RUTGERS

Design and Analysis for Multifidelity Computer Experiments

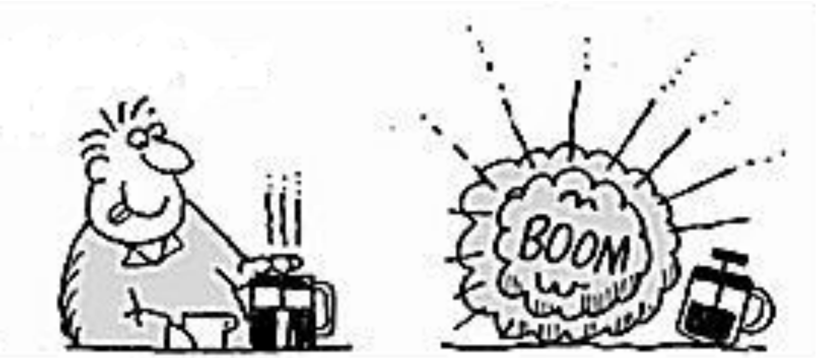
Ying Hung
Department of Statistics and Biostatistics
Rutgers University

Overview

- Introduction to computer experiments
  - Design issues
  - Modeling issues
- Analysis for multifidelity computer experiments
- Improvements based on variable selection

Introduction to computer experiments

- First computer experiments were conducted at Los Alamos National Laboratory to study the behavior of nuclear weapons.
- Computer experiments are becoming popular because many physical experiments are difficult or impossible to perform.



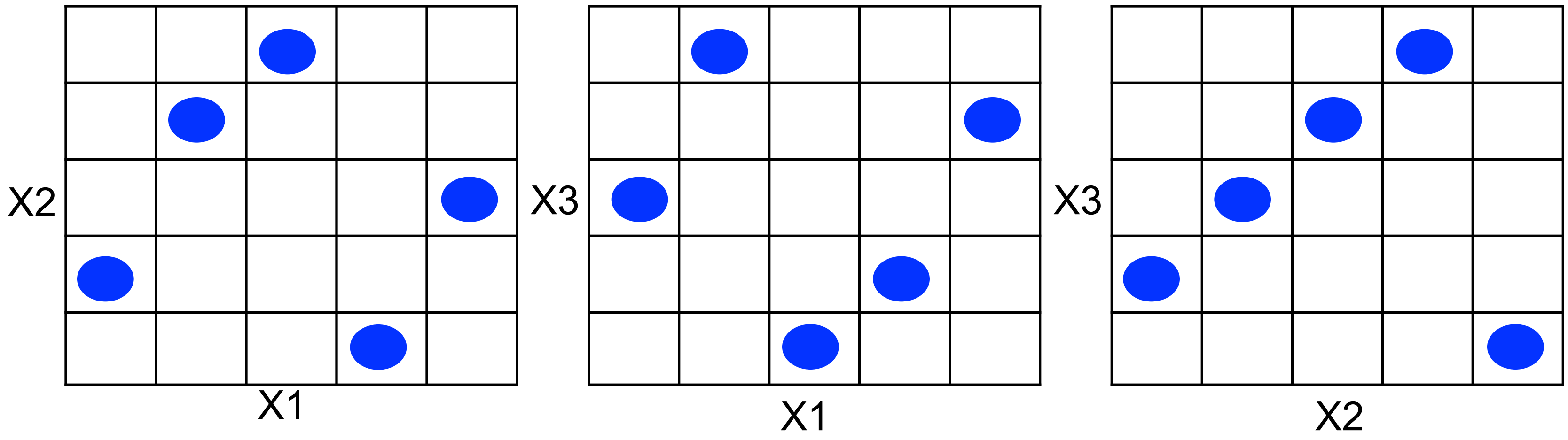
Properties of computer experiments

- Computer experiments refer to those experiments that are performed in computers using physical models and finite element analysis.
- Deterministic outputs (no random error)
  - No replicates required
  - Interpolation
- Large number of variables
- Time-consuming, expensive

Experimental design for computer experiments

- Latin hypercube design (LHD).
  - McKay, Beckman, Conover (1979).
  - Easy to construct.
  - One-dimensional balance.

run	X1	X2	X3
1	1	2	3
2	2	4	5
3	3	5	1
4	4	1	2
5	5	3	4

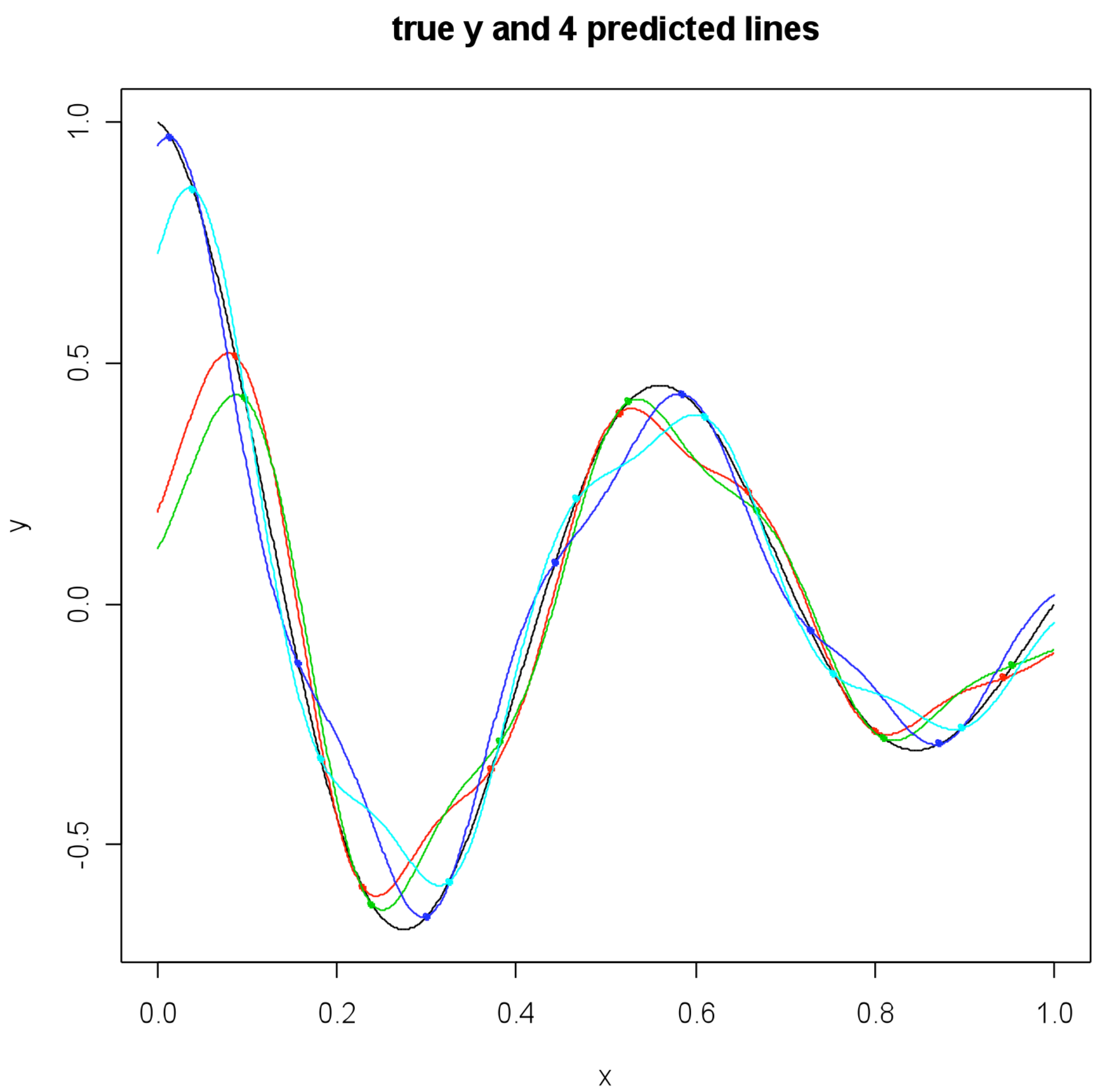


Computer experiments modeling

- Universal kriging $Y(\mathbf{x}) = \mu(\mathbf{x}) + Z(\mathbf{x}),$
  - $\mu(\mathbf{x}) = \sum_{i=0}^{m} \mu_i v_i(\mathbf{x}),$
  - $Z(\mathbf{x})$: a weak stationary stochastic process with mean 0 and covariance function $\sigma^2\psi$.
  - v_i's: known functions, μ_i's: unknown parameters.
  - $cov\{Y(\mathbf{x}+\mathbf{h}), \mathbf{Y}(\mathbf{x})\} = \sigma^{\mathbf{2}}\psi(\mathbf{h})$, where the correlation function $\psi(\mathbf{h})$ is a positive semidefinite function with $\psi(\mathbf{0}) = \mathbf{1}$ and $\psi(-\mathbf{h}) = \psi(\mathbf{h})$.
- Ordinary kriging

$$Y(\mathbf{x}) = \mu_0 + \mathbf{Z}(\mathbf{x})$$

GP example



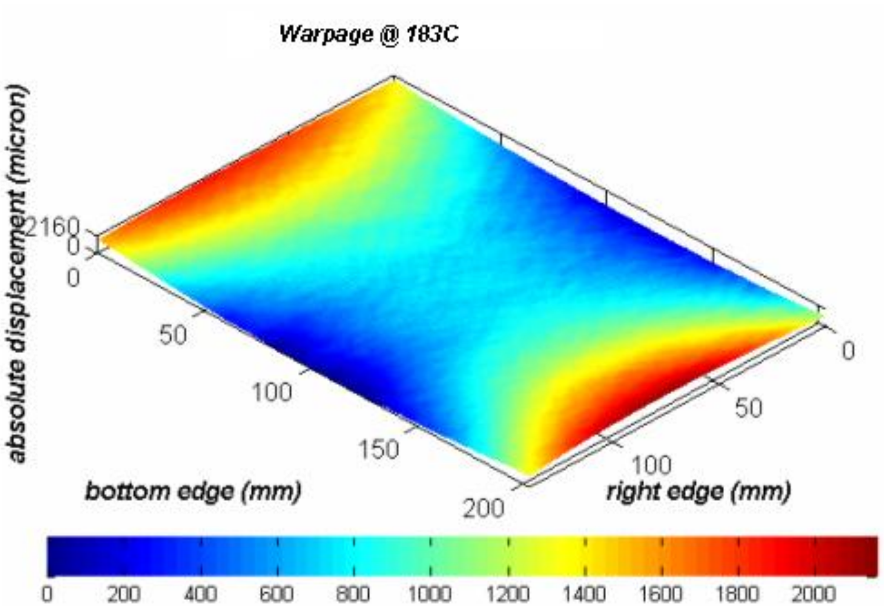
Analysis for Multifidelity Computer Experiments

- Multifidelity computer experiments
  - Physical experiments and computer experiments
  - Computer experiments with different accuracy
- An example in electronic packaging
- Objective:
  - Study effect of initial PWB warpage on low cycle fatigue reliability of solder bumps based on:
    - computer experiments: Finite Element Modeling (FEM)
    - physical experiments: accelerated thermal cycling test
  - How to calibrate?

Finite Element Modeling

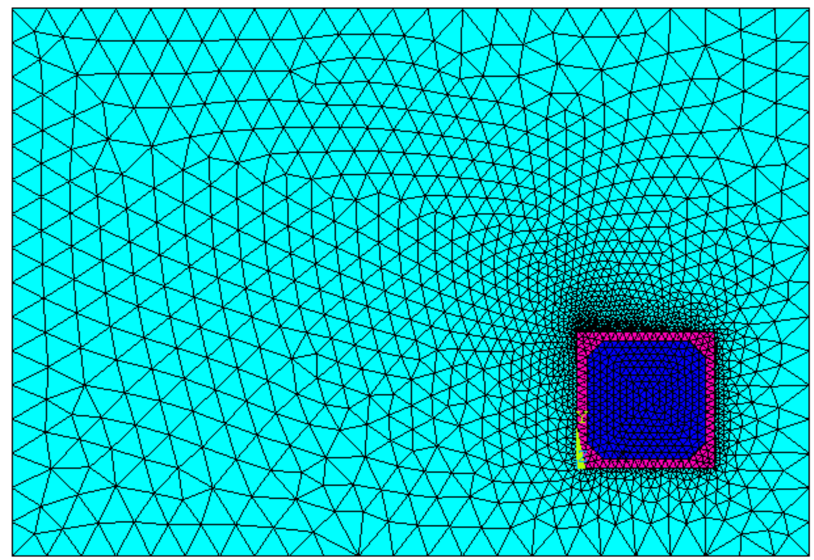
Purpose: To Study How Initial PWB (Printed wiring board) Warpage Affects Solder Bump Fatigue Reliability

- PWB warpage was measured at eutectic temp. and used as initial geometric input to FEM



Max. warpage across PWB = 2158.9 micron

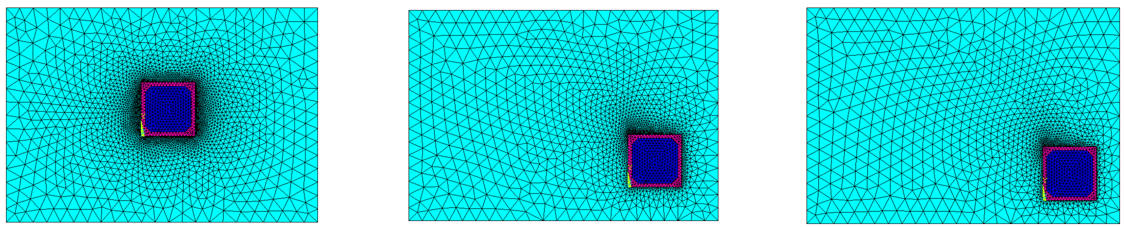
Warpage Measurement of Sample 2 at 183°C



Meshed PWB with 35×35mm PBGA

Study of How PWB Warpage Affects Solder Bump Fatigue

- Case Studies for:
  - Sn-Pb (Tin-Lead) and Sn-Ag-Cu (Lead-Free) Solder Bumps on
    - Two Packages (256-bump 27×27-mm PBGA and 352-bump 35×35-mm PBGA)
      - Each package placed at three different locations:



Location 1 **Location 2** **Location 3**

  - PWB samples can have different initial warpage or can be flat
    - PWBA warpage can be either convex or concave as shown below:

Convex Up (+) **Concave Up (-)**

  - Total 42 cases for each package [including 2 types of solder, 3 chip locations, 3 PWB samples, 2 warpage shapes, and ideal PWB (2 solder types plus 3 package locations) w/o warpage]

Factors studied in FEM

- Factors:

$w_{\max}$	maximum initial PWB warpage at 25°C (mm)	2105.3, 3076.6, 3824.0
w_{shape}	warpage shape	+1: Convex up; -1 Concave up
d_p	package dimension (mm)	27 by 27, 35 by 35
l_p	location of package (mm)	Center, 60-30, Outmost
m_s	solder bump material	Sn-Pb, Lead-free

N_f = fatigue life estimation of solder bumps (cycles)

Data from computer experiments

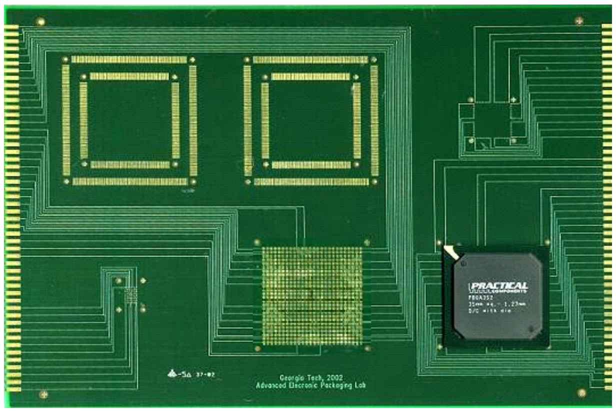
- FEM data (84 runs)

27by27mm	Initial Warp at 25C	-3824	-3076.6	-2105.3	0	2105.3	3076.6	3824
Sn-Pb	Center	1356	1523	1737	1981	1755	1556	1393
	60-30	1438	1618	1823	2005	1846	1644	1481
	Outmost	1477	1652	1861	2034	1884	1691	1518
		-3824	-3076.6	-2105.3	0	2105.3	3076.6	3824
Lead-Free	Center	1618	1807	2042	2215	2061	1845	1660
	60-30	1709	1905	2144	2248	2167	1946	1751
	Outmost	1756	1950	2189	2280	2214	1990	1804

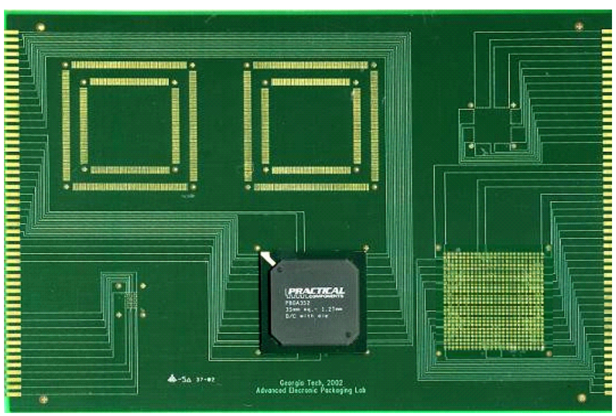
35by35mm		-3824	-3076.6	-2105.3	0	2105.3	3076.6	3824
Sn-Pb	Center	1429	1620	1838	2098	1865	1656	1482
	60-30	1512	1706	1933	2125	1969	1739	1558
	Outmost	1549	1751	1975	2161	1998	1786	1599
		-3824	-3076.6	-2105.3	0	2105.3	3076.6	3824
Lead-Free	Center	1715	1905	2158	2343	2183	1947	1759
	60-30	1841	2006	2251	2380	2278	2054	1885
	Outmost	1882	2063	2304	2415	2341	2097	1929

Experimental Study of Solder Bump Fatigue Reliability Affected by Initial PWB Warpage

Objective: To verify and correlate 3-D finite element simulation results.

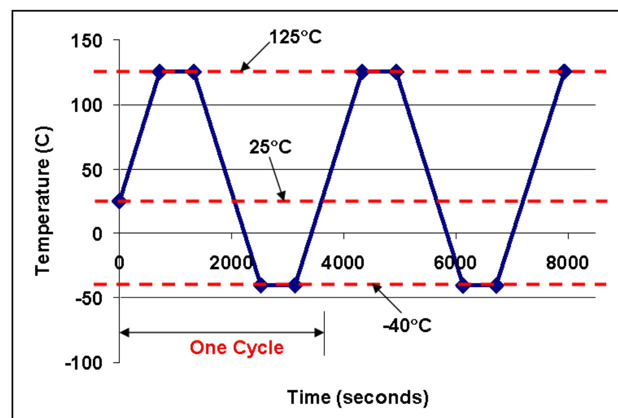


PWB with 35×35 mm PBGA at Location 2



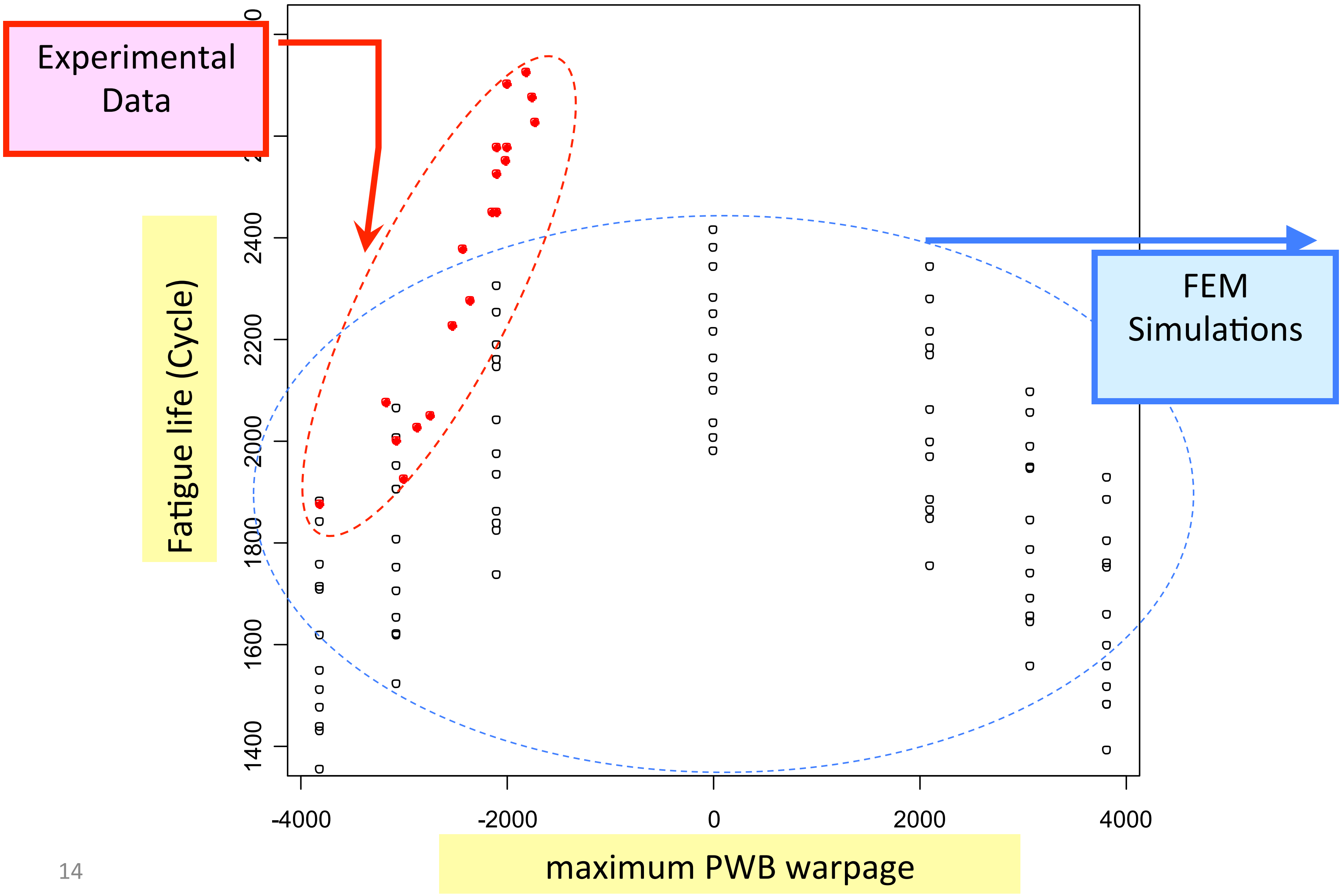
PWB with 35×35 mm PBGA at Location 4

Accelerated Thermal Cycling Test



Standard Thermal Cycling Profile

FEM Simulation vs Experimental Study



Analysis for the two types of data

- Model fitting base on FEM and experimental data:
  - Step 1: Fit kriging model using only simulation data

$$\hat{N}_k(\vec{x}) = 1101.6 + \varphi(\vec{x})^T \Psi^{-1}(N_{out} - 1101.6I)$$

where N_{out} = FEM output data.

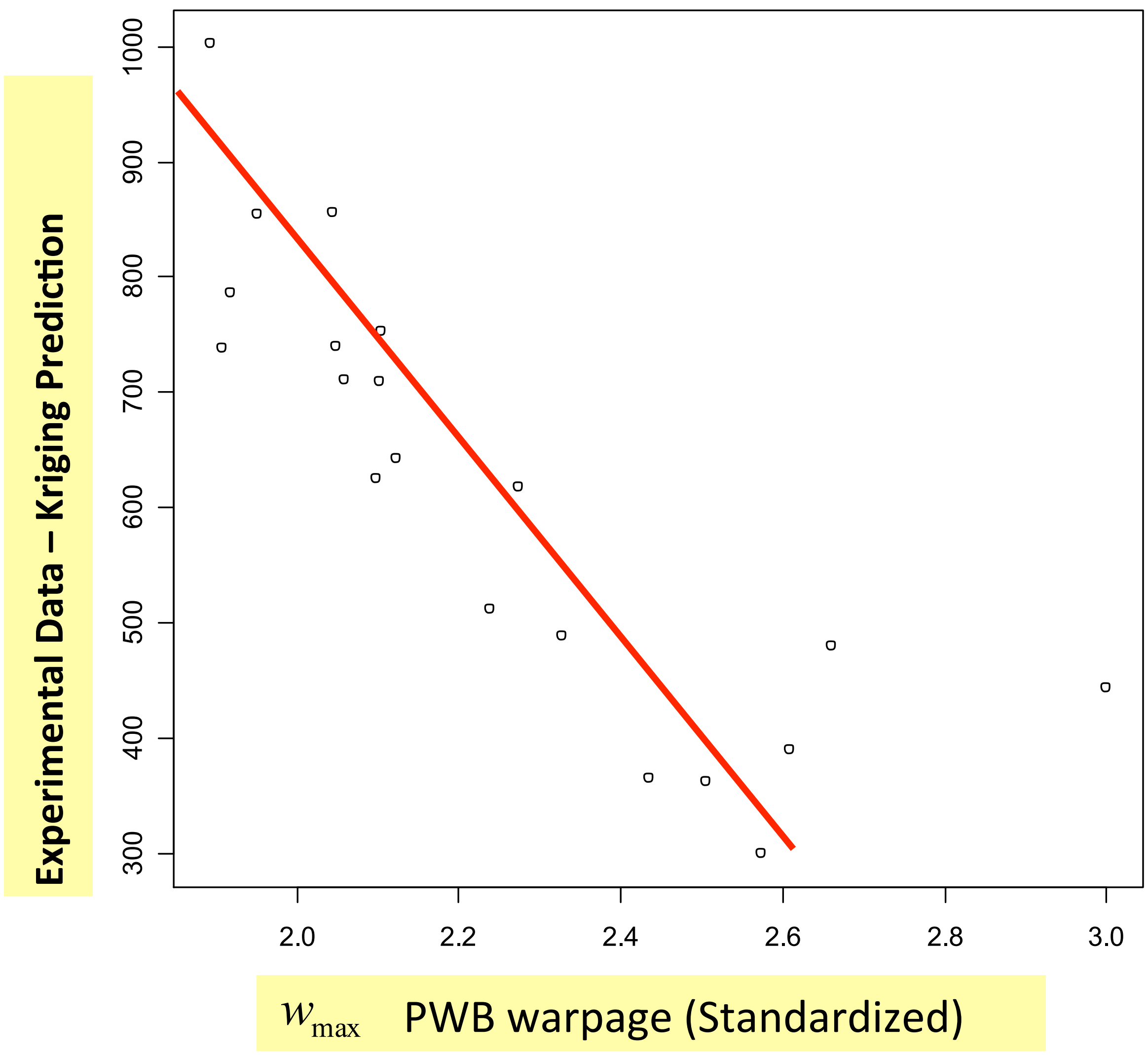
  - Step 2: Calibrate fitted model in Step 1 with experimental data

$$\hat{N}_f(\vec{x}) = 1830.3 - 540w_{\max} + \hat{N}_k(\vec{x})$$

$$= 2931.9 - 540w_{\max} + \phi(\vec{x})^T \Psi^{-1}(N_{out} - 1101.6I)$$

where $\hat{N}_f(\vec{x})$ = fatigue life prediction
$w_{\max}$ = maximum initial PWB warpage at 25°C

Calibration Base on Experimental Data



RUTGERS

Improvement based on variable selection

- Universal kriging $Y(\mathbf{x}) = \mu(\mathbf{x}) + Z(\mathbf{x}),$

  - $\mu(\mathbf{x}) = \sum_{i=0}^{m} \mu_i v_i(\mathbf{x}),$
  - $Z(\mathbf{x})$: a weak stationary stochastic process with mean 0 and covariance function $\sigma^2\psi$.
  - v_i's: known functions, μ_i's: unknown parameters.
  - $cov\{Y(\mathbf{x}+\mathbf{h}), \mathbf{Y}(\mathbf{x})\} = \sigma^{\mathbf{2}}\psi(\mathbf{h})$, where the correlation function $\psi(\mathbf{h})$ is a positive semidefinite function with $\psi(\mathbf{0}) = \mathbf{1}$ and $\psi(-\mathbf{h}) = \psi(\mathbf{h})$.

- Ordinary kriging

$$Y(\mathbf{x}) = \mu_0 + \mathbf{Z}(\mathbf{x})$$

Problems with GP model

- Problems with the ordinary kriging model
  - The prediction can be poor if there are some strong trends.
  - It is not easy to understand the effects of the factors by just looking at the predictor
  - Predictor not robust to the misspecification in the correlation parameters.
- It has been noted that the prediction accuracy and model efficiency of a GP model can be improved by identifying important variables (Welch et al. 1992, Cressie 1993, Martin and Simpson 2005, Gramacy and Lee 2008, Joseph et al. 2008, Stein 2008, Kaufman et al. 2013).

Drawbacks with existing approaches

- Selections are perform based on specific types of a model with convenient by questionable assumptions.
  - A GP mode $y(\boldsymbol{x}) = \mu(\boldsymbol{x}) + Z(\boldsymbol{x}), \quad \mu(\boldsymbol{x}) = \sum_{k=1}^{p} \beta_k x_k = \boldsymbol{f}(\boldsymbol{x})'\boldsymbol{\beta},$
  - Blind kriging selects important variable based only on the mean function of GP models.
  - Linkletter et al. (2006) introduced a variable selection procedure only for the correlation function.
- Computationally intensive

Bayesian variable selection for kriging

- A unified approach that can perform variable selection in a general GP model is attractive but nontrivial. Because the mean function and the correlation structure are not independent. The same variable can appear in either one part or both parts of the model to contribute the effect(s).
- Idea: Using a hierarchical Bayes formulation to connect different effects of the same variables in kriging models.
- Introduce a latent variable into kriging model to indicate if a particular variable is active or not. For those active variables, they can have effect in the mean function and/or in the correlation function.

Bayesian variable selection for kriging

- kriging model:

$$y(\boldsymbol{x}) = \mu(\boldsymbol{x}) + Z(\boldsymbol{x}), \quad \mu(\boldsymbol{x}) = \sum_{k=1}^{p} \beta_k x_k = \boldsymbol{f}(\boldsymbol{x})'\boldsymbol{\beta},$$

- Define a binary vector $\gamma = (\gamma_{1,\ldots,}\gamma_p)'$. Such a vector is used to indicate if a particular variable is active or not.

- Priors:

$$\pi(\beta_k|\gamma_k) = (1-\gamma_k)\delta(0) + \gamma_k DE(0, \tau_k),$$

$$\pi(\theta_k|\gamma_k) = (1-\gamma_k)\delta(0) + \gamma_k Exp(\lambda_k),$$

$$P(\boldsymbol{\gamma}) \propto q^{|\boldsymbol{\gamma}|}(1-q)^{p-|\boldsymbol{\gamma}|},$$

$$\sigma^2 \propto (\sigma^2)^{-\nu_0/2-1} \exp(-1/(2\sigma^2)).$$

Bayesian variable selection for kriging

- This approach is flexible but obtaining the posterior can be computationally difficult because it involves high-dimensional integration.
- With some mild assumptions, we can approximate the posterior by

$$P(\boldsymbol{\gamma}|\boldsymbol{y}) \approx C(\boldsymbol{y})(\sqrt{\sigma^2}w)^{|\boldsymbol{\gamma}|} \times \exp\left(-\frac{1}{2}\min_{\boldsymbol{\beta},\boldsymbol{\theta}} L_\rho(\boldsymbol{\beta},\boldsymbol{\theta})\right).$$

$$L_\rho(\boldsymbol{\beta},\boldsymbol{\theta}) = \log|\boldsymbol{\Phi}(\boldsymbol{\theta})| + \frac{(\boldsymbol{y}-\boldsymbol{F}_{\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}})'\boldsymbol{\Phi}^{-1}(\boldsymbol{\theta})(\boldsymbol{y}-\boldsymbol{F}_{\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}}) + \rho_1\sum_{k\in\boldsymbol{\gamma}}|\beta_k| + \rho_2\sum_{k\in\boldsymbol{\gamma}}\theta_k}{\sigma^2}.$$

- The approximation leads to a double penalized likelihood estimation problem
- Estimation: Coordinate descent algorithm

Summary

- Illustrates how to analyze multifidelity Computer Experiments using a real example.
- Analysis of computer experiments mainly based on GP models, in particular, ordinary kriging model.
- Proposed a Bayesian variable selection framework to improve prediction accuracy and model efficiency.

Dinner is ready!

Thank you!