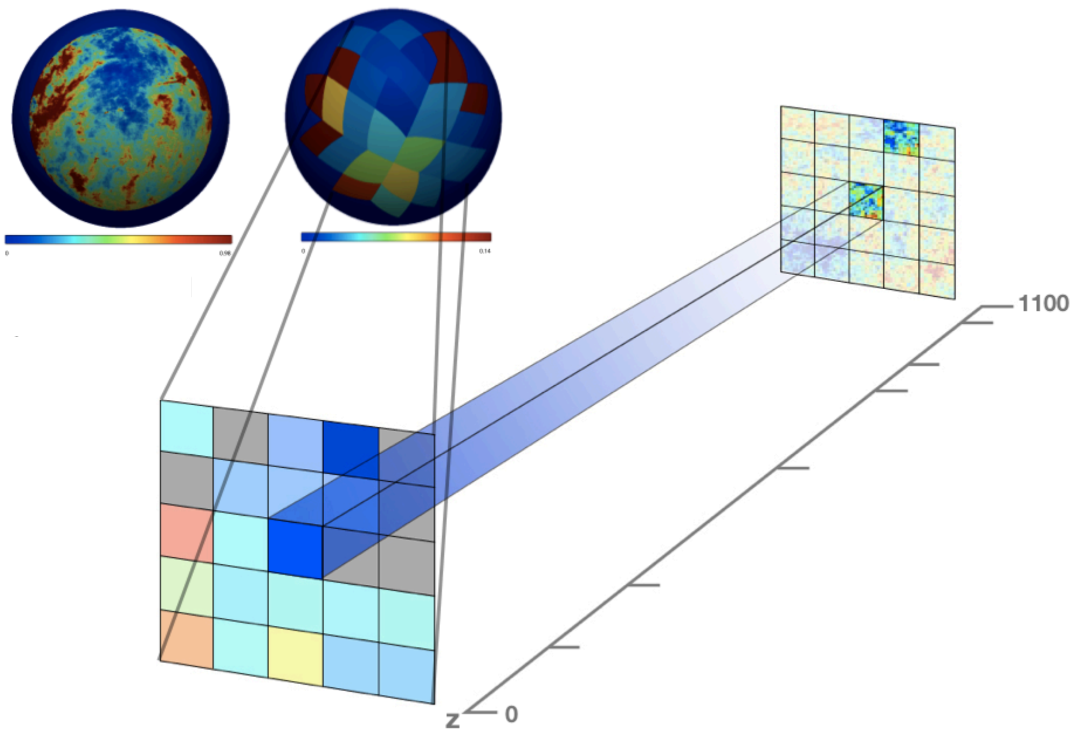
Cosmic Bandits: Exploration versus Exploitation in Cosmological Surveys

Ely D. Kovetz
University of Texas at Austin
LBNL, journal club, Oct. 18th, 2013



Based on: arXiv:1308.1404
and current work
with M. Kamionkowski (JHU)

Introduction: Motivation

Explosion of data from sky surveys.

- Exponential growth in detector size, internet bandwidth, data storage.

  (Moore's law) (Nielsen's law) (Kryder's law)

Examples:

- CMB surveys:

  A factor of up to $O(10^6)$ increase in data from COBE to post-Planck experiments.

- Galaxy surveys:

  LSST: $O(10)$ PB a year. Similar to LHC.

- 21-cm surveys:

  SKA: $O(100)$ PB a year, but 1,000 PB processed every day!

Introduction: Motivation

Experiments are getting (even) harder.

- Require control of systematics over many orders of magnitude.
- Faint signals are overwhelmed by (various) foregrounds.

Examples:

- CMB B-mode surveys:

  Sources of temperature-polarization leakage span many orders of magnitude.

  Inflationary B-modes may show up only at very small amplitudes (if $r \ll 1$).

- High-res imaging:

  With improved sensitivity, confusion noise can be limiting.

- 21-cm surveys:

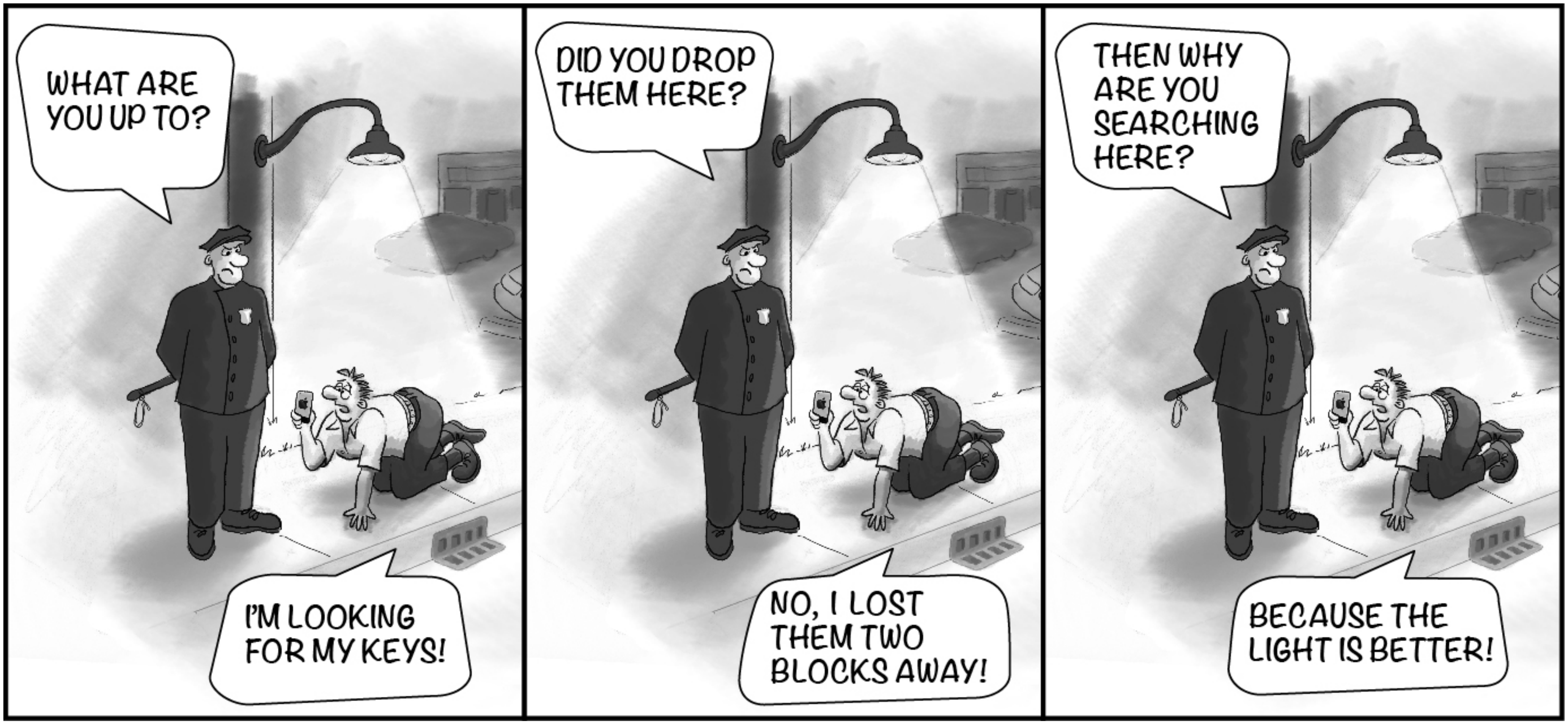
  EoR: foreground roughly 4 orders of magnitude higher than signal.

  Dark Ages: foregrounds up to 7 orders of magnitude higher.

Introduction: Exploration vs. Exploitation

- How to balance the tradeoff between *exploring* and *exploiting*?

  One way...



- Not necessarily stupid.
- Different targets call for different measurements.

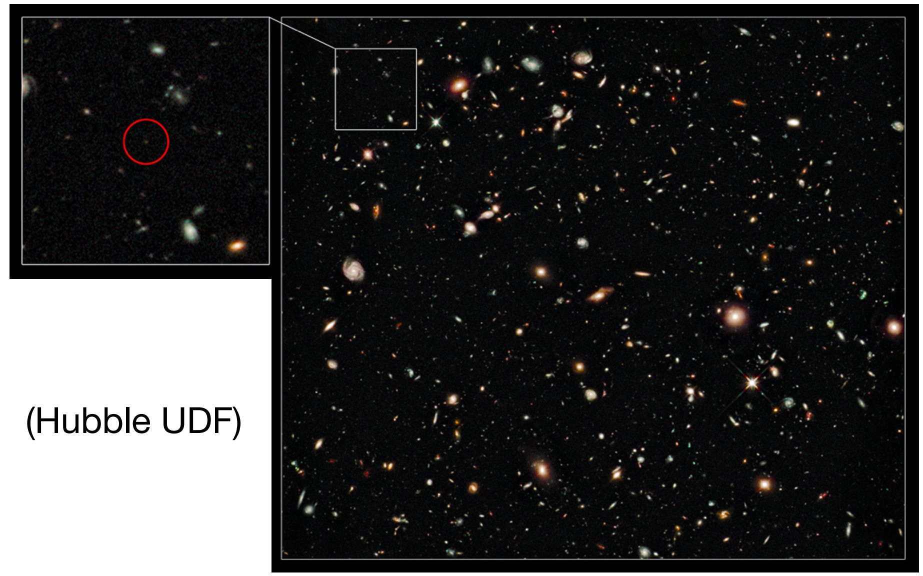
Introduction: Exploration vs. Exploitation

Stochastic measurements and deep-field imaging call for different approaches.

- Deep-field imaging:

Exploration mostly wasted.

Goal of adaptive strategy: --> quickly converge and *exploit*.

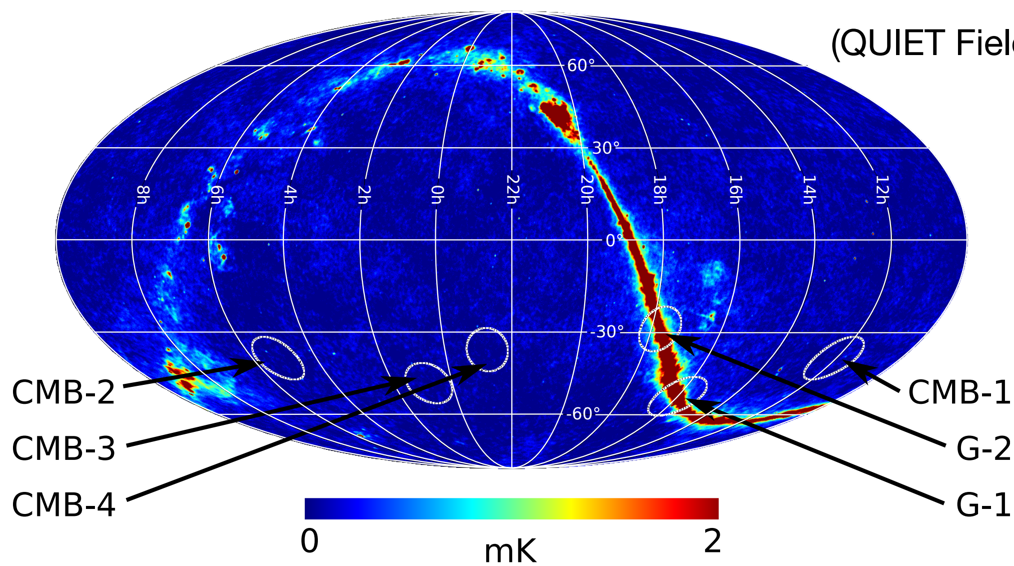


(Hubble UDF)

- Stochastic fluctuations:

Exploration mitigates cosmic variance.

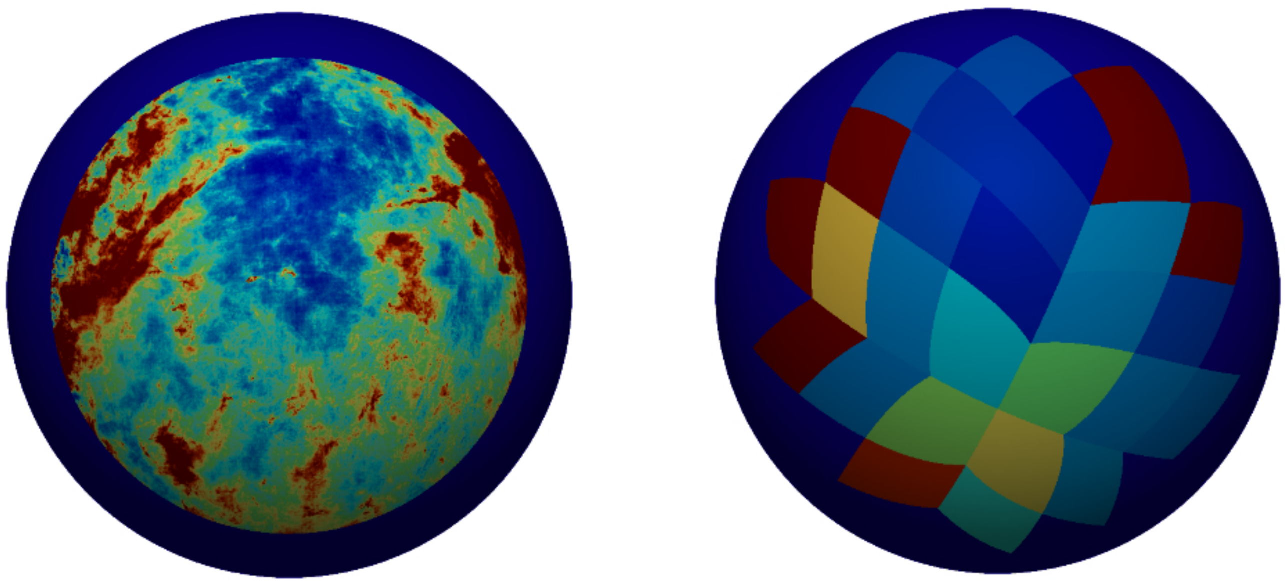
Goal of adaptive strategy: --> find ideal patches to *exploit*.



(QUIET Fields)

Exploration vs. Exploitation: B-mode Surveys

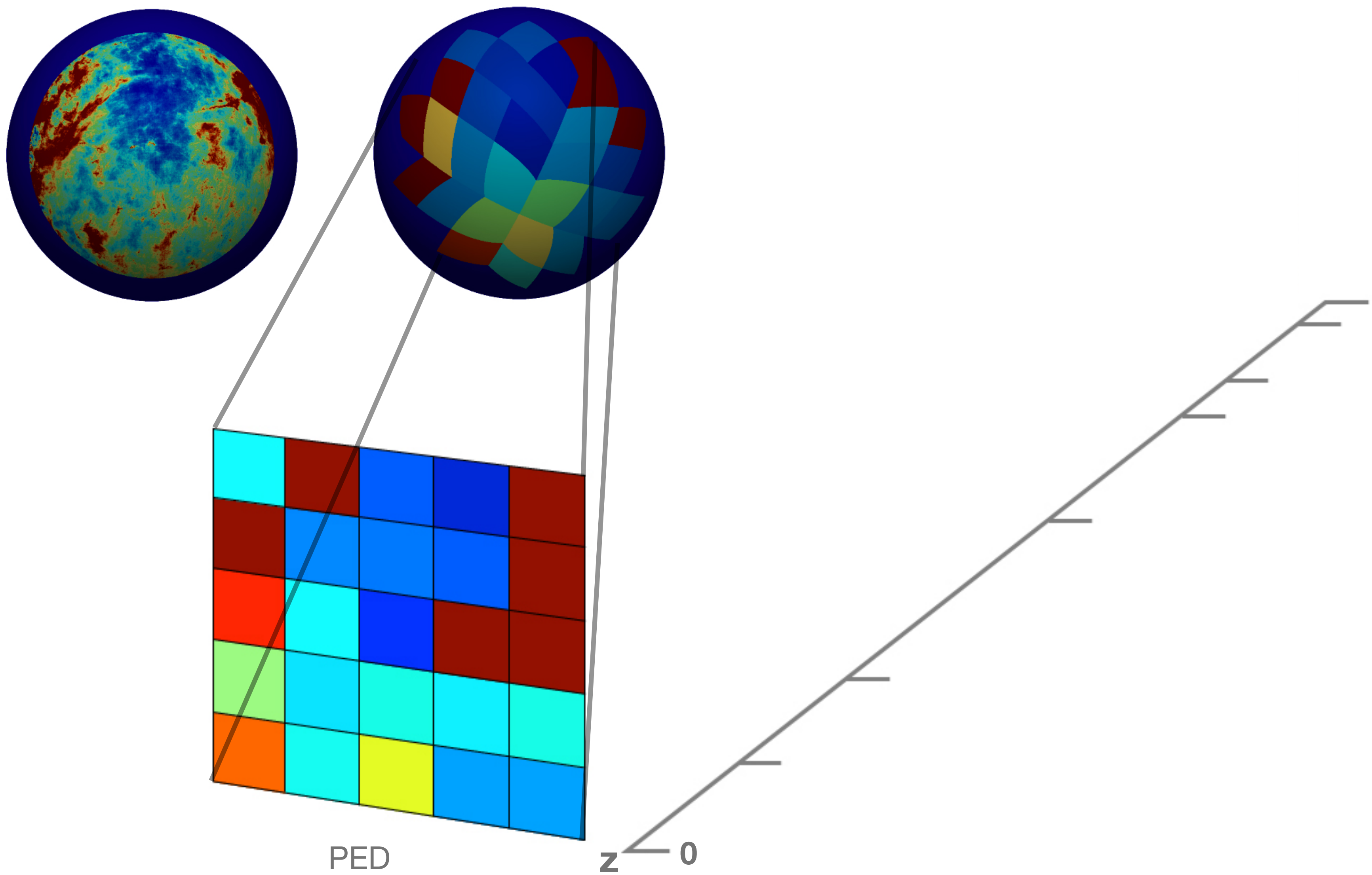
- Consider the case of B-mode detection, which is optimized in small sky patches.
- Tradeoff is between *finding lower-foreground patches* and *integrating over them*.



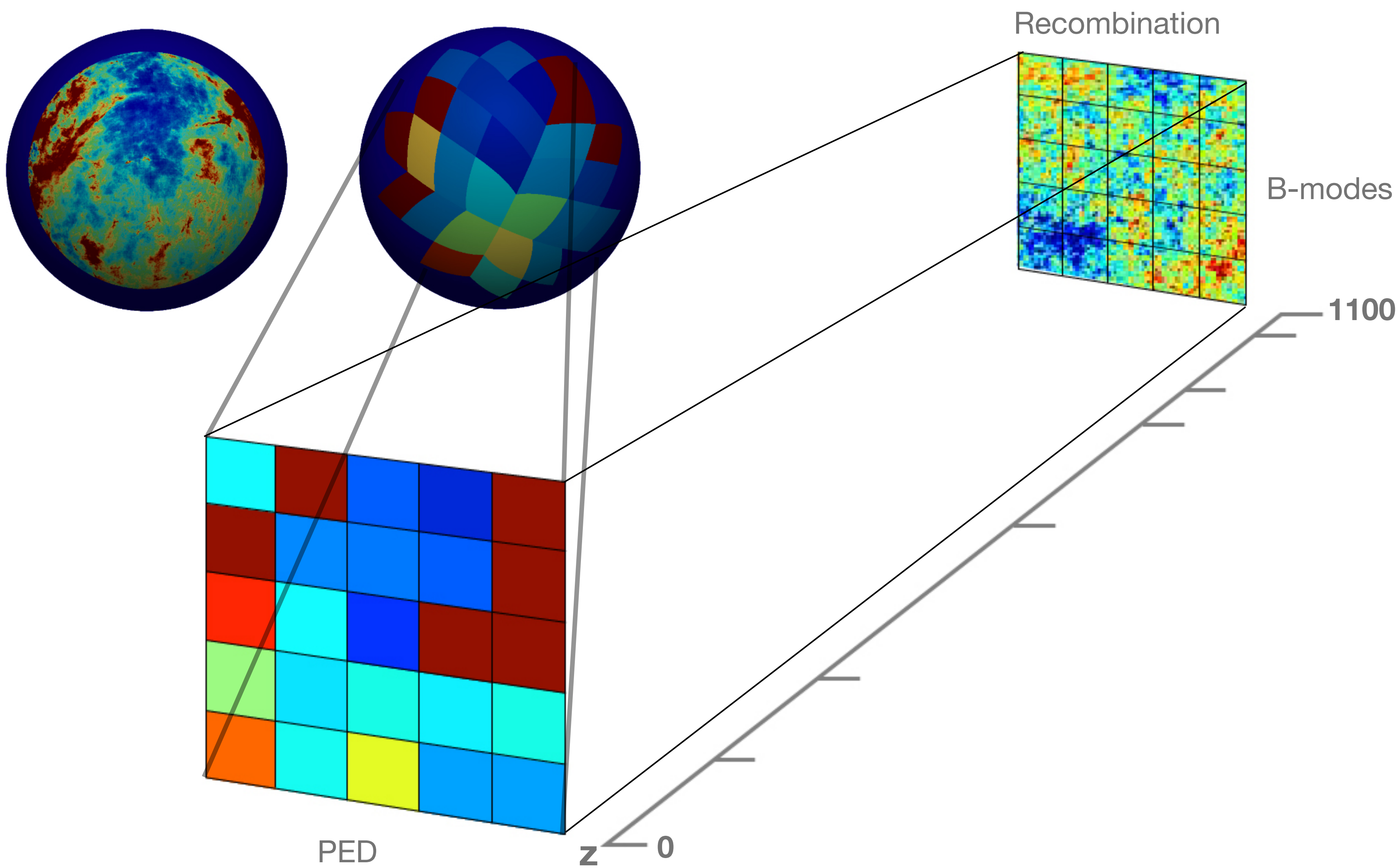
Templates for polarized emission from dust (PED) in the Galaxy at 150GHz

(Clark et al. arXiv:1211.6404)

Exploration vs. Exploitation: B-mode Surveys

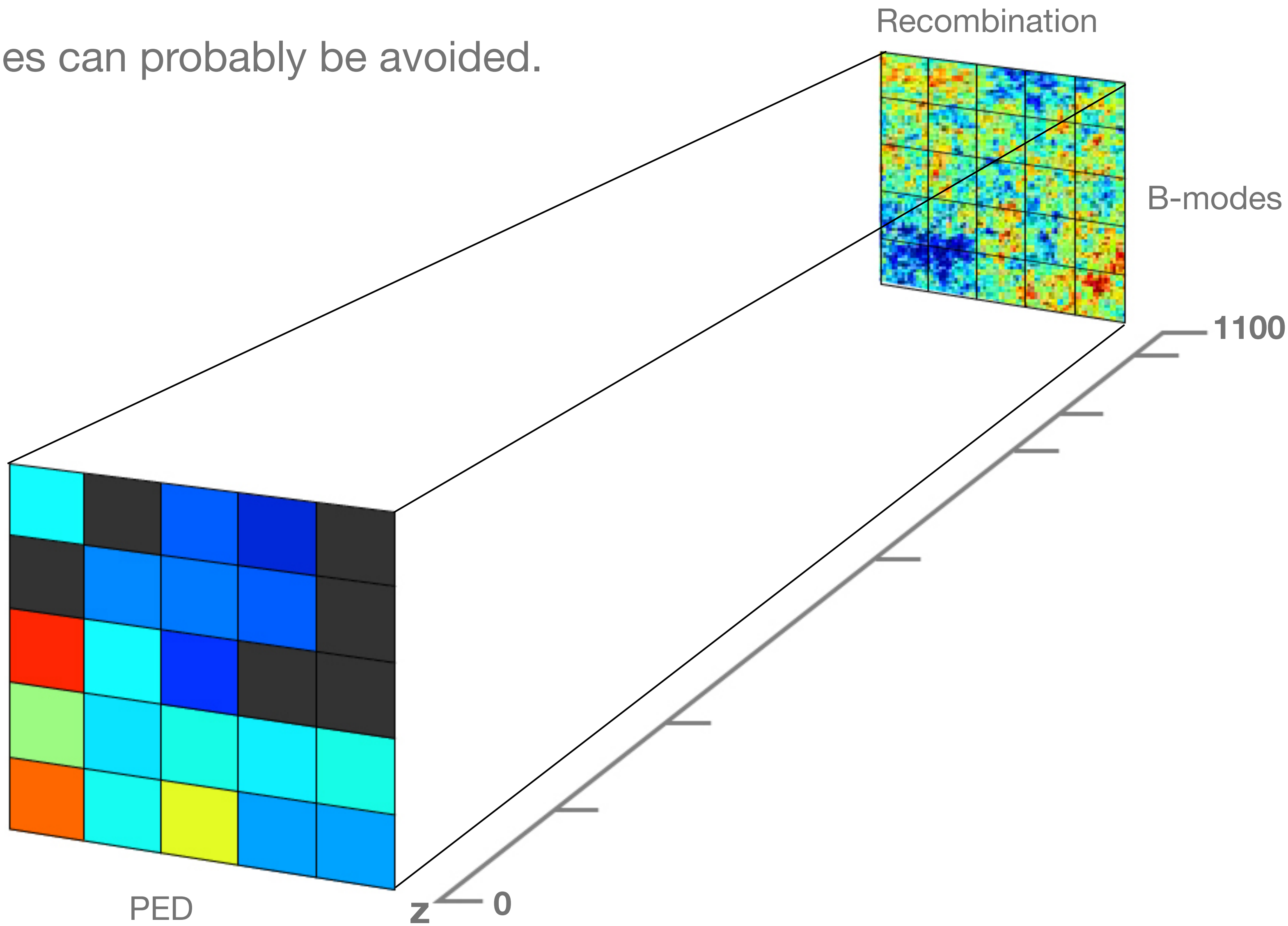


Exploration vs. Exploitation: B-mode Surveys



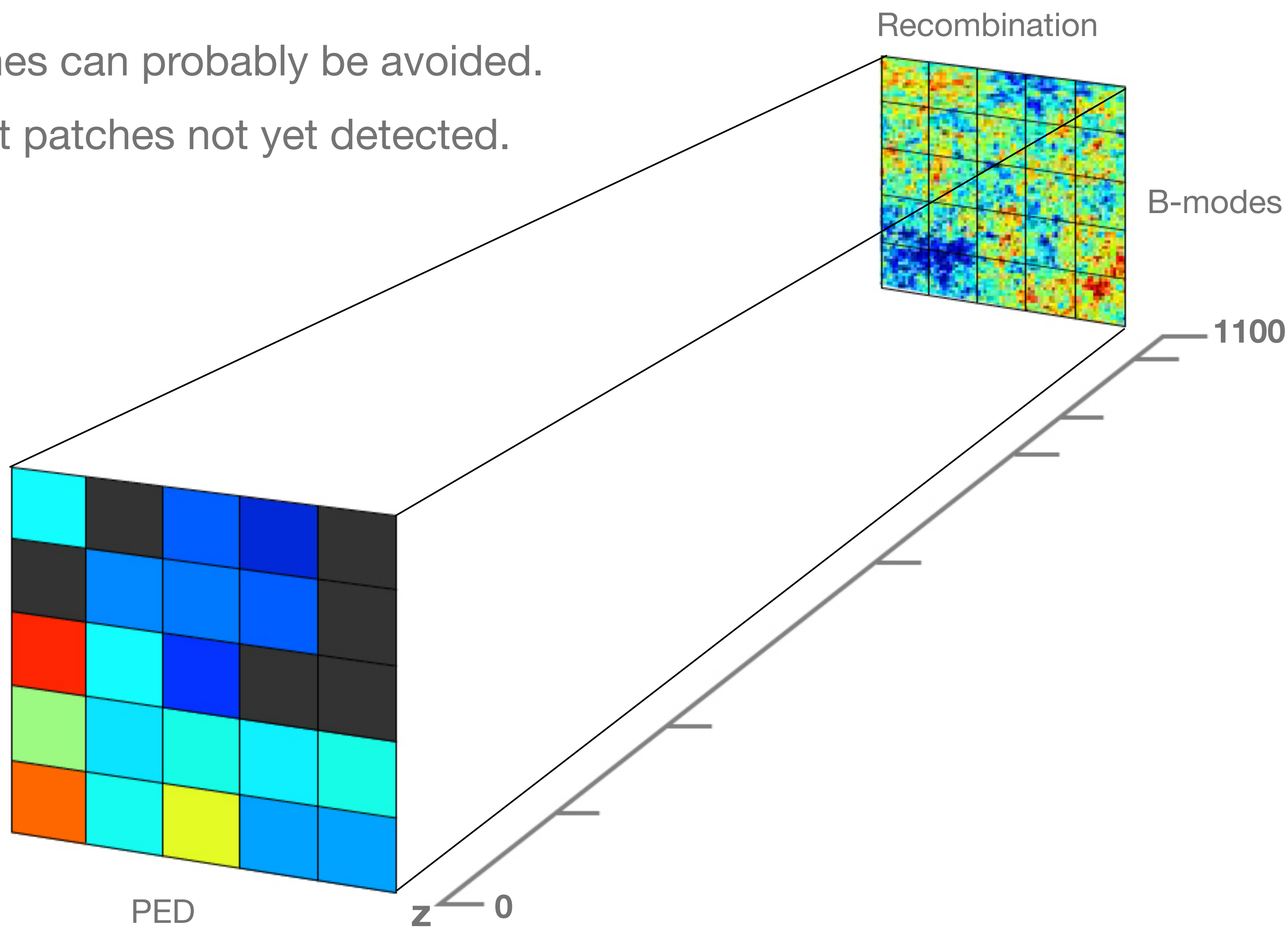
Exploration vs. Exploitation: B-mode Surveys

- Worst patches can probably be avoided.



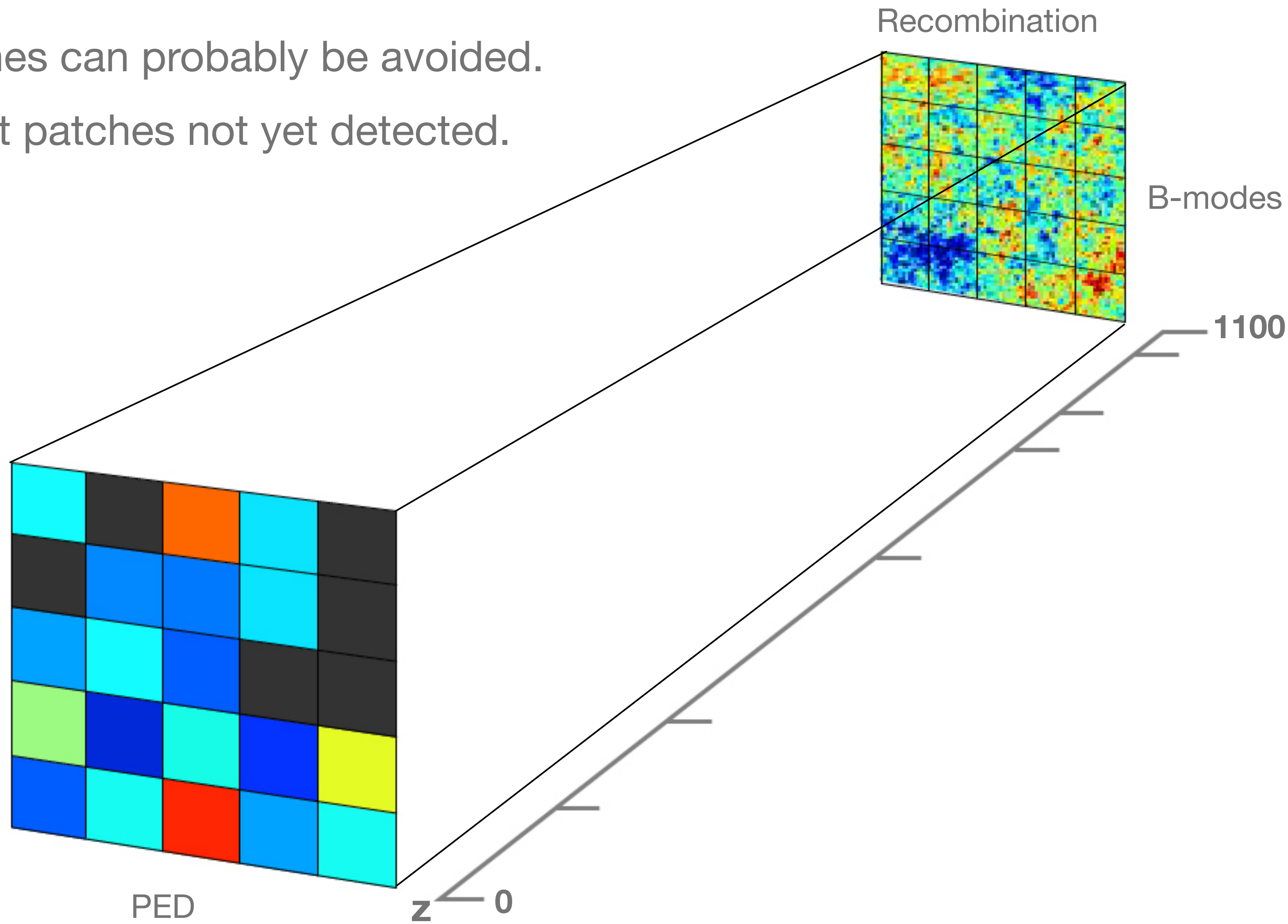
Exploration vs. Exploitation: B-mode Surveys

- Worst patches can probably be avoided.
- But cleanest patches not yet detected.



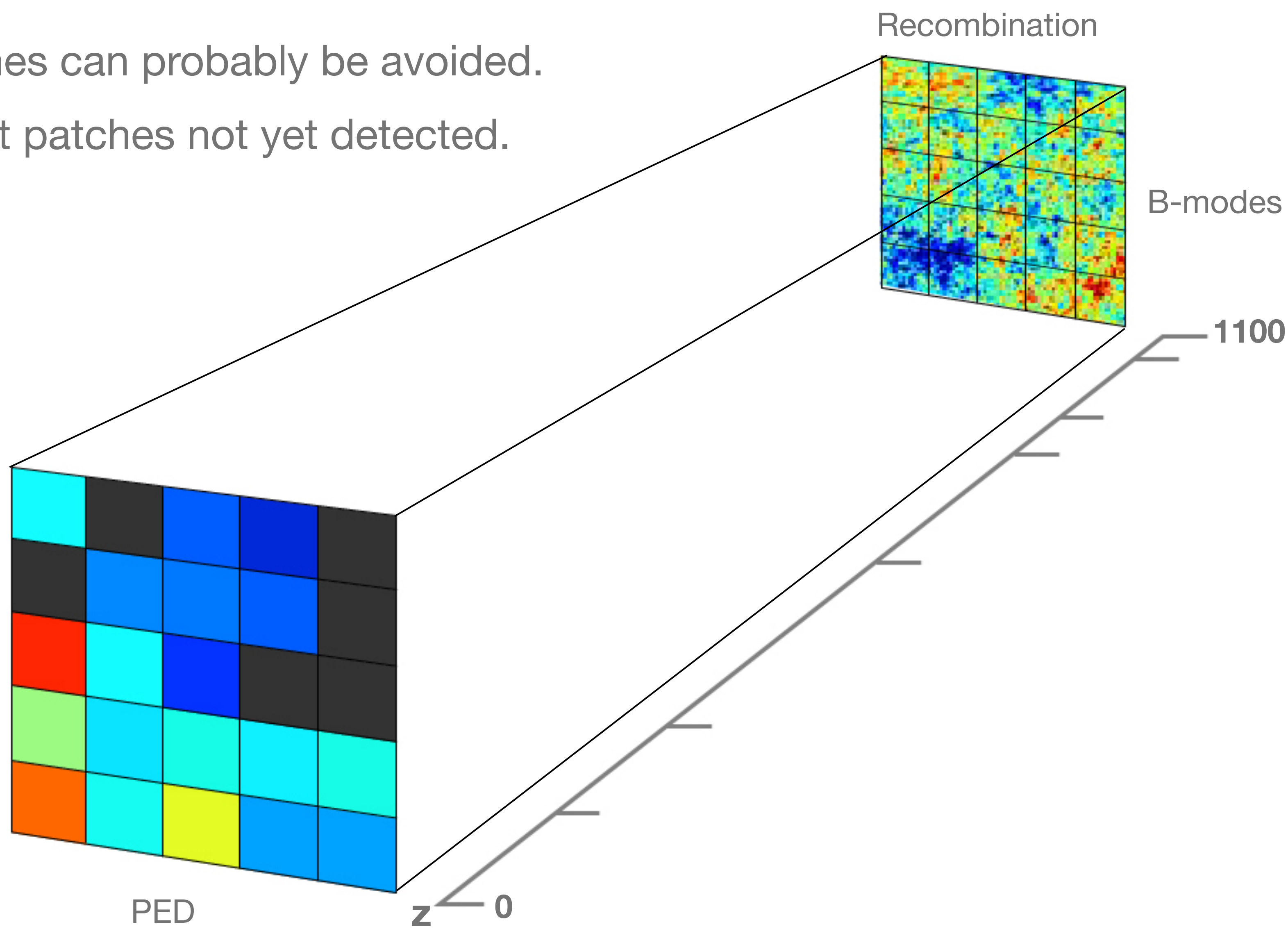
Exploration vs. Exploitation: B-mode Surveys

- Worst patches can probably be avoided.
- But cleanest patches not yet detected.



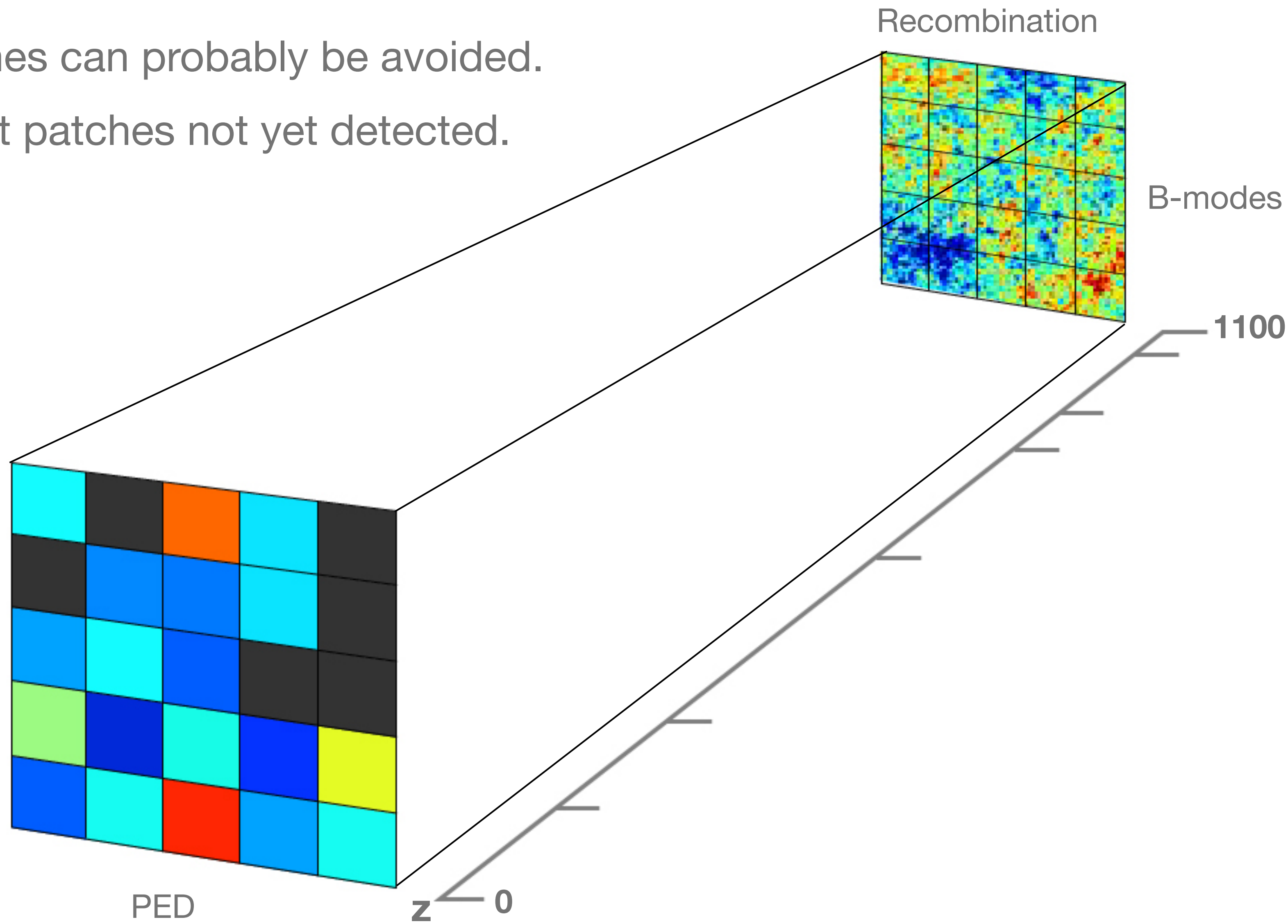
Exploration vs. Exploitation: B-mode Surveys

- Worst patches can probably be avoided.
- But cleanest patches not yet detected.



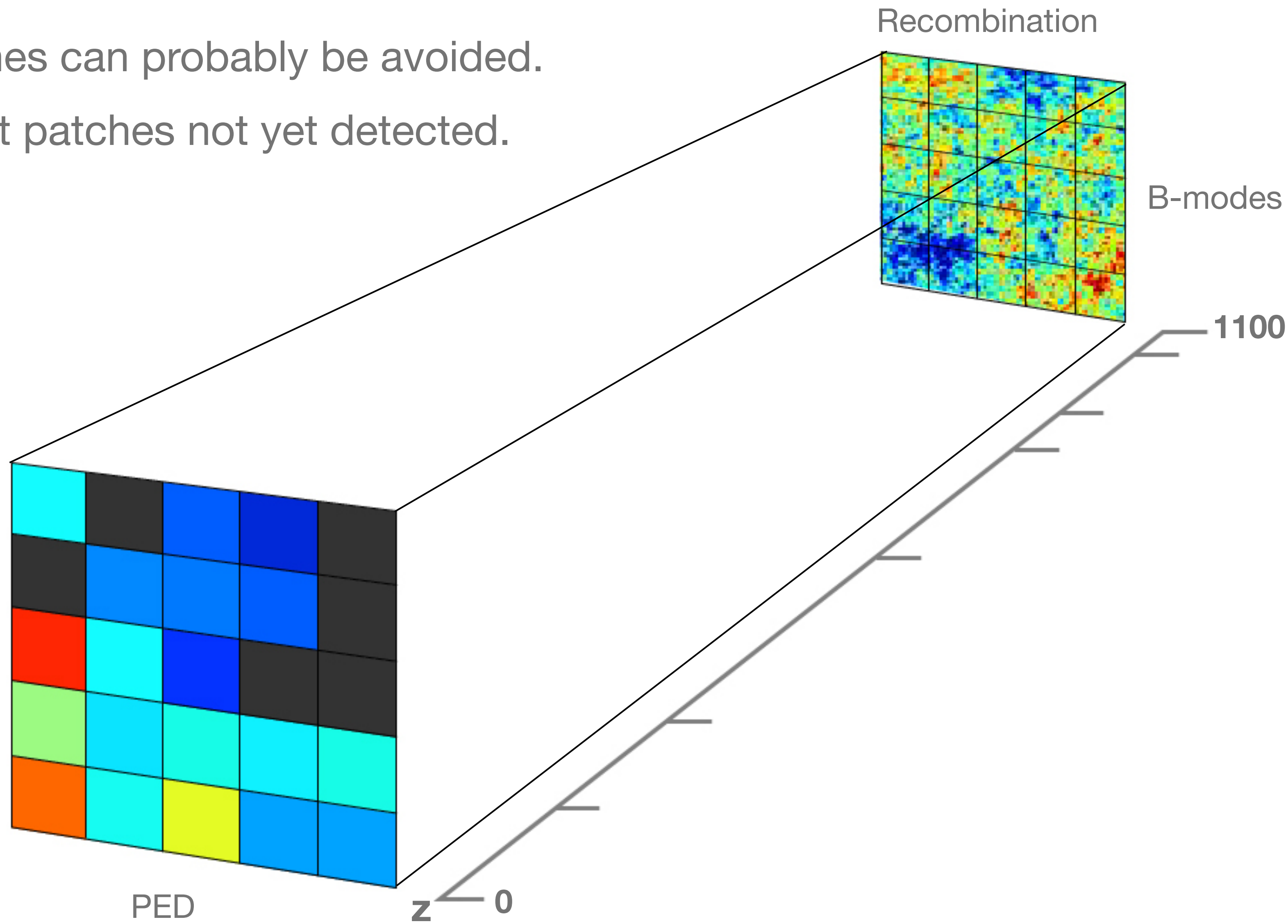
Exploration vs. Exploitation: B-mode Surveys

- Worst patches can probably be avoided.
- But cleanest patches not yet detected.



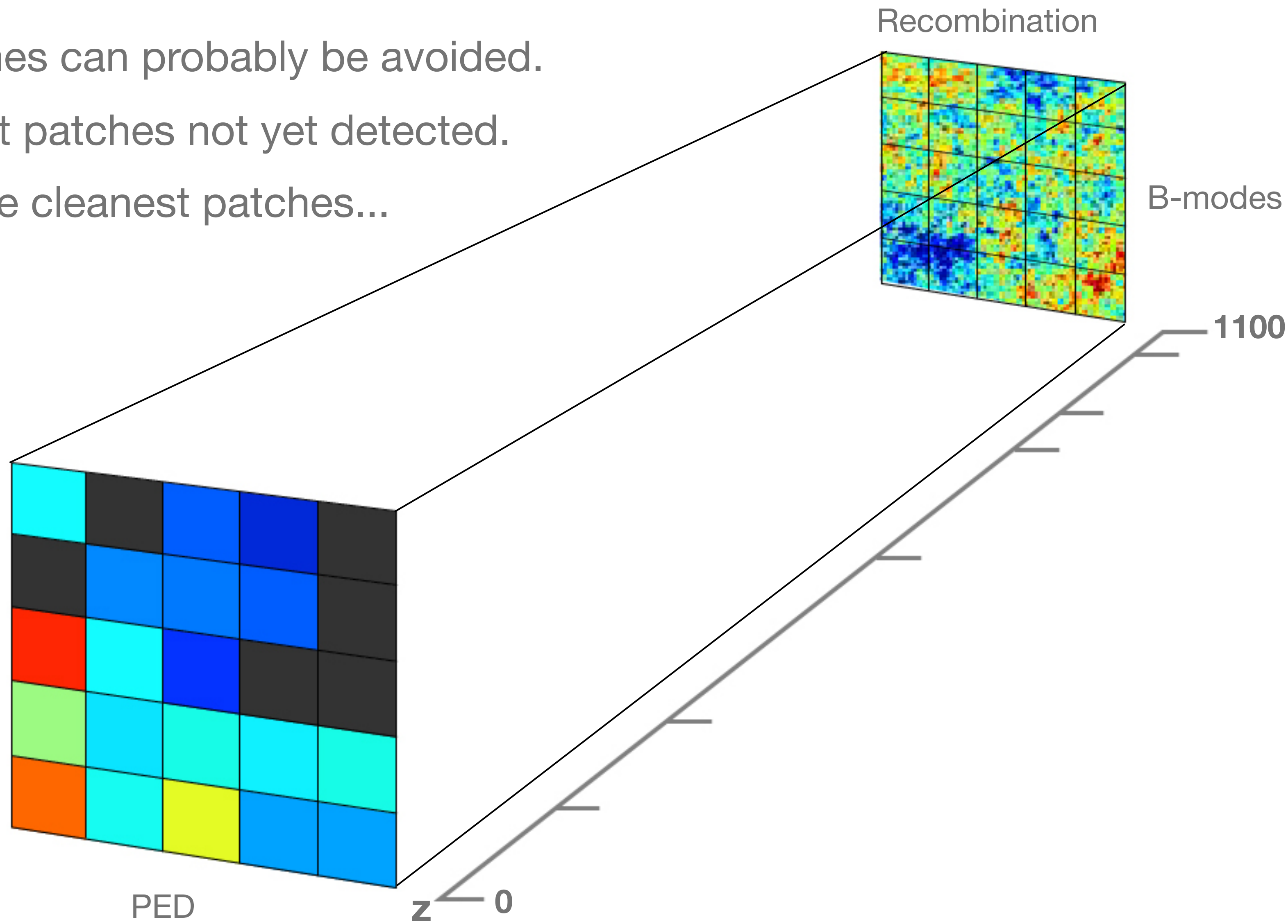
Exploration vs. Exploitation: B-mode Surveys

- Worst patches can probably be avoided.
- But cleanest patches not yet detected.



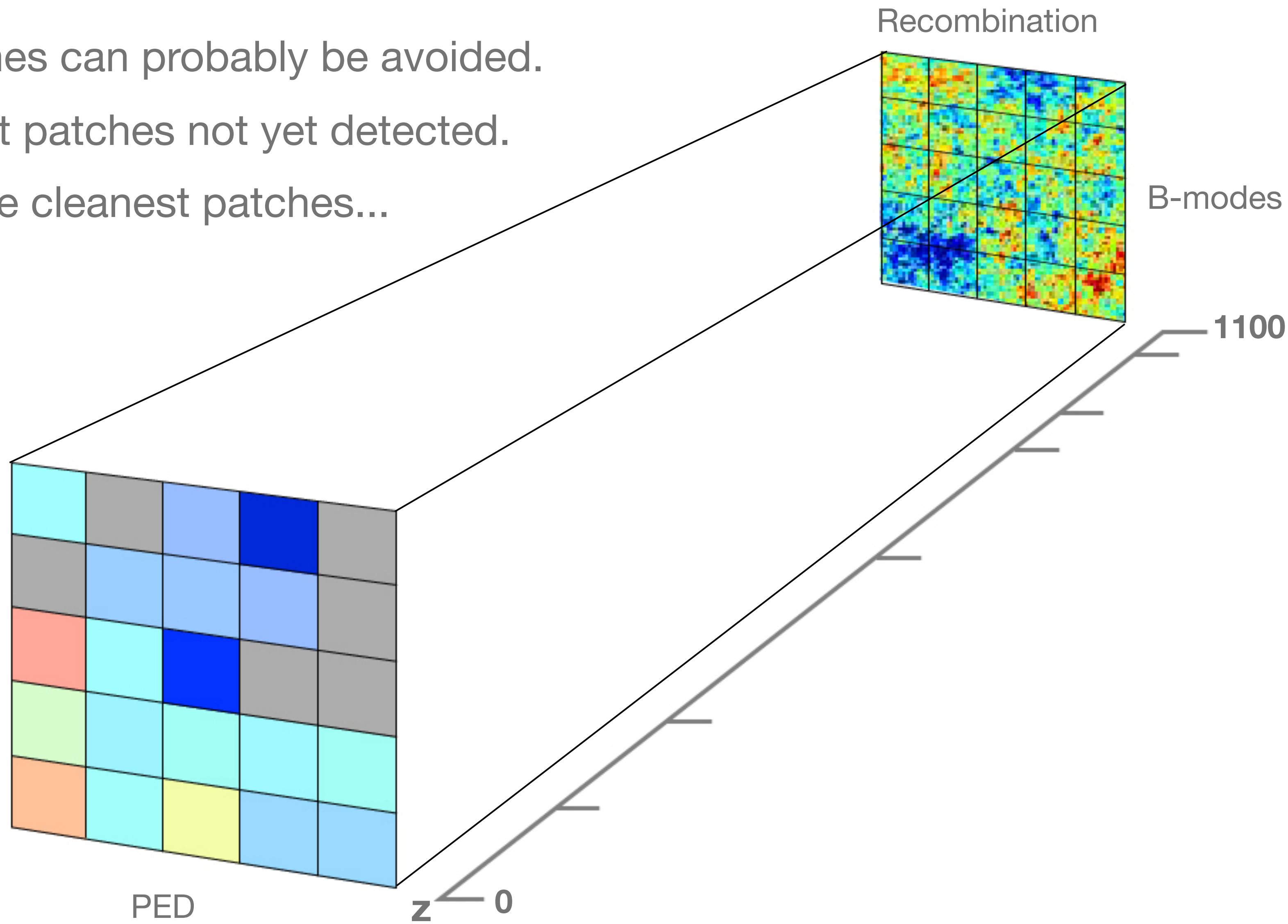
Exploration vs. Exploitation: B-mode Surveys

- Worst patches can probably be avoided.
- But cleanest patches not yet detected.
- If we find the cleanest patches...



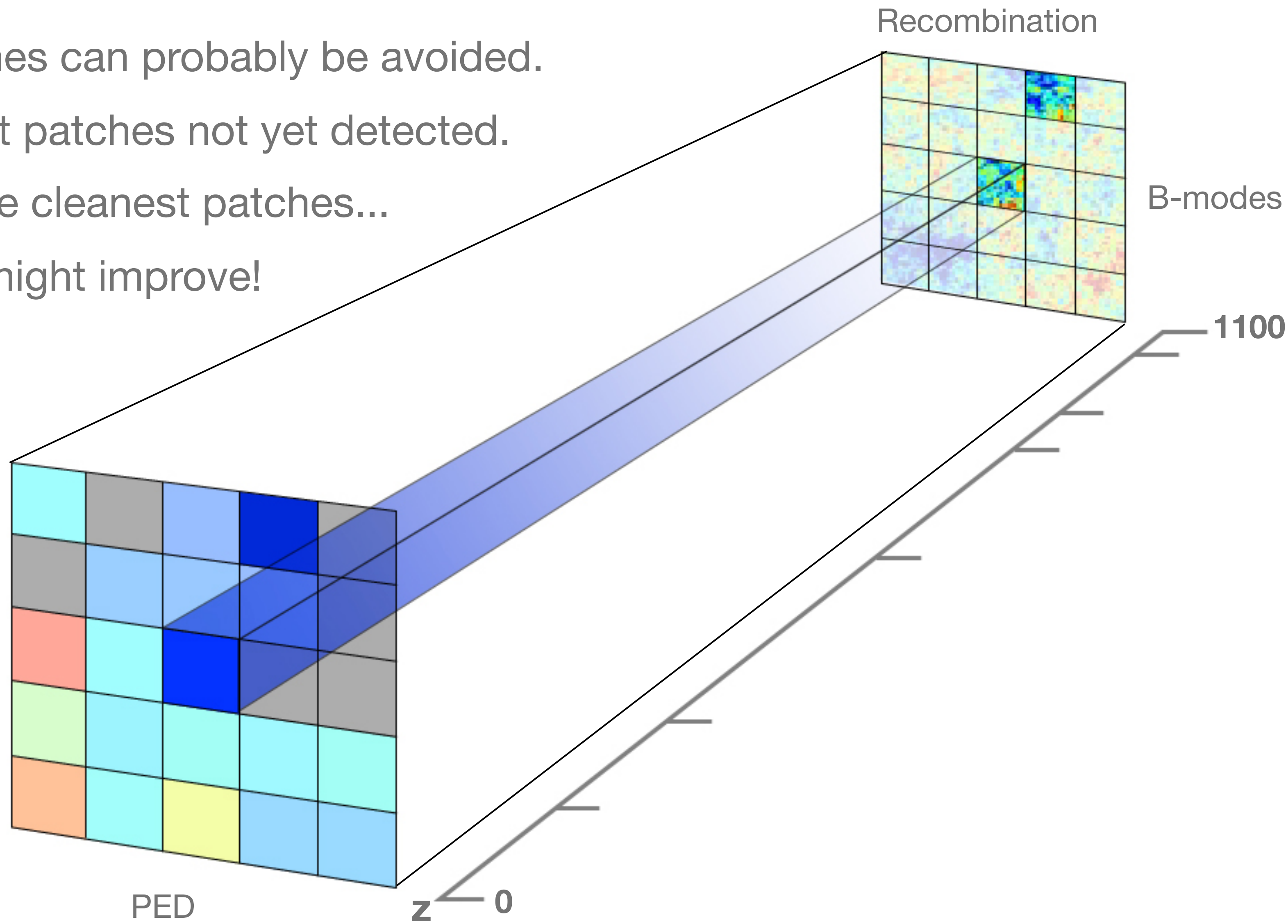
Exploration vs. Exploitation: B-mode Surveys

- Worst patches can probably be avoided.
- But cleanest patches not yet detected.
- If we find the cleanest patches...



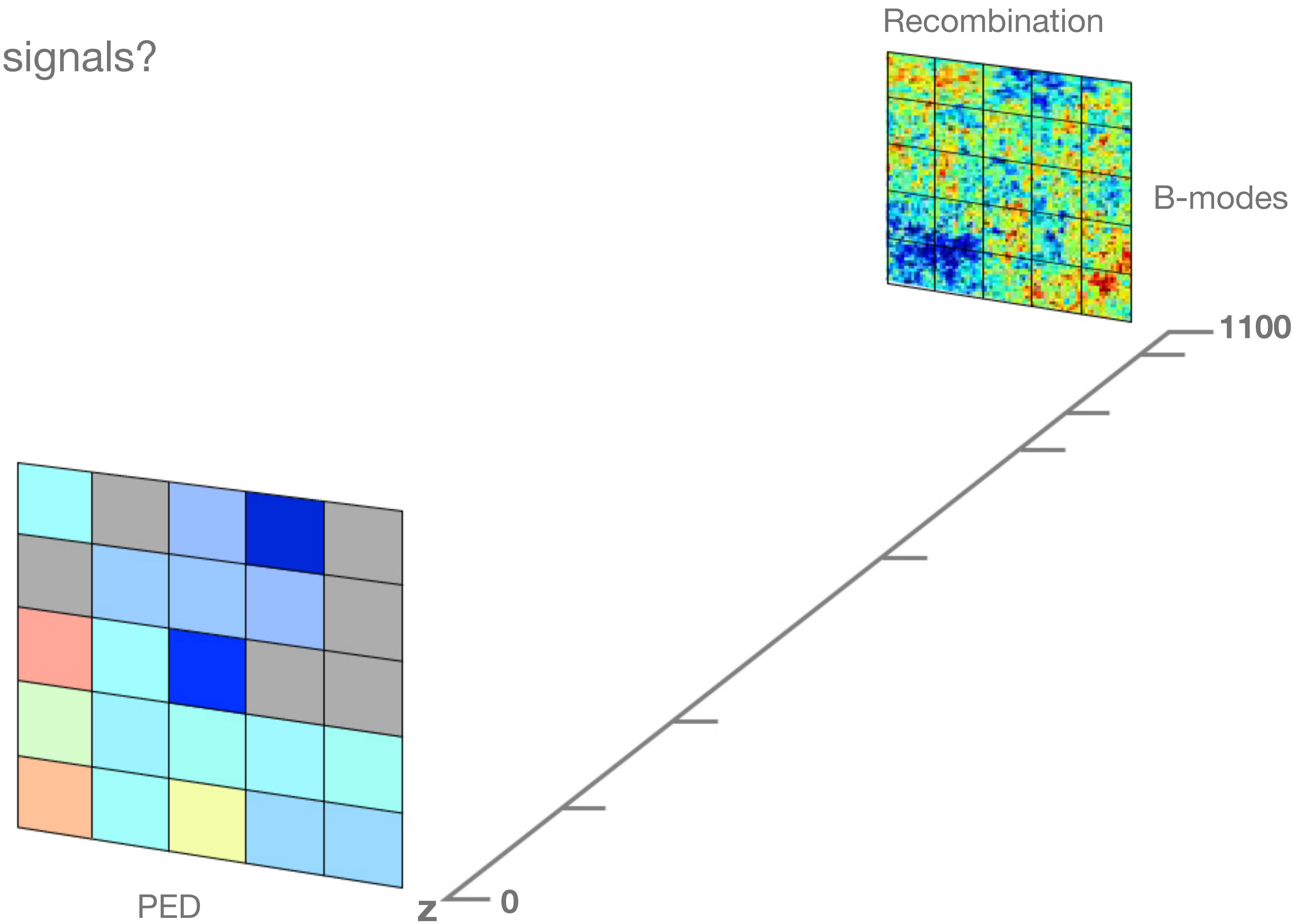
Exploration vs. Exploitation: B-mode Surveys

- Worst patches can probably be avoided.
- But cleanest patches not yet detected.
- If we find the cleanest patches...
- Sensitivity might improve!



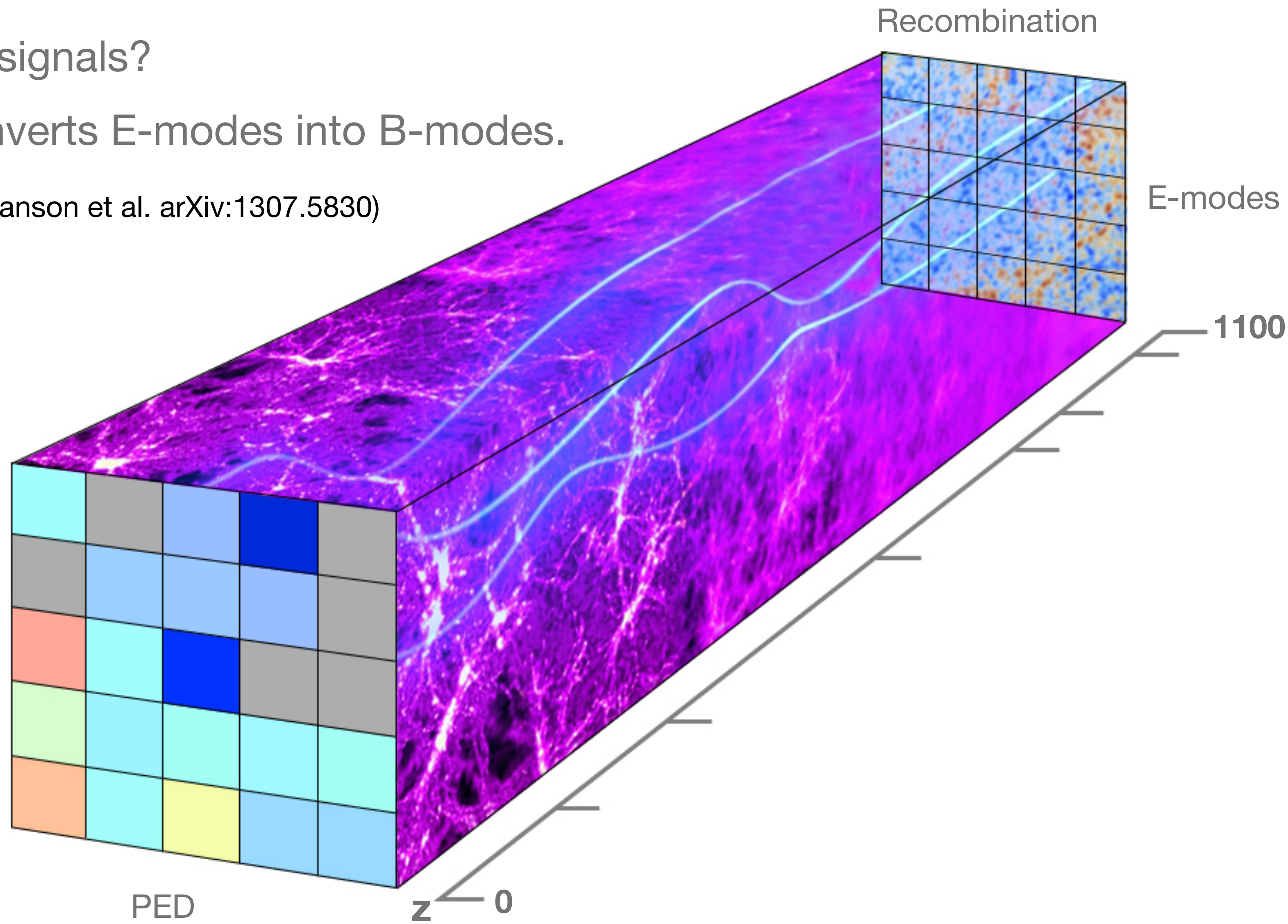
Exploration vs. Exploitation: B-mode Surveys

- Competing signals?



Exploration vs. Exploitation: B-mode Surveys

- Competing signals?
- Lensing converts E-modes into B-modes.
- Detected! (Hanson et al. arXiv:1307.5830)
- De-lensing?



Outline

- The Multi-Armed-Bandit Problem

Outline

- The Multi-Armed-Bandit Problem

- Heuristic Solution Algorithms

Outline

- The Multi-Armed-Bandit Problem
- Heuristic Solution Algorithms
- B-modes: The good, the bad and the ugly

Outline

- The Multi-Armed-Bandit Problem
- Heuristic Solution Algorithms
- B-modes: The good, the bad and the ugly
- Results and Discussion

Outline

- The Multi-Armed-Bandit Problem
- Heuristic Solution Algorithms
- B-modes: The good, the bad and the ugly
- Results and Discussion
- MAB Strategies Elsewhere

The Multi-Armed-Bandit Problem

- Goal:

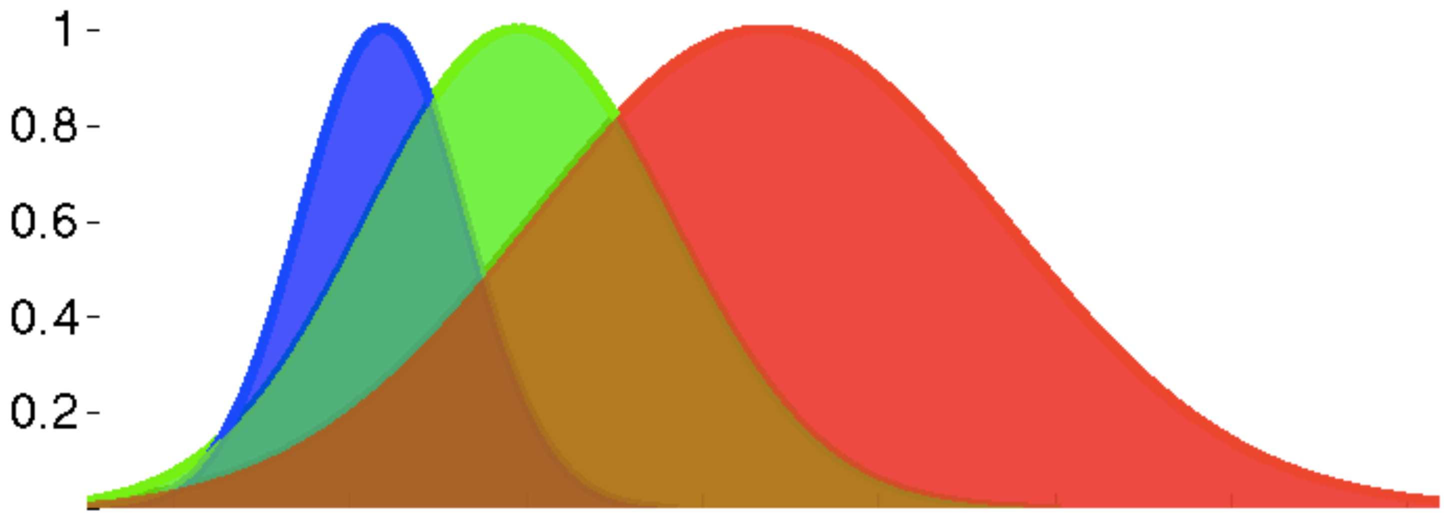
  Facing slots with different odds, maximize winnings.
- With infinite funds, this is easy. You *learn* the odds.
- With a finite number of plays, problem is unsolved.
- Heuristics have been developed and compared.

Multi-Armed-Bandit Strategies

- An MAB strategy:

  Use estimates for the expected rewards of each arm in order to choose action.
- The expected (or *true*) reward $\mu^*(a)$ is called its action-value.
- An action-value estimate is given by the sample-average of previous rewards:

$$\mu_t(a) = \frac{r_1 + r_2 + \cdots + r_{N_t(a)}}{N_t(a)} \qquad \mu_t(a) \underset{N_t(a)\to\infty}{\longrightarrow} \mu^*(a)$$



Multi-Armed-Bandit Strategies

- An MAB strategy:

  Use estimates for the expected rewards of each arm in order to choose action.
- The expected (or *true*) reward $\mu^*(a)$ is called its action-value.
- An action-value estimate is given by the sample-average of previous rewards:

$$\mu_t(a) = \frac{r_1 + r_2 + \cdots + r_{N_t(a)}}{N_t(a)} \qquad \mu_t(a) \underset{N_t(a) \to \infty}{\longrightarrow} \mu^*(a)$$

- To test a strategy:

  The optimal reward is $V^* = \max_a \mu^*(a)$ and we look at the average total regret:

$$R_t = \left\langle \sum_{t=1}^{T} [V^* - \mu_t(a)] \right\rangle = \sum_a \langle N_t(a) \rangle \, \Delta_a \qquad (\Delta_a = V^* - \mu^*(a))$$

gap vs. optimal arm

  where $\langle \ldots \rangle$ is an ensemble average in simulations with known action-values.
- A good strategy ensures smaller counts for larger gaps.

Multi-Armed-Bandit Strategies

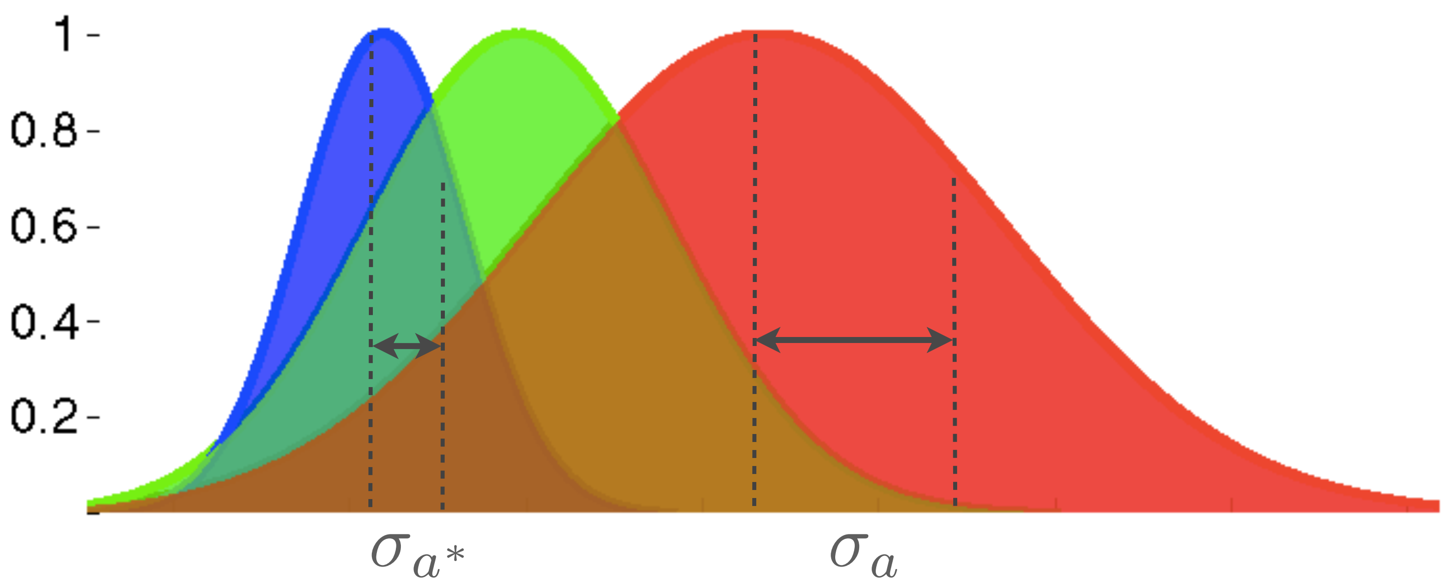
- Lai and Robbins (1985): in the asymptotic limit of infinite number of plays

$$\lim_{t\to\infty} R_t \geq \log t \sum_{a|\Delta_a>0} \frac{\Delta_a}{KL(\mathcal{P}(a)||\mathcal{P}(a^*))}$$

where $KL(\mathcal{P}(a)||\mathcal{P}(a^*))$ is the Kullback-Liebler distance.

- For normal distributions: $\mathcal{N}(\mu^*(a^*), \sigma_{a^*})$ $\mathcal{N}(\mu^*(a), \sigma_a)$

$$KL(\mathcal{P}(a)||\mathcal{P}(a^*)) = \ln\frac{\sigma_{a^*}}{\sigma_a} + \frac{\sigma_a^2 + (\mu^*(a^*) - \mu^*(a))^2}{2\sigma_{a^*}^2} - \frac{1}{2}$$



Outline

- The Multi-Armed-Bandit Problem

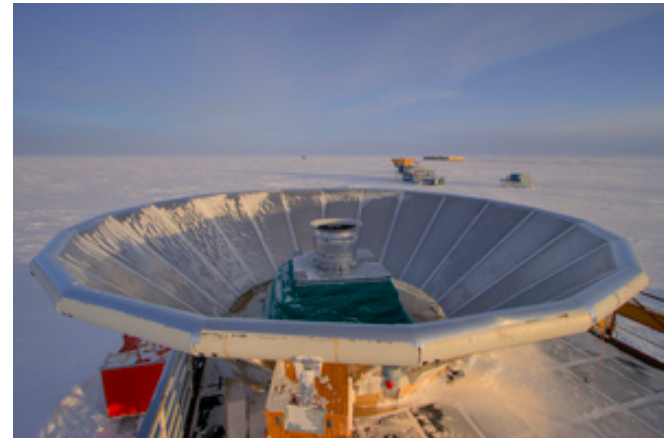
- Heuristic Solution Algorithms

Heuristic Solution Algorithms: Naive

- Uniformly random

For n_a arms: $p_t(a) = 1/n_a$

Never explores --> has linear total regret.



BICEP (larger $f_{\rm sky}$)

- Greedy

$$p_t(a) = 1 \quad \text{if a} = \underset{a'}{\operatorname{argmax}}\ \mu_t(a')$$

Never explores --> has linear total regret.

POLARBEAR (small $f_{\rm sky}$)

Heuristic Solution Algorithms: Forced Exploration

- Trick: use initialization to force exploration!

$$\mu_t(a) = \frac{r_1 + \cdots + r_{N_t(a)}}{N_t(a)} = \mu_{t-1}(a) + \frac{1}{N_t(a)}\left(r_{N_t(a)} - \mu_{t-1}(a)\right)$$

No prior knowledge: $\mu_0(a) = 0 \quad (\text{or} \ll \min\left\{\mu^*(a_i) \ldots \mu^*(a_{n_p})\right\})$

Optimistic initialization: $\mu_0(a) \gg V^*$ (after a while sample-average dominates)

- ϵ - greedy

$$\begin{cases} p_t = 1 - \epsilon & \text{Greedy Arm} \\ p_t = \quad \epsilon & \text{Uniformly Random} \end{cases}$$

Always explores --> has linear total regret.

Limits: $\epsilon \to 0$ (greedy) $\epsilon = 1$ (uniformly random)

- ϵ_t- greedy with a decaying strategy for ϵ --> can achieve logarithmic regret!

Heuristic Solution Algorithms

- Probability matching (Boltzmann)

$$p_t(a) = \frac{e^{\mu_t(a)/\tau}}{\sum_{a' \neq a} e^{\mu_t(a')/\tau}}$$

(W.R. Thompson, 1933)

Limits: $\tau \to 0$ (greedy) $\tau \to \infty$ (uniformly random)

- Upper confidence bound (UCB)

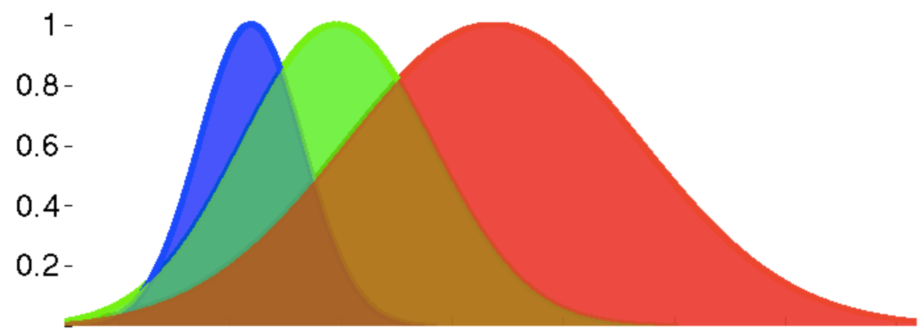
(Auer, Cesa-Bianchi, and Fischer, 2002)

Define a (time-dependent) bound: $\mu^*(a) \leq \mu_t(a) + U_t(a)$

and choose: $a_t = \underset{a}{\operatorname{argmax}} \{\mu_t(a) + U_t(a)\}$

When distributions are normal:

$$U_t(a) = c\,\sigma_a / \sqrt{N_t(a)} \qquad (c > 0)$$



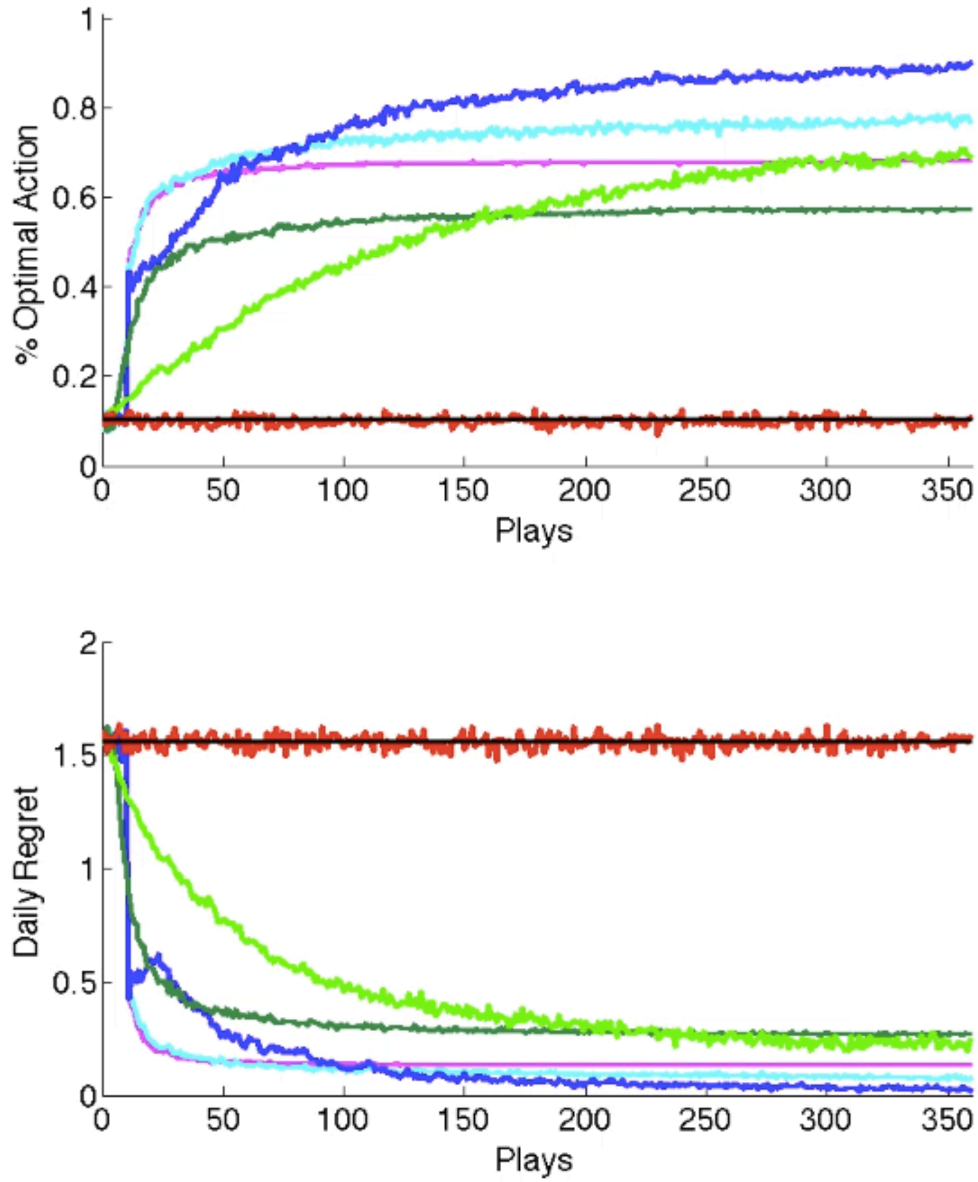
UCB has logarithmic total regret!

Heuristic Solution Algorithms

- MAB Simulation: $\langle \ldots \rangle_{1,000}$

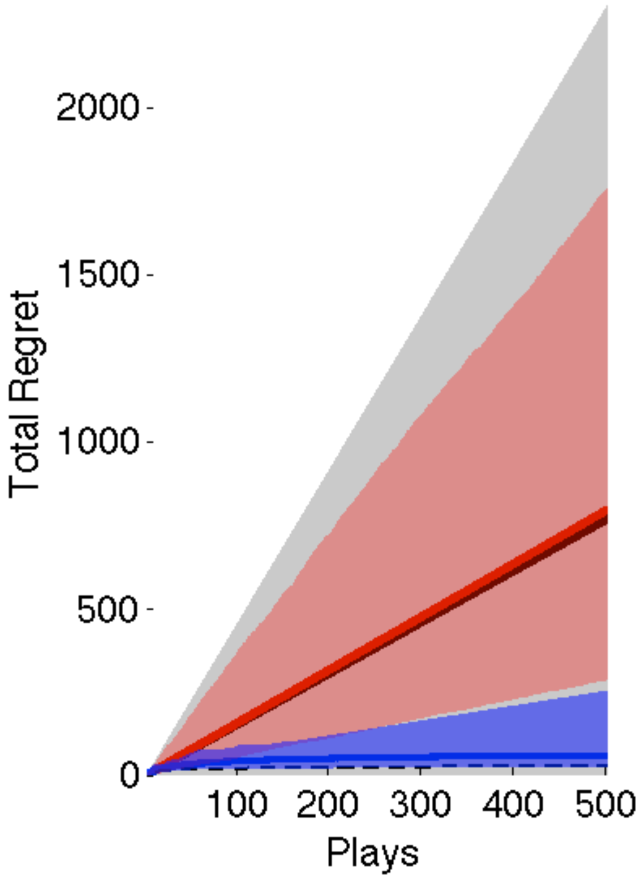
  $n_p = 10;\ \forall a:\ \mathcal{N}(\mu^*(a), \sigma_a)$

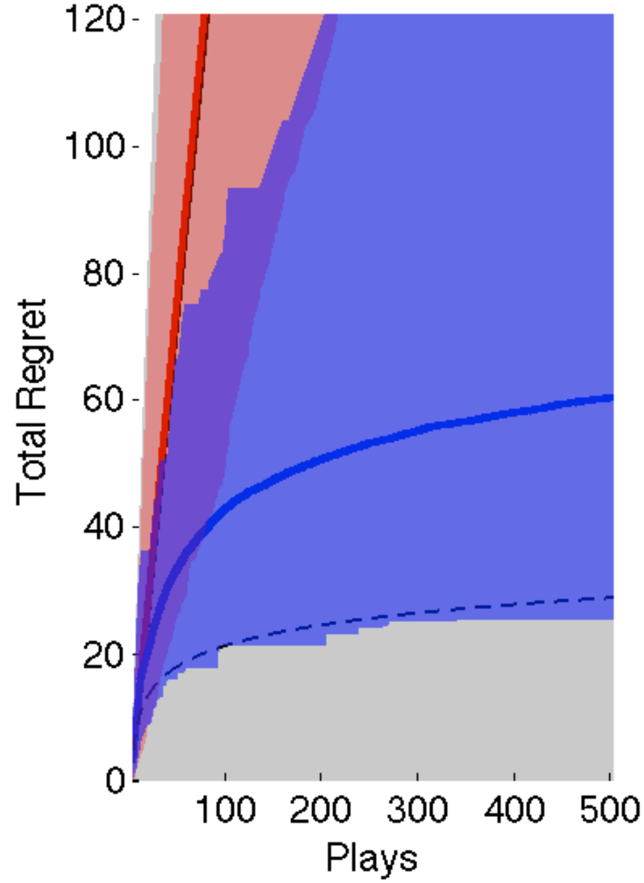
— greedy
— greedy $\epsilon = 1$
— greedy $\epsilon = 0.1$
— greedy decay-ϵ
— UCB
— Boltzmann τ
— opt + greedy



Heuristic Solution Algorithms

- What about the total regret?





- UCB approaches the Lai and Robbins bound!

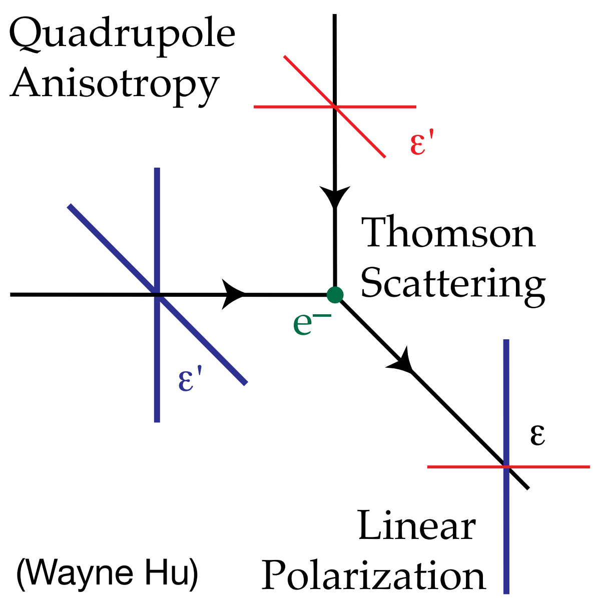
Outline

- The Multi-Armed-Bandit Problem
- Heuristic Solution Algorithms
- B-modes: The good, the bad and the ugly

B-modes: the Fuss

The CMB is (weakly) polarized. Why?

- Induced by radiation anisotropy through Thomson scattering.
- A quadrupole anisotropy in the radiation field is required.
- Happens at the *ionized* $\longleftrightarrow$ *neutral* interface (recombination, reionization)



(Wayne Hu)

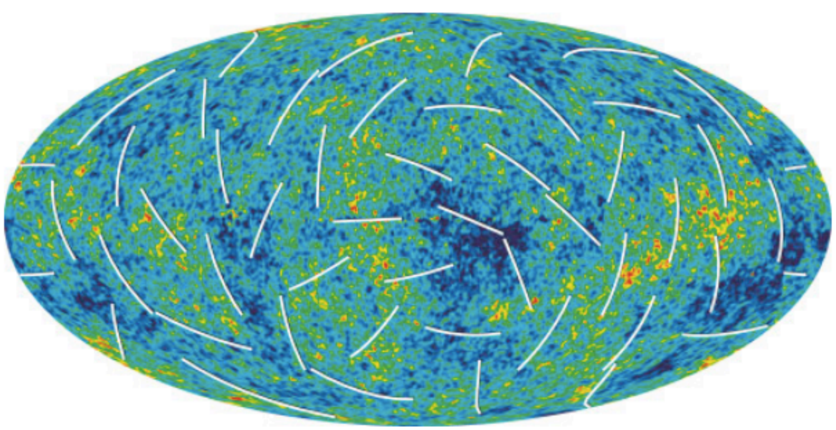
B-modes: the Fuss

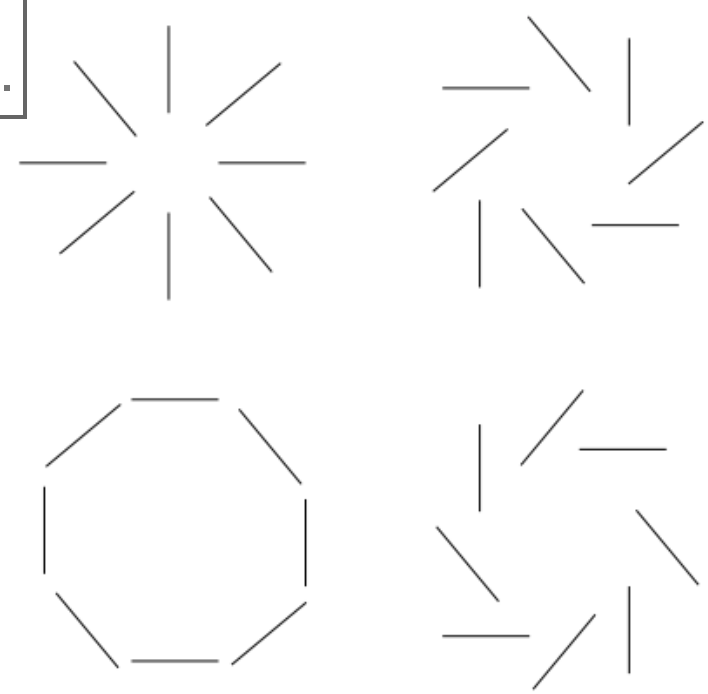
The CMB is (weakly) polarized. Why?

- Induced by radiation anisotropy through Thomson scattering.
- A quadrupole anisotropy in the radiation field is required.
- Happens at the *ionized* ⟷ *neutral* interface (recombination, reionization)
- Can be linearly decomposed into modes: E (gradient) B (curl)

B-modes *can* be generated by inflationary GWs.

Tensor-to-scalar ratio “r”: *holy grail*, *smoking-gun…*

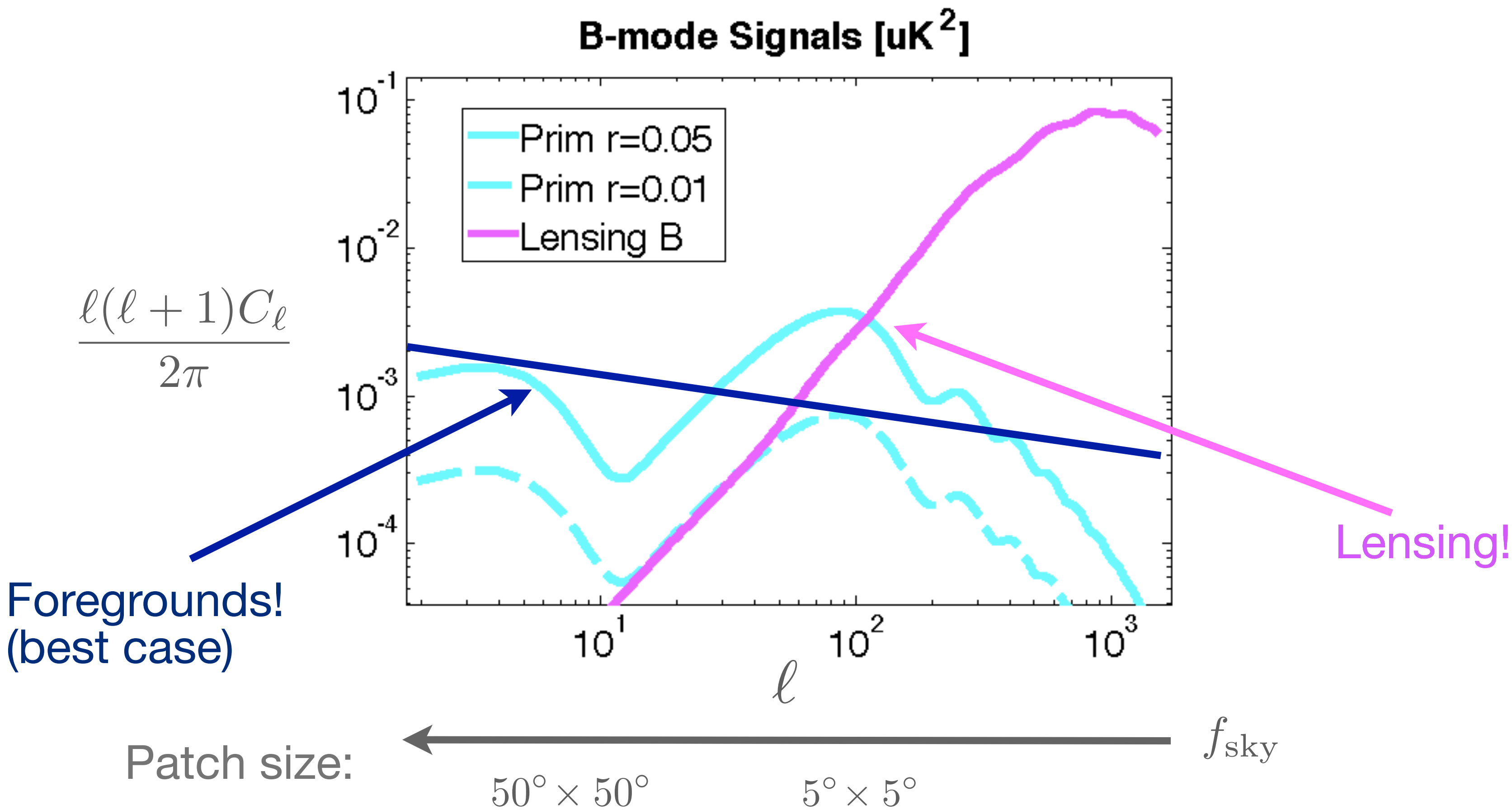




B-modes: Signals

Experimental tradeoffs include:

- Recombination vs. Reionization.
- Frequency coverage vs. Sky Coverage.



B-modes: Instrumental Noise

- Pixel noise: $$\sigma_{\rm pix} = s/\sqrt{t_{\rm pix}}$$

  $s = \mathrm{NET}_{\rm array}$ is determined by noise-equivalent temperature and array size.

  $t_{\rm pix} = T/N_{\rm pix}$ is the observation time for each pixel.

- Define inverse weight per solid angle: $w^{-1} = 4\pi s^2/T$

$\sigma_b^2 = \theta_{\rm fwhm}^2/(8\ln 2)$

- Noise power spectrum: $$C_\ell^N = \frac{\Omega\sigma_{\rm pix}^2}{N_{\rm pix}} e^{\ell^2\sigma_b^2} = \frac{\Omega s^2}{T} e^{\ell^2\sigma_b^2} = f_{\rm sky} w^{-1} e^{\ell^2\sigma_b^2}$$

- Fiducial experiments:

Experiment	$\theta_{\rm fwhm}$ [arcmin]	$f_{\rm sky}$ [%]	T [years]	s=NET$_{\rm array}$ [μK$\sqrt{\rm sec}$]
1	3.5	0.55	2	$\frac{480\sqrt{2}}{\sqrt{1274}} = 19$
2	5	0.06	4	15
3	30	1.52	6	25

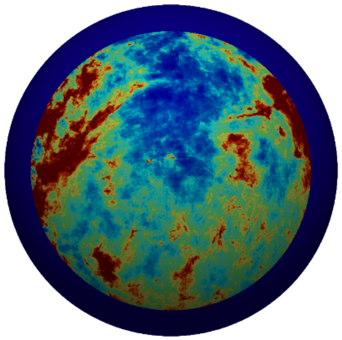
B-modes: Foregrounds

- For simplicity, we focus on a single frequency: 150 GHz.
- Dominant foreground is polarized emission from dust (PED) in the Galaxy.
- Assumption:

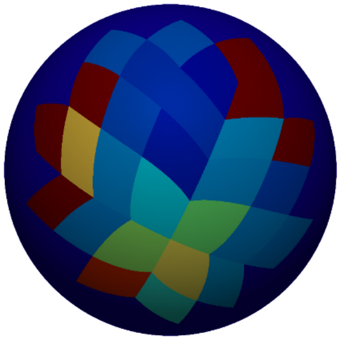
  PED power spectrum obeys a power law

$$\frac{\ell(\ell+1)}{2\pi} C_\ell^D = A\ell^m$$

(Clark et al. arXiv:1211.6404)



$\sqrt{Q^2+U^2}$



$(\sigma_Q^2+\sigma_U^2)/2$

- Assumption:

  Theoretical templates such as FGPol

  provide reasonable ballpark predictions.

- Assumption:

  Can read off the amplitude from the variance in the PED template.

$$\sigma^2 = \frac{1}{4\pi} \sum_{\ell=2}^{\ell_{\max}} (2\ell+1) C_\ell^D B_\ell^2(\theta_s)$$

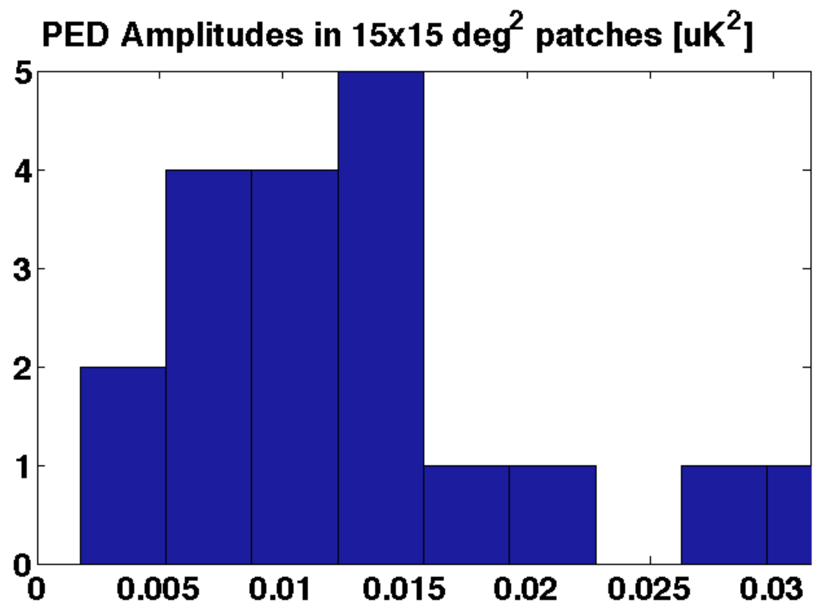
B-modes: Foregrounds

- Assumption:

  There's some percentage of patches better avoided: 67% (outside knowledge)
- Assumption:

  Average dust polarization fraction outside the galactic plane is 3.6% or 10%.

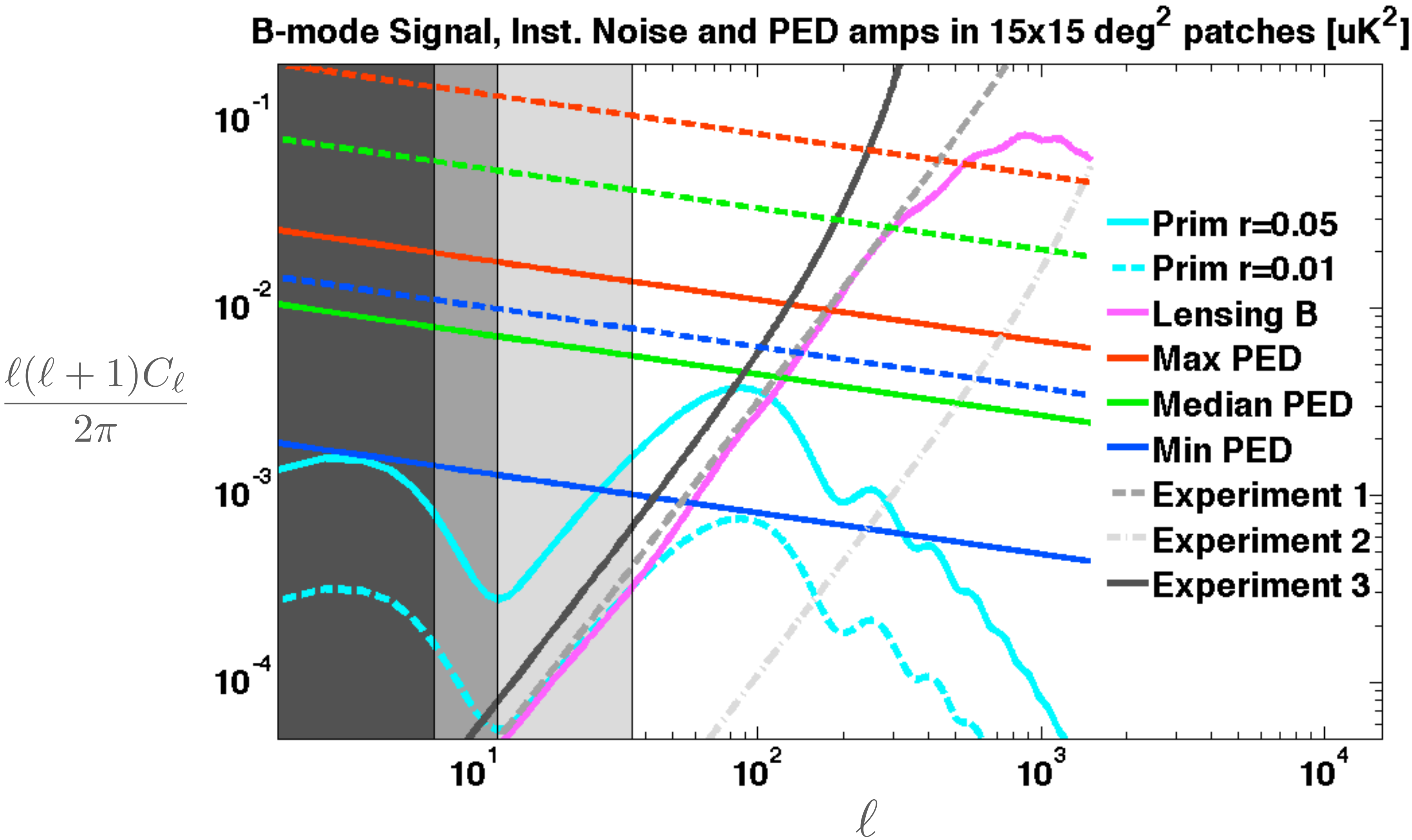
- For a 15-degree patch size (with 3.6% PED normalization), we get:



- There's room for improvement!

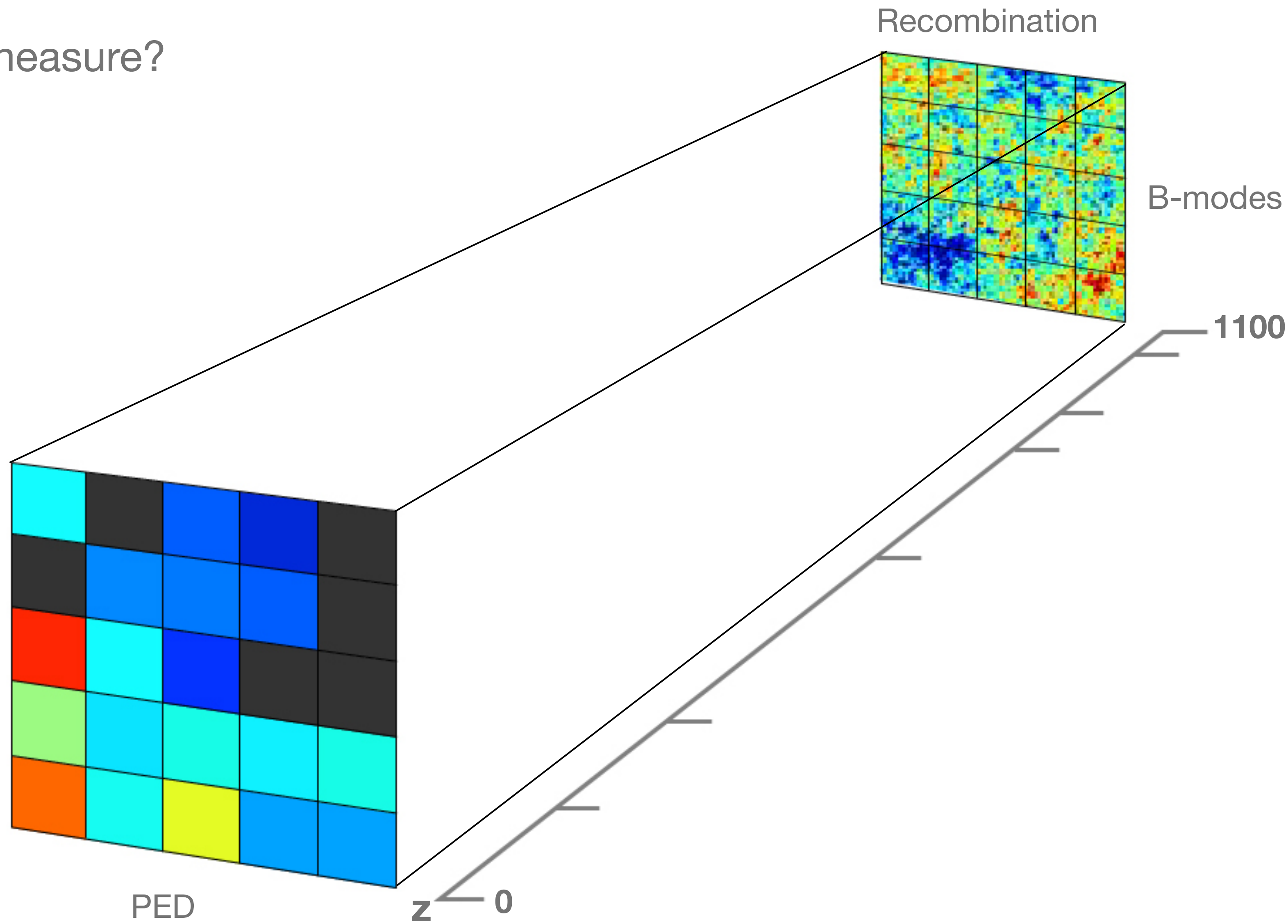
B-modes: The Good, the Bad and the Ugly

- The signal, noise and foregrounds:



B-mode Surveys: Simulations

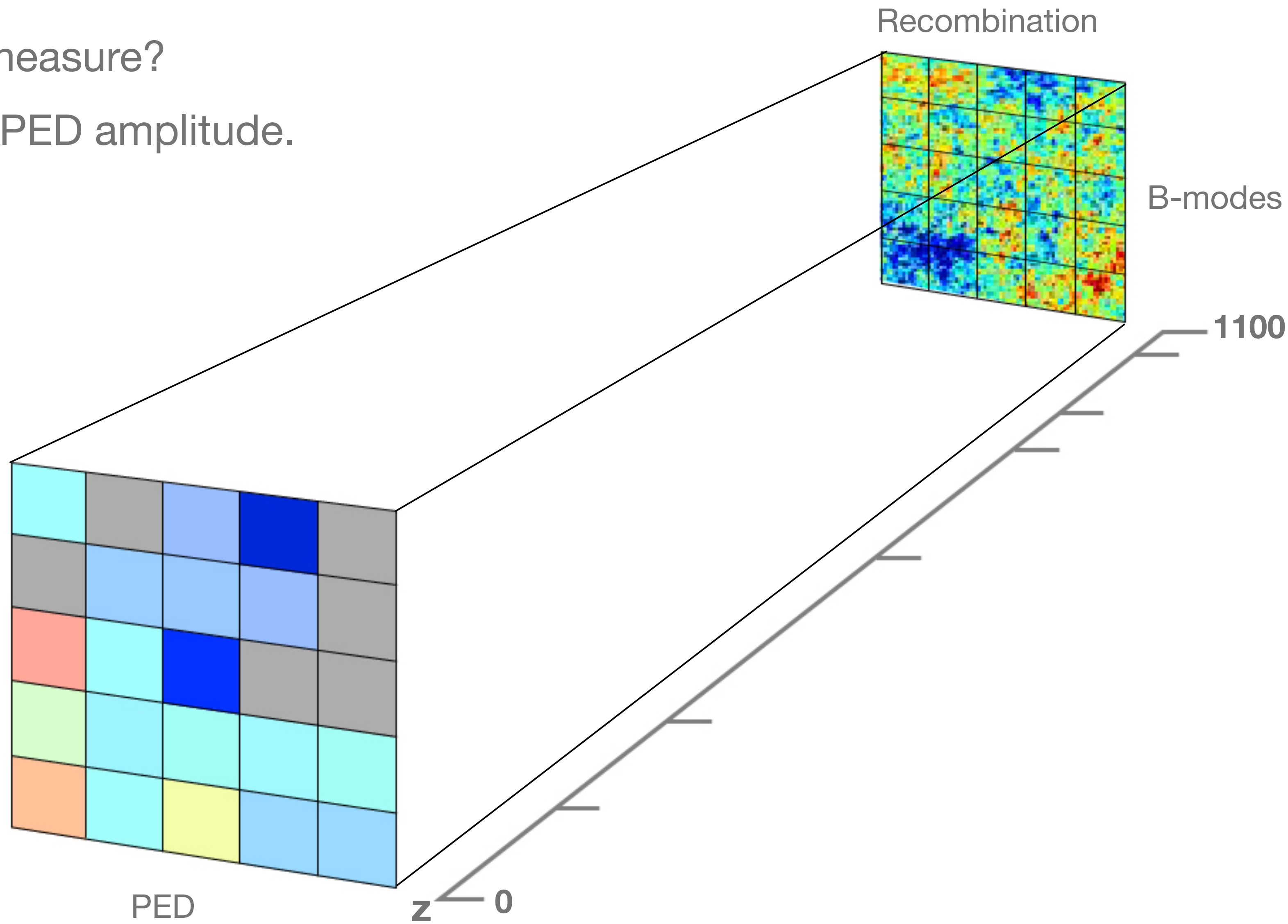
What do we measure?



B-mode Surveys: Simulations

What do we measure?

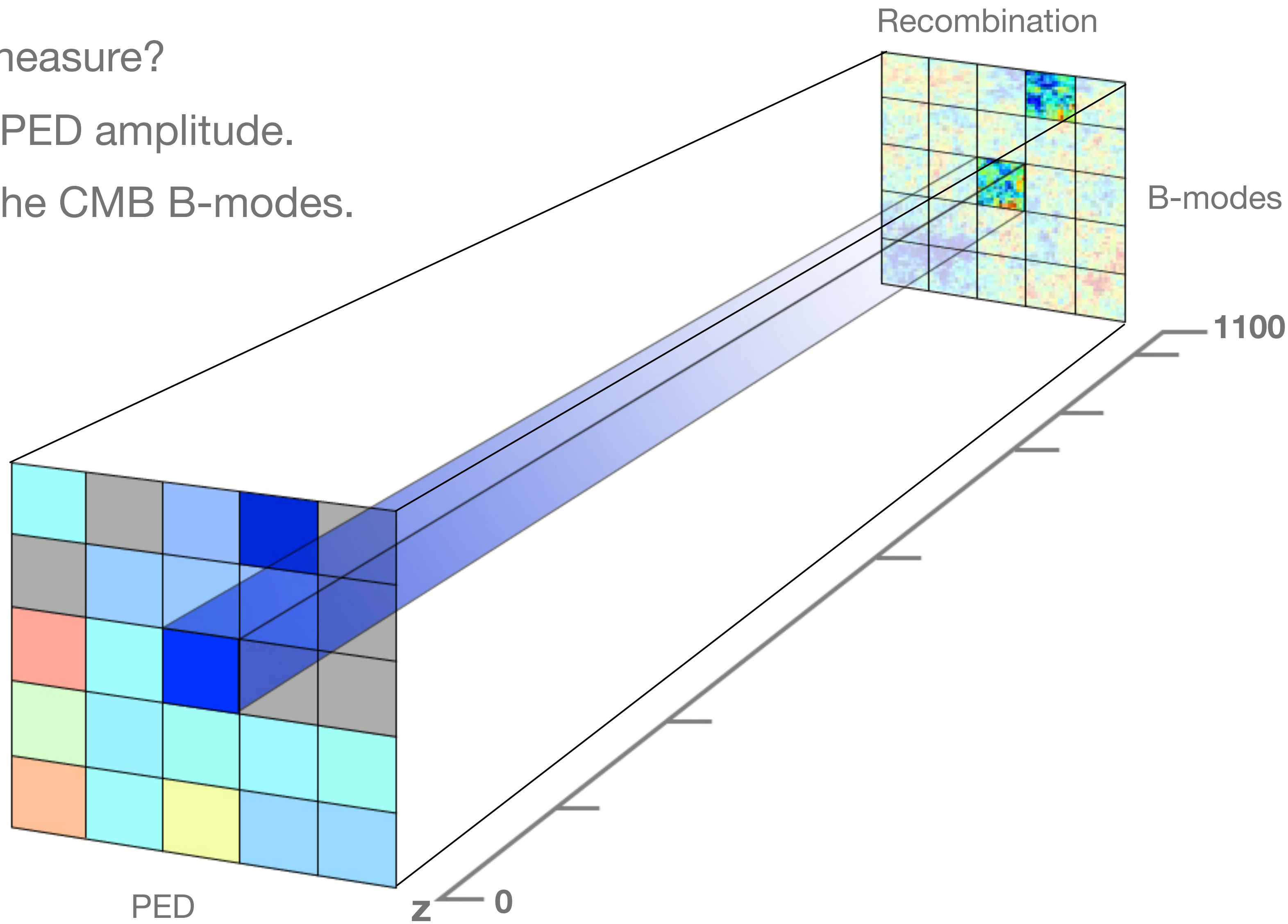
- Initially: the PED amplitude.



B-mode Surveys: Simulations

What do we measure?

- Initially: the PED amplitude.
- Ultimately: the CMB B-modes.



B-modes: Statistics and Simulations

- In a ML analysis, the Fisher forecast for the error in measured amplitude:

$$\frac{1}{\sigma_A^2} = \sum_\ell \left(\frac{\partial C_\ell}{\partial A} \right)^2 \frac{1}{\sigma_\ell^2}$$

- Assumption:

  Likelihood function is Gaussian in the vicinity of its maximum:

$$\sigma_A \longrightarrow \text{``1-sigma''}$$

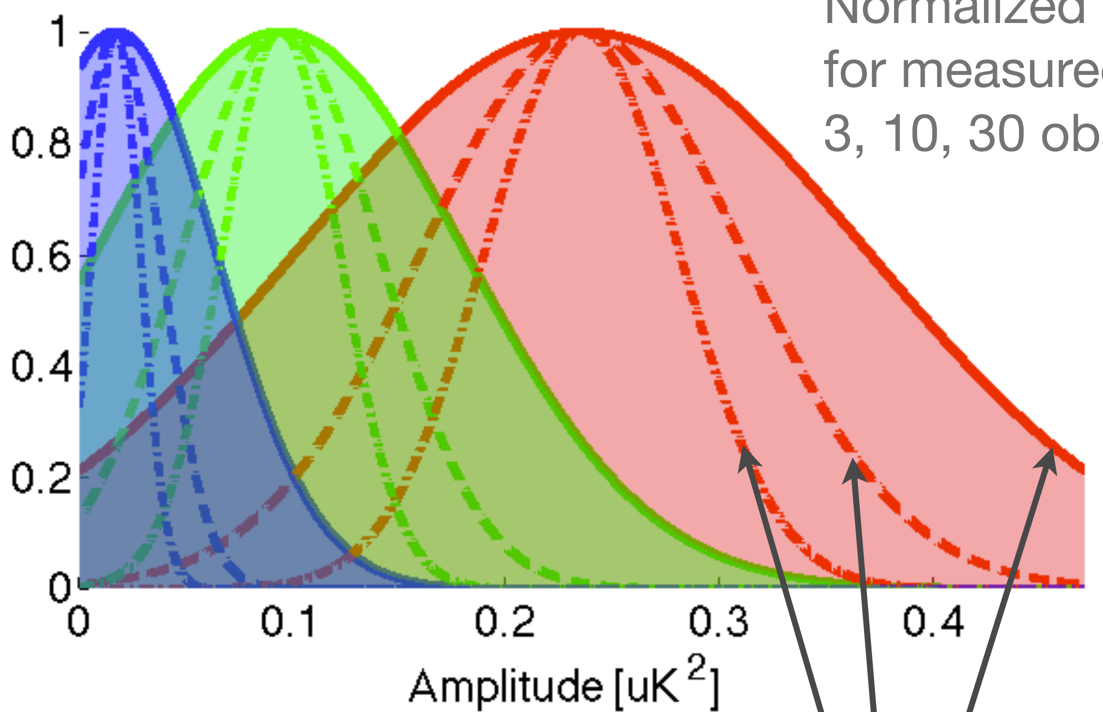
- The ``1-sigma'' error for each multipole:

$$\sigma_\ell^{\widehat{A}_p} = \sqrt{\frac{2}{f_{\rm sky}(2\ell+1)}} \left(\underbrace{A_p \tilde{C}_\ell^D}_{C_\ell^D/A_p} + \underbrace{\alpha C_\ell^L}_{\text{lensing residual}} + \underbrace{f_{\rm sky} w^{-1} e^{\ell^2 \sigma_b^2}}_{C_\ell^N} \right)$$

- Patch error:

$$\sigma^{\widehat{A}_p} = \left[\frac{f_{\rm sky}}{2} \sum_{\ell_{\rm min}}^{\ell_{\rm max}} \frac{(2\ell+1)(\tilde{C}_\ell^D)^2}{\left(A_p \tilde{C}_\ell^D + \alpha C_\ell^L + f_{\rm sky} w^{-1}(t_p) e^{\ell^2 \sigma_b^2} \right)^2} \right]^{-\frac{1}{2}}$$

B-modes: Statistics and Simulations



Normalized likelihood curves for measured amplitudes after 3, 10, 30 observation days

- Patch error:

$$\sigma^{\widehat{A}_p} = \left[\frac{f_{\text{sky}}}{2} \sum_{\ell_{\min}}^{\ell_{\max}} \frac{(2\ell+1)(\tilde{C}_\ell^D)^2}{\left(A_p \tilde{C}_\ell^D + \alpha C_\ell^L + f_{\text{sky}} w^{-1}(t_p) e^{\ell^2 \sigma_b^2}\right)^2} \right]^{-\frac{1}{2}}$$

B-modes: Statistics and Simulations

- How do we judge a survey (bandit) strategy?

  Figure-of-merit is the ``1-sigma'' error for the tensor-to-scalar ratio:

$$\sigma_p^r = \left[\frac{f_{\rm sky}}{2} \sum_{\ell_{\rm min}}^{\ell_{\rm max}} \left(\frac{\sqrt{(2\ell+1)}\tilde{C}_\ell^B}{A_p \tilde{C}_\ell^D + \alpha C_\ell^L + f_{\rm sky}\omega(t_p)^{-1} e^{\ell^2 \sigma_b^2}} \right)^2 \right]^{-\frac{1}{2}}$$

  (when comparing to the null hypothesis)

- After total observing time $T = \sum_{p=1}^{n_p} t_p$:

$$\boxed{\sigma^{\rm r} = \left[\frac{f_{\rm sky}}{2} \sum_{p=1}^{n_p} \sum_{\ell_{\rm min}}^{\ell_{\rm max}} \left(\frac{\sqrt{(2\ell+1)}\tilde{C}_\ell^B}{\underbrace{A_p \tilde{C}_\ell^D}_{\text{foregrounds}} + \underbrace{\alpha C_\ell^L}_{\text{lensing residual}} + \underbrace{f_{\rm sky}\omega(t_p)^{-1} e^{\ell^2 \sigma_b^2}}_{\text{instrumental noise}}} \right)^2 \right]^{-\frac{1}{2}}}$$

Outline

- The Multi-Armed-Bandit Problem
- Heuristic Solution Algorithms
- B-modes: The good, the bad and the ugly
- Results and Discussion

Results: Scenarios

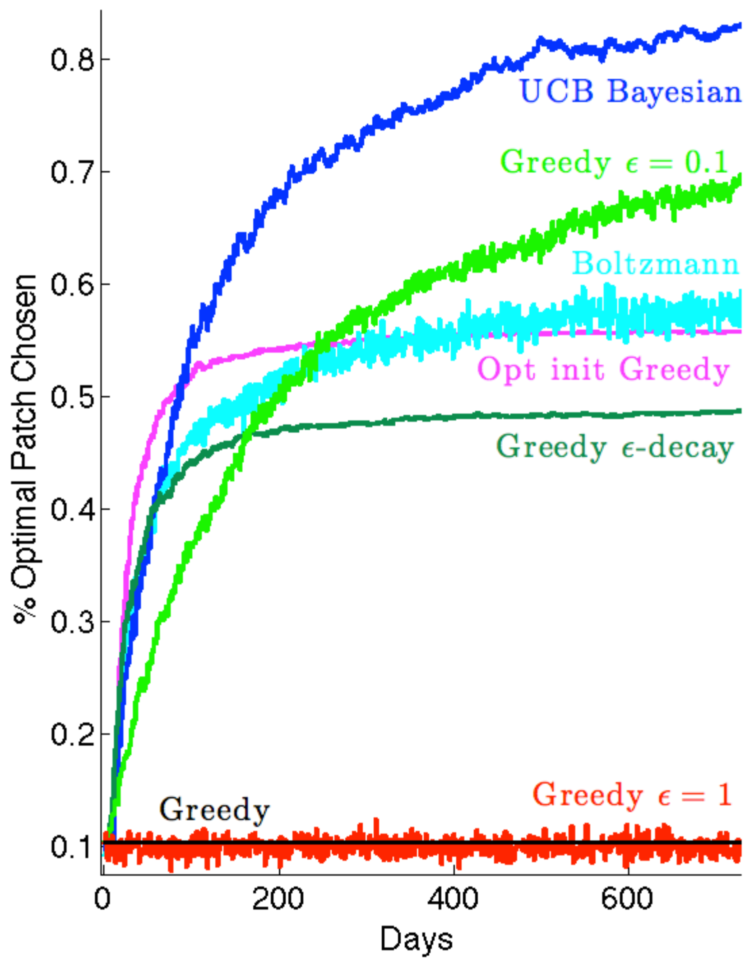
We consider three scenarios:

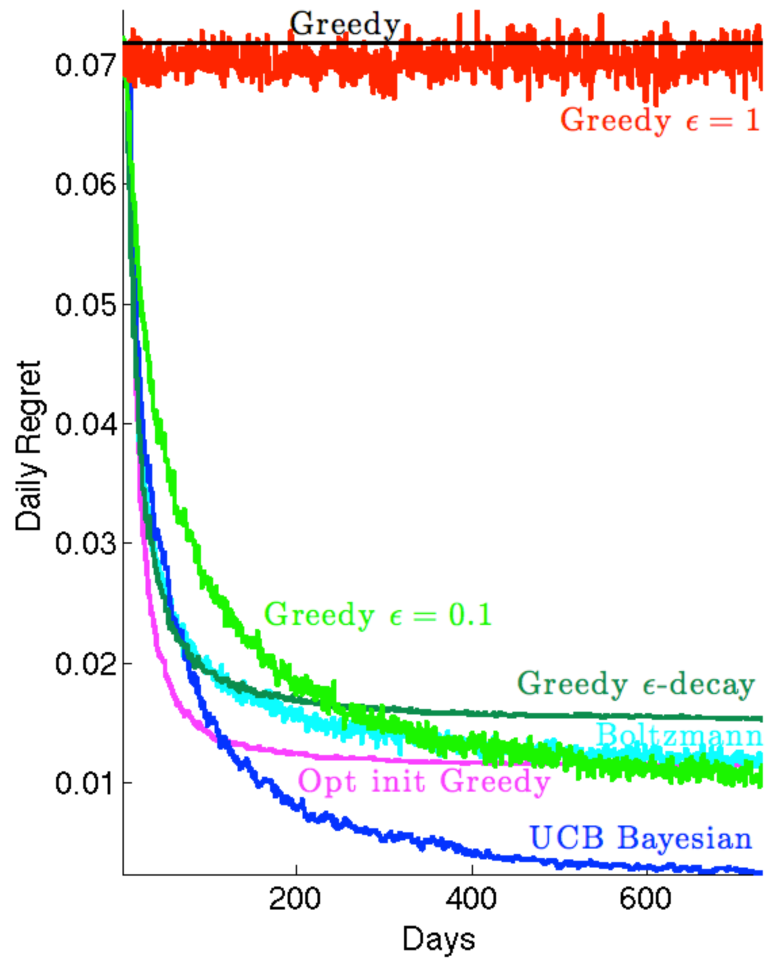
- Pessimistic: normalization to 10%, no de-lensing.
- Conservative: normalization to 3.6%, no de-lensing.
- Optimistic: normalization to 3.6% and 80% de-lensing.

We use 1,000 simulations to acquire an ensemble average for comparison.

Results

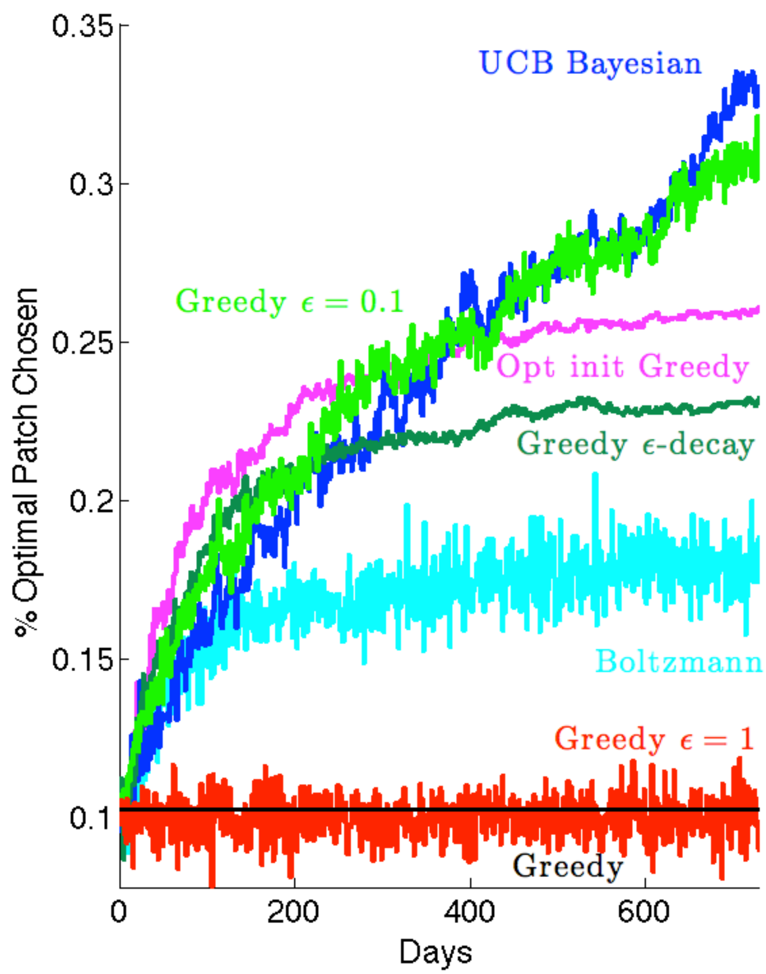
- In the pessimistic scenario (experiment 1):

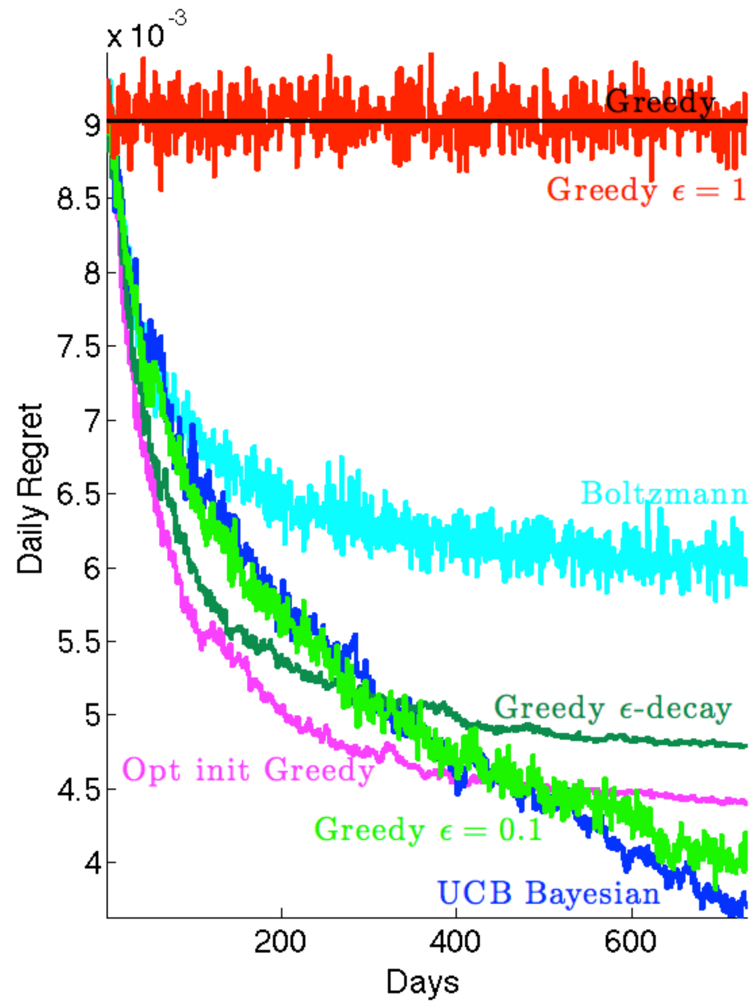




Results

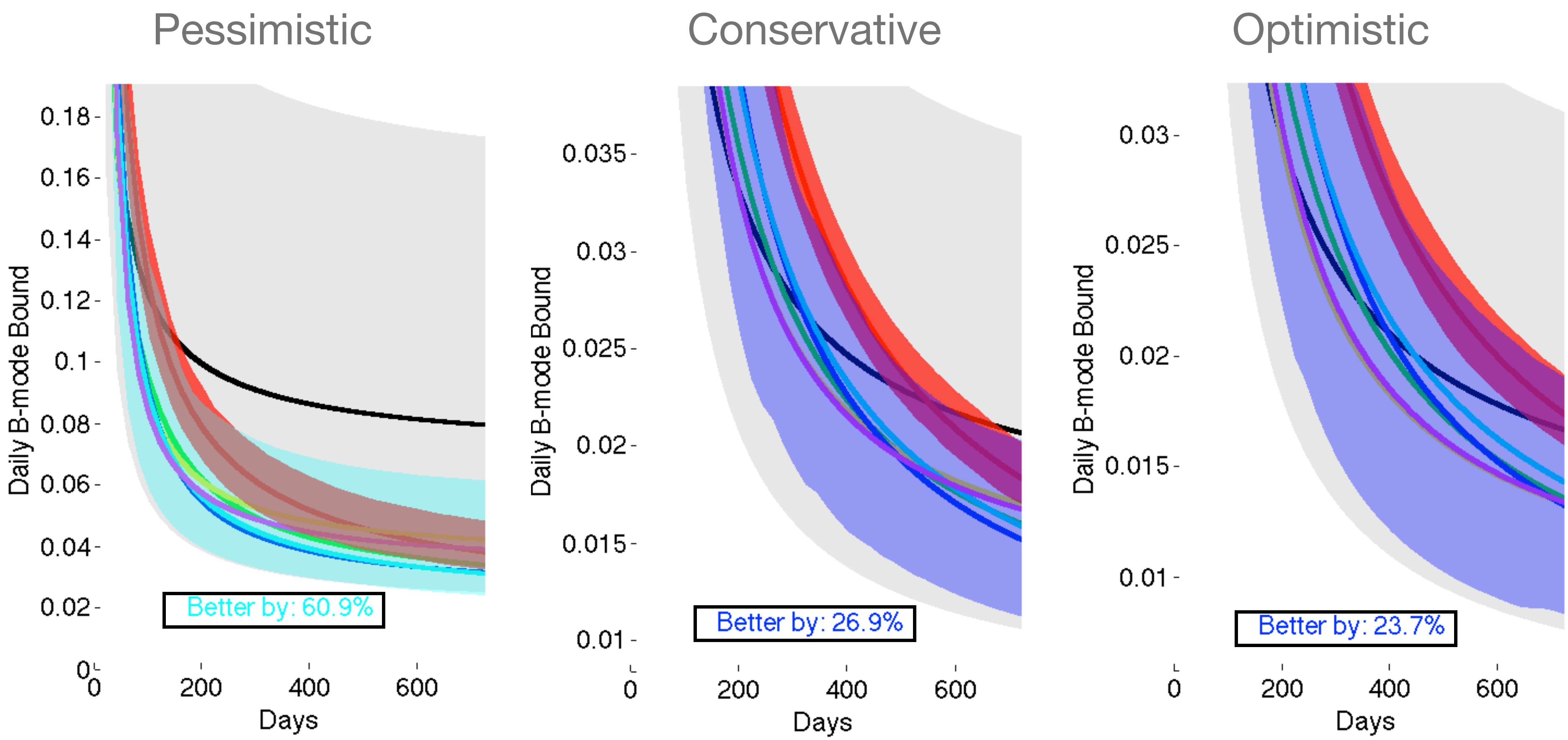
- In the optimistic scenario (experiment 1):





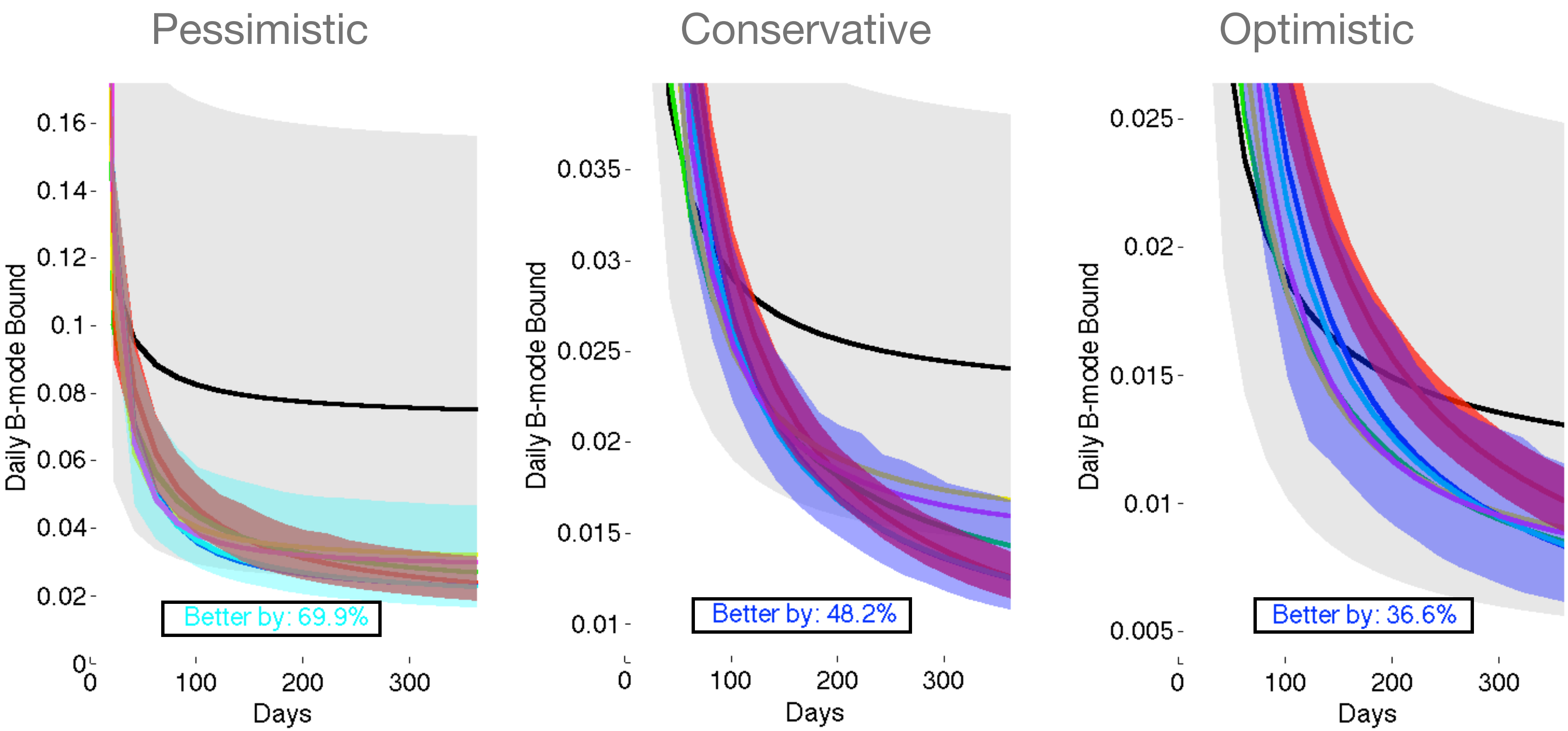
Results: B-mode detection

Experiment	θ_{fwhm} [arcmin]	f_{sky} [%]	T [years]	s=NET$_{\mathrm{array}}$ [μK$\sqrt{\mathrm{sec}}$]	n_p	t_{step} [days]
1	3.5	0.55	2	$\frac{480\sqrt{2}}{\sqrt{1274}} = 19$	10	1



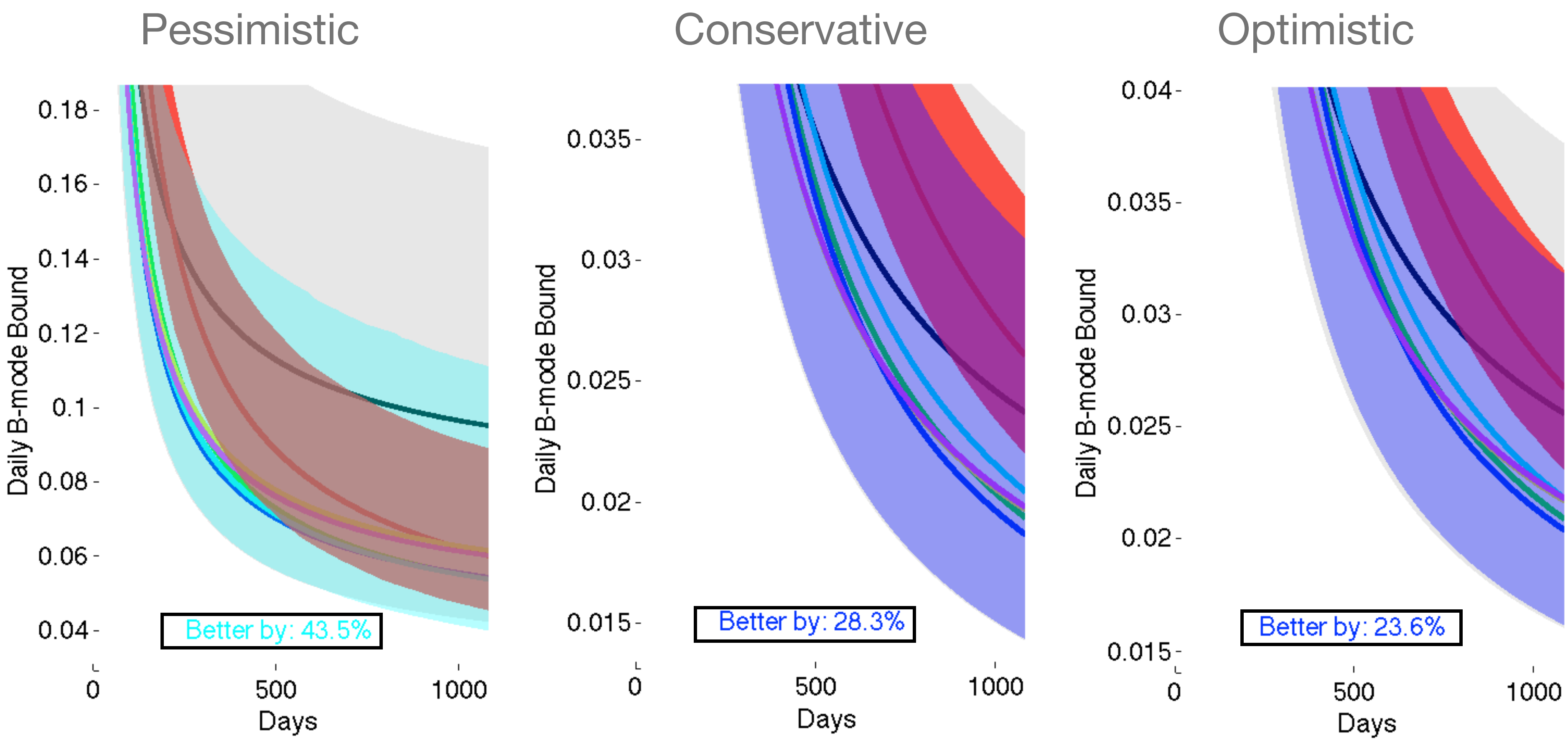
Results: B-mode detection

Experiment	θ_{fwhm} [arcmin]	f_{sky} [%]	T [years]	s=NET$_{\mathrm{array}}$ [μK$\sqrt{\mathrm{sec}}$]	n_p	t_{step} [days]
2	5	0.06	4	15	15	4



Results: B-mode detection

Experiment	θ_{fwhm} [arcmin]	f_{sky} [%]	T [years]	s=NET$_{\mathrm{array}}$ [μK$\sqrt{\mathrm{sec}}$]	n_p	t_{step} [days]
3	30	1.52	6	25	5	2



Discussion

Recap:

- Pessimistic scenario: Up to 75% improvement on average.
- Conservative scenario: Up to 50% improvement on average.
- Optimistic scenario: Up to 40% improvement on average.
- Similar improvements when comparing worst-case performances.
- Improvement in any experiment. Maximized with high resolution + sensitivity.

(Some) Caveats:

- Single frequency. In practice, remain with (non-zero) foreground residuals.
- Cost of moving telescope target. This *should* be taken into account.

Actual performance may deteriorate. However:

- Methods can be optimized further.
- Identification of optimal patches will reap future benefits.

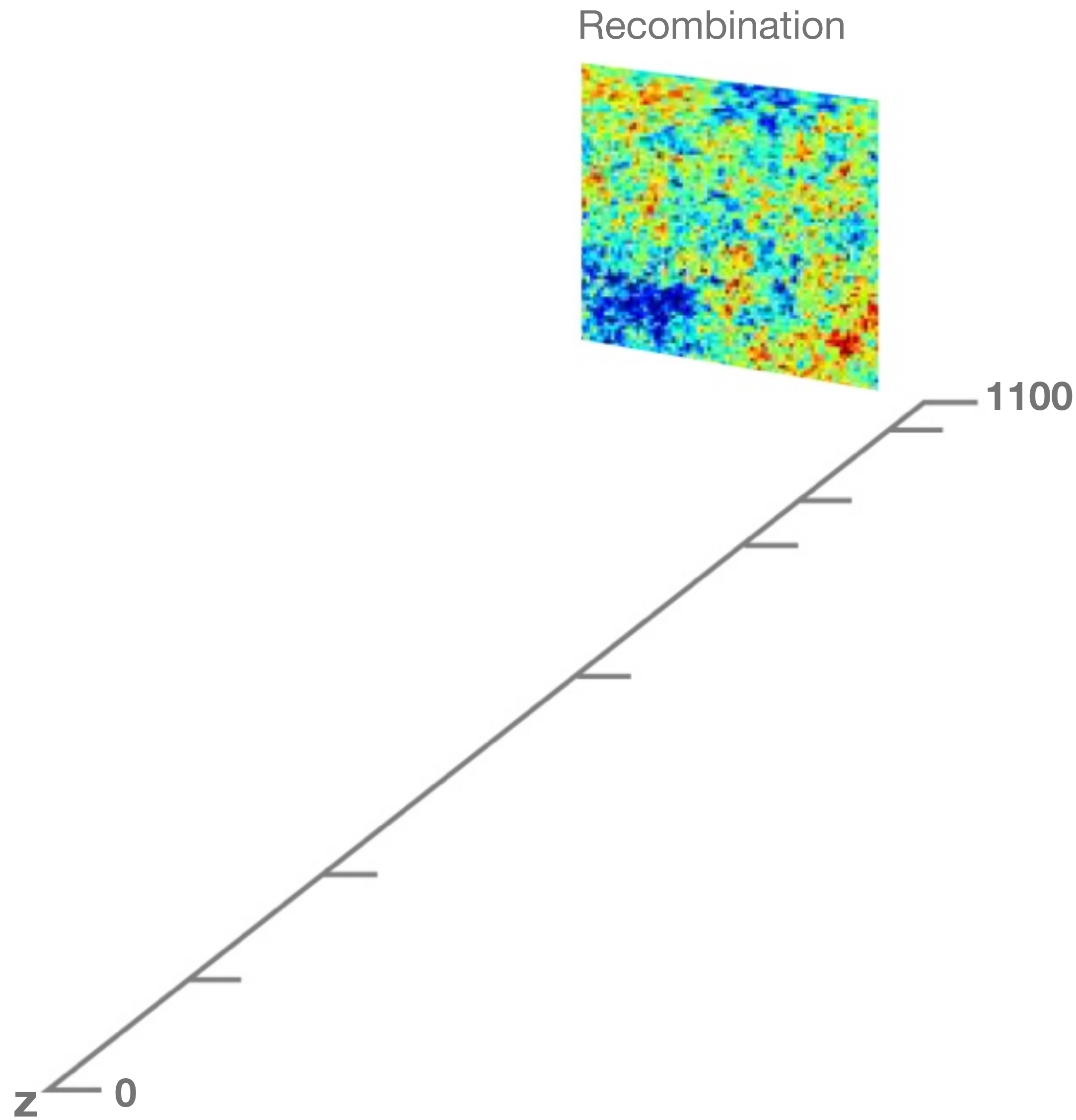
Outline

- The Multi-Armed-Bandit Problem
- Heuristic Solution Algorithms
- B-modes: The good, the bad and the ugly
- Results and Discussion
- MAB Strategies Elsewhere

MAB Strategies Elsewhere

- 21-cm stochastic fluctuations.

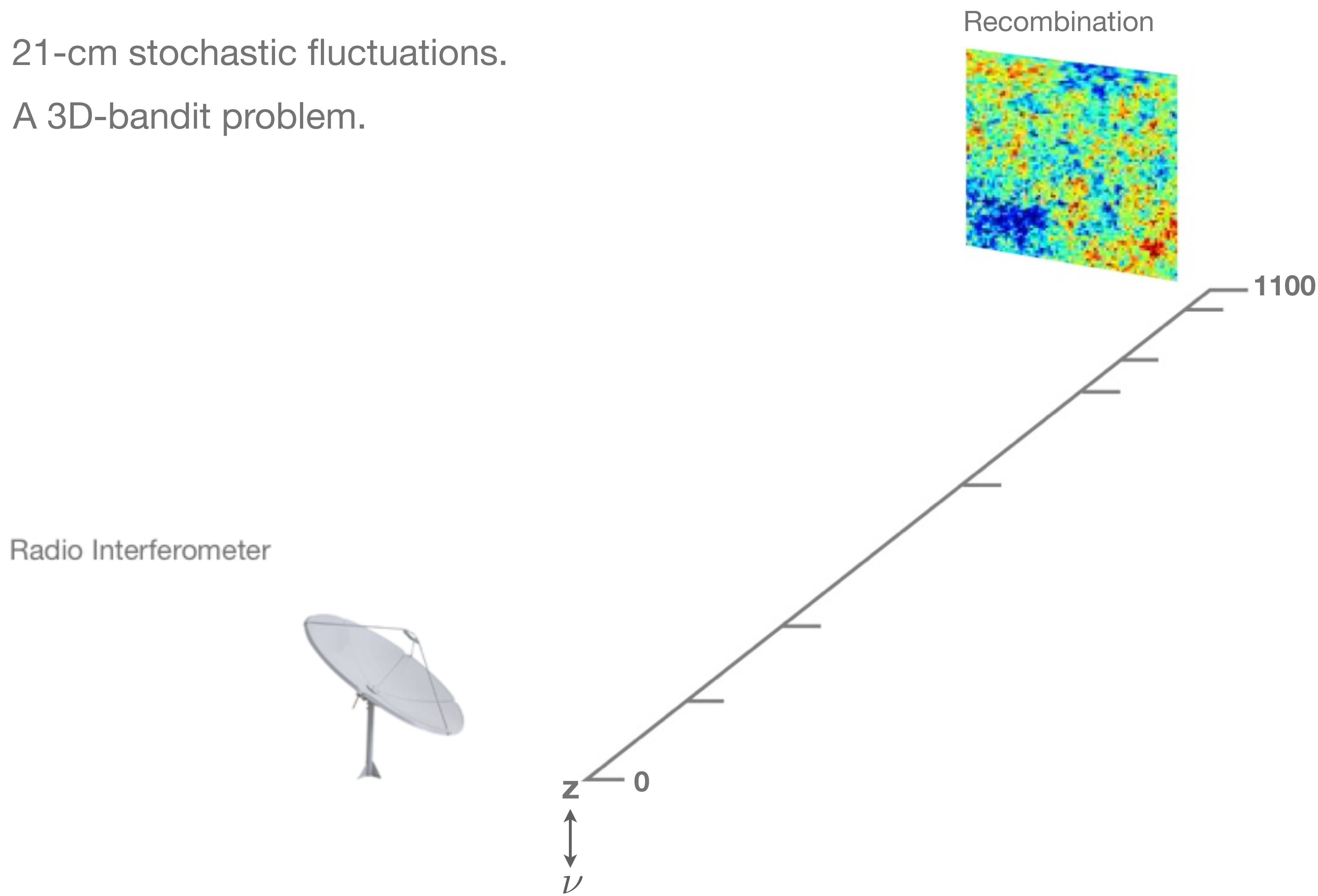
  A 3D-bandit problem.



MAB Strategies Elsewhere

- 21-cm stochastic fluctuations.

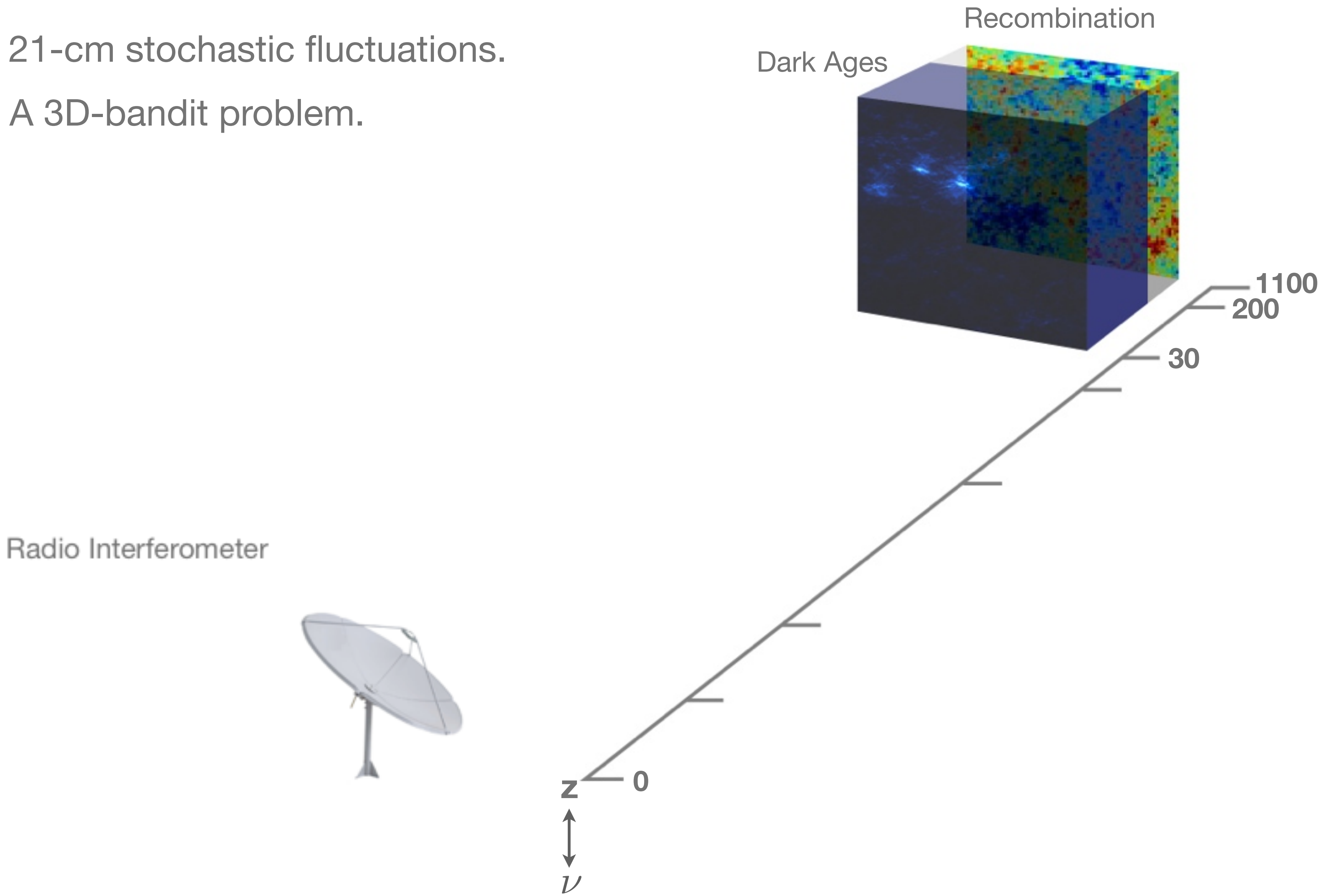
  A 3D-bandit problem.



MAB Strategies Elsewhere

- 21-cm stochastic fluctuations.

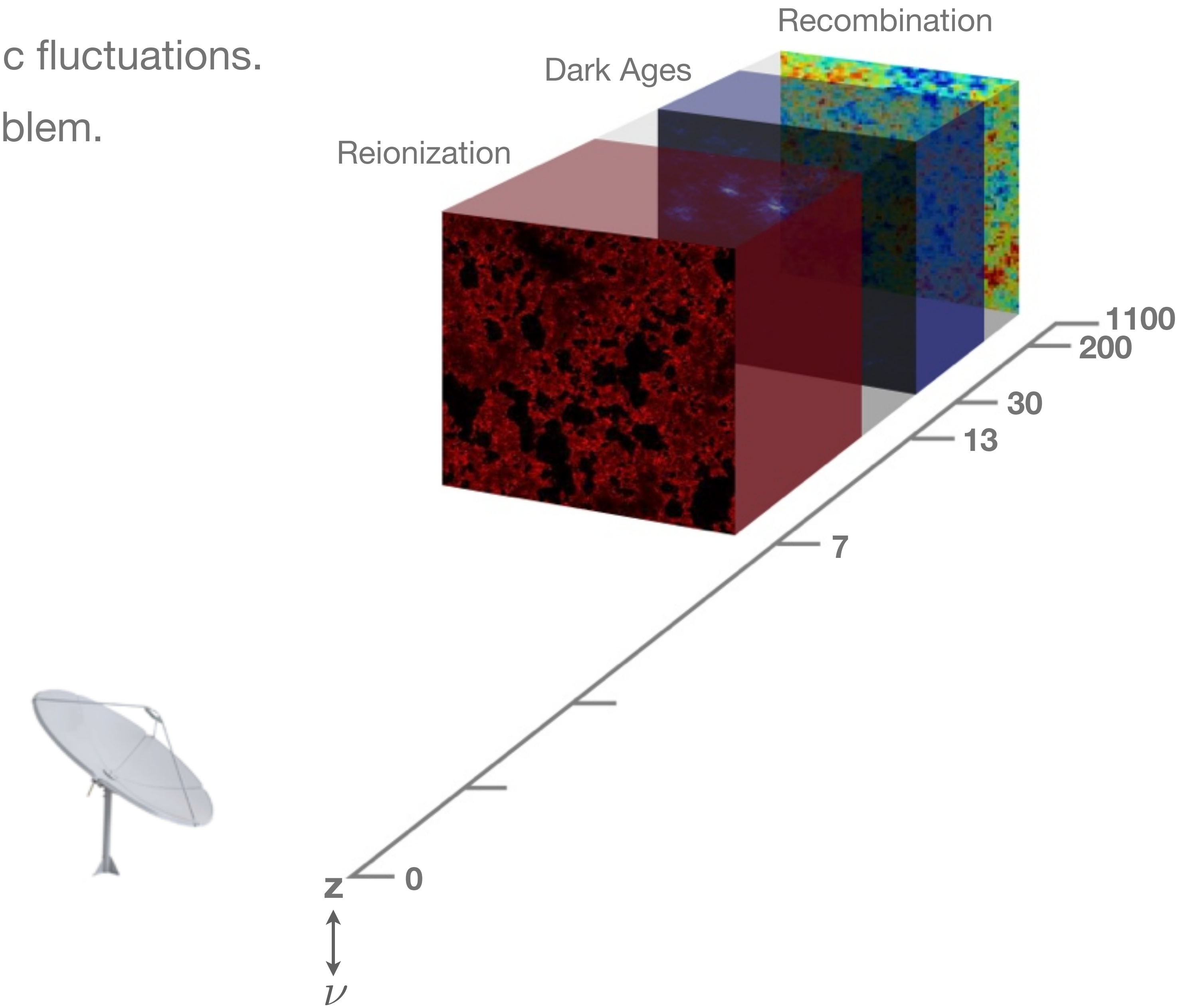
  A 3D-bandit problem.



MAB Strategies Elsewhere

- 21-cm stochastic fluctuations.

  A 3D-bandit problem.



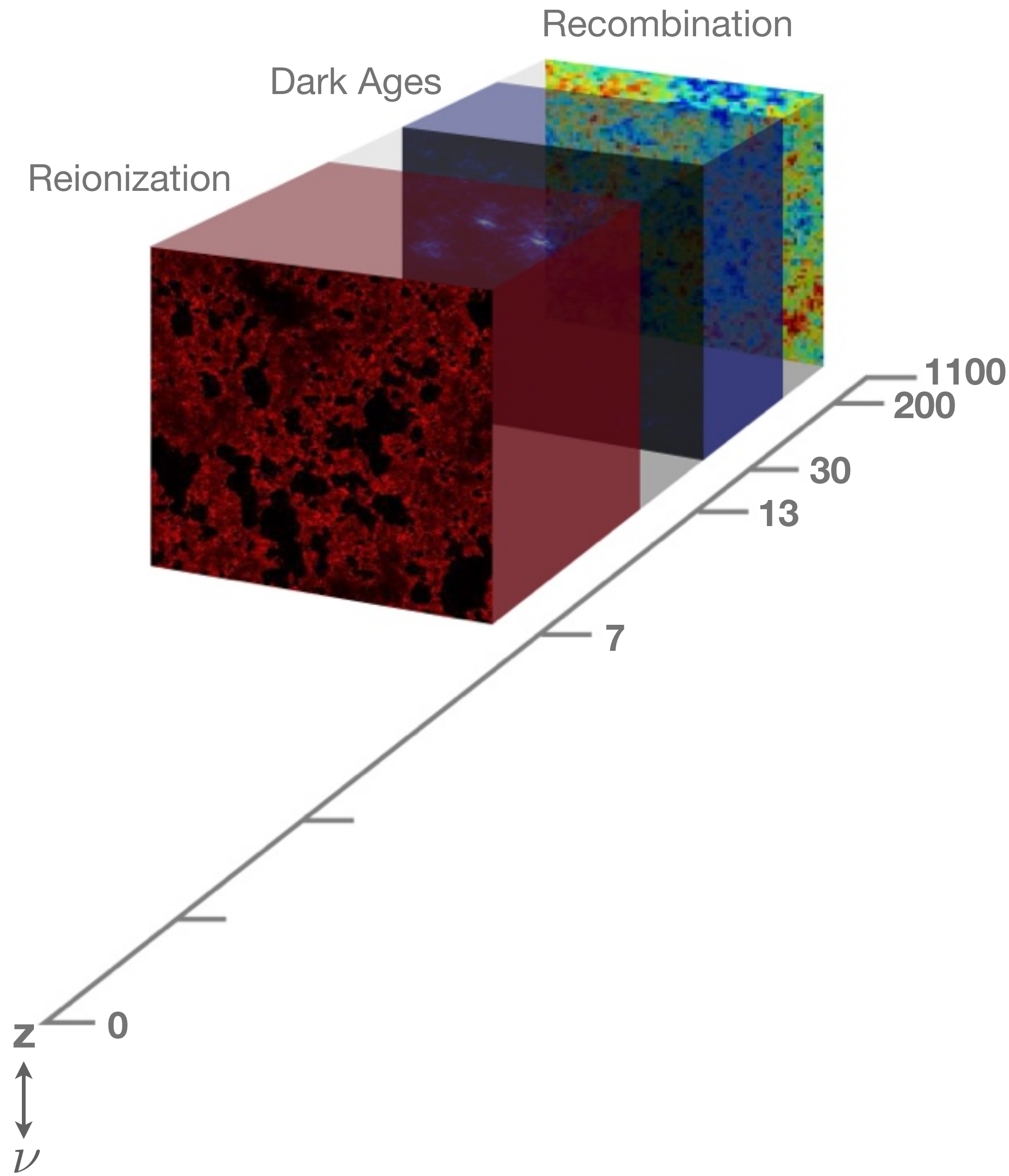
MAB Strategies Elsewhere

- 21-cm stochastic fluctuations.

  A 3D-bandit problem.

  Adds another dimension.

Radio Interferometer



MAB Strategies Elsewhere

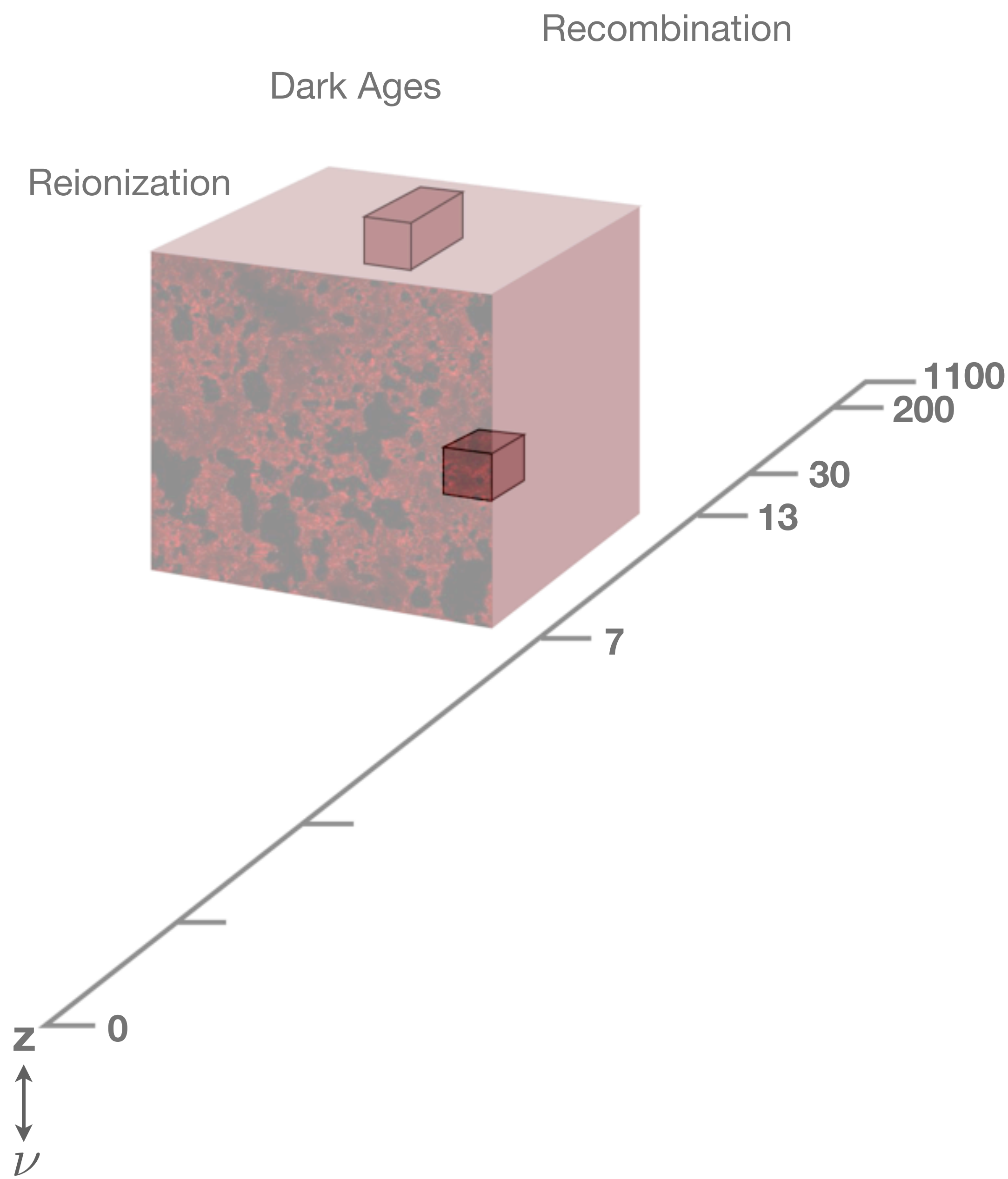
- 21-cm stochastic fluctuations.

  A 3D-bandit problem.

  Adds another dimension.

  MWA: Greedy $\epsilon = 0 \text{ vs. } 1$

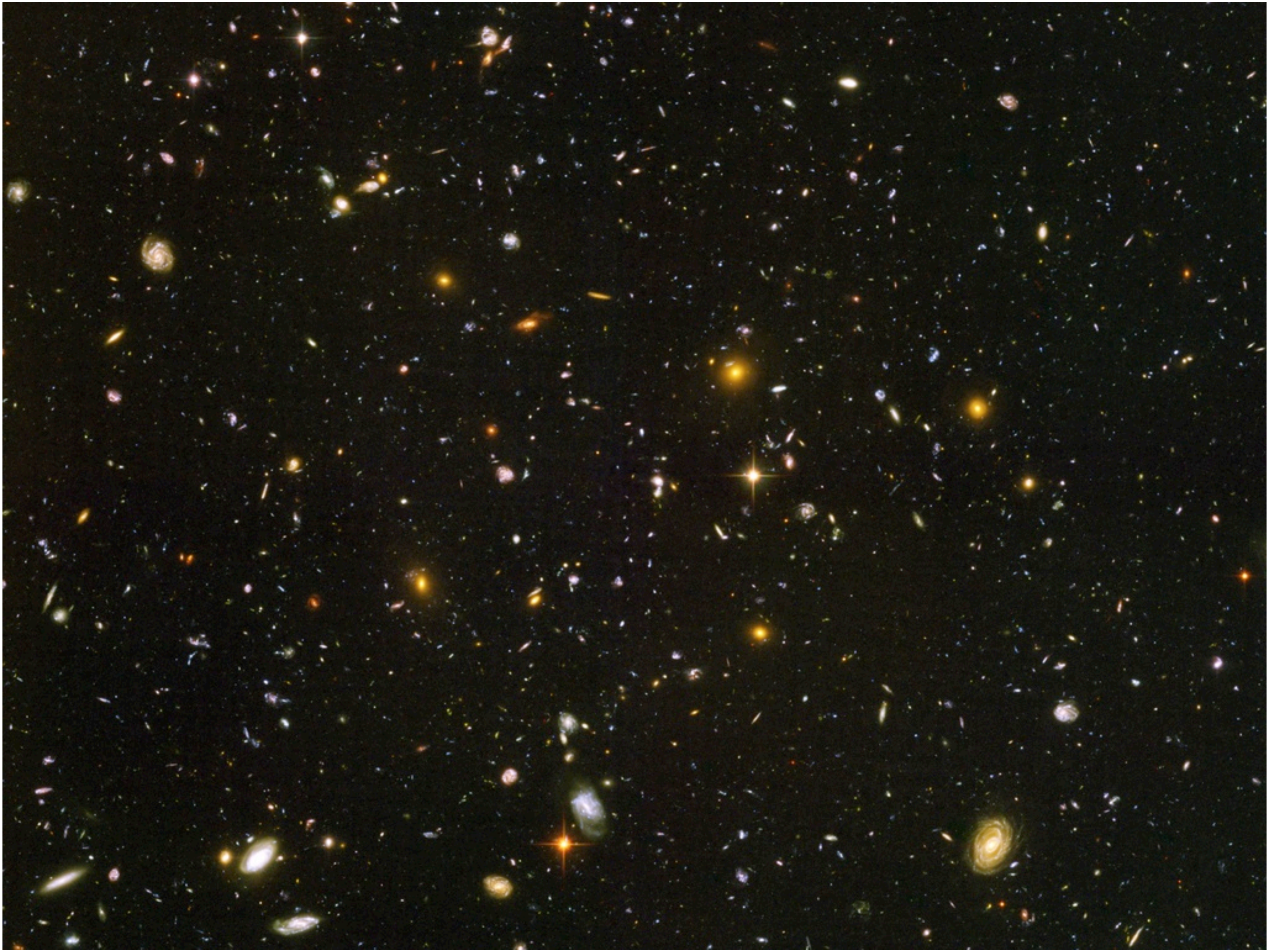
  (Thyagarajan et al. arXiv:1308.0565)

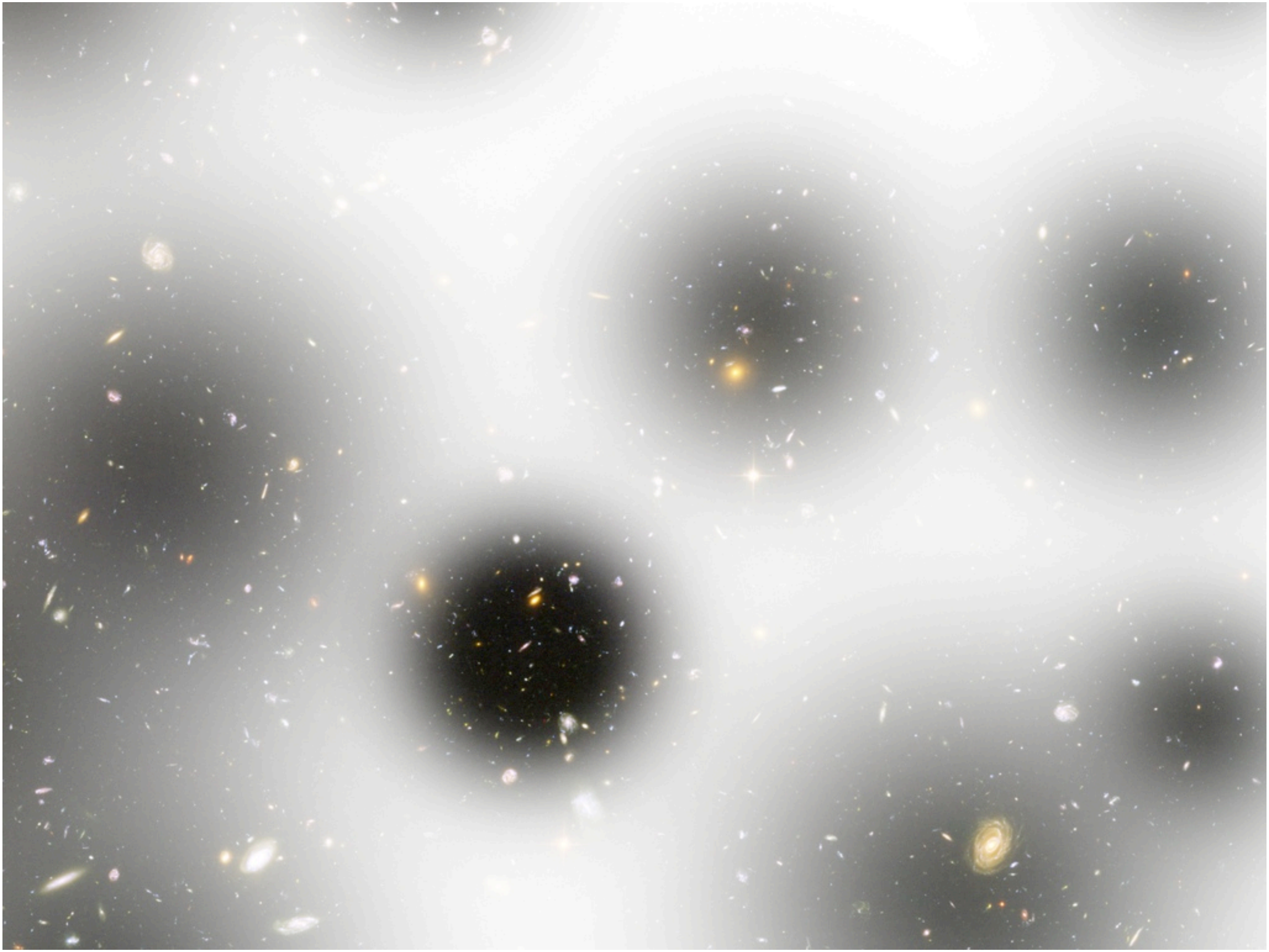


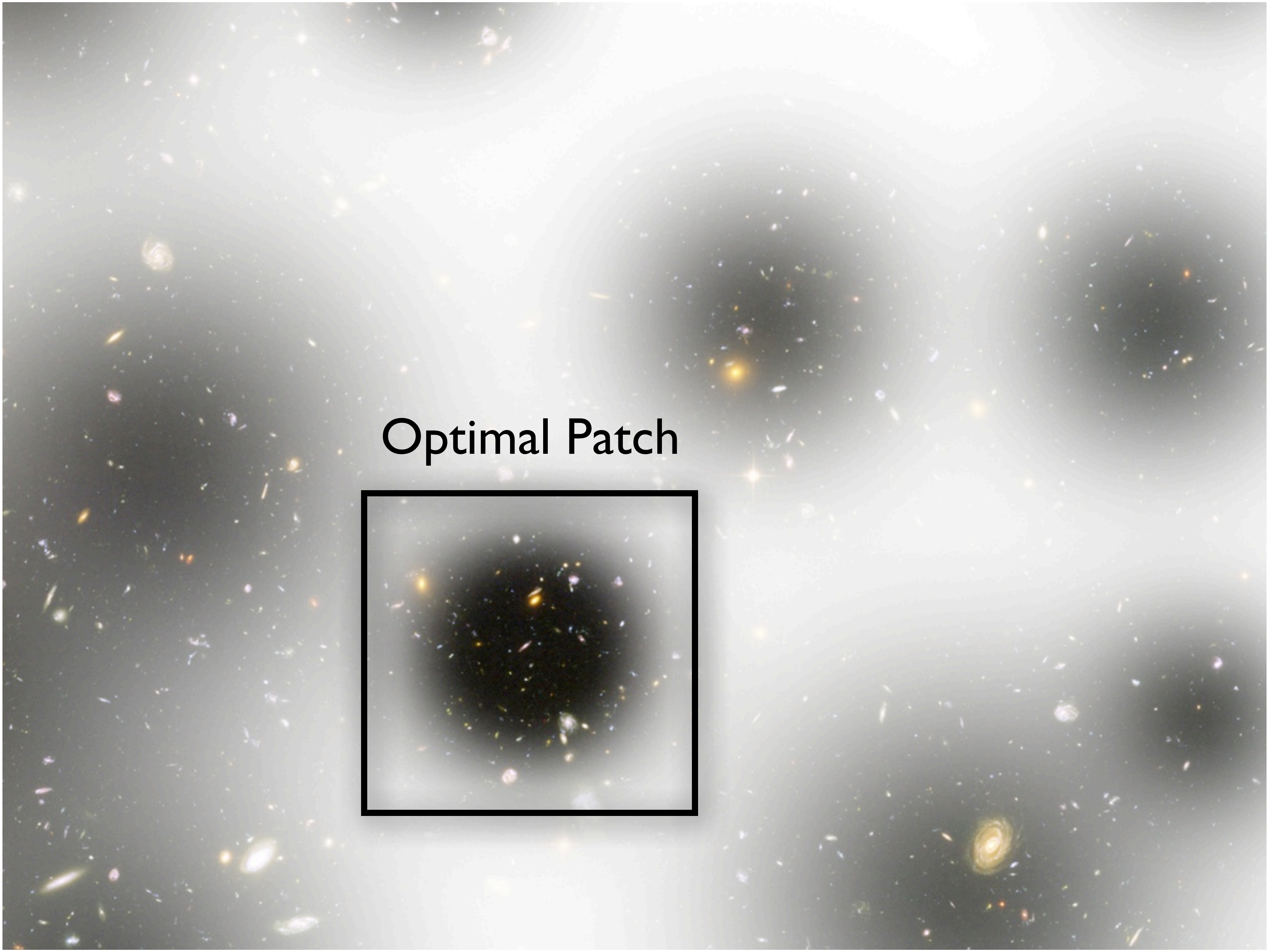
Radio Interferometer

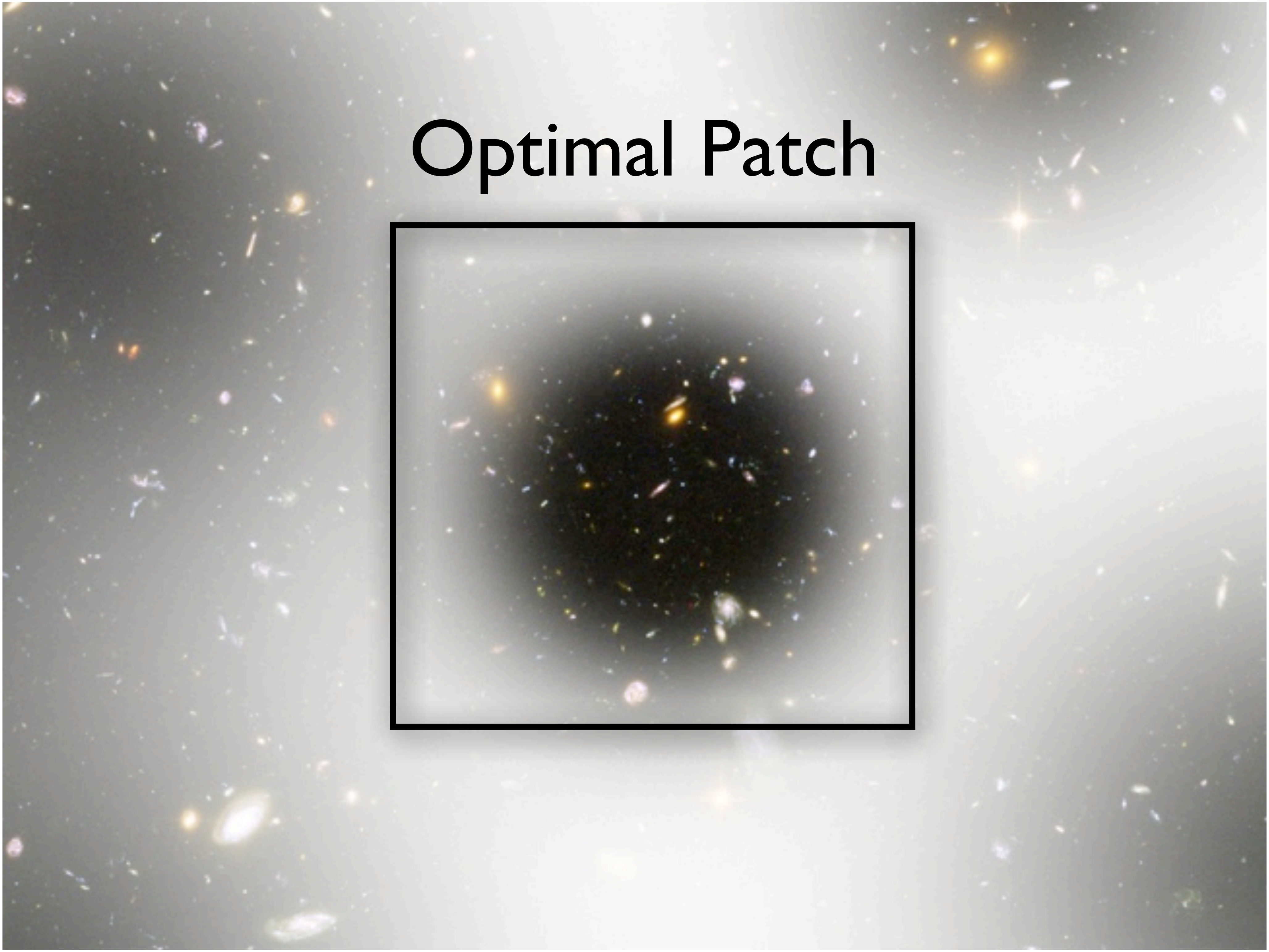
MAB Elsewhere

- Deep-field imaging:
  From HST to JWST?

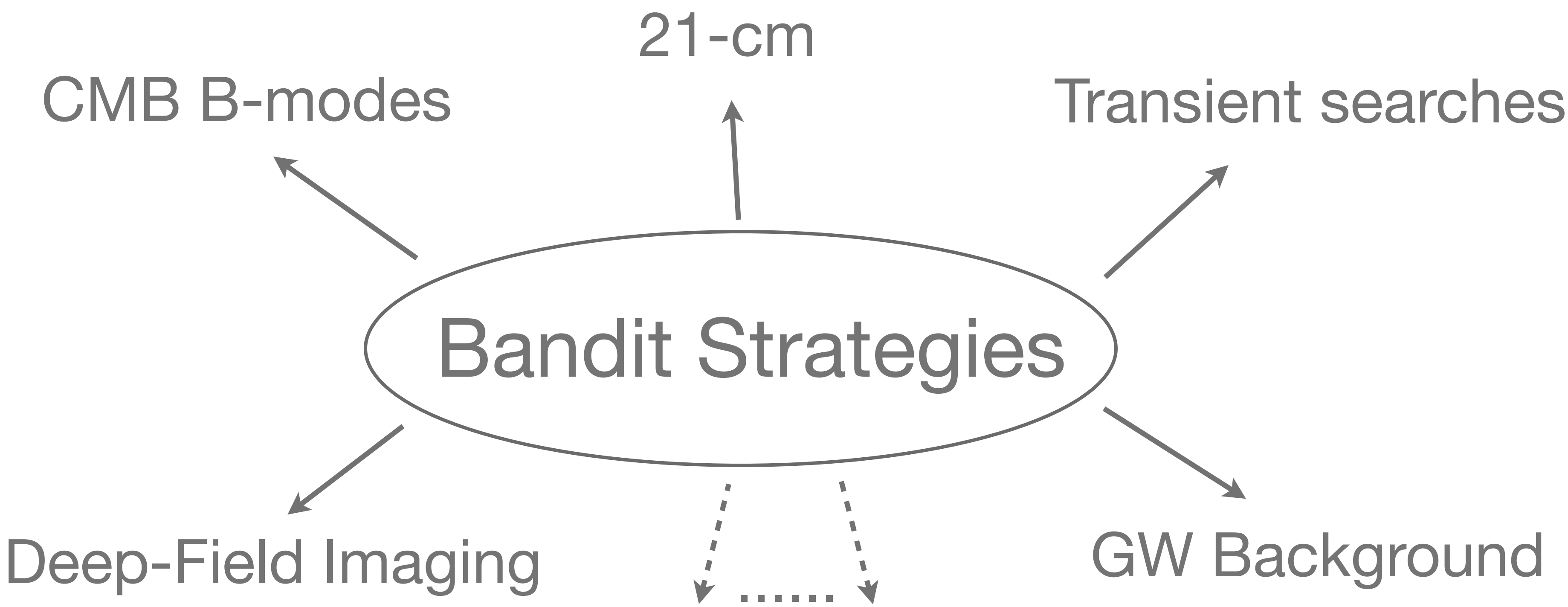





Optimal Patch

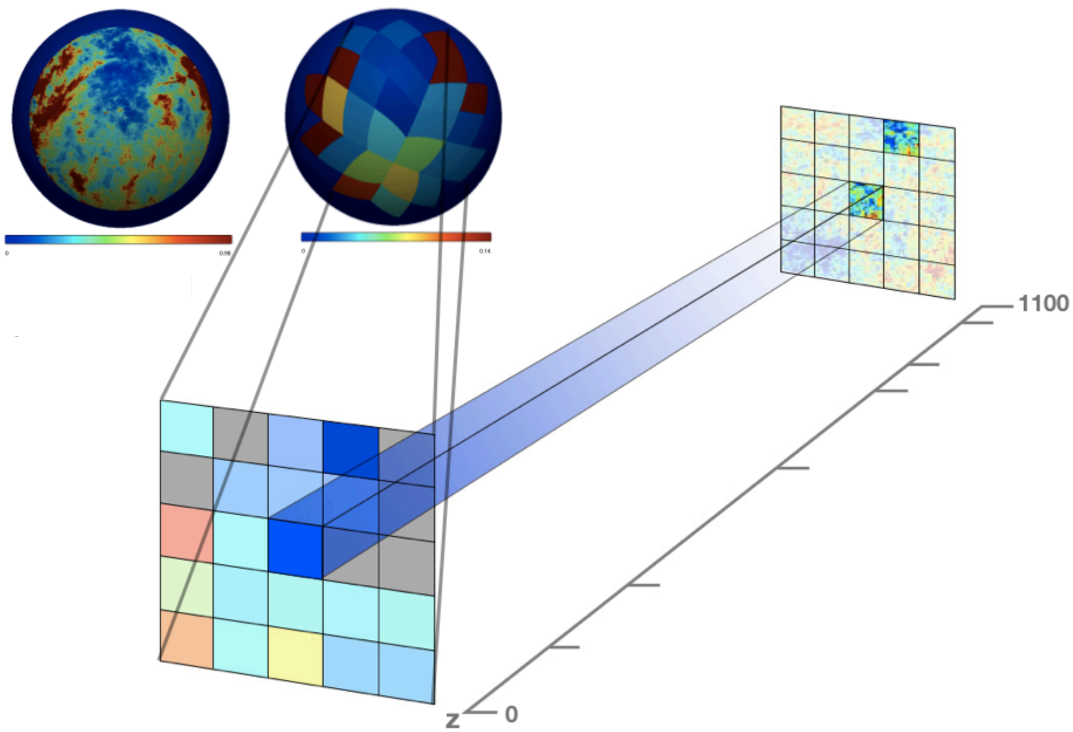

Optimal Patch

Conclusion


21-cm
CMB B-modes
Transient searches
Bandit Strategies
Deep-Field Imaging
......
GW Background

Thank you!

Ely D. Kovetz
University of Texas at Austin
LBNL, journal club, Oct. 18th, 2013



Based on: arXiv:1308.1404
and current work
with M. Kamionkowski (JHU)