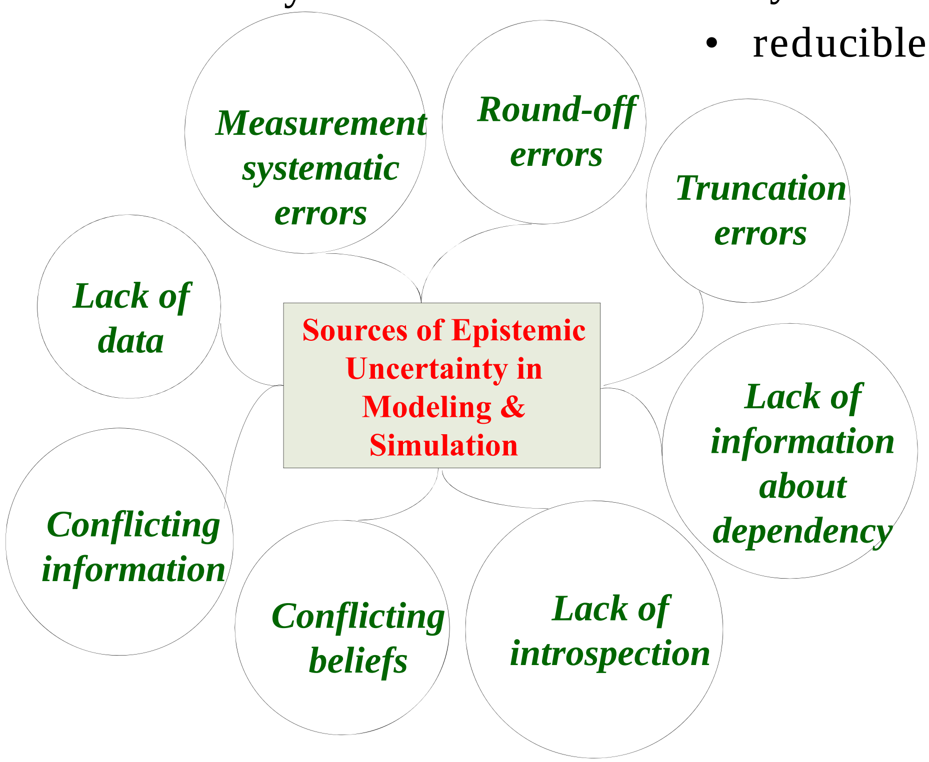
Generalized Interval Probability and Its Applications in Engineering

Yan Wang
Woodruff School of Mechanical Engineering
Georgia Institute of Technology
Atlanta, GA 30332, U.S.A.
yan.wang@me.gatech.edu

Uncertainty in Modeling & Simulation

- ***Aleatory Uncertainty:***
  - *inherent random dispersion* in the system. Also known as:
    - variability
    - random error
    - irreducible uncertainty
- ***Epistemic Uncertainty:***
  - due to *lack of perfect knowledge* about the system. Also known as:
    - incertitude
    - systematic error
    - reducible uncertainty



Imprecise Probability and Its Different Forms

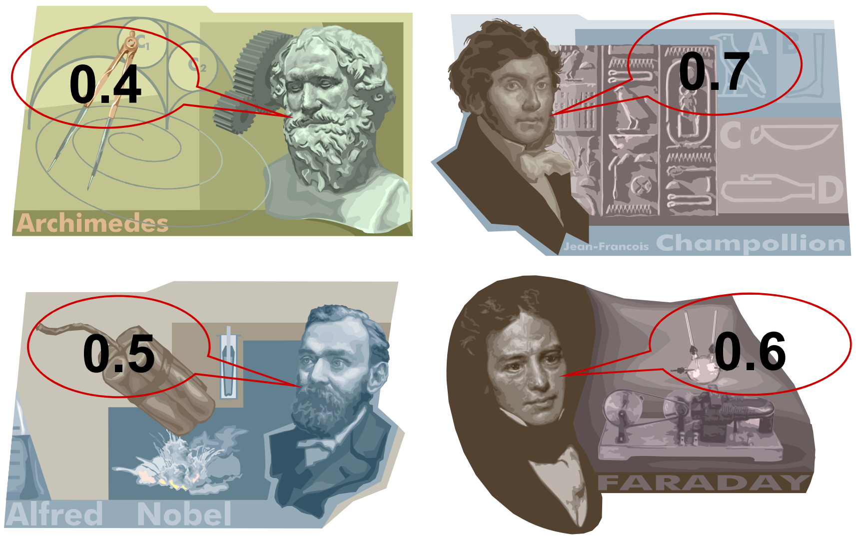
$$\mathbf{P} = [\underline{P}, \overline{P}]$$

- Imprecise probability is a *generalization* and *extension* of the classical probability theory that differentiates *aleatory* and *epistemic* uncertainties explicitly in the probabilistic measure.
- Several forms have been proposed:
  - *Dempster-Shafer evidence theory* [Dempster 1967; Shafer 1976]
  - *Coherent lower prevision* [Walley 1991]
  - *Probability bound analysis* [Ferson et al. 2002]
  - *Possibility theory* [Dubois & Prade 1988]
  - *Fuzzy probability* [Möller & Beer 2004]
  - *F-probability* [Weichselberger 2000]
  - *Random set* [Molchanov 2005]
  - *Cloud* [Neumaier 2004]

Overcome the Limitations of Classical Probability

- **Inconsistency**

$$\mathbf{p}(Netherlands\ wins\ World\ Cup) = [.4, .7]$$



- **Total Ignorance**

$$\mathbf{p}(Netherlands\ wins\ World\ Cup) = [0, 1]$$

- **Indeterminacy**

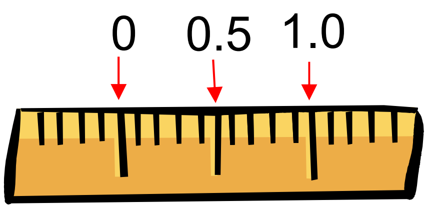
Tony Hayward, CEO of BP, … (was) asked about the top kill (to stop gulf oil spill), Mr. Hayward acknowledged that it was far from a sure fix. "We rate ***the probability of success between 60 percent and 70 percent*** *…"*

-- New York Times, May 24, 2010

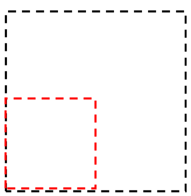
$$\mathbf{p}(success) = [.6, .7]$$

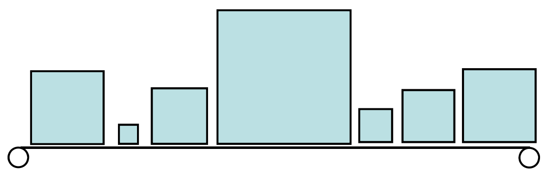
van Fraassen's Cube Factory Paradox

Laplace's *Principle of Insufficient Reason* requires a uniform distribution, which leads to different answers for the same problem.

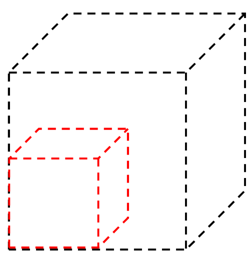


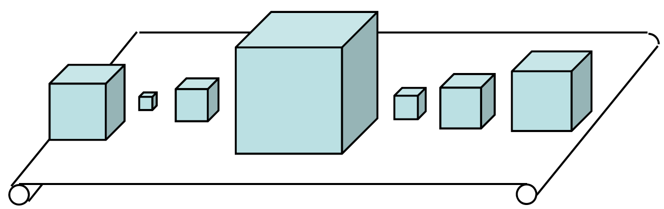
P(***side length*** of a randomly chosen cube $\leq 1/2$)=?





P(***face area*** of a randomly chosen cube $\leq 1/4$)=?





P(***volume*** of a randomly chosen cube $\leq 1/8$)=?

Assumptions in Dutch Book Arguments

"Probability is a choice."

1. You must post all betting quotients of all events at the beginning. The complete knowledge of all outcomes, including their relationships of dependency and mutual exclusiveness, in the world of the discourse is fully expressed in your belief.
   - The degrees of belief on all possible outcomes should be explicit without any hesitation and indeterminacy.
   - Doubt and vagueness are not permitted, even if you have a modest and diffident personality.
   - You should ***choose*** the subjective probability that may not be part of your *true* subjective feeling.

Assumptions in Dutch Book Arguments

"Value-adding activities are irrational."

2. You must accept all bets anyone wants to make at your posted quotients. With the listed prices, decisions of either 'buy' or 'sell' should be made immediately, with probability denoting the fair price at which you will both buy and sell the bet.

- There are only two possible options of decisions, i.e. **'not buy'='sell'**, and **'not sell'='buy'**.
- Non-value-adding activities are deemed as rational. A rational agent will make ***no profit*** through his/her buy and sell activities, whereas there is always a cunning bookie who tries to bankrupt you.

Generalized Interval Probability

- The generalized interval probability $\mathbf{p}: \mathcal{A} \mapsto [0,1]\times[0,1]$ obeys the axioms of Kolmogorov:
  - (1) $0 \le \mathbf{p}(E) \le 1 \; (\forall E \in \mathcal{A})$
  - (2) $\mathbf{p}(\Omega) = [1,1] = 1$
  - (3) For countable mutually disjoint events E_i's,

$$\mathbf{p}\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} \mathbf{p}(E_i)$$

$$\mathbf{p}(E_1 \cup E_2) := \mathbf{p}(E_1) + \mathbf{p}(E_2) - \mathrm{dual}\,\mathbf{p}(E_1 \cap E_2)$$

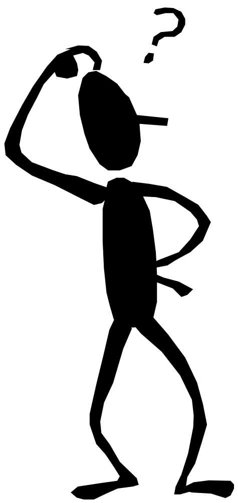
$$\mathbf{p}(E^C) := 1 - \mathrm{dual}\,\mathbf{p}(E)$$

***proper** interval* → $\mathrm{dual}[.1,.2] = [.2,.1]$ ← ***improper** interval*

$\mathrm{dual}[.2,.1] = [.1,.2]$

Generalized Interval vs. Set-based Interval

- Is it possible to calculate with *only zero and positive numbers* to solve all scientific/engineering/financial/… problems?



  - Yes, but the inclusion of negative number is much more convenient.
  - Similar to negative numbers, *improper intervals* can be considered as "negative" intervals.

Generalized Interval

- *Close relatives*
  - Modal Interval [Gardeñes et al. 1986; Vehí et al. 2000]
  - Directed Interval [Dimitrova et al. 1992; Markov 1995; Popova 2001]
  - AE Solution Set [Shary 1995; Kupriyanova 1995; Kreinovich et al. 1996; Goldsztejn 2005]

- classical interval is defined as "set"

$$[a, b]^* = \{x \in \mathbb{R} \mid a \leq x \leq b\}$$

- *Semi-group*: no invertibility

$[.1,.2]^* - [.1,.2]^* = [-.1, .1]^*$

- generalized interval is defined as "pair"

$$\mathbf{x} := [\underline{x}, \overline{x}] \ \ (\underline{x}, \overline{x} \in \mathbb{R})$$

- *Group*

$[.1,.2] - \text{dual}[.1,.2]$

$= [.1,.2] - [.2,.1] = [0,0] = 0$

- The widths of generalized intervals could reduce during calculation

$[1,3] + [2,1] = [3,4]$

$[1,3] + \text{dual}[1,3] = 4$

Generalized Interval for Uncertainty

- The width of an interval $\mathrm{wid}([\underline{x},\overline{x}]) = \left|\overline{x} - \underline{x}\right|$ captures epistemic uncertainty
- Uncertainty is propagated by Kaucher interval arithmetic [Kaucher 1980]

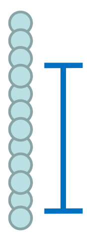
Algebraic Relation:	**Corresponding Logic Interpretation**	**Quantifier of z**	**Range Estimation of z**
[2,3]+[2,4]=[4,7]	$\forall x \in [2,3]^*, \forall y \in [2,4]^*, \exists z \in [4,7]^*\ \ x+y=z$	$\exists$	[4,7] is *complete*
[2,3]+[4,2]=[6,5]	$\forall x \in [2,3]^*, \forall z \in [5,6]^*, \exists y \in [2,4]^*\ \ x+y=z$	$\forall$	[5,6] is *sound*
[3,2]+[2,4]=[5,6]	$\forall y \in [2,4]^*, \exists x \in [2,3]^*, \exists z \in [5,6]^*\ \ x+y=z$	$\exists$	[5,6] is *complete*
[3,2]+[4,2]=[7,4]	$\forall z \in [4,7]^*, \exists x \in [2,3]^*, \exists y \in [2,4]^*\ \ x+y=z$	$\forall$	[4,7] is *sound*

complete = 'no underestimation'   *sound = 'no overestimation'*

Completeness vs. Soundness

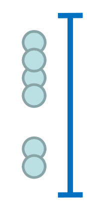
Complete

- All possibilities are included in the variation range estimation.
- No <u>Under</u>-estimation



Sound

- All estimated variations are possible.
- No <u>Over</u>-estimation



Is there any UQ method that can be both complete and sound?

"Thou Shalt Not Lie" (about your estimation).

Kaucher interval arithmetic

[Kaucher 1980]

$$\mathbf{x}+\mathbf{y} := \left[\underline{x}+\underline{y}, \overline{x}+\overline{y}\right]$$

$$\mathbf{x}-\mathbf{y} := \left[\underline{x}-\overline{y}, \overline{x}-\underline{y}\right]$$

$$\mathbf{x}\times\mathbf{y} := \begin{cases}
\left[\underline{x}\underline{y}, \overline{x}\overline{y}\right] & (\underline{x}\ge 0, \overline{x}\ge 0, \underline{y}\ge 0, \overline{y}\ge 0) \\
\left[\underline{x}\underline{y}, \underline{x}\overline{y}\right] & (\underline{x}\ge 0, \overline{x}\ge 0, \underline{y}\ge 0, \overline{y}< 0) \\
\left[\overline{x}\underline{y}, \overline{x}\overline{y}\right] & (\underline{x}\ge 0, \overline{x}\ge 0, \underline{y}< 0, \overline{y}\ge 0) \\
\left[\overline{x}\underline{y}, \underline{x}\overline{y}\right] & (\underline{x}\ge 0, \overline{x}\ge 0, \underline{y}< 0, \overline{y}< 0) \\
\left[\underline{x}\underline{y}, \overline{x}\underline{y}\right] & (\underline{x}\ge 0, \overline{x}< 0, \underline{y}\ge 0, \overline{y}\ge 0) \\
\left[\max\left(\underline{x}\underline{y}, \overline{x}\overline{y}\right), \min\left(\underline{x}\overline{y}, \overline{x}\underline{y}\right)\right] & (\underline{x}\ge 0, \overline{x}< 0, \underline{y}\ge 0, \overline{y}< 0) \\
[0,0] & (\underline{x}\ge 0, \overline{x}< 0, \underline{y}< 0, \overline{y}\ge 0) \\
\left[\overline{x}\overline{y}, \underline{x}\overline{y}\right] & (\underline{x}\ge 0, \overline{x}< 0, \underline{y}< 0, \overline{y}< 0) \\
\left[\underline{x}\overline{y}, \overline{x}\overline{y}\right] & (\underline{x}< 0, \overline{x}\ge 0, \underline{y}\ge 0, \overline{y}\ge 0) \\
[0,0] & (\underline{x}< 0, \overline{x}\ge 0, \underline{y}\ge 0, \overline{y}< 0) \\
\left[\min\left(\underline{x}\overline{y}, \overline{x}\underline{y}\right), \max\left(\underline{x}\underline{y}, \overline{x}\overline{y}\right)\right] & (\underline{x}< 0, \overline{x}\ge 0, \underline{y}< 0, \overline{y}\ge 0) \\
\left[\overline{x}\underline{y}, \underline{x}\underline{y}\right] & (\underline{x}< 0, \overline{x}\ge 0, \underline{y}< 0, \overline{y}< 0) \\
\left[\underline{x}\overline{y}, \overline{x}\underline{y}\right] & (\underline{x}< 0, \overline{x}< 0, \underline{y}\ge 0, \overline{y}\ge 0) \\
\left[\overline{x}\overline{y}, \overline{x}\underline{y}\right] & (\underline{x}< 0, \overline{x}< 0, \underline{y}\ge 0, \overline{y}< 0) \\
\left[\underline{x}\overline{y}, \underline{x}\underline{y}\right] & (\underline{x}< 0, \overline{x}< 0, \underline{y}< 0, \overline{y}\ge 0) \\
\left[\overline{x}\overline{y}, \underline{x}\underline{y}\right] & (\underline{x}< 0, \overline{x}< 0, \underline{y}< 0, \overline{y}< 0)
\end{cases}$$

Kaucher interval arithmetic

[Kaucher 1980]

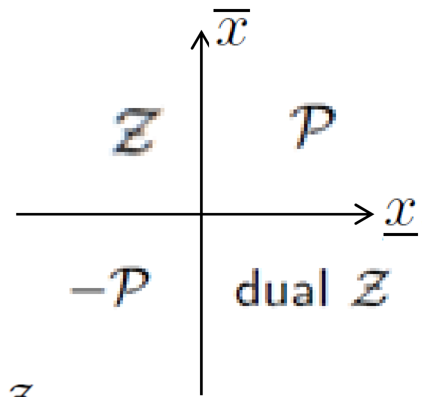
$$\mathbf{x}/\mathbf{y} := \begin{cases} [\underline{x}/\overline{y}, \overline{x}/\underline{y}] & (\underline{x} \geq 0, \overline{x} \geq 0, \underline{y} > 0, \overline{y} > 0) \\ [\overline{x}/\overline{y}, \underline{x}/\underline{y}] & (\underline{x} \geq 0, \overline{x} \geq 0, \underline{y} < 0, \overline{y} < 0) \\ [\underline{x}/\overline{y}, \overline{x}/\overline{y}] & (\underline{x} \geq 0, \overline{x} < 0, \underline{y} > 0, \overline{y} > 0) \\ [\overline{x}/\underline{y}, \underline{x}/\underline{y}] & (\underline{x} \geq 0, \overline{x} < 0, \underline{y} < 0, \overline{y} < 0) \\ [\underline{x}/\underline{y}, \overline{x}/\underline{y}] & (\underline{x} < 0, \overline{x} \geq 0, \underline{y} > 0, \overline{y} > 0) \\ [\overline{x}/\overline{y}, \underline{x}/\overline{y}] & (\underline{x} < 0, \overline{x} \geq 0, \underline{y} < 0, \overline{y} < 0) \\ [\underline{x}/\underline{y}, \overline{x}/\overline{y}] & (\underline{x} < 0, \overline{x} < 0, \underline{y} > 0, \overline{y} > 0) \\ [\overline{x}/\underline{y}, \underline{x}/\overline{y}] & (\underline{x} < 0, \overline{x} < 0, \underline{y} < 0, \overline{y} < 0) \end{cases}$$

Kaucher interval arithmetic

[Kaucher 1980]

$$[\underline{x},\overline{x}] \oplus [\underline{y},\overline{y}] = [\underline{x}+\underline{y},\overline{x}+\overline{y}]$$

$$[\underline{x},\overline{x}] \ominus [\underline{y},\overline{y}] = [\underline{x}-\underline{y},\overline{x}-\overline{y}]$$



	$\boldsymbol{y} \in \mathcal{P}$	$\boldsymbol{y} \in \mathcal{Z}$	$\boldsymbol{y} \in -\mathcal{P}$	$\boldsymbol{y} \in$ dual $\mathcal{Z}$
$\boldsymbol{x} \in \mathcal{P}$	$[\underline{x}\underline{y},\overline{x}\overline{y}]$	$[\overline{x}\underline{y},\overline{x}\overline{y}]$	$[\overline{x}\underline{y},\underline{x}\overline{y}]$	$[\underline{x}\underline{y},\underline{x}\overline{y}]$
$\boldsymbol{x} \in \mathcal{Z}$	$[\underline{x}\overline{y},\overline{x}\overline{y}]$	$[\min\{\underline{x}\overline{y},\overline{x}\underline{y}\},\max\{\underline{x}\underline{y},\overline{x}\overline{y}\}]$	$[\overline{x}\underline{y},\underline{x}\underline{y}]$	0
$\boldsymbol{x} \in -\mathcal{P}$	$[\underline{x}\overline{y},\overline{x}\underline{y}]$	$[\underline{x}\overline{y},\underline{x}\underline{y}]$	$[\overline{x}\overline{y},\underline{x}\underline{y}]$	$[\overline{x}\overline{y},\overline{x}\underline{y}]$
$\boldsymbol{x} \in$ dual$\mathcal{Z}$	$[\underline{x}\underline{y},\overline{x}\underline{y}]$	0	$[\overline{x}\overline{y},\underline{x}\overline{y}]$	$[\max\{\underline{x}\underline{y},\overline{x}\overline{y}\},\min\{\underline{x}\overline{y},\overline{x}\underline{y}\}]$

Kaucher multiplication $[\underline{x},\overline{x}] \otimes [\underline{y},\overline{y}]$ table

$$[\underline{x},\overline{x}] \oslash [\underline{y},\overline{y}] = [\underline{x},\overline{x}] \otimes [1/\overline{y},1/\underline{y}]$$

More about Generalized Interval Probability

- Conditional probability

$$\mathbf{p}\left(E \mid C\right) := \frac{\mathbf{p}\left(E \cap C\right)}{\text{dual}\,\mathbf{p}\left(C\right)} = \left[\frac{\underline{p}\left(E \cap C\right)}{\underline{p}\left(C\right)}, \frac{\overline{p}\left(E \cap C\right)}{\overline{p}\left(C\right)}\right]$$

Wang Y. (2010) Imprecise probabilities based on generalised intervals for system reliability assessment. *International Journal of Reliability & Safety*, **4**(4): 319-342

More about Generalized Interval Probability

- Conditional independence

$$\mathbf{p}\left(A \cap B \mid C\right) = \mathbf{p}\left(A \mid C\right)\mathbf{p}\left(B \mid C\right)$$

- This definition of Independence is Graphoid
  - Symmetry
  - Decomposition
  - *Composition*
  - Weak union
  - Contraction
  - *Reduction*
  - Redundancy
  - Intersection

Wang Y., "Independence in generalized interval probability." *Proc. of 1st Int. Conf. on Vulnerability and Risk Analysis and Management (ICVRAM 2011) and 5th International Symposium on Uncertainty Modeling and Analysis (ISUMA 2011), April 11-13, 2011, Hyattsville, Maryland*, pp.37-44.

More about Generalized Interval Probability

- Generalized Interval Bayes' Rule (GIBR)

$$\mathbf{p}\left(E_i \mid A\right) = \frac{\mathbf{p}\left(A \mid E_i\right)\mathbf{p}\left(E_i\right)}{\sum_{j=1}^{n} \mathrm{dual}\,\mathbf{p}\left(A \mid E_j\right) \mathrm{dual}\,\mathbf{p}\left(E_j\right)}$$

Logic Coherence Constraint (L.C.C.)

For a mutually disjoint event partition, the sum of the interval probabilities is always one.

- e.g. *system state*:

Focal events   Non-focal events

$$\mathbf{p}(down) = [0.2, 0.3] \quad \mathbf{p}(idle) = [0.3, 0.5] \quad \mathbf{p}(busy) = [0.5, 0.2]$$

- logically consistent with precise probability

$$\forall p_1 \in [0.2,0.3]^*, \forall p_2 \in [0.3,0.5]^*, \exists p_3 \in [0.2,0.5]^*, p_1 + p_2 + p_3 = 1$$

L.C.C. also implies ...

- *Avoid sure loss*
  - **IP**: $\sum_i \underline{p}_i \le 1$, $\sum_i \overline{p}_i \ge 1$
  - **GIP**: $\sum_{i\in\mathcal{P}} \underline{p}_i + \sum_{j\in\mathcal{J}} \overline{p}_j \le 1$,

    $\sum_{i\in\mathcal{P}} \overline{p}_i + \sum_{j\in\mathcal{J}} \underline{p}_j \ge 1$
- *Coherence between previsions*
  - **IP**: $\sum_{i\ne k} \underline{p}_i + \overline{p}_k \le 1$ for all k
  - **GIP**: $\sum_{i\in\mathcal{P}} \underline{p}_i + \sum_{j\in\mathcal{J},j\ne k} \overline{p}_j + \underline{p}_k \le 1$ when $k \in \mathcal{J}$

    or $\overline{p}_k + \sum_{i\in\mathcal{P},i\ne k} \underline{p}_i + \sum_{j\in\mathcal{J}} \overline{p}_j \le 1$ when $k \in \mathcal{P}$

Sound but *Incomplete* GIBR

- For example,

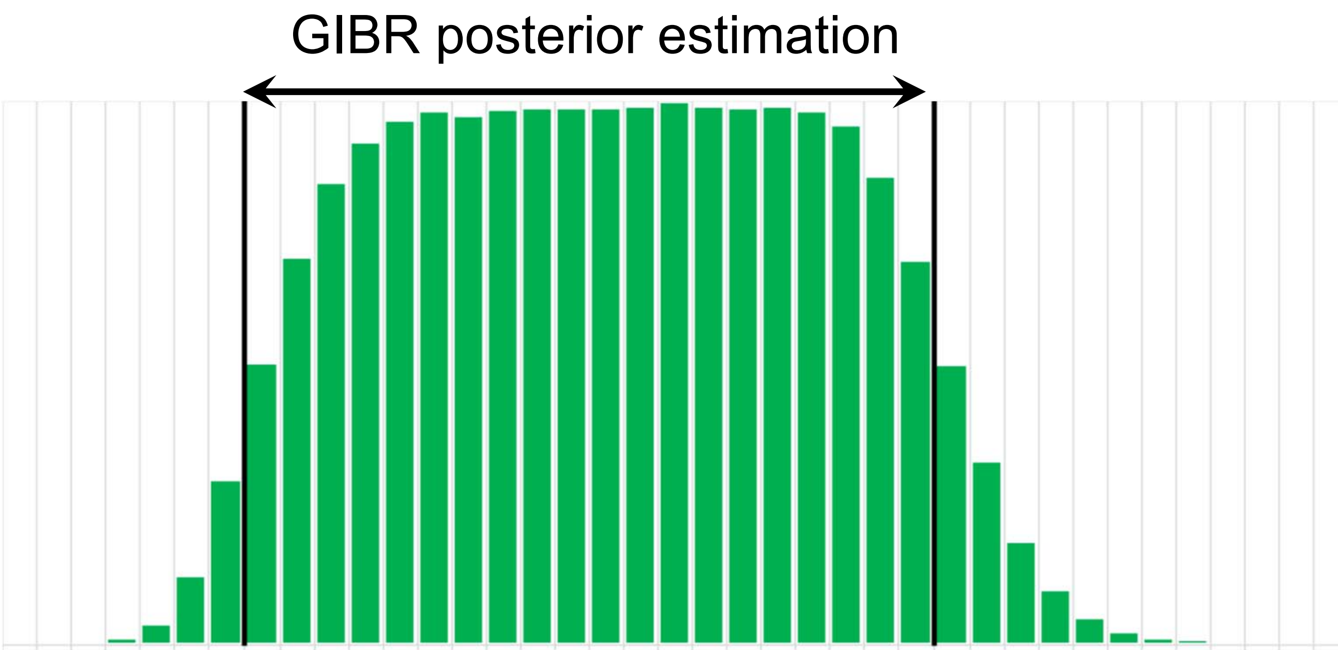
$$P(\gamma|\alpha) = \frac{\begin{bmatrix} P(\alpha|\beta_1) & P(\alpha|\beta_2) & P(\alpha|\beta_3) & P(\alpha|\beta_4) \end{bmatrix} \begin{bmatrix} P(\beta_1|\gamma) \\ P(\beta_2|\gamma) \\ P(\beta_3|\gamma) \\ P(\beta_4|\gamma) \end{bmatrix} \cdot P(\gamma)}{dual\left(\begin{bmatrix} P(\alpha|\beta_1) & P(\alpha|\beta_2) & P(\alpha|\beta_3) & P(\alpha|\beta_4) \end{bmatrix} \begin{bmatrix} P(\beta_1|\gamma) & P(\beta_1|\gamma^C) \\ P(\beta_2|\gamma) & P(\beta_2|\gamma^C) \\ P(\beta_3|\gamma) & P(\beta_3|\gamma^C) \\ P(\beta_4|\gamma) & P(\beta_4|\gamma^C) \end{bmatrix} \begin{bmatrix} P(\gamma) \\ P(\gamma^C) \end{bmatrix} \right)}$$

$$= \frac{\begin{bmatrix} [0.3880,0.9973] \\ [0.7415,0.9024] \\ [0.3667,0.8688] \\ [0.1655,0.4140] \end{bmatrix}^T \begin{bmatrix} [0.1197,0.1826] \\ [0.1662,0.3080] \\ [0.3587,0.4603] \\ [0.3554,0.0491] \end{bmatrix} \cdot [0.3000,0.4500]}{dual\left(\begin{bmatrix} [0.3880,0.9973] \\ [0.7415,0.9024] \\ [0.3667,0.8688] \\ [0.1655,0.4140] \end{bmatrix}^T \begin{bmatrix} [0.1197,0.1826] & [0.0410,0.1930] \\ [0.1662,0.3080] & [0.1527,0.1772] \\ [0.3587,0.4603] & [0.3683,0.4682] \\ [0.3554,0.0491] & [0.4380,0.1616] \end{bmatrix} \begin{bmatrix} [0.3000,0.4500] \\ [0.7000,0.5500] \end{bmatrix} \right)} = [0.3143,0.4658]$$

Blumer, J. (2015) *Cross-Scale Model Validation with Aleatory and Epistemic Uncertainty*, M.S. Thesis, Georgian Institute of Technology

Sound but *Incomplete* GIBR

- GIBR estimation compared with MC sampling

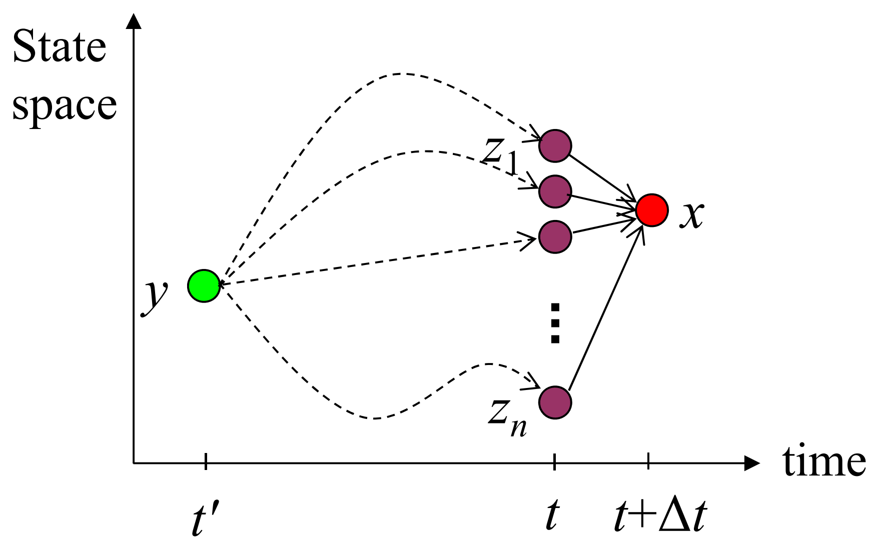


Bin Min	Bin Max	Count
0.2688	0.2764	0
0.2764	0.2840	26
0.2840	0.2915	259
0.2915	0.2991	1581
0.2991	0.3067	6092
0.3067	0.3143	14919
0.3143	0.3218	25788
0.3218	0.3294	35631
0.3294	0.3370	42573
0.3370	0.3446	46288
0.3446	0.3521	48334
0.3521	0.3597	49104
0.3597	0.3673	48662
0.3673	0.3749	49339
0.3749	0.3824	49445
0.3824	0.3900	49386
0.3900	0.3976	49409
0.3976	0.4052	49616
0.4052	0.4128	50013
0.4128	0.4203	49547
0.4203	0.4279	49373
0.4279	0.4355	49548
0.4355	0.4431	49078
0.4431	0.4506	47797
0.4506	0.4582	43053
0.4582	0.4658	35292
0.4658	0.4734	25669
0.4734	0.4809	16649
0.4809	0.4885	9280
0.4885	0.4961	4802
0.4961	0.5037	2176
0.5037	0.5112	826
0.5112	0.5188	328
0.5188	0.5264	91
0.5264	0.5340	19
0.5340	0.5415	6
0.5415	0.5491	1
0.5491	0.5567	0

Generalized Chapman-Kolmogorov Equation

- "First-principles" model of the Markovian property

$$\mathbf{p}(x,t+\Delta t \mid y,t') = \int dz\,\mathbf{p}(x,t+\Delta t \mid z,t)\,\mathbf{p}(z,t \mid y,t')$$



Generalized Differential C-K Equation

- Define *derivative* of generalized interval probability

$$= \int dz \mathbf{p}(x, t+\Delta t \mid z, t)\mathbf{p}(z, t \mid y, t') \quad gCKE$$

$$\frac{\partial}{\partial t}\mathbf{p}(x,t \mid y,t') := \lim_{\Delta t \to 0}\frac{1}{\Delta t}\left\{\mathbf{p}(x,t+\Delta t \mid y,t') - \text{dual}\,\mathbf{p}(x,t \mid y,t')\right\}$$

$$\times \int dz \mathbf{p}(z, t+\Delta t \mid x, t) = 1 \quad L.C.C.$$

$$= \lim_{\Delta t \to 0}\frac{1}{\Delta t}\left\{\int dz \begin{bmatrix} \mathbf{p}(x,t+\Delta t \mid z,t)\mathbf{p}(z,t \mid y,t') \\ -\text{dual}\,\mathbf{p}(z,t+\Delta t \mid x,t)\mathbf{p}(x,t \mid y,t') \end{bmatrix}\right\}$$

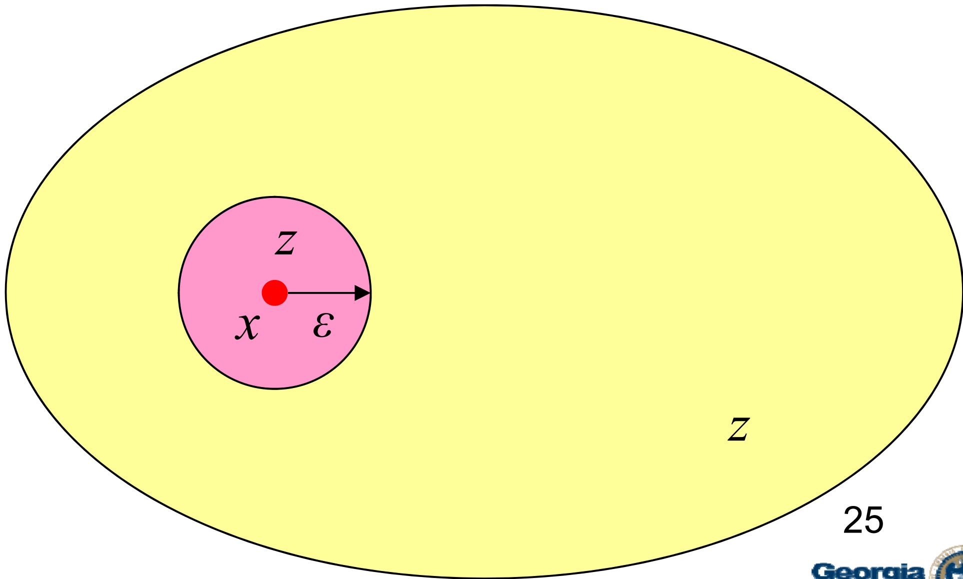
Generalized Differential C-K Equation (cont'd)

$$\frac{\partial}{\partial t}\mathbf{p}(x,t \mid y,t') = \lim_{\Delta t \to 0} \frac{1}{\Delta t}\left\{ \int dz \begin{bmatrix} \mathbf{p}(x,t+\Delta t \mid z,t)\mathbf{p}(z,t \mid y,t') \\ -\text{dual}\,\mathbf{p}(z,t+\Delta t \mid x,t)\mathbf{p}(x,t \mid y,t') \end{bmatrix} \right\}$$

- Spatially divide the state space into two sub-domains with $\int dz[\cdot] = \int_{\|x-z\|\le\varepsilon} dz[\cdot] + \int_{\|x-z\|>\varepsilon} dz[\cdot]$

- Therefore

$$\frac{\partial}{\partial t}\mathbf{p}(x,t \mid y,t') = \mathbf{I}_{\|x-z\|\le\varepsilon} + \mathbf{I}_{\|x-z\|>\varepsilon}$$



Generalized Differential C-K Equation (cont'd)

- Then we have the Gen. Diff. C-K Eq. as

$$\frac{\partial}{\partial t}\mathbf{p}(x,t\mid y,t') = -\operatorname{dual}\sum_{i=1}^{n}\frac{\partial \mathbf{A}_i(x,t)}{\partial x_i}\mathbf{p}(x,t\mid y,t') + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 \mathbf{B}_{ij}(x,t)}{\partial x_i \partial x_j}\mathbf{p}(x,t\mid y,t') + \int dz\mathbf{W}(x\mid z,t)\mathbf{p}(z,t\mid y,t') - \operatorname{dual}\int dz\mathbf{W}(z\mid x,t)\mathbf{p}(x,t\mid y,t')$$

- *Local Drift-Diffusion* - Generalized Fokker-Planck Equation

$$\mathbf{I}_{\|x-z\|\le\varepsilon} = -\operatorname{dual}\sum_{i=1}^{n}\frac{\partial \mathbf{A}_i(x,t)}{\partial x_i}\mathbf{p}(x,t\mid y,t') + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 \mathbf{B}_{ij}(x,t)}{\partial x_i \partial x_j}\mathbf{p}(x,t\mid y,t')$$

- *Global Jump* - Interval Master Equation

$$\mathbf{I}_{\|x-z\|>\varepsilon} = \int dz\mathbf{W}(x\mid z,t)\mathbf{p}(z,t\mid y,t') - \operatorname{dual}\int dz\mathbf{W}(z\mid x,t)\mathbf{p}(x,t\mid y,t')$$

Generalized Fokker-Planck Equation

$$\frac{\partial}{\partial t}\mathbf{p}(x,t \mid y,t') = -\,\mathrm{dual}\sum_{i=1}^{n}\frac{\partial \mathbf{A}_i(x,t)}{\partial x_i}\mathbf{p}(x,t \mid y,t') + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 \mathbf{B}_{ij}(x,t)}{\partial x_i \partial x_j}\mathbf{p}(x,t \mid y,t')$$

- *short-time transition probability density*

$$\mathbf{p}(x',t+\tau \mid x,t) \propto \exp\left(-\frac{1}{2\tau}[x'-x-\mathbf{A}(x,t)\tau]^T\,\mathbf{B}^{-1}(x,t)[x'-x-\mathbf{A}(x,t)\tau]\right)$$

- solved by a *path integral* (PI) method
  - initial distribution $\mathbf{Q}(t_0)$
  - short-time transition probabilities $\mathbf{P}(t_0),\ldots,\mathbf{P}(t_{k-1})$
  - final distribution $\mathbf{Q}(t_k)=\mathbf{P}(t_{k-1})\ldots\mathbf{P}(t_0)\mathbf{Q}(t_0)$

Markov Logic Coherence Constraint

- **Theorem**. Given an interval matrix $\mathbf{P}=[\mathbf{p}_{ij}]_{n\times n}$ ($\mathbf{P}\in \mathrm{KR}^{n\times n}$) and an interval vector $\mathbf{Q}=[\mathbf{q}_i]_{n\times 1}$ ($\mathbf{Q}\in \mathrm{KR}^{n}$) with their respective elements as generalized interval probabilities with $\sum_{i=1}^{n}\mathbf{p}_{ij}=1$ and $\sum_{i=1}^{n}\mathbf{q}_i=1$, if $\mathbf{T}=\mathbf{PQ}=[\mathbf{q}_i]_{n\times 1}$, then the elements of $\mathbf{T}=[\mathbf{t}_i]_{n\times 1}$ also satisfy $\sum_{i=1}^{n}\mathbf{t}_i=1$.

Wang Y. (2015) Stochastic dynamics simulation with generalized interval probability. *International Journal of Computer Mathematics*, **92**(3): 623-642

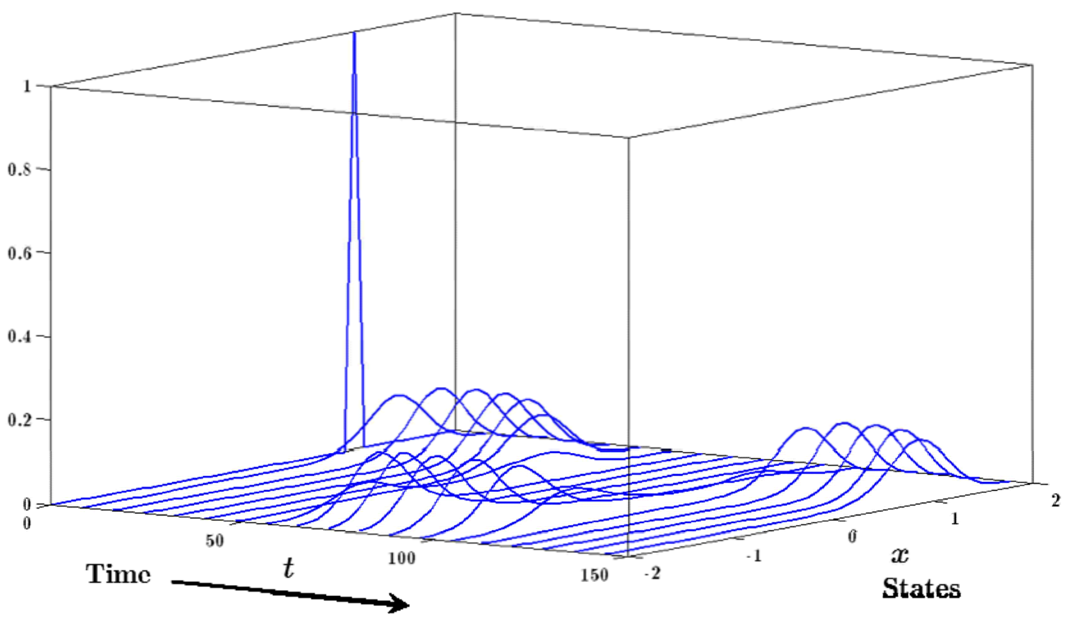
Gen. F-P Equation – Example 1

- Bi-stable Stochastic Resonance System

$$dx/dt = c_1 x - c_2 x^3 + a_0 \sin(2\pi f_0 t) + N(t) \qquad with\ N(t) = \sqrt{2B}\,\xi(t)$$

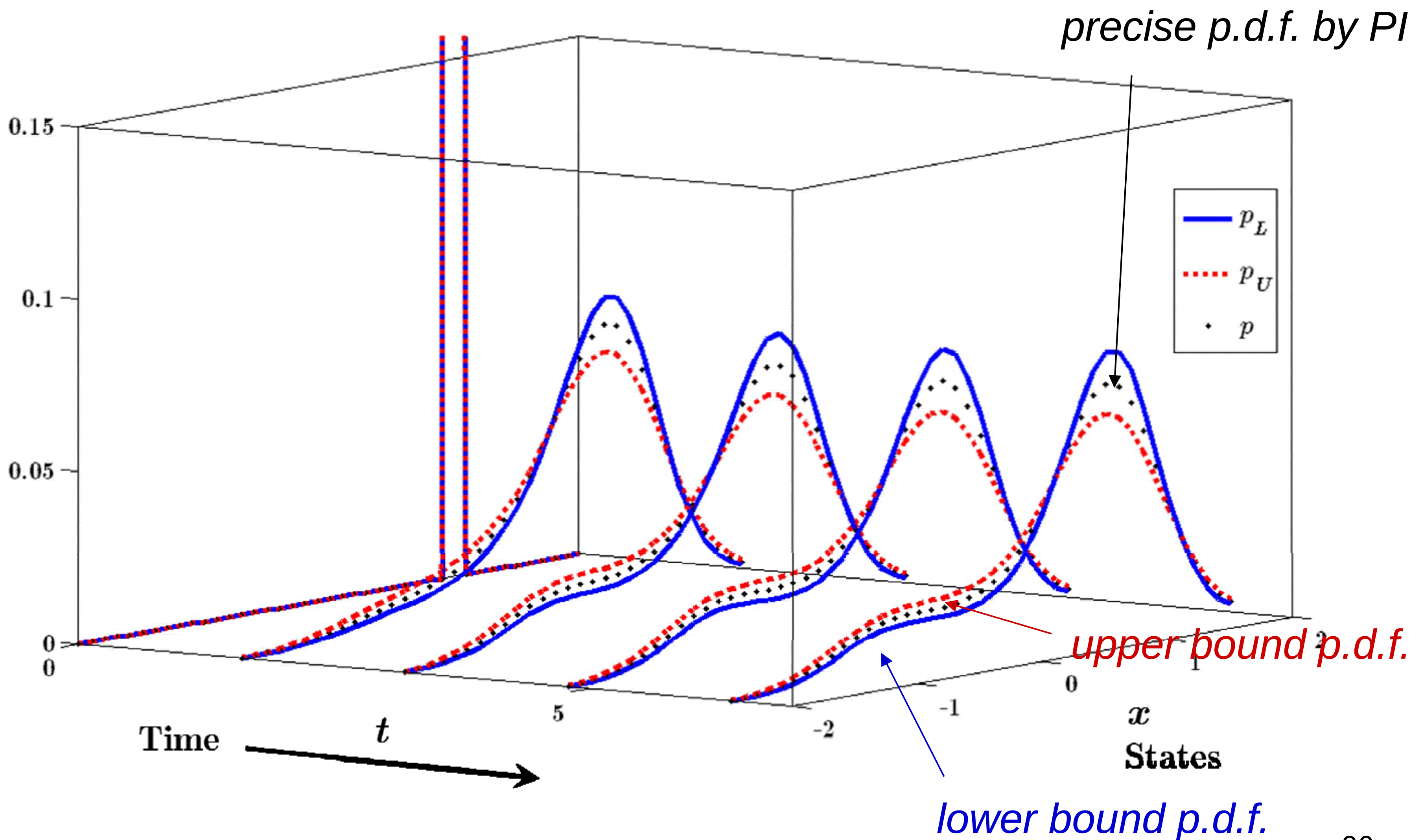
- Solve the equivalent F-P equation

$$\frac{\partial}{\partial t} p(x,t) = -\frac{\partial}{\partial x}\big(A p(x,t)\big) + \frac{\partial^2}{\partial x^2}\big(B p(x,t)\big) \qquad with\ A = c_1 x - c_2 x^3 + a_0 \sin(2\pi f_0 t)$$



Gaussian initial dist.
$c_1 = c_2 = 1$
$a_0 = 1$
$f_0 = 0.01$
$B = 0.31$
$\mathbf{A} = [A - 0.1, A + 0.1]$
$\mathbf{B} = [B - 0.031, B + 0.031]$

Gen. F-P Equation – Example 1 (cont'd)



Gen. F-P Equation – Example 2

- Van der Pol Oscillator

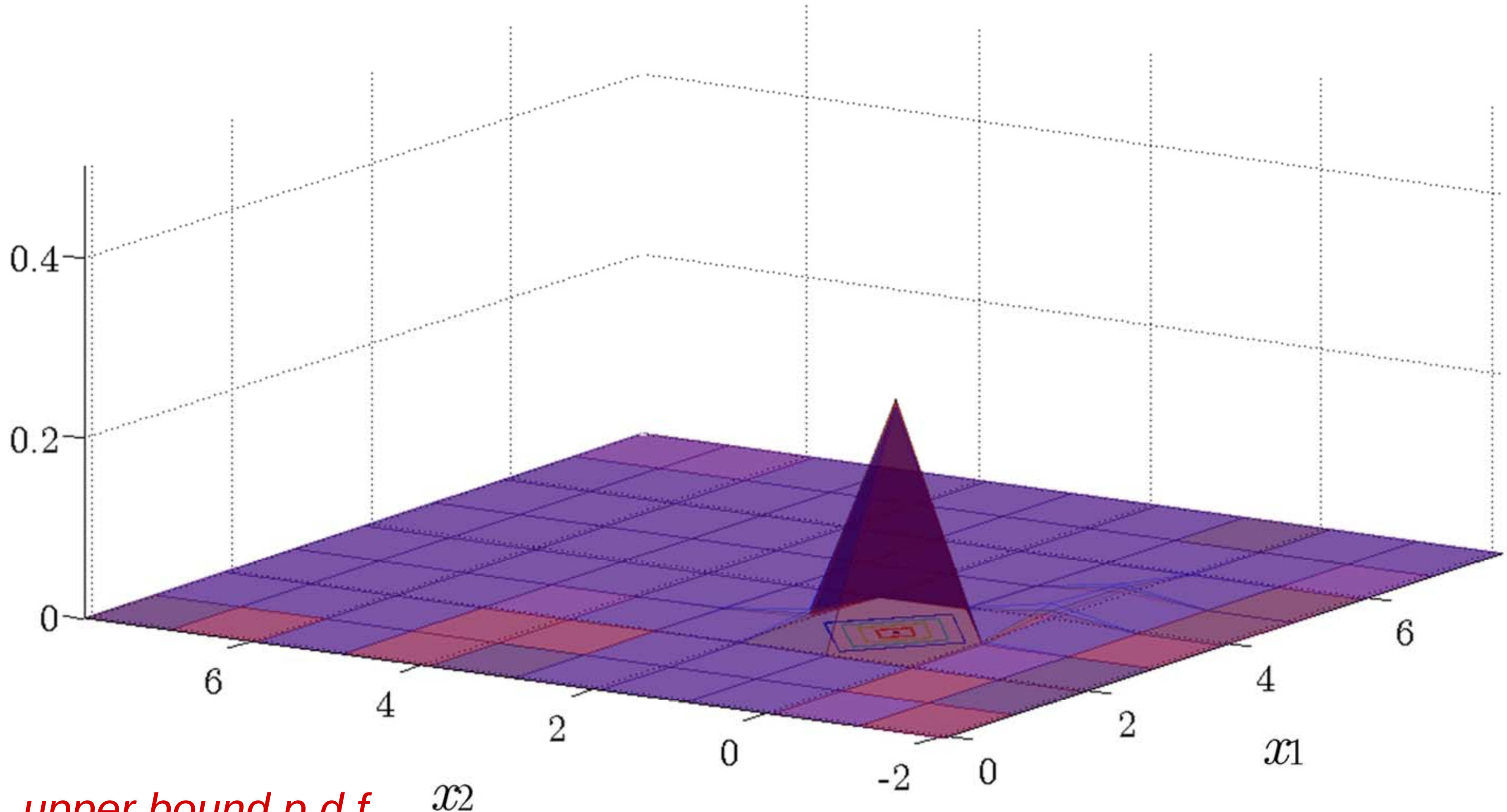
$$\frac{\partial}{\partial t}p(x_1,x_2,t) = -\frac{\partial}{\partial x_1}\left(g_1(x_1,x_2)p(x_1,x_2,t)\right) - \frac{\partial}{\partial x_2}\left(g_2(x_1,x_2)p(x_1,x_2,t)\right) + D\frac{\partial^2}{\partial x_2^2}p(x_1,x_2,t)$$

$$A = \begin{bmatrix} g_1(x_1,x_2) = x_2 \\ g_2(x_1,x_2) = 2\zeta\omega_0(1-\varepsilon x_1^2)x_2 - \omega_0^2 x_1 \end{bmatrix}$$

$$B = \begin{bmatrix} D & 0 \\ 0 & Dx_2^2 \end{bmatrix}$$

ζ=0.05
ω_0=1
ε=1
D=0.1
A=A±0.1A
B=B±0.1B

Gen. F-P Equation – Example 2 (cont'd)



upper bound p.d.f.

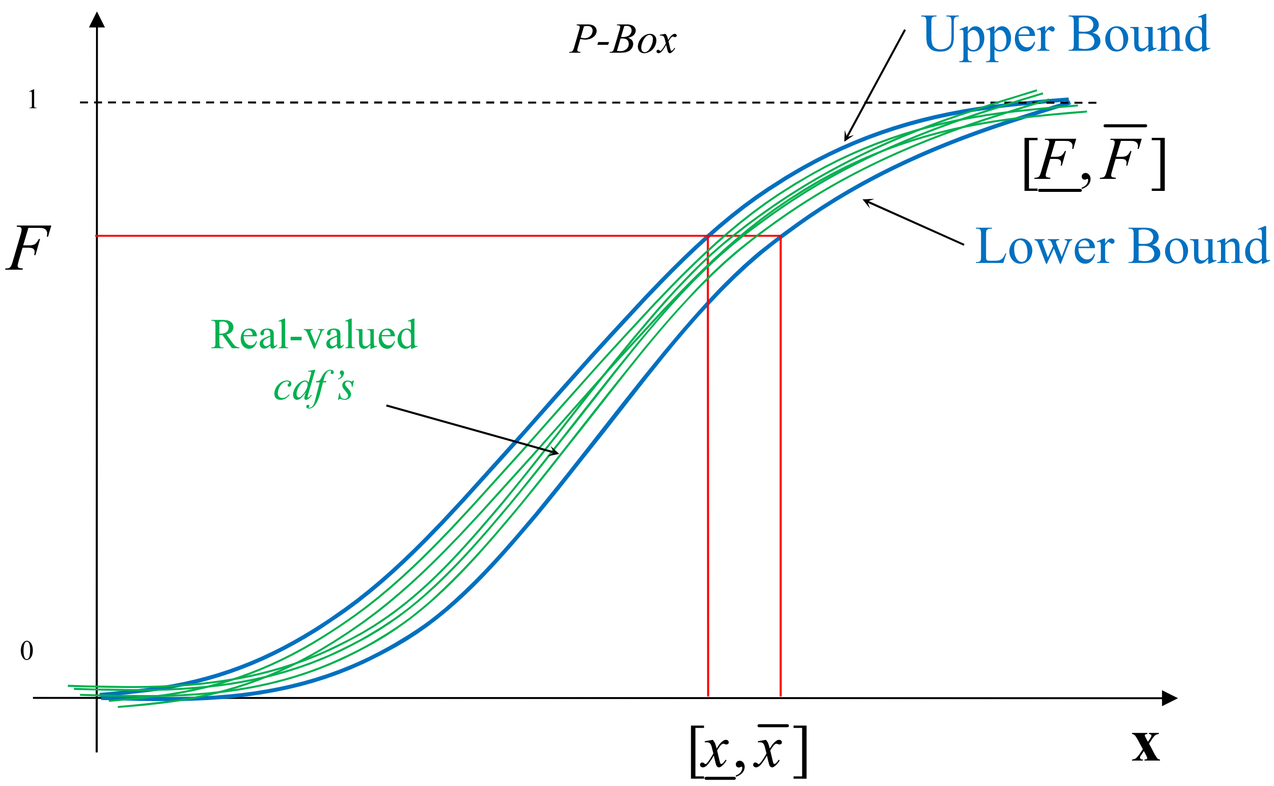
real-valued p.d.f.

lower bound p.d.f.

Δt=0.125

t=0~1.0

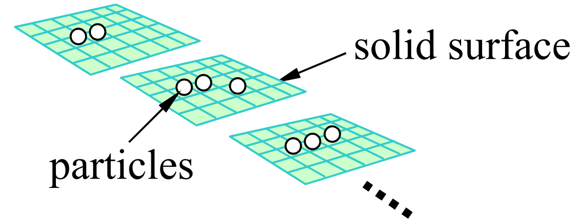
Random Set Sampling



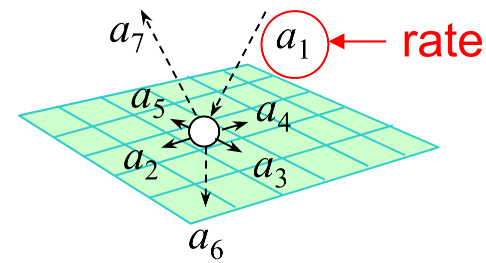
Kinetic Monte Carlo (KMC)

- KMC is a type of discrete-event stochastic simulation methods
- KMC is extensively used in physics and chemistry
- KMC
  - defines a discrete set of **states** of the system (i.e. all possible configurations)
  - simulates state transitions between states which are triggered by **events** (also called *processes* in chemistry-oriented literature) that cause state changes.

- e.g. vapor deposition



(a) discrete **states**



(b) possible **events**

- e.g. $2H_2O \leftrightarrow 2H_2 + O_2$

states: combinations of # *of species*

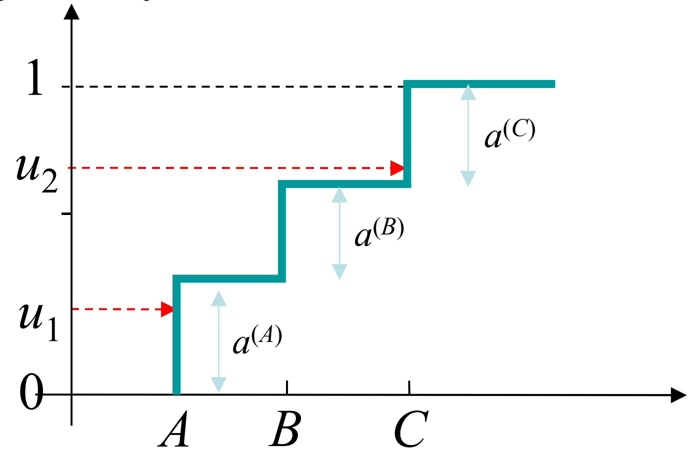
H_2O	n	n-2	n-4	…
H_2	0	2	4	…
O_2	0	1	2	…

events: reactions

propensity → a_1: $\rightarrow$ a_2: $\leftarrow$

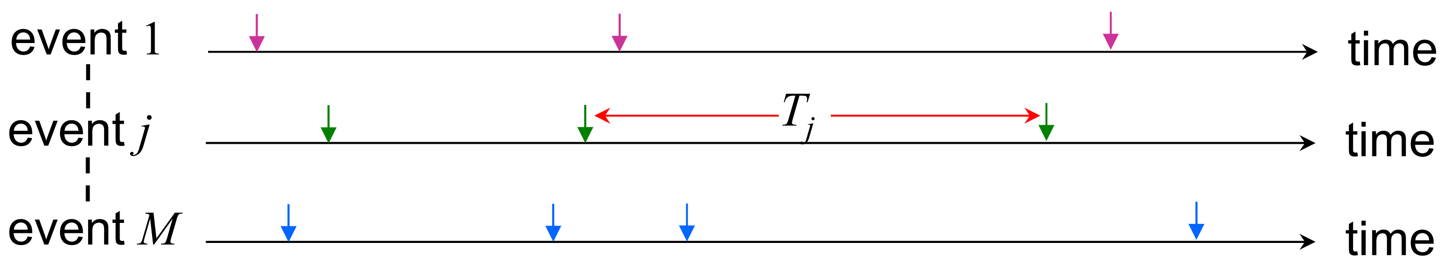
KMC Sampling Algorithm

- For each iteration
  - Randomly choose an event out of all possible ones
  - Update the system state as a result of the event
  - Randomly sample a time duration Δt of the event
  - Update the system clock: $t = t + \Delta t$



Uncertainty associated with rate a is modeled by interval $[a_L, a_U]$

How KMC Works – event selection



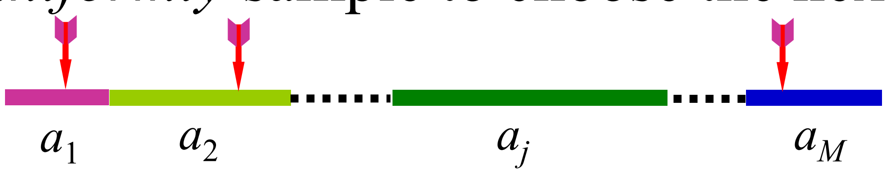
- The inter-arrival time T_j of event j is assumed to be *exponentially* distributed as

$$T_j \sim Exponential(a_j)$$

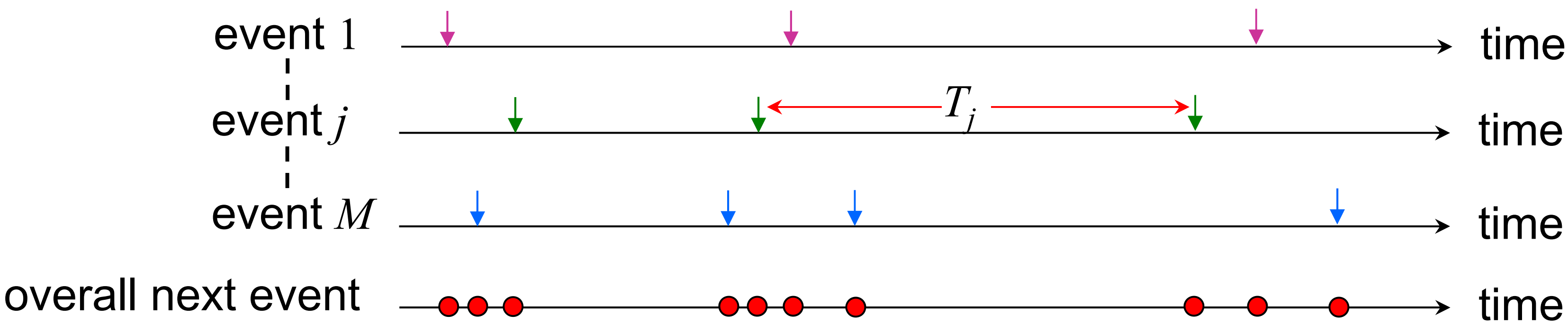
- The probability that the inter-arrival time of event j is the minimum among M *independent* events

$$\Pr[T_j = \min(T_1, \ldots, T_M)] = a_j / (a_1 + \cdots + a_M)$$

- Therefore, *uniformly* sample to choose the next event



How KMC Works – clock advancement



- The inter-arrival time T_j of event j is

$$P\left(T_j \le \tau\right) = 1 - \exp\left(-a_j \tau\right) \quad (j = 1, \ldots, M)$$

- The earliest time $T^{(1)}=\min(T_1,\ldots,T_M)$ of any event occurs is

$$P\left(T^{(1)} \le \tau\right) = 1 - P\left(T^{(1)} > \tau\right) = 1 - \prod_{j=1}^{M} P\left(T_j > \tau\right) = 1 - \exp\left(-\left(\sum_{j=1}^{M} a_j\right)\tau\right)$$

- Therefore, *exponentially* sample to advance clock

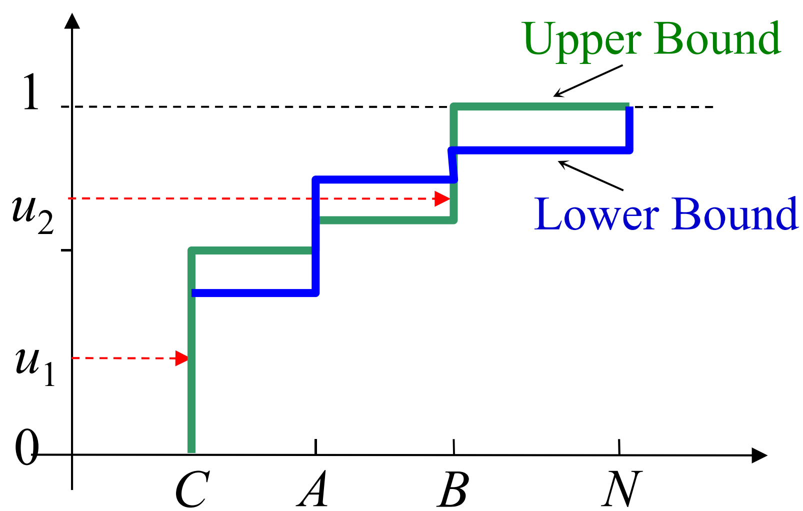
$$T_L = -\ln\rho / \sum_{j=1}^{M} a_j \quad \text{where} \quad \rho \sim Uniform(0,1)$$

Uncertainty in KMC Simulation

- Uncertain **rate** or **propensity** a_j's
- Sources
  - kinetic rate in biochemical reactions varies over time instead of a constant because of the macromolecular crowding effect [Berry 2002; Schnell & Turner 2002]
    - Existing solution: use fractal & Zipf-Mandelbrot relationships
  - biochemical processes such as diffusion, translocation, protein synthesis and folding do not occur instantaneously and are often affected by spatial homogeneities [Bratsun et al. 2005; Barrio et al. 2006; Burrage et al. 2007]
    - Existing solution: delay stochastic simulation algorithm

Multi-Event Algorithm – a simple illustration

- Given events A,B,C with the respective rates $\boldsymbol{a}_1$=[1,3], $\boldsymbol{a}_2$=[1,3] and $\boldsymbol{a}_3$=[4,5] as proper or precise intervals
  1. Sort rates based on uncertain level (widths of intervals) in the ascending order $\boldsymbol{a}_3$=[4,5], $\boldsymbol{a}_1$=[1,3], $\boldsymbol{a}_2$=[1,3]
  2. "flip" in an alternating pattern (apply dual) $\boldsymbol{a}'_3$=[4,5], $\boldsymbol{a}'_1$=[3,1], $\boldsymbol{a}'_2$=[1,3]
  3. Introduce a *null event* N (in this example, rate $\boldsymbol{a}_N$=[1,0]) so that $a_0=\boldsymbol{a}'_1+\boldsymbol{a}'_2+\boldsymbol{a}'_3+\boldsymbol{a}'_N$ is a precise number (here a_0=[9,9])
  4. Build empirical c.d.f. with probability mass $\boldsymbol{p}_j = \boldsymbol{a}'_j/a_0$ for event j
  5. As shown below, sample u~*Uniform*(0,1) and choose either one or two events to fire



[Wang (2013), *Soft Comp.*, **17**(8)]

Multi-Event Algorithm

- For M possible events with interval rates $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_M$, the assorted ones are $\boldsymbol{a}^{(1)}, \boldsymbol{a}^{(2)}, \ldots, \boldsymbol{a}^{(M)}$, where

$$\mathrm{wid}\left(\boldsymbol{a}^{(1)}\right) \le \mathrm{wid}\left(\boldsymbol{a}^{(2)}\right) \le \cdots \le \mathrm{wid}\left(\boldsymbol{a}^{(M)}\right)$$

- Define *-sum, denoted by $\sum^*$, recursively as

$$\sum_{j=1}^{*J} \boldsymbol{a}^{(j)} := \mathrm{dual}\left(\sum_{j=1}^{*J-1} \boldsymbol{a}^{(j)}\right) + \boldsymbol{a}^{(J)} \quad \left(J = 2,\ldots,M\right) \quad \text{with } \sum_{j=1}^{*J=1} \boldsymbol{a}^{(j)} = \boldsymbol{a}^{(1)}$$

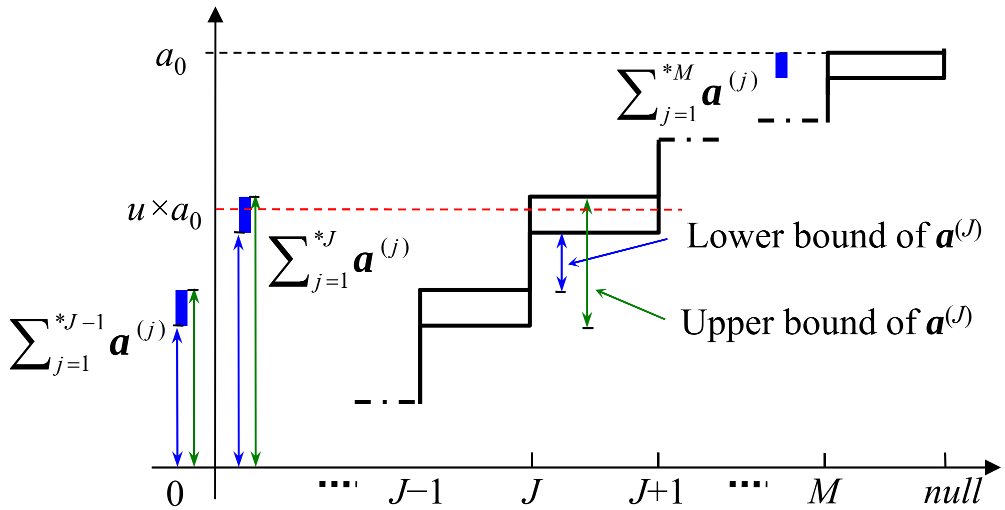
Theorem. If generalized intervals $\boldsymbol{x}$ and $\boldsymbol{y}$ are not improper and $\mathrm{wid}\left(\boldsymbol{x}\right) \le \mathrm{wid}\left(\boldsymbol{y}\right)$, then $\mathrm{dual}\,\boldsymbol{x} + \boldsymbol{y}$ is not improper and $\mathrm{wid}\left(\mathrm{dual}\,\boldsymbol{x} + \boldsymbol{y}\right) \le \mathrm{wid}\,\boldsymbol{y}$.

- Therefore,

$$\mathrm{wid}\left(\sum_{j=1}^{*J-1} \boldsymbol{a}^{(j)}\right) \le \mathrm{wid}\,\boldsymbol{a}^{(J-1)} \le \mathrm{wid}\,\boldsymbol{a}^{(J)} \quad \left(J = 2,\ldots,M\right)$$

Multi-Event Algorithm – cont'd

- In other words, the c.d.f. always has the following shape



- With the rate of the *null* event calculated as $\boldsymbol{a}_{null} = \left[0, \mathrm{wid}\left(\sum_{j=1}^{*M} \boldsymbol{a}^{(j)}\right)\right]$, the real-valued sum is

$$a_0 = \mathrm{dual}\left(\sum_{j=1}^{*M} \boldsymbol{a}^{(j)}\right) + \left[0, \mathrm{wid}\left(\sum_{j=1}^{*M} \boldsymbol{a}^{(j)}\right)\right]$$

- The intent that "*the maximum and minimum possible probabilities to fire event J are specified by* $\left[\underline{a}^{(J)} / a_0, \overline{a}^{(J)} / a_0\right]$" is maintained!

Best- and Worst-Case Clock Advancement

$$[t_L, t_U] \leftarrow [t_L, t_U] + [T_L, T_U]$$

updated current time increment

- The shortest time for a set of events to fire is when all of them occur at the same time → lower bound.
- The longest time for a set of events to fire is when they occur one by one → upper bound

Best- and Worst-Case Clock Advancement

$$[t_L, t_U] \leftarrow [t_L, t_U] + [T_L, T_U]$$

updated current time increment

- The least time for a set of events to fire is when all of them occur at the same time.
- The lower bound of clock advancement is

$$T_L = -\ln\rho / \sum_{j=1}^{M} \overline{a}_j \quad \text{where} \quad \rho \sim Uniform(0,1)$$

Best- and Worst-Case Clock Advancement

- The longest possible time for the set of events to fire is when the events occur consecutively one by one.

$$T^{(n)} = X^{(1)} + X^{(2)} + \cdots + X^{(n)}$$

with exponential inter-arrival times

$$X^{(1)} \sim Exponential(\underline{\alpha}_0),\ X^{(2)} \sim Exponential(\underline{\alpha}_1), \ldots,\ X^{(n)} \sim Exponential(\underline{\alpha}_{n-1})$$

where
$$\begin{cases} \underline{\alpha}_0 = \sum_{i=1}^{N} \underline{a}_i \\ \underline{\alpha}_1 = \sum_{i=1}^{N} \underline{a}_i - \underline{a}^{(1)} \\ \cdots \\ \underline{\alpha}_{n-1} = \sum_{i=1}^{N} \underline{a}_i - \sum_{j=1}^{n-1} \underline{a}^{(j)} \end{cases}$$

The c.d.f. of $T^{(n)}$ is

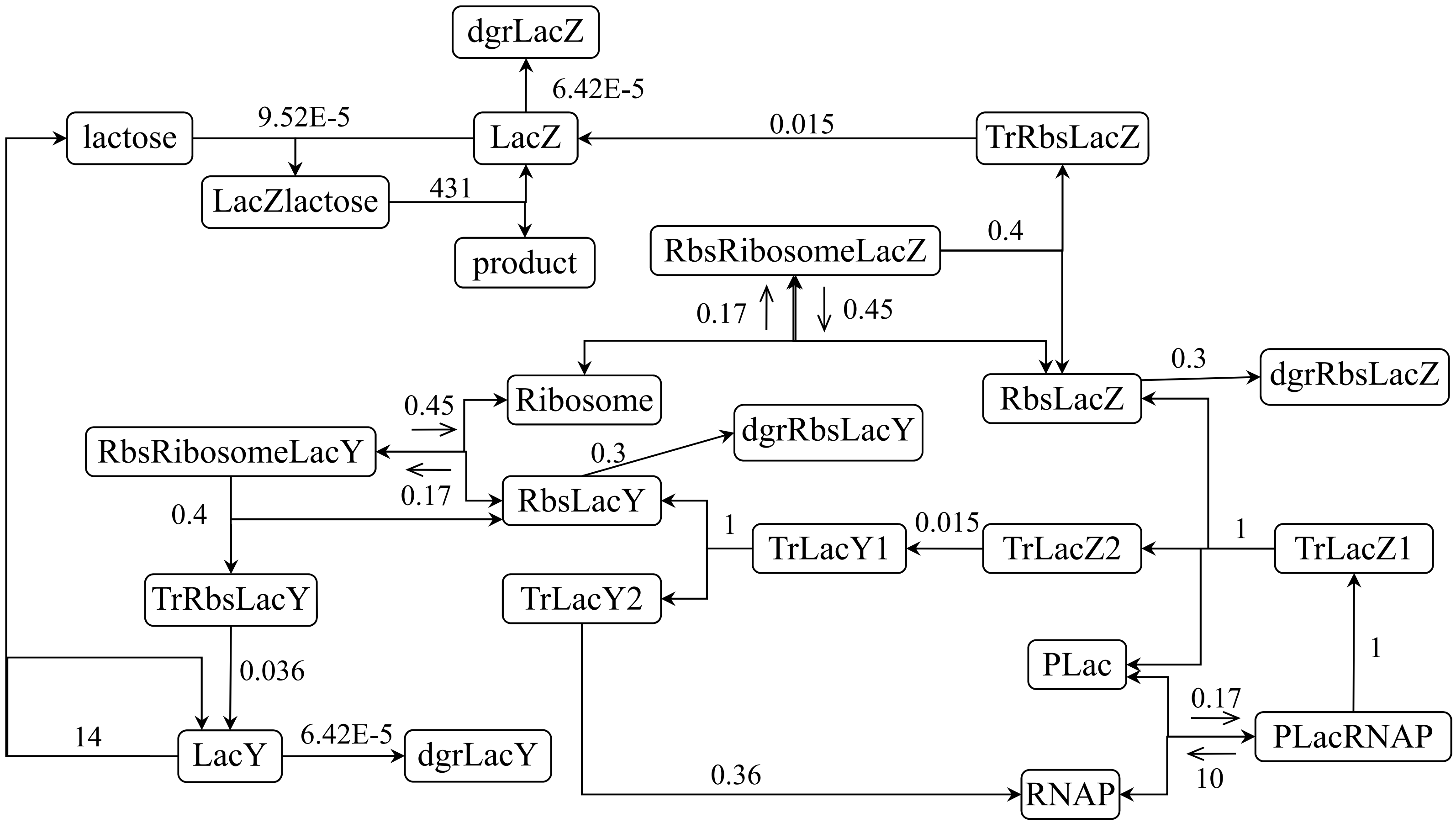
$$P\left(T^{(n)} \le \tau\right) = 1 - P\left(T^{(n)} > \tau\right) = 1 - \sum_{j=0}^{n-1} A_j \exp\left(-\underline{\alpha}_j \tau\right)$$

where $A_j = \prod_{\substack{i=0 \\ i \ne j}}^{n-1} \frac{\underline{\alpha}_i}{\underline{\alpha}_i - \underline{\alpha}_j}$

- Yet a simpler sampling of the upper bound of clock advancement is

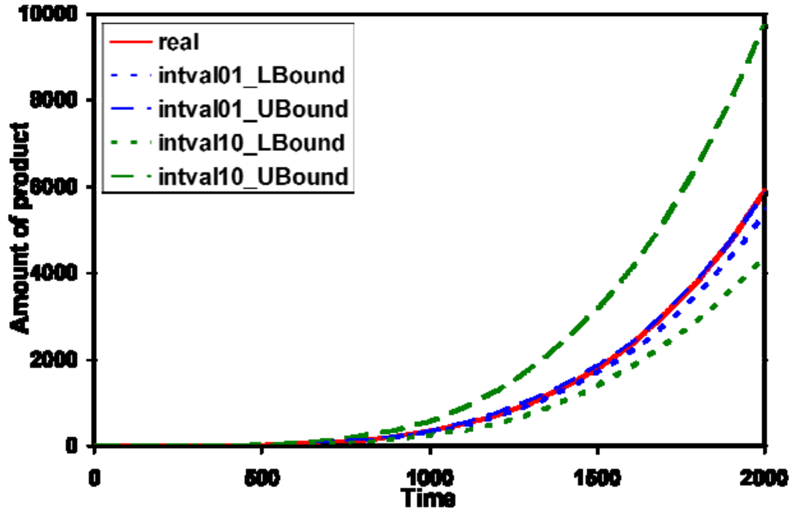
$$T_U = -\ln \rho \left[\sum_{j=0}^{n-1} 1 / \underline{\alpha}_j \right]$$

Reactions of LacZ and LacY proteins in *E. coli*

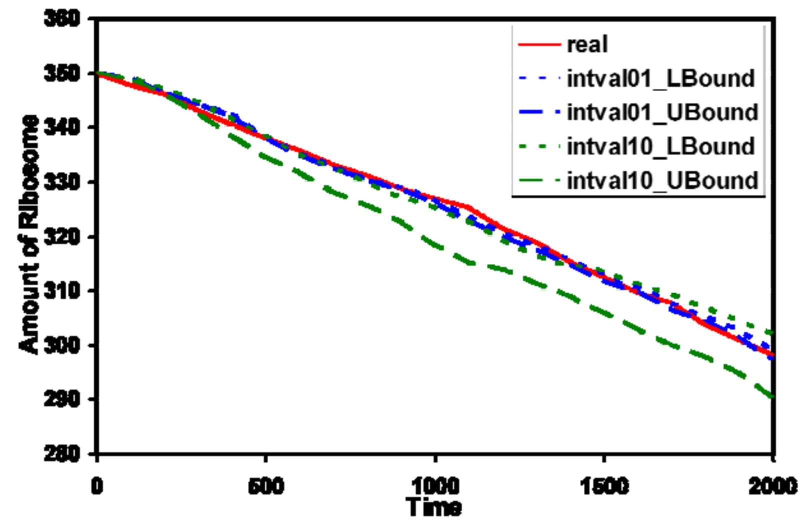


- Rates have the ranges of $\pm 1\%$ and $\pm 10\%$ of the nominal
- Results with averages of 20 runs are compared

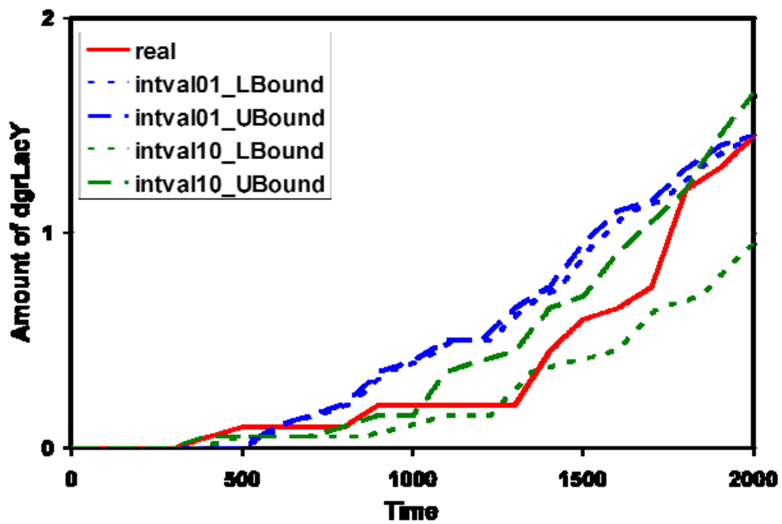
Reactions of LacZ and LacY proteins in *E. coli*



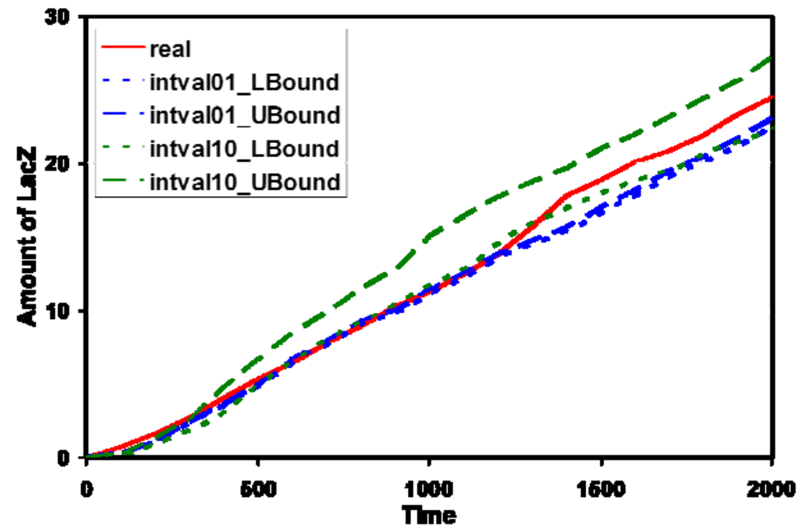
(a) product



(b) Ribosome



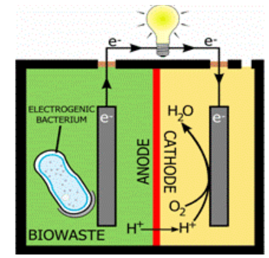
(c) dgrLacY



(d) LacZ

Georgia Tech

Reliable Kinetic Monte Carlo

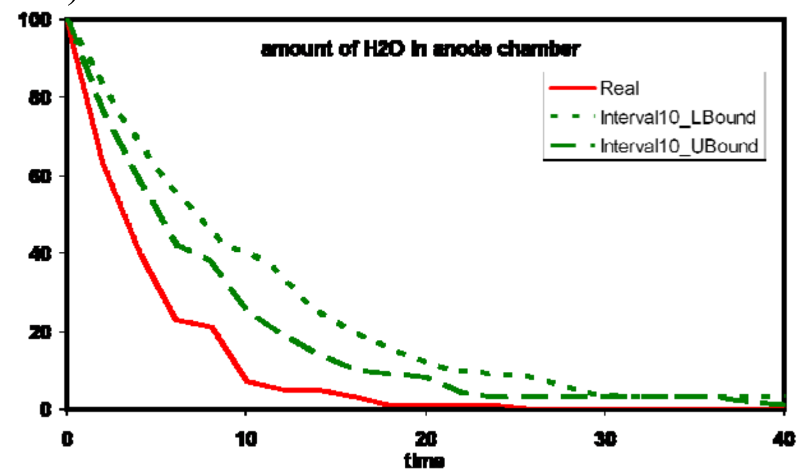


- Sensitivity analysis on-the-fly with *random set sampling*
  - Implemented in SPPARKS (e.g. Microbial Fuel Cell)

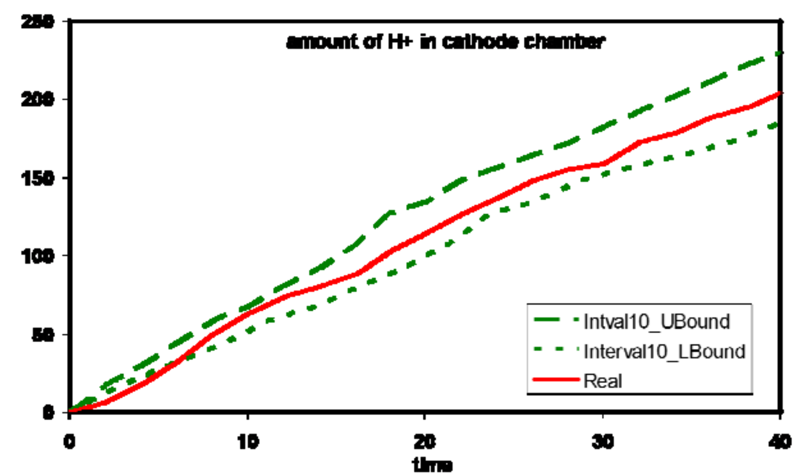
anode chamber

cathode chamber

Event type	Species and reactions	Rate constant
R1: water dissociation	$H_2O \leftrightarrow OH^- + H^+$	10^1 ±10%
R2: carbonic acid dissociation	$CO_2 + H_2O \leftrightarrow HCO_3^- + H^+$	10^1 ±10%
R3: acetic acid dissociation	$AcH \leftrightarrow Ac^- + H^+$	10^1
R4: reduced thionine first dissociation	$MH_3^+ \leftrightarrow MH_2 + H^+$	10^1
R5: reduced thionine second dissociation	$MH_4^{2+} \leftrightarrow MH_3^+ + H^+$	10^1
R6: acetate with oxidized mediator	$Ac^- + MH^+ + NH_4^+ + H_2O \rightarrow X_{Ac} + MH_3^+ + HCO_3^- + H^+$	10^1 ±10%
R7: oxidation double protonated mediator	$MH_4^{2+} \rightarrow MH^+ + 3H^+ + 2e^-$	10^1
R8: oxidation single protonated mediator	$MH_3^+ \rightarrow MH^+ + 2H^+ + 2e^-$	10^1
R9: oxidation neutral mediator	$MH_2 \rightarrow MH^+ + H^+ + 2e^-$	10^1
R10: proton diffusion through PEM	$H^+ \rightarrow H\_^+$	10^{-2}
R11: electron transport from anode to cathode	$e^- \rightarrow e\_^-$	10^{-2}
R12: reduction of oxygen with current generated	$2H\_^+ + 1/2O_2\_ + 2e\_^- \rightarrow H_2O\_$	10^5
R13: reduction of oxygen with current generated	$O_2\_ + 4e\_^- + 2H_2O\_ \rightarrow 4OH\_^-$	10^3



(a) H_2O in anode chamber



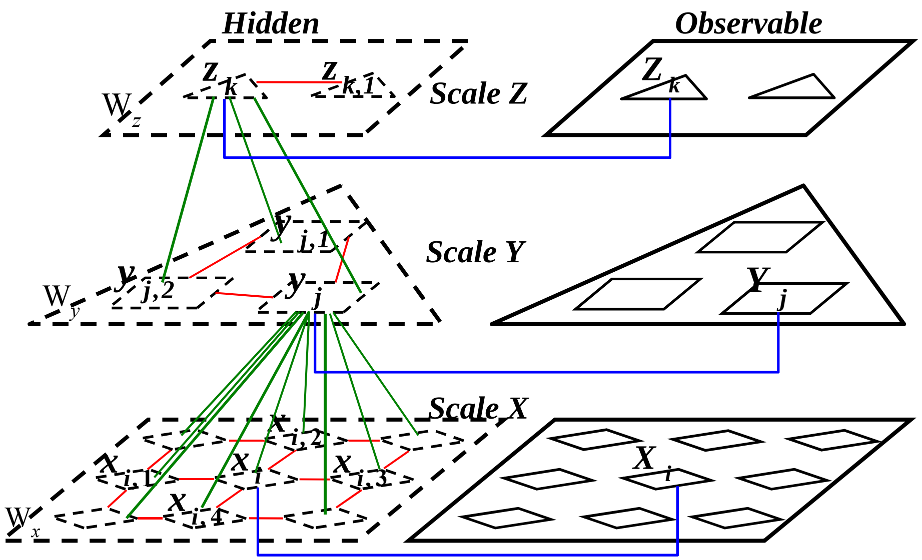
(b) H+ in cathode chamber

More Applications of GIP

- Multiscale uncertainty quantification based on *Generalized Hidden Markov model*
- Generalized interval for global sensitivity analysis
- Molecular dynamics with on-the-fly sensitivity analysis

Multiscale Uncertainty Quantification

- ***Generalized Hidden Markov Model* (GHMM)**



- Multiscale information assimilation
  - Single-Scale Single-Point observation
  - Single-Scale Multi-Point observation
  - Multi-Scale Multi-Point observation
- Carbon nano-tube (CNT) composite actuator design example

major contributors of epistemic uncertainty in multiscale analysis:
- lack of data
- inconsistent observations
- measurement errors

microscale

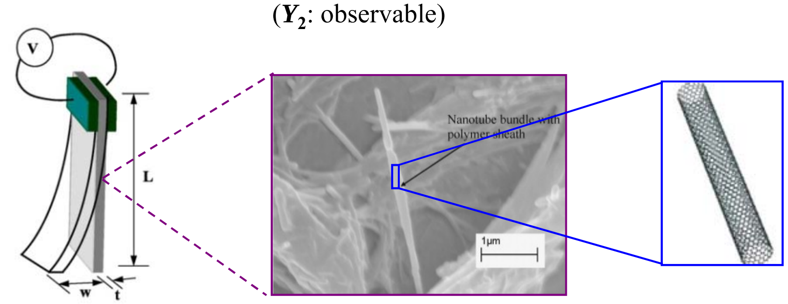
z: bending strain rate of actuator (Z: observable)

mesoscale

y_1: conductivity of composite (1%CNT) (Y_1: observable)
y_2: conductivity of composite (2%CNT) (Y_2: observable)

nanoscale

x: resistivity of single CNT

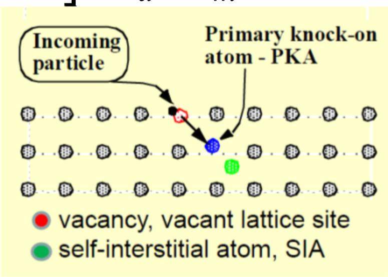


[Wang (2011) *J. Mech. Des.*, **133**(3), 031004]

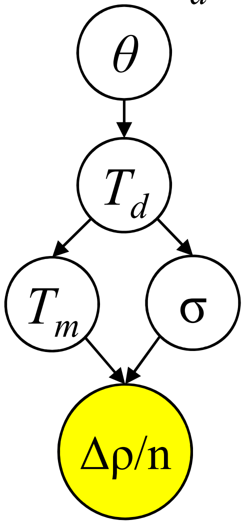
Cross-Scale Model Validation

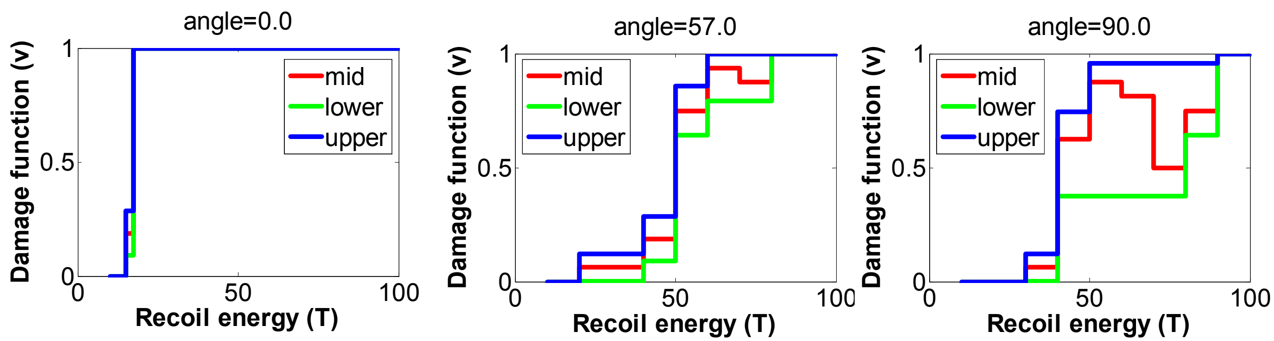
❑Bayesian model validation

$$\mathbf{p}(\theta \mid \Delta\rho / n) = \frac{\mathbf{p}(\theta)\int\int\left[\mathbf{p}(\Delta\rho / n \mid \sigma, T_m)\mathbf{p}(\sigma, T_m \mid T_d)\mathbf{p}(T_d \mid \theta)\right] dT_d dT_m}{\text{dual}\int\int\int\left[\mathbf{p}(\Delta\rho / n \mid \sigma, T_m)\mathbf{p}(\sigma, T_m \mid T_d)\mathbf{p}(T_d \mid \theta)\mathbf{p}(\theta)\right] dT_d dT_m d\theta}$$



- Δρ/n: electrical resistivity change 🡐 *Measureable*
- σ: total displacement cross section
- T_m: maximum possible level of transferred energy
- T_d: damage threshold 🡐 MD





[Tallman, Blumer, Wang, & McDowell, *IDETC/CIE2014*]

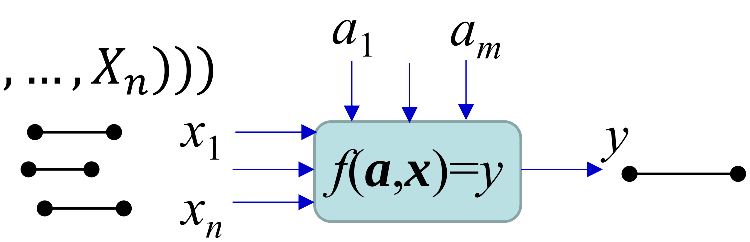
Generalized Interval for Global Sensitivity Analysis

- *High efficiency* (*No probability distribution, No sampling*)
- Hartley-like measure in generalized information theory [Klir 2006]

$$M(\mathbf{Y}) = log_2(1 + width(\mathbf{Y}))$$

- Sensitivity index

$$M(\mathbf{Y}|x_i) = log_2(1 + width(\mathbf{Y}(X_1, \dots, x_i, \dots, X_n)))$$

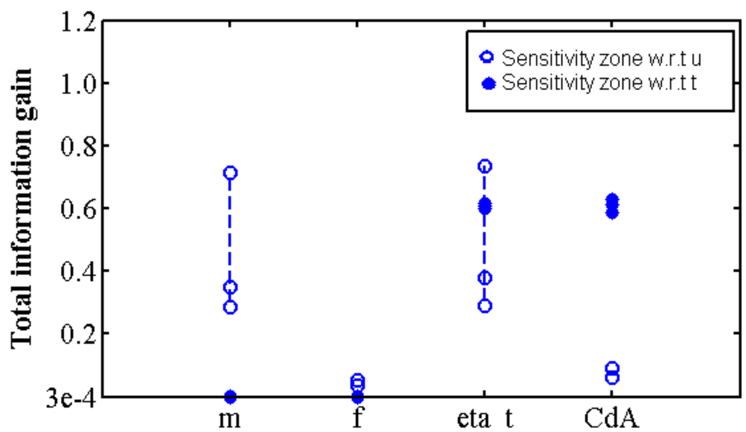


  - First-order information gain

$$I_i = \frac{M(\mathbf{Y}) - M(\mathbf{Y}|x_i)}{M(\mathbf{Y})}$$

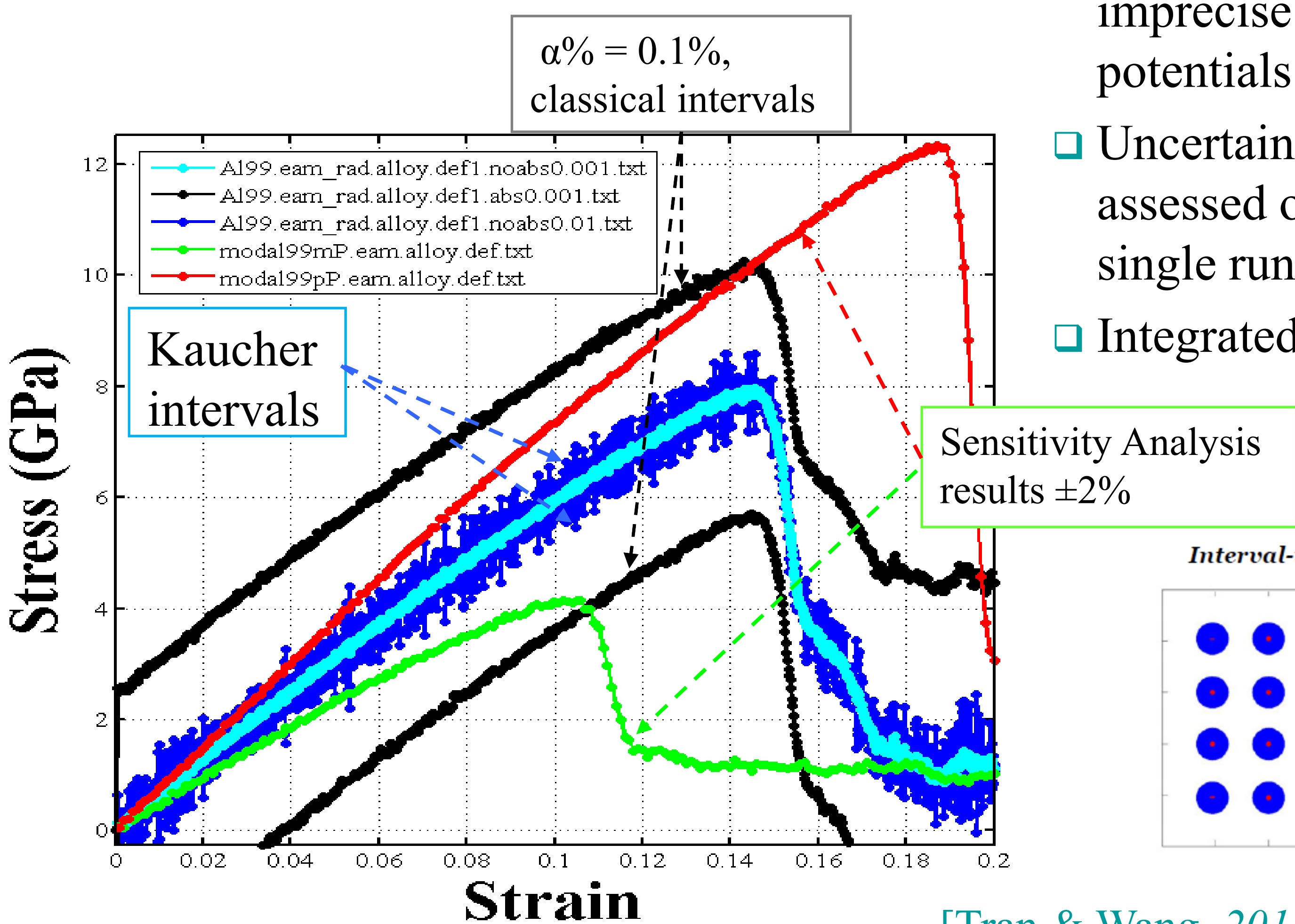
  - Second-order information gain

[Hu, Wang, Cheng, and Zhong (2015) *J. Mech. Des.* **137**(4): 041701]

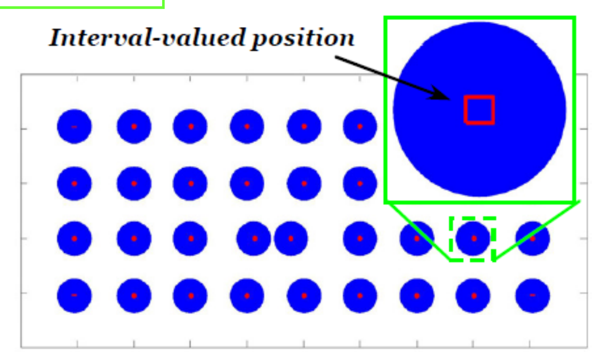


Georgia Tech

Reliable Molecular Dynamics
– *interval interatomic potentials*



- Interval-valued positions and velocities as a result of imprecise interatomic potentials
- Uncertainty effect is assessed on-the-fly with single run of simulation
- Integrated with LAMMPS



[Tran & Wang, *2017, Comp. Mat. Sci.*]

Georgia Tech

Thanks!

Dr. Ola Batarseh
Dr. Jie Hu
Dr. Anh Tran
Dr. Aaron Tallman
Mr. Joel Blumer

- Wang Y. (2010) “Imprecise probabilities based on generalized intervals for system reliability assessment,” *International Journal of Reliability & Safety*, **4**(4): 319-342
- Wang Y., "Independence in Generalized Interval Probability," *Proceedings of 1st International Conference on Vulnerability and Risk Analysis and Management (ICVRAM 2011) and 5th International Symposium on Uncertainty Modeling and Analysis (ISUMA 2011), April 11-13, 2011, Hyattsville, Maryland*, pp.37-44
- Wang Y. (2011) “Multiscale Uncertainty Quantification Based on a Generalized Hidden Markov Model,” *ASME Journal of Mechanical Design*, **133**(3): 031004(1-10)
- Wang Y. (2013) Reliable kinetic Monte Carlo simulation based on random set sampling. *Soft Computing*, **17**(8): 1439-1451.
- Wang Y. (2013) Generalized Fokker-Planck equation with generalized interval probability. *Mechanical Systems & Signal Processing*, **37**(1-2):92-104.
- Wang Y. (2015) Stochastic dynamics simulation with generalized interval probability. *International Journal of Computer Mathematics*, **92**(3): 623-642
- Hu J., Wang Y., Cheng A., and Zhong Z. (2015) Sensitivity analysis in quantified interval constraint satisfaction problems. *Journal of Mechanical Design*, **137**(4): 041701.
- Hu J., Wang Y., Cheng A., and Zhong Z. (2015) An extended Kalman filtering mechanism based on generalized interval probability. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B*, **1**(3), 031002.
- Tran A.V. and Wang Y. (2017) Reliable molecular dynamics: Uncertainty quantification using interval analysis in molecular dynamics simulation. *Computational Materials Science*, **127**: 141-160.
- Batarseh O.G. and Wang Y. (2018) An interval-based approach to model input uncertainty in M/M/1 simulation. *International Journal of Approximate Reasoning*, **95**: 46-61.