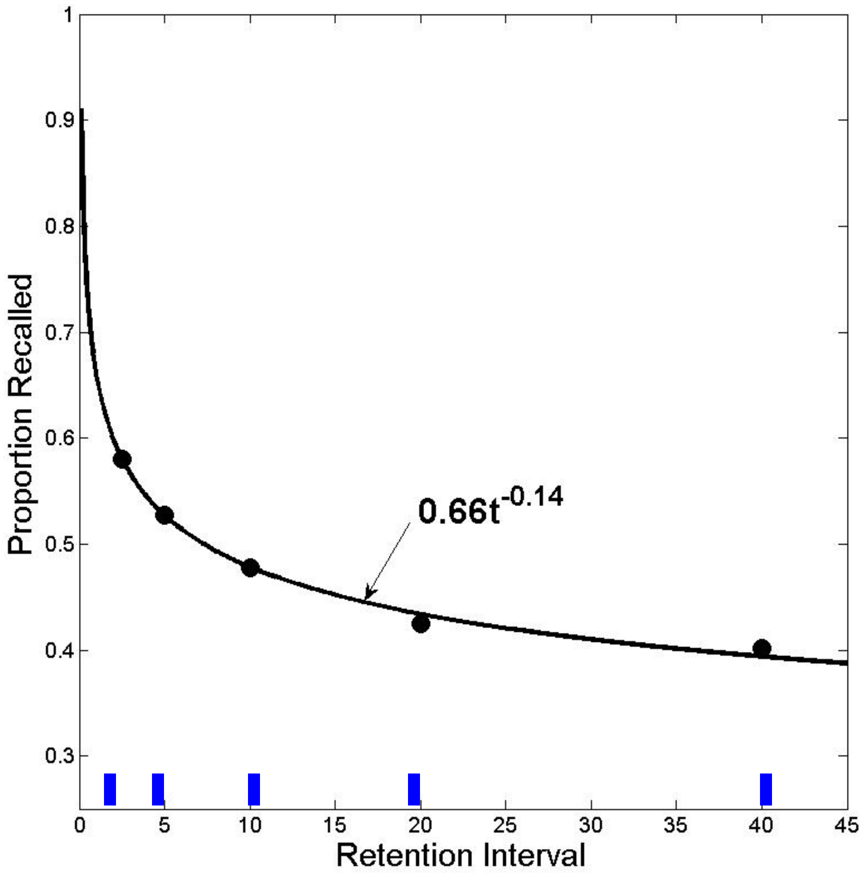
Optimal Experiment Design for Model Discrimination

Jay Myung & Mark Pitt
Ohio State University

Australian Mathematical Psychology Conference (AMPC2009)
University of Newcastle, Newcastle, Australia (Feb 2009)

Wixted and Ebbesen (1991) Experiment



Experimental Design:

$(t_1, t_2, t_3, t_4, t_5) = (2.5, 5, 10, 20, 40)$ (geometric design)

The Issue

- A goal of experimentation is to distinguish between models
- Experimental design relies on experimenter knowledge and intuitions about the models and experimental setup (stimuli, task, presentation schedule)
- With formal models, it should be possible to assist in this enterprise by identifying designs that best discriminate between them
- This is achieved by optimizing the values of a chosen independent variable

Formal Definition of Design Optimization Problem

An optimal design, intuitively, should maximize extent of model dissimilarity between models (Atkinson & Donev, 1992), formally,

- $$d^* = \underset{d \in D}{\operatorname{argmax}}\{U(d)\}$$
- $$U(d) = \sum_{m=1}^{K} p(m) \iint u(d, y_m, \theta_m)\, p(y_m \mid \theta_m, d)\, p(\theta_m \mid d)\, dy_m d\theta_m$$

$$(e.g.,\ u(d, y_m, \theta_m) = \sum_{k \neq m}^{K} SSE(y_m, prd_k))$$

The *U(d)* represents the expected (i.e., average) utility of a design *d*, where the expectation is taken *over all the unknowns*: models under consideration, model parameters, and possible experimental observations.

Finding the optimal design *d**, however, is a nontrivial undertaking due to the requirements of optimization and high dimensional integration.

Design Optimization via Density Simulation

Following Muller (1999) and Muller, Sanso & de Iorio (2004), we adopt a simulation-based Bayesian approach to design optimization.

The ***Bayesian "trick"*** is to treat *d* as a random variable, to view *U(d)* as a density over design space, and to define an artificial distribution h(.) as

$$h(d, y_{m's}, \theta_{m's}) = \alpha \left[\sum_{m=1}^{K} p(m) u(d, y_m, \theta_m) \right] \prod_{m=1}^{K} p(y_m \mid \theta_m, d) p(\theta_m \mid d)$$

The marginal distribution $h(d)$ obtained after integrating out $y_{m's}$ and $\theta_{m's}$ turns out to be equal to $U(d)$, up to a proportionality constant,

$$\begin{aligned} h(d) &= \int \cdots \int h(d, y_{m's}, \theta_{m's}) dy_1 \cdots dy_K d\theta_1 \cdots d\theta_K \\ &= \alpha U(d) \end{aligned}$$

Formal Definition of Design Optimization Problem

An optimal design, intuitively, should maximize extent of model dissimilarity between models (Atkinson & Donev, 1992), formally,

- $d^* = \underset{d \in D}{\text{argmax}}\{U(d)\}$
- $U(d) = \sum_{m=1}^{K} p(m) \iint u(d, y_m, \theta_m) p(y_m \mid \theta_m, d) p(\theta_m \mid d)\, dy_m d\theta_m$

$$(e.g.,\ u(d, y_m, \theta_m) = \sum_{k \neq m}^{K} SSE(y_m, prd_k))$$

The *U(d)* represents the expected (i.e., average) utility of a design *d*, where the expectation is taken *over all the unknowns*: models under consideration, model parameters, and possible experimental observations.

Finding the optimal design *d**, however, is a nontrivial undertaking due to the requirements of optimization and high dimensional **integration**.

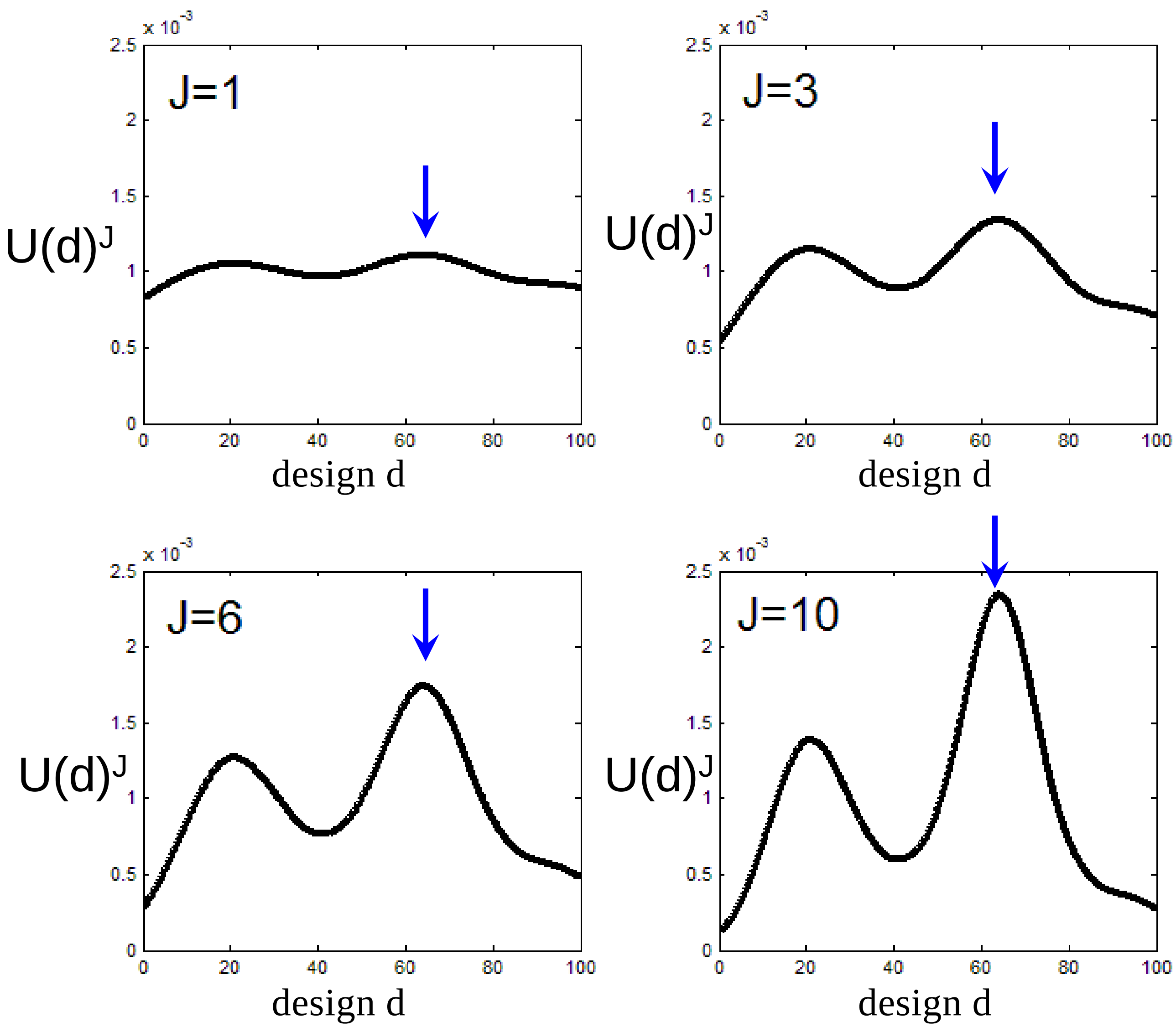
To improve computational efficiency, we augment the artificial distribution with J (e.g., =10) independent samples of $y_{m's}$ and $\theta_{m's}$ given a fixed *d* as follows

$$H_J(d, y_{mj's}, \theta_{mj's}) = \alpha_J \prod_{j=1}^{J} h(d, y_{mj's}, \theta_{mj's})$$

Marginalizing the above density $\mathrm{H}_J(.)$ over all $y_{mj's}$ and $\theta_{mj's}$ then yields

$$H_J(d) = \alpha_J U(d)^J$$

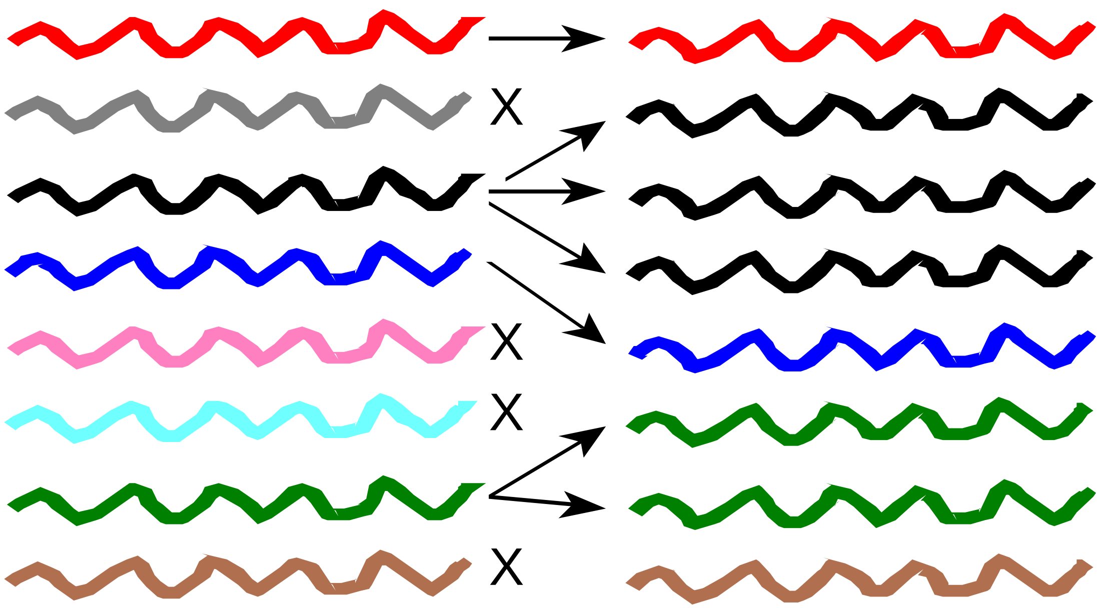
Optimization via “Sharpening-up”



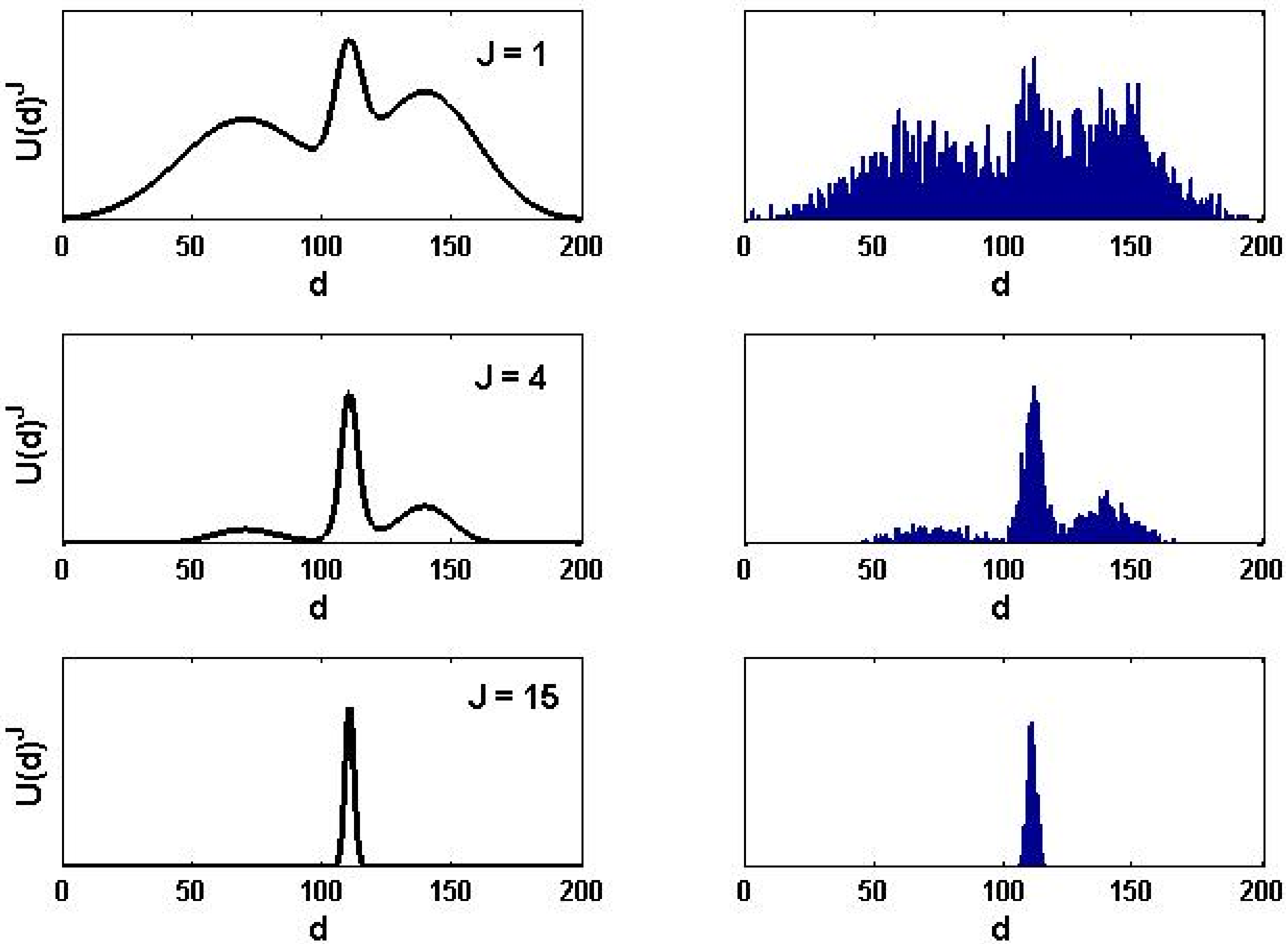
Density Simulation via Sequential Monte Carlo (SMC)

(Amzal, Bois, Parent, Robert, 2006; Doucet, de Freitas & Gordon, 2001)

SMC (or particle filter) is the machinery used in the sampling scheme. It consists of a *population* of "parallel-running & interacting" MCMC chains, called *particles*, that get eliminated or multiplied according to an evolutionary process.



Target density SMC samples



Optimal Designs for Discriminating Retention Models

A central issue in memory research is to determine the exact course of forgetting over time. Of the dozens of retention functions, the following two have received considerable attention:

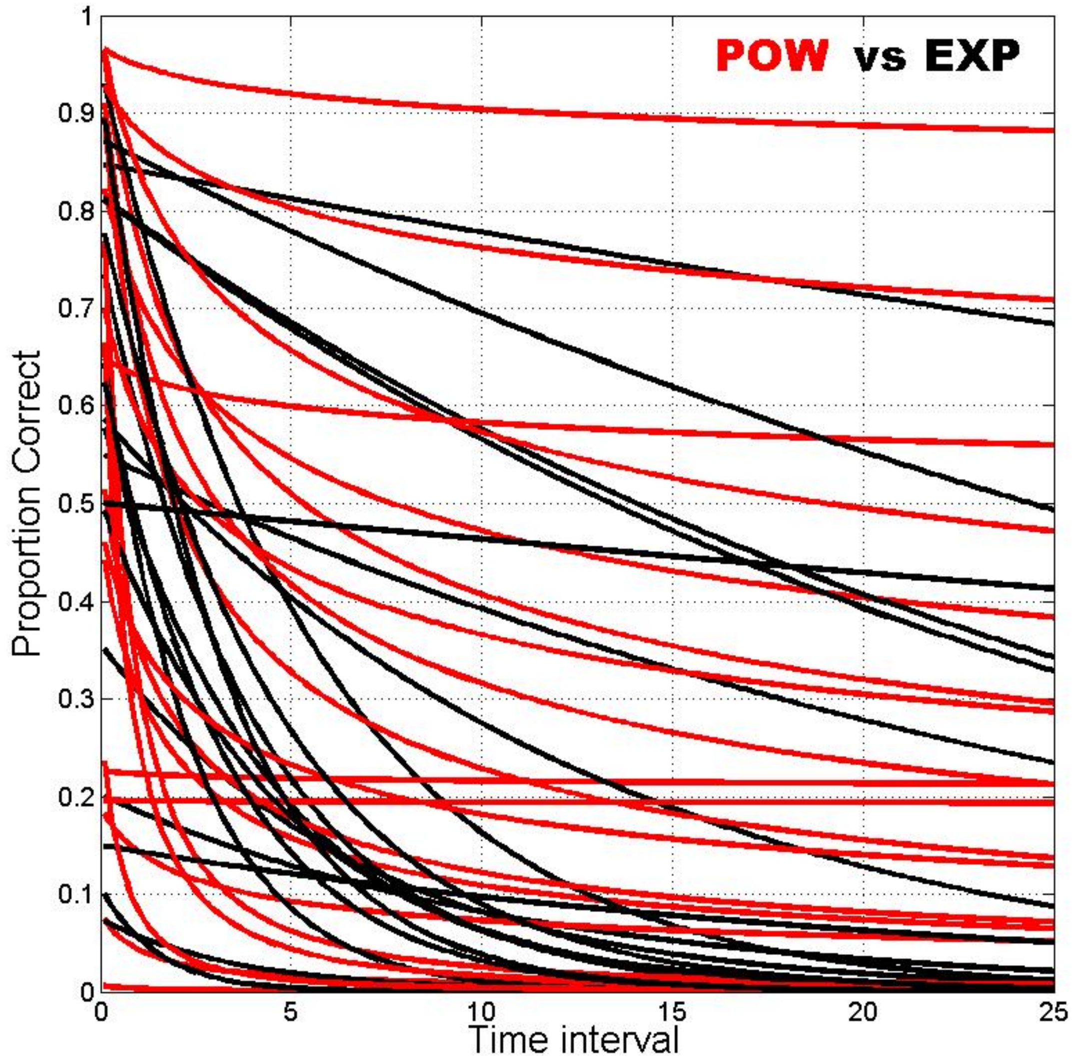
Power model (POW): $p_i = a(t_i + 1)^{-b}$

Exponential model (EXP): $p_i = a\exp(-bt_i)$

where p_i is the proportion of correct recall given retention time t_i and a and b are two parameters.

Design Optimization (DO) Problem:

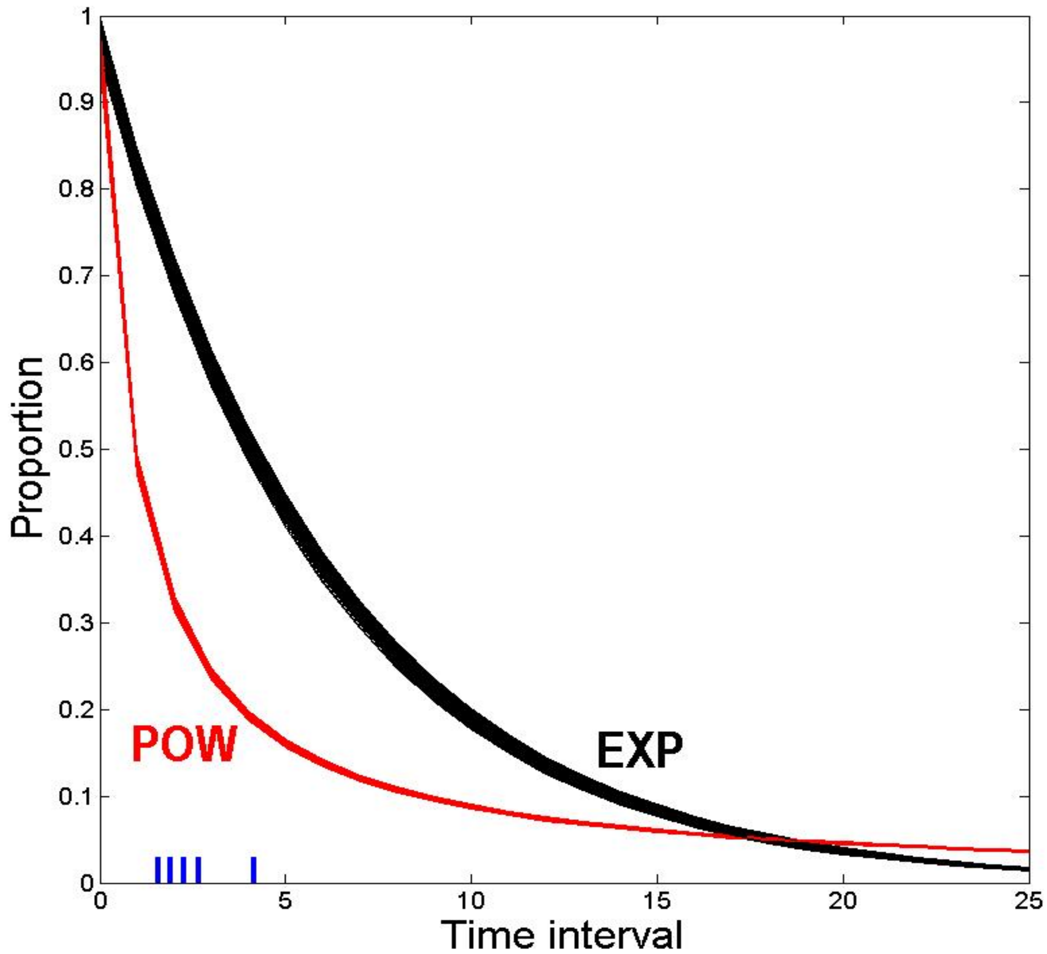
The problem is to find a design $d = \{t_1,\ldots,t_N\}$ consisting of N time intervals that best discriminates the two models.



POW: $p = a(t+1)^{-b} \quad (0 < a < 1; 0 < b < 3)$

EXP: $p = ae^{-bt} \quad (0 < a < 1; 0 < b < 3)$

Intuitive Design Optimization



Restricted parameter ranges

POW: $p = a(t+1)^{-b}$ $(0.95 < a < 1; 1.00 < b < 1.01)$

EXP: $p = ae^{-bt}$ $(0.95 < a < 1; 0.16 < b < 0.17)$

Encorporating Model Selection into Design Optimization

Model Recovery Rate as Utility Function:

Under this utility, the goal is to identify a design that maximizes the expected model recovery rate between two models, A and B, averaged across observations and parameters:

$$U(d) = E[I_A(d) + I_B(d)]$$

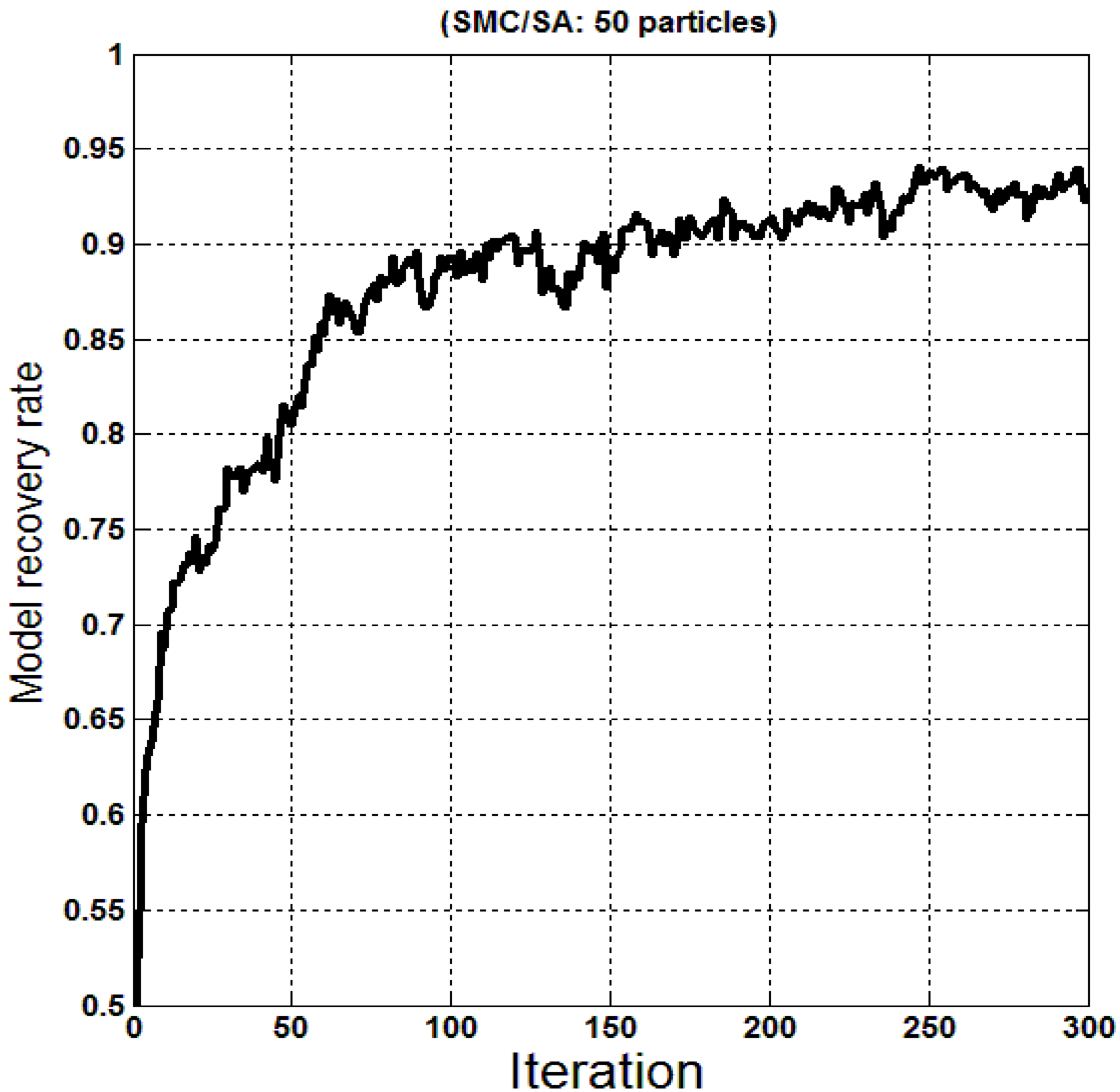
where

$$I_A(d) = 1 \quad \text{if} \quad MDL_A < MDL_B$$
$$= 0 \quad \text{otherwise}$$

with

$$MDL = -\ln f(y \mid \hat{\theta}) + \frac{k}{2}\ln\left(\frac{n}{2\pi}\right) + \ln\int\sqrt{|I(\theta)|}\,d\theta$$

Typical Run of SMC/SA Algorithm



How Good Are the “Optimal” Designs?

N	Type	Design (chosen time intervals)	Model discriminability (% model recovery)
3	**Linear**	**(1, 2, 3)**	**58.3**

How Good Are the "Optimal" Designs?

N	Type	Design (chosen time intervals)	Model discriminability (% model recovery)
3	**Linear**	**(1, 2, 3)**	**58.3**
	Geometric	**(1, 2, 4)**	**61.9**

How Good Are the "Optimal" Designs?

N	Type	Design (chosen time intervals)	Model discriminability (% model recovery)
3	**Linear**	**(1, 2, 3)**	**58.3**
	Geometric	**(1, 2, 4)**	**61.9**
	Optimal	**(9.53, 26.9, 252)**	**86.3**

How Good Are the "Optimal" Designs?

N	Type	Design (chosen time intervals)	Model discriminability (%model recovery)
3	**Linear**	**(1, 2, 3)**	**58.3**
	Geometric	**(1, 2, 4)**	**61.9**
	Optimal	**(9,53, 26.9, 252)**	**86.3**
5	**Linear**	**(1, 2, 3, 4, 5)**	**65.3**
	Geometric	**(1, 2, 4, 8, 16)**	**76.5**

How Good Are the "Optimal" Designs?

N	Type	Design (chosen time intervals)	Model discriminability (%model recovery)
3	**Linear**	**(1, 2, 3)**	**58.3**
	Geometric	**(1, 2, 4)**	**61.9**
	Optimal	**(9,53, 26.9, 252)**	**86.3**
5	**Linear**	**(1, 2, 3, 4, 5)**	**65.3**
	Geometric	**(1, 2, 4, 8, 16)**	**76.5**
	W&E (1991)	**(2.5, 5, 10, 20, 40)**	**79.2**

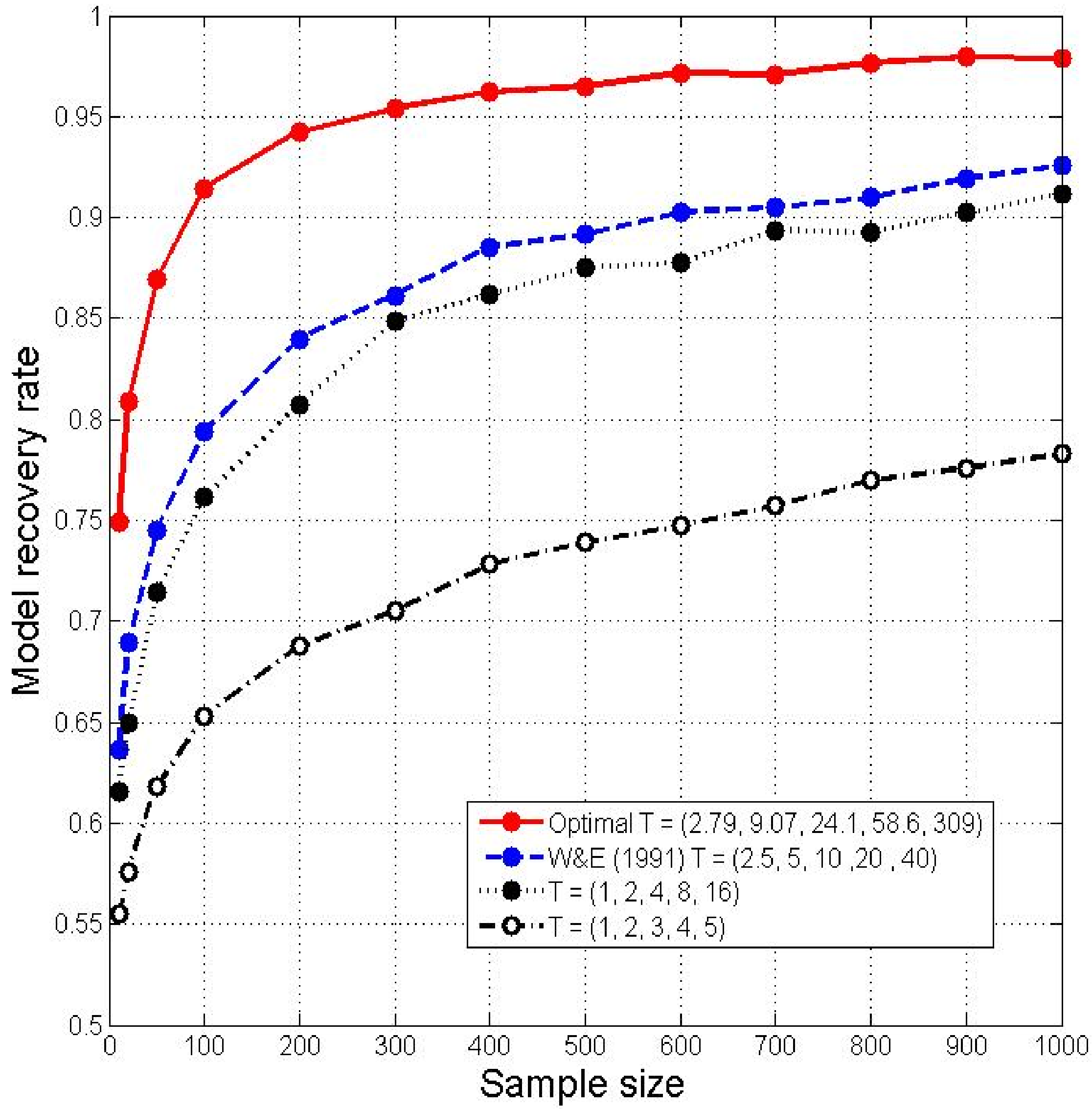
How Good Are the "Optimal" Designs?

N	Type	Design (chosen time intervals)	Model discriminability (%model recovery)
3	**Linear**	**(1, 2, 3)**	**58.3**
	Geometric	**(1, 2, 4)**	**61.9**
	Optimal	**(9,53, 26.9, 252)**	**86.3**
5	**Linear**	**(1, 2, 3, 4, 5)**	**65.3**
	Geometric	**(1, 2, 4, 8, 16)**	**76.5**
	W&E (1991)	**(2.5, 5, 10, 20, 40)**	**79.2**
	Optimal	**(2.79, 9.07, 24.1, 58.6, 309)**	**91.5**

- DO found a design that outperforms the W&E design by more than 10% (91.5% vs 79.2%)

- DO found solutions with three time points that are *even better* than the five time-point design of Wixted & Ebbesen (86.3% vs 79.2%)

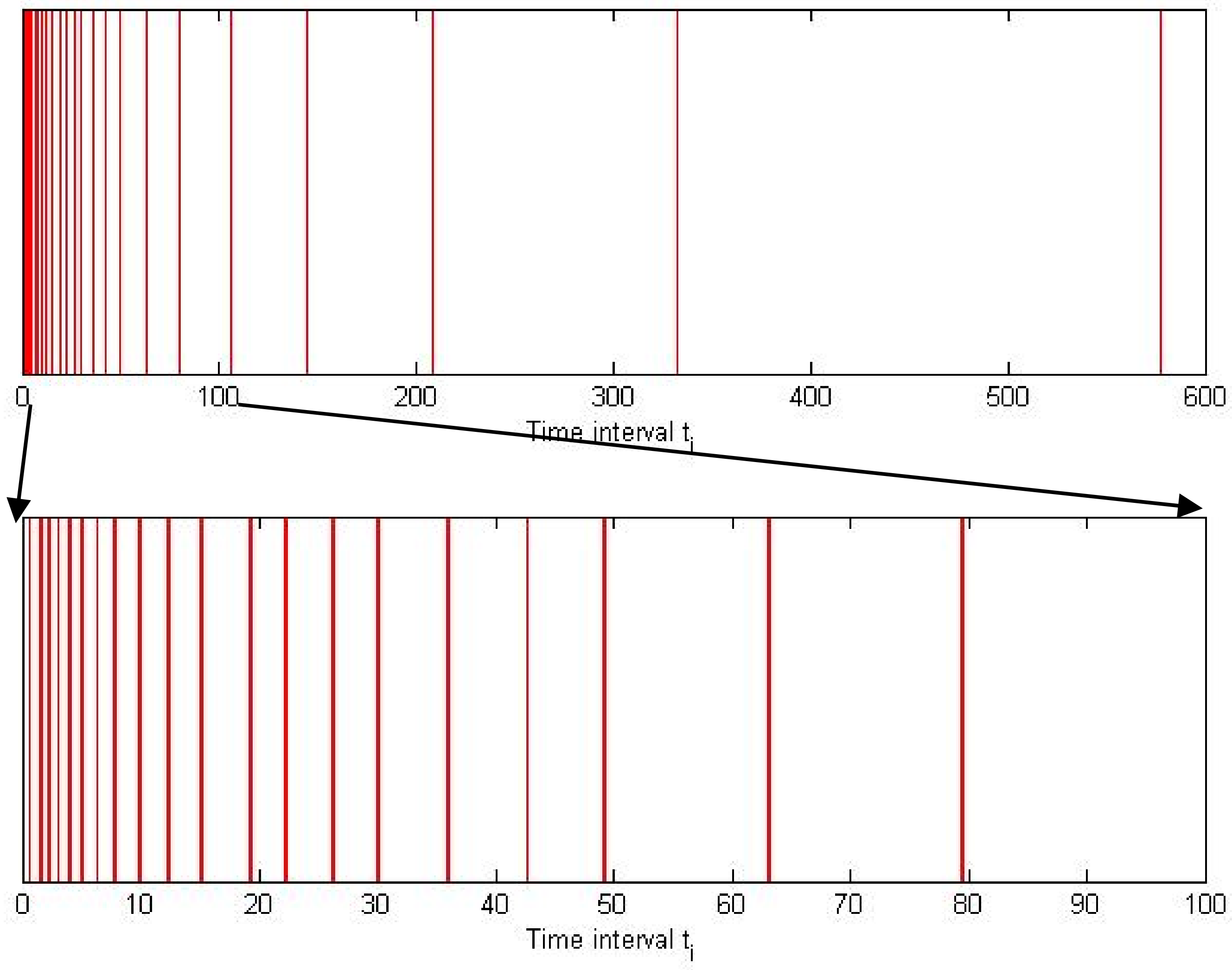
“Power Curves” of Designs



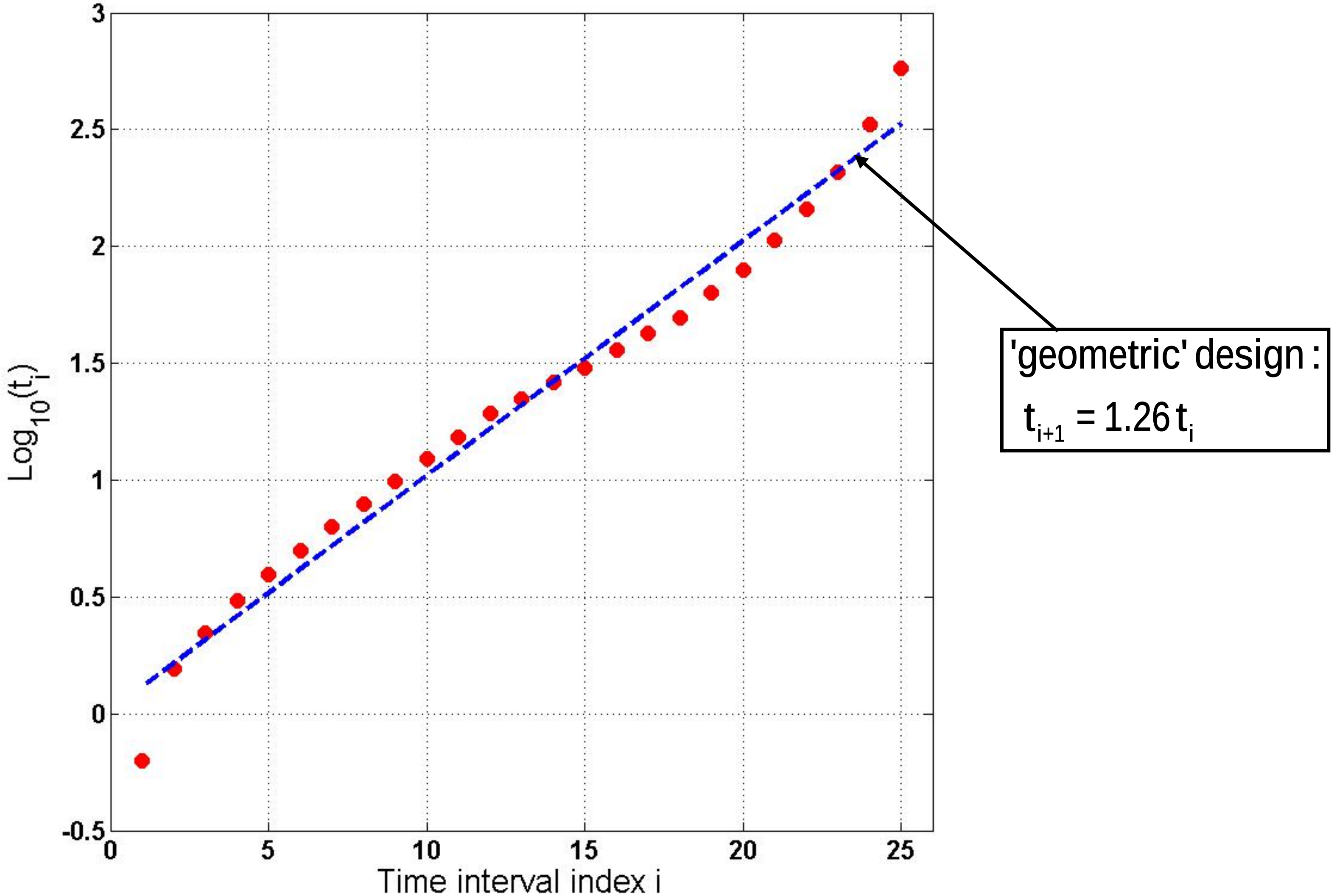
How Good Are the "Optimal" Designs?

N	Type	Design	%Recovery
3	Comparison	(1, 2, 3)	58.3
	Comparison	(1, 2, 4)	61.9
	Optimal	**(9.53, 26.9, 252)**	**86.3**
5	Comparison	(1, 2, 3, 4, 5)	65.3
	Comparison	(1, 2, 4, 8, 16)	76.5
	W&E (1991)	**(2.5, 5, 10, 20, 40)**	**79.2**
	Optimal	**(2.79, 9.07, 24.1, 58.6, 309)**	**91.5**
25	**Comparison**	**(1, 2, 3, 4, …, 24, 25)**	**87.2**
	Optimal	**(0.63, 1.56, 2.22, …, 333, 578)**	**97.3**

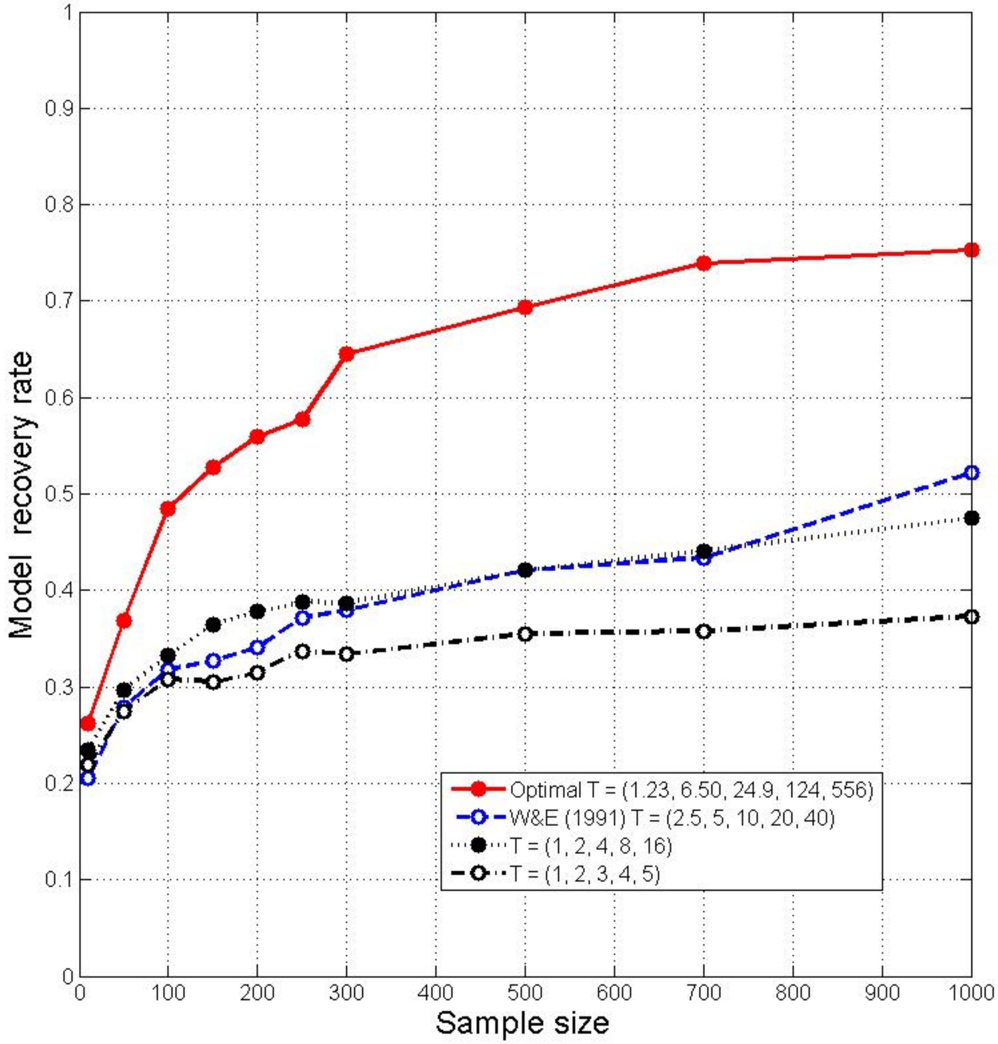
25-Time Point Optimal Design



25-Time Point Optimal Design



Discriminating Six Retention Models



Model	Equation
Power (POW)	$p = a(t+1)^{-b}$
Exponential (EXP)	$p = ae^{-bt}$
Hyperbolic (HYP)	$p = a/(a+t^b)$
Power with asymptote (POWA)	$p = a(t+1)^{-b} + c$
Exponential with asymptote (EXPA)	$p = ae^{-bt} + c$
Exponential with exponent (EXPE)	$p = ae^{-bt^c}$

Conclusions

- Design optimization provides a means of
  - identifying experimental designs that are likely to be most successful in discriminating between formal models

- Current work in adaptive design optimization
  - Mutual information based utility function
  - Trans-dimensional design optimization using Dirichlet Process

Adaptive Design Optimization (ADO)

- Bayesian sequential decision framework
  - Repeat an experiment multiple times
  - Improve the design of the next experiment using knowledge gained from previous experiment