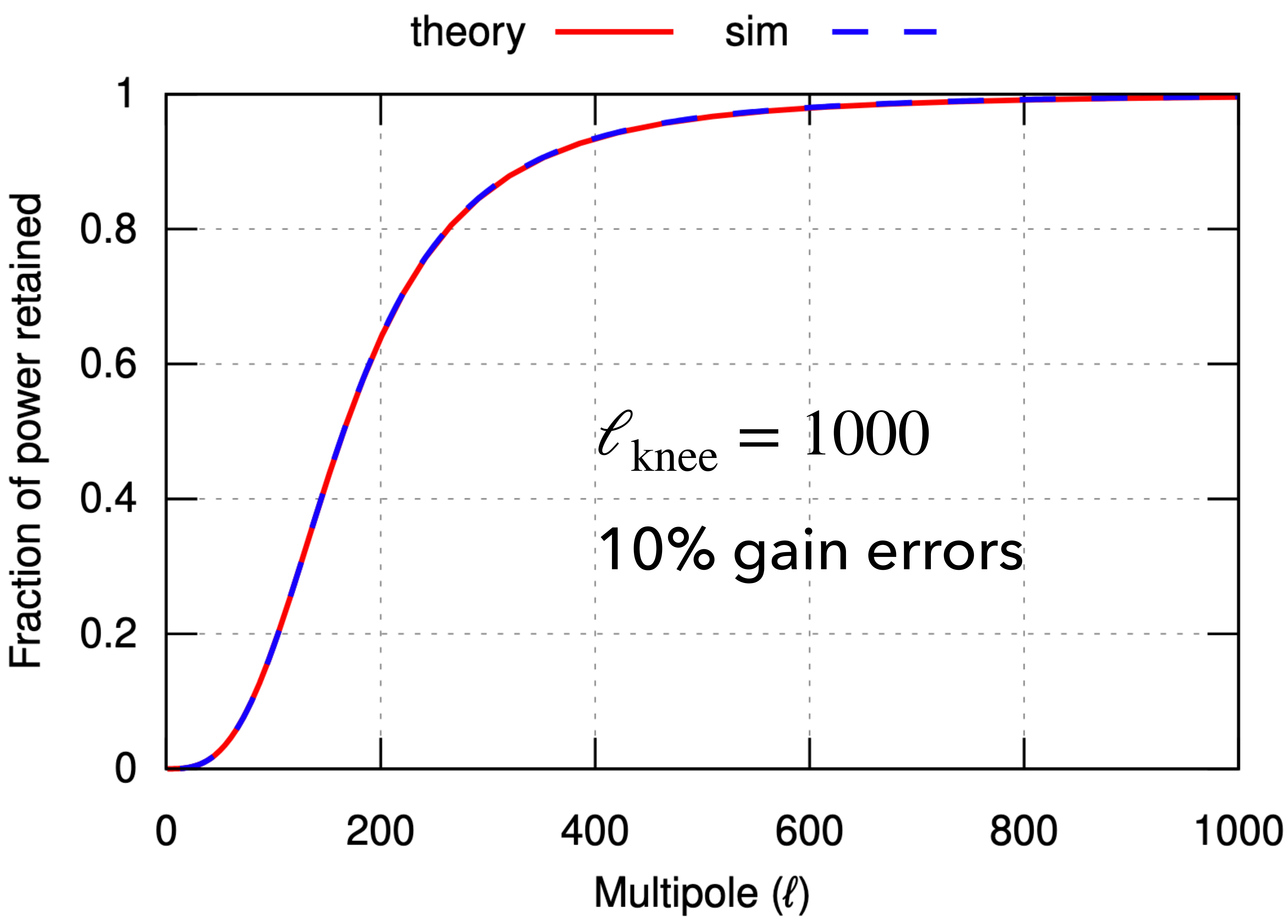
- $\langle \hat{s} \rangle = g_1 g_2 \frac{g_1 + g_2}{g_1^2 + g_2^2} \frac{1 - \alpha_f}{1 - \frac{2 g_1 g_2}{g_1^2 + g_2^2} \alpha_f} m$

Naess, Louis, 2023 (2210.02243)

TRANSFER FUNCTION FROM GAIN ERRORS

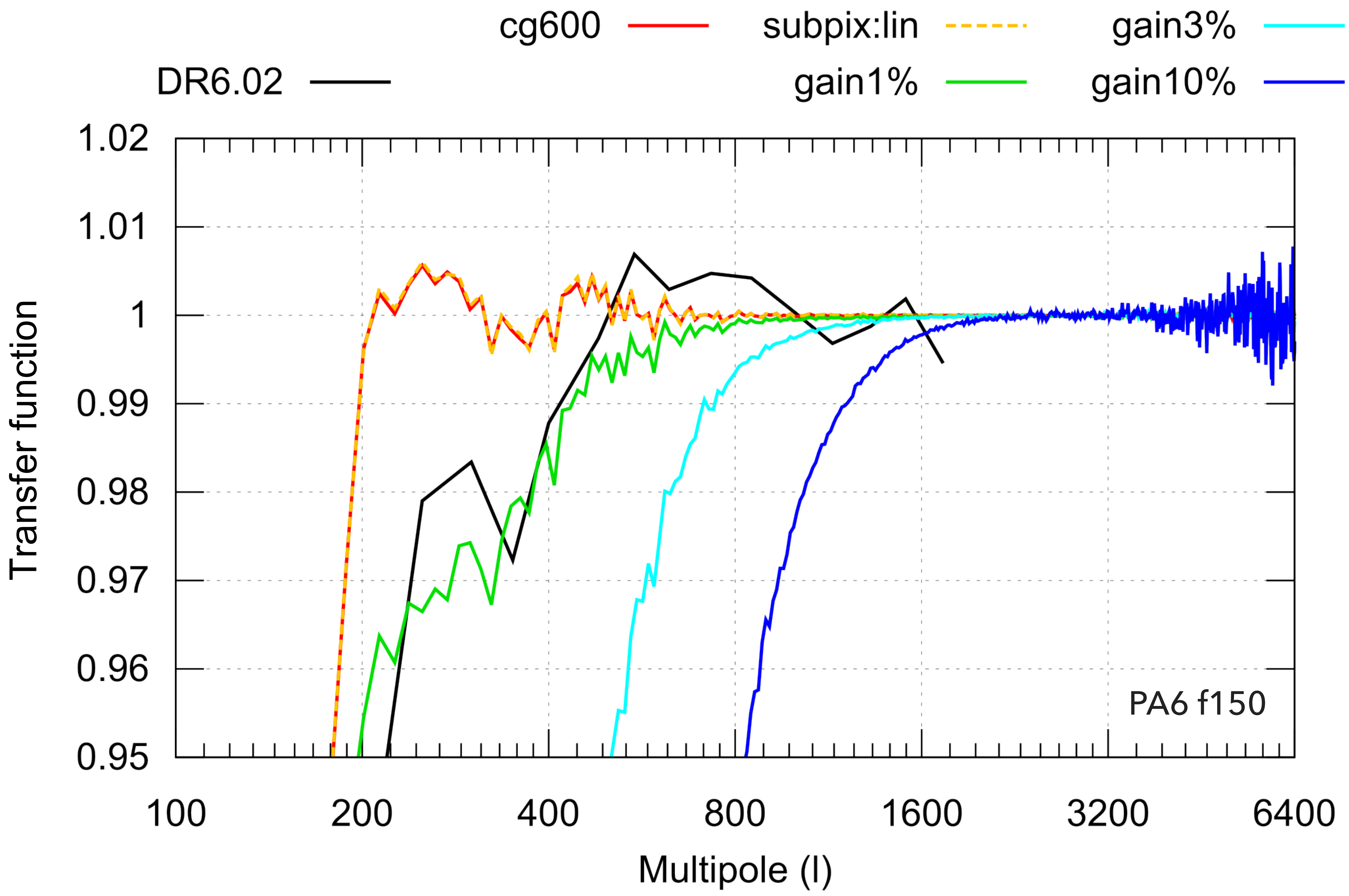
- $\langle \hat{s} \rangle = (P^\top G^{-1} N^{-1} G^{-1} P)^{-1} P^\top G^{-1} N^{-1} G^{-1} G P s$

- 1D toy model with 2 detectors
  - $N_f = A_f \begin{pmatrix} 1 & \alpha_f \\ \alpha_f & 1 \end{pmatrix}, \; G = \begin{pmatrix} g_1 & 0 \\ 0 & g_2 \end{pmatrix}$
- $\langle \hat{s} \rangle = g_1 g_2 \dfrac{g_1 + g_2}{g_1^2 + g_2^2} \dfrac{1 - \alpha_f}{1 - \frac{2 g_1 g_2}{g_1^2 + g_2^2} \alpha_f} m$



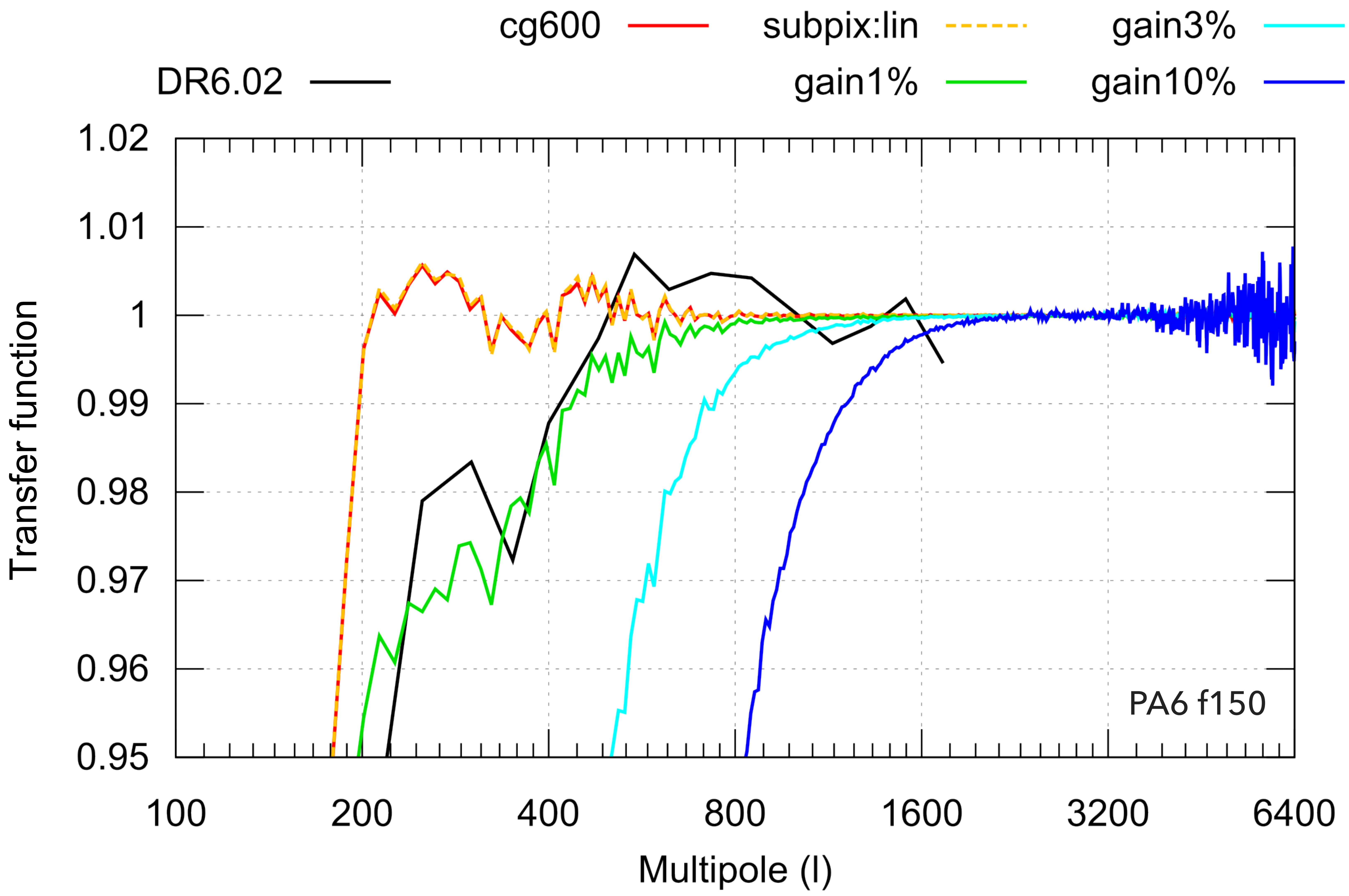
Naess, Louis, 2023 (2210.02243)

MAPMAKING SYSTEMATICS



Naess et al, 2025 (2503.14451)

MAPMAKING SYSTEMATICS



Final ACT DR6 transfer function consistent with ~1% relative gain errors between detectors

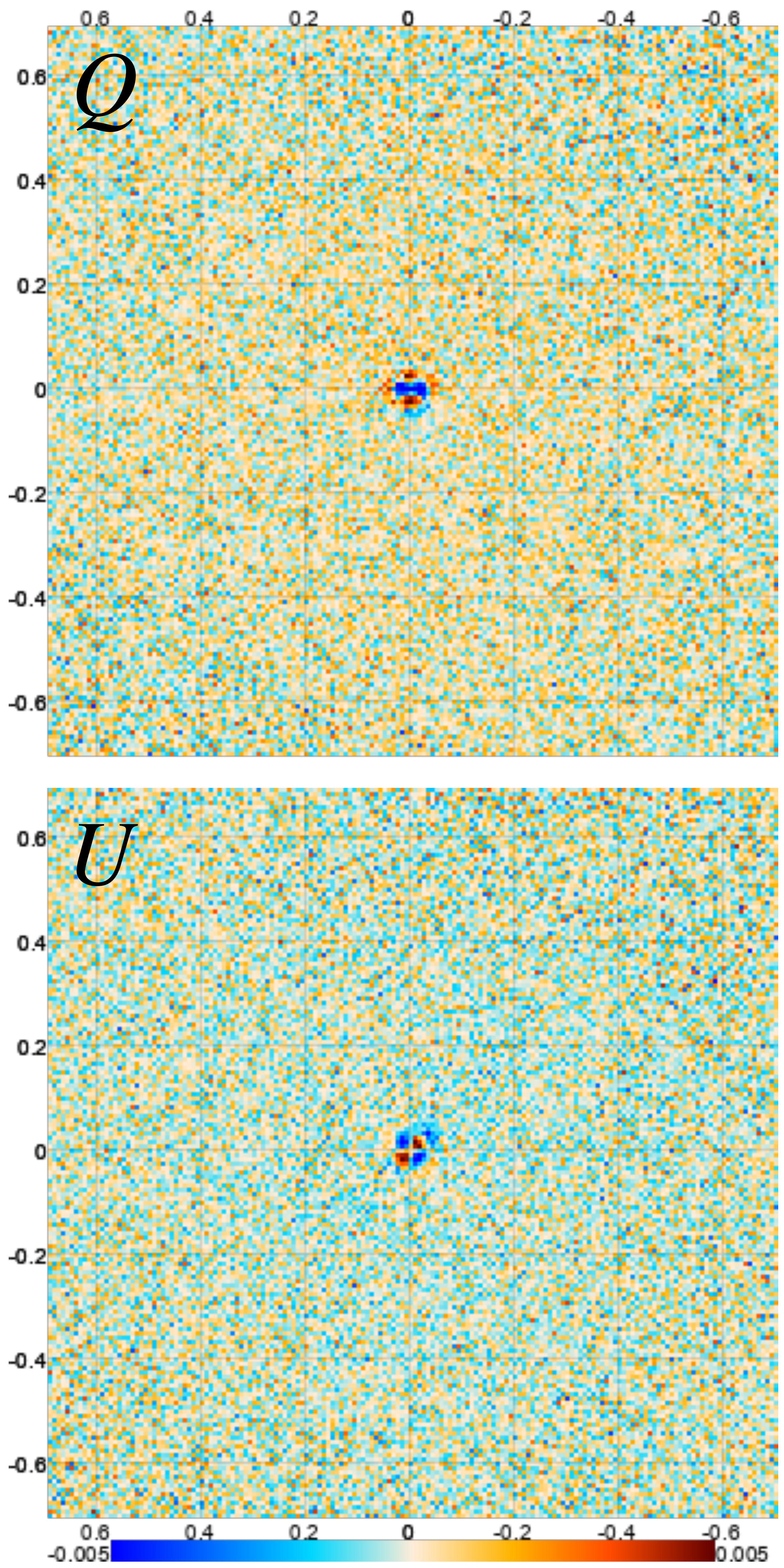
- Difficult to improve much with either planet- or atmospheric-based calibration
- Improvements will likely require dedicated calibration hardware

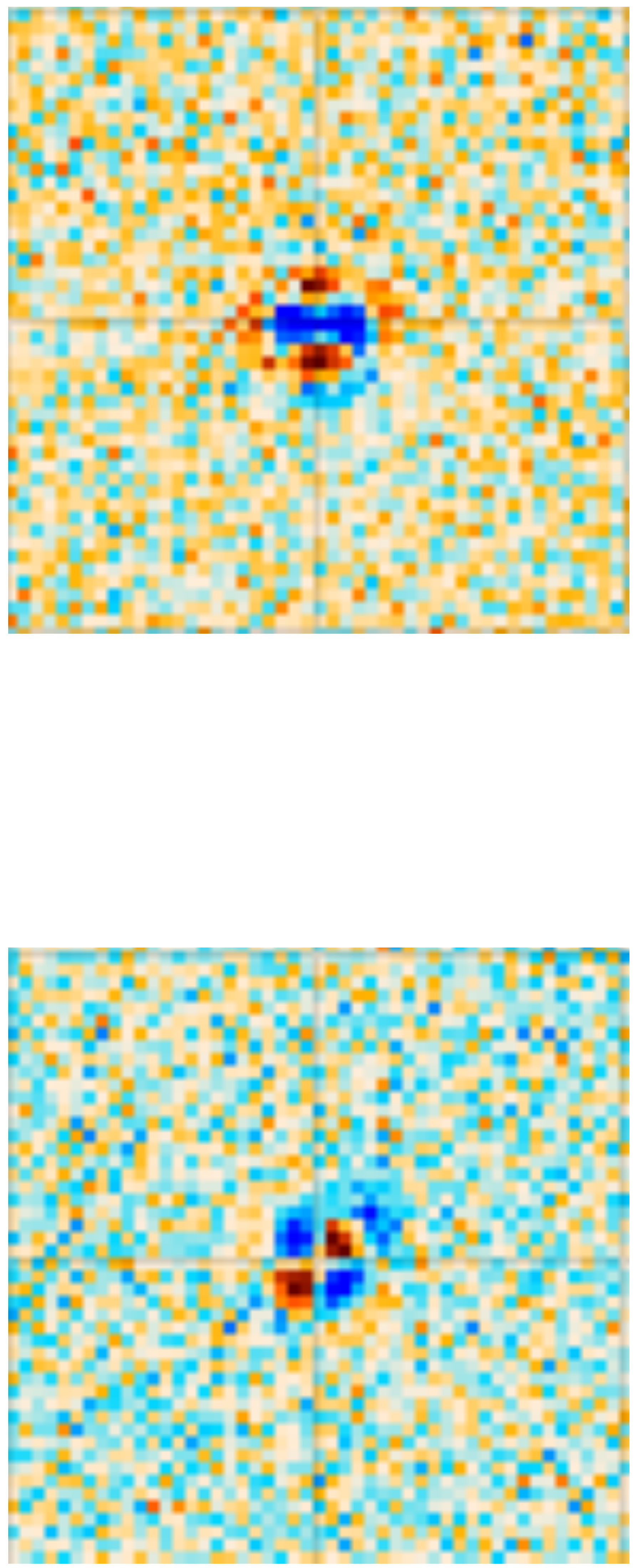
Naess et al, 2025 (2503.14451)

TEMPERATURE-TO-POLARIZATION LEAKAGE

Optical non-idealities and errors in the data model cause spurious polarized signal

- Quantified using planet observations



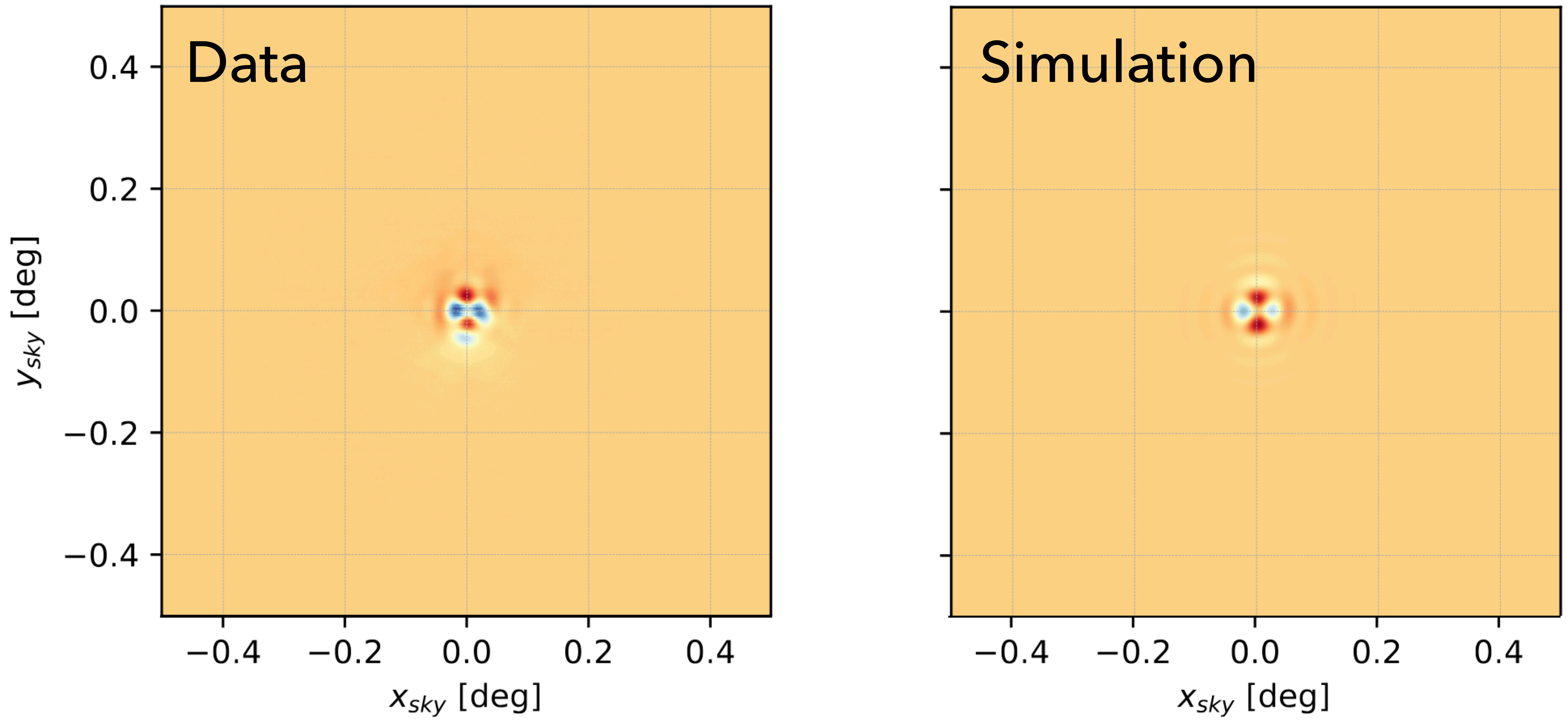


0.7x0.7 deg^2 maps of 25 Uranus observations

OPTICAL SIMULATIONS ARE DIFFICULT

Simulations reveal that the bulk of the leakage at small scales is caused by geometry of the optics

- Leakage at large angular scales appears to caused by unknown mis-modeling (e.g. gain errors)

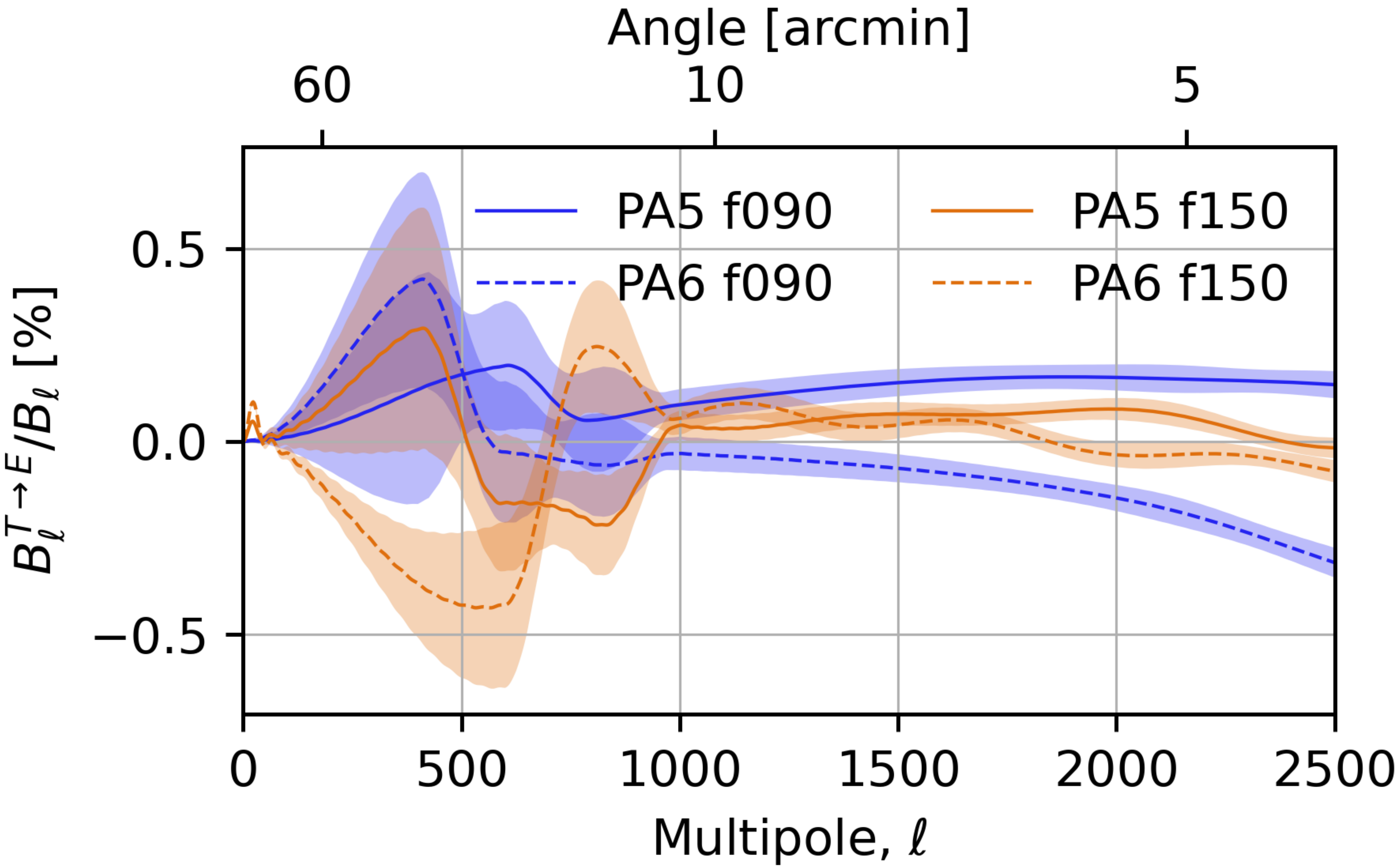


- Simulation by Roberto Puddu

GRASP simulation of single detector on the center of the focal plane

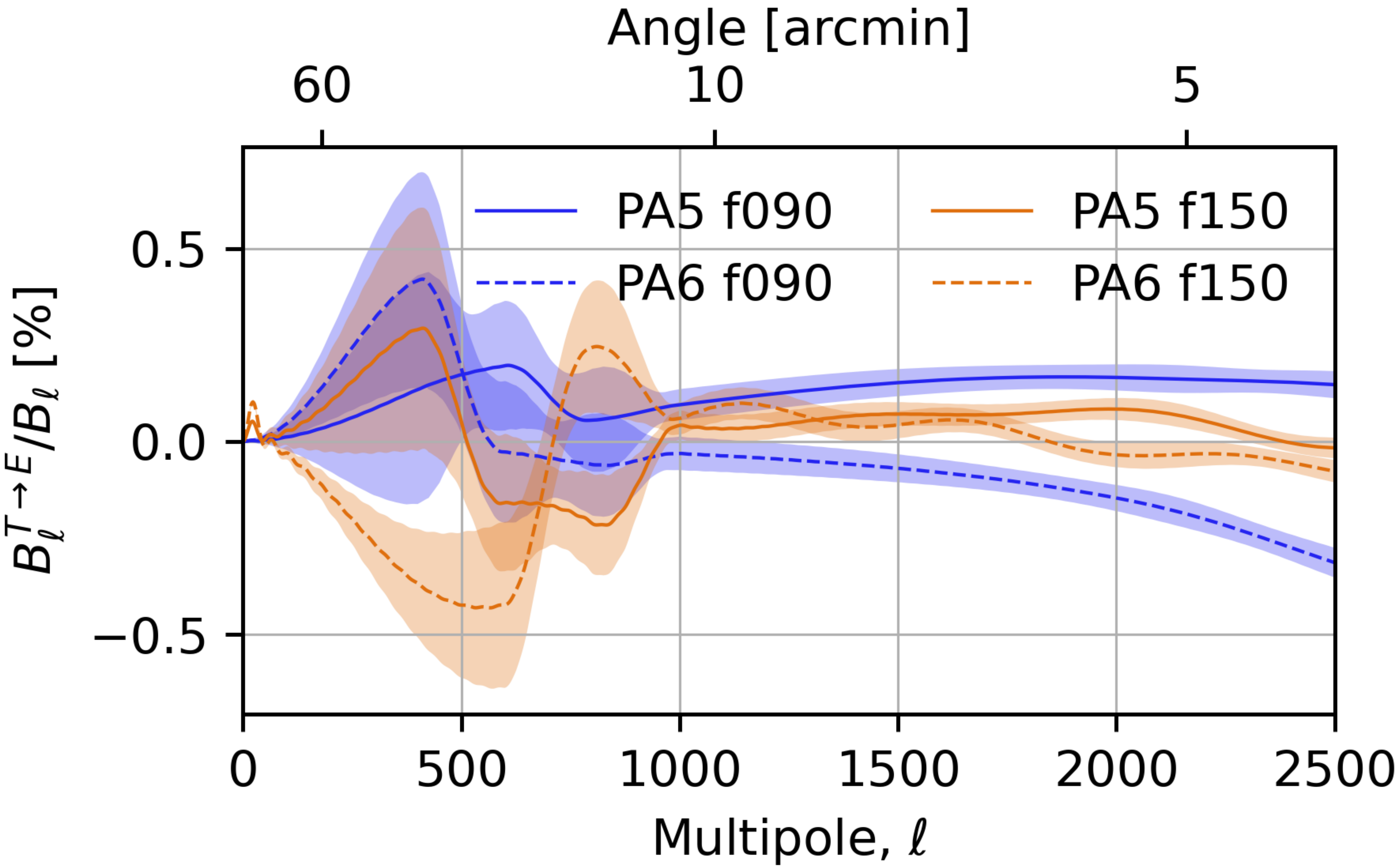
No gain errors simulated, just the mirrors and lenses

BEAM LEAKAGE MODEL



BEAM LEAKAGE MODEL

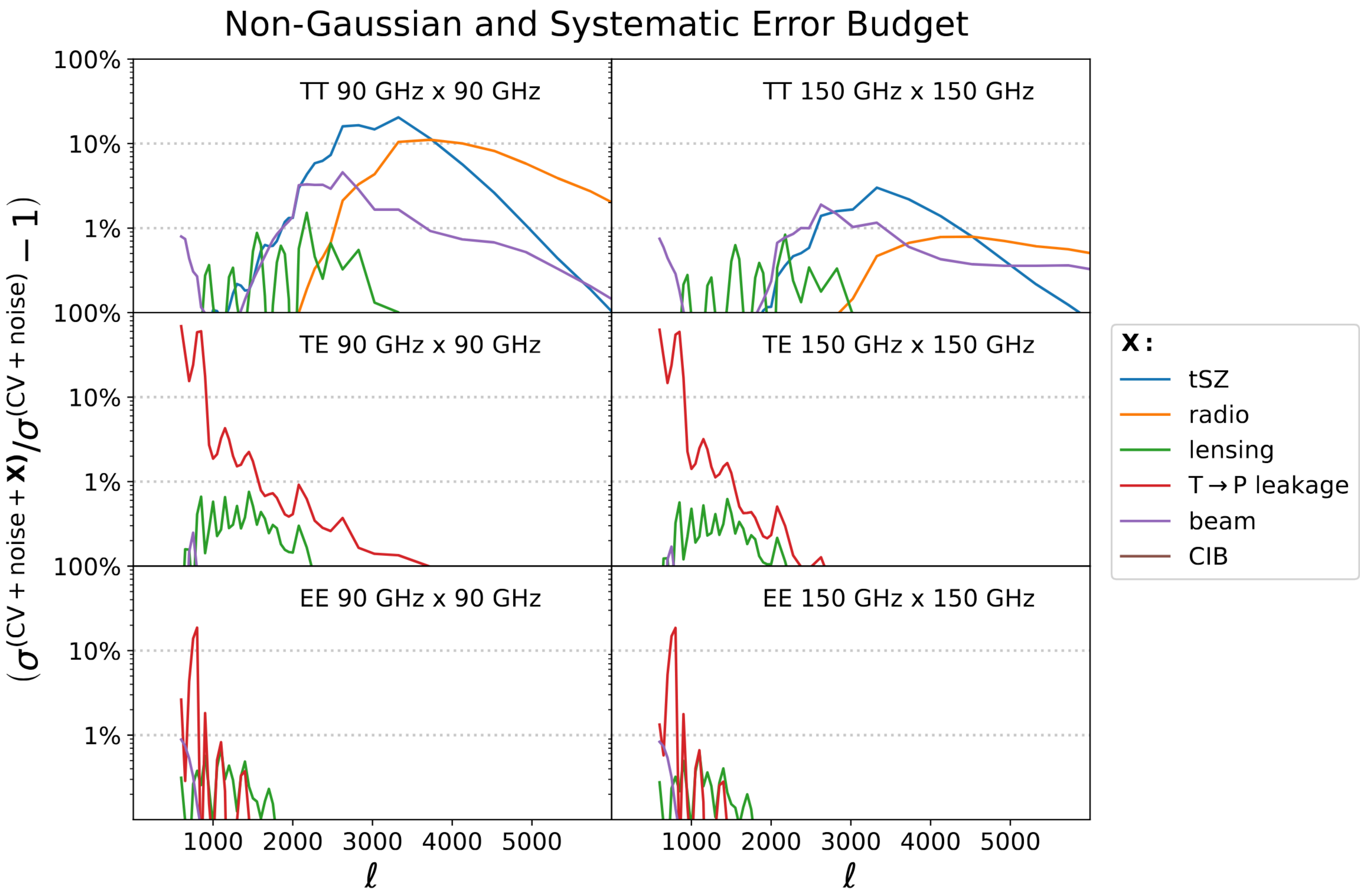
Difficult to get good SNR measurements of the leakage at large angular scales



BEAM LEAKAGE MODEL

Difficult to get good SNR measurements of the leakage at large angular scales

- Leakage error dominates error budget at large scales



Louis et al, 2025 (2503.14452)

POLARIZED BEAM

$$\begin{pmatrix} \boxed{B^{II}} & B^{IQ} & B^{IU} & B^{IV} \\ B^{QI} & B^{QQ} & B^{QU} & B^{QV} \\ B^{UI} & B^{UQ} & B^{UU} & B^{UV} \\ B^{VI} & B^{VQ} & B^{VU} & B^{VV} \end{pmatrix}$$

POLARIZED BEAM

$$\begin{pmatrix} B^{II} & B^{IQ} & B^{IU} & B^{IV} \\ B^{QI} & B^{QQ} & B^{QU} & B^{QV} \\ B^{UI} & B^{UQ} & B^{UU} & B^{UV} \\ B^{VI} & B^{VQ} & B^{VU} & B^{VV} \end{pmatrix}$$

POLARIZED BEAM

$$\begin{pmatrix} B^{II} & B^{IQ} & B^{IU} & B^{IV} \\ B^{QI} & B^{QQ} & B^{QU} & B^{QV} \\ B^{UI} & B^{UQ} & B^{UU} & B^{UV} \\ B^{VI} & B^{VQ} & B^{VU} & B^{VV} \end{pmatrix}$$

POLARIZED BEAM

$$\begin{pmatrix} B^{II} & B^{IQ} & B^{IU} & B^{IV} \\ B^{QI} & B^{QQ} & B^{QU} & B^{QV} \\ B^{UI} & B^{UQ} & B^{UU} & B^{UV} \\ B^{VI} & B^{VQ} & B^{VU} & B^{VV} \end{pmatrix}$$

- How do we directly measure our coupling to the polarized sky?
  - Polarized astrophysical sources

POLARIZED BEAM

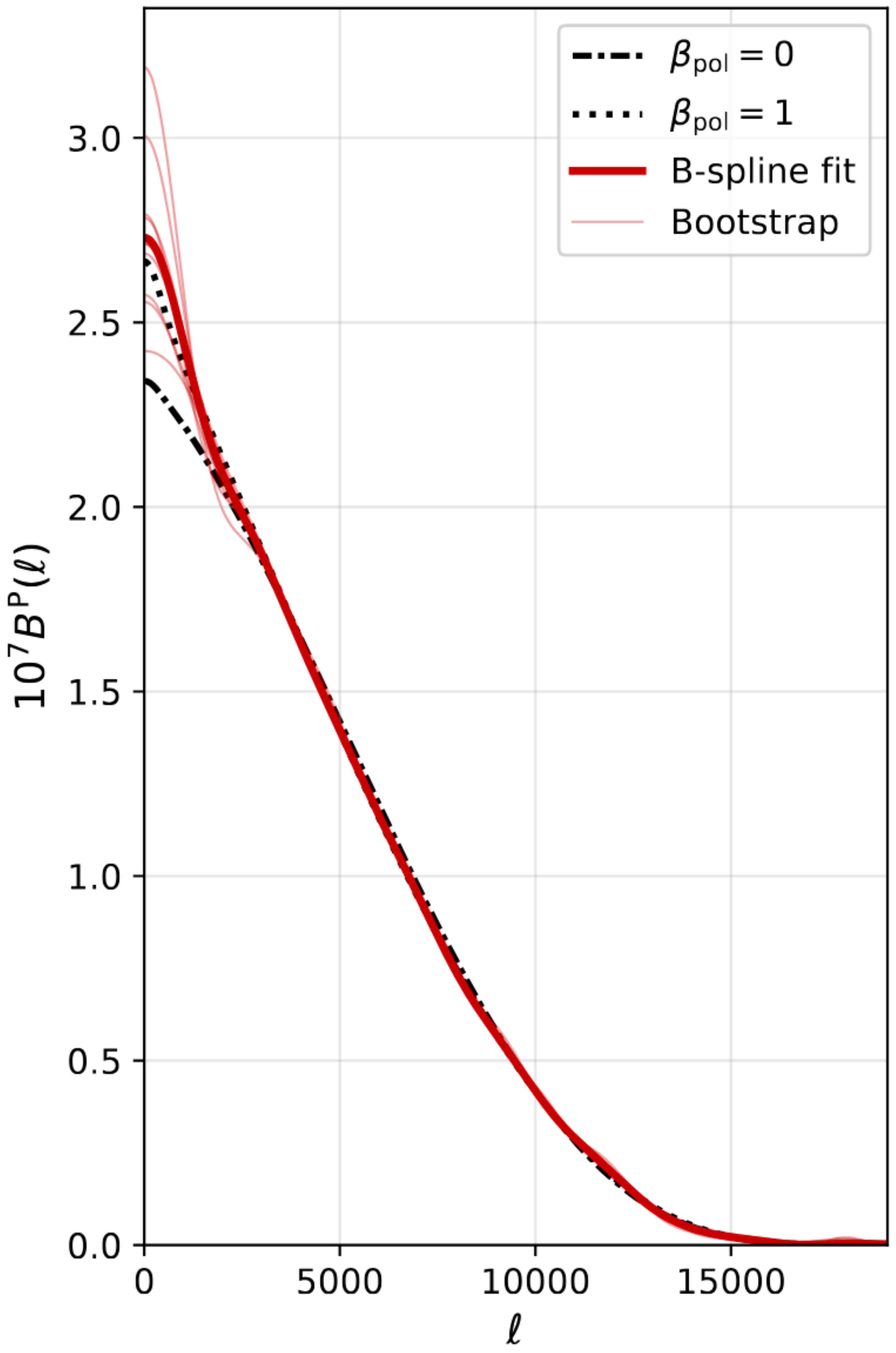
$$\begin{pmatrix} B^{II} & B^{IQ} & B^{IU} & B^{IV} \\ B^{QI} & B^{QQ} & B^{QU} & B^{QV} \\ B^{UI} & B^{UQ} & B^{UU} & B^{UV} \\ B^{VI} & B^{VQ} & B^{VU} & B^{VV} \end{pmatrix}$$

- How do we directly measure our coupling to the polarized sky?
  - Polarized astrophysical sources

POLARIZED BEAM

$$\begin{pmatrix} B^{II} & B^{IQ} & B^{IU} & B^{IV} \\ B^{QI} & B^{QQ} & B^{QU} & B^{QV} \\ B^{UI} & B^{UQ} & B^{UU} & B^{UV} \\ B^{VI} & B^{VQ} & B^{VU} & B^{VV} \end{pmatrix}$$

- How do we directly measure our coupling to the polarized sky?
  - Polarized astrophysical sources

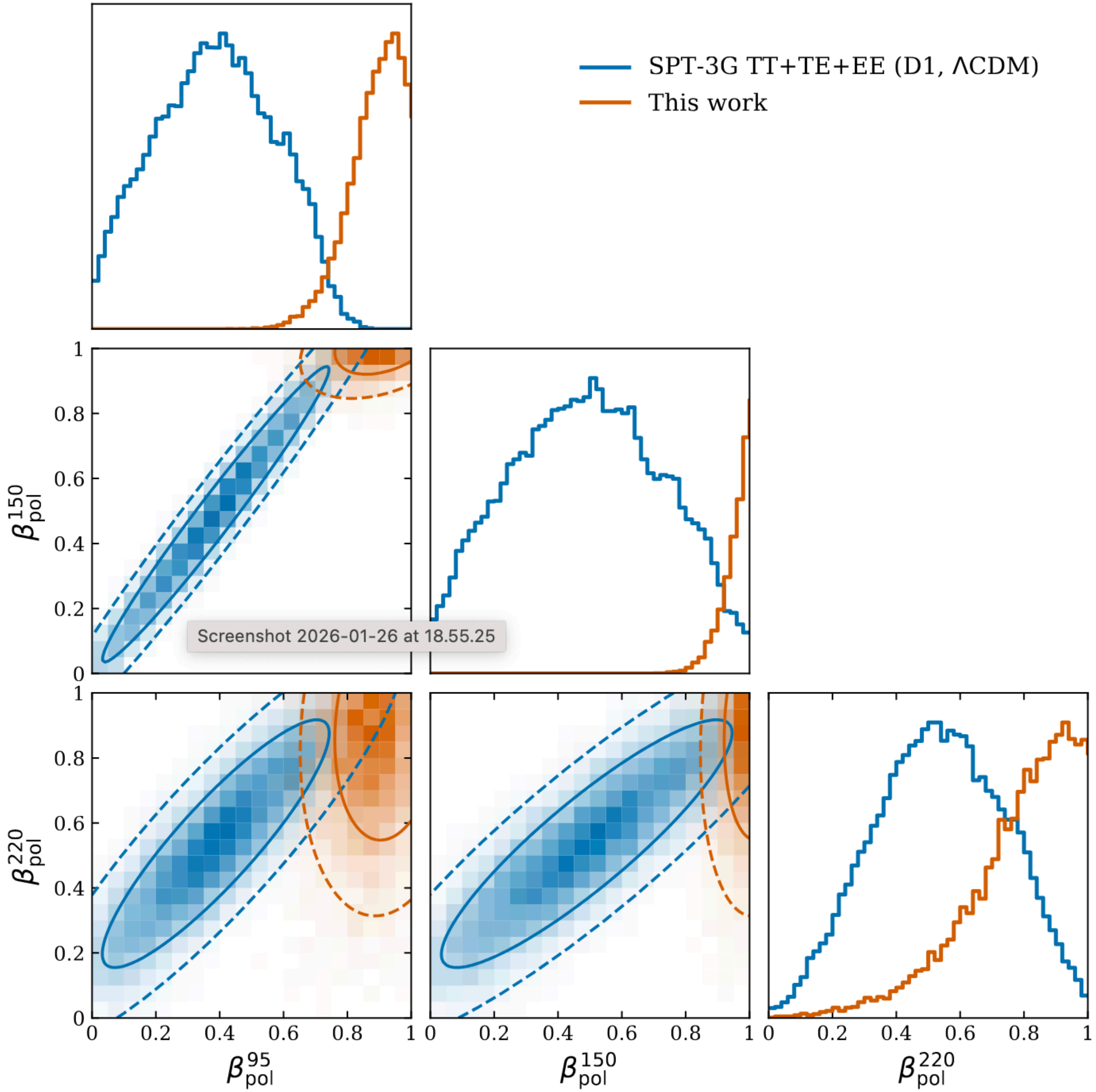


de Haan et al., 2026 (2602.06334)

POLARIZED BEAM

$$\begin{pmatrix} B^{II} & B^{IQ} & B^{IU} & B^{IV} \\ B^{QI} & B^{QQ} & B^{QU} & B^{QV} \\ B^{UI} & B^{UQ} & B^{UU} & B^{UV} \\ B^{VI} & B^{VQ} & B^{VU} & B^{VV} \end{pmatrix}$$

- How do we directly measure our coupling to the polarized sky?
  - Polarized astrophysical sources

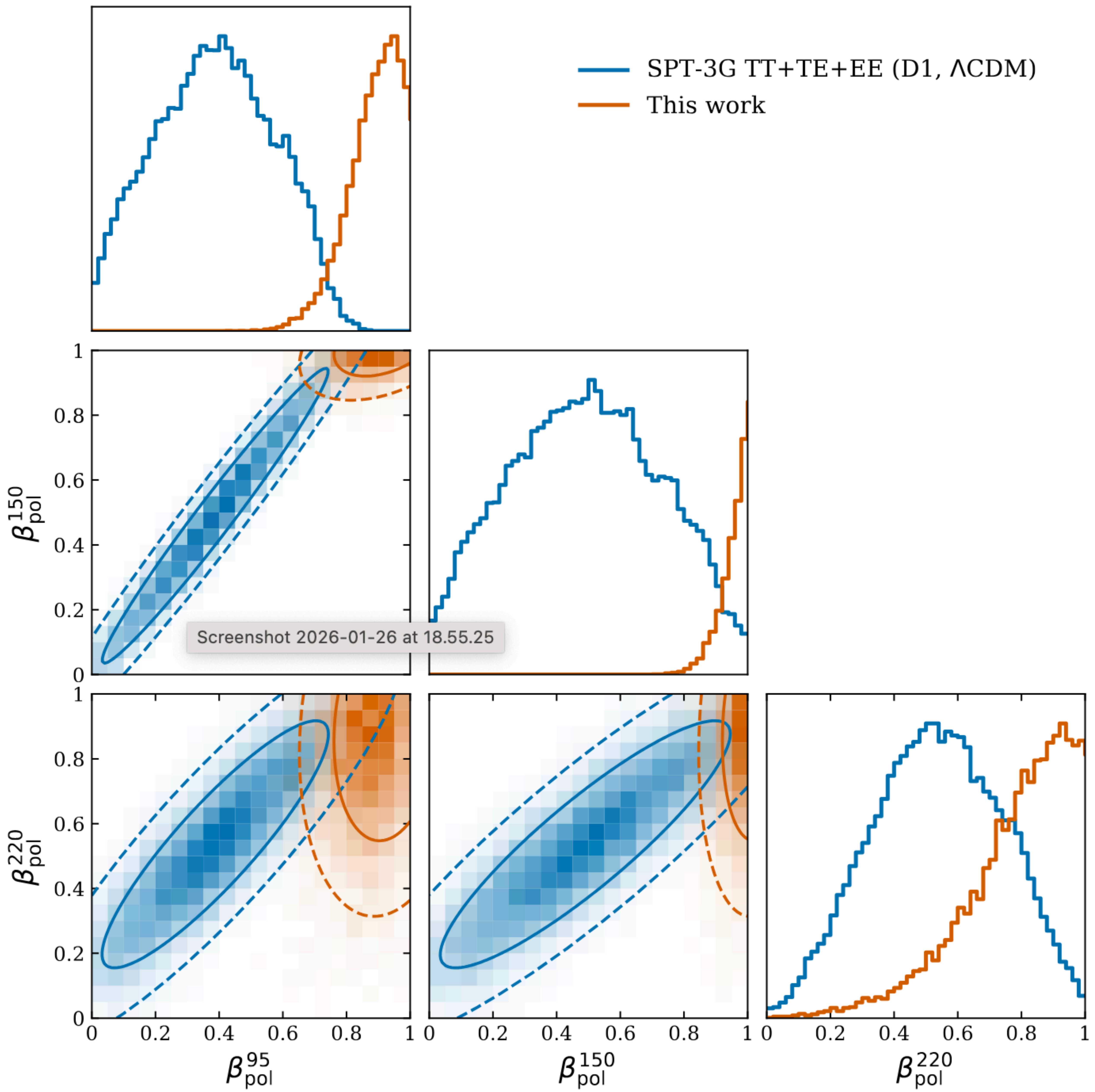


de Haan et al., 2026 (2602.06334)

POLARIZED BEAM

$$\begin{pmatrix} B^{II} & B^{IQ} & B^{IU} & B^{IV} \\ B^{QI} & B^{QQ} & B^{QU} & B^{QV} \\ B^{UI} & B^{UQ} & B^{UU} & B^{UV} \\ B^{VI} & B^{VQ} & B^{VU} & B^{VV} \end{pmatrix}$$

- How do we directly measure our coupling to the polarized sky?
    - Polarized astrophysical sources
    - "Repurpose" calibration sources designed for polarization angle calibration, e.g. drone calibration

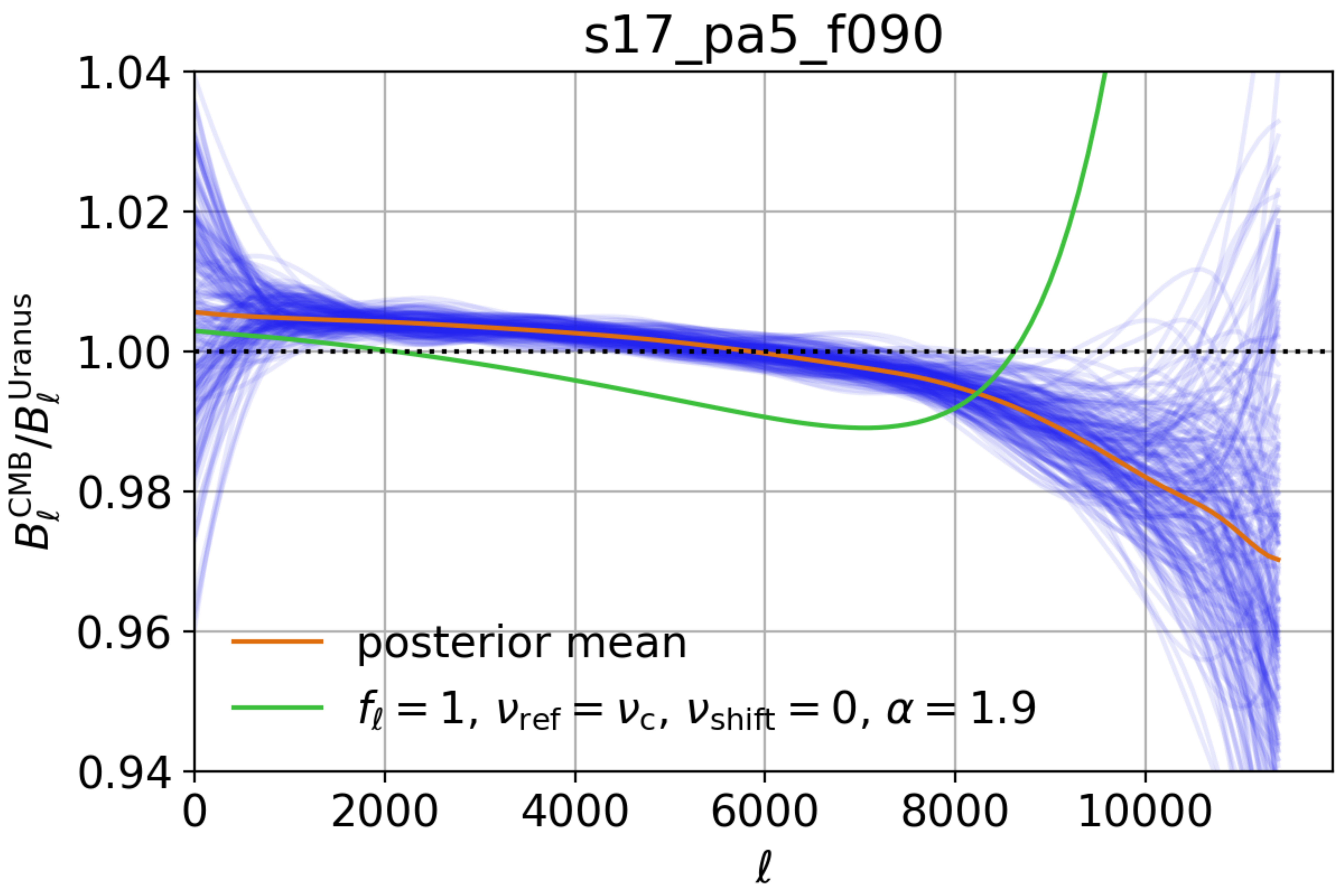


de Haan et al., 2026 (2602.06334)

BEAM CHROMATICITY

Frequency-dependence of the beam slightly changes beam shape for each sky component

- $B_\ell^c = \int B_\ell(\nu)\, I^c(\nu) \tau(\nu) \mathrm{d}\nu$

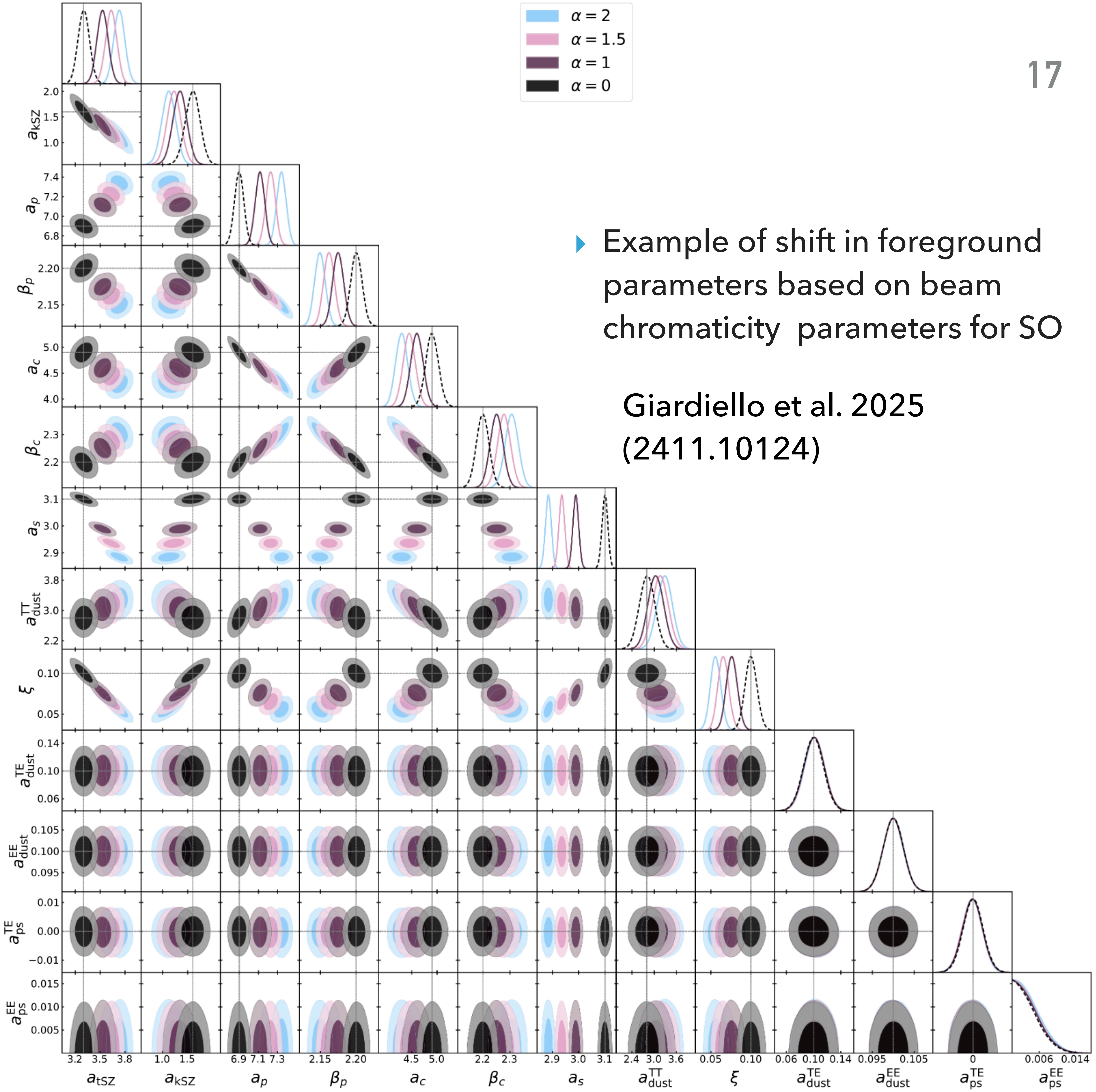


- Example of color-corrected beam for the CMB sky component

EFFECT ON PARAMETERS

Frequency-dependence of the beam slightly changes beam shape for each sky component

- $B_\ell^c = \int B_\ell(\nu)\, I^c(\nu) \tau(\nu) \mathrm{d}\nu$
- How to measure $B_\ell(\nu)$?
  - For ACT DR6 determined solely from planet observations with priors from optical simulations
  - Future: measure beams jointly from planets, quasars and dusty sources



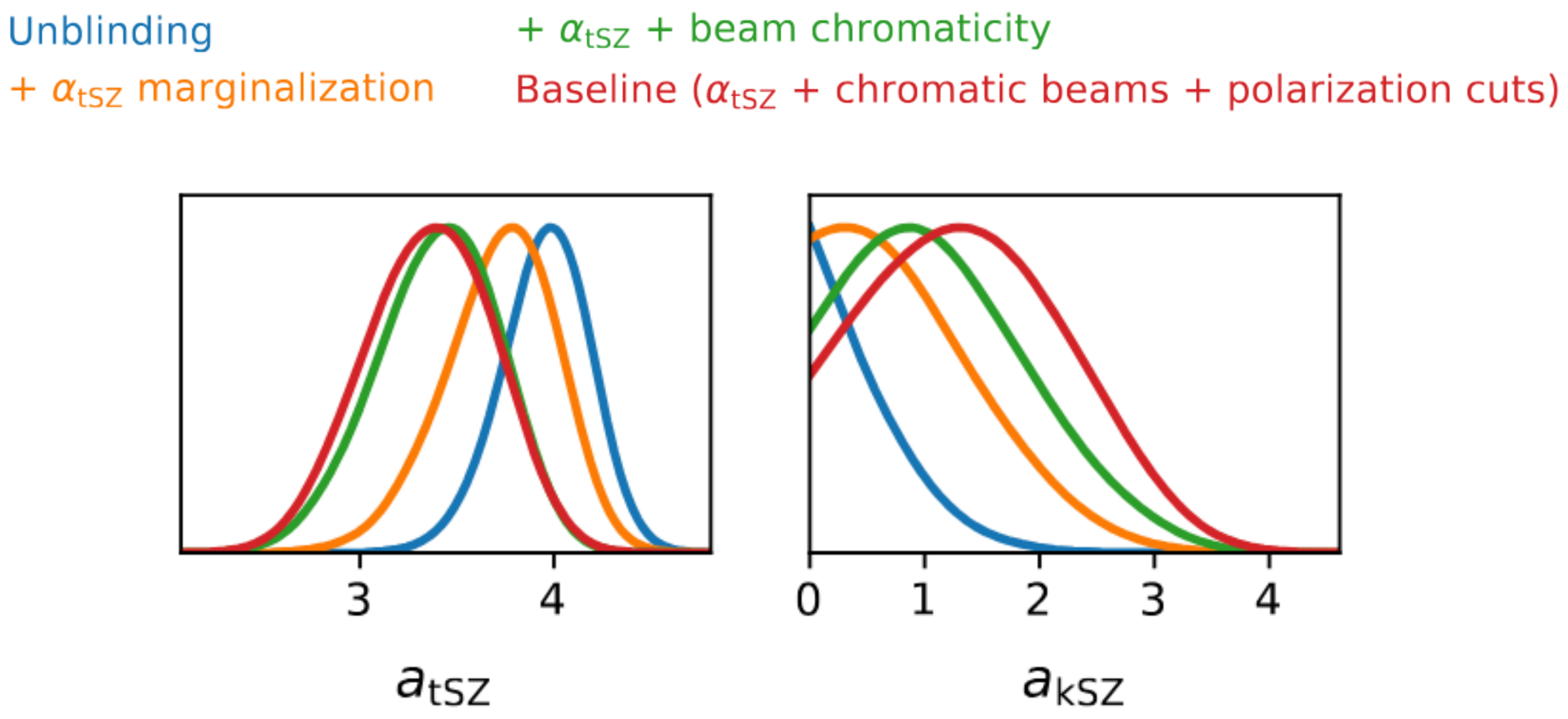
- Example of shift in foreground parameters based on beam chromaticity parameters for SO

Giardiello et al. 2025 (2411.10124)

EFFECT ON PARAMETERS

Frequency-dependence of the beam slightly changes beam shape for each sky component

- $$B^c_\ell = \int B_\ell(\nu)\, I^c(\nu)\tau(\nu)\mathrm{d}\nu$$
- How to measure $B_\ell(\nu)$?
  - For ACT DR6 determined solely from planet observations with priors from optical simulations
  - Future: measure beams jointly from planets, quasars and dusty sources



- ACT DR6 shifts in foreground parameters due to beam chromaticity

Louis et al, 2025 (2503.14452)

MARGINALIZATION OVER NUISANCE PARAMETERS

- $P(\vec{\theta}\,|\,\vec{d}) = \int P(\vec{\theta}, \vec{\phi}\,|\,\vec{d})\,\mathrm{d}\vec{\phi}$

- $\vec{\theta}$: Cosmological parameters, e.g. *r*
- $\vec{d}$: Data (noisy sky maps for each frequency band)
- $\vec{\phi}$: Nuisance parameters (sky realization, instrumental systematics, etc.)

MARGINALIZATION OVER NUISANCE PARAMETERS

- $P(\vec{\theta}\,|\,\vec{d}) = \int P(\vec{\theta}, \vec{\phi}\,|\,\vec{d})\,\mathrm{d}\vec{\phi}$

- $\vec{\theta}$: Cosmological parameters, e.g. r
- $\vec{d}$: Data (noisy sky maps for each frequency band)
- $\vec{\phi}$: Nuisance parameters (sky realization, instrumental systematics, etc.)

- Option 1: sample from $P(\vec{\theta}, \vec{\phi}\,|\,\vec{d})$ using MCMC

MARGINALIZATION OVER NUISANCE PARAMETERS

Option 1: sample from $P(\vec{\theta}, \vec{\phi} \mid \vec{d})$
using MCMC

MARGINALIZATION OVER NUISANCE PARAMETERS

Option 1: sample from $P(\vec{\theta}, \overrightarrow{\phi} \,|\, \vec{d})$ using MCMC

- Default approach: find low-dimensional parameterization for each systematic deemed important enough to include

MARGINALIZATION OVER NUISANCE PARAMETERS

Option 1: sample from $P(\vec{\theta}, \overrightarrow{\phi} \,|\, \vec{d})$ using MCMC

- Default approach: find low-dimensional parameterization for each systematic deemed important enough to include
- Does this approach scale to upcoming experiments?
  - Largely solved by moving to gradient-based sampling, e.g. Candl differentiable likelihood (Balkenhol et al. 2024 2401.13433)

MARGINALIZATION OVER NUISANCE PARAMETERS

Option 1: sample from $P(\vec{\theta}, \overrightarrow{\phi} \,|\, \vec{d})$ using MCMC

- Default approach: find low-dimensional parameterization for each systematic deemed important enough to include
- Does this approach scale to upcoming experiments?
  - Largely solved by moving to gradient-based sampling, e.g. Candl differentiable likelihood (Balkenhol et al. 2024 2401.13433)

- Low-dimensional parameterizations all over the place:
  - Beam and filters, B_ℓ and F_ℓ
  - passband errors $\tau(\nu + \delta\nu)$

MARGINALIZATION OVER NUISANCE PARAMETERS

Option 1: sample from $P(\vec{\theta}, \vec{\phi} \,|\, \vec{d})$ using MCMC

- Default approach: find low-dimensional parameterization for each systematic deemed important enough to include
- Does this approach scale to upcoming experiments?
  - Largely solved by moving to gradient-based sampling, e.g. Candl differentiable likelihood (Balkenhol et al. 2024 2401.13433)

- Low-dimensional parameterizations all over the place:
  - Beam and filters, B_ℓ and F_ℓ
  - passband errors $\tau(\nu + \delta\nu)$
- Finding these models is often very time-consuming
  - Sometimes simple parameterizations are not possible → we cut data

MARGINALIZATION OVER NUISANCE PARAMETERS

Option 1: sample from $P(\vec{\theta}, \overrightarrow{\phi} \,|\, \vec{d})$ using MCMC

- Default approach: find low-dimensional parameterization for each systematic deemed important enough to include
- Does this approach scale to upcoming experiments?
  - Largely solved by moving to gradient-based sampling, e.g. Candl differentiable likelihood (Balkenhol et al. 2024 2401.13433)

- Low-dimensional parameterizations all over the place:
  - Beam and filters, B_ℓ and F_ℓ
  - passband errors $\tau(\nu + \delta\nu)$
- Finding these models is often very time-consuming
  - Sometimes simple parameterizations are not possible → we cut data
- Simulating the systematic is often easier

MARGINALIZATION OVER NUISANCE PARAMETERS

- $P(\vec{\theta}\,|\,\vec{d}) = \int P(\vec{\theta}, \vec{\phi}\,|\,\vec{d})\,\mathrm{d}\vec{\phi}$

- $\vec{\theta}$: Cosmological parameters, e.g. r

- $\vec{d}$: Data (noisy sky maps for each frequency band)

- $\vec{\phi}$: Nuisance parameters (sky realization, instrumental systematics, etc.)

- Option 1: sample from $P(\vec{\theta}, \vec{\phi}\,|\,\vec{d})$ using MCMC

MARGINALIZATION OVER NUISANCE PARAMETERS

- $P(\vec{\theta}\,|\,\vec{d}) = \int P(\vec{\theta}, \vec{\phi}\,|\,\vec{d})\,\mathrm{d}\vec{\phi}$

- $\vec{\theta}$: Cosmological parameters, e.g. *r*
- $\vec{d}$: Data (noisy sky maps for each frequency band)
- $\vec{\phi}$: Nuisance parameters (sky realization, instrumental systematics, etc.)

- Option 1: sample from $P(\vec{\theta}, \vec{\phi}\,|\,\vec{d})$ using MCMC

- Option 2: directly target $P(\vec{\theta}\,|\,\vec{d})$ using simulation-based inference
  - Well-suited for systematics-dominated analyses, like CMB searches for r
  - NB. really target $P(\vec{\theta}\,|\,\vec{x})$, where $\vec{x} = f(\vec{d})$ is a compressed representation of the data that suppresses part of its complexity

NEURAL POSTERIOR ESTIMATION

NEURAL POSTERIOR ESTIMATION

Density estimator $q_{\vec{\lambda}}(\vec{\theta}\,|\,\vec{x})$ to approximate $P(\vec{\theta}\,|\,\vec{x})$

- Typically q is a (small) neural network, e.g. a normalizing flow

$$\arg\min_{\vec{\lambda}} D_{\mathrm{KL}}\left(P(\vec{\theta},\vec{x}) \;||\; q_{\vec{\lambda}}\big(\theta\,|\,\vec{x}\big)P(\vec{x}) \right)$$

NEURAL POSTERIOR ESTIMATION

Density estimator $q_{\vec{\lambda}}(\vec{\theta}\,|\,\vec{x})$ to approximate $P(\vec{\theta}\,|\,\vec{x})$

- Typically q is a (small) neural network, e.g. a normalizing flow

$$\arg\min_{\vec{\lambda}} D_{\mathrm{KL}}\left(P(\vec{\theta},\vec{x}) \;||\; q_{\vec{\lambda}}\big(\theta\,|\,\vec{x}\big)P(\vec{x}) \right)$$

$$= \arg\min_{\vec{\lambda}} \int P(\vec{\theta},\vec{x}) \log \frac{P(\vec{\theta}\,|\,\vec{x})P(\vec{x})}{q_{\vec{\lambda}}\big(\vec{\theta}\,|\,\vec{x}\big)P(\vec{x})} \mathrm{d}\vec{\theta}\,\mathrm{d}\vec{x}$$

NEURAL POSTERIOR ESTIMATION

Density estimator $q_{\vec{\lambda}}(\vec{\theta}\,|\,\vec{x})$ to approximate $P(\vec{\theta}\,|\,\vec{x})$

- Typically q is a (small) neural network, e.g. a normalizing flow

$$\arg\min_{\vec{\lambda}} D_{\mathrm{KL}}\left(P(\vec{\theta},\vec{x})\ ||\ q_{\vec{\lambda}}\big(\theta\,|\,\vec{x}\big)P(\vec{x})\right)$$

$$= \arg\min_{\vec{\lambda}} \int P(\vec{\theta},\vec{x})\log\frac{P(\vec{\theta}\,|\,\vec{x})P(\vec{x})}{q_{\vec{\lambda}}\big(\vec{\theta}\,|\,\vec{x}\big)P(\vec{x})}\mathrm{d}\vec{\theta}\,\mathrm{d}\vec{x}$$

$$= \arg\min_{\vec{\lambda}} \int -P(\vec{\theta},\vec{x})\log q_{\vec{\lambda}}\big(\vec{\theta}\,|\,\vec{x}\big)\mathrm{d}\vec{\theta}\,\mathrm{d}\vec{x}$$

NEURAL POSTERIOR ESTIMATION

Density estimator $q_{\vec{\lambda}}(\vec{\theta}\,|\,\vec{x})$ to approximate $P(\vec{\theta}\,|\,\vec{x})$

- Typically q is a (small) neural network, e.g. a normalizing flow

$$\arg\min_{\vec{\lambda}} D_{\mathrm{KL}}\left(P(\vec{\theta},\vec{x})\ ||\ q_{\vec{\lambda}}\big(\theta\,|\,\vec{x}\big)P(\vec{x})\right)$$

$$= \arg\min_{\vec{\lambda}} \int P(\vec{\theta},\vec{x})\log\frac{P(\vec{\theta}\,|\,\vec{x})P(\vec{x})}{q_{\vec{\lambda}}\big(\vec{\theta}\,|\,\vec{x}\big)P(\vec{x})}\mathrm{d}\vec{\theta}\,\mathrm{d}\vec{x}$$

$$= \arg\min_{\vec{\lambda}} \int -P(\vec{\theta},\vec{x})\log q_{\vec{\lambda}}\big(\vec{\theta}\,|\,\vec{x}\big)\mathrm{d}\vec{\theta}\,\mathrm{d}\vec{x}$$

$$\approx \arg\min_{\vec{\lambda}} \sum_{i=1}^{N_{\mathrm{sim}}} -\log q_{\vec{\lambda}}\big(\vec{\theta}_i\,|\,\vec{x}_i\big)$$, sum over $\{\theta_i, x_i\}_{i=1}^{N_{\mathrm{sim}}}$

NEURAL POSTERIOR ESTIMATION

Density estimator $q_{\vec{\lambda}}(\vec{\theta}\,|\,\vec{x})$ to approximate $P(\vec{\theta}\,|\,\vec{x})$

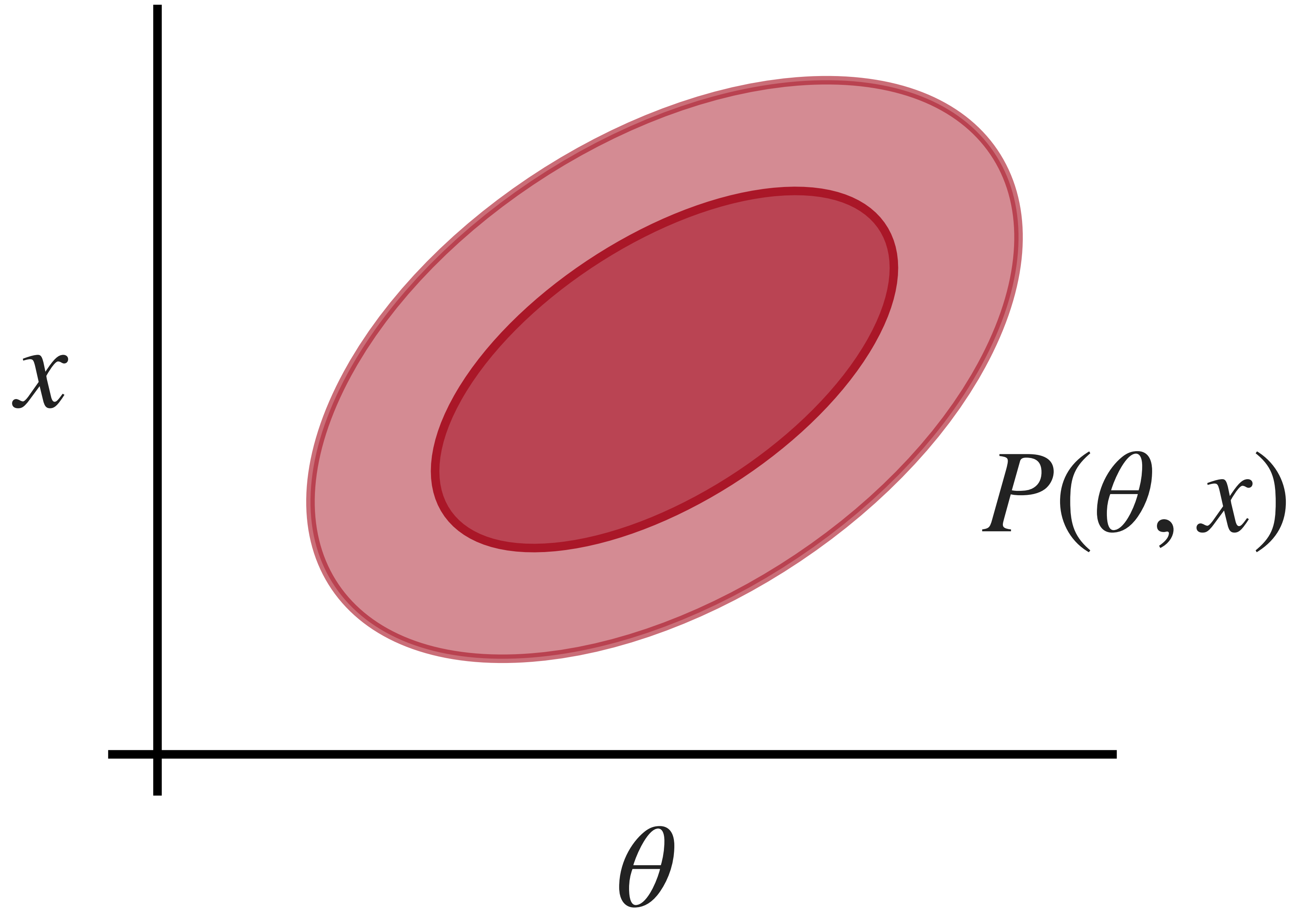
- Typically q is a (small) neural network, e.g. a normalizing flow

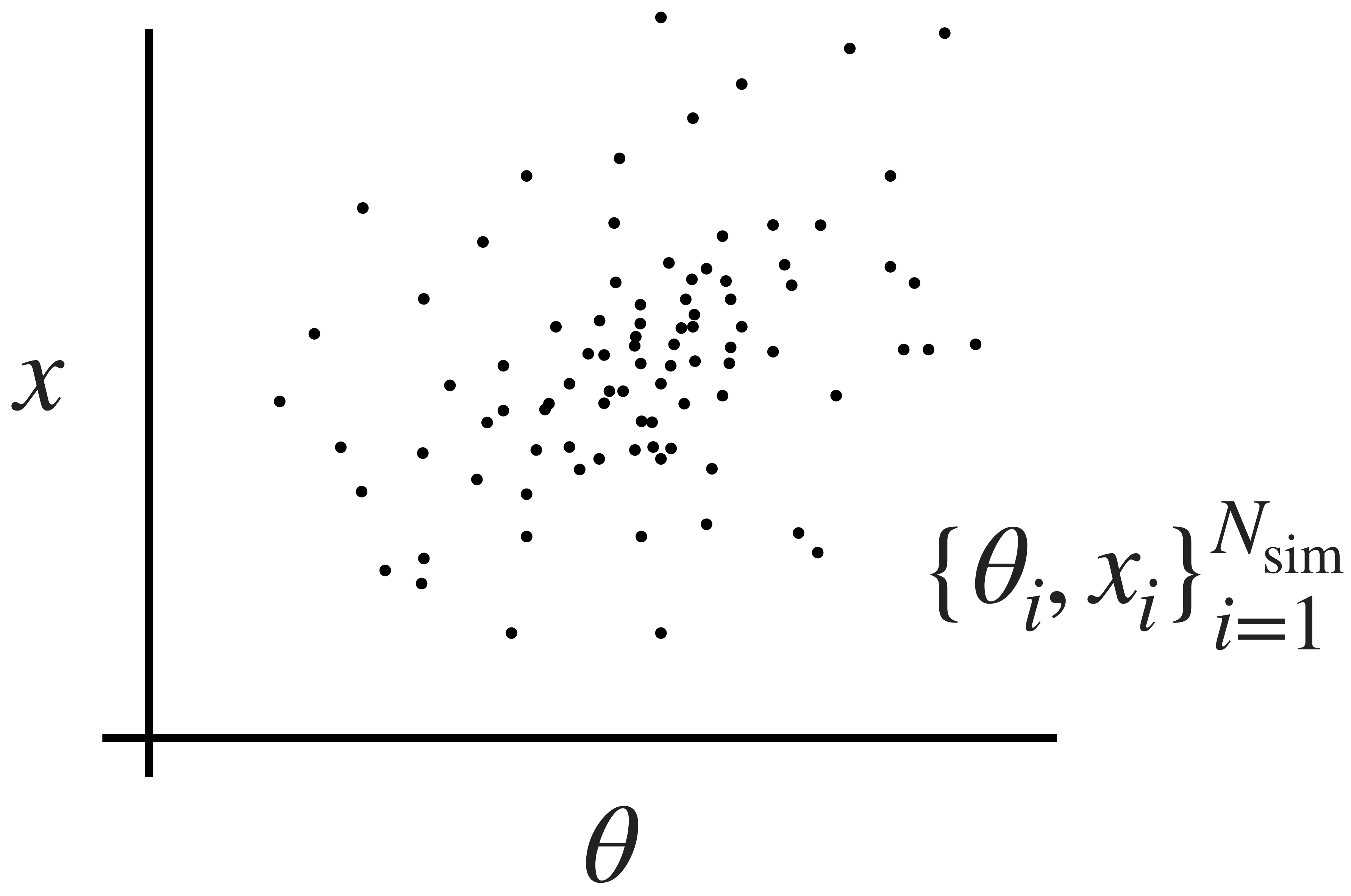
$$\arg\min_{\vec{\lambda}} D_{\mathrm{KL}}\left(P(\vec{\theta},\vec{x}) \;||\; q_{\vec{\lambda}}\left(\theta\,|\,\vec{x}\right)P(\vec{x})\right)$$

$$= \arg\min_{\vec{\lambda}} \int P(\vec{\theta},\vec{x})\log \frac{P(\vec{\theta}\,|\,\vec{x})P(\vec{x})}{q_{\vec{\lambda}}\left(\vec{\theta}\,|\,\vec{x}\right)P(\vec{x})}\mathrm{d}\vec{\theta}\,\mathrm{d}\vec{x}$$

$$= \arg\min_{\vec{\lambda}} \int -P(\vec{\theta},\vec{x})\log q_{\vec{\lambda}}\left(\vec{\theta}\,|\,\vec{x}\right)\mathrm{d}\vec{\theta}\,\mathrm{d}\vec{x}$$

$$\approx \arg\min_{\vec{\lambda}} \sum_{i=1}^{N_{\mathrm{sim}}} -\log q_{\vec{\lambda}}\left(\vec{\theta}_i\,|\,\vec{x}_i\right)$$, sum over $\{\theta_i, x_i\}_{i=1}^{N_{\mathrm{sim}}}$





NEURAL POSTERIOR ESTIMATION

Density estimator $q_{\vec{\lambda}}(\vec{\theta}\,|\,\vec{x})$ to approximate $P(\vec{\theta}\,|\,\vec{x})$

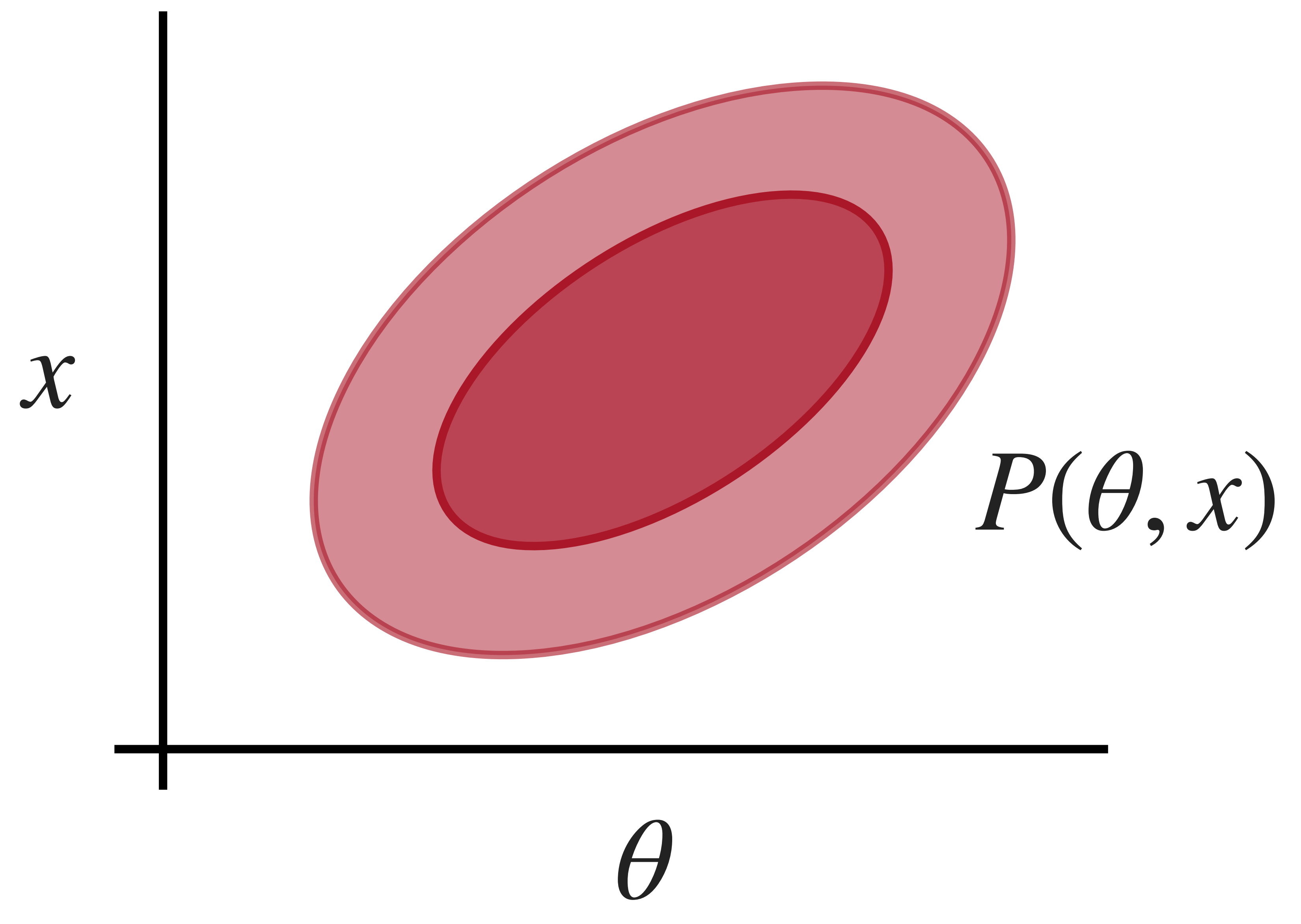
- Typically q is a (small) neural network, e.g. a normalizing flow

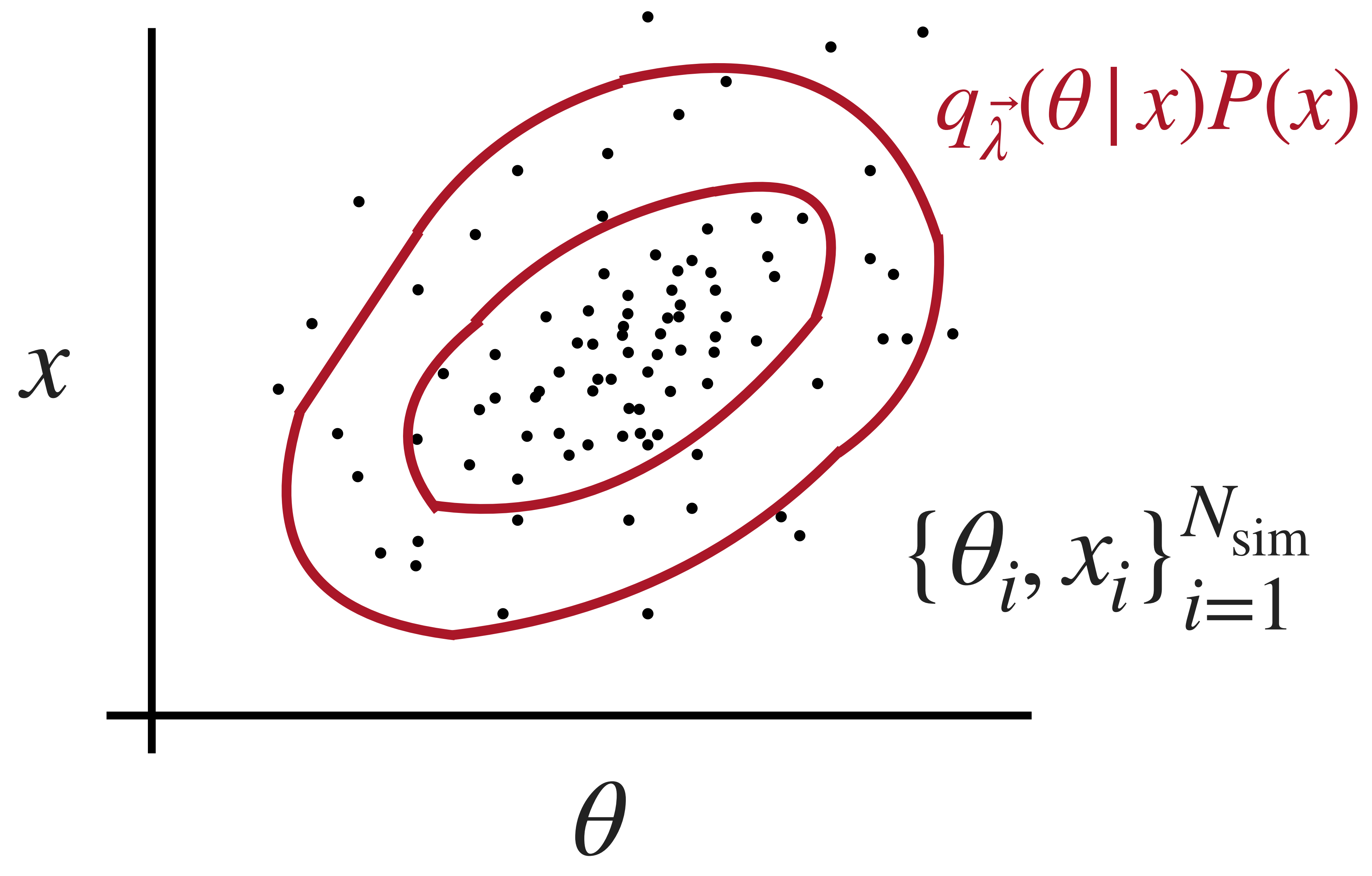
$$\arg\min_{\vec{\lambda}} D_{\mathrm{KL}}\left(P(\vec{\theta},\vec{x})\ ||\ q_{\vec{\lambda}}\big(\theta\,|\,\vec{x}\big)P(\vec{x})\right)$$

$$= \arg\min_{\vec{\lambda}} \int P(\vec{\theta},\vec{x})\log\frac{P(\vec{\theta}\,|\,\vec{x})P(\vec{x})}{q_{\vec{\lambda}}\big(\vec{\theta}\,|\,\vec{x}\big)P(\vec{x})}\mathrm{d}\vec{\theta}\,\mathrm{d}\vec{x}$$

$$= \arg\min_{\vec{\lambda}} \int -P(\vec{\theta},\vec{x})\log q_{\vec{\lambda}}\big(\vec{\theta}\,|\,\vec{x}\big)\mathrm{d}\vec{\theta}\,\mathrm{d}\vec{x}$$

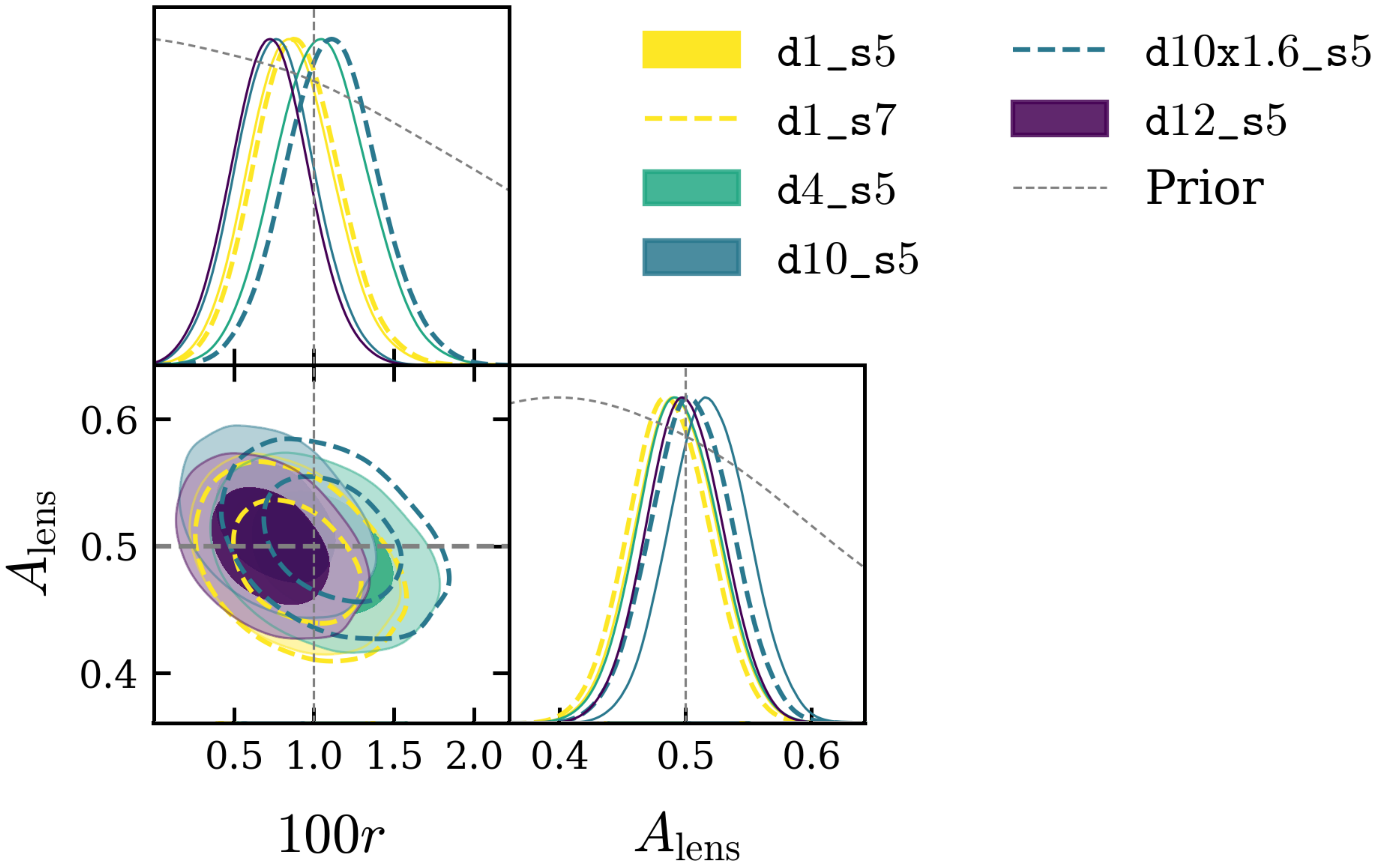
$$\approx \arg\min_{\vec{\lambda}} \sum_{i=1}^{N_{\mathrm{sim}}} -\log q_{\vec{\lambda}}\big(\vec{\theta}_i\,|\,\vec{x}_i\big)$$, sum over $\{\theta_i, x_i\}_{i=1}^{N_{\mathrm{sim}}}$





PROOF-OF-CONCEPT FOCUSED ON FOREGROUND CLEANING

- Unbiased r inference with complicated foregrounds using SBI trained on very simple foreground simulations
    - Exploits SBI's ability to do inference with aggressively compressed data vectors



AJD, K. Surrao, A. E. Bayer, A. E. Adler, N. Dachlythra, S. Azzoni and J. C. Hill (2512.16869)

GENERATING

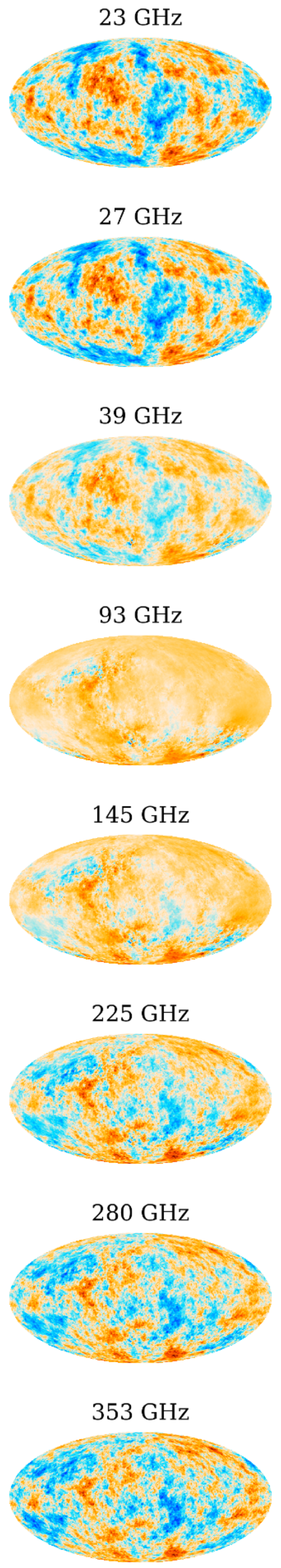
$$\{\vec{\theta}_i, \vec{x}_i\}_{i=1}^{N_{\rm sim}}$$

$$\vec{\theta}_i \sim P(\vec{\theta})$$

Draw from prior

→

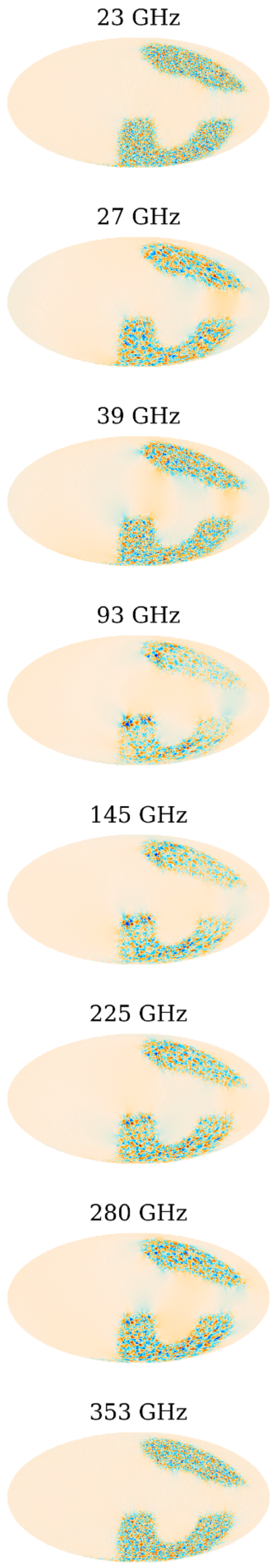
$$\overrightarrow{\phi}_i \sim P(\overrightarrow{\phi} \,|\, \vec{\theta}_i)$$



Draw sky realization

→

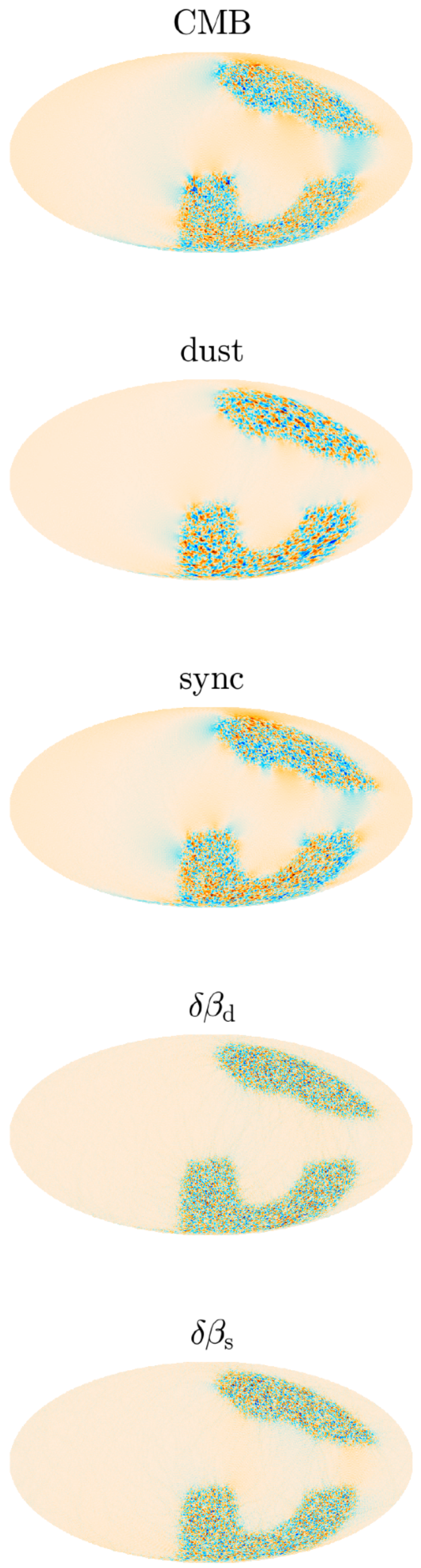
$$\vec{d}_i \sim P(\vec{d} \,|\, \overrightarrow{\phi}_i)$$



Convolve with beam, filter, add noise, mask

→

$$\vec{x}_i = f(\vec{d}_i)$$



→

$$\left\{ \begin{array}{c} \hat{C}_\ell^{\rm CMB \times CMB} \\ \hat{C}_\ell^{\rm CMB \times dust} \\ \vdots \\ \hat{C}_\ell^{{\rm sync} \times \delta_{\beta_{\rm s}}} \\ \hat{C}_\ell^{\delta_{\beta_{\rm s}} \times \delta_{\beta_{\rm s}}} \end{array} \right\}$$

15 spectra,
165 bins total

Apply Needlet-ILC, compute *B*-mode cross-spectra

NUISANCE PARAMETERS IN SBI

NUISANCE PARAMETERS IN SBI

Consider parameter of interest θ_1 and nuisance parameter θ_2

- We directly target $P(\theta_1 \,|\, \vec{x}) = \int \mathrm{d}\theta_2 P(\theta_1, \theta_2 \,|\, \vec{x})$

$$\arg\min_{\vec{\lambda}} D_{\mathrm{KL}}\left(P(\theta_1 \,|\, \vec{x}) P(\vec{x}) \;||\; q_{\vec{\lambda}}(\theta_1 \,|\, \vec{x}) P(\vec{x}) \right)$$

NUISANCE PARAMETERS IN SBI

Consider parameter of interest θ_1 and nuisance parameter θ_2

- We directly target $P(\theta_1 | \vec{x}) = \int \mathrm{d}\theta_2 P(\theta_1, \theta_2 | \vec{x})$

$$\arg\min_{\vec{\lambda}} D_{\mathrm{KL}}\left(P(\theta_1 | \vec{x}) P(\vec{x}) \;||\; q_{\vec{\lambda}}\left(\theta_1 | \vec{x}\right) P(\vec{x}) \right)$$

$$= \arg\min_{\vec{\lambda}} \int P(\theta_1 | \vec{x}) P(\vec{x}) \log \frac{P(\theta_1 | \vec{x}) P(\vec{x})}{q_{\vec{\lambda}}\left(\theta_1 | \vec{x}\right) P(\vec{x})} \mathrm{d}\theta_1 \, \mathrm{d}\vec{x}$$

NUISANCE PARAMETERS IN SBI

Consider parameter of interest θ_1 and nuisance parameter θ_2

- We directly target $P(\theta_1 \,|\, \vec{x}) = \int \mathrm{d}\theta_2 P(\theta_1, \theta_2 \,|\, \vec{x})$

$$\arg\min_{\vec{\lambda}} D_{\mathrm{KL}}\left(P(\theta_1 \,|\, \vec{x})P(\vec{x}) \;||\; q_{\vec{\lambda}}\big(\theta_1 \,|\, \vec{x}\big) P(\vec{x}) \right)$$

$$= \arg\min_{\vec{\lambda}} \int P(\theta_1 \,|\, \vec{x})P(\vec{x}) \log \frac{P(\theta_1 \,|\, \vec{x})P(\vec{x})}{q_{\vec{\lambda}}\big(\theta_1 \,|\, \vec{x}\big) P(\vec{x})} \mathrm{d}\theta_1 \, \mathrm{d}\vec{x}$$

$$= \arg\min_{\vec{\lambda}} \int - P(\theta_1, \theta_2, \vec{x}) \log q_{\vec{\lambda}}\big(\theta_1 \,|\, \vec{x}\big) \mathrm{d}\theta_1 \mathrm{d}\theta_2 \, \mathrm{d}\vec{x}$$

NUISANCE PARAMETERS IN SBI

Consider parameter of interest θ_1 and nuisance parameter θ_2

- We directly target $P(\theta_1 \,|\, \vec{x}) = \int \mathrm{d}\theta_2 P(\theta_1, \theta_2 \,|\, \vec{x})$

$$\arg\min_{\vec{\lambda}} D_{\mathrm{KL}}\left(P(\theta_1 \,|\, \vec{x})P(\vec{x}) \;||\; q_{\vec{\lambda}}\big(\theta_1 \,|\, \vec{x}\big) P(\vec{x}) \right)$$

$$= \arg\min_{\vec{\lambda}} \int P(\theta_1 \,|\, \vec{x})P(\vec{x}) \log \frac{P(\theta_1 \,|\, \vec{x})P(\vec{x})}{q_{\vec{\lambda}}\big(\theta_1 \,|\, \vec{x}\big) P(\vec{x})} \mathrm{d}\theta_1 \, \mathrm{d}\vec{x}$$

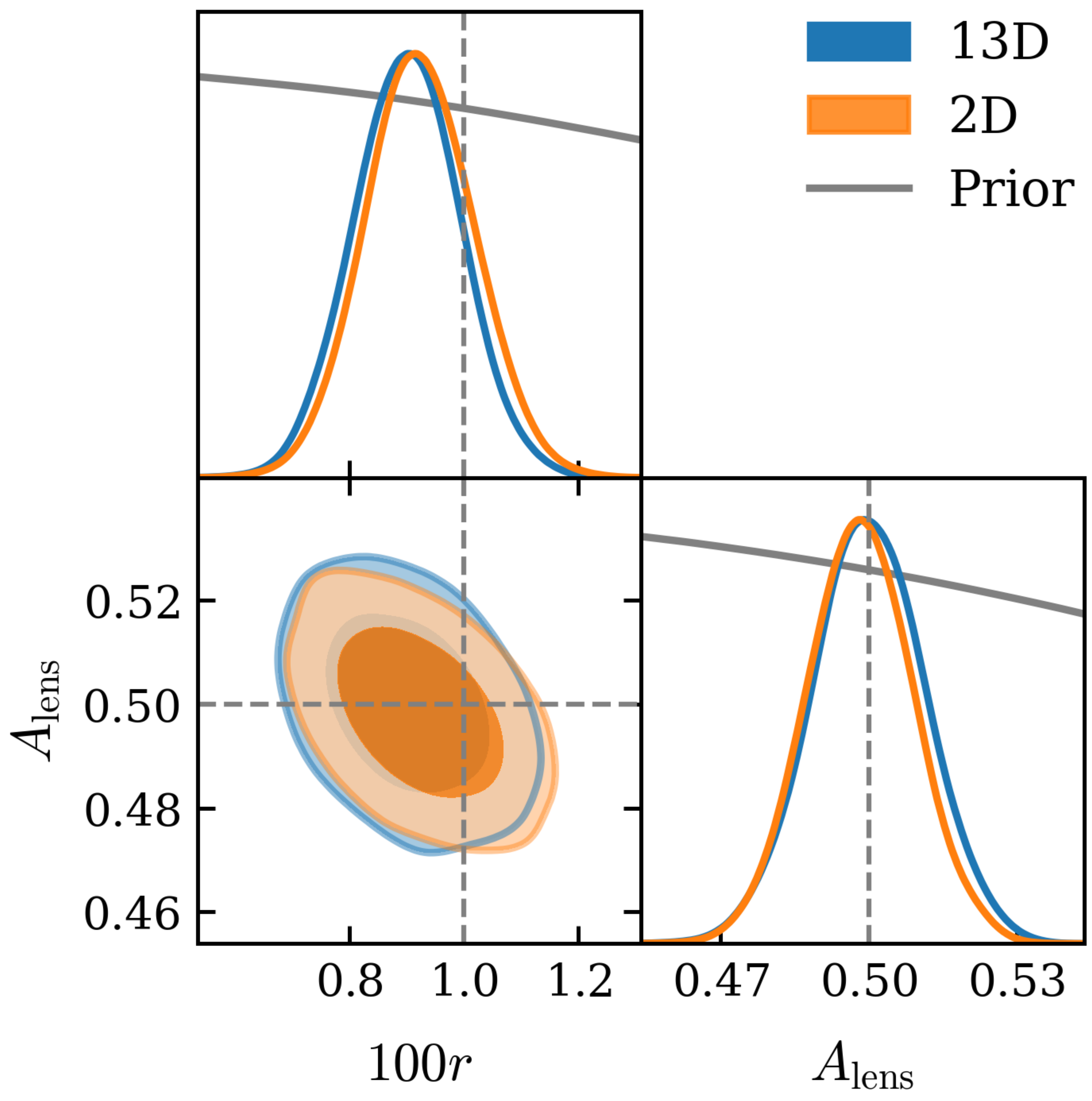
$$= \arg\min_{\vec{\lambda}} \int - P(\theta_1, \theta_2, \vec{x}) \log q_{\vec{\lambda}}\big(\theta_1 \,|\, \vec{x}\big) \mathrm{d}\theta_1 \mathrm{d}\theta_2 \, \mathrm{d}\vec{x}$$

$$\approx \arg\min_{\vec{\lambda}} \sum_{i=1}^{N_{\mathrm{sim}}} - \log q_{\vec{\lambda}}\big(\vec{\theta}_{1,i} \,|\, \vec{x}_i\big), \text{ sum over } \{\theta_{1,i}, x_i\}_{i=1}^{N_{\mathrm{sim}}}$$

NUISANCE PARAMETERS

Consistent results for r and $A_{\rm lens}$ regardless of whether the normalizing flow is trained to predict the posterior of all 13 parameters or just r and $A_{\rm lens}$

- With SBI, including nuisance parameters is straightforward
  - Define prior distribution of additional parameters and includes their effect in the simulations
  - No architectural changes to the normalizing flow network are needed, provided it has sufficient capacity



CONCLUSIONS

- With improvements in SNR, we should be careful with modeling errors in the standard CMB data reduction (e.g. mapmaking)
- The properties of the beam (leakage, polarized components, chromaticity) start to dominate the systematic error budget → Need tailored calibration strategies
- For certain CMB analyses, e.g. *B*-mode searches for *r*, simulation-based inference provides a promising way to deal with systematics