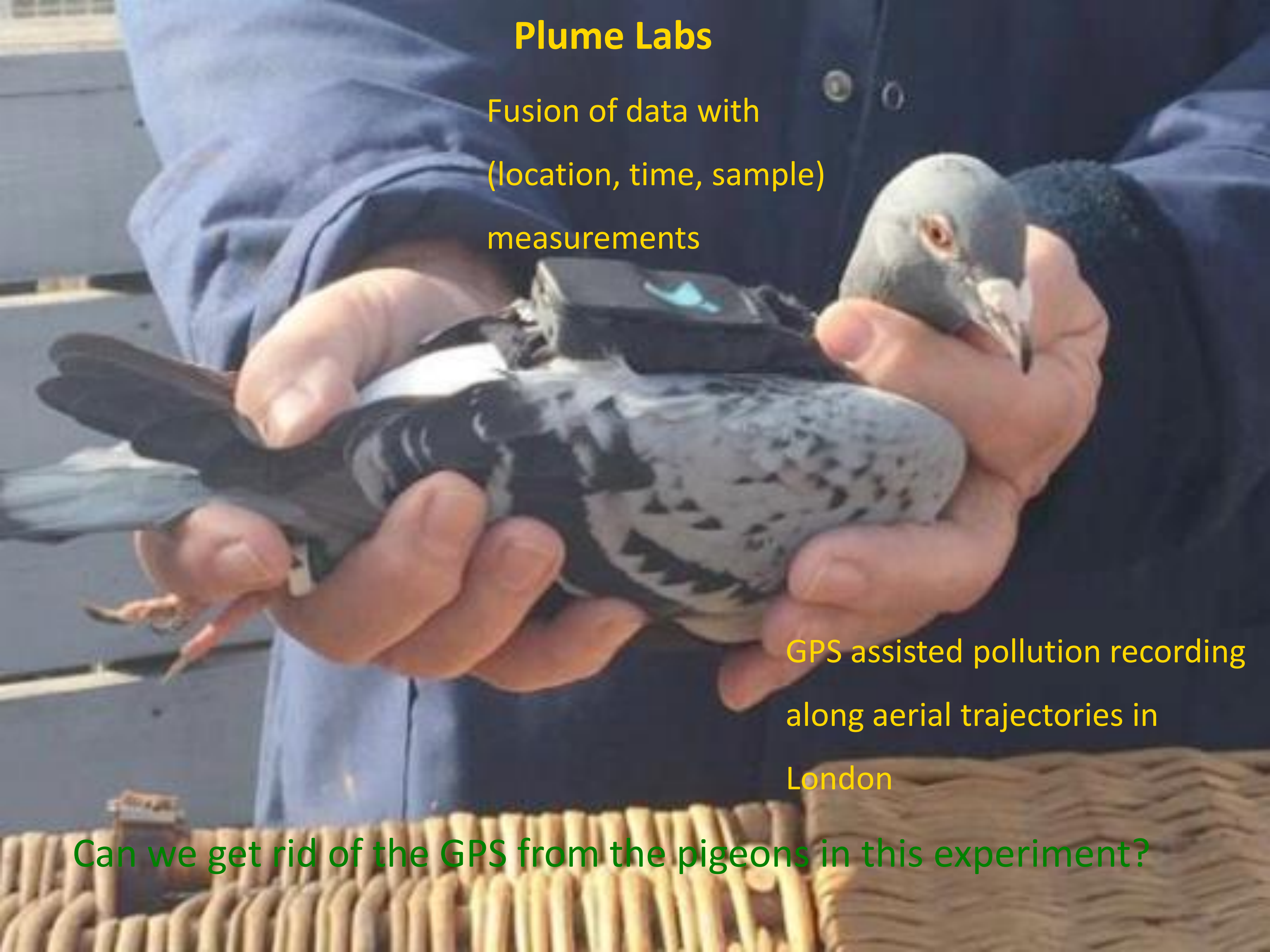
Bandlimited Field Estimation from Samples Recorded by Location-Unaware Sensors

Animesh Kumar

Electrical Engineering

Indian Institute of Technology Bombay, India



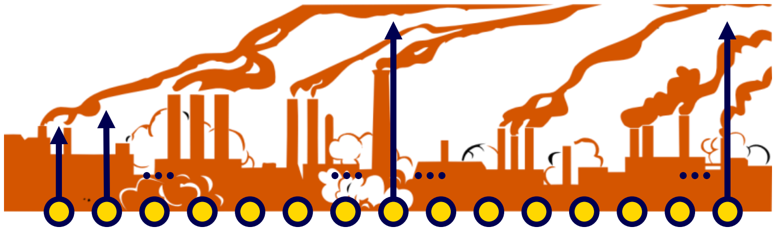
Plume Labs

Fusion of data with (location, time, sample) measurements

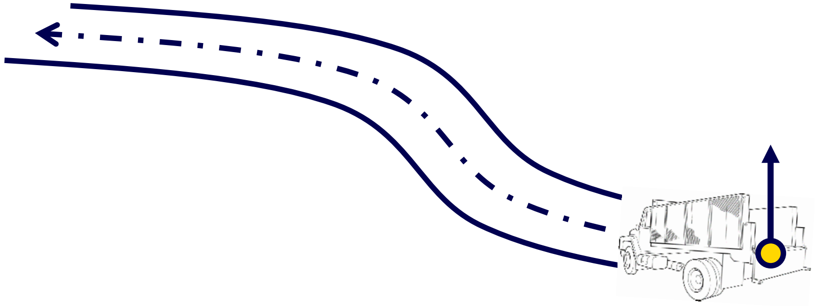
GPS assisted pollution recording along aerial trajectories in London

Can we get rid of the GPS from the pigeons in this experiment?

Spatial sampling is everywhere

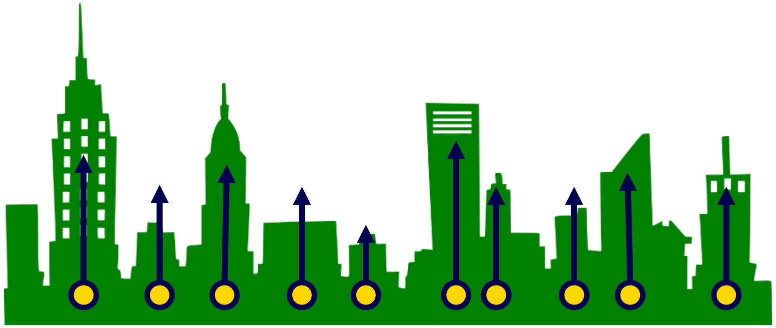


Spatial field sampling with an array of sensors



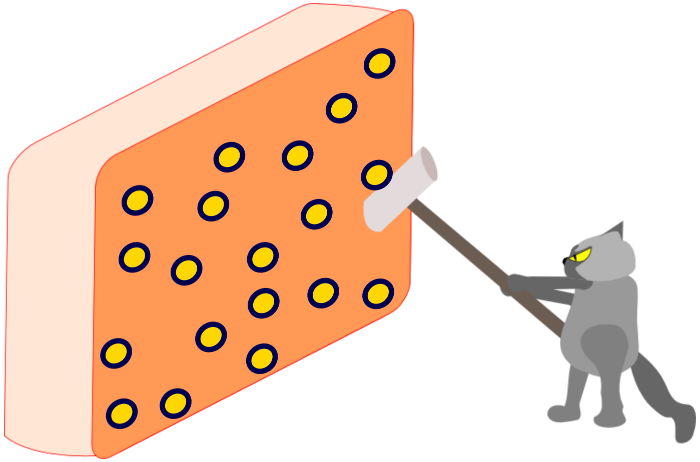
Sampling along a path with vehicle

[Unnikrishnan-Vetterli’2012]



Sensing with the Internet of Things

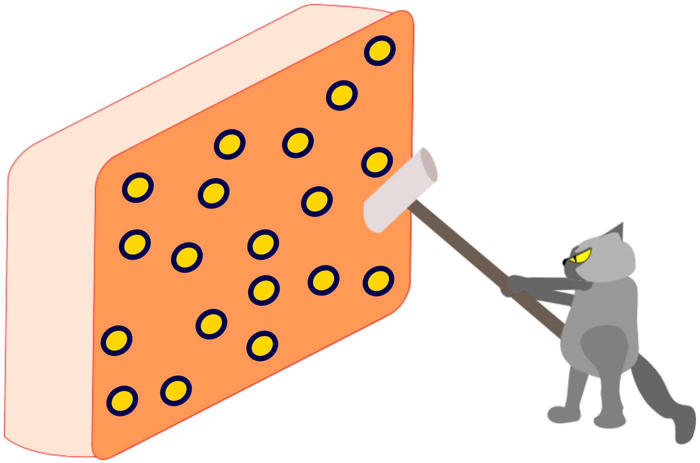
[Zanella-Bui-Castellani et al.’2014]



Randomly sprayed smart-dust/paint

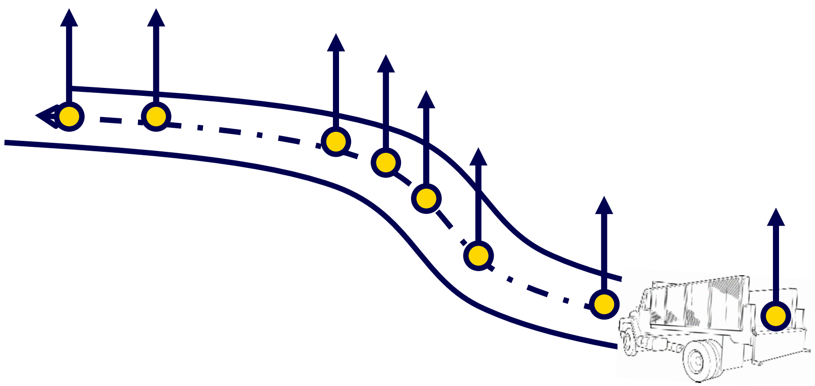
[Kahn-Katz-Pister’1999]

Spatial sampling models of interest



scattering model, motivated by smart-dust/smart-paint

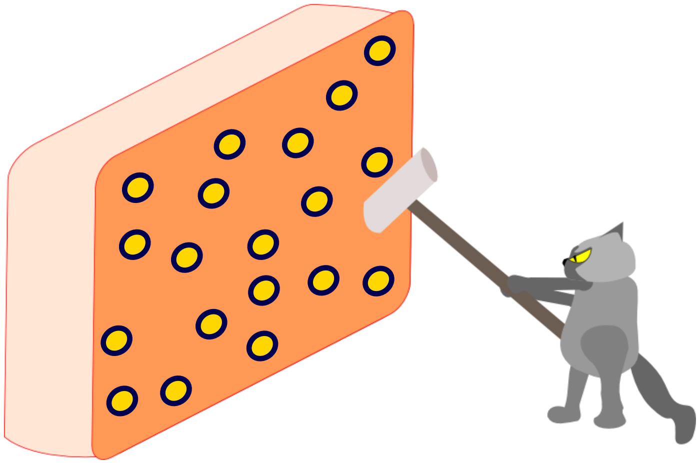
[Kahn-Katz-Pister’1999]



sampling on a path model, motivated by mobile sampling

[Unnikrishnan-Vetterli’2012]

Spatial sampling with "unknown" location

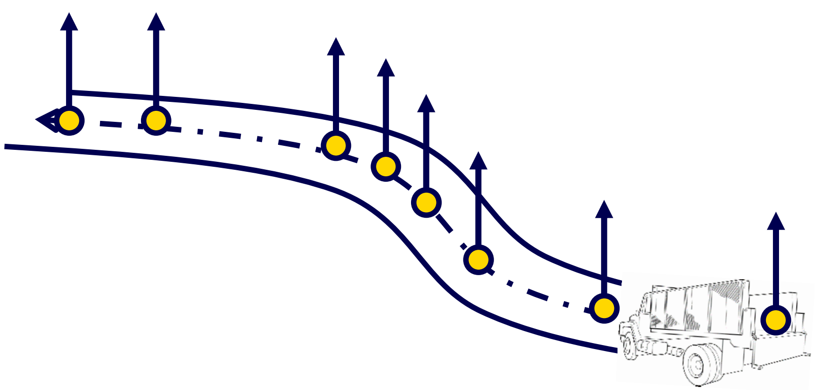


uniformly scattered smart-dust/smart-paint

If no localization algorithm is used

↑ Saves power, cost, and resources

↓ Location unawareness



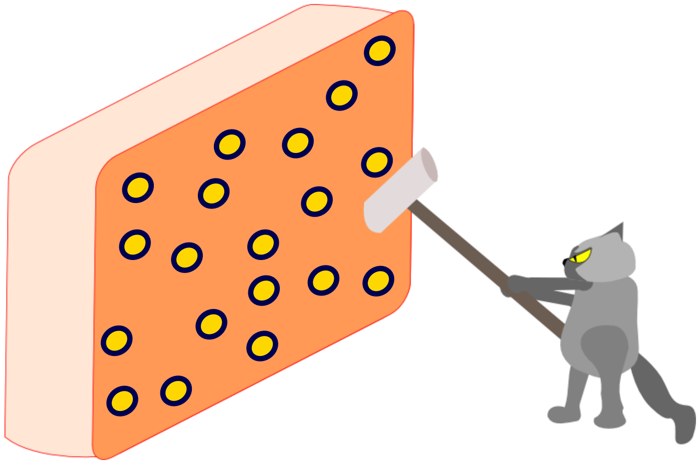
sampling on a path model

No GPS is used

↑ Saves power and cost

↓ Location unawareness

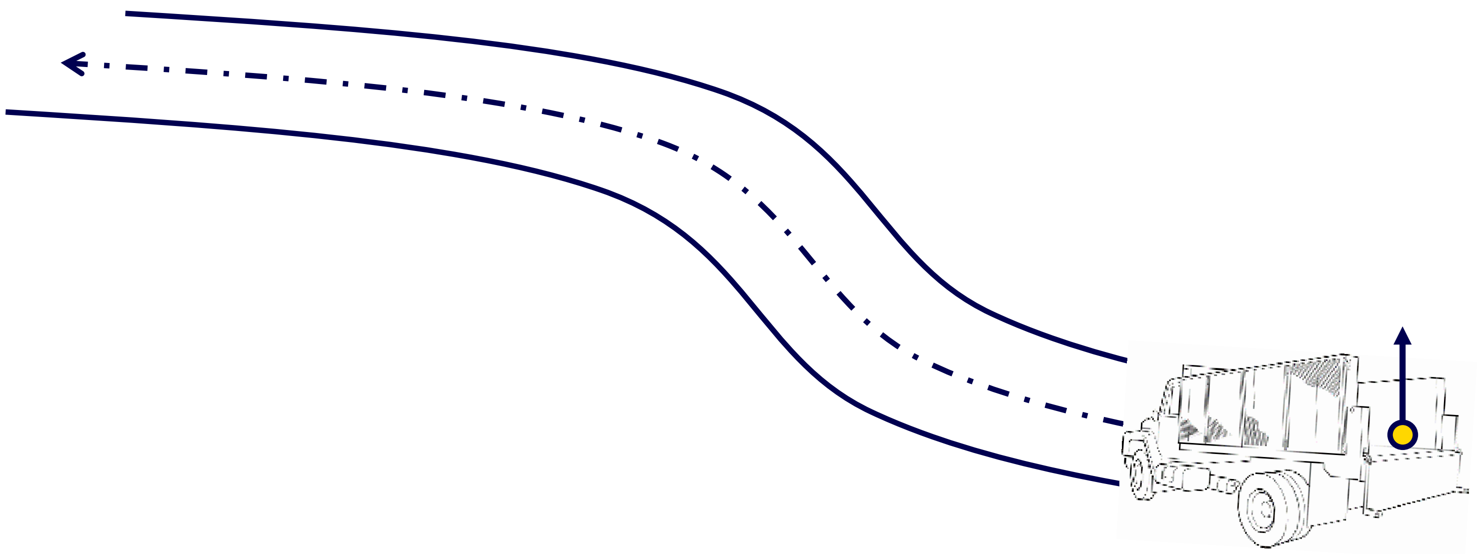
Scattering model and field estimation



◊ Location-unawareness models the scattering of sensors. Sensor locations are assumed to be uniformly distributed and independent

◊ Underlying structure such as bandlimitedness, smoothness, or sparsity will be an aid in field estimation

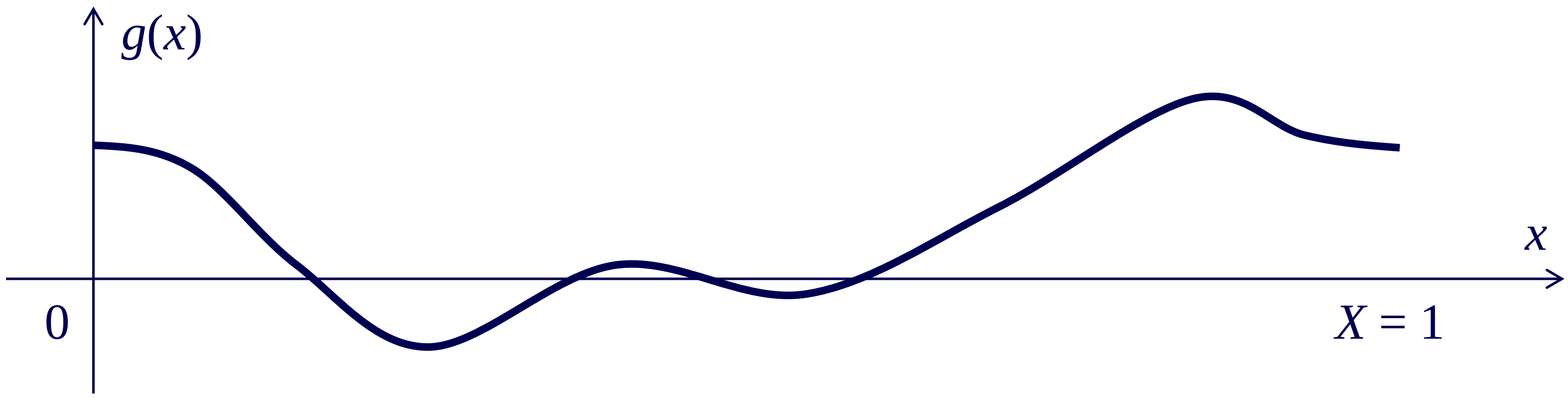
Mobile sampling model and field estimation



◊ Location-unawareness models unknown mobile-sensor’s speed profile, no location tracking, or the absence of GPS

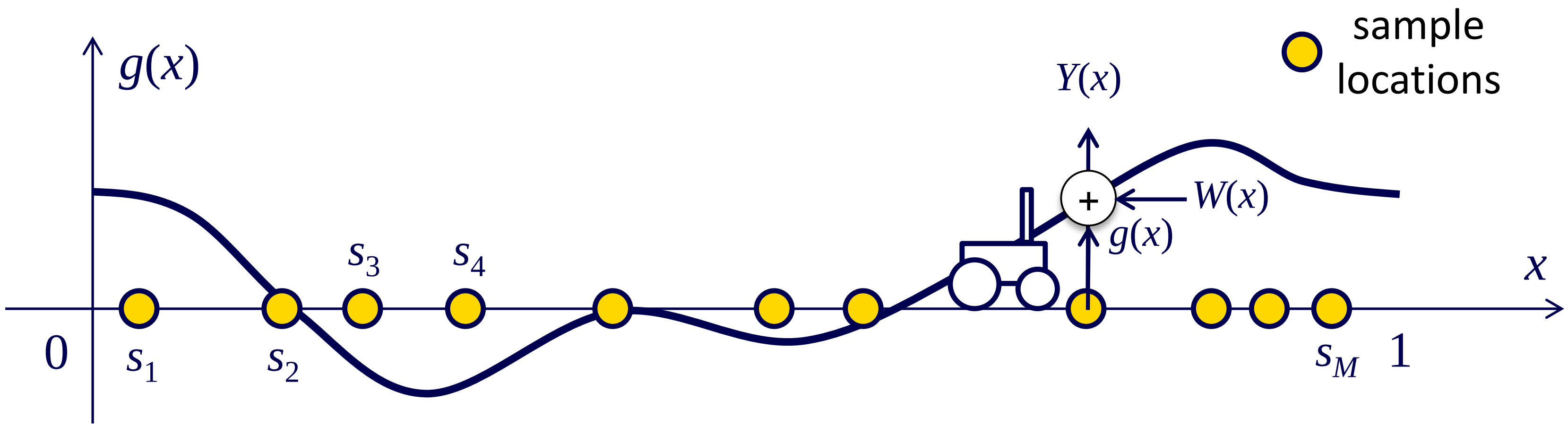
◊ Underlying structure such as bandlimitedness, smoothness, or sparsity will be an aid in field estimation

Spatial acquisition problem of interest



Consider the acquisition problem, where a smooth field in a finite interval has to be sampled or estimated. Field is assumed to be fixed with time during the measurement process

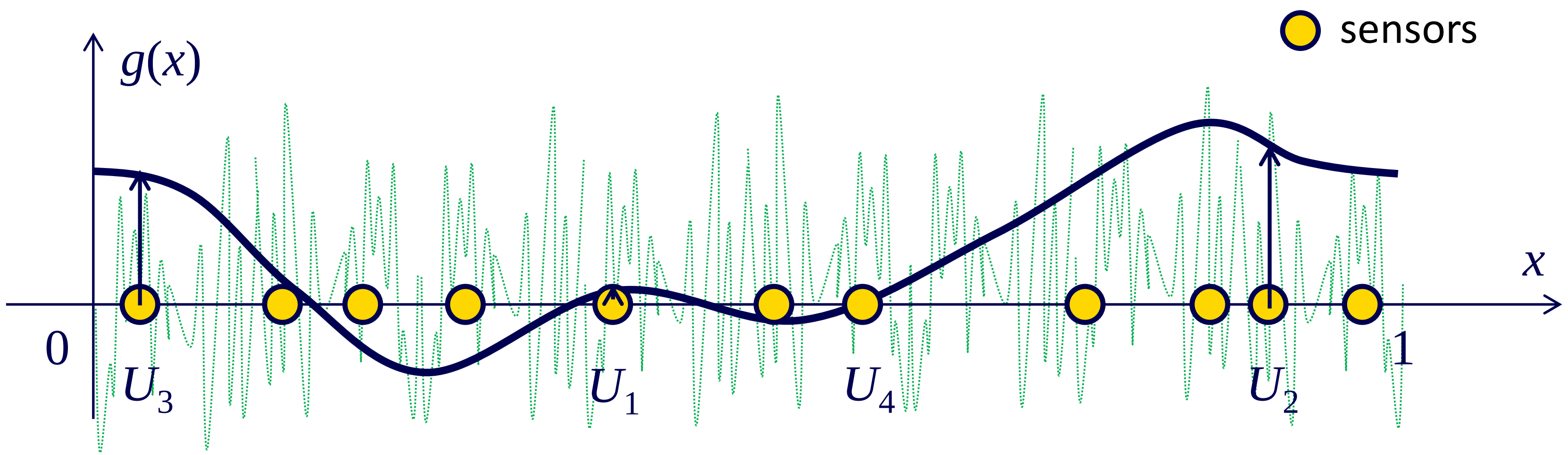
Acquisition: the adversaries



Adversaries during field acquisition include:

- ◊ Unknown sampling locations $s_1, s_2, s_3, \ldots$
- ◊ Additive measurement noise, afflicting the field sample measurement process
- ◊ Unknown bandwidth of $g(x)$ (an essential ingredient of Nyquist style sampling)
- ◊ Quantization of samples recorded by the sensor

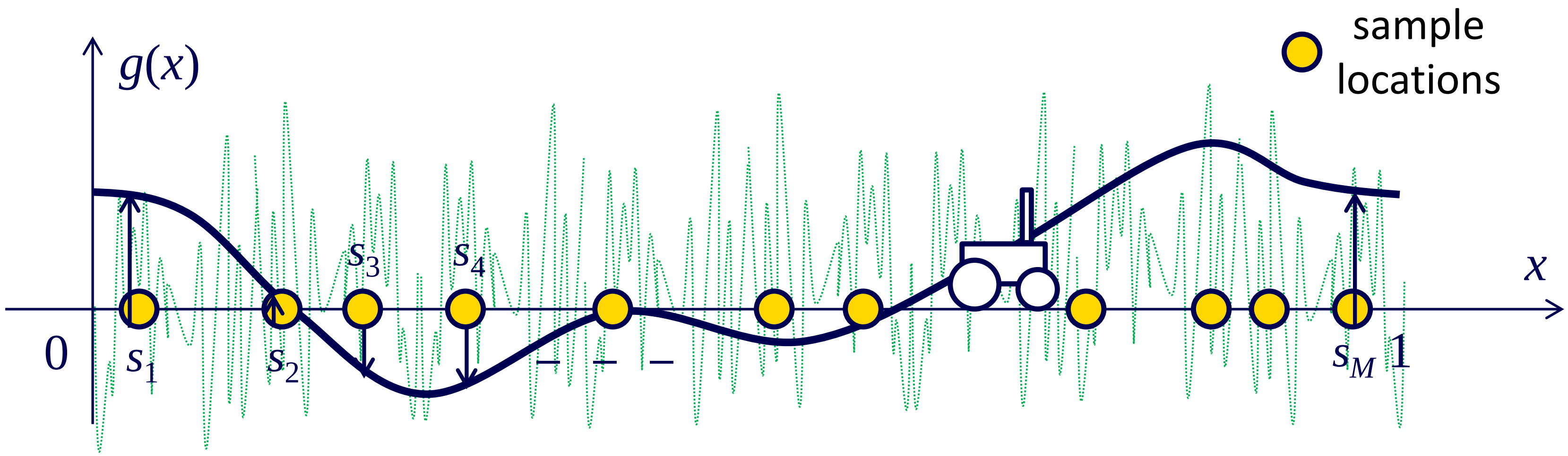
Acquisition setup with scattered sensors



A smart-dust sensor network samples a bandlimited field such that

- ◊ The locations $U_1, U_2, \ldots$ are obtained from scattering of uniformly distributed sensor locations in the interval $[0,1]$
- ◊ There is measurement-noise, with zero-mean and finite variance, that is independent and identically distributed but unknown in distribution

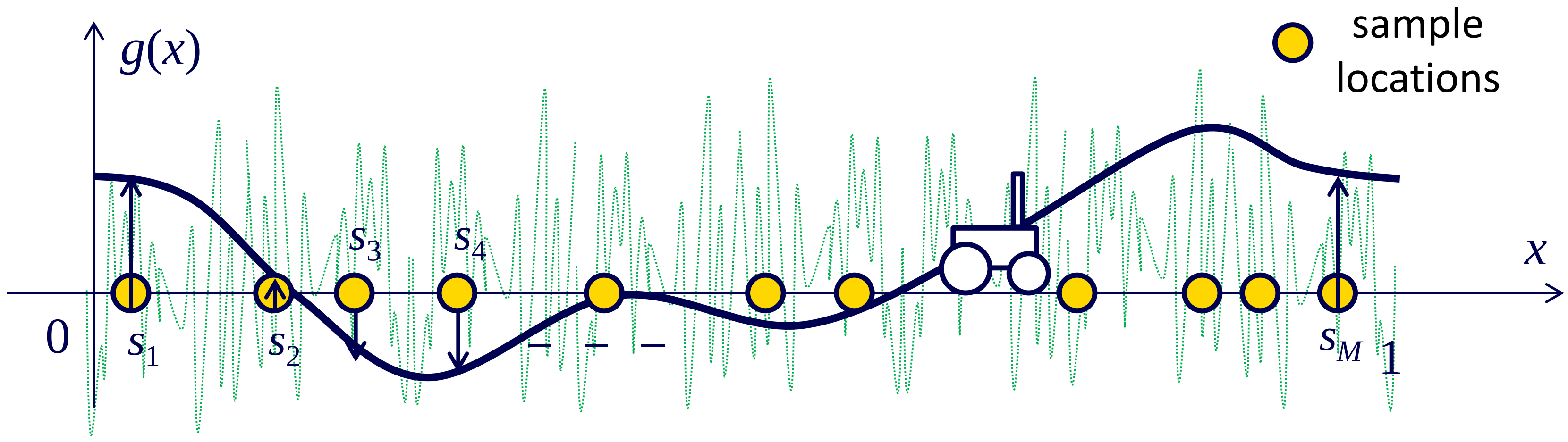
Acquisition setup with mobile sensors



A mobile-sensor samples a bandlimited field at unknown locations such that

- Due to lack of good models for mobile sensor trajectories, the distribution governing S_1, S_2, … is unknown
- The unknown locations s_1, s_2, … are obtained from an unknown renewal process
- There is measurement-noise, with zero-mean and finite variance, that is independent and identically distributed but unknown in distribution

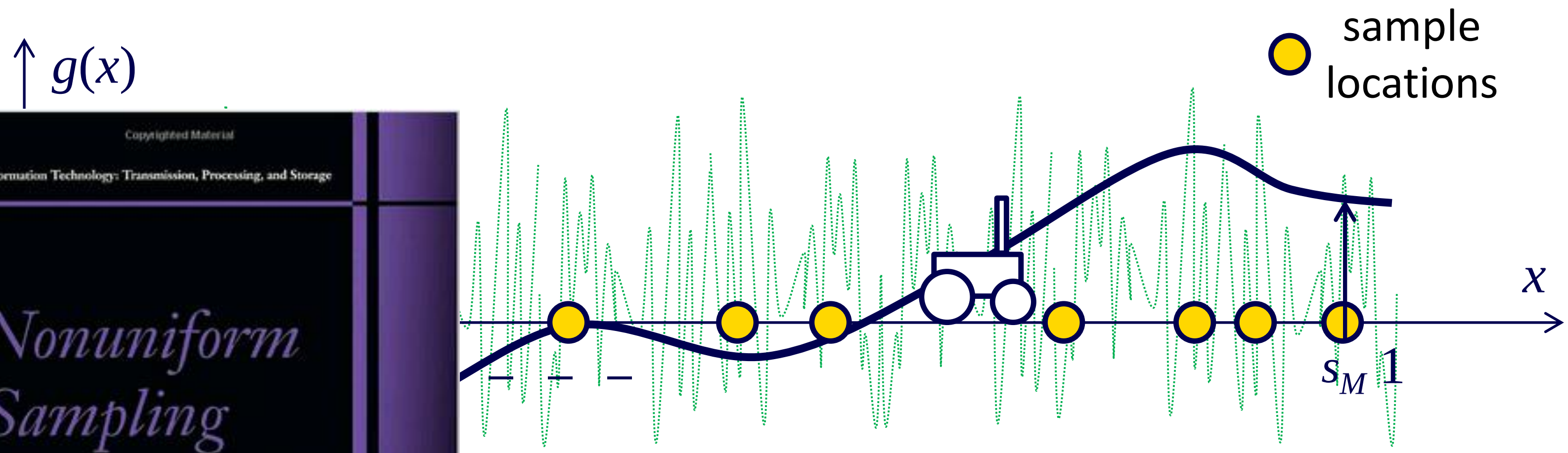
Key tradeoff: oversampling versus unawareness

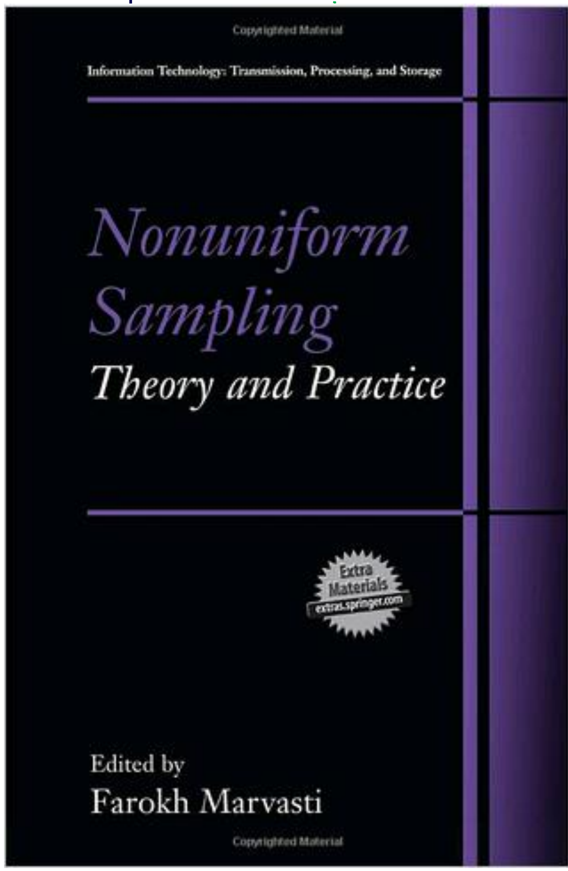


The mobile sensor (or scattered sensors) collects a lot of readings – there is oversampling above the known bandwidth

With a lot of noise affected measurements, without the knowledge of noise distribution and with unknown sample locations, can the field be estimated with a guarantee on accuracy (SNR)?

Location awareness is not there in prior art





a bandlimited field such that

obtained from a renewal process but they are unknown

mobile sensor trajectories, the distribution governing $S_1, S_2, \ldots$

listribution, with zero-mean and finite variance, is unknown

nality) of the field $g(x)$ is known

◊ The field does not evolve or change with time during the measurement process

Related work

- ◊ Simultaneous localization and mapping (SLAM) [**Montemerlo-Thrun-Koller-Wegbreit'2002, …**]
- ◊ Recovery of (narrowband) discrete-time bandlimited signals from samples taken at unknown locations **[Marziliano and Vetterli'2000]**
- ◊ Recovery of a bandlimited signal from a finite number of ordered nonuniform samples at unknown sampling locations **[Browning'2007]**.
- ◊ Estimation of periodic bandlimited signals in the presence of random sampling location under the following model **[Nordio, Chiasserini, and Viterbo'2008]**
    - Estimation of bandlimited signal from noisy samples on a location set obtained by random perturbation of equi-spaced deterministic grid

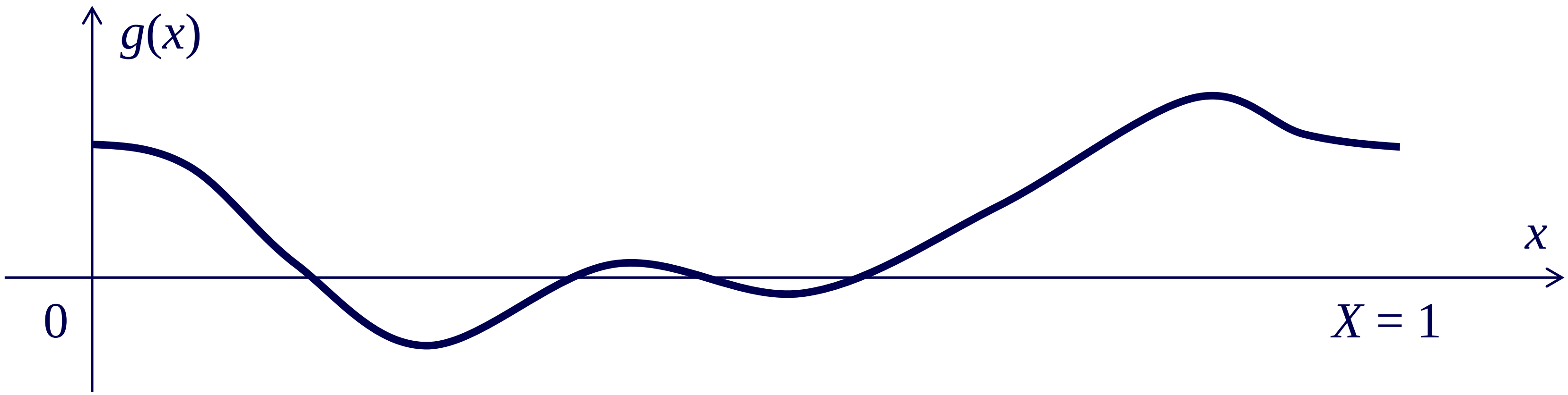
Organization

- ◊ Field model, sampling models, and distortion
- ◊ Field estimation with mobile sensor sampling on an unknown renewal process
- ◊ Field estimation with scattered smart dust sensor network
- ◊ Conclusions and open questions

Organization

- ◊ Field model, sampling models, and distortion
- ◊ Field estimation with mobile sensor sampling on an unknown renewal process
- ◊ Field estimation with scattered smart dust sensor network
- ◊ Conclusions and open questions

Field model

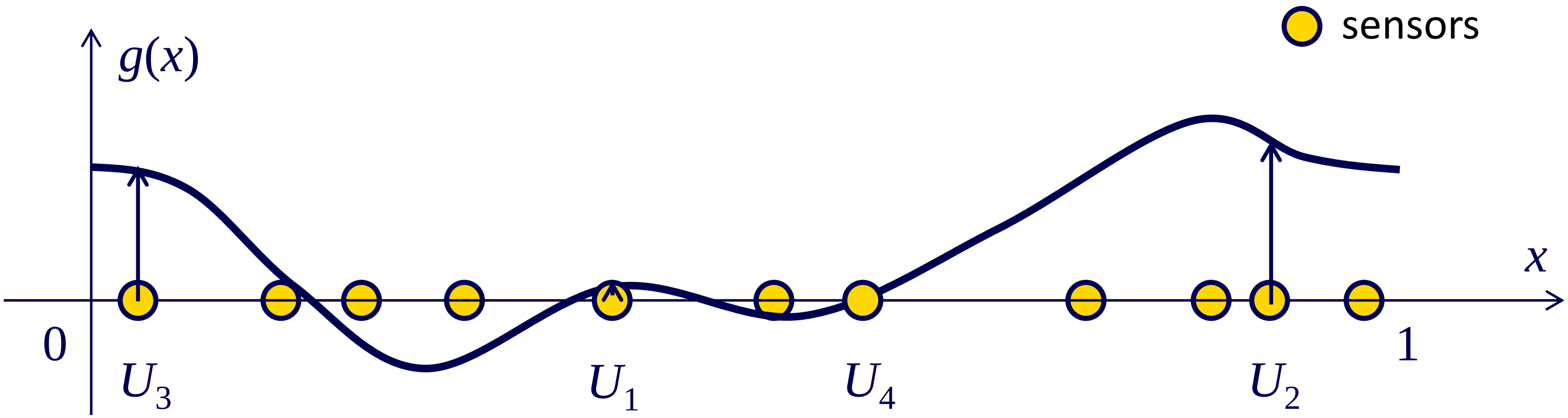


We assume that a periodic extension of the field $g(x)$ is bandlimited, that is, $g(x)$ is given by a finite number of Fourier series coefficients, (WLOG) $|g(x)| \leq 1$, and $X = 1$

$$g(x) = \sum_{k=-b}^{b} a[k] \exp(j2\pi kx)$$

The bandwidth will be assumed to be known. A polynomial basis can also be used to model smooth fields, and analytical treatment will be similar

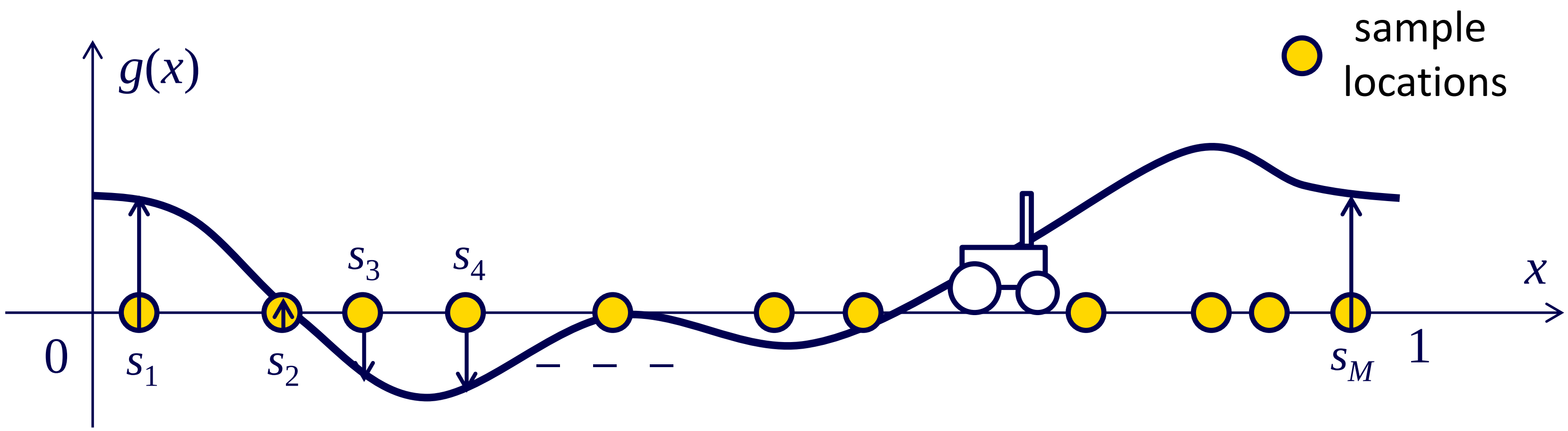
Location-unawareness in scattered sensors



Location-unaware mobile-sensor samples the bandlimited field at unknown locations such that

- The locations $U_1, U_2, \ldots$ are obtained from uniformly distributed random variables
- Uniform distribution models the random scattering phenomenon

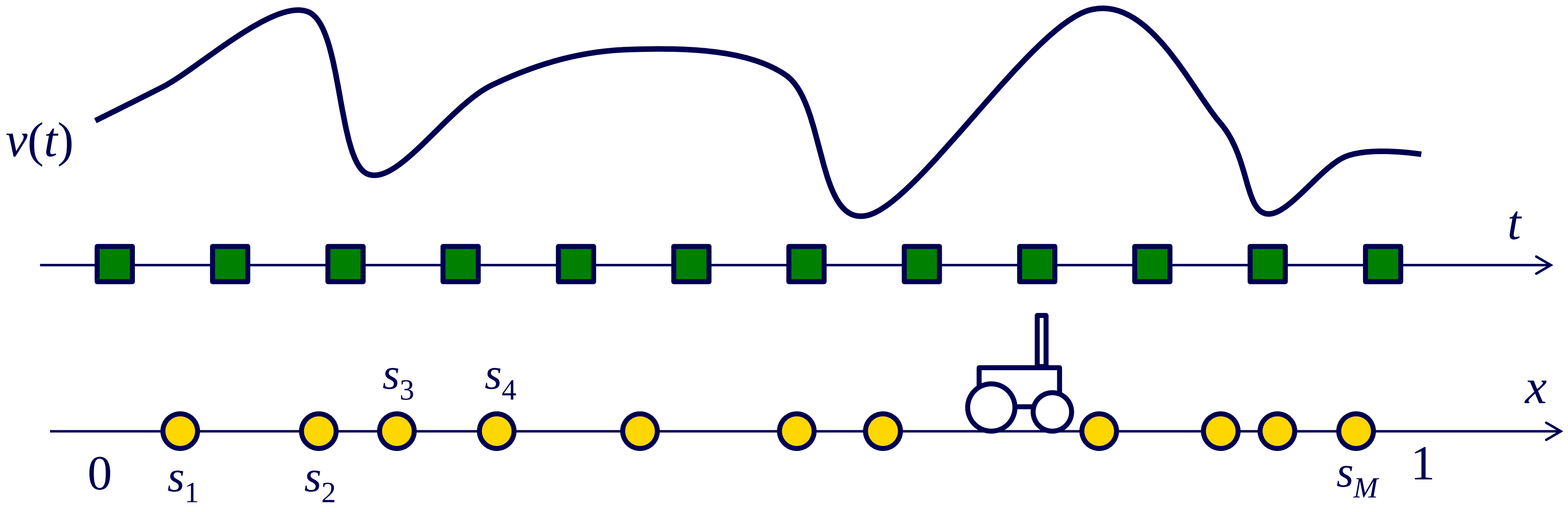
Location-unawareness in mobile sensing



Location-unaware mobile-sensor samples the bandlimited field at unknown locations such that

- The locations $S_1, S_2, \ldots$ are obtained from an unknown renewal process. The distribution governing $S_1, S_2, \ldots$ is unknown
- $\mathbb{E}(X_1) = 1/n$, where n is large (oversampling)
- It is assumed that $nX_i := n(S_i - S_{i-1})$ is bounded in $(0, \lambda]$

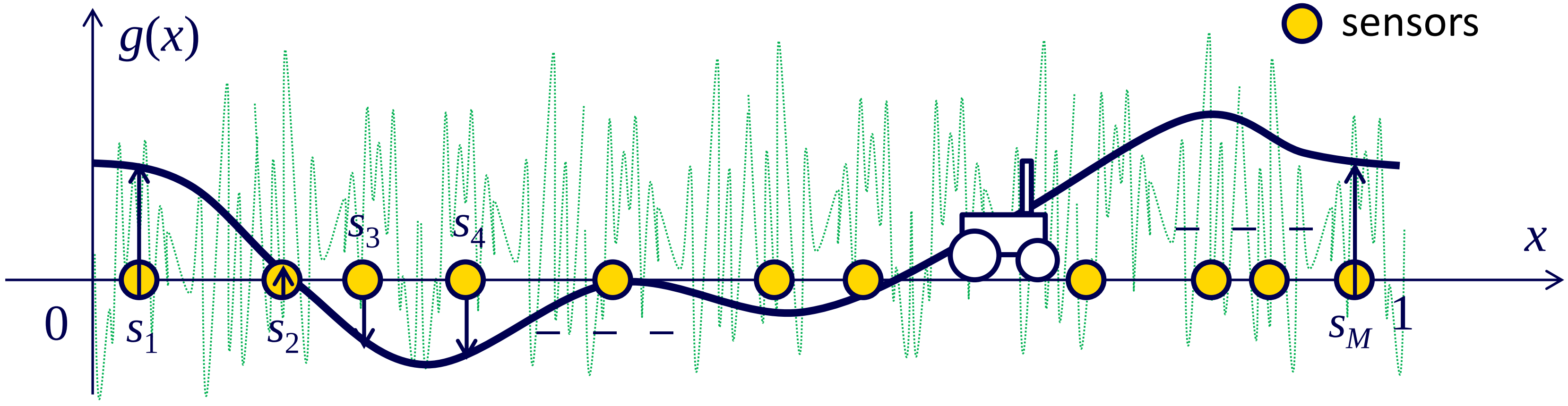
Comments on the renewal process model



A mobile sensor with variable speed profile $v(t)$ and uniform sampling rate in time will result in irregularly spaced (unknown) sampling locations

- $\mathbb{E}(X_1) = 1/n$, where n is large, would imply that the sampling rate is high in time
- $nX_i = n(S_i - S_{i-1})$ is bounded in $(0, \lambda]$ implies that the maximum speed of the sensor is finite

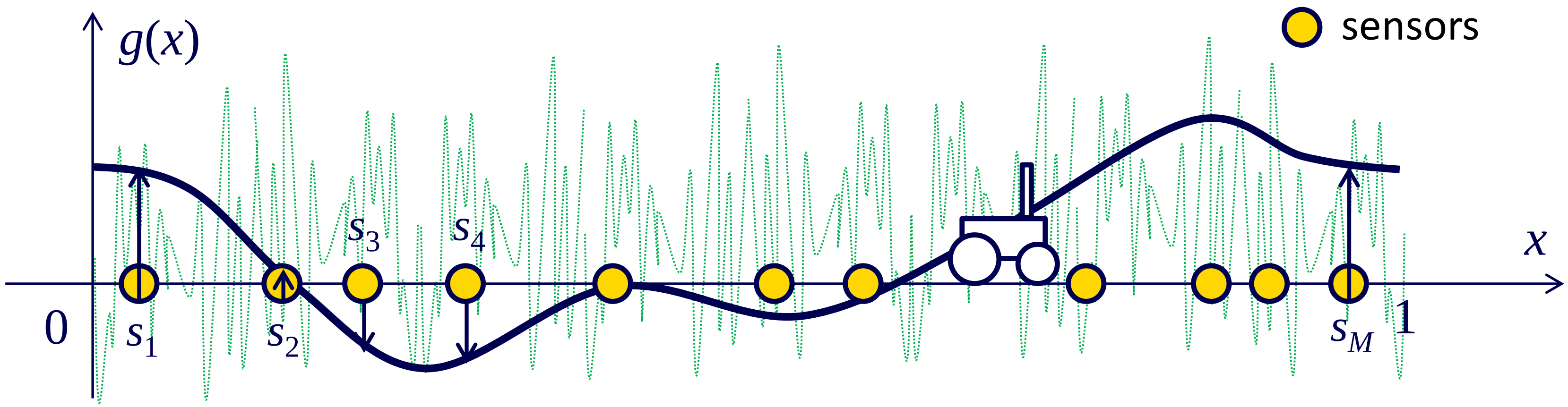
Noise model



The additive measurement-noise is independent, zero-mean, and independent of the spatial sampling process

- $W(S_1), W(S_2), \ldots, W(S_M), X_1, X_2, \ldots, X_M$, that is, noise and inter-sample distances are independent
- The variance (power) of the measurement-noise is finite

Observations made and distortion criterion



$Y(S_1) = g(S_1)+W(S_1)$, $Y(S_2) = g(S_2)+W(S_2)$, …, $Y(S_M) = g(S_M)+W(S_M)$ is collected without the knowledge of $(S_1, S_2, \ldots, S_M)$

We wish to estimate $g(x)$ and measure the performance of estimate against the average mean-squared error. If $\hat{G}(x)$ is the estimate with Fourier series $A[k]$ then

$$D := \mathbb{E}\left[\int_0^1 \left|\widehat{G}(x) - g(x)\right|^2 \mathrm{d}x\right] = \sum_{k=-b}^{b} \mathbb{E}\left[|A[k] - a[k]|^2\right]$$

Organization

- ◊ Field model, sampling models, and distortion
- ◊ Field estimation with mobile sensor sampling on unknown renewal process
- ◊ Field estimation with scattered smart dust sensor network
- ◊ Conclusions and open questions

Estimate of the spatial field's Fourier series

◊ Assume that the bandwidth b of the field is known. Recall the field model

$$g(x) = \sum_{k=-b}^{b} a[k] \exp(j2\pi kx)$$

◊ From noise-affected samples of the field at unknown locations S_1, S_2, …, the $(2b+1)$ Fourier series coefficients can be estimated as follows

$$\widehat{A}[k] = \frac{1}{M} \sum_{i=1}^{M} \{g(S_i) + W(S_i)\} \exp\left(-j2\pi k \frac{i}{M}\right)$$

◊ In other words, S_i is approximated as i/M in the above estimate. Compare with

$$a[k] = \int_0^1 g(x) \exp(-j2\pi kx) \mathrm{d}x$$

Interpretation of the estimate

$$a[k] = \int_0^1 g(x)\exp(-j2\pi kx)\mathrm{d}x$$

the coefficient of interest

$$\frac{1}{M}\sum_{i=1}^{M} g\left(\frac{i}{M}\right)\exp\left(-j2\pi k\frac{i}{M}\right)$$

its M-point Riemman approximation

$$\frac{1}{M}\sum_{i=1}^{M} g(S_i)\exp\left(-j2\pi k\frac{i}{M}\right)$$

$g(S_i) \approx g(i/M)$

$$\widehat{A}[k] = \frac{1}{M}\sum_{i=1}^{M}\{g(S_i) + W(S_i)\}\exp\left(-j2\pi k\frac{i}{M}\right)$$

estimate with location unaware noise-affected samples

◊ In other words, S_i is approximated as i/M in the above estimate. This may work well since, loosely speaking, $S_i = X_1 + X_2 + \ldots + X_i$ is composed of independent and identically distributed random variables with mean $1/n$

Main result: $\hat{A}[k]$ has a distortion of $O(1/n)$

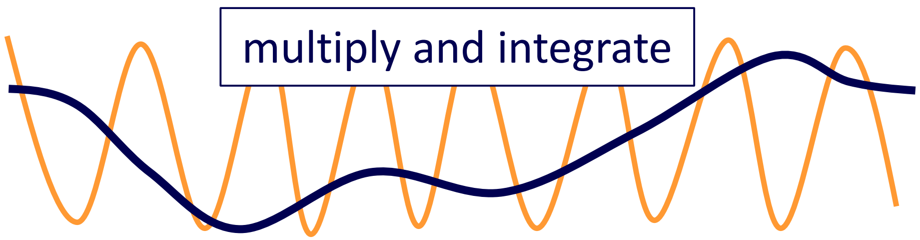
Let $\widehat{A}[k] = \frac{1}{M}\sum_{i=1}^{M}\{g(S_i) + W(S_i)\}\exp\left(-j2\pi k\frac{i}{M}\right)$ and S_i, $i = 1, 2, 3, \ldots$ be derived from a renewal process with mean $1/n$ and $nX_1 < \lambda$. The noise process is additive, zero-mean, finite-variance, and independent of sampling locations. Then,

$$\mathbb{E}\left[\left|\widehat{A}[k] - a[k]\right|^2\right] \leq \frac{C}{n}$$

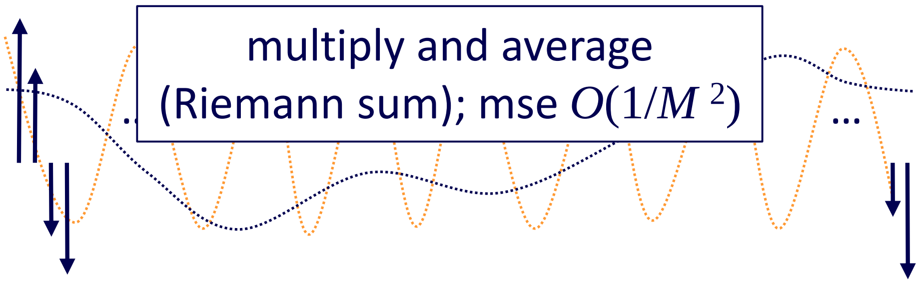
And, $D_{\text{gen}} \leq (2b+1)C/n$, since there are $(2b+1)$ Fourier coefficients to be estimated **[Kumar'2016]**

$$D := \mathbb{E}\left[\int_0^1 \left|\widehat{G}(x) - g(x)\right|^2 \mathrm{d}x\right] = \sum_{k=-b}^{b}\mathbb{E}\left[|A[k] - a[k]|^2\right]$$

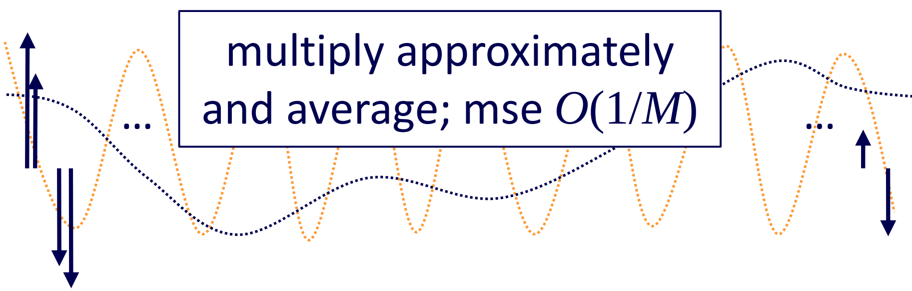
Three step coupling to justify the estimate



$$a[k] = \int_0^1 g(x) \exp(-j2\pi kx) \mathrm{d}x$$



$$\frac{1}{M} \sum_{i=1}^{M} g\left(\frac{i}{M}\right) \exp\left(-j2\pi k \frac{i}{M}\right)$$



$$\frac{1}{M} \sum_{i=1}^{M} g(S_i) \exp\left(-j2\pi k \frac{i}{M}\right)$$

Ignore the noise;
mse $O(1/M)$)

$$\widehat{A}[k] = \frac{1}{M} \sum_{i=1}^{M} \{g(S_i) + W(S_i)\} \exp\left(-j2\pi k \frac{i}{M}\right)$$

Main result

If the bandwidth of the field $g(x)$ is known, an estimate $\hat{G}(x)$ of measurement-noise affected spatial field samples $Y(S_1), Y(S_2), \ldots, Y(S_M)$ at unknown renewal process generated locations is developed which has a distortion given by

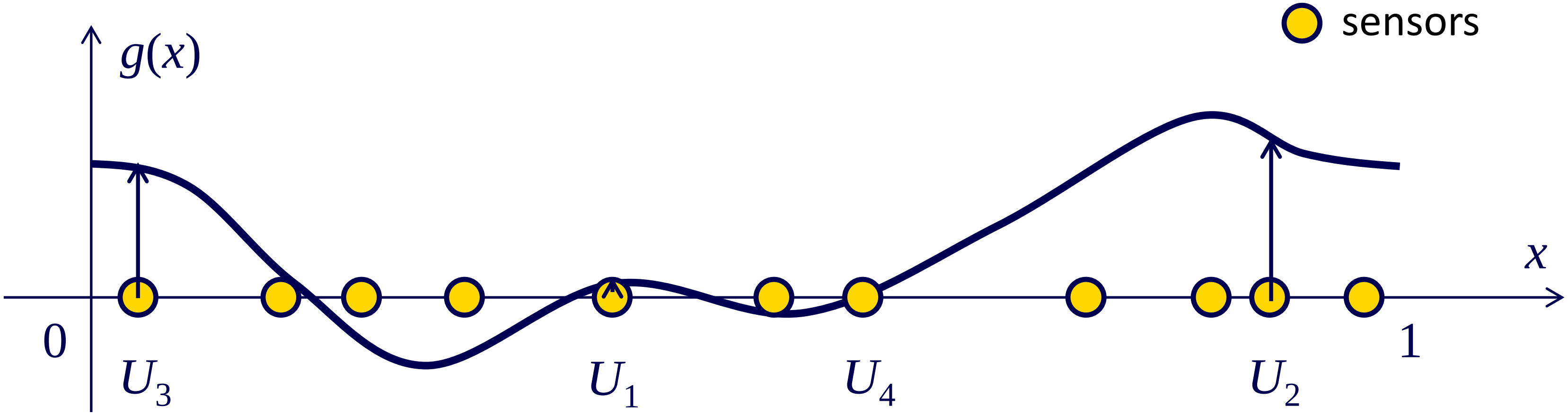
$$\mathbb{E}\left(|\hat{G}(x) - g(x)|^2\right) = O(1/n)$$

A consequence of unknown renewal distribution on inter-sample distances ensures that our field estimate is universal

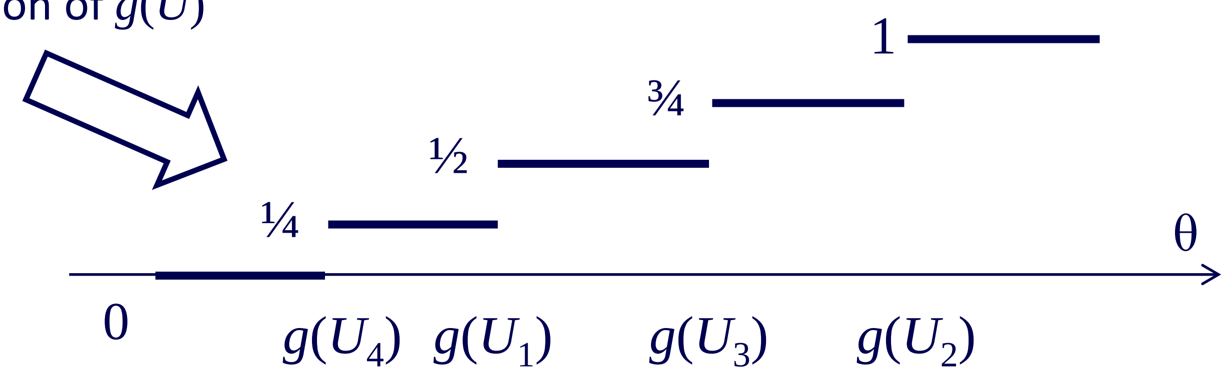
Organization

- ◊ Field model, sampling models, and distortion
- ◊ Field estimation with mobile sensor sampling on unknown renewal process
- ◊ Field estimation with scattered smart dust sensor network
- ◊ Conclusions and open questions

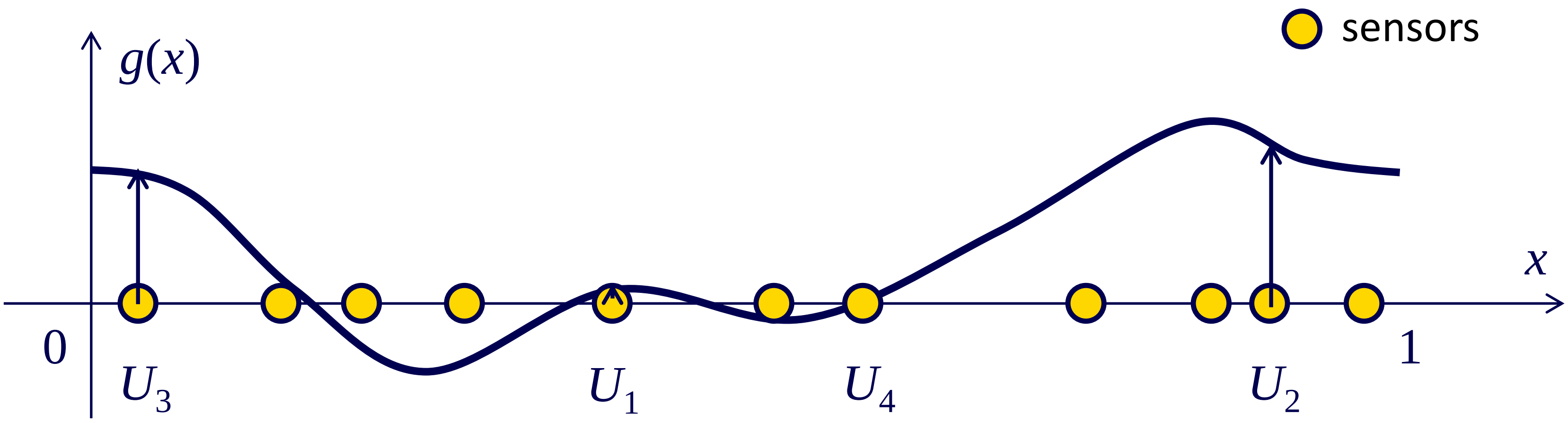
Impossible to infer $g(x)$ from $g(U_1)$, $g(U_2)$, …



Effectively, we are just collecting the empirical distribution or histogram of $g(U_1)$, $g(U_2)$, …, $g(U_n)$ and, in the limit of large n, the task is to estimate $g(x)$ from the distribution of $g(U)$



Impossible to infer $g(x)$ from $g(U_1)$, $g(U_2)$, …



Consider the statistic

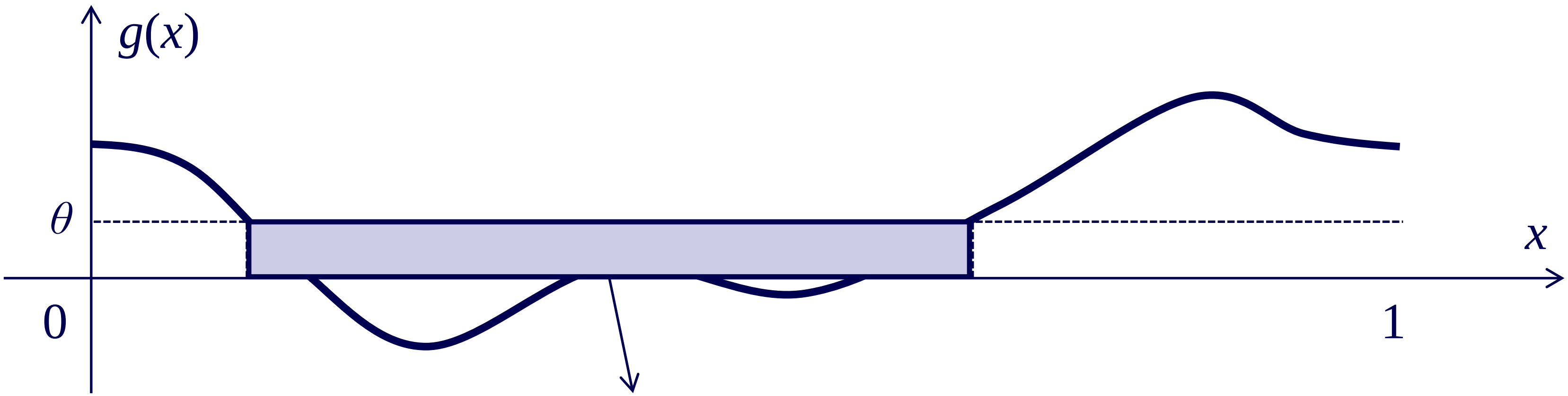
$$F_{g,n}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}[g(U_i) \leq \theta]$$

◊ Then $F_{g,n}(\theta)$, x in set of reals and $g(U_1)$, $g(U_2)$, …, $g(U_n)$ are statistically equivalent

◊ By the Glivenko Cantelli theorem, $F_{g,n}(\theta)$ converges almost surely to Prob$(g(U) \leq \theta)$ for each θ in set of real numbers **[van der Vaart'1998]**

Intuition into the limit of $F_{g,n}$

So what does Prob($g(U) \leq \theta$), for x in set of real numbers, looks like?

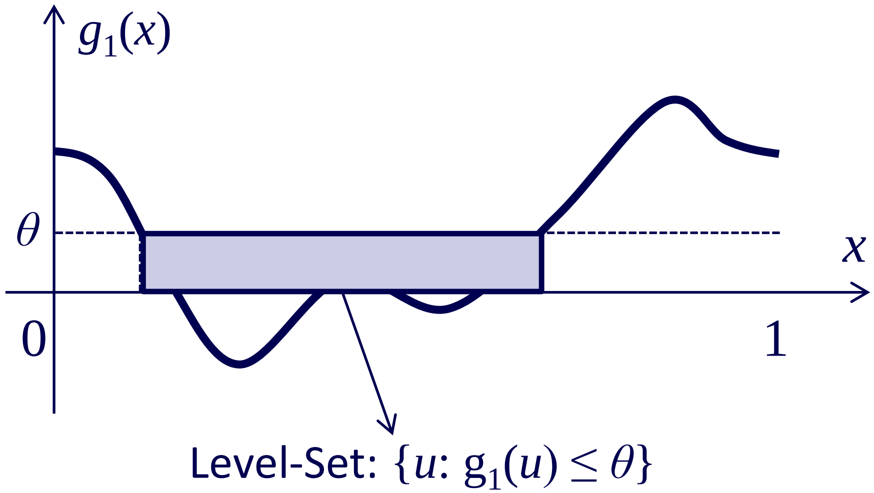


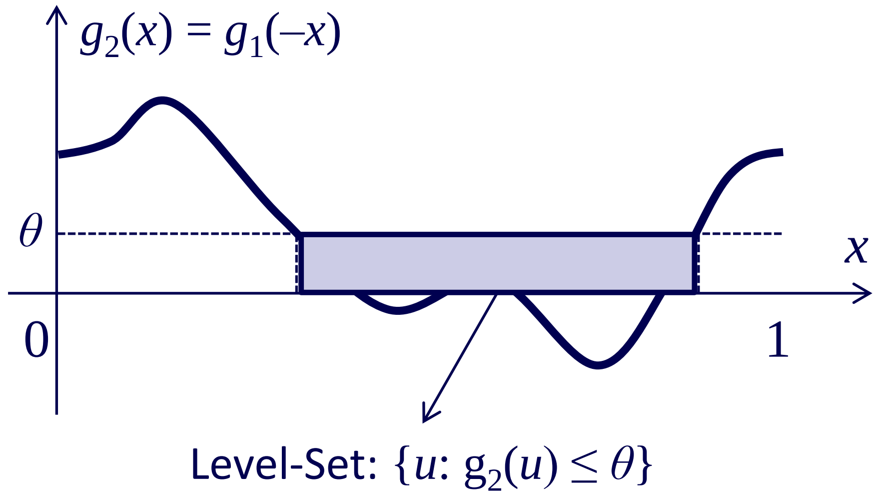
Level-Set: $\{u\text{: } g(u) \leq \theta\}$

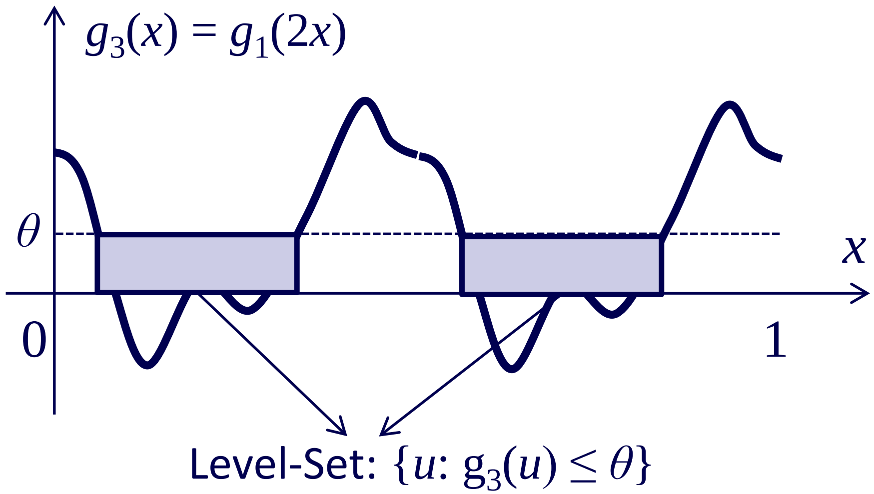
◊ Prob($g(U) \leq \theta$) for each θ is the probability of U belonging in the level-set.

Thus, it is simply the length (measure) of level-set

Prob($g(U) \leq \theta$) does not give $g(x)$ uniquely







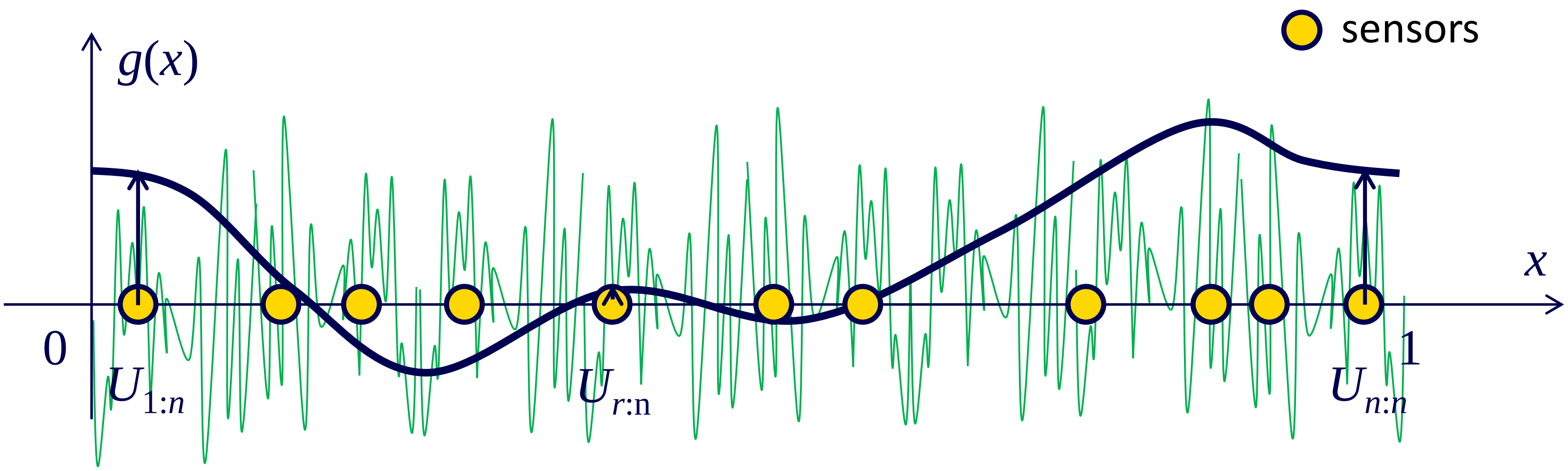
◊ The length (measure) of the level-sets is the same in the three cases for every θ

◊ Thus, the observed samples alone do not lead to a unique reconstruction of the field

Ordered samples in smart-dust sensor network

◊ If the **order** (left to right) of sample locations in the smart-dust sensor network is known and field is affected by independent measurement noise, a consistent estimate $\hat{G}(x)$ for the field of interest can be understood as follows

◊ In this setup, $g(U_{1:n}), g(U_{2:n}), \ldots, g(U_{n:n})$ are available



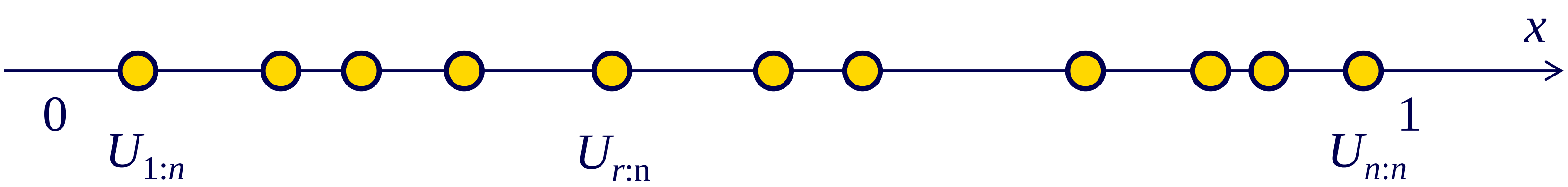
Ordered uniformly distributed variables

◊ Ordered uniformly distributed random variables are related to a Poisson process

◊ A Poisson process with rate n will realize M points in the interval $[0,1]$

◊ Conditioned on $M = m$, the renewal process $S_1, S_2, \ldots, S_m$ will be distributed as ordered uniformly distributed random variables $U_{1:m}, U_{2:m}, \ldots, U_{m:m}$

◊ Finally $M/n = 1$ with high probability. It is not surprising that the $O(1/n)$ distortion holds true for this setup as well



The Poisson setup and the uniform setup are related but not the same

Main result: $\hat{A}_{\text{scat}}[k]$ has a distortion of $O(1/n)$

Theorem: Let Fourier series coefficient estimates for $g(x)$ be obtained as

$$\widehat{A}_{\text{scat}}[k] = \frac{1}{n}\sum_{i=1}^{n}\{g(U_{i:n}) + W(U_{i:n})\}\exp\left(-j2\pi k\frac{i}{n}\right)$$

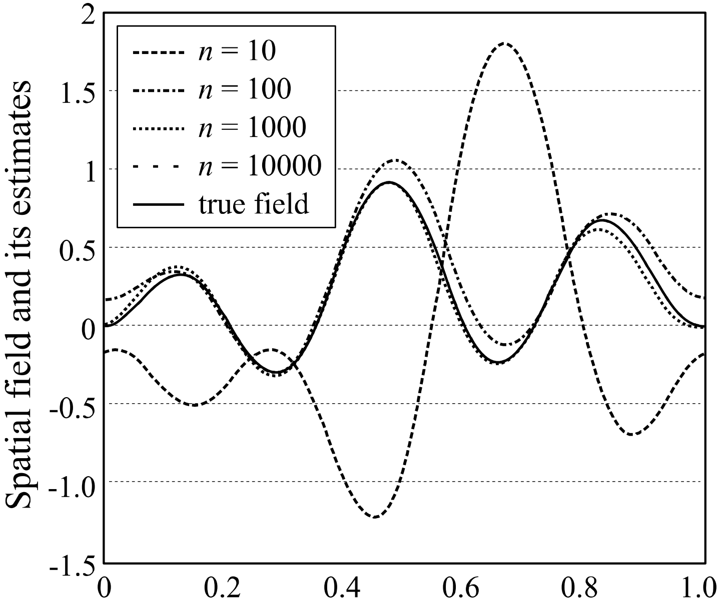
Then the average mean-squared error (distortion) between $g(x)$ and its estimate $\hat{G}_{\text{scat}}(x)$ with Fourier series coefficients above is bounded by

$$\mathbb{E}\left[\int_0^1 |\widehat{G}_{\text{scat}}(x) - g(x)|^2 \mathrm{d}x\right] \le (2b+1)\left[\frac{\pi^2 b^2}{n} + O\left(\frac{1}{n\sqrt{n}}\right) + \frac{16\pi^2 b^2}{n^2} + \frac{\sigma^2}{n}\right]$$
$$= O\left(\frac{1}{n}\right)$$

where is the σ^2 variance of the additive noise **[Kumar'2015]**

Simulation results

Fourier series coefficients are selected by random realizations of $\mathrm{Unif}[-1,1]$ variable and bandwidth parameter was selected as $b = 3$

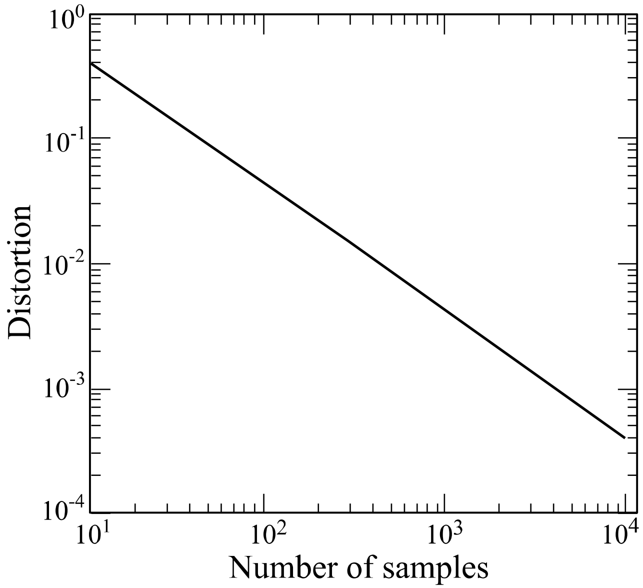


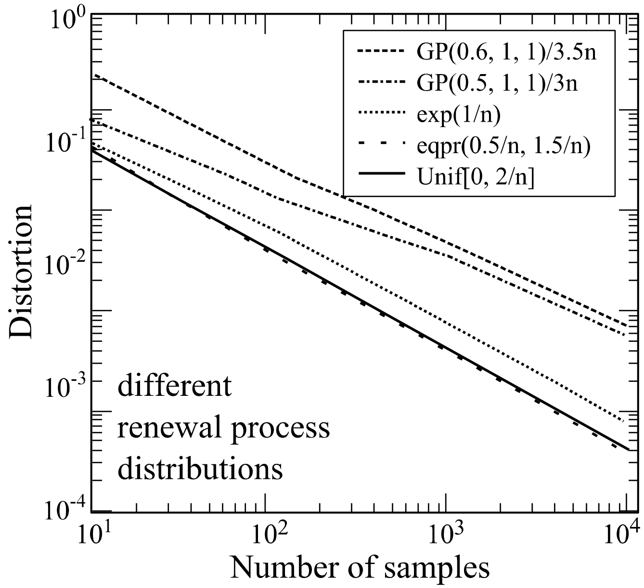
n

Renewal distribution	Noise distribution
$\mathrm{Unif}[0, 2/n]$	$\mathrm{Unif}[-1, 1]$

Simulation results

Fourier series coefficients are selected by random realizations of Unif[–1,1] variable and bandwidth parameter was selected as $b = 3$. Various renewal distributions were used to obtain unknown locations of samples





Noise distribution

Unif[-1, 1]

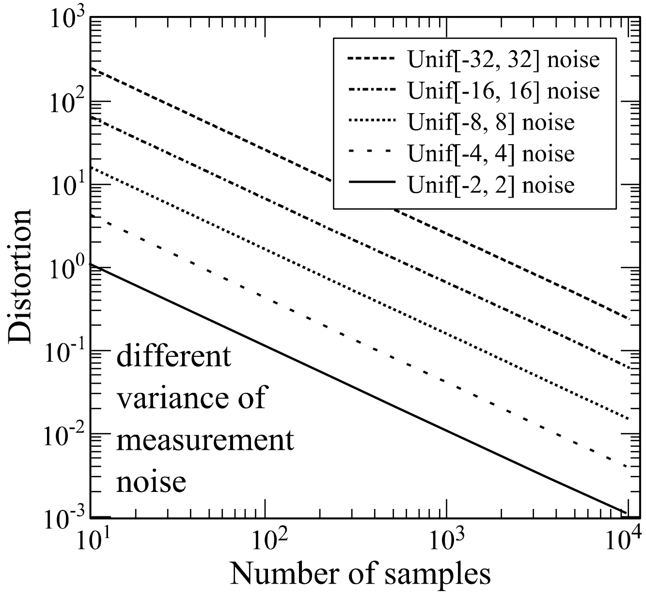
$$f_X(x) = \frac{1}{\sigma}\left(1 + \xi\frac{x-\mu}{\sigma}\right)^{-1/(\xi+1)}, \quad x \geq \mu$$

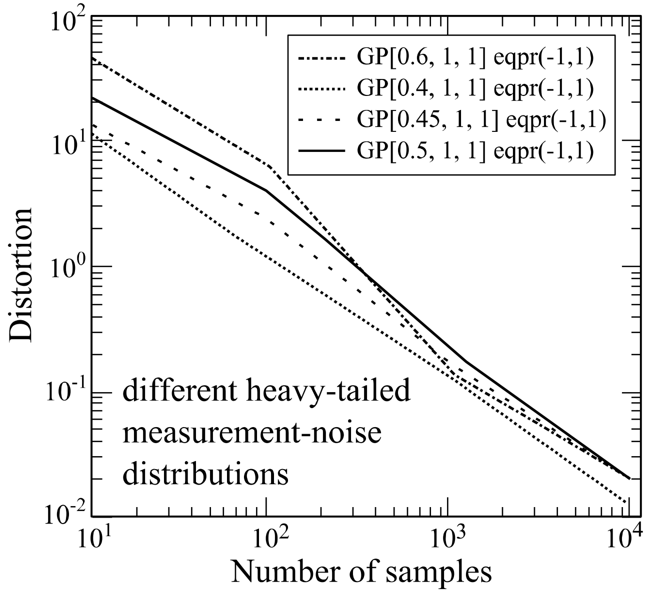
$$\text{mean} = \mu + \frac{\sigma}{1-\xi}; \quad \text{variance is finite for} \xi < 1/2$$

Generalized Pareto
$GP(\xi, \mu, \sigma)$
distribution

Simulation results

Fourier series coefficients are selected by random realizations of $\mathrm{Unif}[-1,1]$ variable and bandwidth parameter was selected as $b = 3$. Various noise distributions were used





Renewal distribution

$\mathrm{Unif}[0, 2/n]$

Organization

- ◊ Field model, sampling models, and distortion
- ◊ Field estimation with mobile sensor sampling on unknown renewal process
- ◊ Field estimation with scattered smart dust sensor network
- ◊ Conclusions and open questions

Conclusions

- ◊ Recovery of bandlimited fields, with samples taken on sampling locations governed by unknown renewal process was shown. It was shown that the average mean-squared error decreases as $O(1/n)$ where n is the average number of field samples collected. The estimation process also accounts for additive independent noise; the noise statistics need not be known.
- ◊ The obtained estimate and ensuing results are universal, and they do not depend either on the renewal process distribution or on the noise distribution

Future work in location-unaware sampling

- Estimates are not minimum risk. Or, techniques for finding Maximum likelihood estimates will be useful. It will also be interesting to study different measures of risk (than mean-square)
- It is expected that $O(1/n)$ distortions obtained are optimal. It will be interesting to show the same
- Extension of these results to more classes of fields (FRI, finite-support, orthogonal spaces, non-bandlimited fields)
- What is the effect of quantization?
- In two dimensions: fusion of data from multiple mobile sensors, which sample the field along independent (disjoint) trajectories needs to be studied
- Connections if any with SLAM!

Publications

1. A. Kumar, "Bandlimited Spatial Field Sampling with Mobile Sensors in the Absence of Location Information", ISIT 2016 and submitted to IEEE Trans on Information Theory **(mobile-sensor based sensing)**
2. A. Kumar, "On Bandlimited Signal Reconstruction From the Distribution of Unknown Sampling Locations", IEEE Transactions on Signal Processing (**smart-dust based sensing**)