DARK ENERGY SPECTROSCOPIC INSTRUMENT
U.S. Department of Energy Office of Science

IFAE
Institut de Física d'Altes Energies

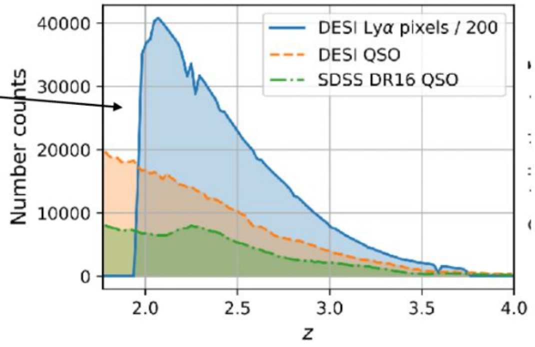
Improving the accuracy of small-scale Lyman-α forest measurements

Martine Lokken
In collaboration with Andreu Font Ribera

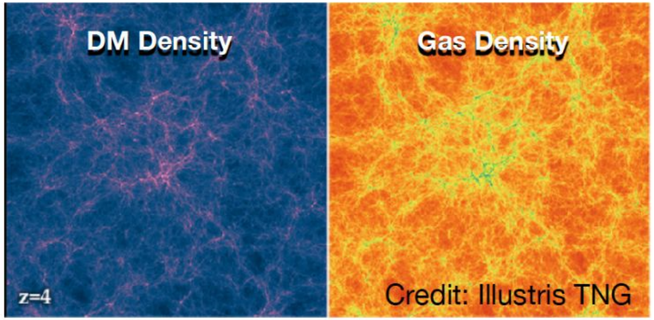
BCN-MAD meeting, winter 2025, CIEMAT

The Lyman-α forest tells us about early-universe dark matter structure and gas physics

- Ground-based observations (i.e., DESI): z~2-4
- Flux decrements probe low-to-intermediate densities (saturating at δ~10)
- At scales larger than ~1 Mpc, gas traces the dark matter (pressure support at smaller scales)



BAO results paper, DESI Collaboration 2024

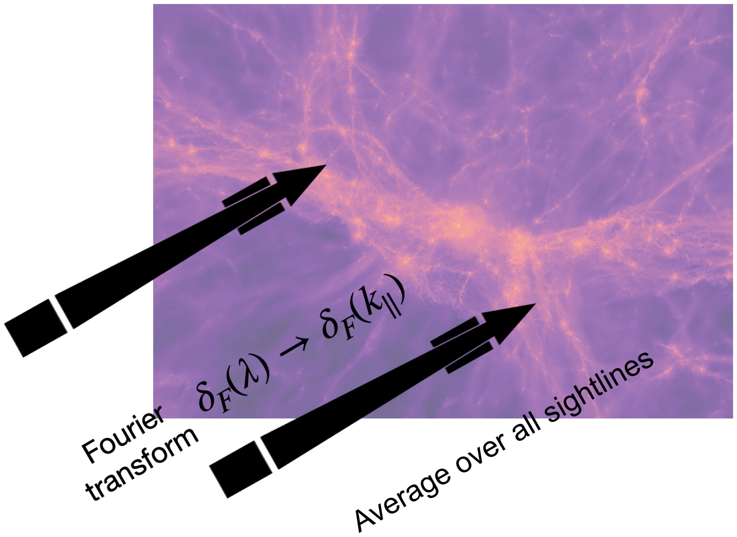


Key goal: robust measurement of matter power spectrum amplitude and slope → constrain dark matter models, neutrino mass, Ω_M

One-dimensional power spectrum (P_{1D})

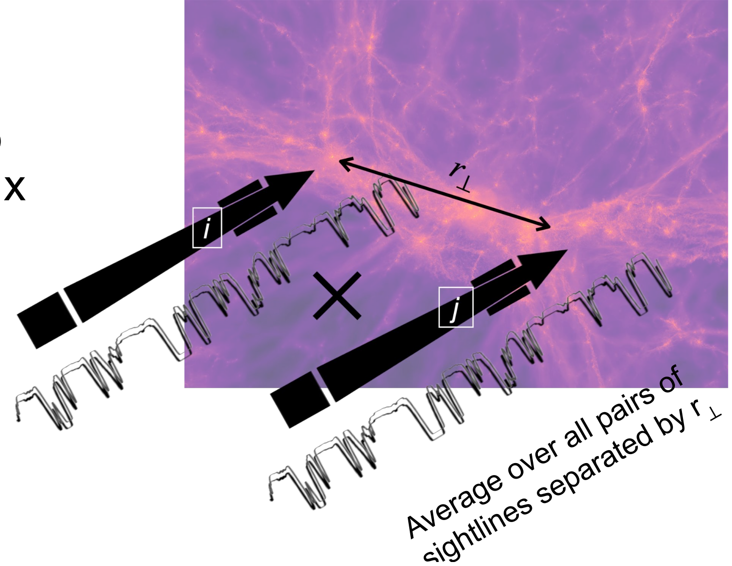
Currently two approaches:

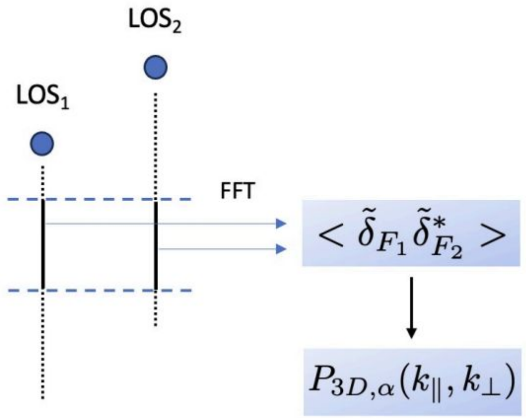
- FFT estimator
  - **Masked pixels** impact the power spectrum, effects are incompletely modeled
  - Much faster
- Optimal quadratic estimator
  - Masked pixels are fully accounted for
  - Pixels also inverse-variance-weighted
  - Much slower



Incorporating 3D information: P_x

- Same two approaches possible (FFT & quadratic)
- Scales with $N^2 \rightarrow$ quadratic estimator may be prohibitively slow
- Different quasar redshifts add complication
- Current FFT approach
  - Successful first measurements, yet…
  - Cutting large fraction of data
  - **Modeling of mask is incomplete**

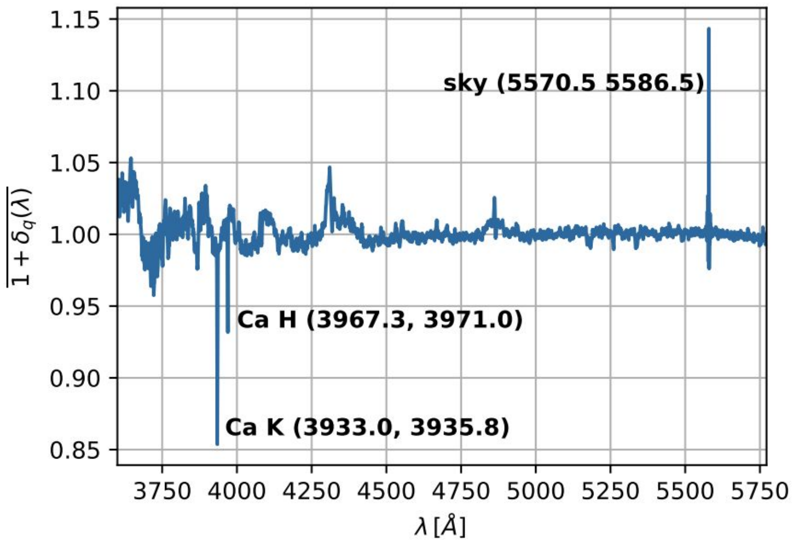




Abdul-Karim+ 2024

Mask Modeling: Motivation

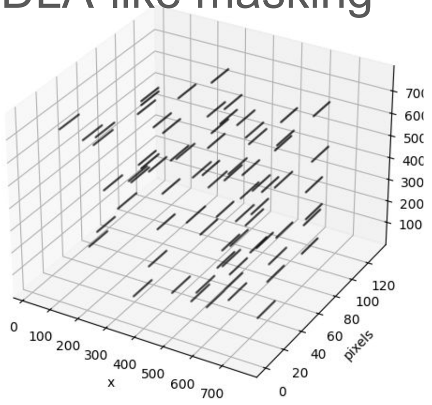
- Many reasons to mask pixels:
  - Damped line absorbers (DLAs)
  - Missing / corrupted data
  - Emission/absorption lines from the atmosphere and galaxy
- In practice, masking = setting pixel to 0
- P1D or Px via FFT estimator → FFT sees this as true signal variation → biases the signal
- Current approach calculates scale-dependent multiplicative correction from mocks
- Challenges with this approach:
  - Difficult to perfectly mimic the masking used in data
  - Some features in the data will not be in the mocks; these can be impacted (e.g. smoothed) by the mask



Spectrum including a sky line and Galactic Calcium lines (Ramirez-Perez 2023)

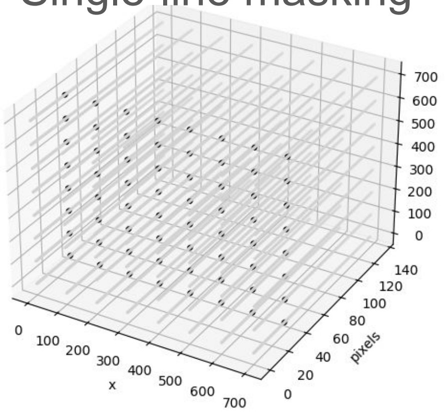
Incorporating masking into mock spectra from MP-Gadget sims*

DLA-like masking

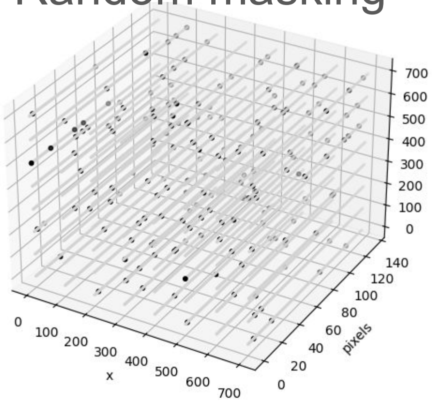


Small DLA width: 20 pixels = 5 Mpc

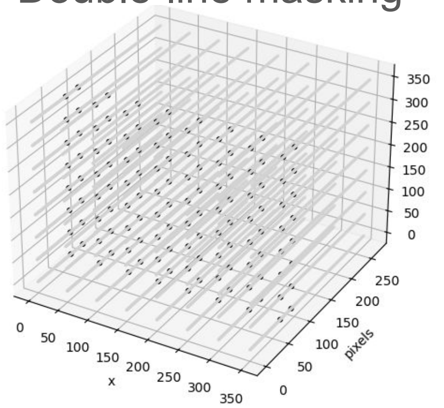
Single-line masking

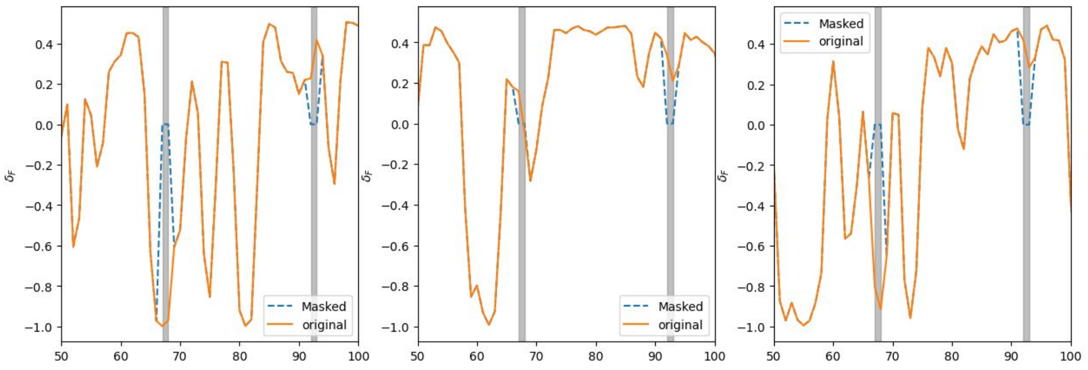


Random masking



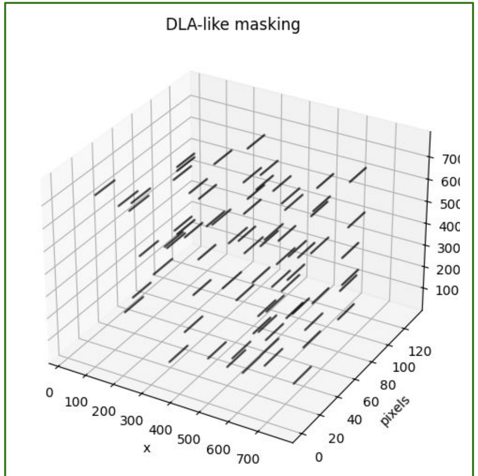
Double-line masking



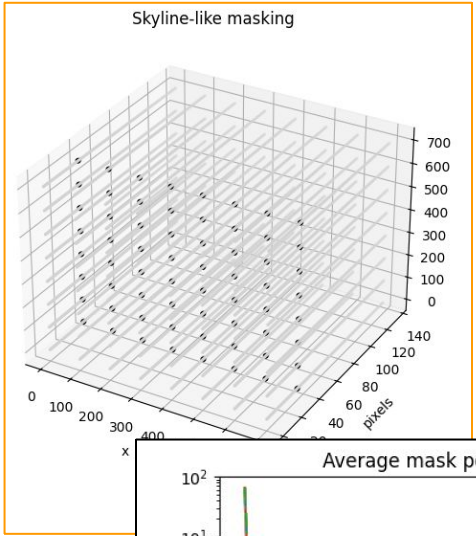


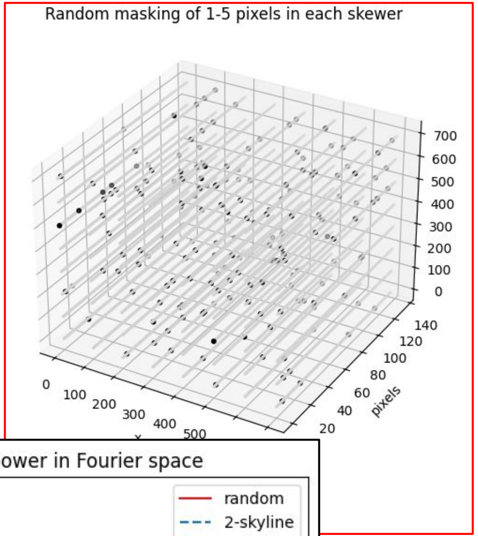
*Pederson+ 2021

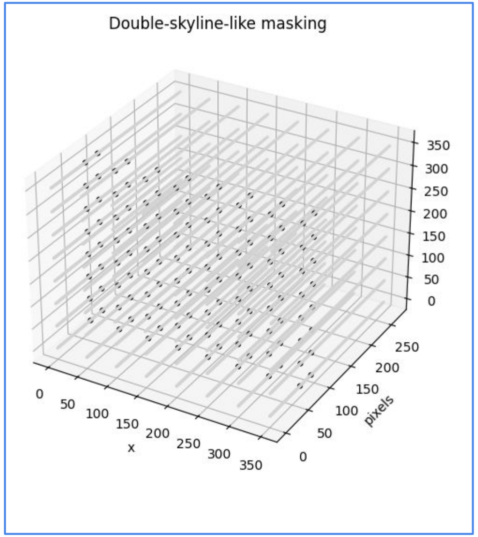
Incorporating masking into mock spectra from MP-Gadget sims*



DLA width: 20 pixels = 5 Mpc

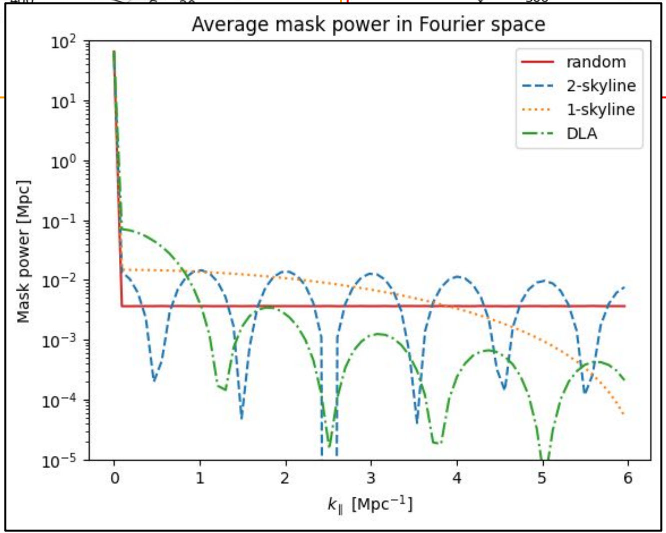






Average power of mask over all lines-of-sight:

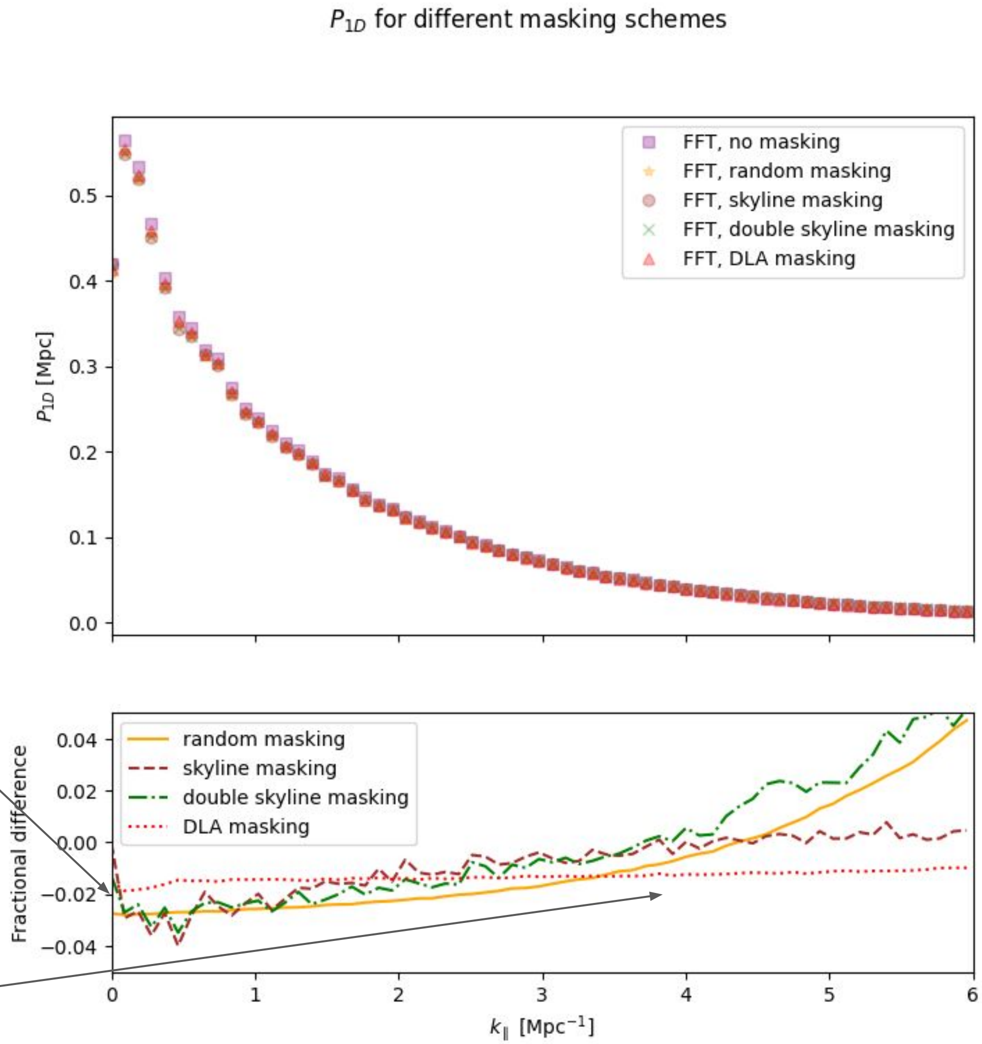
$$\langle |FFT(mask)|^2 \rangle$$



Effects on P1D

Keeping total number of masked pixels equal → similar low-k deficit

Different scale-dependent behavior



Modeling the mask effects – Methods

Continuous case:

As done in other fields (e.g. CMB), predict the masked signal by convolving Fourier-transformed weights with the theory P1D:

$$\left\langle |f_m|^2 \right\rangle = \int \frac{dk}{2\pi} |w(k_m - k)|^2 R^2(k) P_{1\mathrm{D}}(k)$$

- $\left\langle |f_m|^2 \right\rangle$: Expected value of variance of the Fourier modes
- $|w(k_m - k)|^2$: Weights/mask* in Fourier space
- $R^2(k)$: Resolution
- $P_{1\mathrm{D}}(k)$: True P1D

*Masking is a sub-case of weighting:

- In masking, weight = 1 or 0
- Can optimize the FFT estimators by noise-weighting pixels, and use this approach to model the impact

In practice, on the hydro boxes:

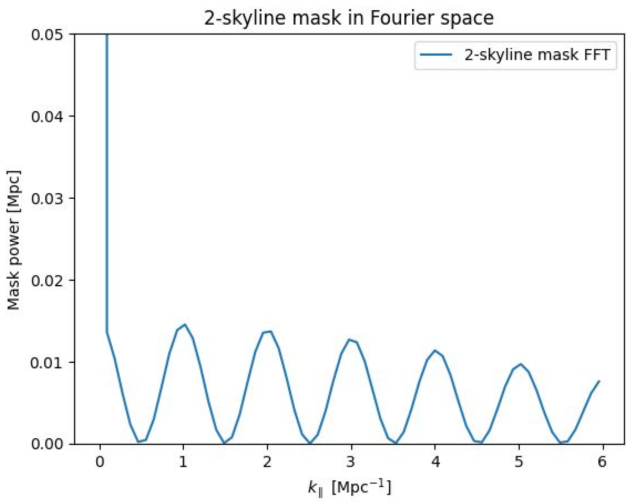
Average over the weight/mask FFT for many different spectra (index *q*)

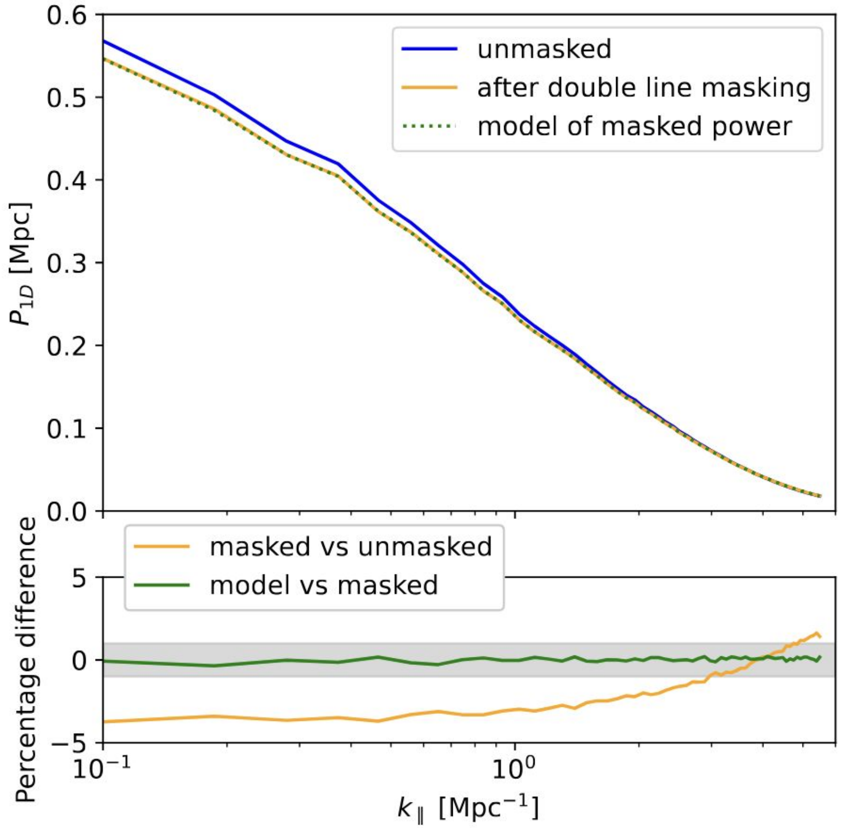
$$W_m = A \sum_q \frac{1}{N} |w_m^q|^2$$

$$\left\langle \hat{\Theta}_m \right\rangle = \sum_n W_{m-n} P_{1\mathrm{D}}(k_n)$$

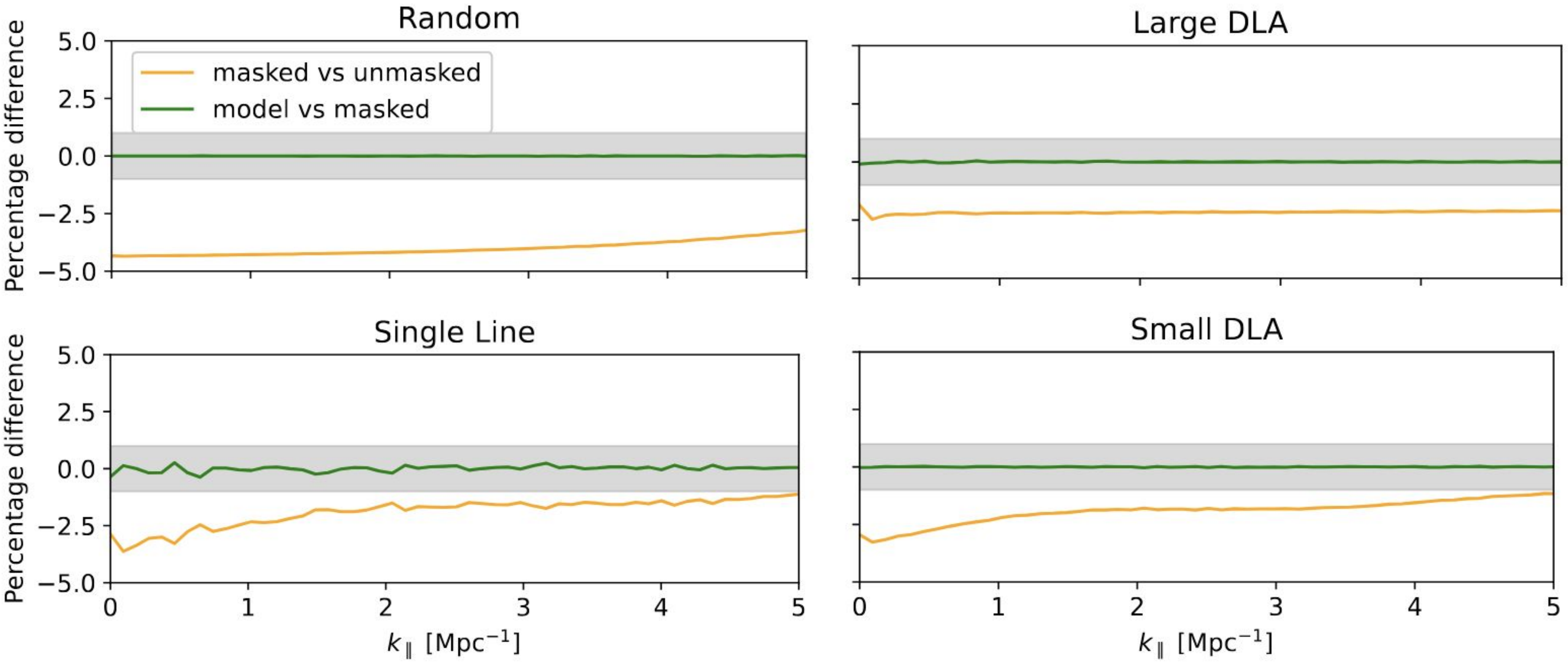
Calculate the **expected value** for the masked P1D by convolving *W* with the **theory for true P1D**

Modeling the masks – Results (double-line case)



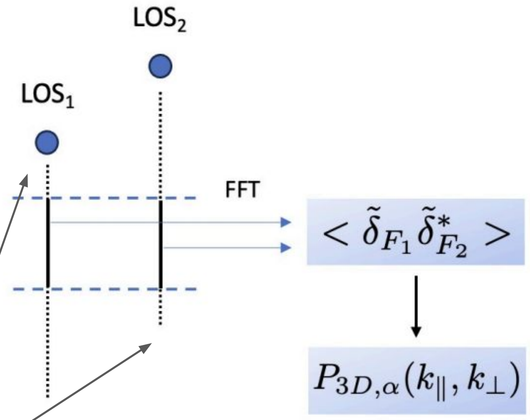


Modeling the masks – Results (all)



Zero-padding spectra

- Makes spectra periodic at the boundaries (zeros on either side) → better for FFT
- Preserves more of the data
- We can model the padding exactly, as it acts just like a mask



Add long list of consecutive zeros to spectra ends

Conclusions and Future Work

Modeling the exact mask impacts through convolution removes the mask bias

Modeling can also be used for non-binary weights, allows for optimization of FFT estimators

Same formalism works for P_x and would allow us to use all the data, overcoming past limitations

Next steps:

- incorporate into DESI inference pipelines for P_{1D} and P_x:

  → convolve the theory predictions with exact data mask, before comparing with data

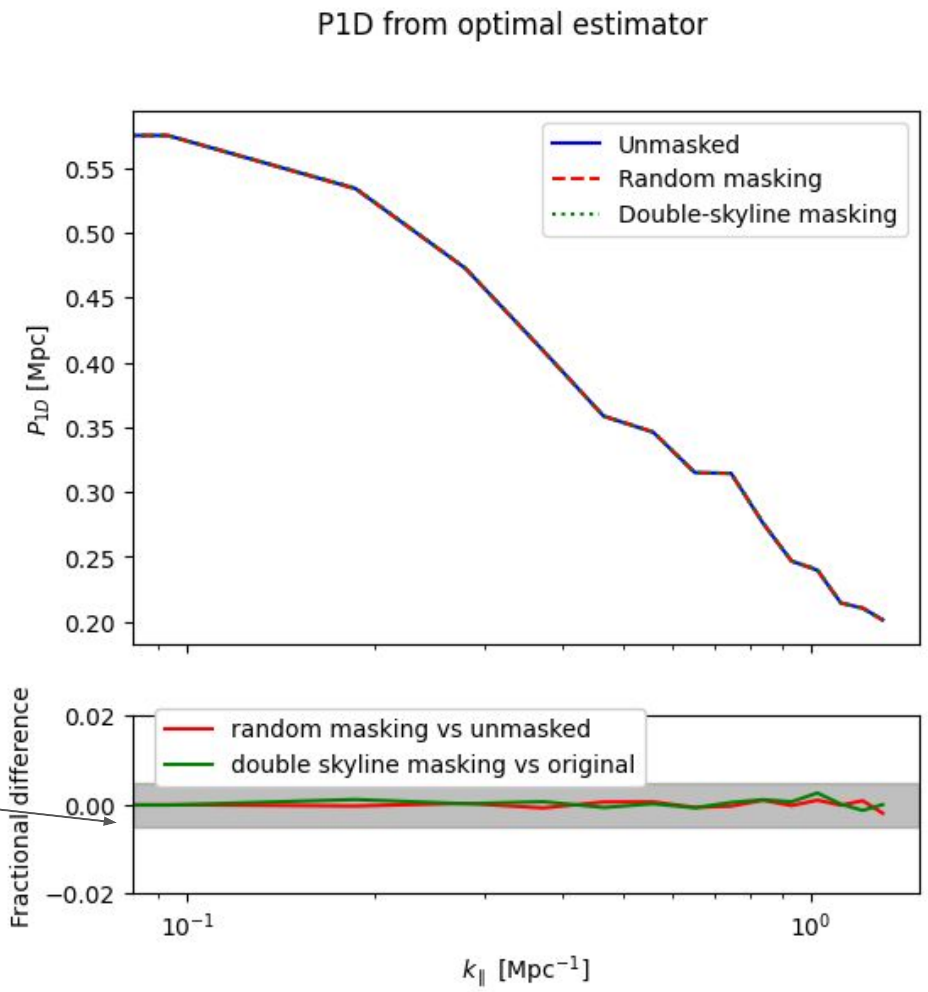
- work on modeling of metal contaminants and other systematics

Extra slides

Optimal quadratic estimator is unbiased when you assign the masked pixels very high noise…

…but is much more computationally expensive

Differences within 1%



Optimal quadratic estimator for Px (proof-of-concept)

- Narrow bin in rperp
- tiny fraction of box
- Bottleneck is speed