Texas A&M University
Genomic Signal Processing Laboratory

Intrinsically Bayesian Robust Operators and Optimal Experimental Design Based on Uncertainty Quantification

Edward R. Dougherty

Department of Electrical and Computer Engineering
Texas A&M University

The Core Engineering Problem: Synthesis

- The fundamental problem: given a system, synthesize an operator to operate on the system in a desirable manner.
  - Classification: Make decisions.
  - Filtering (regression): Make optimal estimates; for instance, given a degraded signal, estimate the true signal.
  - Control: Alter the dynamical behavior of the system.
- Synthesis starts with a mathematical theory constituting the relevant scientific knowledge and the theory is used to derive an optimal operator for some objective.
- Science constructs models; engineering operates on them.

Synthesis Protocol

- 1. Construct the mathematical model.
- 2. Define a class of operators.
- 3. Define the optimization problem via a cost function.
- 4. Solve the optimization problem.

Wiener Filtering

- 1. Model: two jointly WSS random processes, true signal *Y* and observation signal *X*.
  - Autocorrelation, $E[X(t)X(t')]$, is a function of $t - t'$
- 2. Operator class consists of linear filters on the observed signal:

  $$\psi(X)(t) = \sum_{a \leq s \leq b} g(t, s)X(s)$$

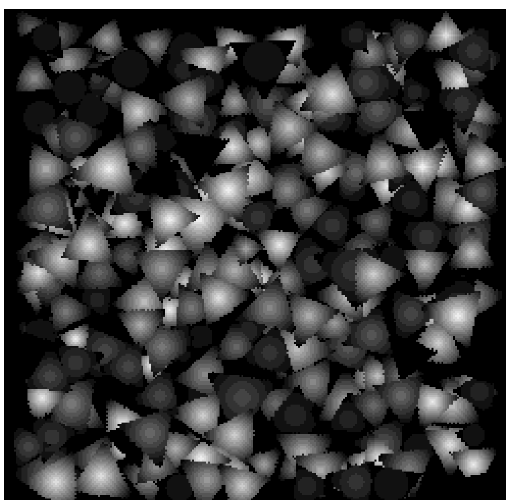
- 3. Cost is the mean-square error between *Y* and $\psi(X)$:

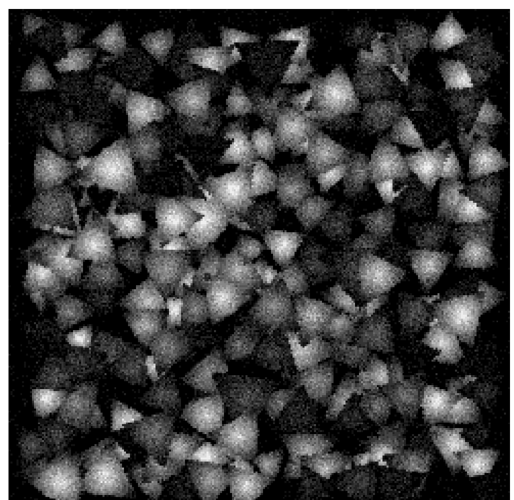
  $$\text{MSE} = \sum_{-\infty \leq t \leq \infty} |Y(t) - \psi(X)(t)|^2$$

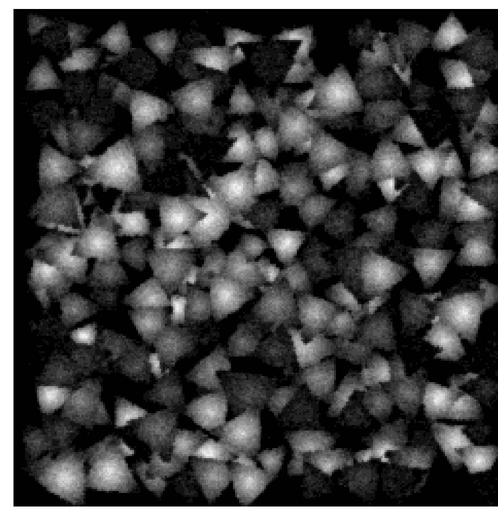
- 4. Optimization: the Fourier transform of the optimal *g* is given in terms of the power spectral density of the observed signal and the cross power spectral density, which are the Fourier transforms of the auto- and cross-correlation functions, respectively:
- $$G(\omega) = S_{YX}(\omega)/S_X(\omega)$$

Filtering Blurred Signal + Noise

- Filter model
  - Original signal: s
  - Blurred signal + white noise: $B(s) + n$
  - Filtered noisy signal: $\psi[B(s) + n]$
- Linear optimization problem
  - Find ψ to minimize distance between $\psi[B(s) + n]$ and s



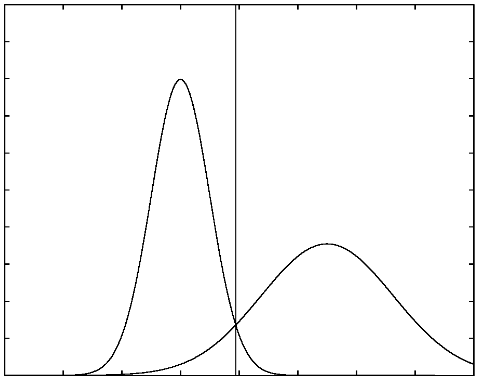




Optimal Classification

- 1. Construct the feature-label distribution.
- 2. The operators consist of classifiers (decisions).
- 3. The cost is classifier error.
- 4. An optimal operator is given by a Bayes classifier.

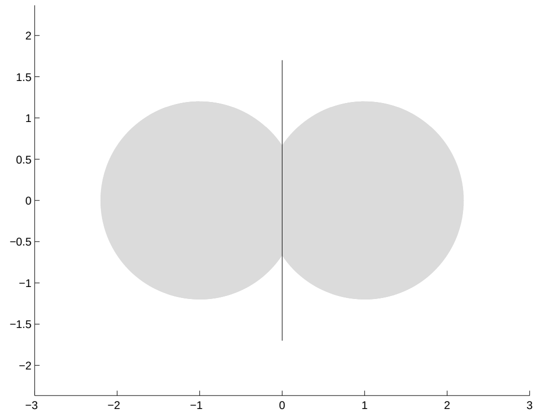
- Class conditional densities:
  $f(\mathbf{x}|0)$ and $f(\mathbf{x}|1)$.



- *Bayes Classifier* $\psi_{\mathrm{Bay}}(\mathbf{x}) = 1$ if $f(\mathbf{x}|1) \geq f(\mathbf{x}|0)$;
  $\psi_{\mathrm{Bay}}(\mathbf{x}) = 0$ if $f(\mathbf{x}|1) < f(\mathbf{x}|0)$.

Gaussian Model

- 1. Feature-label distribution: two Gaussian class-conditional distributions.
- 2. The operators consist of classifiers (decisions).
- 3. The cost is classifier error.
- 4. An optimal operator is given by a Bayes classifier.
  - Linear discriminant for equal covariance matrices.
  - Quadratic discriminant for unequal covariance matrices.



Uncertainty Class of Models

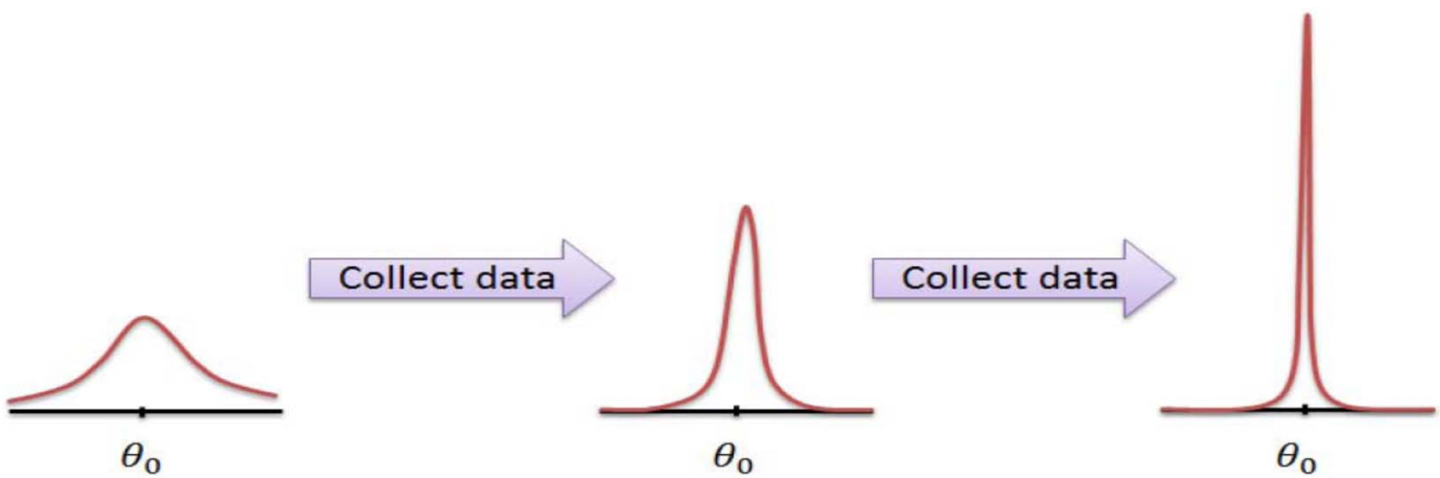
- With complex systems, model parameters cannot be accurately estimated due to lack of data – Small Data.
- Partial knowledge provides *uncertainty class* of models.
  - $\Theta = \{\theta_1, \theta_2, \ldots, \theta_m\}$
  - Model uncertainty is not about randomness within a model, but that we can only identify a class of possible models.
- Example for filtering:
  - Model: $X(t) = \sin(at) + N(t)$; $N(t) \sim N(0, \sigma^2)$.
  - Uncertain model: σ unknown.
- Example for classification:
  - Model: $f(\mathbf{x}|0) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
  - Uncertain model: $\boldsymbol{\mu}$ or $\boldsymbol{\Sigma}$, possibly both, unknown.

Prior and Posterior Distributions

- Assume is a *prior distribution* $\pi(\theta)$ governing the likelihood that some model θ is the full (true) model.
  - Prior distribution encodes our belief concerning the likelihood that any θ corresponds to the full model.
  - $\pi(\theta)$ may be often uniform.
  - Prior construction requires transforming knowledge into a form that is relevant to modeling uncertainty.
- Given a random sample (new data) from the full model, we deduce the posterior distribution $\pi^*(\theta) = \pi(\theta|S)$.
  - The posterior distribution characterizes our understanding of the uncertainty class: prior knowledge + data.

Prior + Data = Posterior

- The prior distribution represents the state of our knowledge prior to the data; the posterior represents the state of our knowledge after joining the prior with the data.
  - Data reduces the uncertainty in the prior – less variance.



Texas A&M University Genomic Signal Processing Laboratory

Intrinsically Bayesian Robust Operator

- An IBR operator minimizes the expected value of the cost among all operators in a class Ψ:

- $E_\Theta[C_\theta(\psi_{\mathrm{IBR}})] = \min\{E_\Theta[C_\theta(\psi)], \psi \in \Psi\}$.

- $E_\Theta[C_\theta(\psi)] = C_{\theta 1}(\psi)\pi(\theta_1) + C_{\theta 2}(\psi)\pi(\theta_2) + \ldots + C_{\theta m}(\psi)\pi(\theta_m)$

- $E_\Theta[C_\theta(\psi)] = \int_\Theta C_\theta(\psi)\pi(\theta)d\theta$

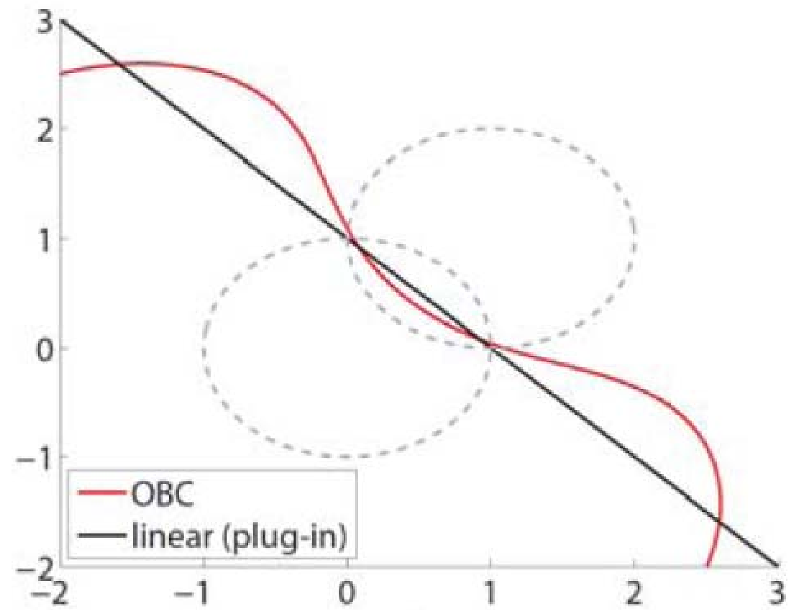
- If the operator class is finite, then the minimum is over a finite set.

Optimal Bayesian Classifier

- *Optimal Bayesian classifier* (*OBC*) uses the posterior.
  - Gaussian model: $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \rightarrow \pi^*(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \pi(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid S)$
- $E_\Theta[\varepsilon_\theta[\psi_{\text{OBC}}]] = \min\{E_\Theta[\varepsilon_\theta[\psi]], \psi \in \Psi\}$
- *Effective* class conditional densities:
  - $f(\mathbf{x}|0; \Theta) = E_\Theta[f(\mathbf{x}|0; \theta)]$; expectation with respect to π^*
  - $f(\mathbf{x}|1; \Theta) = E_\Theta[f(\mathbf{x}|1; \theta)]$; expectation with respect to π^*
- *Bayes Classifier* $\psi_{\text{OBC}}(\mathbf{x}) = 1$ if $f(\mathbf{x}|1;\Theta) \geq f(\mathbf{x}|0;\Theta)$;
  $\psi_{\text{OBC}}(\mathbf{x}) = 0$ if $f(\mathbf{x}|1;\Theta) < f(\mathbf{x}|0;\Theta)$.

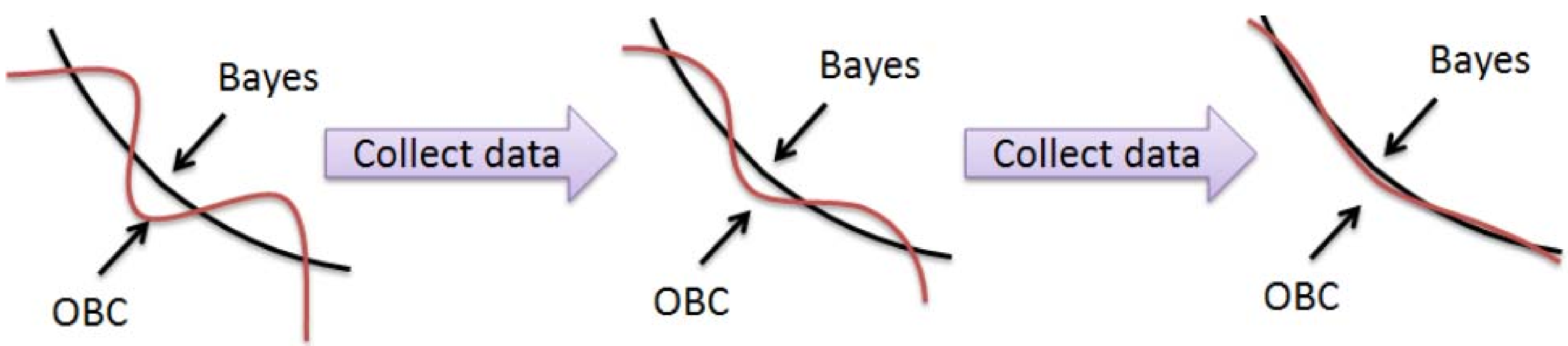
OBC for Gaussian Model

- Polynomial Optimal Bayesian Classifier (red line)
  - Dotted lines are level curves for the Gaussian class-conditional densities corresponding to the expected means and (equal) covariances for a given posterior.
  - Black solid line is linear classifier corresponding to the Bayes classifier for the expected mean and covariance parameters (naive).



Convergence of the OBC

- In the Gaussian and multinomial models, as $n \to \infty$ the OBC converges to the Bayes classifier for the true feature-label distribution.



Synthesis Protocol with Uncertainty

- 1. Construct the mathematical model.
- 2. Define a class of operators.
- 3. Define optimization problem via a cost function.
- 4. Solve optimization via model characteristics.
- 5. Identify the uncertainty class.
- 6. Construct a prior distribution.
- 7. State the IBR optimization problem.
- 8. Construct the effective characteristics.
- 9. Prove that IBR optimization is solved by replacing the model characteristics by effective characteristics.

IBR Wiener Filtering

- Uncertainty class: jointly WSS random processes.
- Optimization: minimize $E_{\Theta}[\text{MSE}_{\theta}]$ between true signal and filtered observation signal.
  - Expected because the signals depend on θ.
- 4. *Effective power spectra* are Fourier transforms of the expected auto- and cross-correlation functions: $S_{\Theta,X}(\omega) = \mathcal{F}[E_{\Theta}[r_{\theta,X}]](\omega)$ and $S_{\Theta,YX}(\omega) = \mathcal{F}[E_{\Theta}[r_{\theta,YX}]](\omega)$. The Fourier transform of the IBR weighting function is
- $$G_{\Theta}(\omega) = \text{S}_{\Theta,YX}(\omega)/S_{\Theta.X}(\omega)$$

IBR Wiener Protocol

- 5. The uncertainty class is defined by the uncertain parameters in the autocorrelation and cross-correlation functions.
- 6. A prior/posterior distribution is constructed for these parameters.
- 7. IBR optimization: minimize the expected MSE.
- 8. Effective characteristics: effective power spectra.
- 9. Prove that the IBR optimization problem is solved by replacing the model characteristics by the effective characteristics.

Measuring Uncertainty

- Uncertainty in parameters that do not affect operator design is no concern.
- Requirements of an uncertainty measure
  - Measure should be based on the objective of the model.
  - Uncertainty should be quantified in terms of the expected cost it induces with respect to the operator class.
- Entropy does not satisfy these requirements.

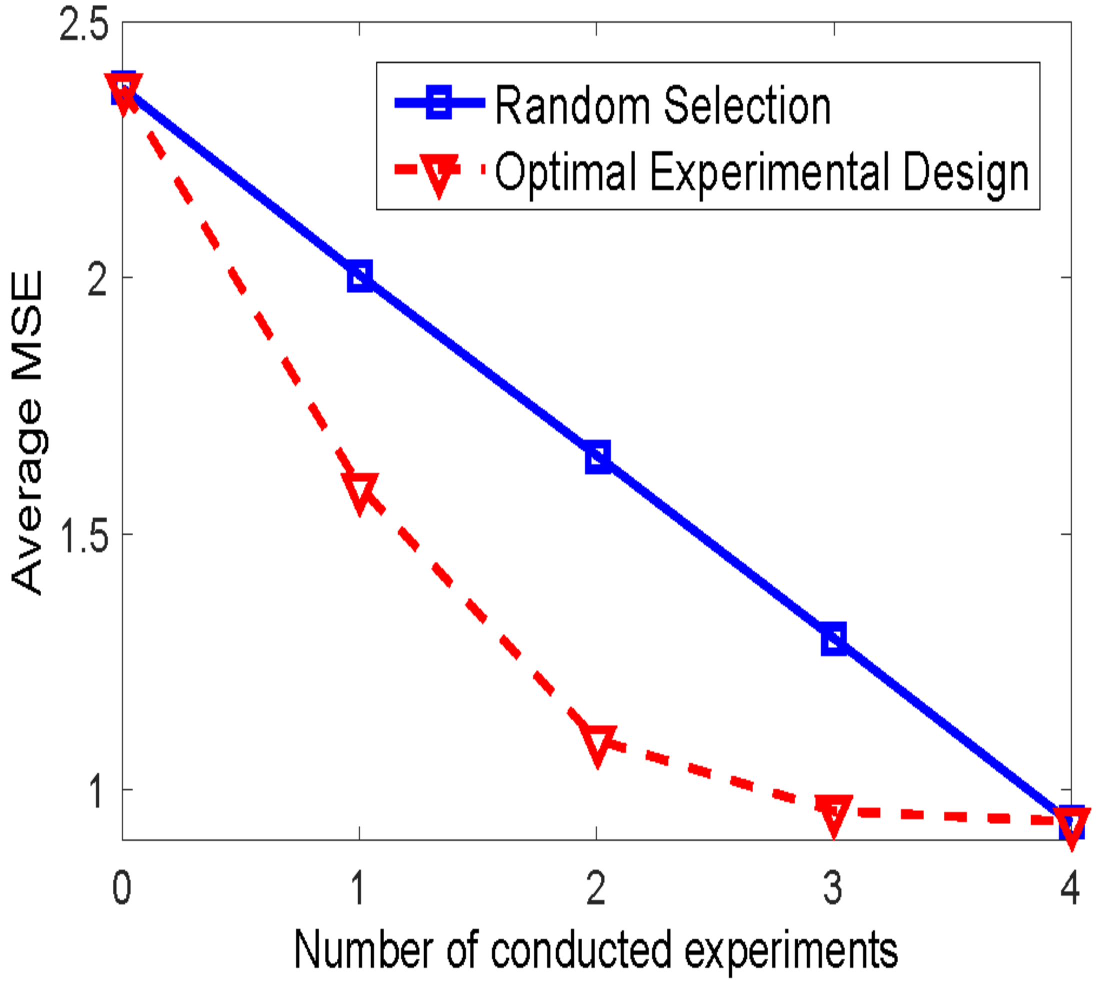
Objective Cost of Uncertainty

- IBR optimality: $E_\Theta[C_\theta(\psi_{IBR})] \le E_\Theta[C_\theta(\psi)]$ for any ψ.
- If ψ_θ is optimal for θ, then $C_\theta(\psi_\theta) \le C_\theta(\psi_{IBR})$.
- For any $\theta \in \Theta$, the *objective cost of uncertainty* (*OCU*) relative to θ is $\mathrm{OCU}(\theta) = C_\theta(\psi_{IBR}) - C_\theta(\psi_\theta)$.
- OCU for the uncertainty class, OCU(Θ), is the OCU relative to the full model, which is not known.
- *Mean objective cost of uncertainty* (MOCU):
- $\mathrm{MOCU}(\Theta) = E_\Theta[\mathrm{OCU}(\theta)] = E_\Theta[C_\theta(\psi_{IBR}) - C_\theta(\psi_\theta)]$

Optimal Experimental Design

- If MOCU ≈ 0, then, on average, $C_\theta(\psi_{IBR}) \approx C_\theta(\psi_\theta)$.
  - If prior is concentrated around full model (plus some regularity conditions), expect IBR to be close to optimal.
- To get a new posterior, among the set of possible experiments, choose the experiment with minimum expected remaining MOCU given the experiment.
  - For each possible experiment, compute remaining MOCU for all possible outcomes, average these MOCUs, take the minimum of these averages, and do experiment. Iterate.
- Result *optimal experimental design* relative to the cost function and the objective uncertainty.

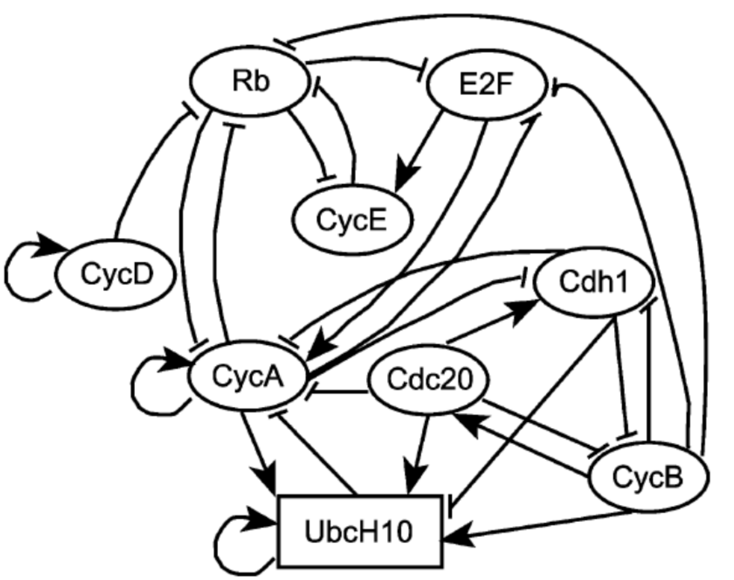
Optimal versus Random – Wiener



Uncertain Mammalian Cell Cycle Network

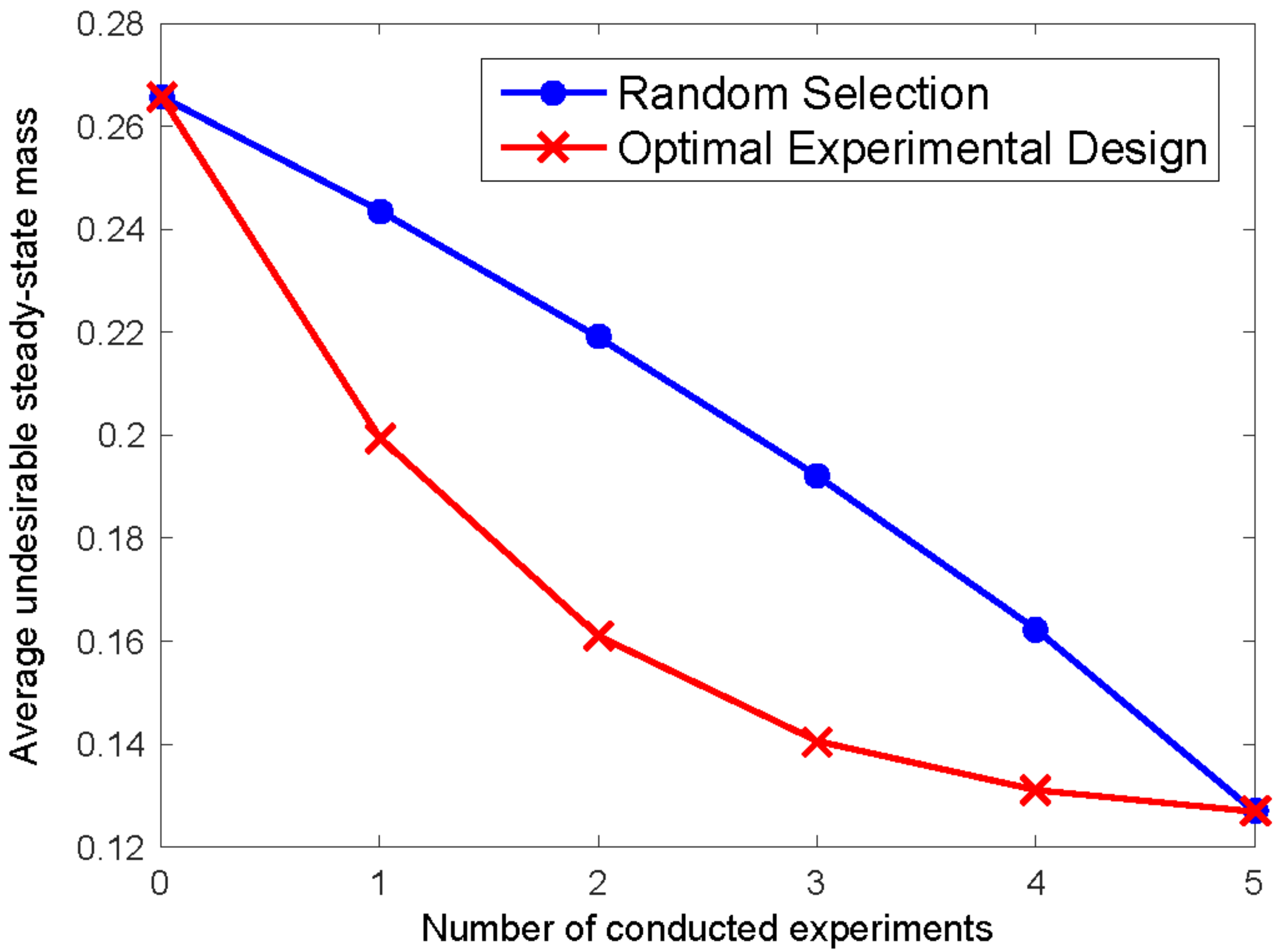
- Binary model of mammalian cell cycle.
  - Construct is network model, plus assumption of binary relations.
  - Data are used to estimate model relations.
  - Uncertainty class from multiple possible constructs or relations.
  - Objective: if Rb = CycD = 0, the cell cycles in the absence of a growth factor, so alter logic.

Product	Predictors
CycD	Input
Rb	$(\overline{CycD} \wedge \overline{CycE} \wedge \overline{CycA} \wedge \overline{CycB})$
E2F	$(\overline{Rb} \wedge \overline{CycA} \wedge \overline{CycB})$
CycE	$(E2F \wedge \overline{Rb})$
CycA	$(E2F \wedge \overline{Rb} \wedge \overline{Cdc20} \wedge \overline{(Cdh1 \wedge Ubc)}) \vee (CycA \wedge \overline{Rb} \wedge \overline{Cdc20} \wedge \overline{(Cdh1 \wedge Ubc)})$
Cdc20	CycB
Cdh1	$(\overline{CycA} \wedge \overline{CycB}) \vee (Cdc20)$
Ubc	$(\overline{Cdh1}) \vee (Cdh1 \wedge Ubc \wedge (Cdc20 \vee CycA \vee CycB))$
CycB	$(\overline{Cdc20} \wedge \overline{Cdh1})$



Texas A&M University
Genomic Signal Processing Laboratory

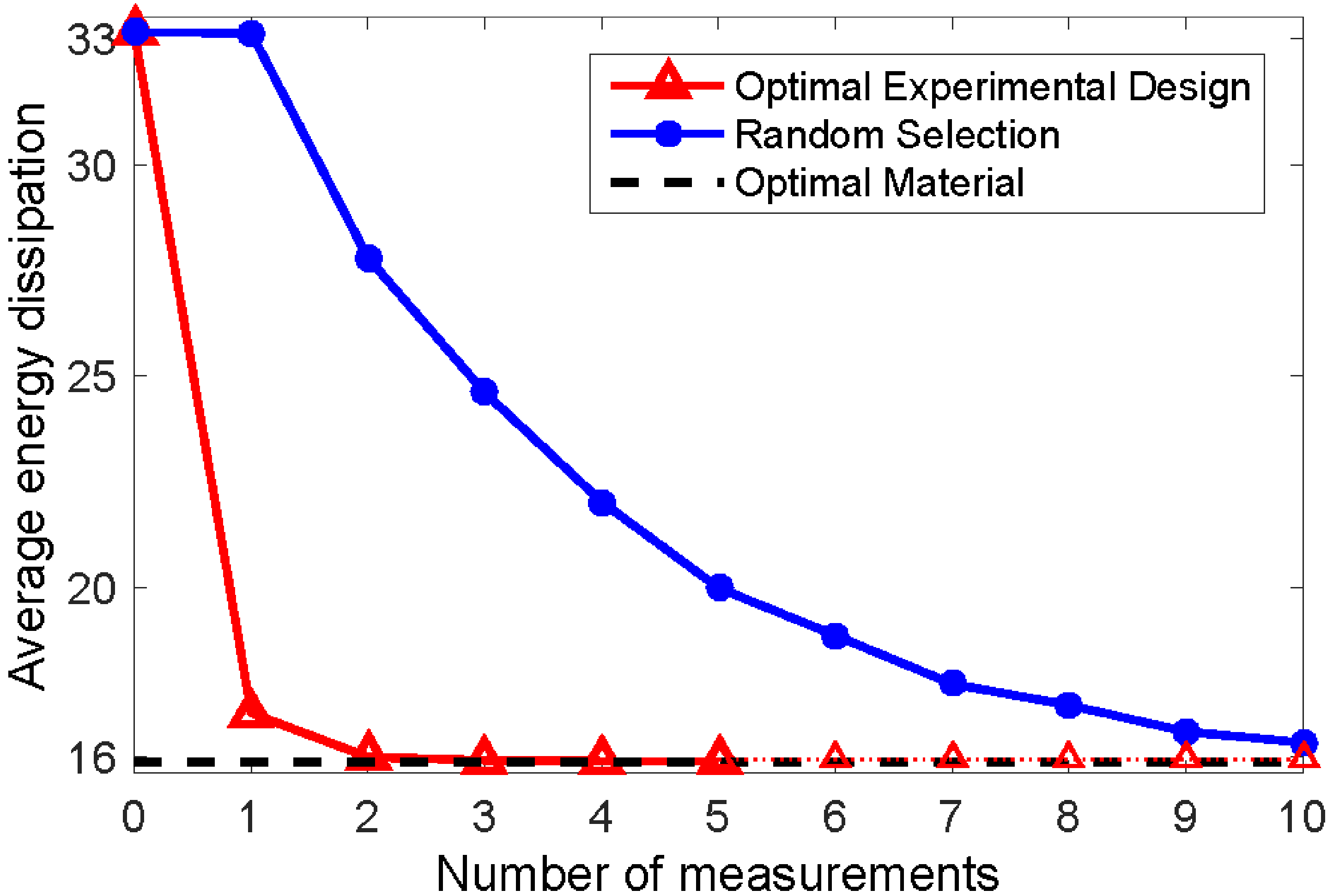
Optimal versus Random – Gene Network



Design Shape Memory Alloy

- Goal: Design a shape memory alloy with low dissipation energy.
  - Dissipate little energy (heat) under load.
- Problem: Design a material with the lowest energy dissipation at a specific temperature where the aim of the experimental design is to suggest the best dopant (added impurity) and concentration for the next measurement.

Optimal versus Random – Materials



Issues for Design under Uncertainty

- Prior Construction
  - Transforming scientific knowledge into constraints and probabilistic knowledge relevant to the design problem.
- Computation – especially for non-Gaussian models
  - Efficient MCMC algorithms.
  - Efficient architectures.
  - Model compression while keeping relevant information.
- Building models compatible with experiments.
- **There is no notion of scientific validity because there is no actual model to experimentally validate.**
  - Modeling our knowledge serves as a practical construct from which to derive an “optimal” action in the world.