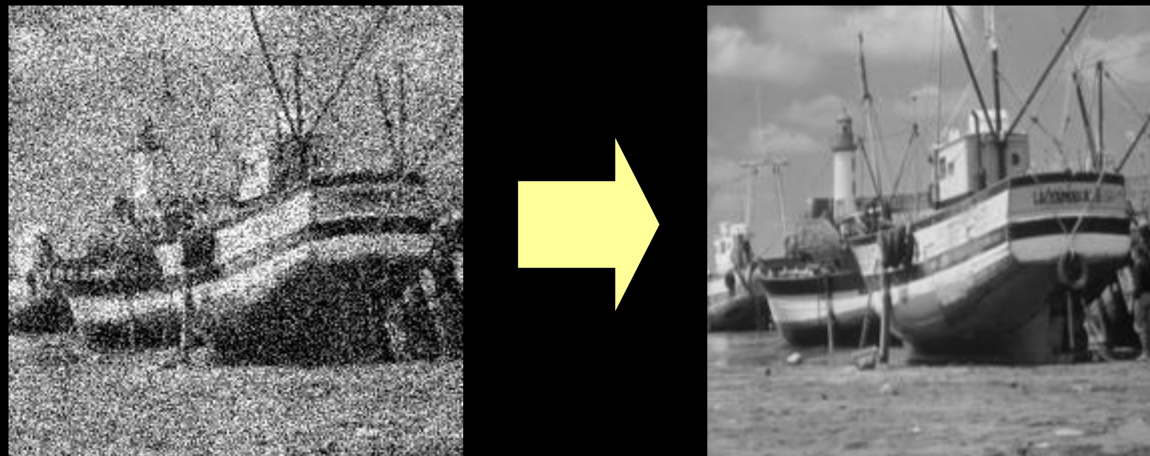
Patch Complexity, Finite Pixel Correlations and Optimal Denoising

Anat Levin, Boaz Nadler,
Fredo Durand and Bill Freeman

Weizmann Institute, MIT CSAIL

Image denoising

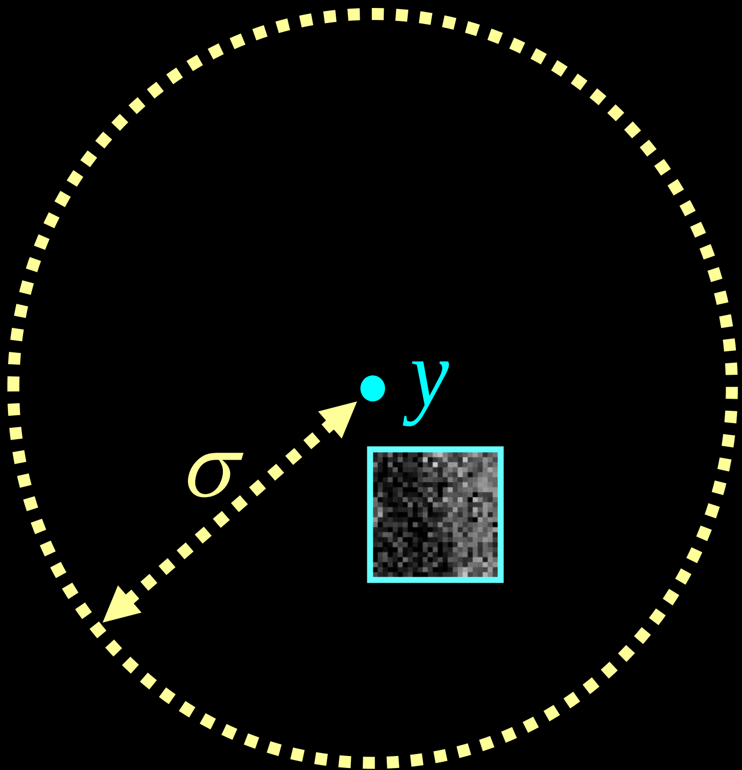


Many research efforts invested, and results harder and harder to improve: reaching saturation?

- **What uncertainty is inherent in the problem?**
- **How further can we improve results?**

Denoising Uncertainty

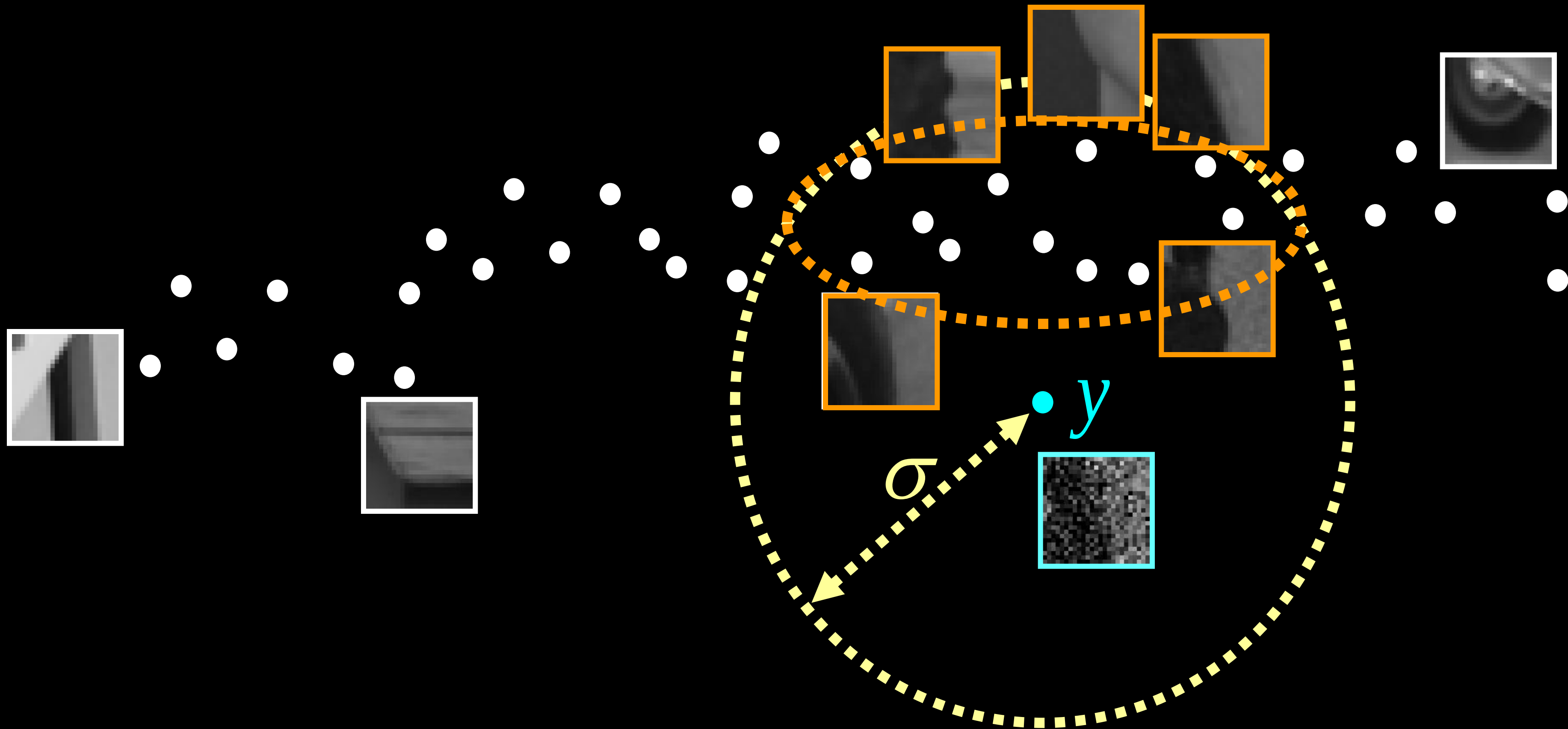
What is the volume of all clean x images that can explain a noisy image y?



Denoising Uncertainty

What is the volume of all clean images x that can explain a noisy image y?

Multiple clean images within noise level.



Denoising limits- prior work

- **Signal processing assumptions (Wiener filter, Gaussian priors)**

- **Limits on super resolution- numerical arguments, no prior** *[Baker&Kanade 02]*

- **Sharp bounds for perfectly piecewise constant images** *[Korostelev&Tsybakov 93, Polzehl&Spokoiny 03]*

- **Non-local means- asymptotically optimal for infinitely large images. No analysis of finite size images.** *[Buades,Coll&Morel. 05]*

- **Natural image denoising limits, but many assumptions which may not hold in practice and affect conclusions.** *[Chatterjee and Milanfar 10]*

MMSE denoising bounds

$$\mathrm{MMSE} = \int p(y)\, V_{x|y} = \int p(y) \int p(x \mid y)(x_c - \mu(y))^2\, dx dy$$

MMSE= conditional variance, achieved by the conditional mean

MMSE with the exact p(x) (and not with heuristics used in practice), is the *optimal possible denoising. By definition.*

Using internal image statistics or class specific information might provide practical benefits, but cannot perform better than the MMSE. *By definition!*

MMSE with a finite support

$$\text{MMSE}_{\text{d}} = \int p(y_{w_d})\, V_{x_{w_d}|y_{w_d}}$$
$$= \int p(y_{w_d}) \int p(x_{w_d} \mid y_{w_d})(x_c - \mu_d(y))^2\, dx dy$$

MMSE_{d} best possible result of any algorithm which can utilize a d=k x k window w_d around a pixel of interest

e.g. spatial kernel size in bilateral filter,
patch size in non-parametric methods

Non Local Means: effective support = entire image

Estimating denoising bounds in practice

$$\mathrm{MMSE} = \int p(y) \int p(x \mid y)(x_c - \mu(y))^2 \, dx dy$$

Challenge: Compute MMSE without knowing p(x)?

The trick *[Levin&Nadler CVPR11]*:

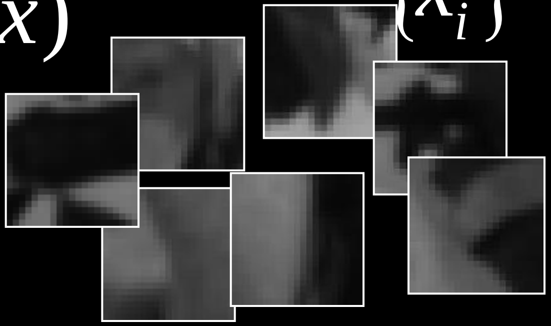
We don't know p(x) but we can sample from it

Evaluate MMSE non parametrically

Sample mean:

$$\{x_i\} \sim p(x)$$

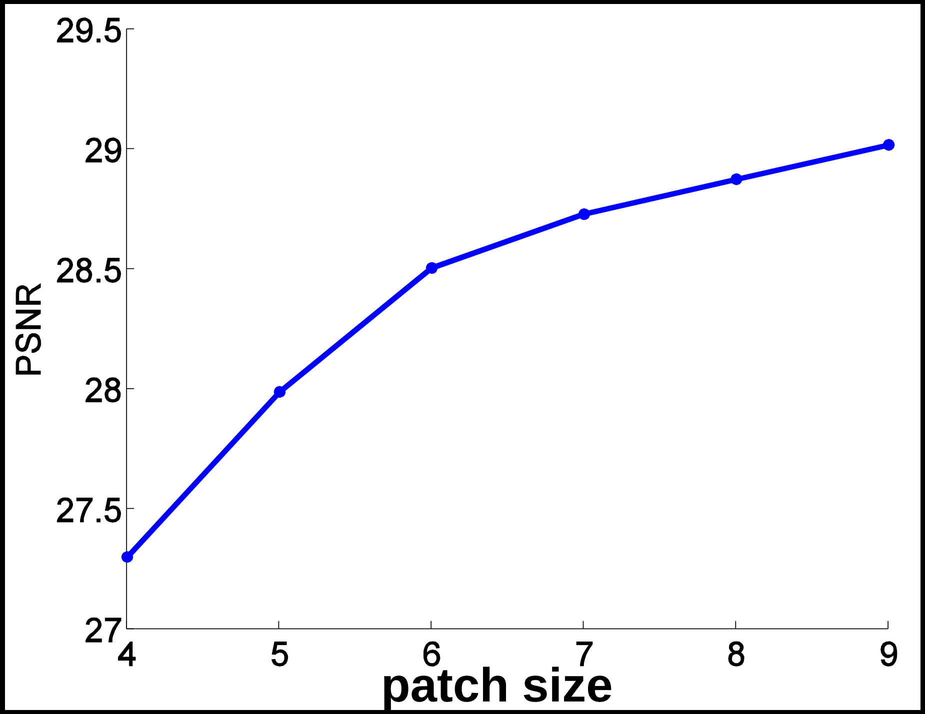
$$\{x_i\}$$



$$\hat{\mu}(y) = \frac{\frac{1}{N}\sum_i p(y \mid x_i) x_{i,c}}{\frac{1}{N}\sum_i p(y \mid x_i)}$$

MMSE as a function of patch size

$\sigma = 35$

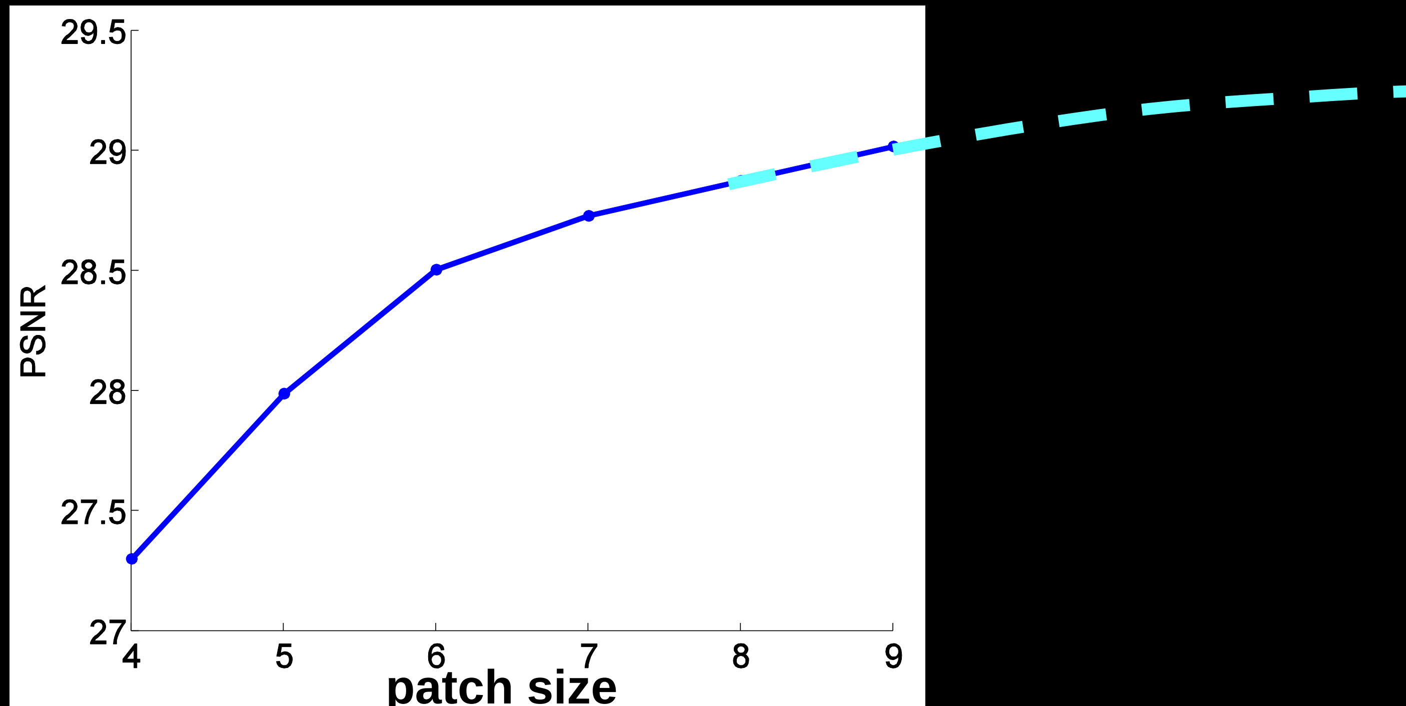


[Levin&Nadler CVPR11]**:**

For small patches/ large noise, non parametric approach can accurately estimate the MMSE.

MMSE as a function of patch size

$$\sigma = 35$$



How much better can we do by increasing window size?

Towards denoising bounds

Questions:

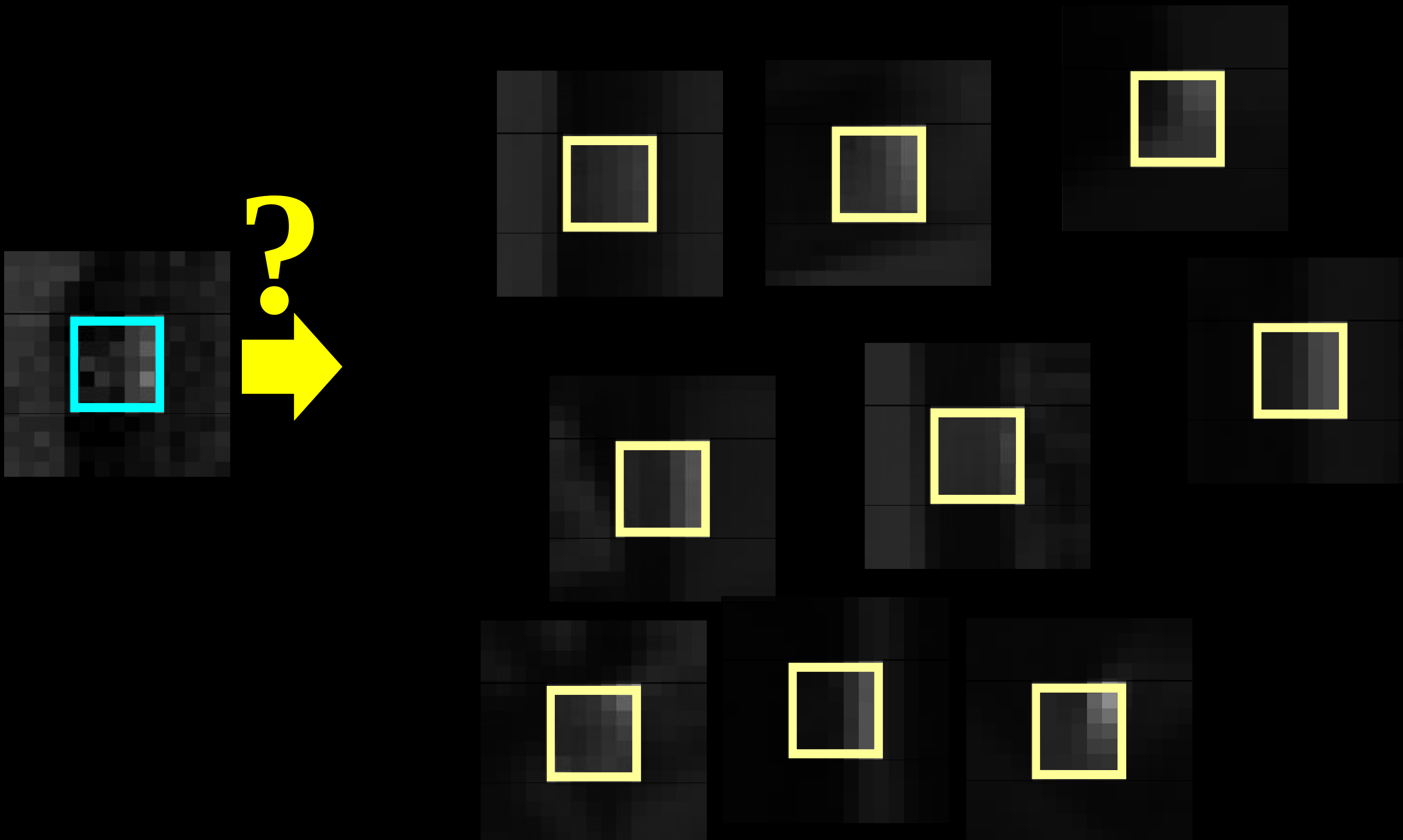
- ***For non-parametric methods:***

  How does the difficulty in finding nearest neighbors relates to the potential gain, and how can we make a better usage of a given database size?

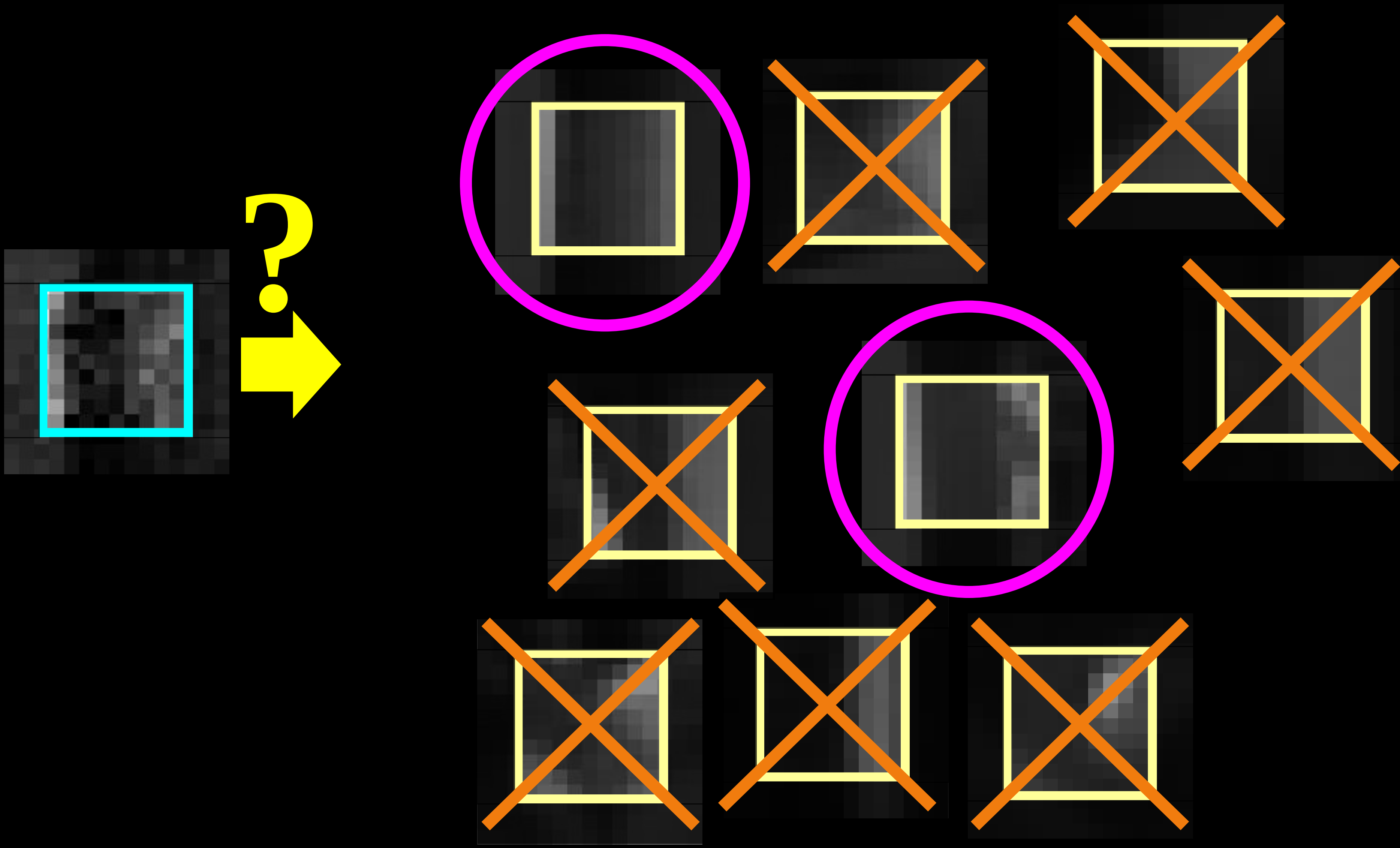
- ***For any possible method:***

  Computational issues aside, what is the optimal possible restoration? Can we achieve zero error?

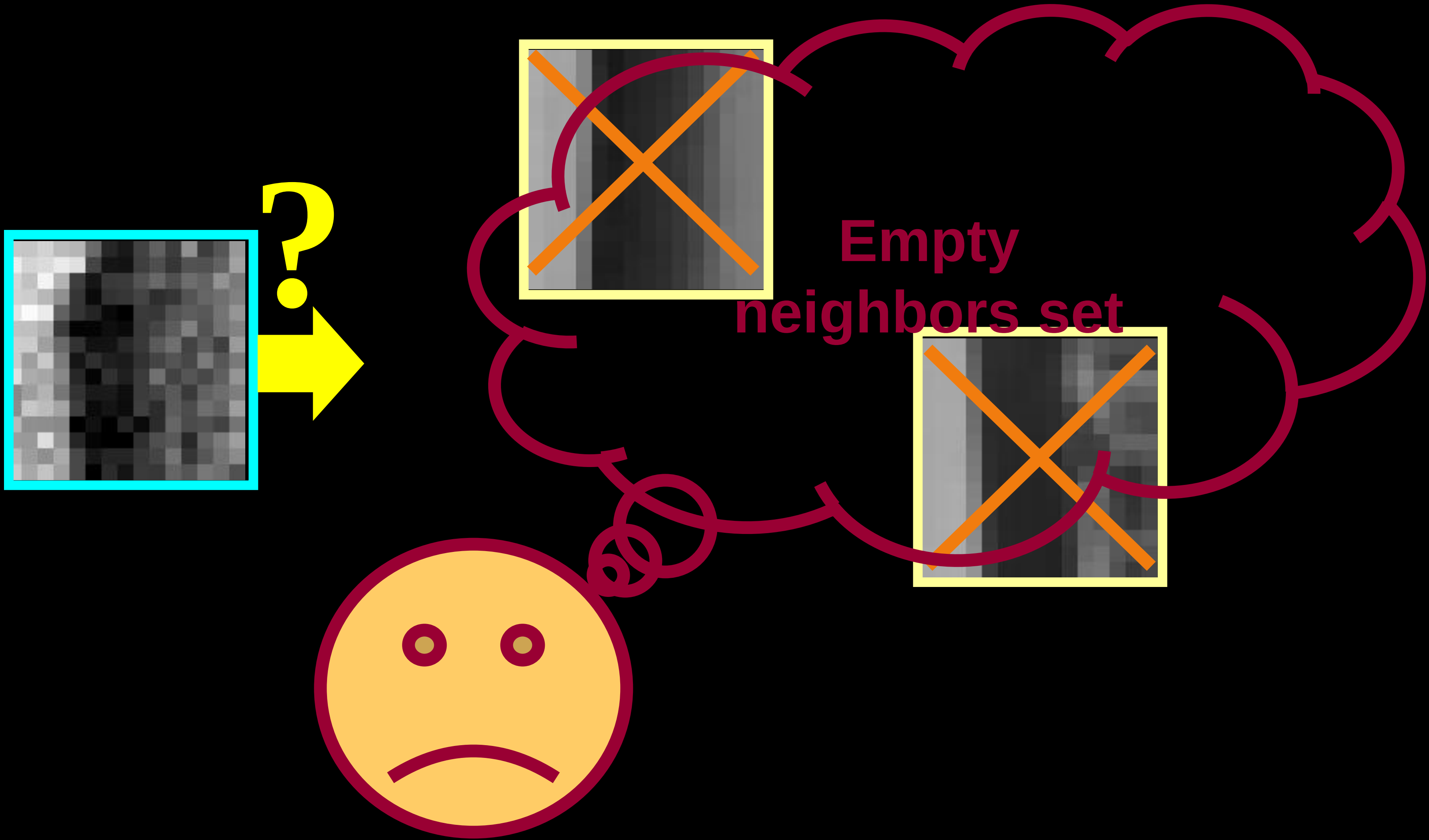
Patch Complexity



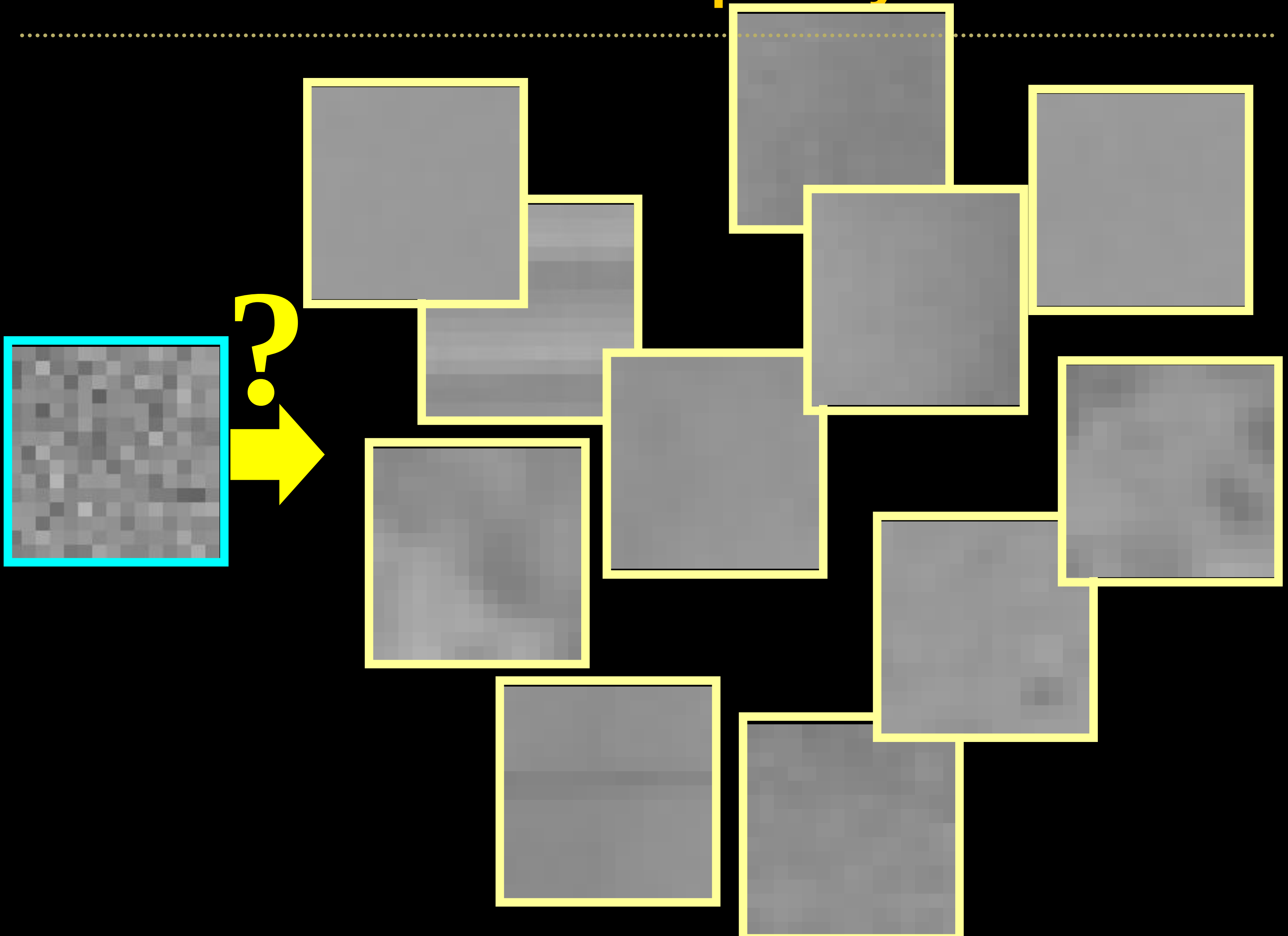
Patch Complexity



Patch Complexity



Patch Complexity



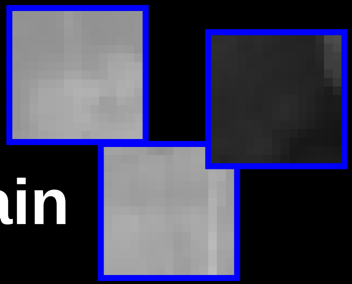
Patch complexity v.s. PSNR gain

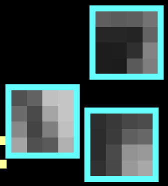
Law of diminishing return:

When an increase in patch width requires many more training samples, the performance gain is smaller.

Smooth regions:

***Easy* to increase support, *large* gain**





Textured regions:

***Hard* to increase support, *small* gain**

Adaptive patch size selection in denoising algorithms.

See paper

Towards denoising bounds

Questions:

- ***For non-parametric methods:***

  How does the difficulty in finding nearest neighbors relates to the potential gain, and how can we make a better usage of a given database size?

- ***For any possible method:***

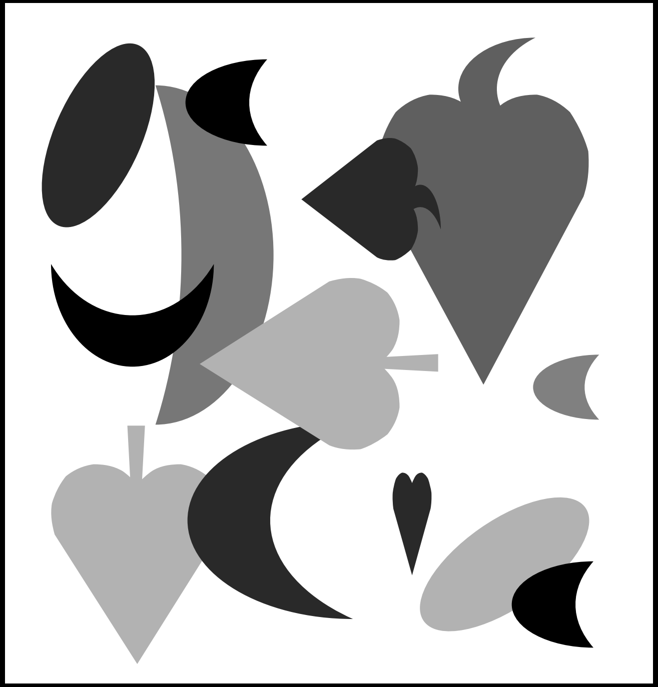
  Computational issues aside, what is the optimal possible restoration? Can we achieve zero error?

  - What is the convergence rate as a function of patch size?

The Dead Leaves model *[Matheron '68]*

Image = random collection of finite size piece-wise constant regions

Region intensity = random variable with uniform distribution

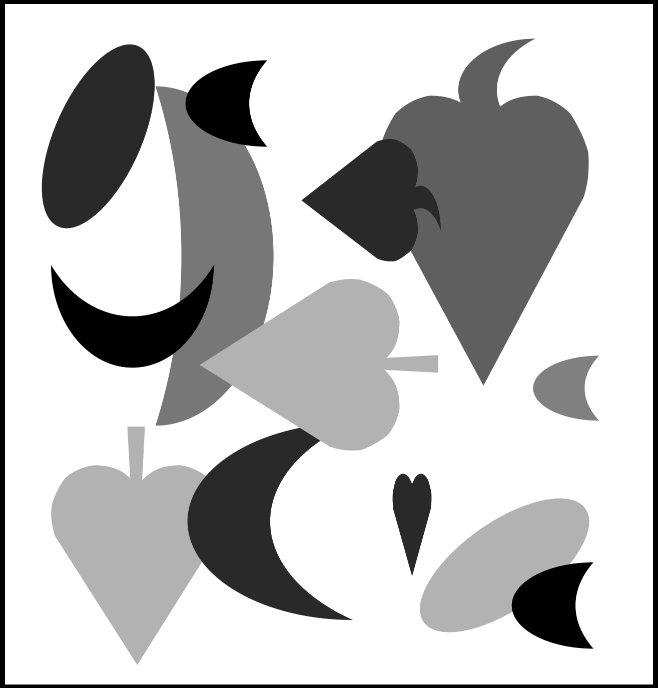


Best possible denoising: average all observations within a segment

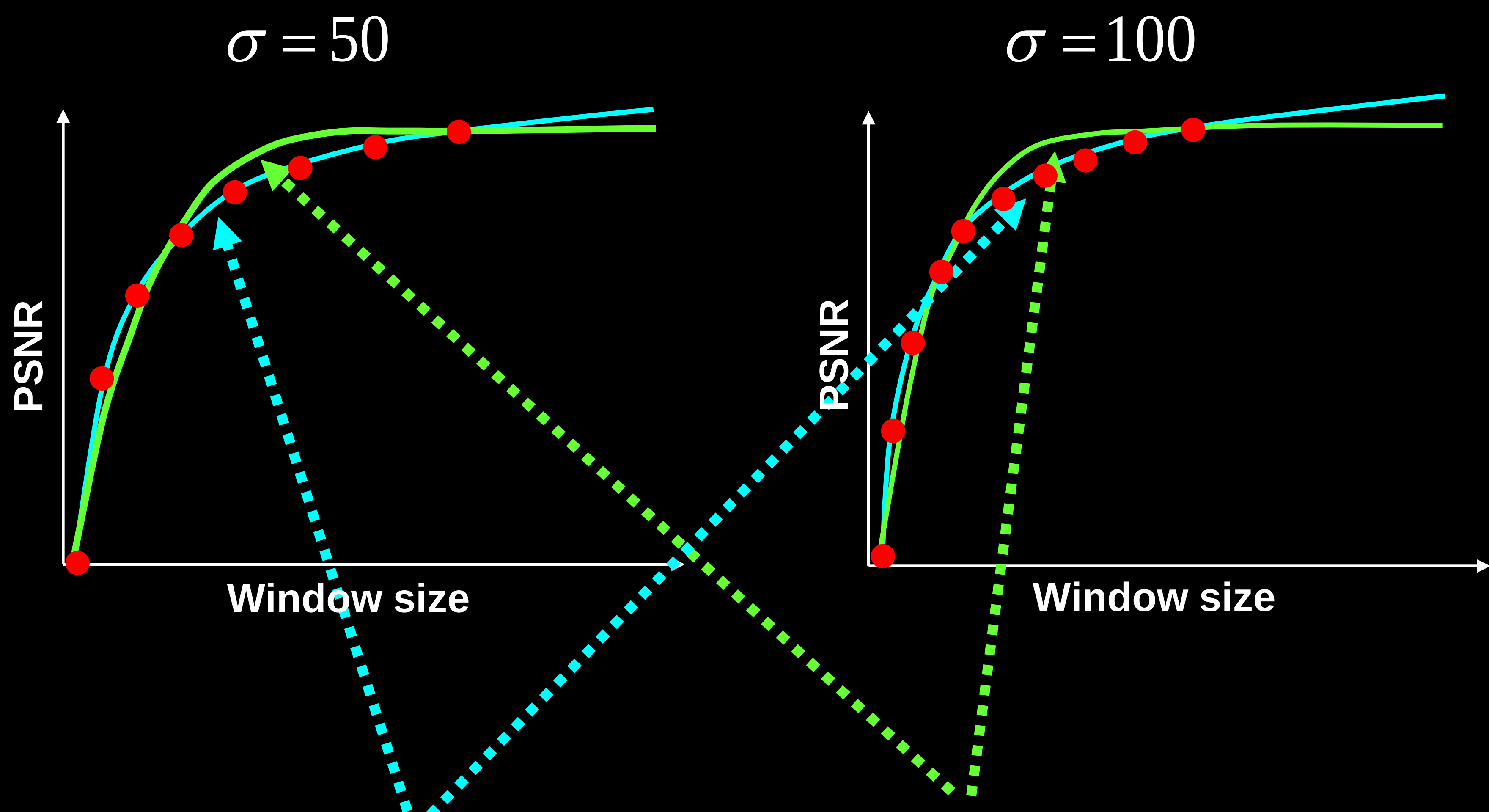
Optimal denoising in the Dead Leaves model

Scale invariance + dead leaves ⇒ **power law convergence**

$$\mathrm{MMSE}_d \approx \mathrm{MMSE}_\infty + \frac{c}{d}$$



Empirical PSNR v.s. window size



Good fit with a power law

$$\mathrm{MMSE}_d = \mathrm{e} + \frac{\mathrm{c}}{\mathrm{d}}$$

Poor fit with an exponential curve (implied by Markov models)

$$\mathrm{MMSE}_d = \mathrm{e} + \mathrm{c}\mathrm{r}^{\mathrm{d}}$$

Extrapolating optimal PSNR

$$\text{MMSE}_d \approx \text{MMSE}_\infty + \frac{c}{d}$$

σ	35	50	75	100
Extrapolated bound	**30.6**	**28.8**	**27.3**	**26.3**
KSVD	28.7	26.9	25.0	23.7
BM3D	30.0	28.1	26.3	25.0
EPLL	29.8	28.1	26.3	25.1

Future sophisticated denoising algorithms appear to have modest room for improvement: ~ 0.6-1.2dB

Summary: inherent uncertainty of denoising

Non-parametric methods: Law of diminishing return

- When increasing patch size requires a significant increase in training data, the gain is low
- Correlation with new pixels makes it easier to find samples AND makes them more useful
- Adaptive denoising

For any method:

Optimal denoising as a function of window size follows a power law convergence

- Scale invariance, dead leaves
- Extrapolation predicts denoising bounds