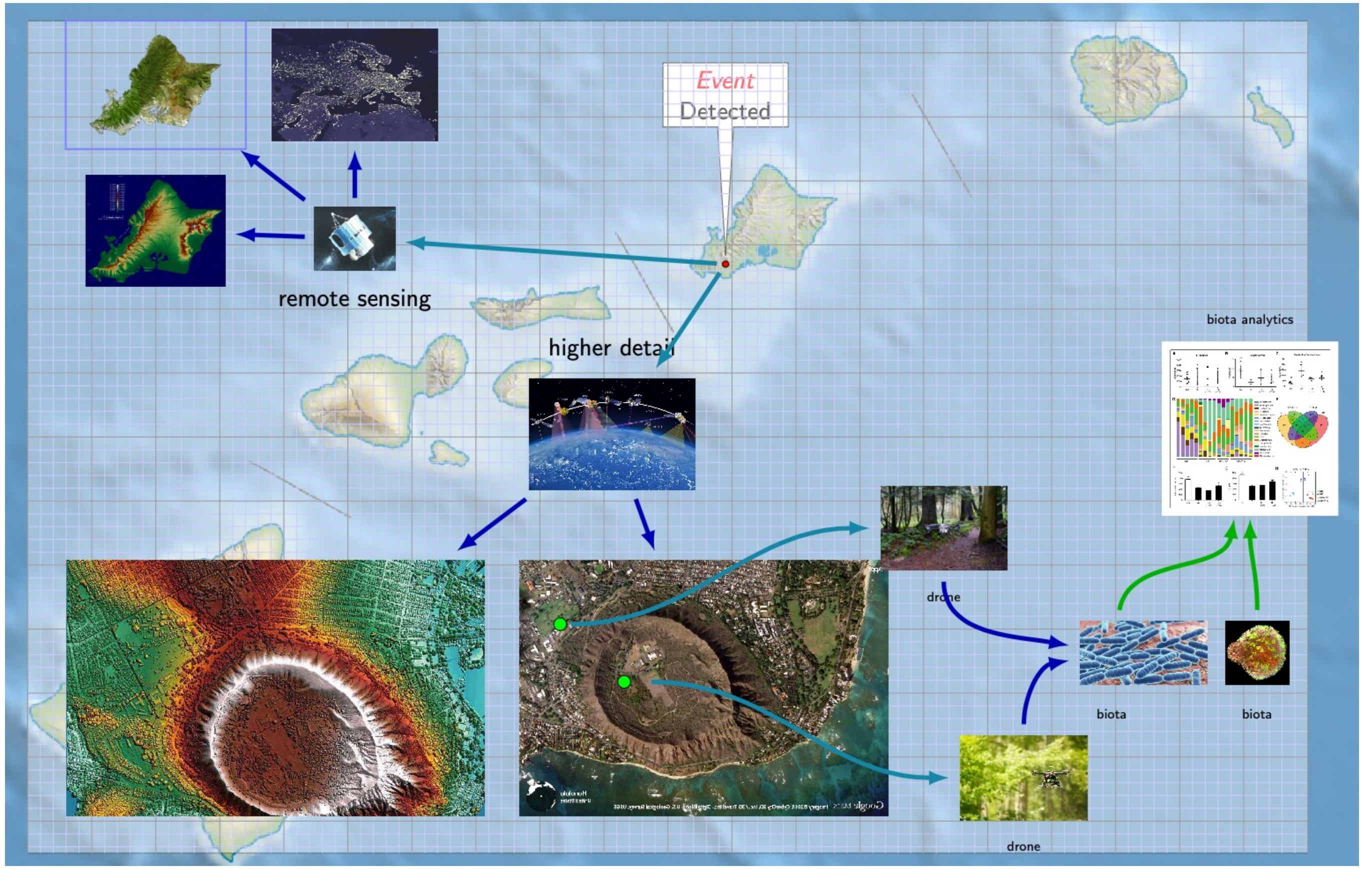
# Autonomous Sensing and Information Gathering

John W. Fisher III

Massachusetts Institute of Technology

# Data Fusion, Information Gathering, Uncertainty Quantification and Risk

- Multiple sensing modalities provide *complementary information* of the environment, but…
    - modeling complexity, computational resources, multi-physics relations, etc.
- Structured probabilistic models provide a consistent mathematical representation for data fusion (inference), uncertainty quantification, and risk.
- Bayesian optimal experiment design (BOED) is a natural framework for managing uncertainty, risk, and resource expenditures

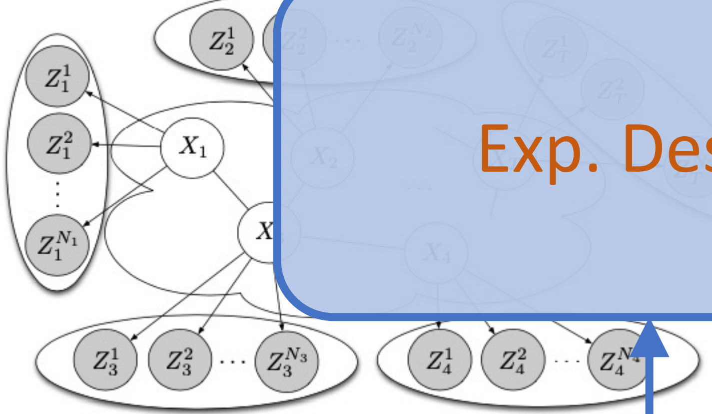


# Sequential Bayesian Optimal Experiment Design

Exp. Design

Execute Experiment

Inference

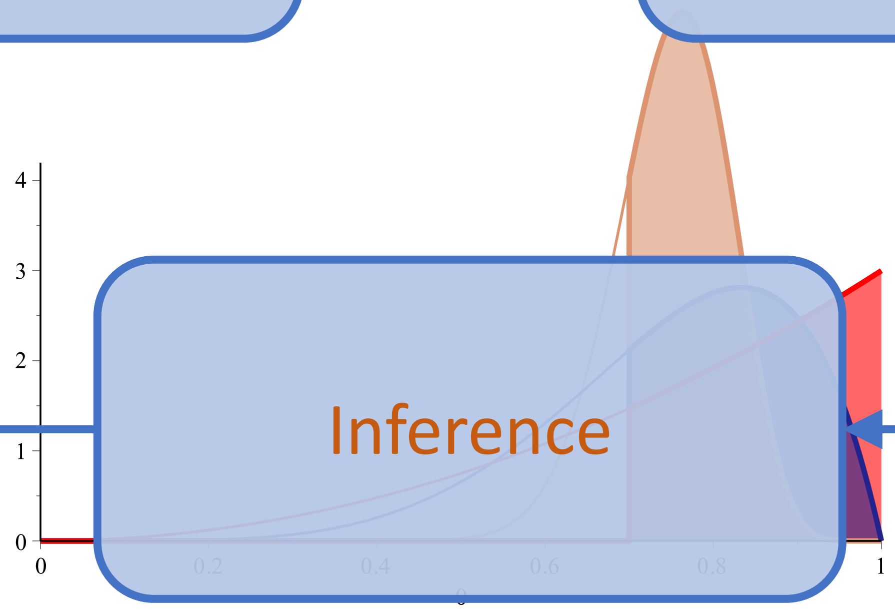


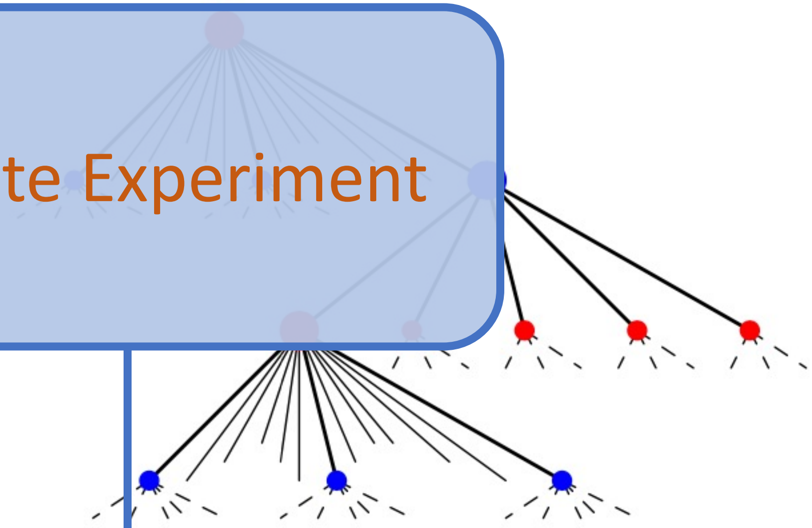
Probabilistic Graphical Models

- the graph encodes probabilistic relations
- analysis algorithms exploit structural properties

Inference, UQ and Risk

- estimates AND confidence
- Computation needed to *estimate* quantities of interest.





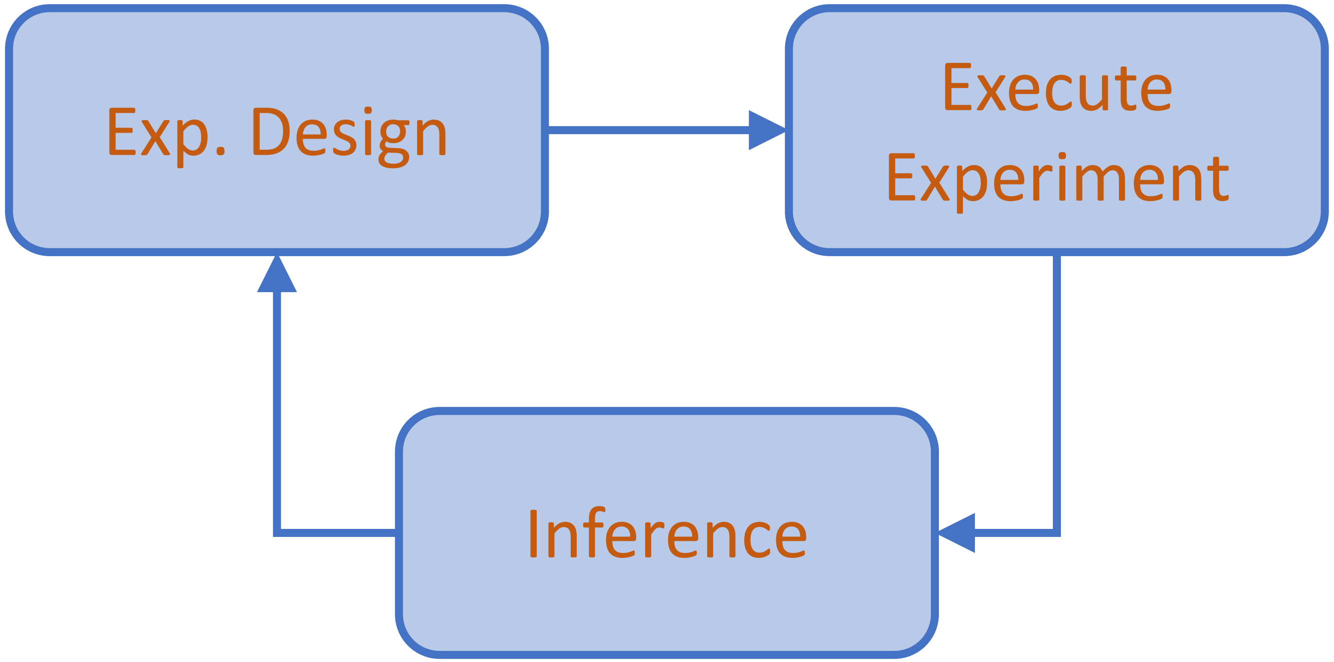
Information Planning

- uses PGM to assess information and risk
- incorporates resource expenditures

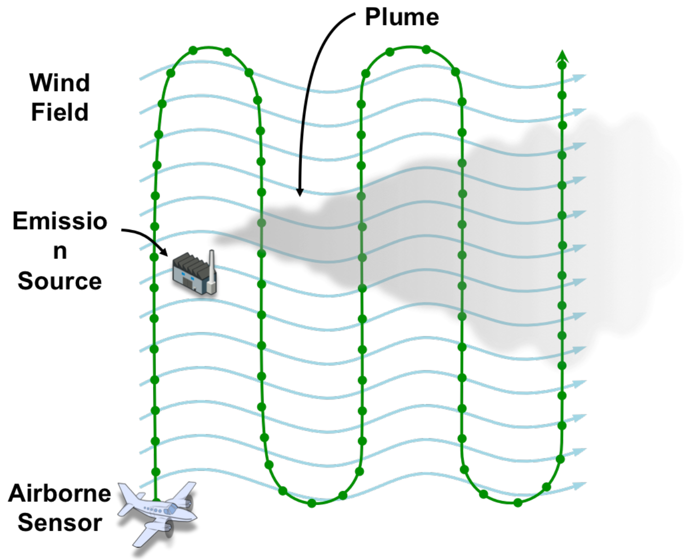
# Sequential Bayesian Optimal Experiment Design

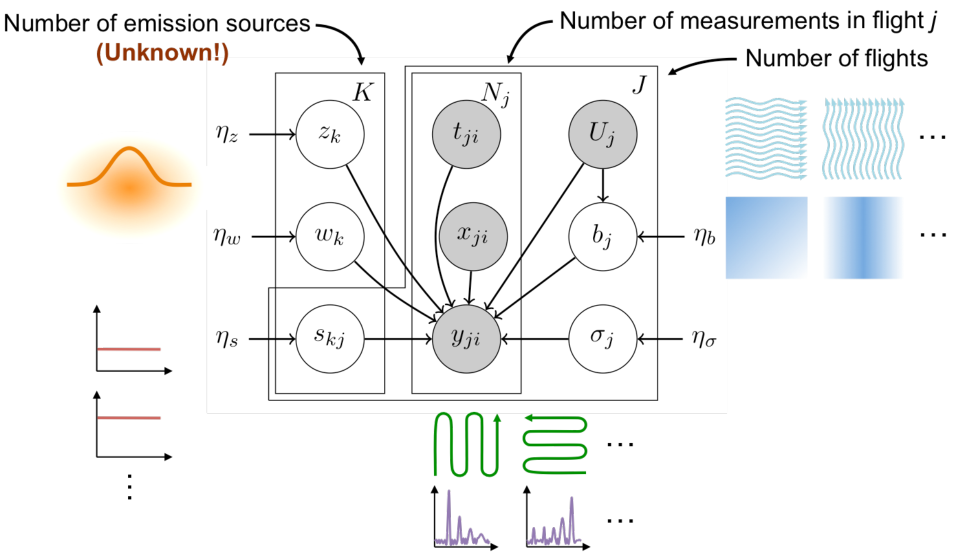
Iteratively select ***informative designs***

- Design: Value of information (VoI) analysis
- Execute: Collect sensor measurements
- Infer: Update beliefs/understanding



# Source Localization with Complex Transport




**Goal:** detect an *unknown* *number* of sources via airborne sensors

**Challenges:**

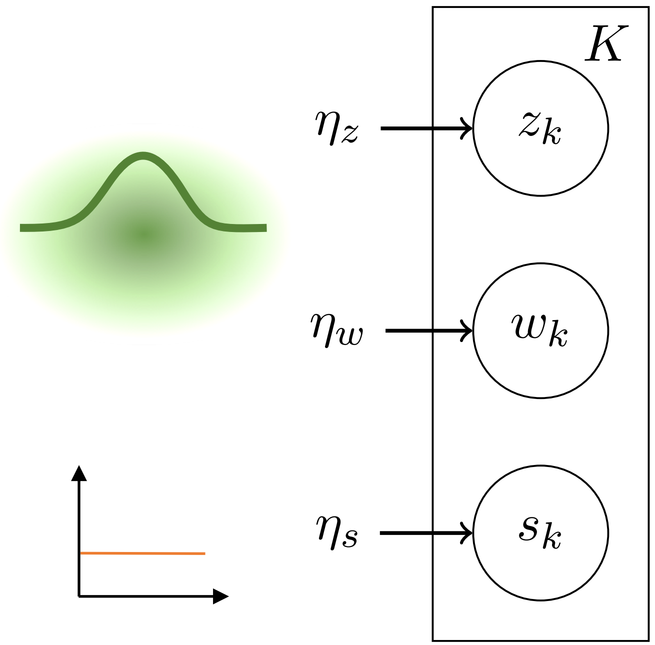
- Complex transport phenomenon
- Very low SNR
- Uncertain nuisance parameters matter

**Technical Approach:** PGM that combines computationally efficient propagation models, flexible source models, and BNP background models

- RJMCC inference for uncertainty in the *number* of sources
- Gaussian process background model (*irregular spatial sampling*)
- CPAB for stochastic transport from noisy wind measurements
- *Information-driven* multi-flight flight planning

# Probabilistic Graphical Model Representation

Number of emission sources
**(Unknown!)**



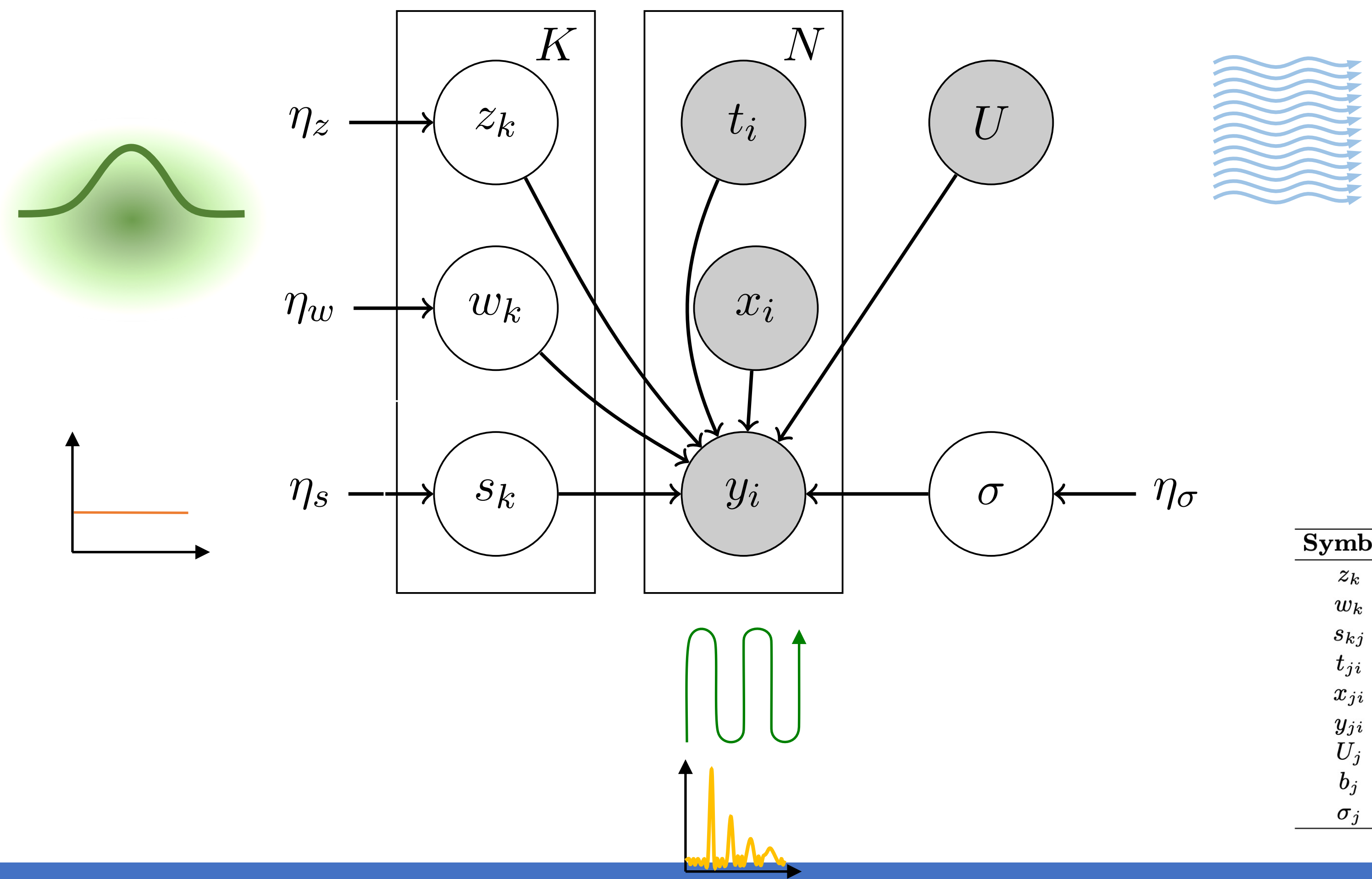

| Symbol | Definitions |
|---|---|
| $z_k$ | 2D mean location of source $k$ |
| $w_k$ | Location std. dev. ("width") of source $k$ |
| $s_{kj}$ | Emission rate of source $k$ in flight $j$ |
| $t_{ji}$ | Time of the $i$-th measurement in flight $j$ |
| $x_{ji}$ | Location of the $i$-th measurement in flight $j$ |
| $y_{ji}$ | Value of the $i$-th measurement in flight $j$ |
| $U_j$ | Wind field data for flight $j$ |
| $b_j$ | Background contributions to measured $\boldsymbol{y}_j$ |
| $\sigma_j$ | Sensor noise std. dev. in flight $j$ |

# Probabilistic Graphical Model Representation

Number of emission sources

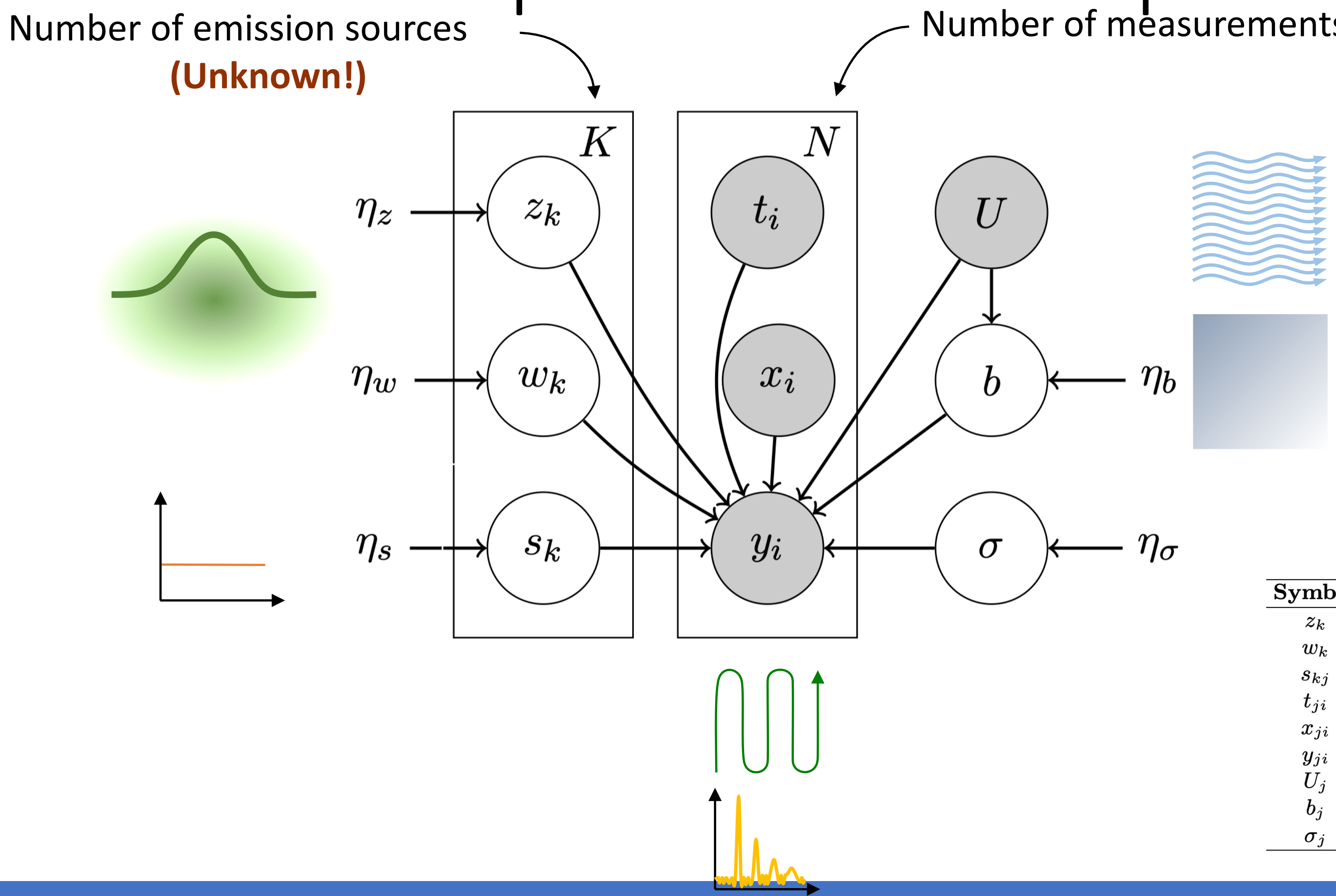
**(Unknown!)**

Number of measurements



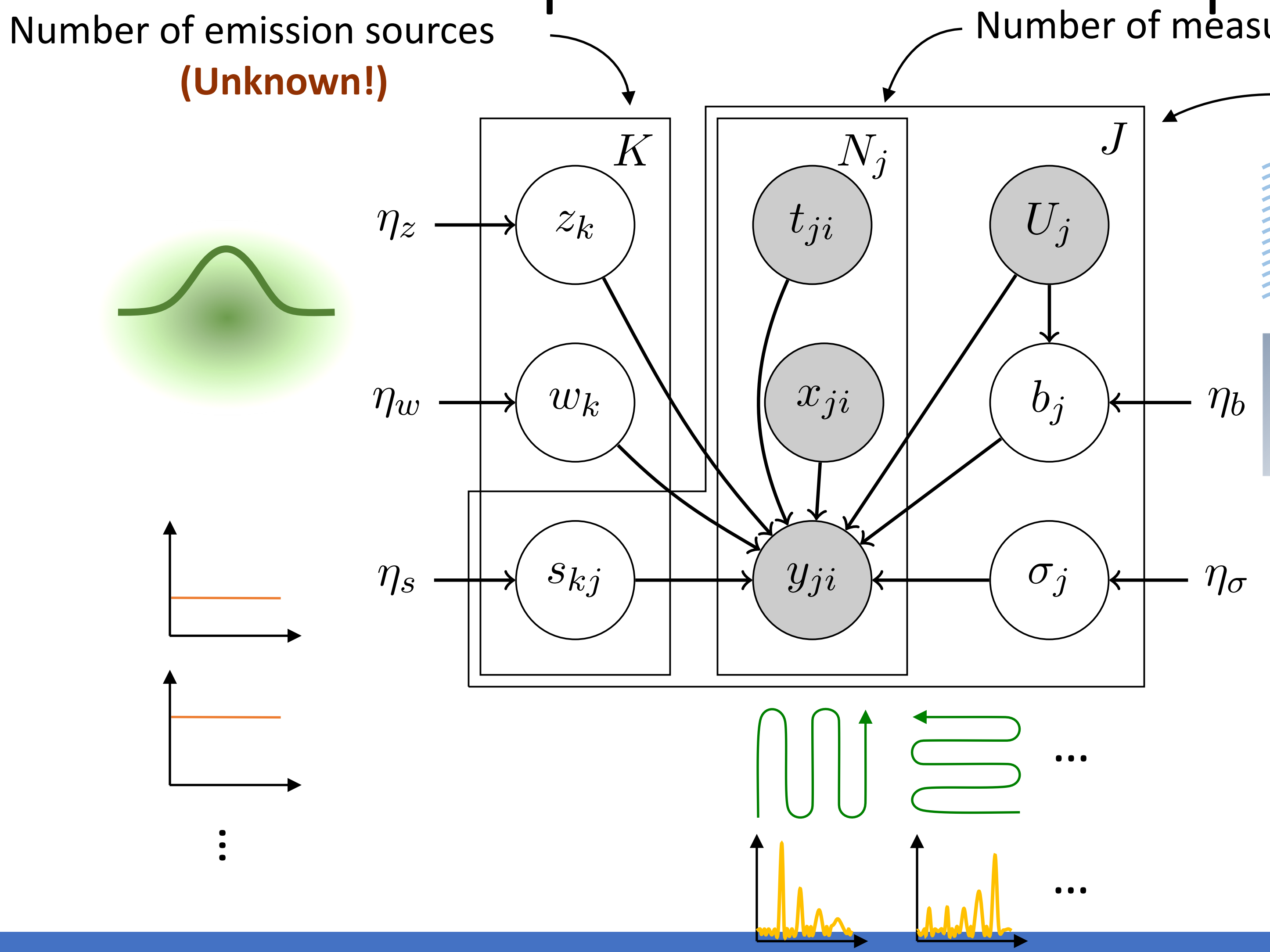

| Symbol | Definitions |
|---|---|
| $z_k$ | 2D mean location of source $k$ |
| $w_k$ | Location std. dev. ("width") of source $k$ |
| $s_{kj}$ | Emission rate of source $k$ in flight $j$ |
| $t_{ji}$ | Time of the $i$-th measurement in flight $j$ |
| $x_{ji}$ | Location of the $i$-th measurement in flight $j$ |
| $y_{ji}$ | Value of the $i$-th measurement in flight $j$ |
| $U_j$ | Wind field data for flight $j$ |
| $b_j$ | Background contributions to measured $\boldsymbol{y}_j$ |
| $\sigma_j$ | Sensor noise std. dev. in flight $j$ |

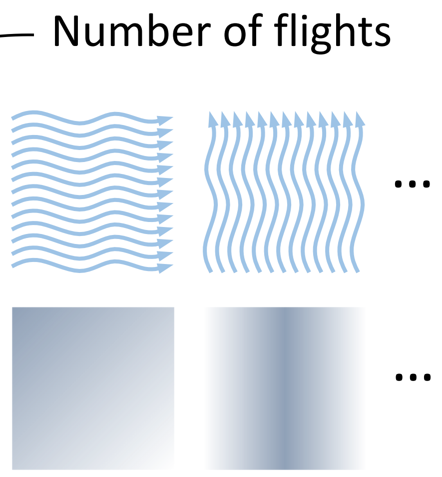
# Probabilistic Graphical Model Representation




| Symbol | Definitions |
|---|---|
| $z_k$ | 2D mean location of source $k$ |
| $w_k$ | Location std. dev. ("width") of source $k$ |
| $s_{kj}$ | Emission rate of source $k$ in flight $j$ |
| $t_{ji}$ | Time of the $i$-th measurement in flight $j$ |
| $x_{ji}$ | Location of the $i$-th measurement in flight $j$ |
| $y_{ji}$ | Value of the $i$-th measurement in flight $j$ |
| $U_j$ | Wind field data for flight $j$ |
| $b_j$ | Background contributions to measured $\boldsymbol{y}_j$ |
| $\sigma_j$ | Sensor noise std. dev. in flight $j$ |

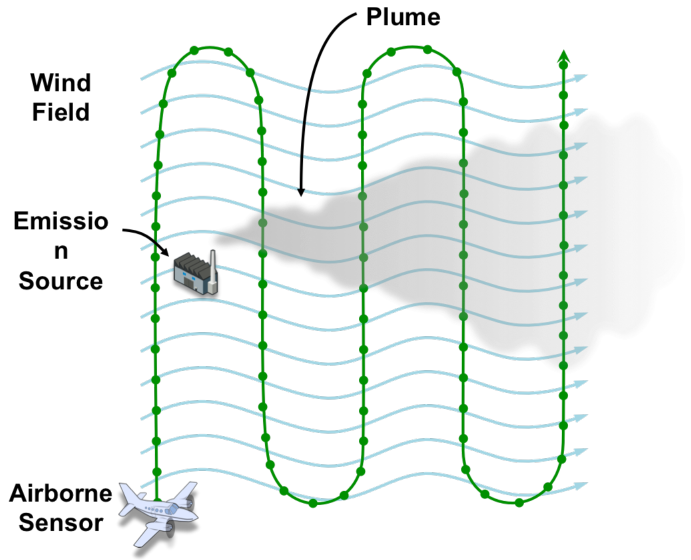
# Probabilistic Graphical Model Representation

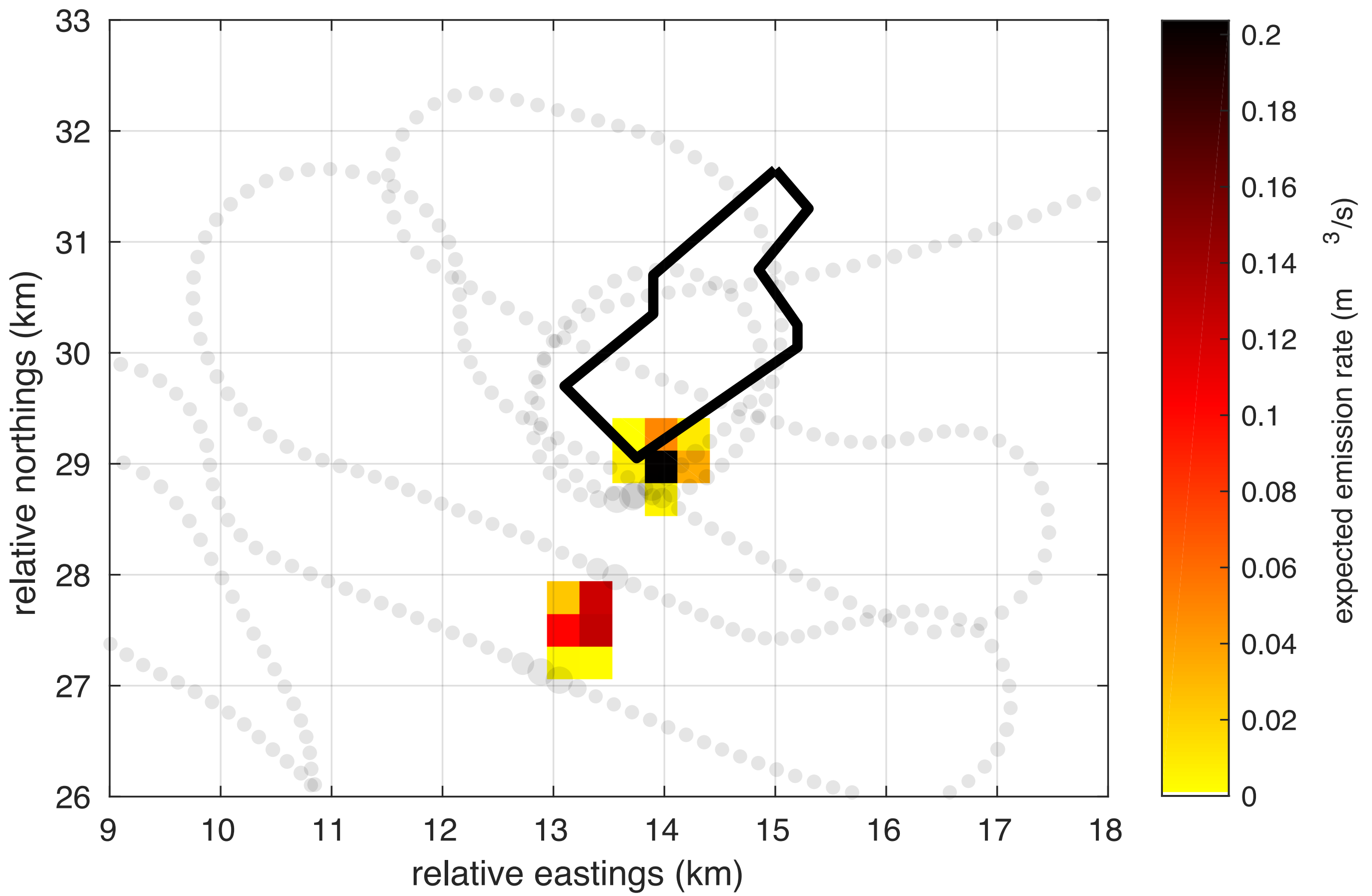



| Symbol | Definitions |
|---|---|
| $z_k$ | 2D mean location of source $k$ |
| $w_k$ | Location std. dev. ("width") of source $k$ |
| $s_{kj}$ | Emission rate of source $k$ in flight $j$ |
| $t_{ji}$ | Time of the $i$-th measurement in flight $j$ |
| $x_{ji}$ | Location of the $i$-th measurement in flight $j$ |
| $y_{ji}$ | Value of the $i$-th measurement in flight $j$ |
| $U_j$ | Wind field data for flight $j$ |
| $b_j$ | Background contributions to measured $\boldsymbol{y}_j$ |
| $\sigma_j$ | Sensor noise std. dev. in flight $j$ |

# Expected Emission Rate




**Expected Source Emissions Originating in Area $\mathcal{A}$:**

$$f(\boldsymbol{\theta};\ \mathcal{A}) = \sum_k s_{kj} \int_{\mathcal{A}} \mathcal{N}(z;\ z_k, \mathbf{I}w_k^2)\,\mathrm{d}z$$

# Bayesian Experimental Design with Variable Cost

Sue Zheng, Jason Pacheco, John W. Fisher III

# Bayesian Experimental Design

Unknown $X \sim p(x)$, Outcome $Y \sim p_a(y \mid x)$

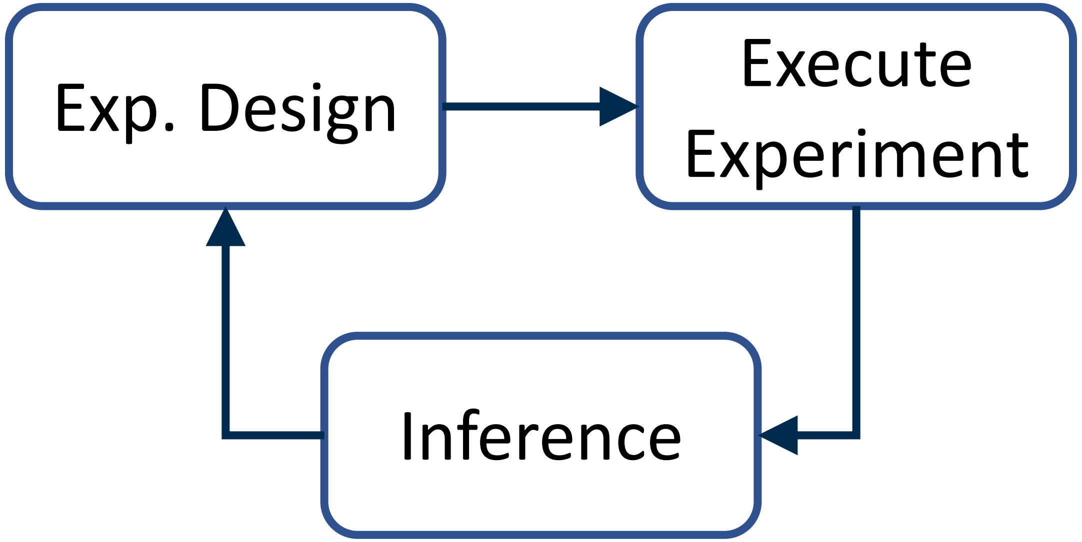
- Experimental Design:

$$a^* = \text{argmax}_a \, \mathrm{I}_a(X;Y) \mid D_t)$$

- Execute: Collect observations, $Y = y$

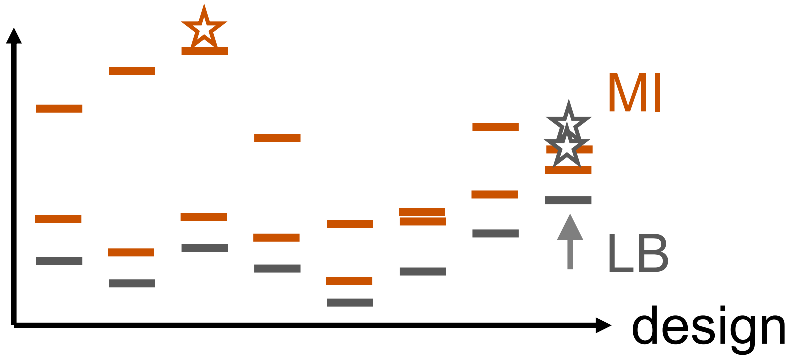
- Infer: Update beliefs

$$p(x \mid D_t) \text{ where } D_t = \{a_\tau, y_\tau\}_{\tau=1}^{t}$$



# Challenges to BED

- Mutual Information $\mathrm{I}_a(X;Y|D_t)$ is intractable
- Exacerbated by number of designs
- No known unbiased estimator




Either upper bound $\mathbb{E}[\hat{u}] \geq \mathrm{I}$,or lower bound $\mathbb{E}[\hat{l}] \leq \mathrm{I}$

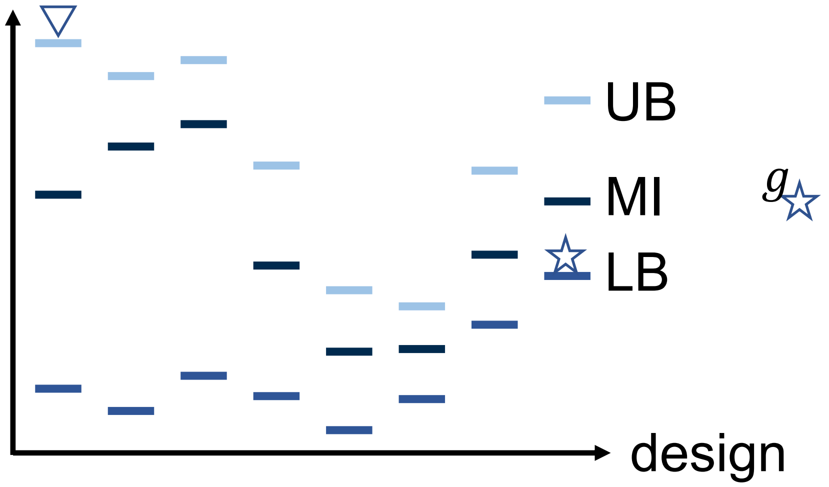
# Adaptive Allocation of Resources

Upper ***and*** lower bounds at little additional cost gives guarantee relative to optimal

$$\mathrm{I}_a - \mathrm{I}^* \geq l_a - \max_{a' \in \mathcal{A} \setminus a} u_{a'} := g_a$$

Allocate computation to promising designs

Trade computation for performance

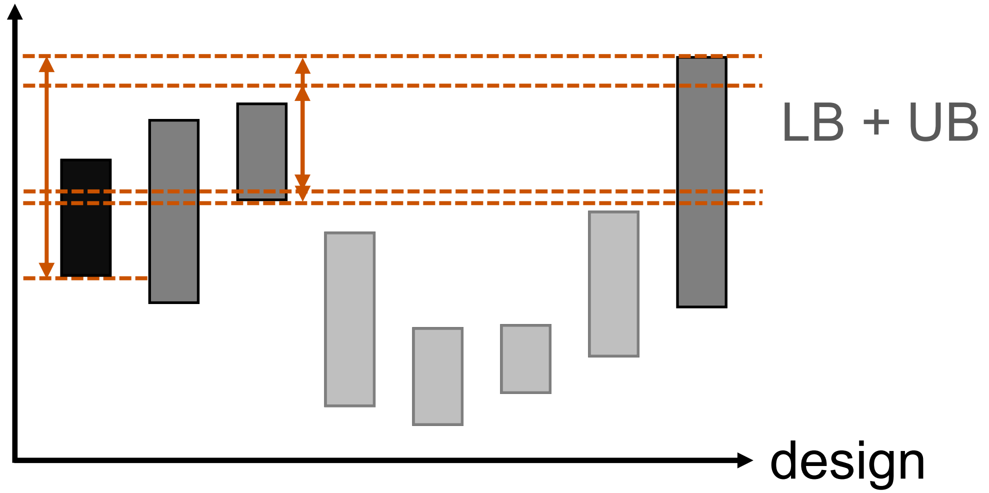



$$\hat{l} = \frac{1}{N} \sum_{n=1}^{N} \log \frac{p(y_n \mid x_n)}{\boxed{\frac{1}{N} \sum_{m=1}^{N} p(y_n \mid x_m)}}$$

$$\hat{u} = \frac{1}{N} \sum_{n=1}^{N} \log \frac{p(y_n \mid x_n)}{\boxed{\frac{1}{N-1} \sum_{m \neq n} p(y_n \mid x_m)}}$$

# Cost-Aware BED with Iterative Refinement

- Two-sided bounds
  - Provide guarantee relative to optimal
  - Focus computation
  - Trade off performance/computation
- Iterative Approach
  - Loosely bound all
  - Select high marginal utility, mu $= \frac{\Delta_a g^*}{c_a}$
    cost-sensitive allocation

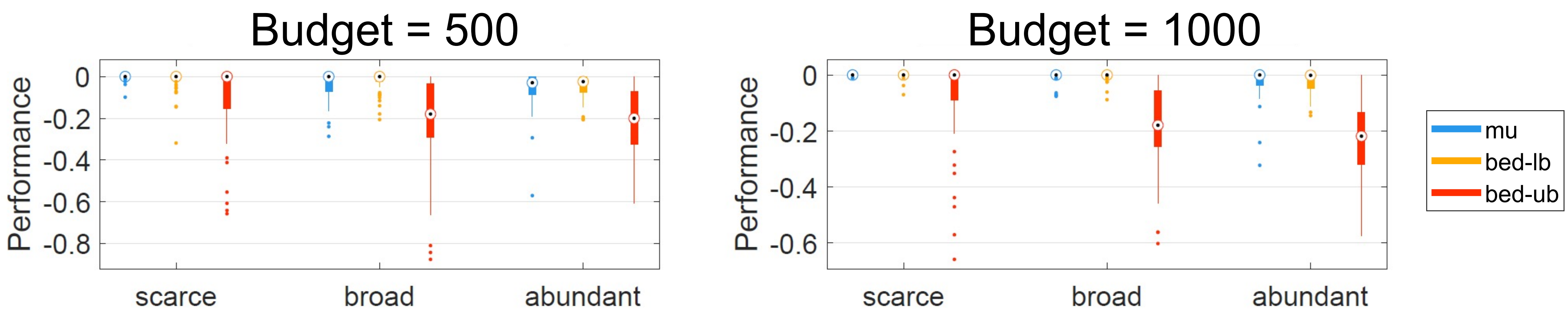


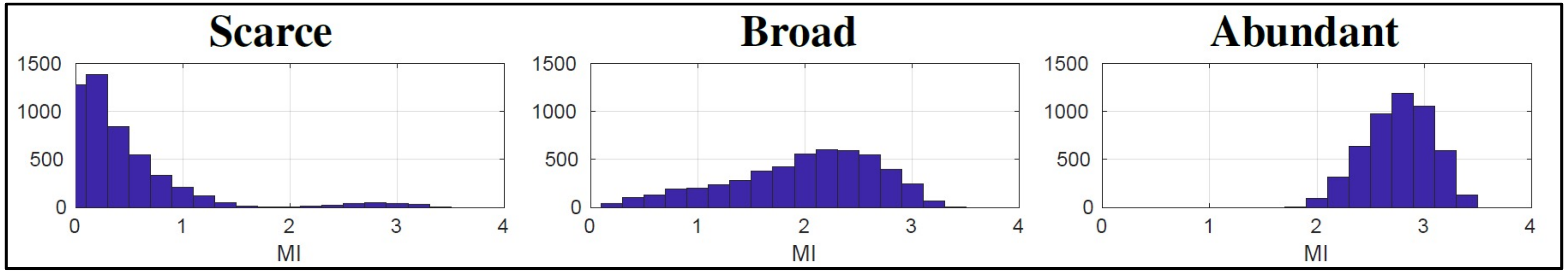

$\Delta_a g^*$ → Update to *highest* guarantee

$c_a$ → Bound evaluation, sampling

# Iterative Refinement of Bounds

$I_{\hat{a}} - I^*$



# Conclusion and Next Steps

BOED and Information Gathering

- Two-sided bounds give performance guarantees
  - Iterative approach minimizes computation
  - Knapsack-based procedure gives additional computational savings
- Trade off computational resources with performance guarantee
- Integration probabilistic machine learning with human domain expertise within BOED

Items I did not discuss

- VoI-based design of *macro-actions* with *performance guarantee*s
- Integration of deep learning with *interpretable* probabilistic transformations for data fusion

# ~~Expected~~ Current Impact

Existing Projects

- BOED for design oof lead-perovskite solar cells
- Well routing and gas-lift optimization for management offshore petroleum infrastructure
- Detection of biota response to radiological exposure (national lab collaboration)
- Sensor array design for detection of SNM *without* reconstruction
- Detection of undersea volcanic plumes
- Your problems?