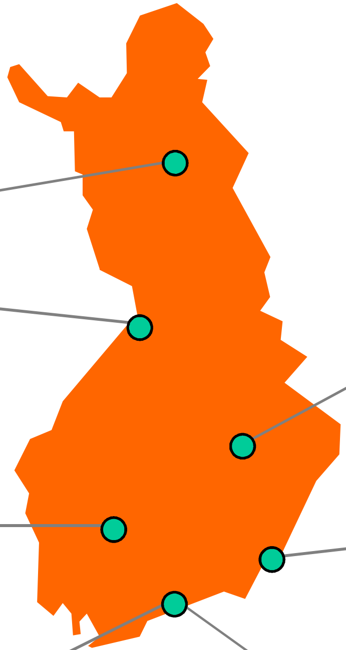
# Discretization-invariant Bayesian inversion and Besov space priors


Samuli Siltanen
Tampere University of Technology


RICAM 28.10.2008

# Finnish Centre of Excellence in Inverse Problems Research



OULUN YLIOPISTO UNIVERSITY of OULU

UNIVERSITY OF KUOPIO

TAMPERE UNIVERSITY OF TECHNOLOGY

LAPPEENRANTA UNIVERSITY OF TECHNOLOGY

TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF HELSINKI
Rolf Nevanlinna Institute
Research institute of mathematics, computer science and statistics

http://math.tkk.fi/inverse-coe/

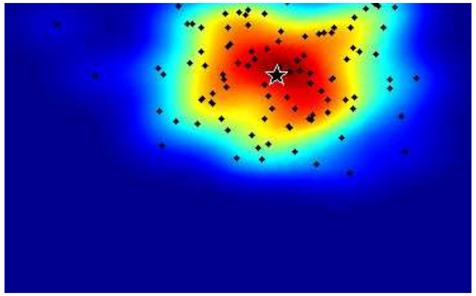
# This is a joint work with

**Matti Lassas**

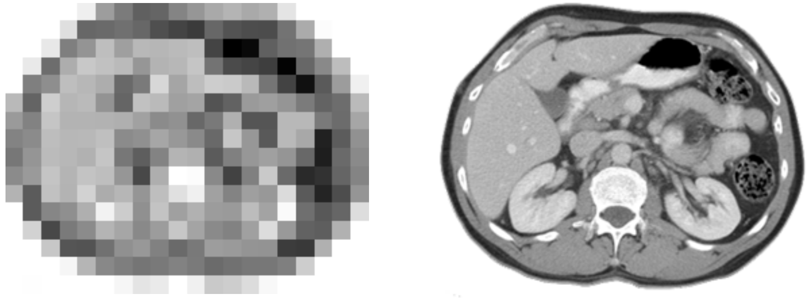
Helsinki University of Technology
Finland

**Eero Saksman**
University of Helsinki
Finland

This research work has been supported by
**Finnish Technology Agency, Academy of Finland, Instrumentarium Imaging, GE Healthcare, and Palodex Group**



1. Bayesian inversion



2. Discretization-invariance

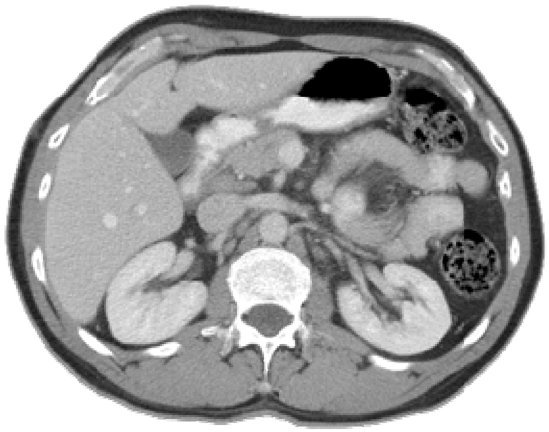
3. Regularization results


4. Besov space priors

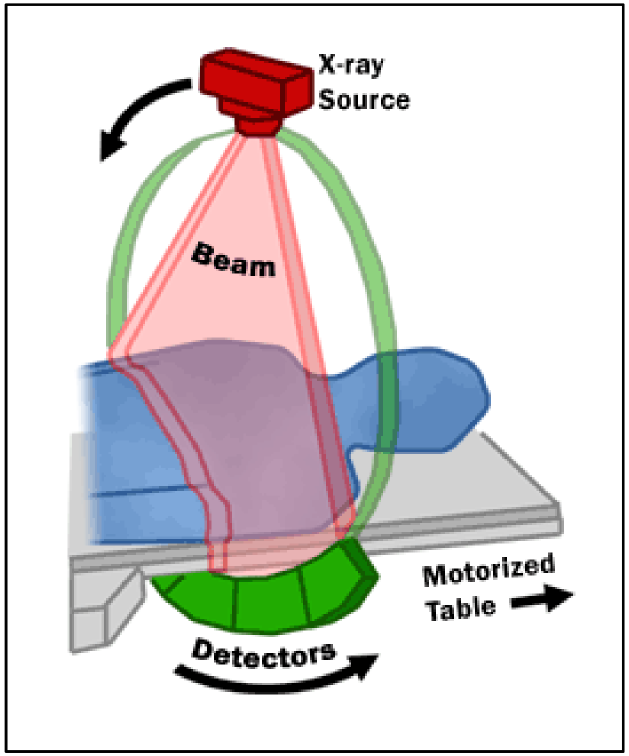
$$e^{-\alpha\|u\|_{B^1_{11}}}$$

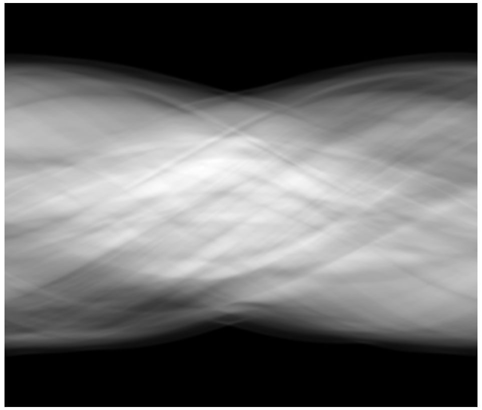
# We discuss indirect measurements of continuum quantities

Example: Radon transform in X-ray tomography



*U*
(attenuation coefficient)



*M*
(noisy sinogram)

$$M = AU + \mathcal{E}$$

# The continuum measurement model must be discretized for practical inversion

We simulate physics by the **continuum model**

$$M = AU + \mathcal{E},$$

where $\mathcal{E}$ is white noise. Here $M(\,\cdot\,,\omega)$ and $U(\,\cdot\,,\omega)$ are random functions defined on $\mathbb{R}^d$.

Data is a realization $M_k(\omega_0) \in \mathbb{R}^k$ described by the **practical measurement model**

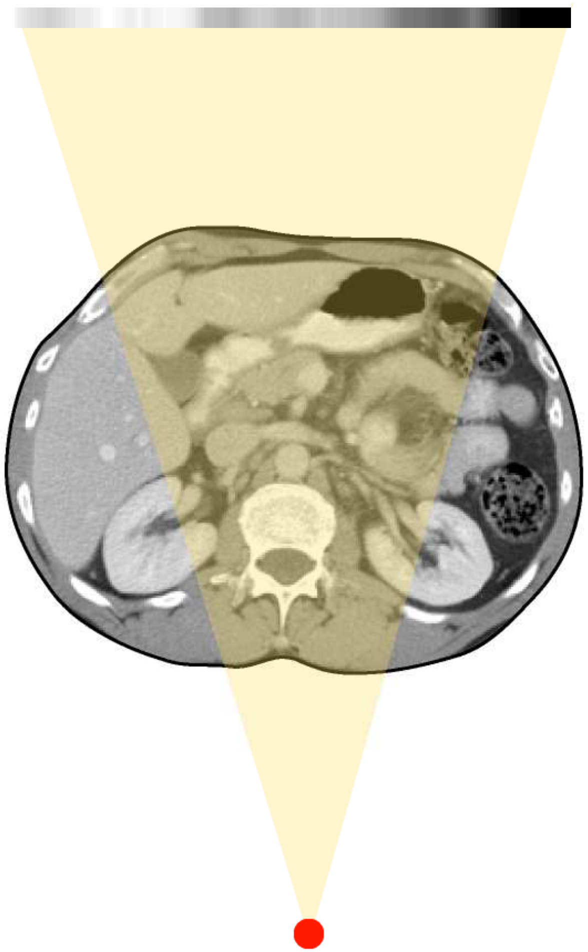
$$M_k = P_k AU + P_k \mathcal{E}.$$

Numerical work needs the **computational model**

$$M_{kn} = P_k AU_n + P_k \mathcal{E},$$

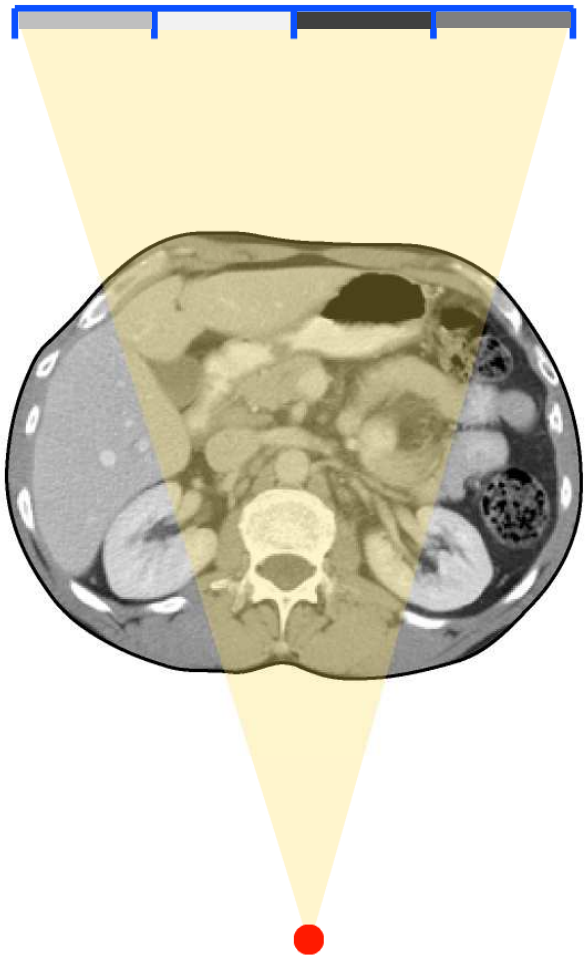
where $U_n = T_n U$ is a discretization of $U$.

# Continuum model for tomography:



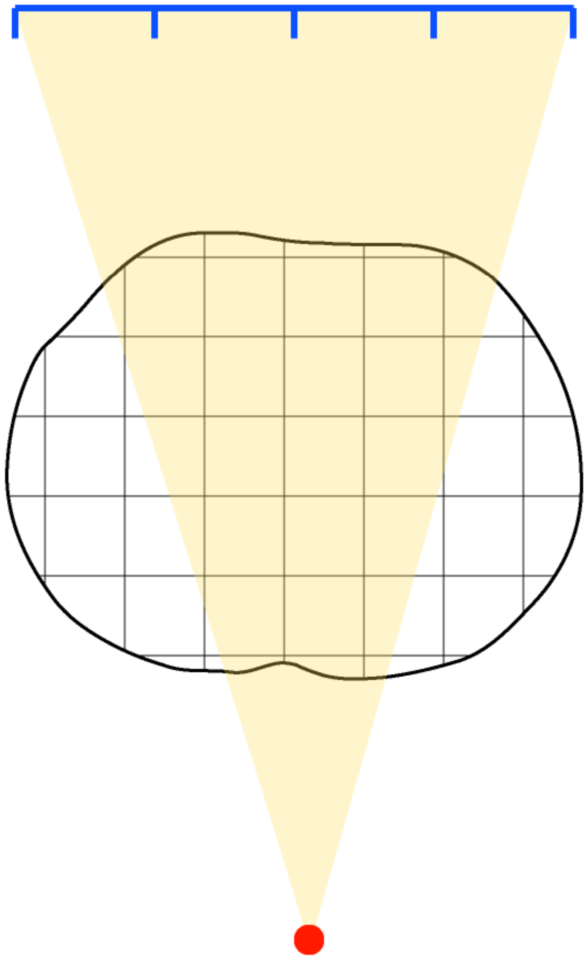
$$M = AU + \mathcal{E}$$

# Practical measurement model:



$$M_k = A_k U + \mathcal{E}_k = P_k A U + P_k \mathcal{E}$$

# Computational model:



$$M_{kn} = P_k A U_n + \mathcal{E}_k$$

# The finite dimensions n and k are independent

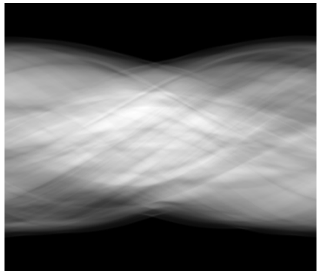
The linear projection operator $P_k$ is related to the measurement device and has $k$-dimensional range.

The linear projection operator $T_n$ is related to the computational discretization of the unknown and has $n$-dimensional range.

# Inverse problem: given a measurement, estimate the quantity U

X-ray tomography: Given measurement

$$M_k(\omega_0) = $$



,

estimate X-ray coefficient U in a two-dimensional slice.

We consider Bayesian estimation of U using a discrete conditional mean estimate (and confidence intervals).

# Bayes formula combines measured and *a priori* information

The posterior distribution corresponding to the finite-dimensional computational model is

$$\pi_{kn}(u_n \,|\, m_{kn}) \sim \Pi_n(u_n)\ \exp(-\frac{1}{2}\|m_{kn} - P_k A u_n\|_2^2),$$

where the prior distribution $\Pi_n$ assigns high probability to functions $u_n$ that are expected in light of *a priori* information, and the likelihood distribution

$$\exp(-\frac{1}{2}\|m_{kn} - P_k A u_n\|_2^2)$$

measures data misfit.

# In this work we estimate U by a discrete conditional mean estimate

Posterior distribution from computational model:

$$\pi_{kn}(u_n \,|\, m_{kn}) = \frac{\Pi_n(u_n)\ \mathrm{exp}(-\frac{1}{2}\|m_{kn} - P_k A u_n\|_2^2)}{\Upsilon_{kn}(m_{kn})}.$$

Conditional mean estimate is defined by

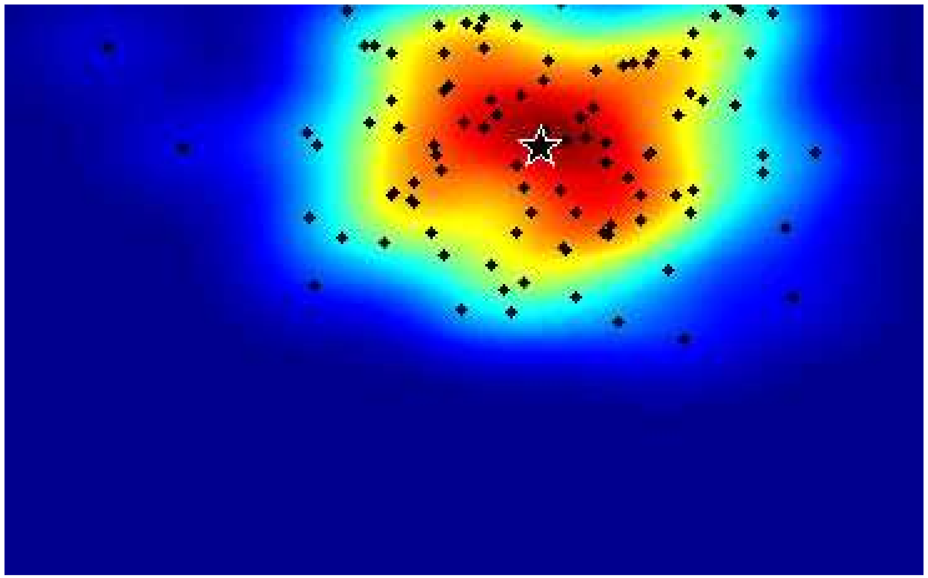
$$\mathbf{u}_{kn}^{\mathrm{CM}} := \int_{\mathbb{R}^n} u_n \pi_{kn}(u_n \,|\, m_k)\, du_n,$$

where the realization $m_k := M_k(\omega_0)$ is data from the practical measurement model.

# The conditional mean estimate is often computed by Markov chain Monte Carlo

$$\mathbf{u}_{kn}^{\mathrm{CM}} = \int_{\mathbb{R}^n} u_n \pi_{kn}(u_n \,|\, m_k)\, du_n \approx \frac{1}{N}\sum_{j=1}^{N} u_n^{(j)}$$

The samples $u_n^{(j)}$ are drawn from the posterior distribution using e.g. Metropolis-Hastings method or Gibbs sampler.



# Conditional mean estimates have been applied to various inverse problems

**Image restoration**

1991 Besag, York & Mollié

**Geological prospecting**

1998 Nicholls

2003 Andersen, Brooks & Hansen

**Atmospheric and ionospheric remote sensing**

1995 Markkanen *et al.*

1997 Nygrén, Markkanen, Lehtinen

1999 D´Ambrogi, Mäenpää & Markkanen

1999 Tamminen

2004 Haario, Laine, Lehtinen, Saksman & Tamminen

**Medical X-ray tomography**

1997 Battle, Cunningham & Hanson

2003 Kolehmainen *et al.*

2003 Siltanen *et al.*

**Electrical impedance imaging**

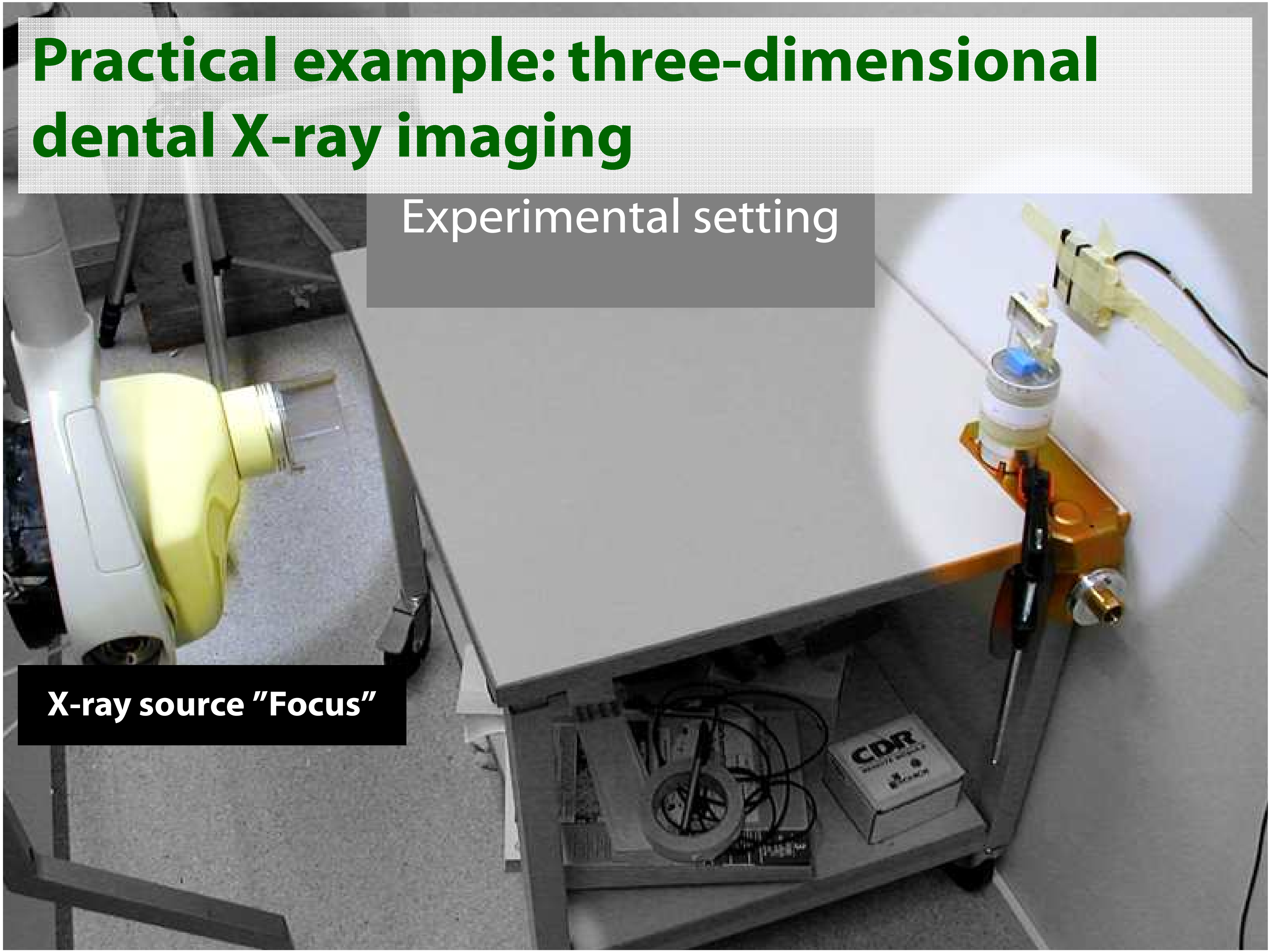
2000 Kaipio, Kolehmainen, Somersalo & Vauhkonen

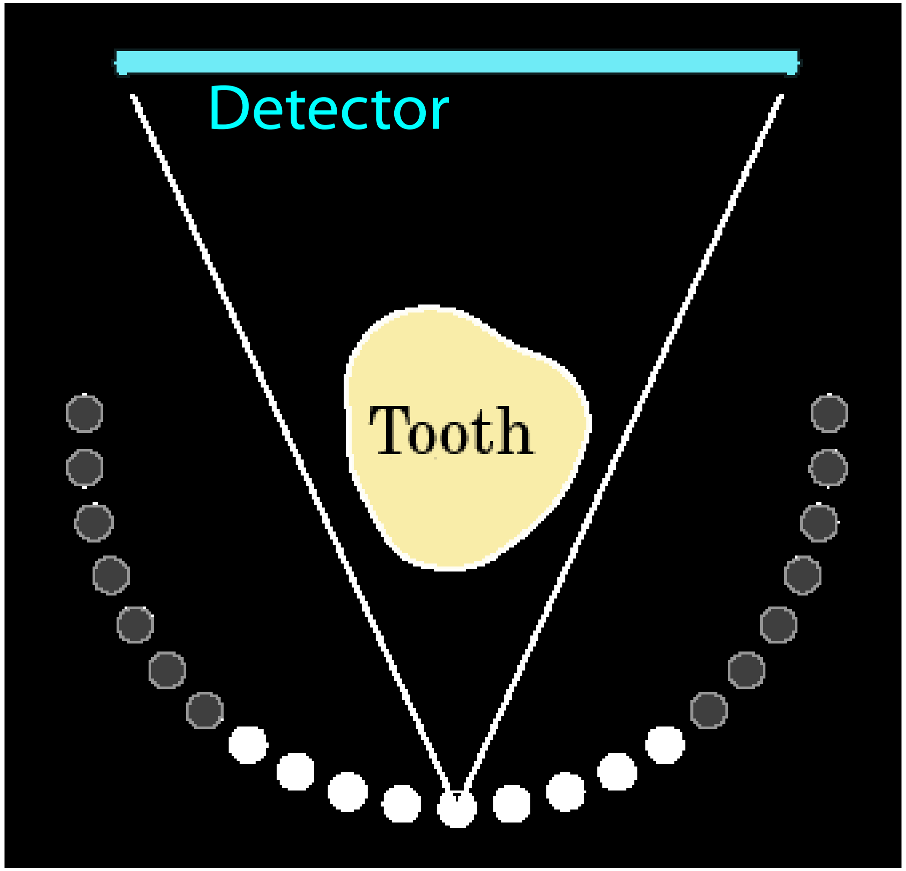
2001 Andersen, Brooks & Hansen

**Personal positioning using cell phone data**

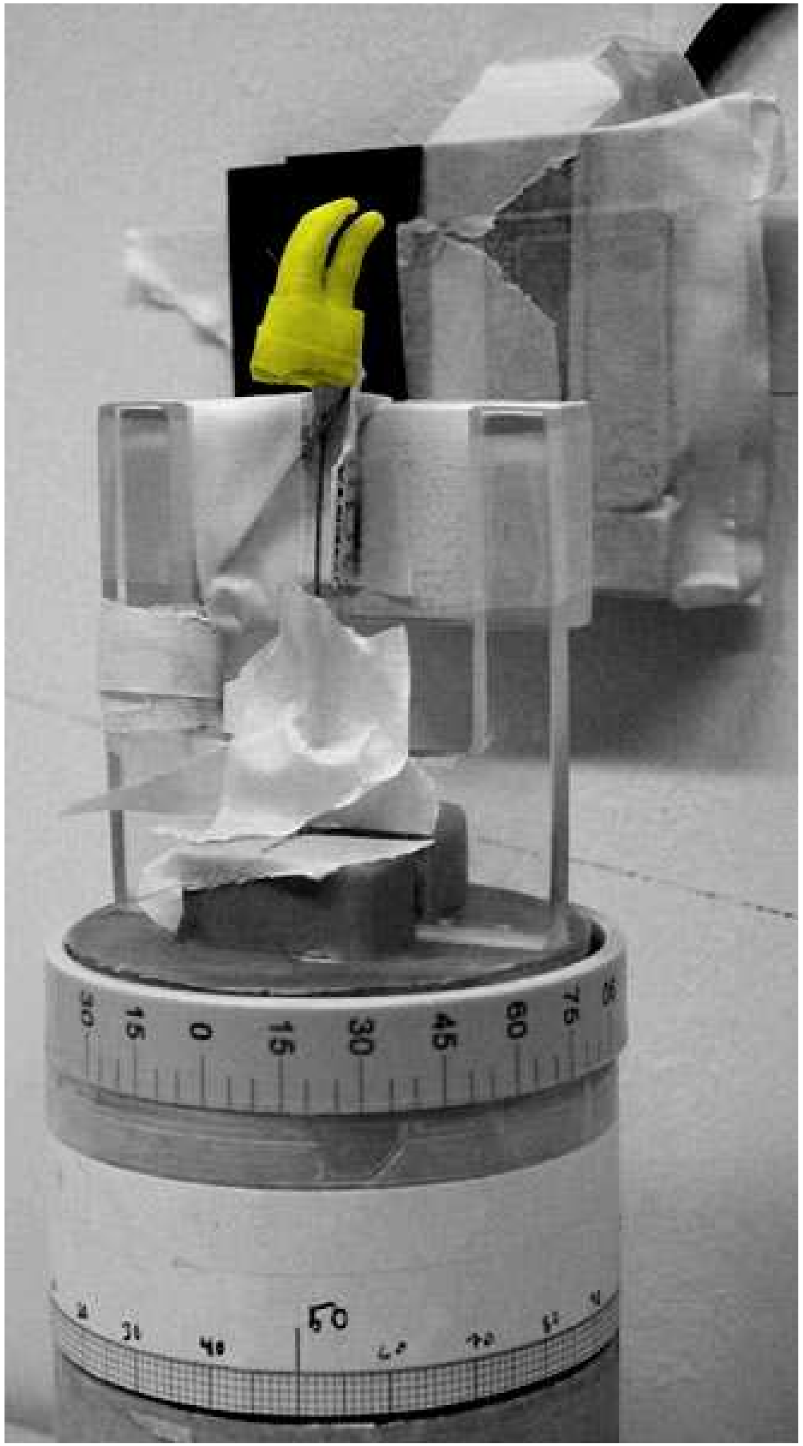
2006 Petrovich and Piché

# Practical example: three-dimensional dental X-ray imaging






X-ray source positions



# We use total variation prior with non-negativity constraint

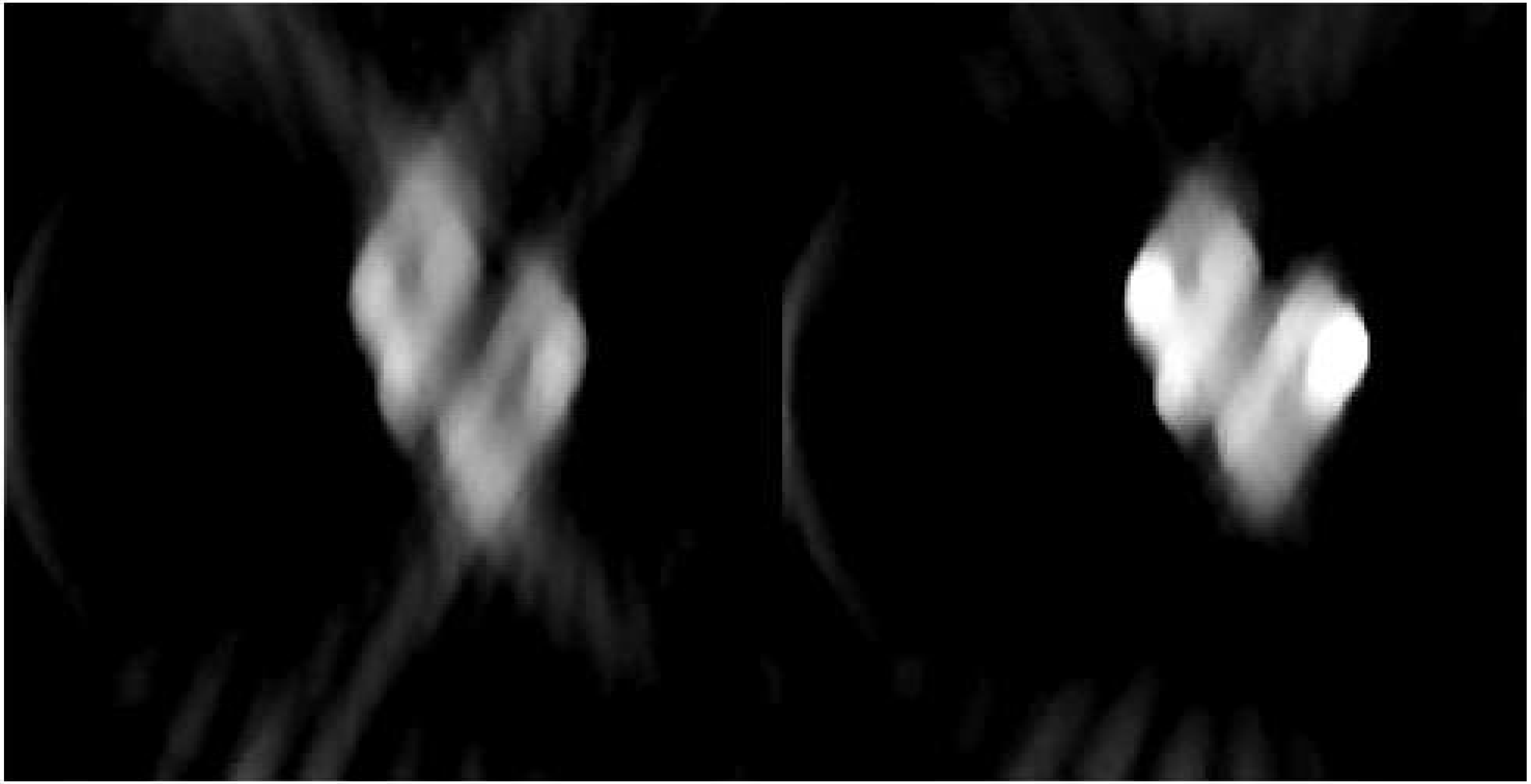
$$\pi_{kn}(u_n \,|\, m_{kn}) = \frac{\Pi_n(u_n)\ \exp(-\frac{1}{2}\|m_{kn} - P_k A u_n\|_2^2)}{\Upsilon_{kn}(m_{kn})},$$

where the prior is given by

$$\Pi_n(u_n) = \chi_{u_n \geq 0} \exp\left( -\alpha_n \sum_{x_\ell, x_\nu \text{ neighbors}} |u_n(x_\ell) - u_n(x_\nu)|_\beta \right).$$

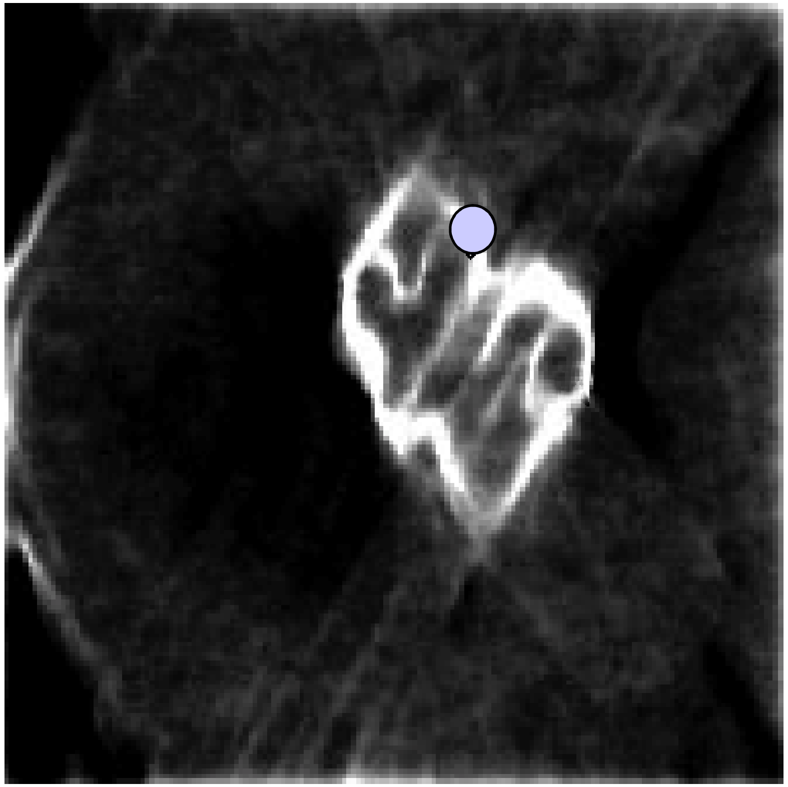
# We can compute several kinds of estimates from the posterior

MAP | Conditional mean

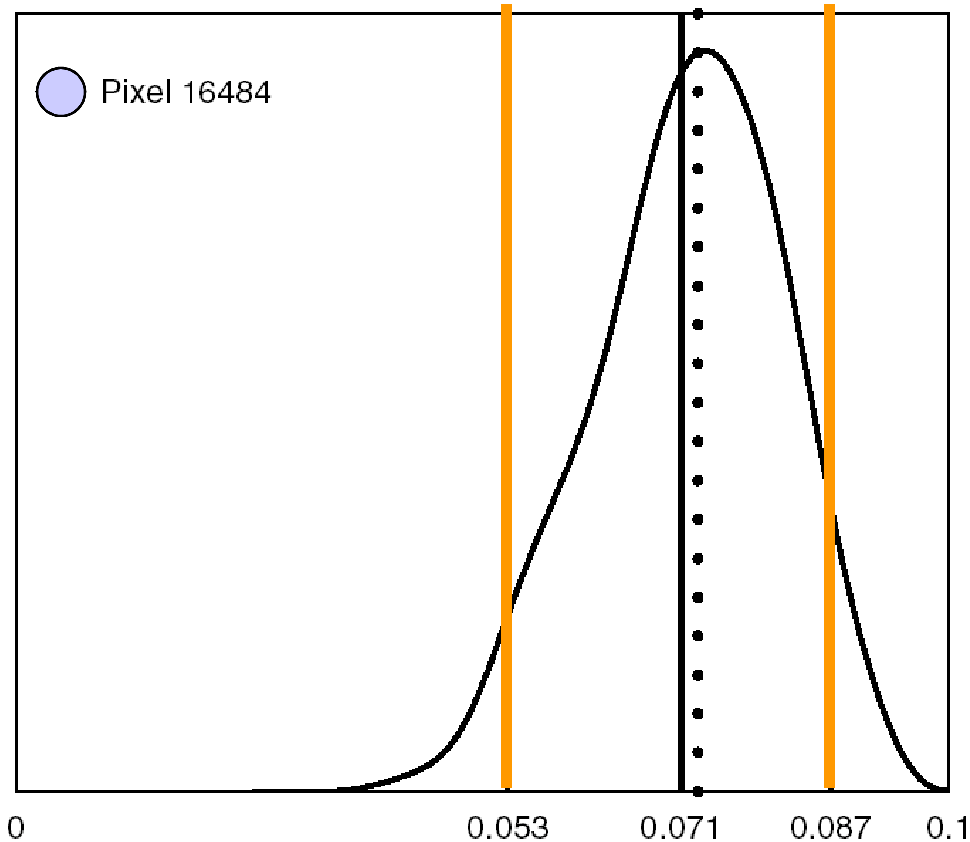


# The Bayesian approach allows further statistical inference

Posterior variance



90% confidence limits




[Kolehmainen *et al.* 2003], thanks to Instrumentarium Imaging

# TV prior seems to work for tomography. However, there is a problem.

Let $Y_n$ be the space of piecewise linear functions on $[0,1]$ with basis $\{\psi_j^n\}$ defined by

$$\psi_j^n(\frac{k}{n}) = \delta_{jk}.$$

We say that $U_n$ is total variation prior in $Y_n$ if

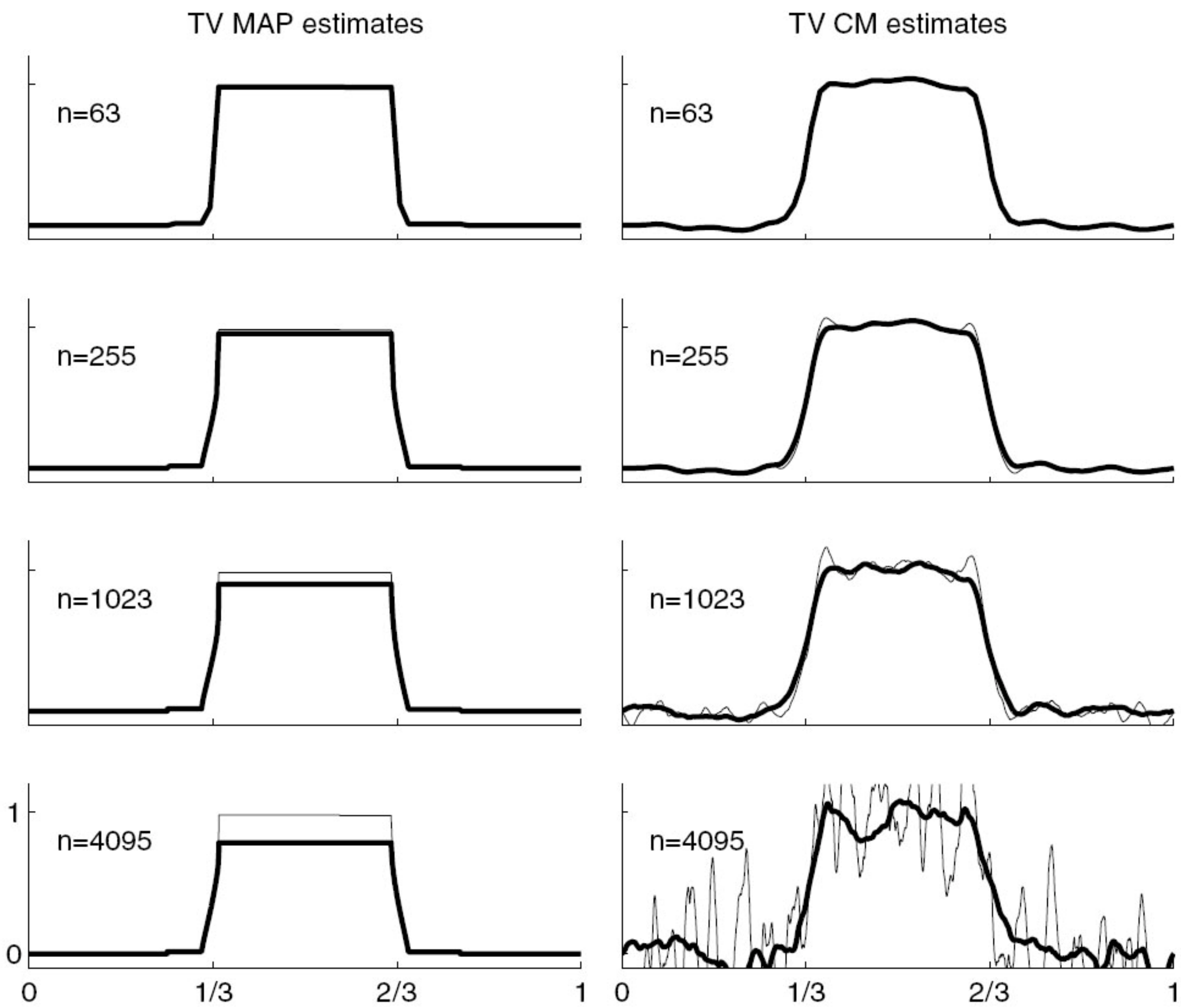
$$U_n(t,\omega) = \sum_{j=1}^{n} u_j^n(\omega)\psi_j^n(t), \quad \omega \in \Omega,$$

where $U_1^n, \dots, U_n^n$ are random numbers with probability density function

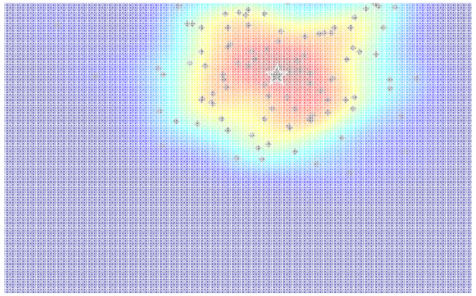
$$\pi(u_1^n, \dots, u_n^n) = c_n \exp(-\alpha_n \|\partial_t u_n\|_{L^1(0,1)}).$$

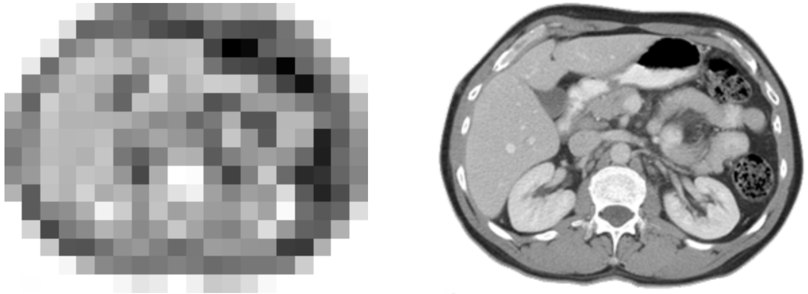
How should $\alpha_n$ be chosen for $n \to \infty$?

# **Total variation prior is not discretization-invariant!** (Lassas & S 2004)




Thin line:
$\alpha_n \sim 1.$
Thick line:
$\alpha_n \sim \sqrt{n}.$





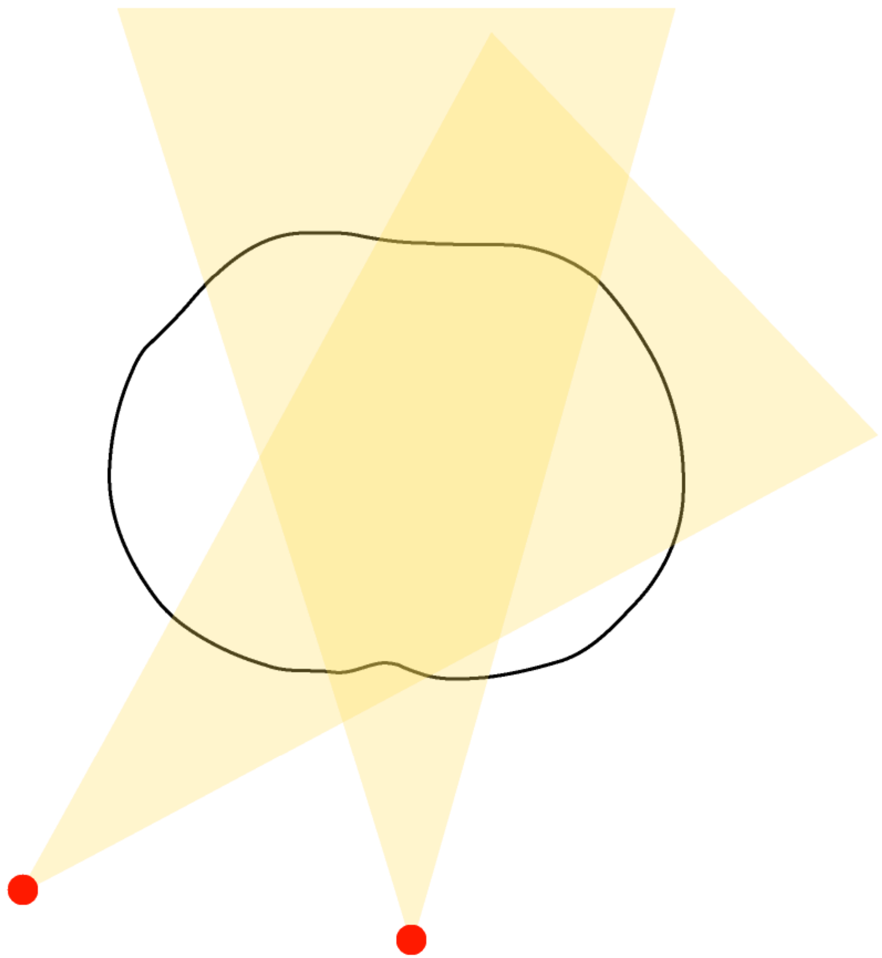
1. Bayesian inversion

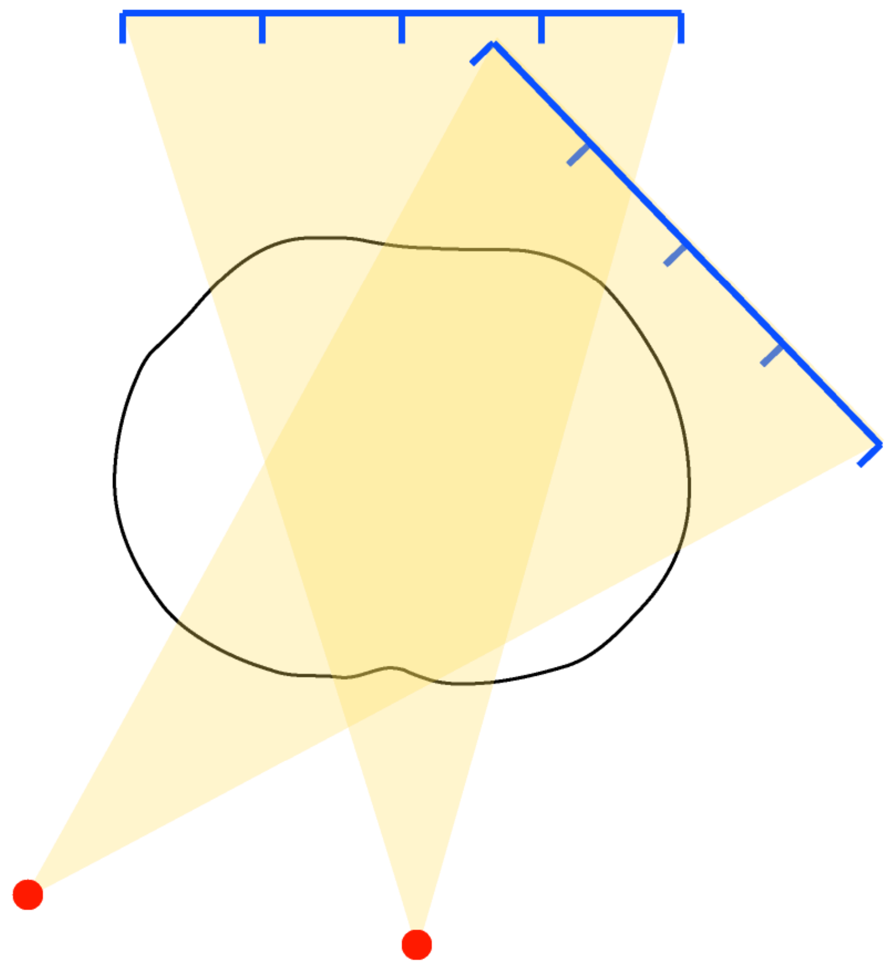
2. Discretization-invariance

3. Regularization results

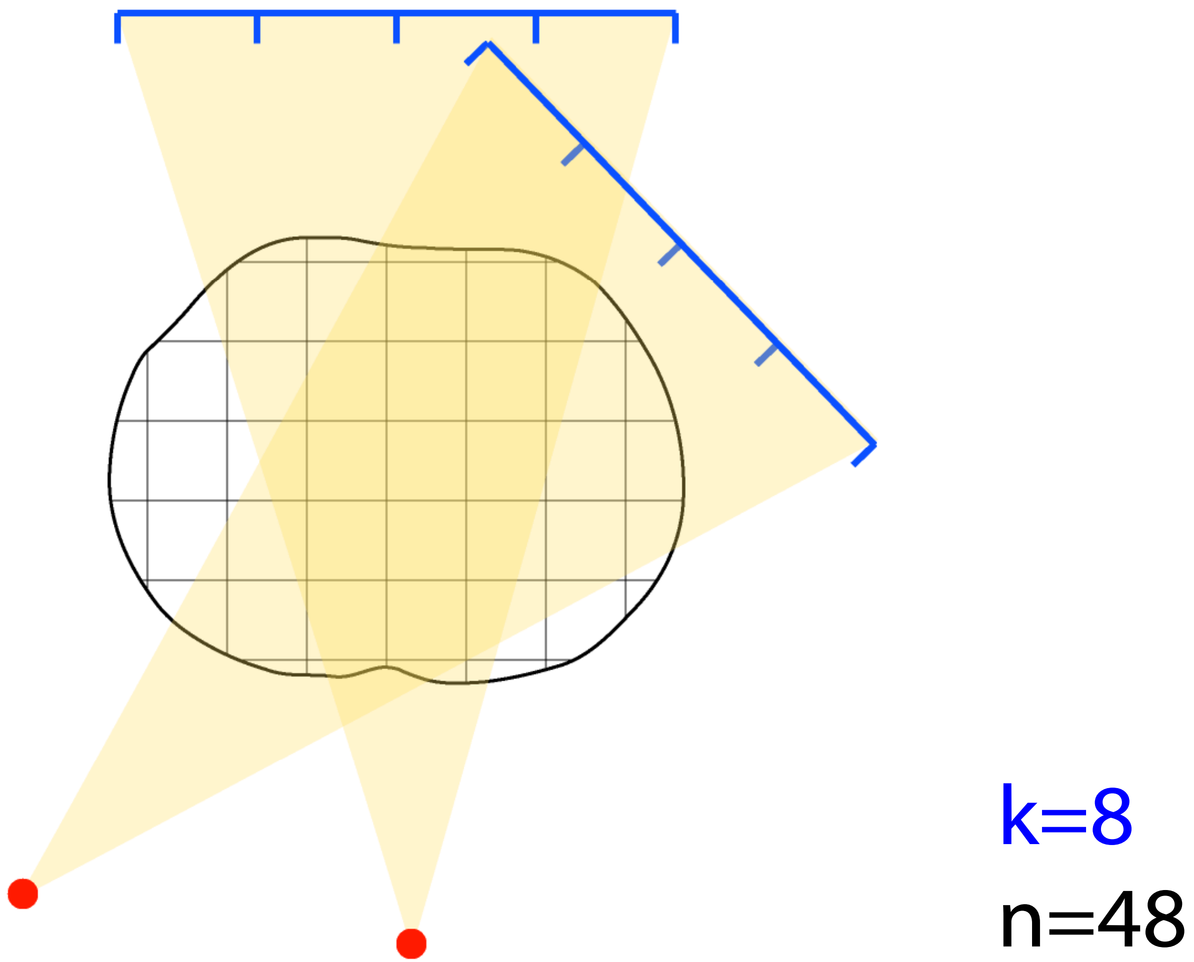
4. Besov space priors

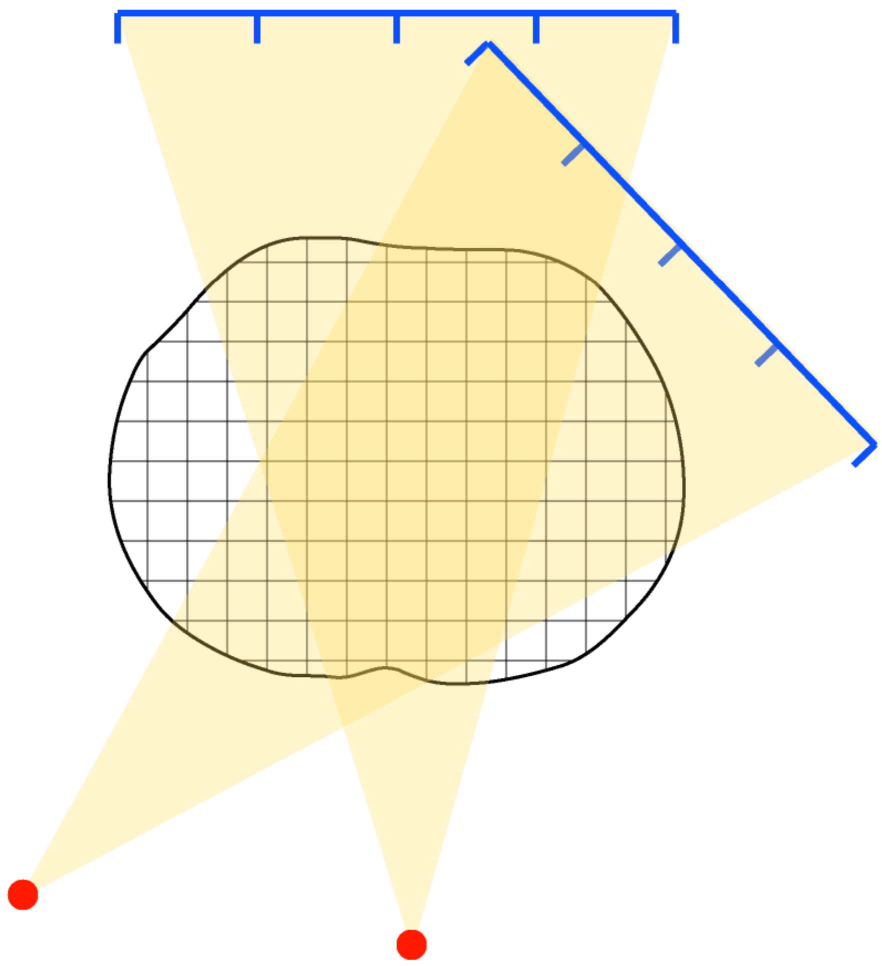

$$e^{-\alpha\|u\|_{B^1_{11}}}$$





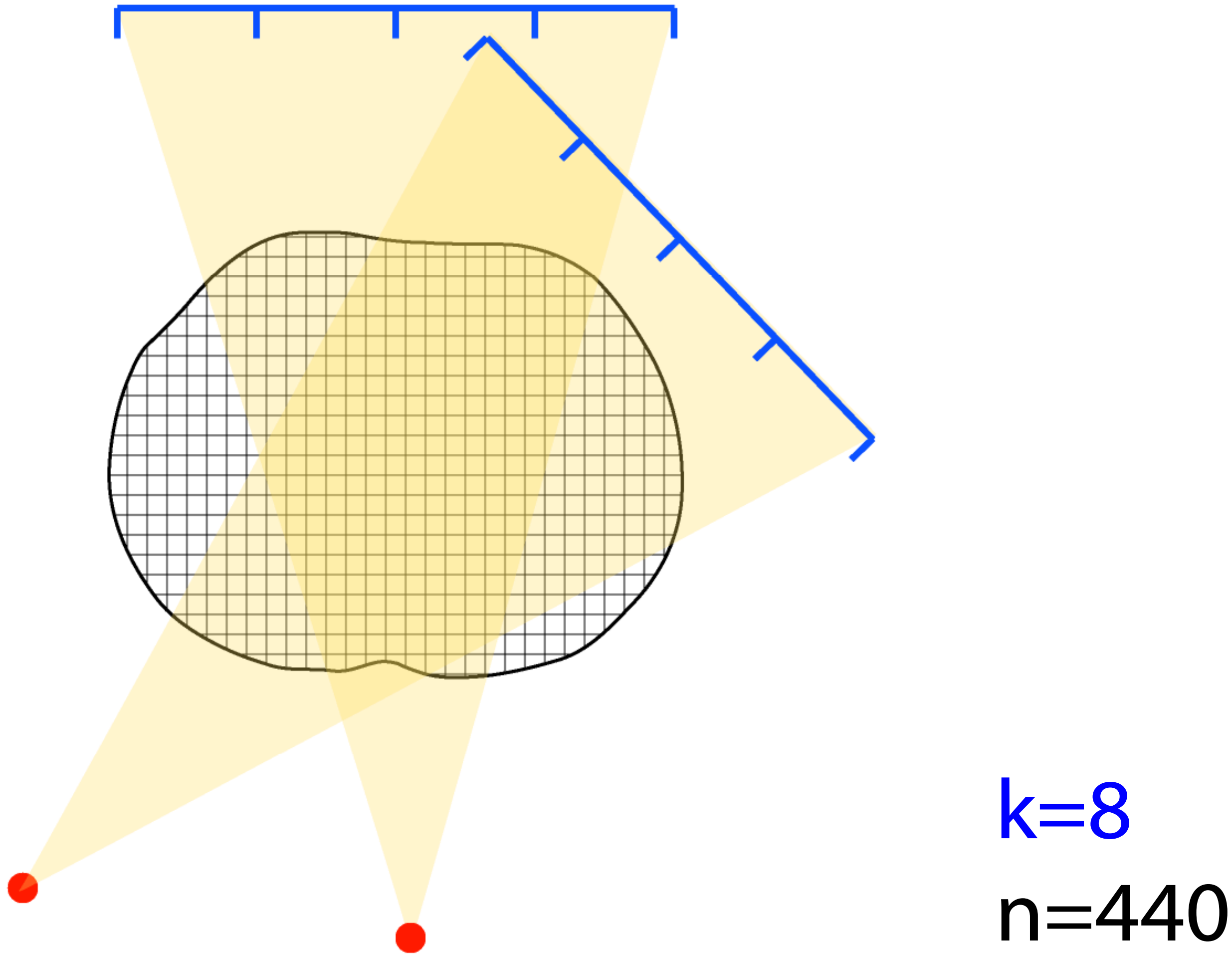
k=8


k=8
n=48

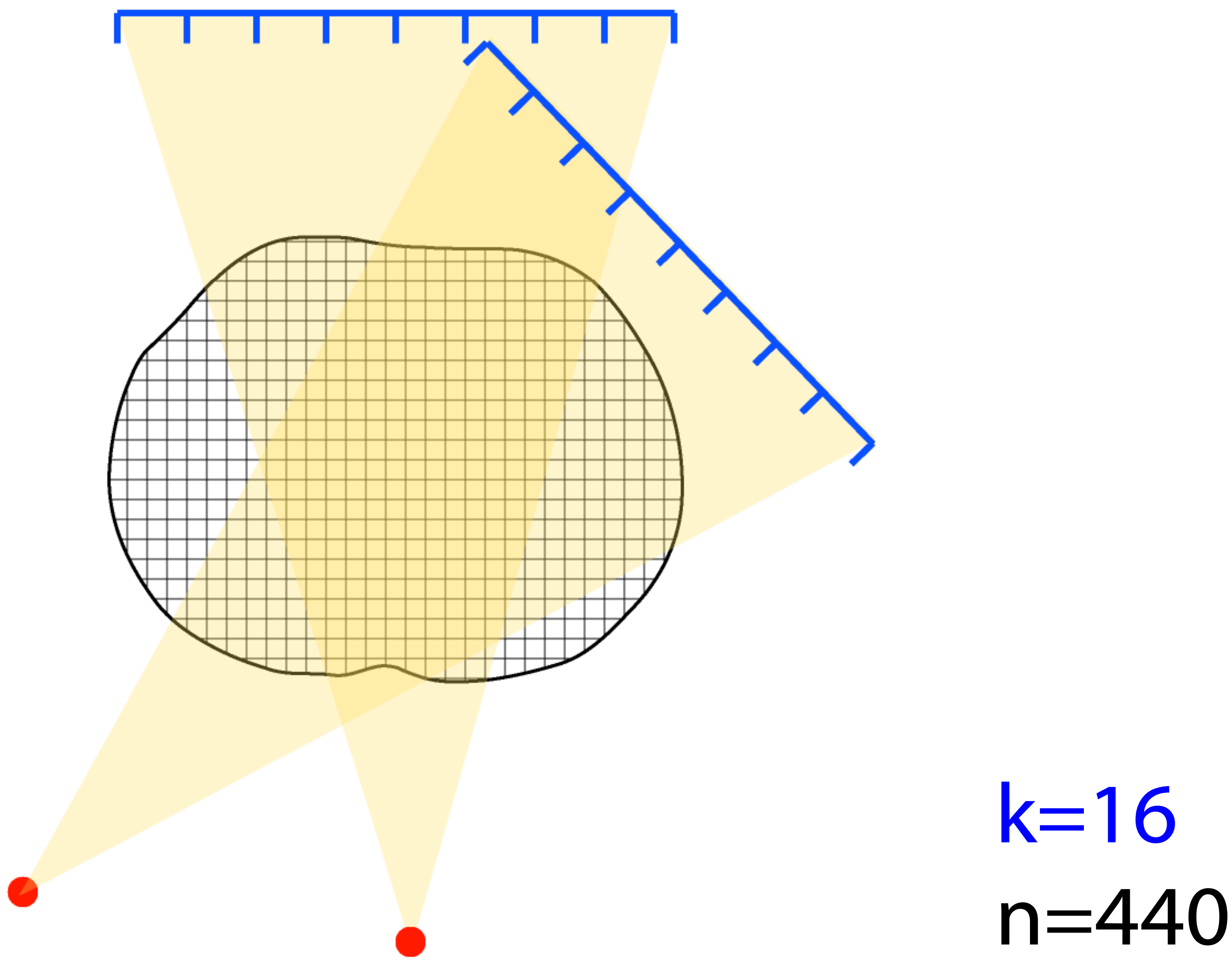


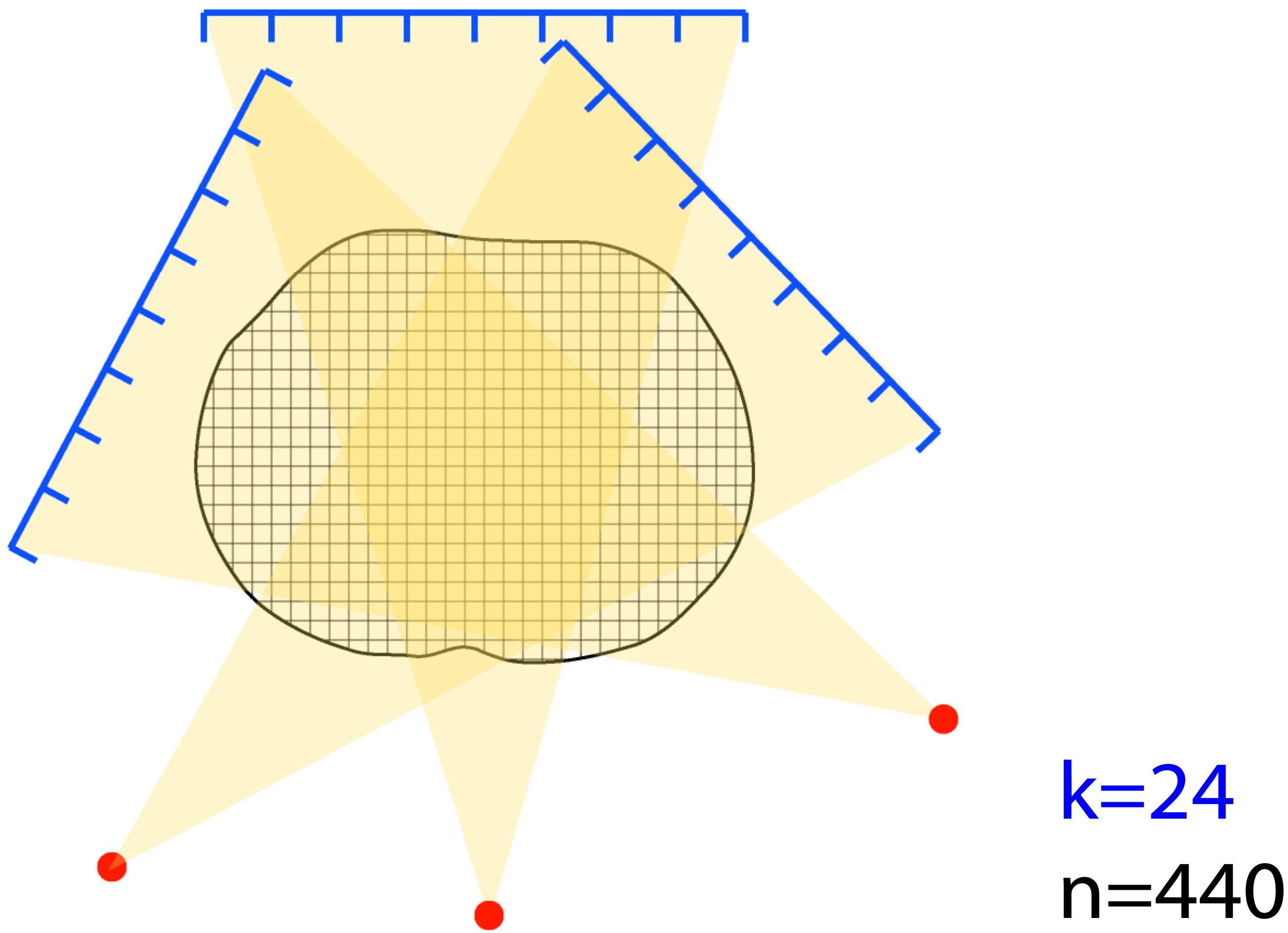
k=8

n=156



k=8
n=440

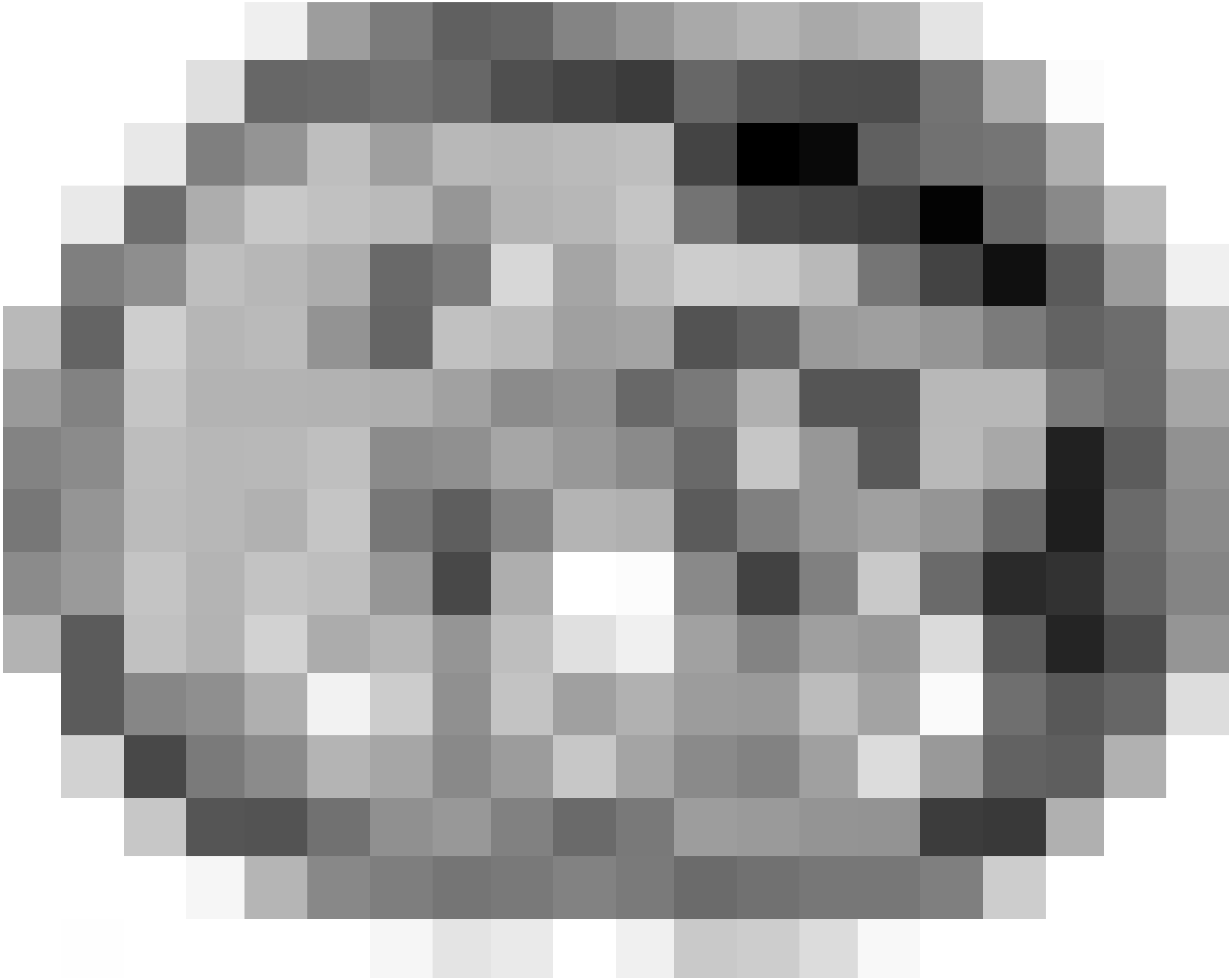

k=16
n=440

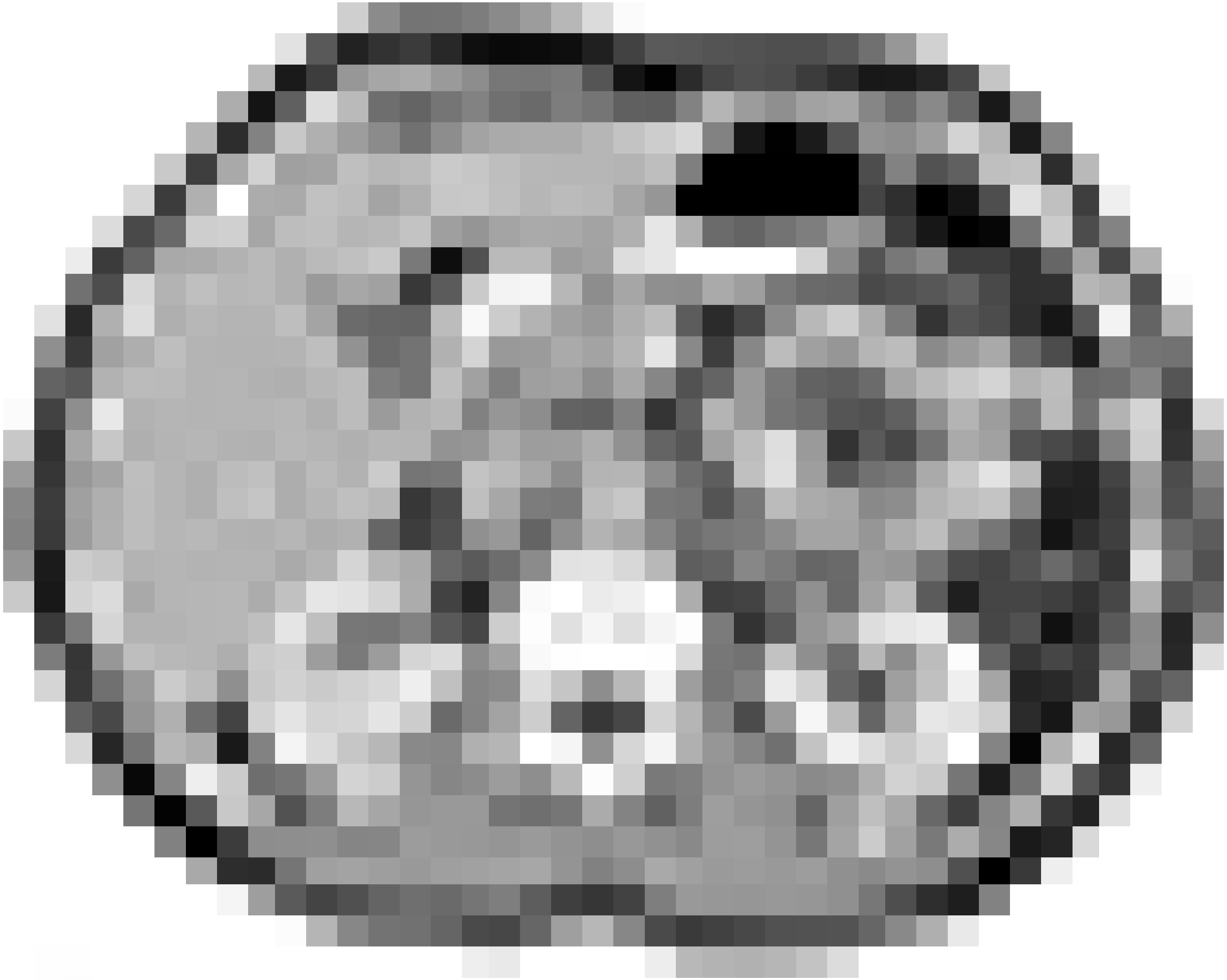

k=24
n=440

**This is the central idea of studying discretization-invariance:**

**The numbers n and k are independent.**

**For the Bayesian inversion strategy to work, the conditional mean estimates must converge as n or k or both tend to infinity.**





# We look for discretization-invariant choices of prior distributions

Recall the conditional mean estimate:

$$\mathbf{u}_{kn}^{\mathsf{CM}} := \int_{\mathbb{R}^n} u_n \pi_{kn}(u_n \,|\, M_k(\omega_0))\, du_n.$$

Possible problems with using $\mathbf{u}_{kn}^{\mathsf{CM}}$:

**Problem 1.** Estimates $\mathbf{u}_{kn}^{\mathsf{CM}}$ diverge as $n \to \infty$.

**Problem 2.** Estimates $\mathbf{u}_{kn}^{\mathsf{CM}}$ diverge as $k \to \infty$.

**Problem 3.** Prior distributions do not express the same prior information for all $n$.

Any choice of $T_n$ and $\Pi_n$ that avoids problems 1–3 is called discretization-invariant.

# Our results continue the tradition of infinite-dimensional statistical inversion

1970 Franklin
1984 Mandelbaum
1989 Lehtinen, Päivärinta and Somersalo
1991 Fitzpatrick
1995 Luschgy
2002 Lasanen
2005 Piiroinen

We achieve discretization invariance for Gaussian and some non-Gaussian prior distributions. Furthermore, we consider realistic measurements.

# Why is discretization invariance useful for finite-dimensional problems?

Sometimes in Bayesian inversion it is necessary to perform computations at two different (finite) resolutions.

For instance,

- Statistical error modeling of Kaipio and Somersalo
- Delayed acceptance Markov chain Monte Carlo

In such case it is important to transform prior information consistently between the two grids.

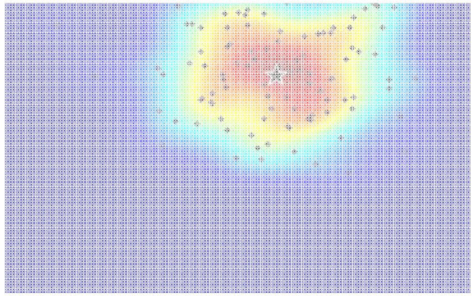
# Our proofs are based on Banach spaces and the concept of reconstructor

We use the following diagram of spaces:

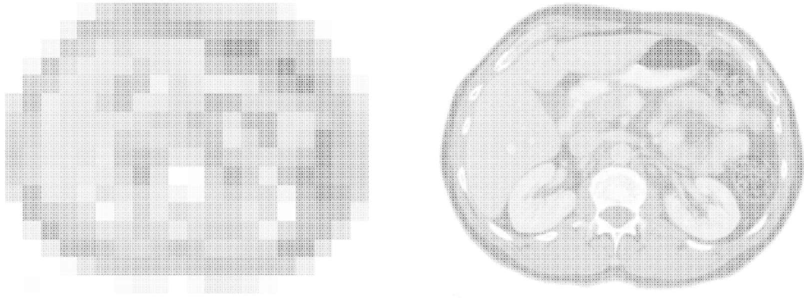
$$\begin{array}{ccccccccc} Y & \xrightarrow{A} & S^1 & \subset & S^{1/2} = Z & \subset & S \\ \Cup & & & & & & \Cup \\ U(\omega_1) & & & & & & \mathcal{E}(\omega_2) \end{array}$$

The reconstructor $\mathcal{R}_{M_{kn}}(U_n | \cdot\ )$ takes the measurement data $m_k = M_k(\omega_0)$ to the mean $\mathbf{u}_{kn}^{\mathsf{CM}}$.

The infinite-dimensional model $M = AU + \mathcal{E}$ has a reconstructor $\mathcal{R}_M(U | \cdot\ )$ as well.



1. Bayesian inversion



2. Discretization-invariance

3. Regularization results


4. Besov space priors

$$e^{-\alpha\|u\|_{B^1_{11}}}$$

# The infinite-dimensional limit case can be rigorously defined

Express U as a random Fourier series

$$\sum_{j=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} c_{j\ell}(\omega) e^{ij\theta + i\ell\psi},$$

and define projections $T_n$ as simple truncations.

We work with the following diagram:

$$\begin{array}{ccccccc} H^{-1}(\mathbb{T}^2) & \xrightarrow{A} & H^2(\mathbb{T}^2) & \subset & L^2(\mathbb{T}^2) & \subset & H^{-2}(\mathbb{T}^2) \\ \Cup & & & & & & \Cup \\ U(\omega_1) & & & & & & \mathcal{E}(\omega_2) \end{array}$$

Note: applying the classical *smoothness prior* results in realizations of U being continuous with probability zero.

**Theorem** (Lassas, Saksman and S 2008)
Bayesian inversion using Gaussian smoothness prior is discretization-invariant:

$$\lim_{k,n\to\infty} \mathbf{u}_{kn}^{\mathrm{CM}} = \lim_{k,n\to\infty} \mathcal{R}_{M_{kn}}(U_n \,|\, M_k(\omega_0)) = \mathcal{R}_M(U \,|\, M(\omega_0))$$

# Bayesian inversion with smoothness prior is related to Tikhonov regularization

In the Gaussian case, the posterior mean

$$\mathbf{u}_{kn}^{\mathsf{CM}} = \int_{\mathbb{R}^n} u_n \exp(-\alpha \|u_n\|_{H^1}^2) \exp(-\frac{1}{2}\|m_k - P_k A u_n\|_2^2)\, du_n$$

coincides with the MAP estimate

$$\arg\max_{u_n} \left[\exp(-\frac{1}{2}\|m_k - P_k A u_n\|_2^2 - \alpha\|u_n\|_{H^1}^2)\right],$$

where $m_k = M_k(\omega_0)$ is measurement data.

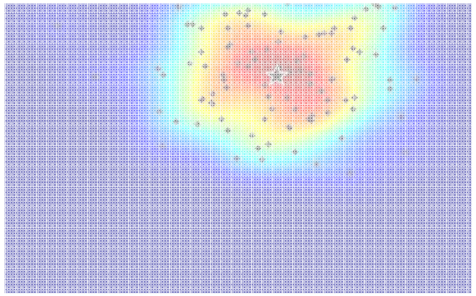
Consider Tikhonov regularization in the form

$$\mathbf{u}_{kn}^{\mathsf{T}} = \arg\min_{u_n} \left[\frac{1}{2}\|m_k - P_k A u_n\|_2^2 + \alpha(\|u_n\|_2^2 + \|\nabla u_n\|_2^2)\right],$$

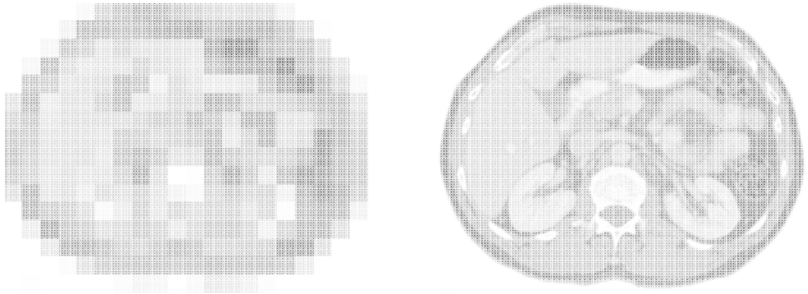
where $\alpha > 0$ is the regularization parameter. It follows that $\mathbf{u}_{kn}^{\mathsf{T}} = \mathbf{u}_{kn}^{\mathsf{CM}} = \mathcal{R}_{M_{kn}}(U_n \,|\, M_k(\omega_0))$.

Much of the above results about
Gaussian smoothness priors
are due to Sari Lasanen and Petteri Piiroinen.

The essential new contribution here is
using the more realistic measurement model.



1. Bayesian inversion



2. Discretization-invariance

3. Regularization results


4. Besov space priors

$$e^{-\alpha\|u\|_{B^1_{11}}}$$

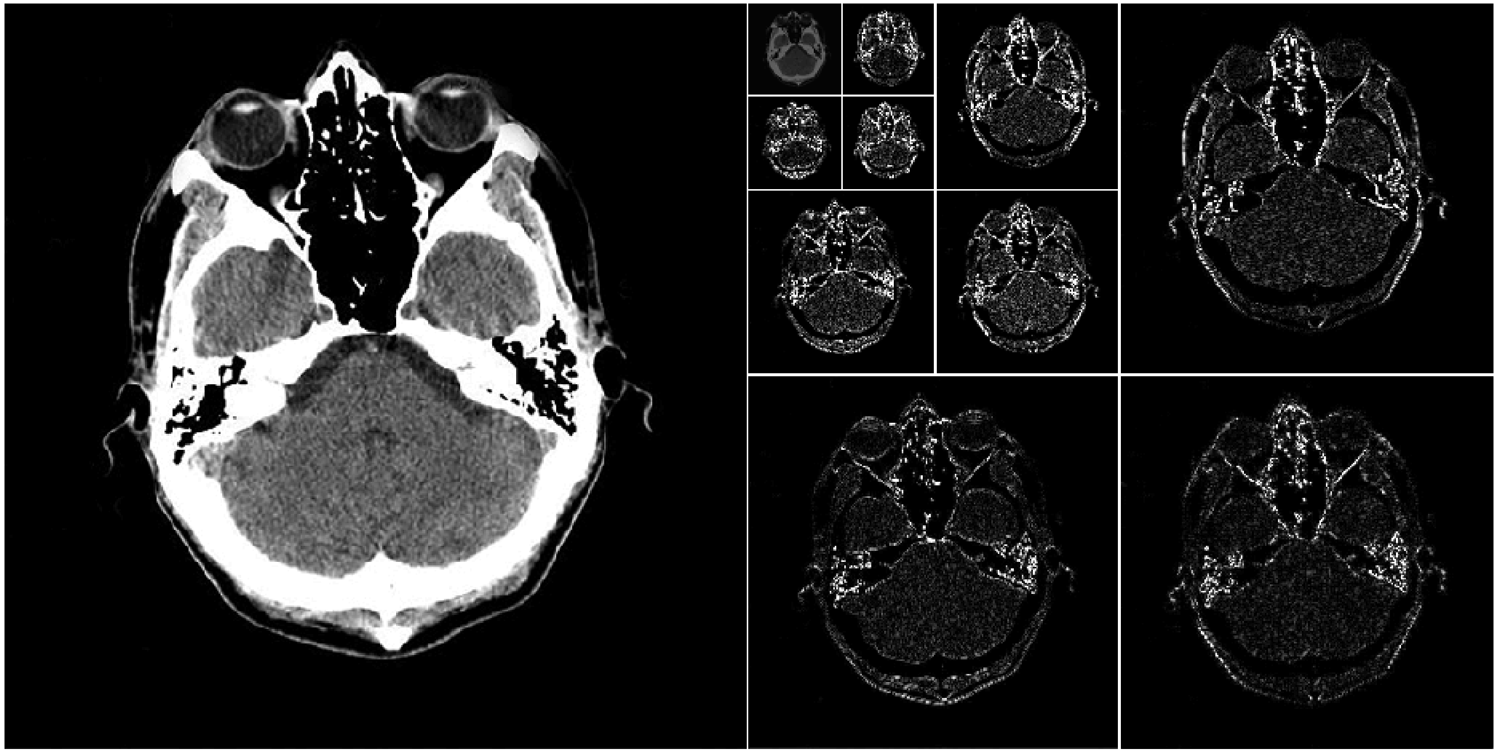
# How to express edge-preserving prior information invariantly?

We suggest replacing the problematic formal total variation density $\exp(-\alpha\|\nabla u\|_1)$ by the formal Besov space density $\exp(-\alpha\|u\|_{B^1_{11}})$.

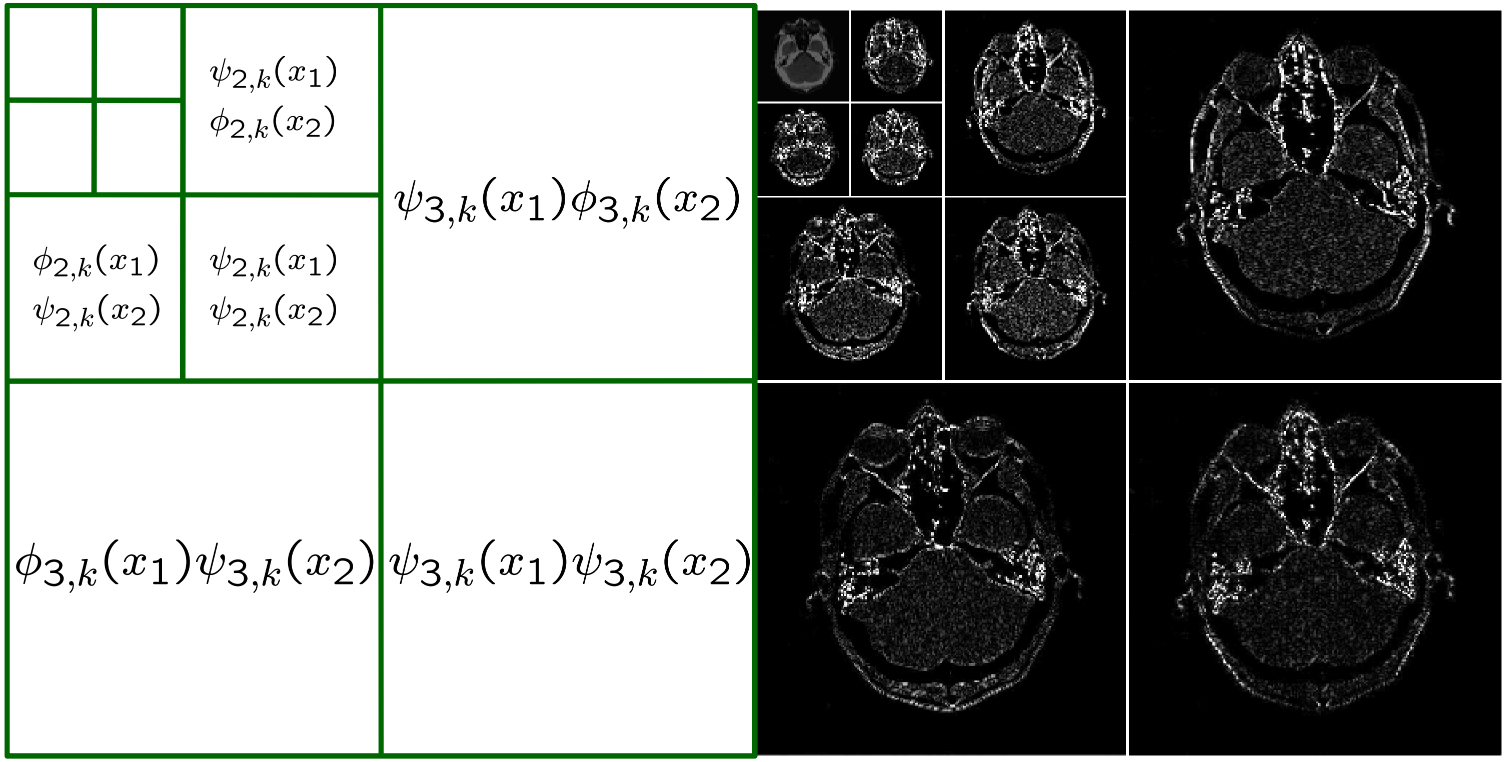
In dimension $d = 2$ the diagram of spaces is

$$\begin{array}{ccccc} B^{-2}_{11}(\mathbb{T}^2) & \xrightarrow{A} & B^{2}_{11}(\mathbb{T}^2) & \subset & B^{-2}_{\infty\infty}(\mathbb{T}^2) \\ \Cup & & & & \Cup \\ U(\omega_1) & & & & \mathcal{E}(\omega_2) \end{array}$$

# Wavelet transform divides a function into details at different scales



# We introduce a convenient renumbering of the basis functions




$$f(x) = \sum_{\ell=1}^{\infty} c_\ell \psi_\ell(x)$$

# Besov space norms can be written in terms of wavelet coefficients

The function

$$f(x) = \sum_{\ell=1}^{\infty} c_\ell \psi_\ell(x)$$

belongs to $B^s_{pq}(\mathbb{T}^d)$ if and only if

$$2^{js} 2^{dj(\frac{1}{2}-\frac{1}{p})} \left( \sum_{\ell=2^{jd}}^{2^{(j+1)d}-1} |c_\ell|^p \right)^{1/p} \in \ell^q(\mathbb{N}).$$

In particular, $f \in B^1_{11}(\mathbb{T}^2)$ if and only if

$$\sum_{\ell=1}^{\infty} |c_\ell| < \infty.$$

# Computation of the CM estimate reduces to sampling from well-known densities

$B^1_{11}(\mathbb{T}^2)$ prior: write $U$ in wavelet basis as

$$U = \sum_{\ell=1}^{\infty} X_\ell \psi_\ell$$

with each $X_\ell$ distributed independently $\sim \exp(-|x|)$.

Posterior distribution of $U_n$ takes the following form in terms of wavelet coefficients $x_1, \ldots, x_n$:

$$C \exp\left( -\frac{1}{2} \| M_k(\omega_0) - A \sum_{\ell=1}^{n} x_\ell \psi_\ell \|^2_{L^2(\mathbb{T}^2)^2} - \alpha \sum_{\ell=1}^{n} |x_\ell| \right)$$

Direct and inverse wavelet transforms are easy and quick to compute.

**Theorem** (Lassas, Saksman and S 2008) Bayesian inversion using $B^1_{11}(\mathbb{T}^2)$ Besov prior is discretization-invariant. More precisely:

Assume that $A : \mathcal{D}'(\mathbb{T}^2) \to C^\infty(\mathbb{T}^2)$ is a bounded linear operator. Let $t < \tilde{t} < -1$, $r > r_1 > 1$ and $\tau > 0$. Assume that $m = M(\omega_0) \in B^{-r_1}_{11}(\mathbb{T}^2)$.

Then

$$\|\mathcal{R}_{M_{kn}}(U_n|m_k) - \mathcal{R}_M(U|m)\|_{B^t_{11}(\mathbb{T}^2)} \leq C[k^{-\tau} + n^{-(\tilde{t}-t)/2}].$$

# We look for discretization-invariant choices of prior distributions

Recall the conditional mean estimate:

$$\mathbf{u}_{kn}^{\mathrm{CM}} := \int_{\mathbb{R}^n} u_n \pi_{kn}(u_n \,|\, M_k(\omega_0))\, du_n.$$

Possible problems with using $\mathbf{u}_{kn}^{\mathrm{CM}}$:

**Problem 1.** Estimates $\mathbf{u}_{kn}^{\mathrm{CM}}$ diverge as $n \to \infty$.

**Problem 2.** Estimates $\mathbf{u}_{kn}^{\mathrm{CM}}$ diverge as $k \to \infty$.

**Problem 3.** Prior distributions do not express the same prior information for all $n$.

Any choice of $T_n$ and $\Pi_n$ that avoids problems 1–3 is called discretization-invariant.

**We show some Besov prior computations to give a flavor of how they work.**

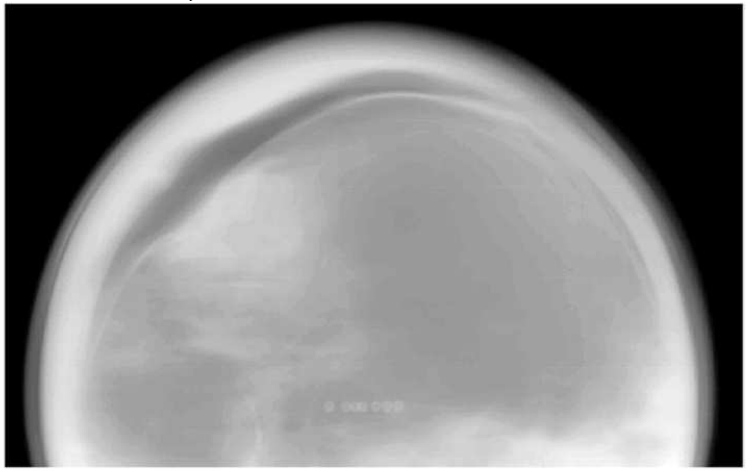
**However, the following examples are maximum a posteriori estimates only.**

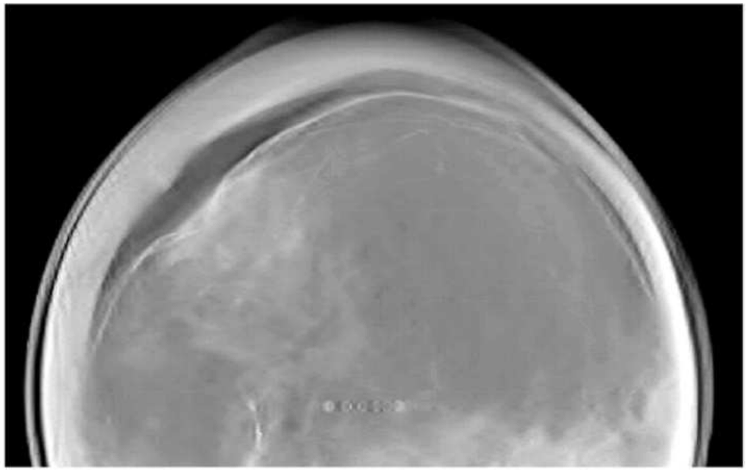
**Computation of conditional mean estimates is a work in progress.**

# Limited angle tomography results for X-ray mammography

[Rantala *et al.* 2006]
Thanks to GE Healthcare

Tomosynthesis



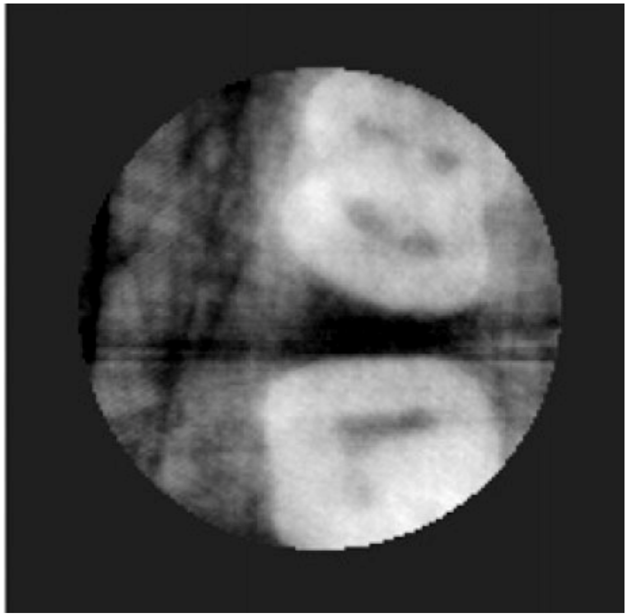


Besov prior

# Local tomography results for dental X-ray imaging

$\Lambda$-tomography

MAP with $B^{1/2}_{3/2,3/2}$ prior



[Niinimäki, S and Kolehmainen 2007]
Thanks to Palodex Group

# Empirical Bayes methodology for specifying Besov prior parameters

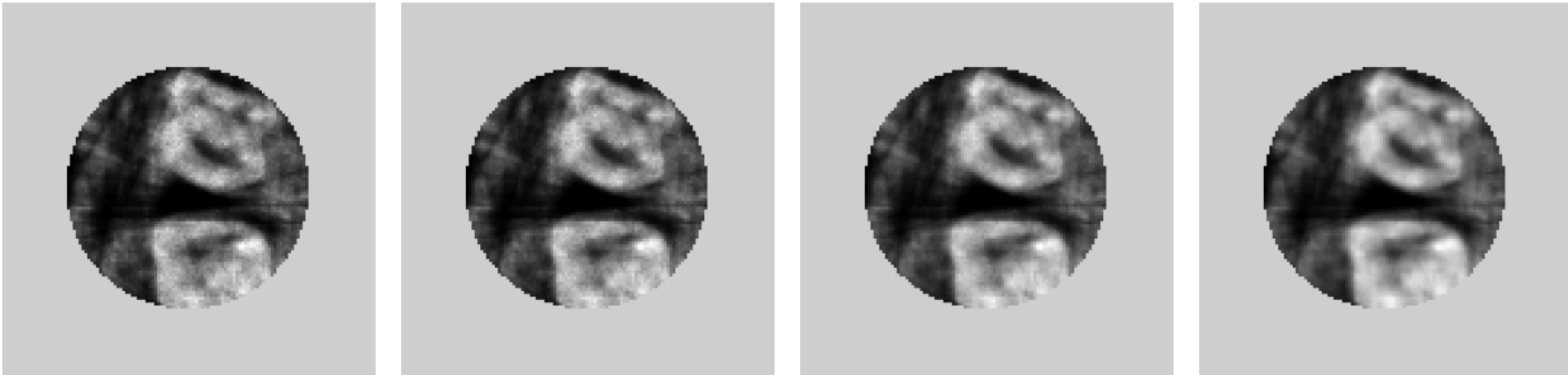


Figure 16: Reconstructions from in vitro dental data. From left: $s = 0$, $s = 0.4$, $s = 0.8$, $s = 1.2$.

[Vänskä, Lassas and S 2008]
Thanks to Palodex Group

# References

**Siltanen et al. 2003**
Statistical inversion for X-ray tomography with few radiographs 1: General theory,
Phys Med Biol 48 pp 1437-1463

**Kolehmainen et al. 2003**
Statistical inversion for X-ray tomography with few radiographs 2:
Application to dental radiology,
Phys Med Biol 48 pp 1465-1490

**Lassas and Siltanen 2004**
Can one use total variation prior for edge-preserving Bayesian inversion?
Inverse Problems 20, pp. 1537-1563

**Kolehmainen et al. 2006**
Parallelized Bayesian inversion for 3-D dental X-ray imaging
IEEE Transactions on Medical Imaging 25(2), pp. 218-228.

**Rantala et al. 2006**
Wavelet-based reconstruction for limited angle X-ray tomography
IEEE Transactions on Medical Imaging 25(2), pp. 210-217

**Niinimäki, Siltanen and Kolehmainen 2007**
Bayesian multiresolution method for local tomography in dental X-ray imaging.
Physics in Medicine and Biology 52, pp. 6663-6678.

**Vänskä, Lassas and Siltanen**
Statistical X-ray tomography using empirical Besov priors.
To appear in International Journal of Tomography & Statistics.

**Lassas, Saksman and Siltanen (submitted)**
Discretization invariant Bayesian inversion and Besov space priors

You can download the references at

**www.siltanen-research.net**