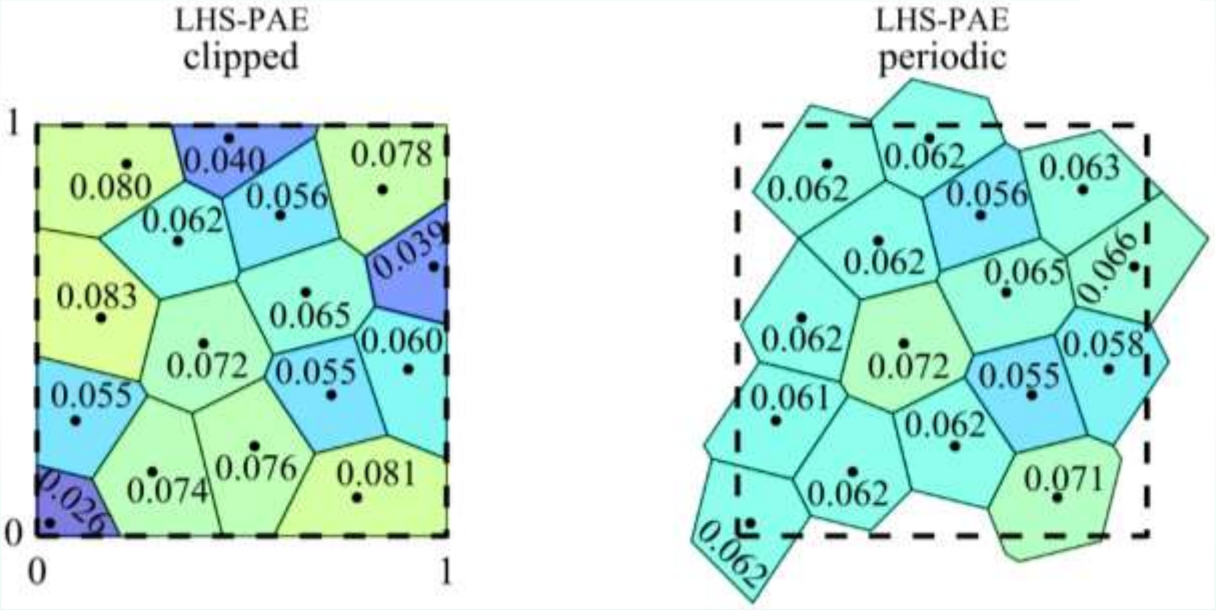
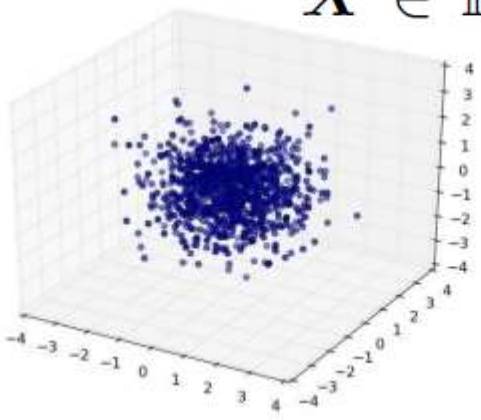
Application of Voronoi Weights in Monte Carlo Integration with a Given Sampling Plan

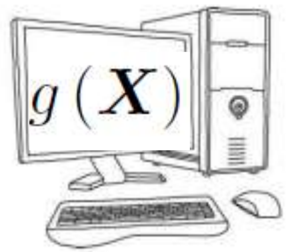


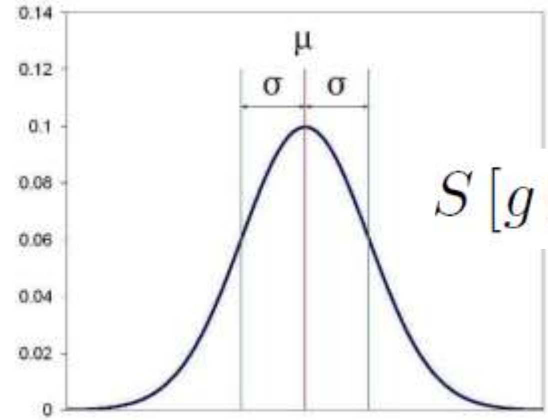
Miroslav Vořechovský, Václav Sadílek, Jan Eliáš
Brno University of Technology, Czech Republic

Computer (or physical) experiments

$\boldsymbol{X} \in \mathbb{R}^{N_{\text{var}}}$







$S[g(\boldsymbol{x})]$

- **Random inputs**
  defined by their (joint) probability distribution
- optimal selection of representative points from design space

- **Outputs**
  characteristics of the resulting probability distribution (mean, st. dev., …), sensitivity or the probability of failure of structure/system

Computer (or physical) experiments

- Target statistical parameter

$$\mathrm{E}[S\left[g\left(\boldsymbol{X}\right)\right]] = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} S\left[g\left(\boldsymbol{x}\right)\right] \mathrm{d}F_{\boldsymbol{X}}\left(\boldsymbol{x}\right)$$

$$\mathrm{d}F_{\boldsymbol{X}}\left(\boldsymbol{x}\right) = f_{\boldsymbol{X}}\left(\boldsymbol{x}\right) \cdot \mathrm{d}x_1 \, \mathrm{d}x_2 \, \cdots \, \mathrm{d}x_{N_{\mathrm{var}}}$$

- Numerical evaluation of the integral involves $i = 1, \ldots, N_{\mathrm{sim}}$ points associated with weights

$$\mathrm{E}[S\left[g\left(\boldsymbol{X}\right)\right]] \approx \sum_{i=1}^{N_{\mathrm{sim}}} S\left[g(\boldsymbol{x}_i)\right] \cdot w_i$$

Computer (or physical) experiments

- Target statistical parameter

$$\mathrm{E}[S\,[g\,(\boldsymbol{X})]] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} S\,[g\,(\boldsymbol{x})]\,\mathrm{d}F_{\boldsymbol{X}}\,(\boldsymbol{x})$$

$$\mathrm{d}F_{\boldsymbol{X}}\,(\boldsymbol{x}) = f_{\boldsymbol{X}}\,(\boldsymbol{x}) \cdot \mathrm{d}x_1\,\mathrm{d}x_2 \cdots \mathrm{d}x_{N_{\mathrm{var}}}$$

- Numerical evaluation of the integral involves $i = 1, \ldots, N_{\mathrm{sim}}$ points associated with weights

$$\mathrm{E}[S\,[g\,(\boldsymbol{X})]] \approx \sum_{i=1}^{N_{\mathrm{sim}}} S\,[g(\boldsymbol{x}_i)] \cdot w_i$$

- Usually all points share equal probability w_i

$$\mathrm{E}[S\,[g\,(\boldsymbol{X})]] \approx \frac{1}{N_{\mathrm{sim}}} \sum_{i=1}^{N_{\mathrm{sim}}} S\,[g(\boldsymbol{x}_i)]$$

Computer (or physical) experiments

- Target statistical parameter

$$\mathrm{E}[S[g(\boldsymbol{X})]] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} S[g(\boldsymbol{x})] \, \mathrm{d}F_{\mathbf{X}}(\boldsymbol{x})$$

$$\mathrm{d}F_{\mathbf{X}}(\boldsymbol{x}) = f_{\mathbf{X}}(\boldsymbol{x}) \cdot \mathrm{d}x_1 \, \mathrm{d}x_2 \cdots \mathrm{d}x_{N_{\mathrm{var}}}$$

- Rewrite the integral in terms of *U* – integrate over unit hypercube with **<u>uniform</u>** density:

$$\mathrm{E}[S[g(\boldsymbol{X})]] = \int_0^1 \dots \int_0^1 S[g(\boldsymbol{x})] \; \mathrm{d}C(u_1, \dots, u_{N_{\mathrm{var}}})$$

$$= \int_{[0,1]^{N_{\mathrm{var}}}} S[g(\boldsymbol{x})] \prod_{v=1}^{N_{\mathrm{var}}} \mathrm{d}U_v$$

Computer (or physical) experiments

- Target statistical parameter

$$\mathrm{E}[S[g(\boldsymbol{X})]] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} S[g(\boldsymbol{x})]\,\mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x})$$

$$\mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) = f_{\boldsymbol{X}}(\boldsymbol{x}) \cdot \mathrm{d}x_1\,\mathrm{d}x_2 \cdots \mathrm{d}x_{N_{\mathrm{var}}}$$

- Consider independent uniform variables U_i with the following copula:

$$C(u_1, \ldots, u_{N_{\mathrm{var}}}) = \mathrm{P}(U_1 \le u_1, \ldots, U_{N_{\mathrm{var}}} \le u_{N_{\mathrm{var}}}) = \prod_{v=1}^{N_{\mathrm{var}}} u_v$$

- They represent probabilities $F_{X_v} = U_v$

$$\boldsymbol{x} = \{x_1, \ldots, x_{N_{\mathrm{var}}}\} = \{F_1^{-1}(u_1), \ldots, F_{N_{\mathrm{var}}}^{-1}(u_{N_{\mathrm{var}}})\}$$

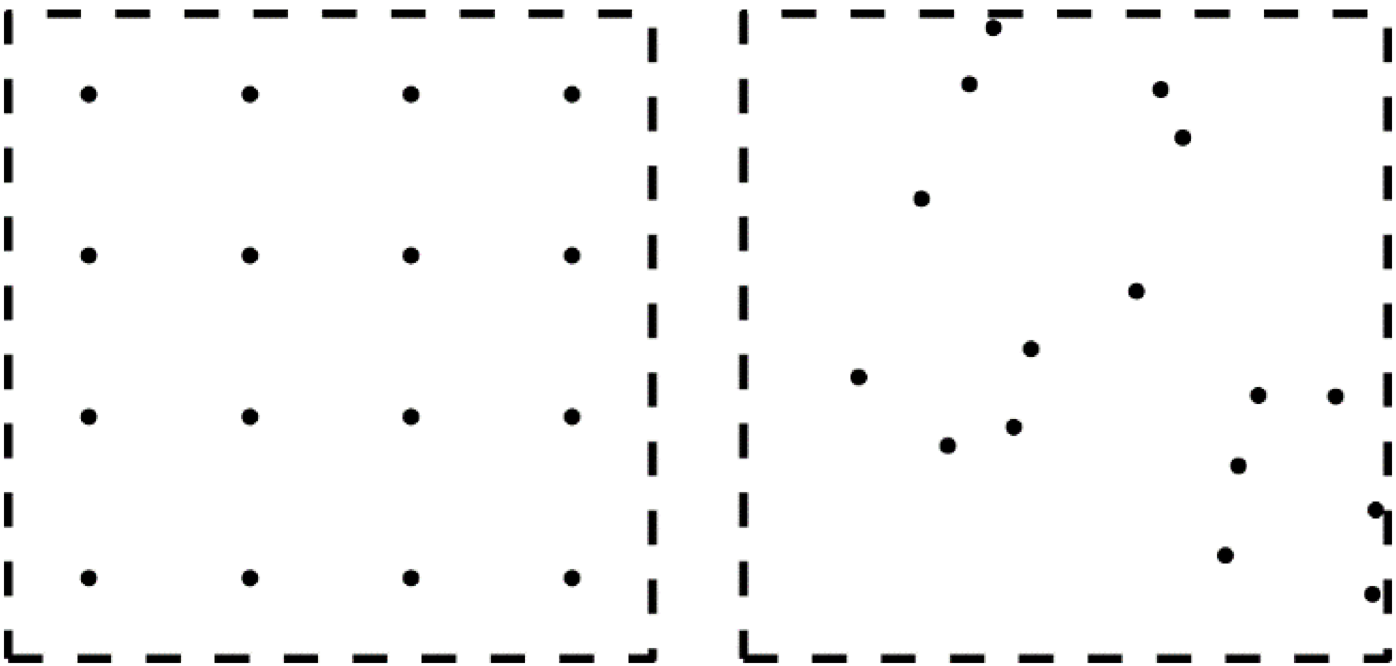
- The joint CDF reads:

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_v F_{X_v} = \prod_v U_v, \text{ and } \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_v \mathrm{d}U_v$$

The weights for fixed points

$$\mathrm{E}[S\left[g\left(\boldsymbol{X}\right)\right]] \approx \frac{1}{N_{\mathrm{sim}}} \sum_{i=1}^{N_{\mathrm{sim}}} S\left[g(\boldsymbol{x}_i)\right] \qquad \mathrm{E}[S\left[g\left(\boldsymbol{X}\right)\right]] \approx \sum_{i=1}^{N_{\mathrm{sim}}} S\left[g(\boldsymbol{x}_i)\right] \cdot w_i$$

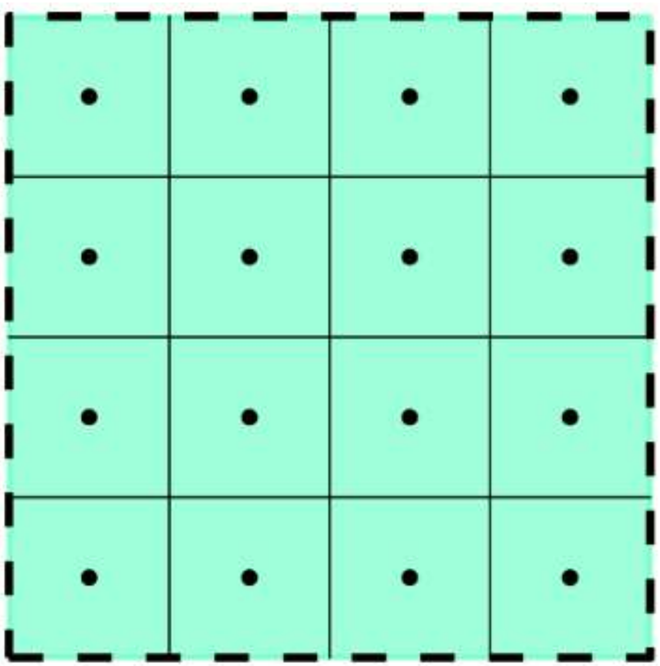
- Equal weights $1/N_{\mathrm{sim}}$?

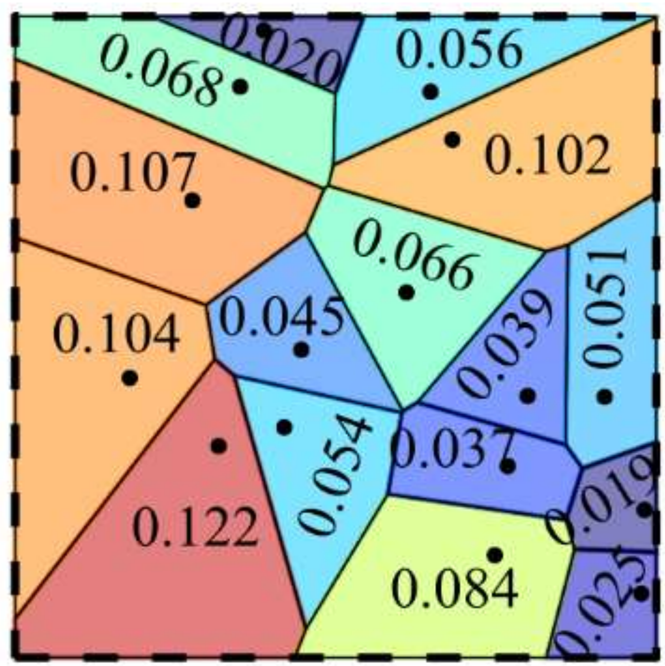


The weights for fixed points

$$\mathrm{E}[S\left[g\left(\boldsymbol{X}\right)\right]] \approx \frac{1}{N_{\mathrm{sim}}} \sum_{i=1}^{N_{\mathrm{sim}}} S\left[g(\boldsymbol{x}_i)\right] \qquad \mathrm{E}[S\left[g\left(\boldsymbol{X}\right)\right]] \approx \sum_{i=1}^{N_{\mathrm{sim}}} S\left[g(\boldsymbol{x}_i)\right] \cdot w_i$$

- Equal weights $1/N_{\mathrm{sim}}$?

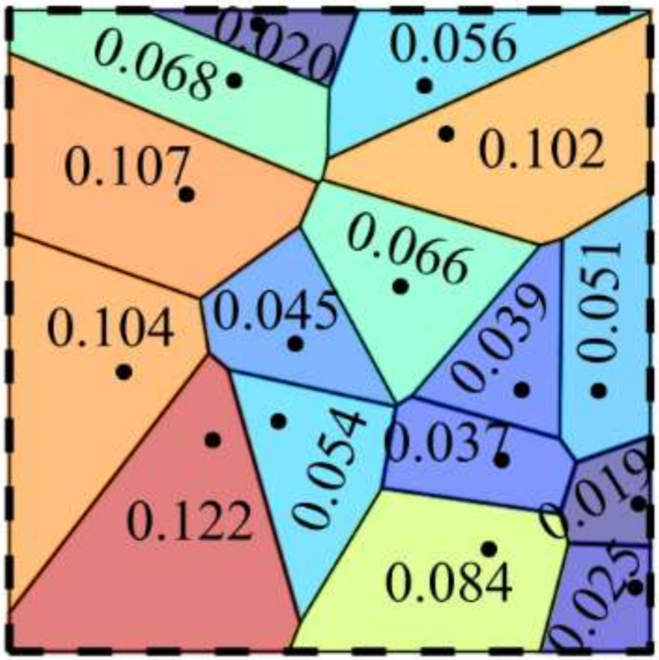




The weights for fixed points

$$\mathrm{E}[S[g(\boldsymbol{X})]] \approx \frac{1}{N_{\mathrm{sim}}} \sum_{i=1}^{N_{\mathrm{sim}}} S[g(\boldsymbol{x}_i)] \qquad \mathrm{E}[S[g(\boldsymbol{X})]] \approx \sum_{i=1}^{N_{\mathrm{sim}}} S[g(\boldsymbol{x}_i)] \cdot w_i$$

- The idea: spatial distribution of points can be used for selection of associated probability
- Weights are obtained using (*Voronoi diagrams* – a tessellation into cells) in the design domain (unit hypercube)
- Weights are the surfaces/volumes around points
- **Post-processing** of existing design & results (extract more information from an existing bad design)



Periodic extension of the design space

- Vořechovský, M. and J. Eliáš. *Improved formulation of Audze-Eglajs criterion for space-filling designs.*
  In: Proc. of 12th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP12, Vancouver, Canada, 2015.

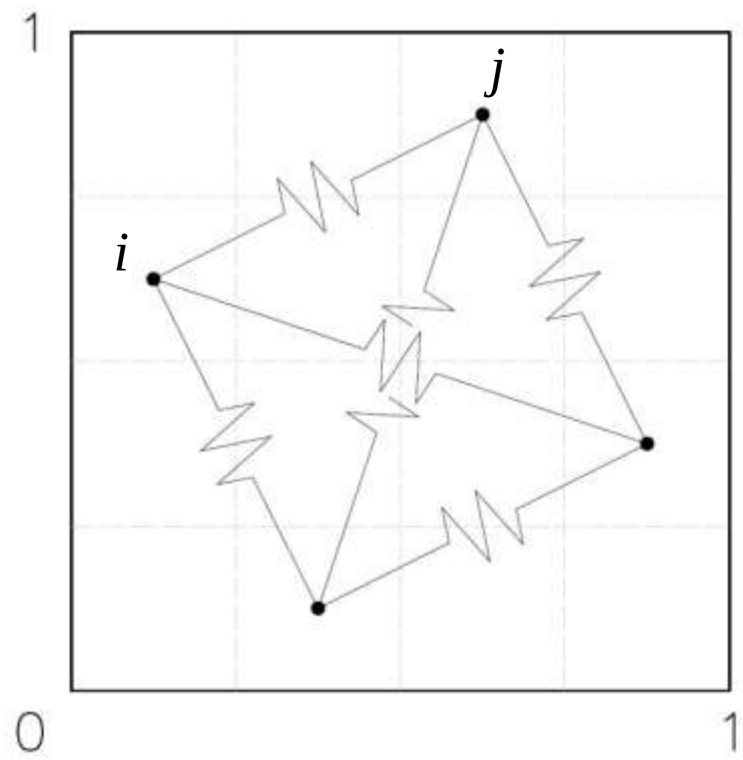
- Eliáš, J. and M. Vořechovský. Modication of the Audze-Eglajs criterion to achieve a uniform distribution of sampling points.
  Advances in Engineering Software, under review.

Audze-Eglājs (AE) criterion

- Potential energy of the system

$$E^{\mathrm{AE}} = \sum_{i=1}^{N_{\mathrm{sim}}} \sum_{j=i+1}^{N_{\mathrm{sim}}} \frac{1}{L_{ij}^2}$$

- Mutually closer points → higher contribution to the overall energy of the system
- Optimization (min. energy) → points tend to shift away from each other



AUDZE, P. P. – EGLĀJS, V. O. New approach for planning out of experiments. Problems of Dynamics and Strengths. Zinatne Publishing House, 1977, Vol 35, 104–107.

Audze-Eglājs (AE) criterion

- Potential energy of the system

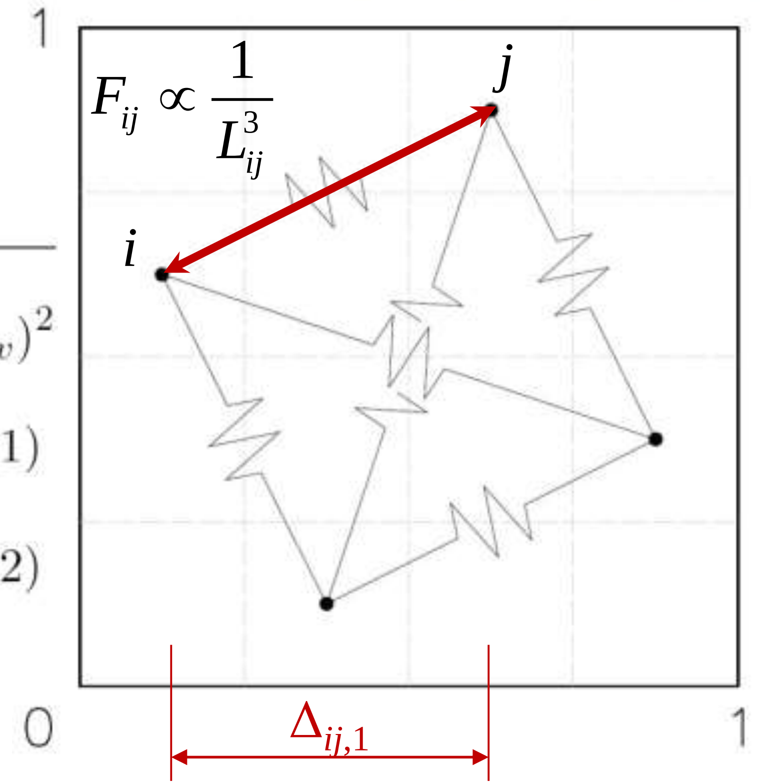
$$E^{\mathrm{AE}} = \sum_{i=1}^{N_{\mathrm{sim}}} \sum_{j=i+1}^{N_{\mathrm{sim}}} \frac{1}{L_{ij}^2}$$

$$L_{ij} = L(\boldsymbol{u}_i, \boldsymbol{u}_j) = \sqrt{\sum_{v=1}^{N_{\mathrm{var}}} (u_{i,v} - u_{j,v})^2} = \sqrt{\sum_{v=1}^{N_{\mathrm{var}}} (\Delta_{ij,v})^2} \quad (1)$$

where

$$\Delta_{ij,v} = |u_{i,v} - u_{j,v}| \quad (2)$$

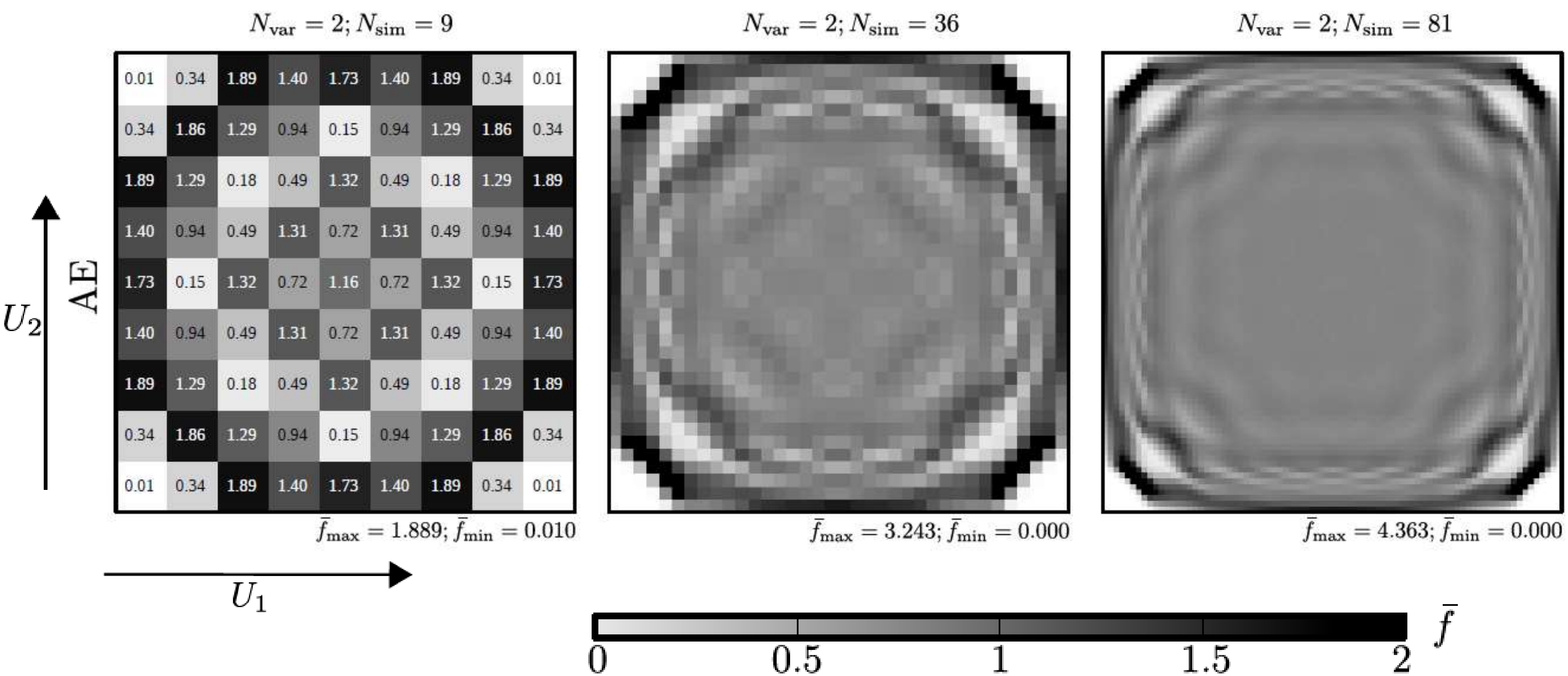
$$F_{ij} = -\frac{\mathrm{d}\,E_{ij}^{\mathrm{AE}}}{\mathrm{d}\,L_{ij}} = -\frac{\mathrm{d}\,\frac{1}{L_{ij}^2}}{\mathrm{d}\,L_{ij}} = \frac{2}{L_{ij}^3}$$



- Does the criterion really prioritize <u>uniform distribution</u> of sampling points in space?

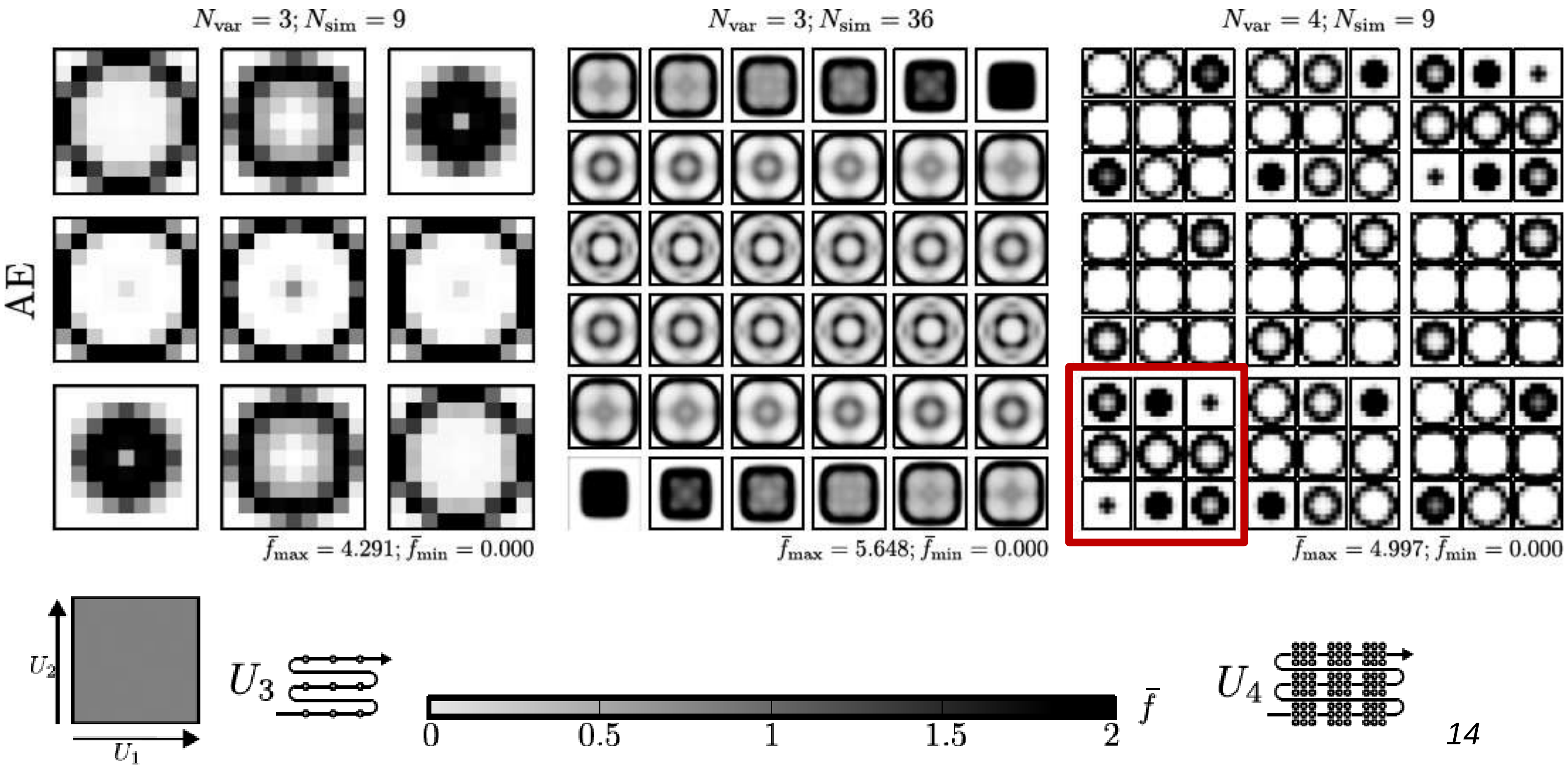
Measurement of uniformity

2D domain covered by AE-LHS-optimized sample



Measurement of uniformity

3D and 4D domains covered by AE-LHS



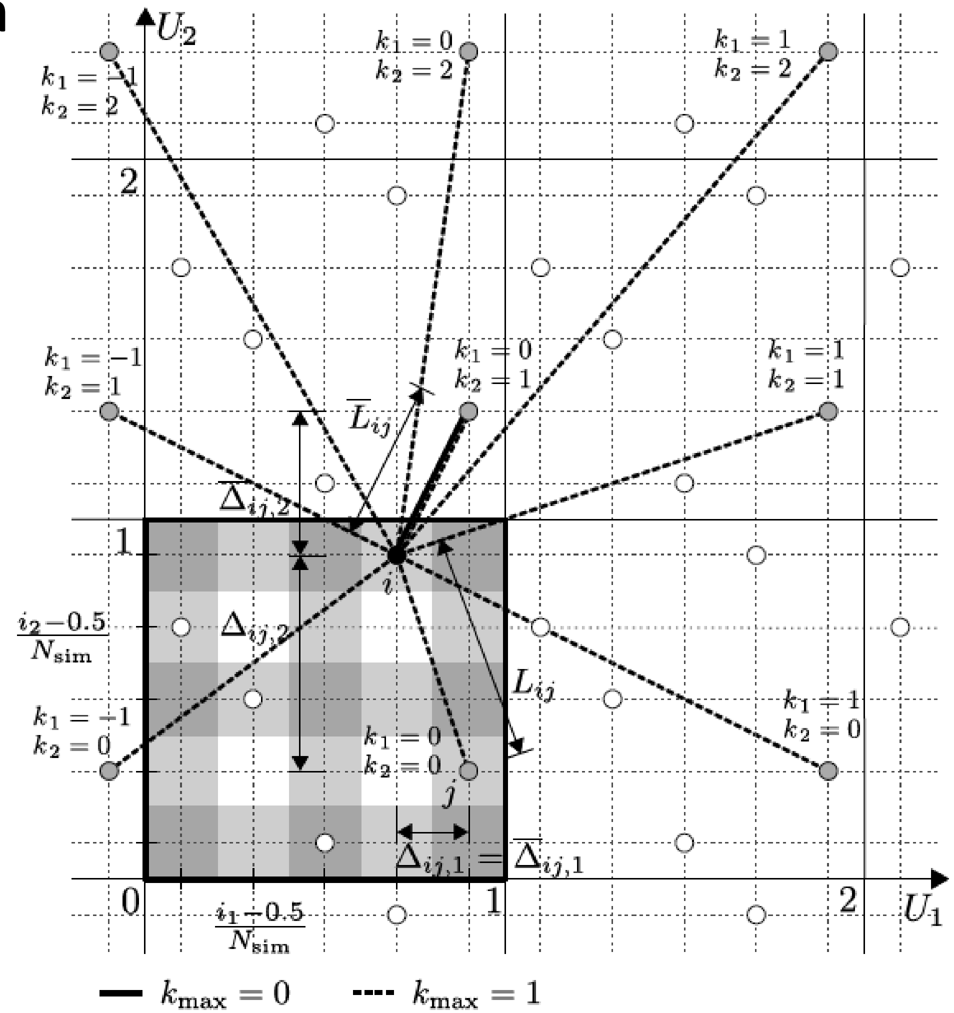
Periodic Audze-Eglājs (PAE) criterion

- Potential energy of the system

$$E^{\text{PAE}} = \sum_{i=1}^{N_{\text{sim}}} \sum_{j=i+1}^{N_{\text{sim}}} \frac{1}{\overline{L}_{ij}^{2}}$$

- The shortest distance

$$\overline{L}_{ij} = \sqrt{\sum_{v=1}^{N_{\text{var}}} [\min(\Delta_{ij,v}, 1 - \Delta_{ij,v})]^2}$$

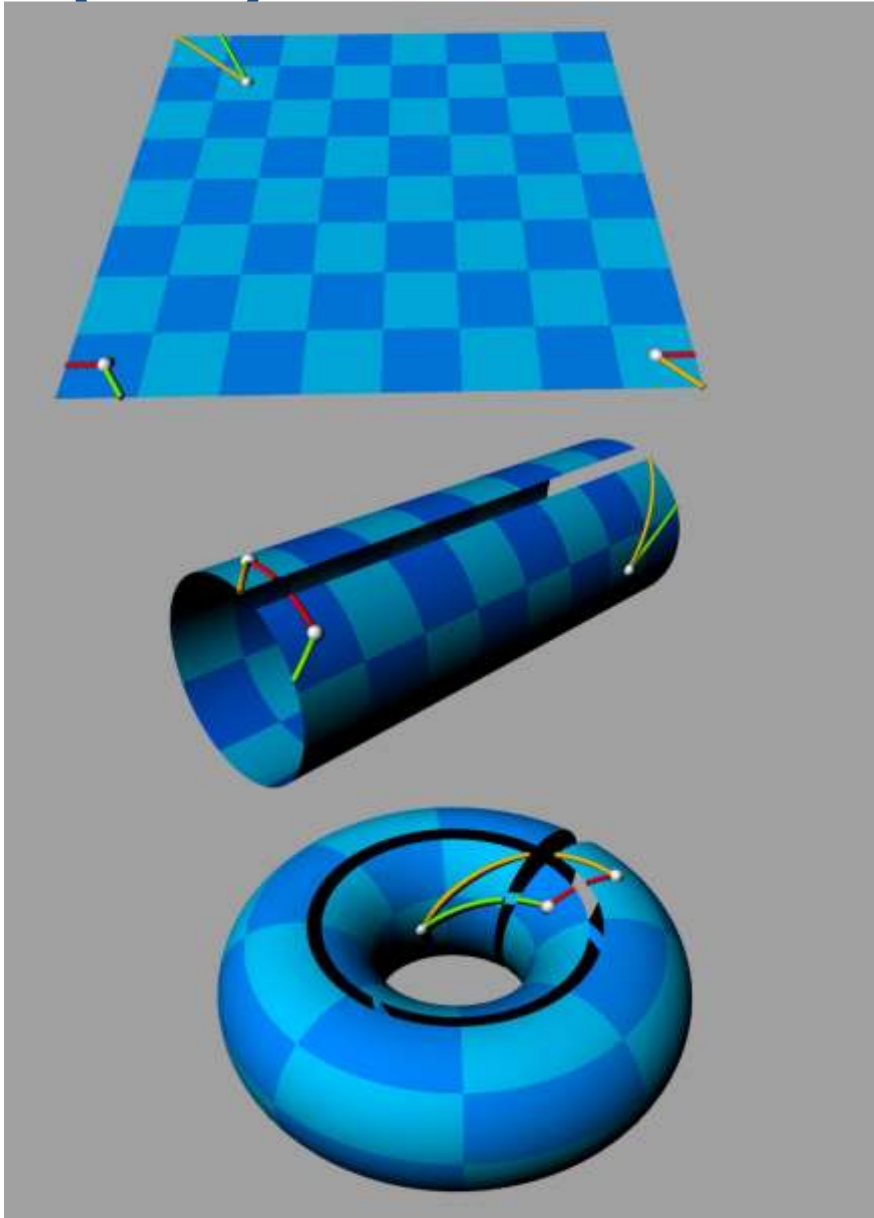


Periodic Audze-Eglājs (PAE) criterion

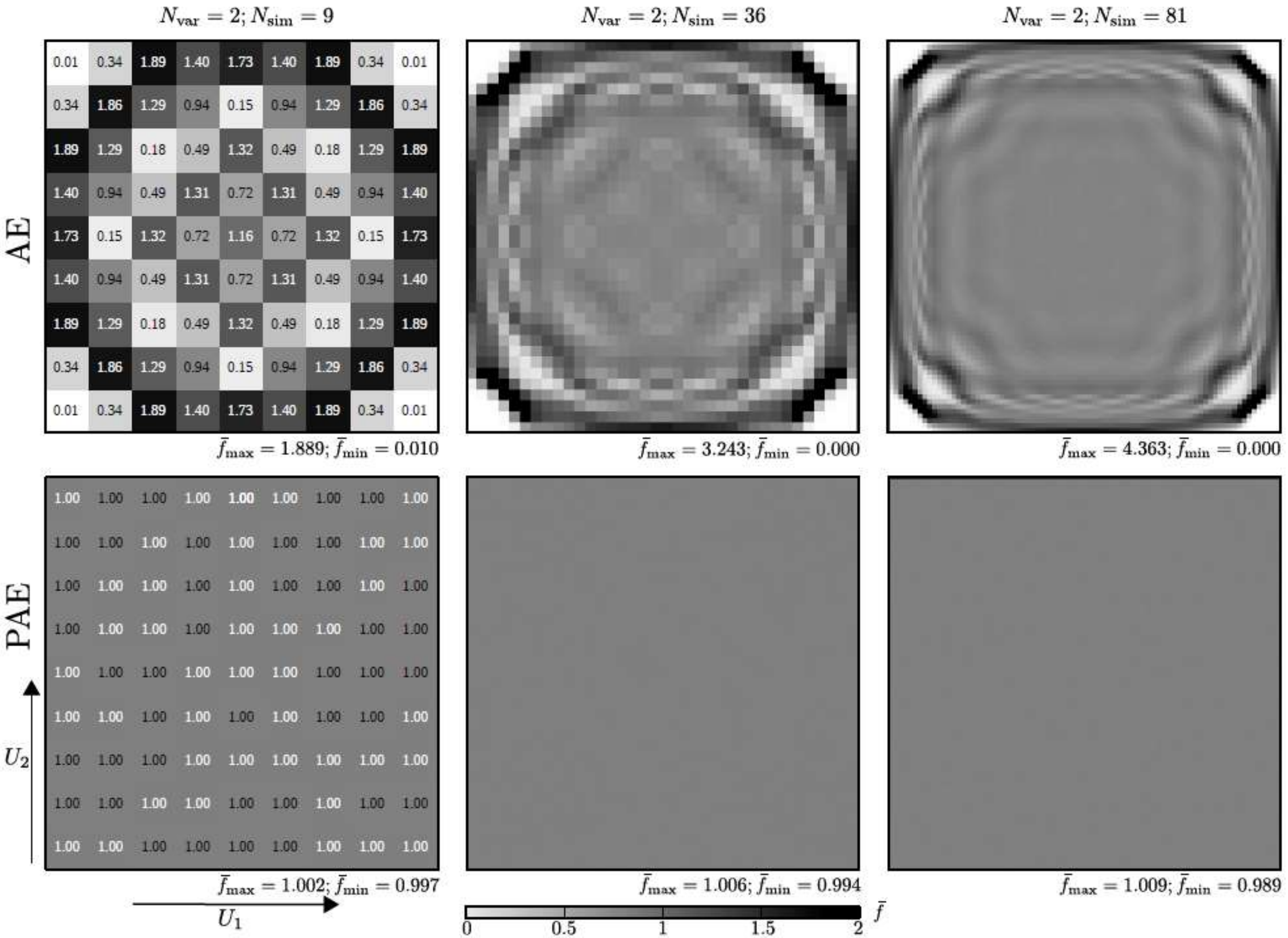
- Potential energy of the system

$$E^{\mathrm{PAE}} = \sum_{i=1}^{N_{\mathrm{sim}}} \sum_{j=i+1}^{N_{\mathrm{sim}}} \frac{1}{\overline{L}_{ij}^{2}}$$

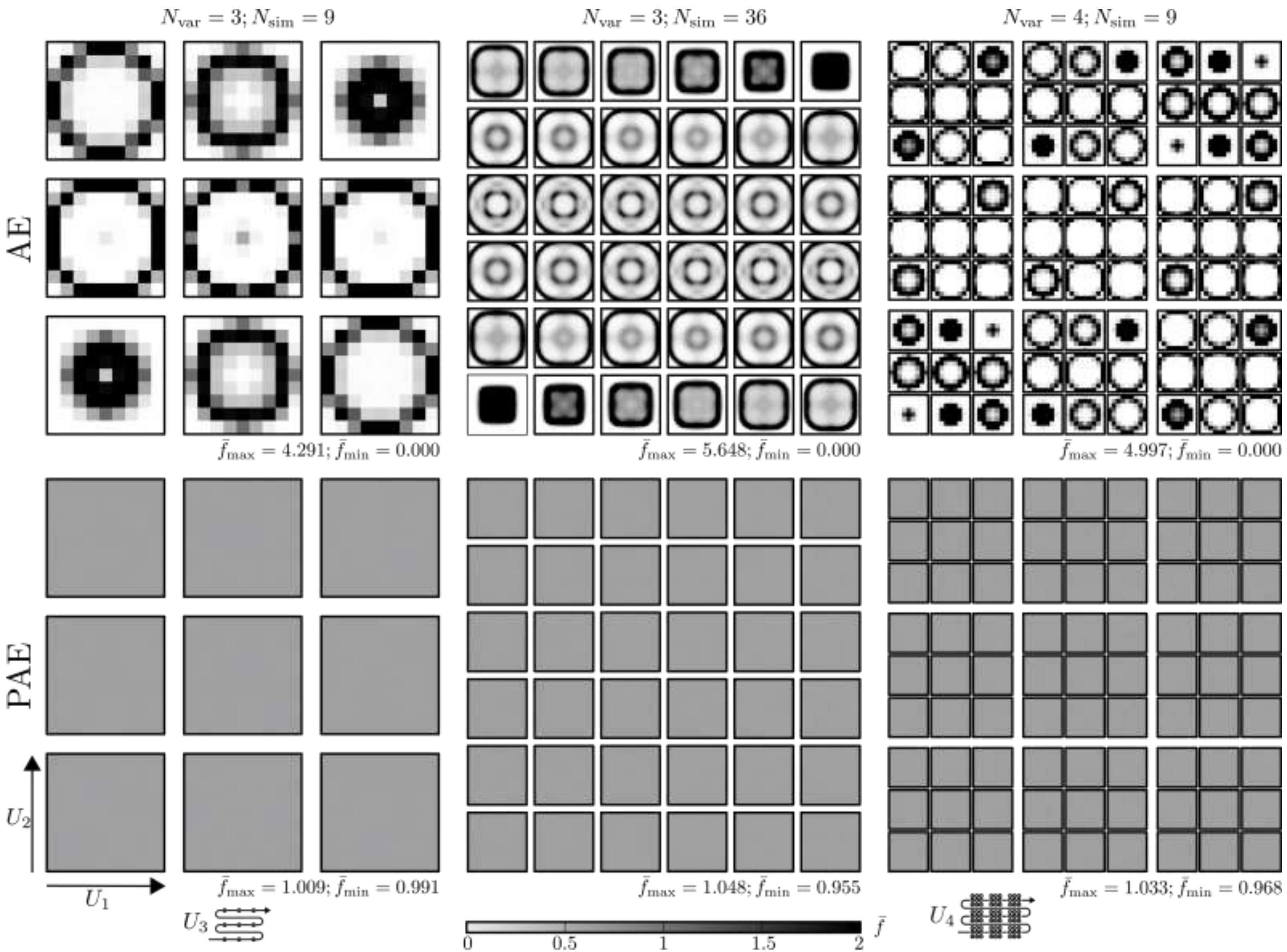
- The shortest distance for each pair is considered
- Does the new criterion prioritize uniform distribution of sampling points in space
- The source of uniformity lies in the invariance of PAE wrt to translation



Measurement of uniformity



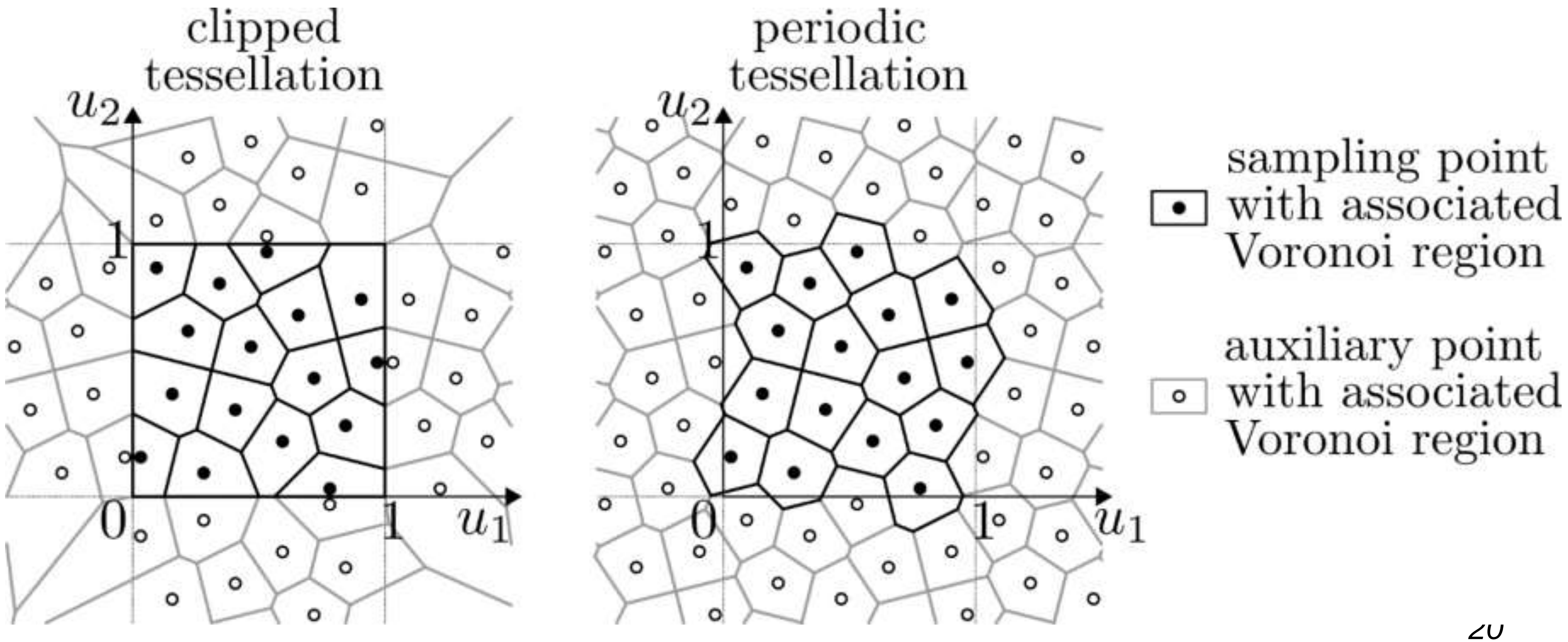
Measurement of uniformity



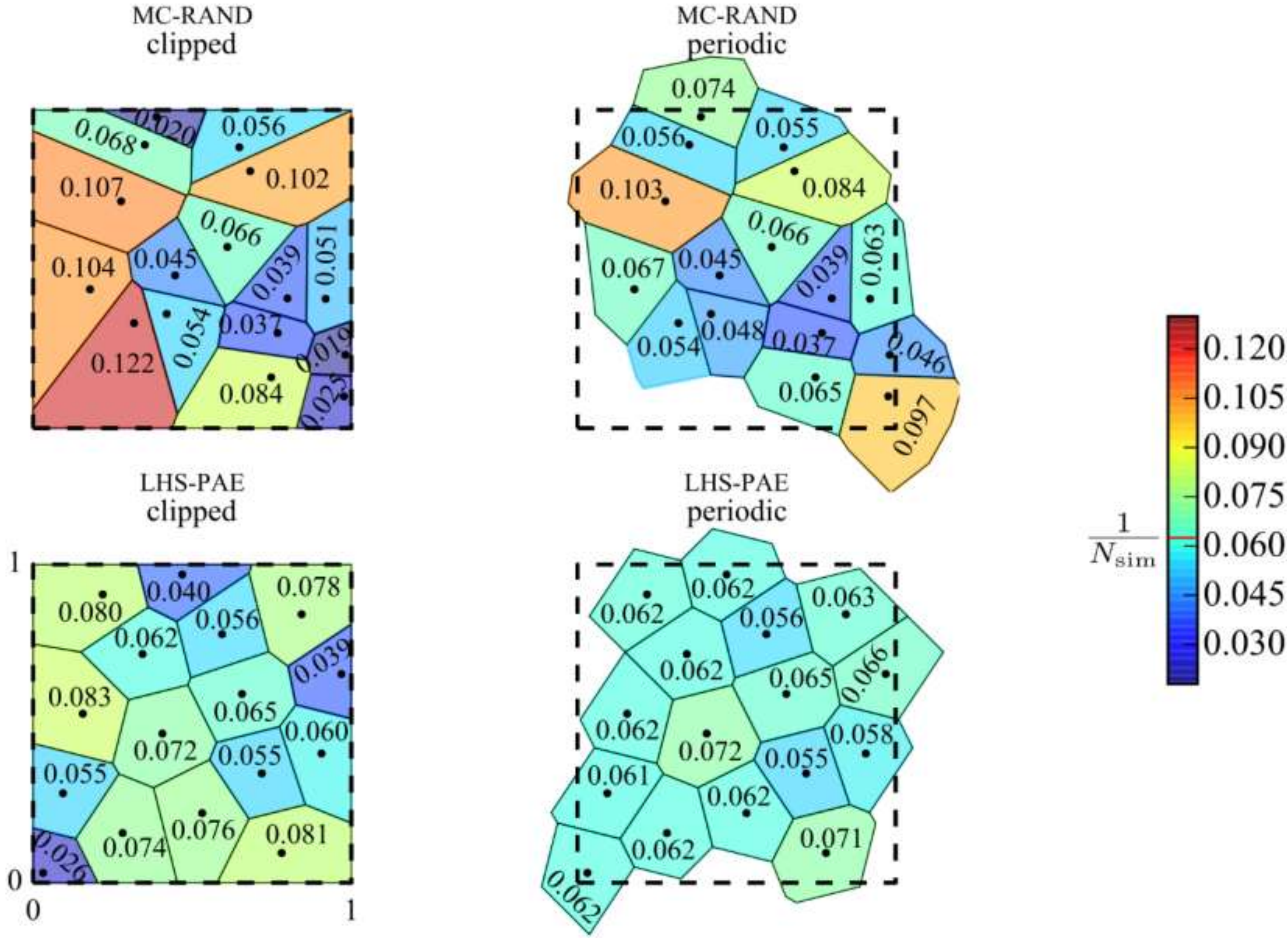
The Voronoi weights

Clipped vs. Periodic tessellation

- Reuse the thought



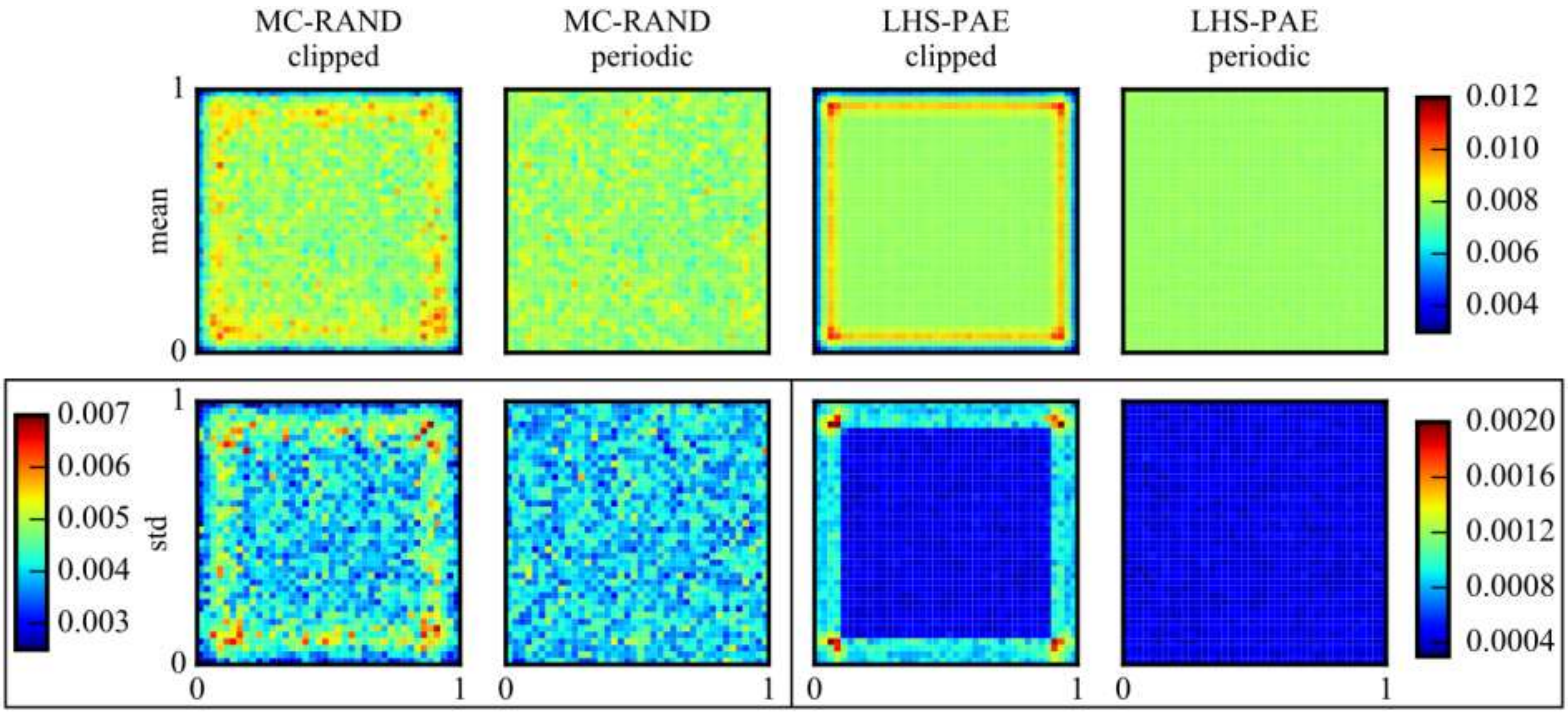
Clipped vs. Periodic tessellation



Clipped vs. Periodic tessellation

- Spatial distribution (frequency) of weights
- Systematic error close to boundaries

The tessellation results in systematic appearance of underestimated regions near the boundaries followed by regions with over-weighted regions.



Numerical examples

Numerical examples

- Three functions (frequently used transformations)

$$g_{\text{sum}}(\boldsymbol{X}) = \sum_{v=1}^{N_{\text{var}}} X_v^2$$

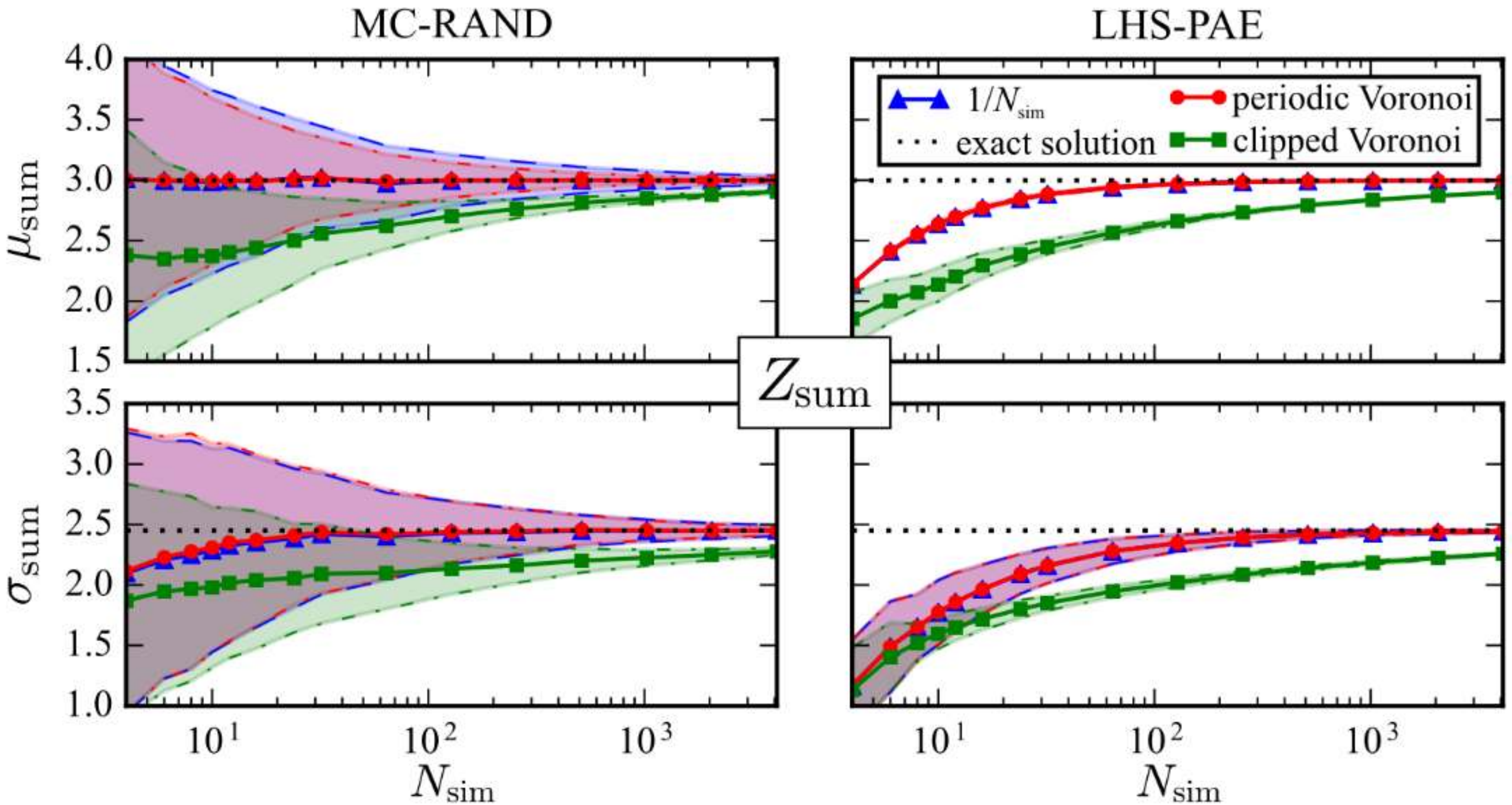
$$g_{\text{exp}}(\boldsymbol{X}) = \sum_{v=1}^{N_{\text{var}}} \exp\left(-X_v^2\right)$$

$$g_{\text{prod}}(\boldsymbol{X}) = \prod_{v=1}^{N_{\text{var}}} X_v$$

$$N_{\text{var}} = \{2,3,5,9\}$$

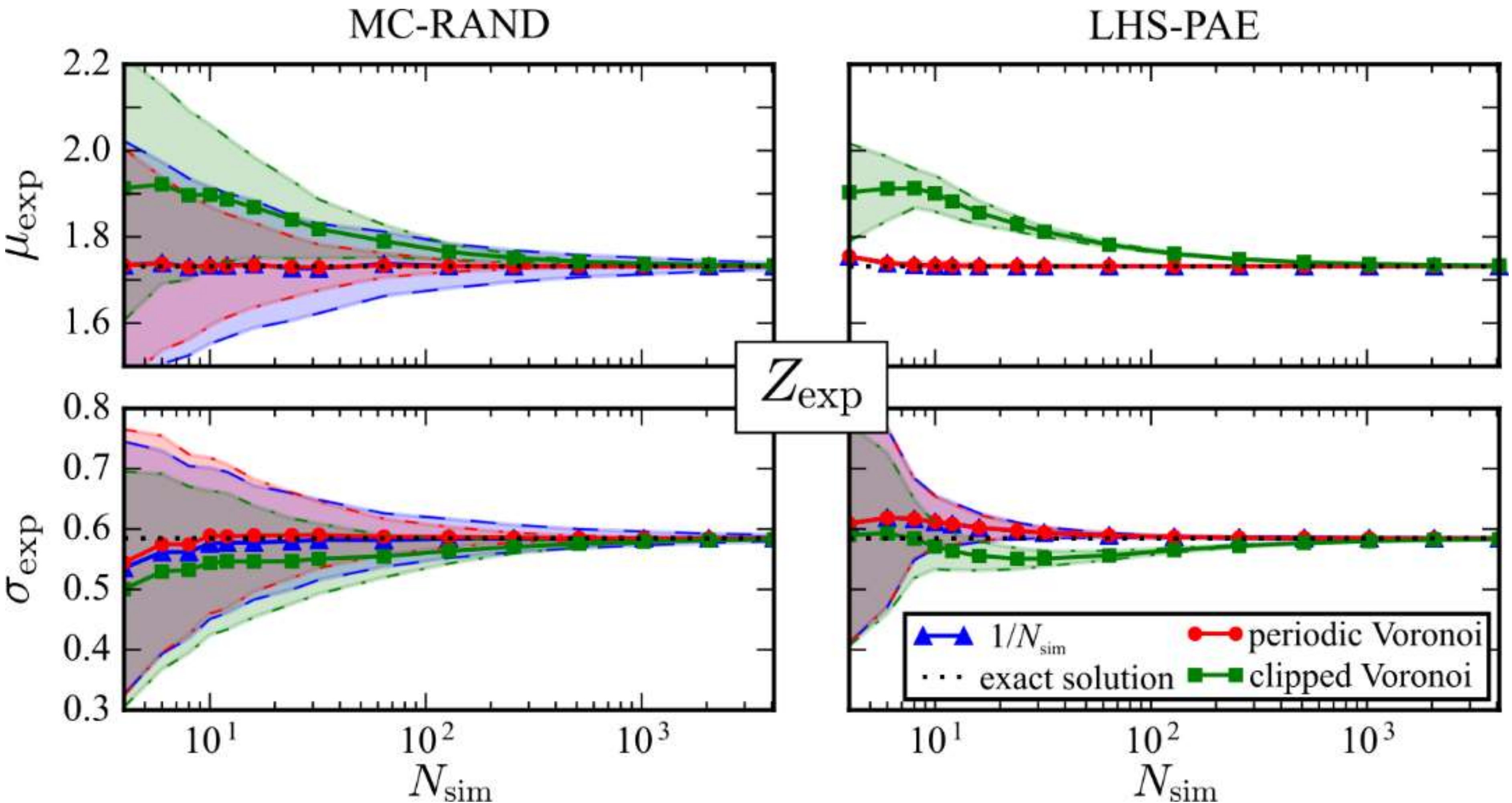
Numerical examples

$$g_{\mathrm{sum}}(\boldsymbol{X}) = \sum_{v=1}^{N_{\mathrm{var}}} X_v^2 \qquad \mu_{\mathrm{sum}} = N_{\mathrm{var}} \qquad \sigma_{\mathrm{sum}} = \sqrt{2N_{\mathrm{var}}}$$



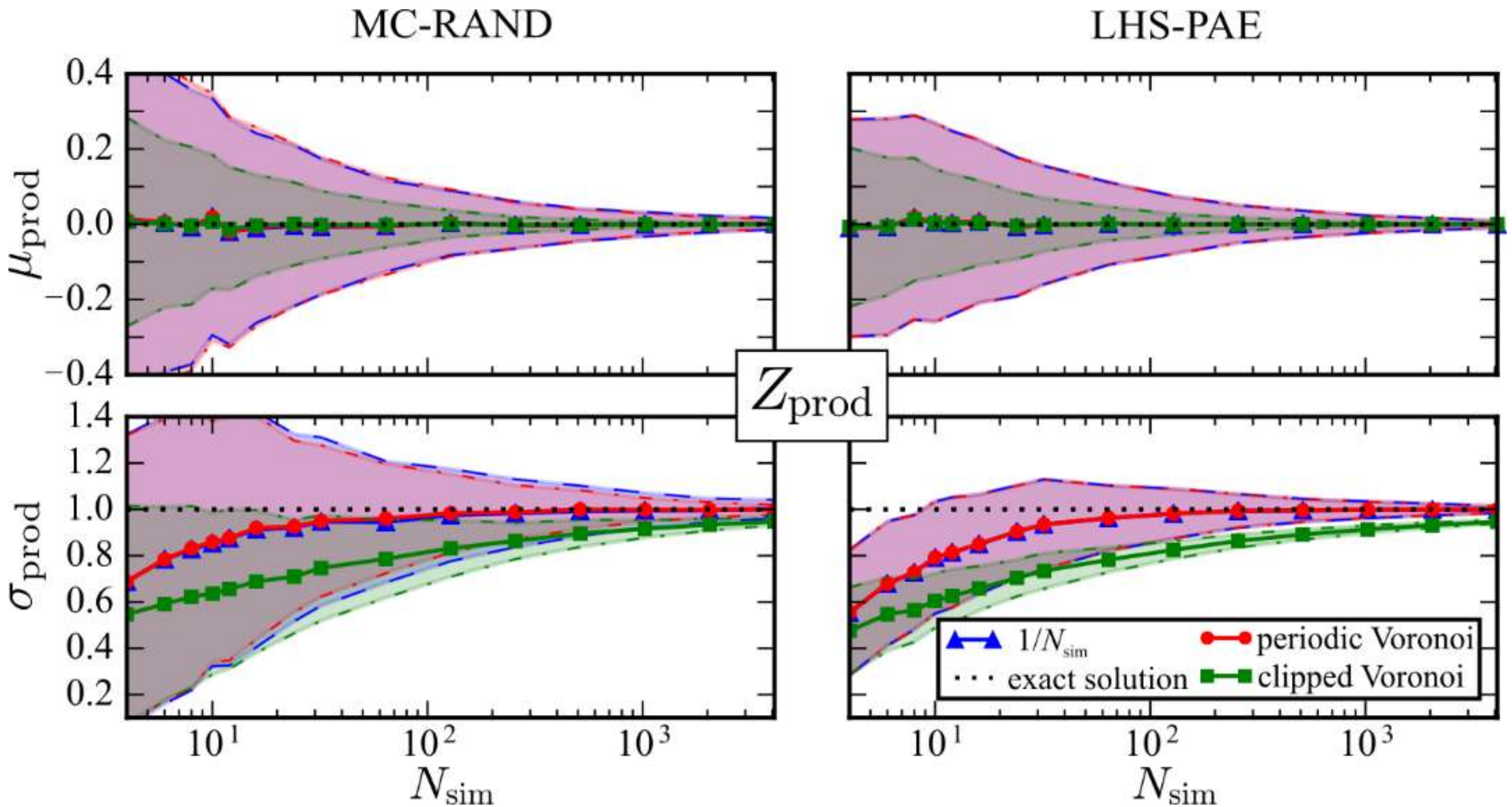
Numerical examples

$$g_{\exp}(\boldsymbol{X}) = \sum_{v=1}^{N_{\text{var}}} \exp(-X_v^2) \qquad \mu_{\exp} = \frac{\sqrt{3}}{3} N_{\text{var}} \qquad \sigma_{\exp} = \sqrt{N_{\text{var}}} \sqrt{\frac{\sqrt{5}}{5} - \frac{1}{3}}$$



Numerical examples

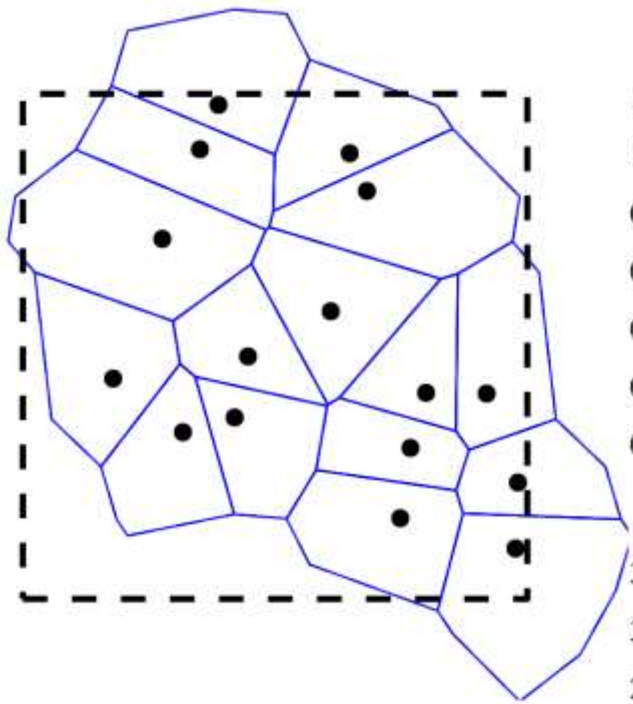
$$g_{\text{prod}}(\boldsymbol{X}) = \prod_{v=1}^{N_{\text{var}}} X_v \qquad \mu_{\text{prod}} = 0 \qquad \sigma_{\text{prod}} = 1$$

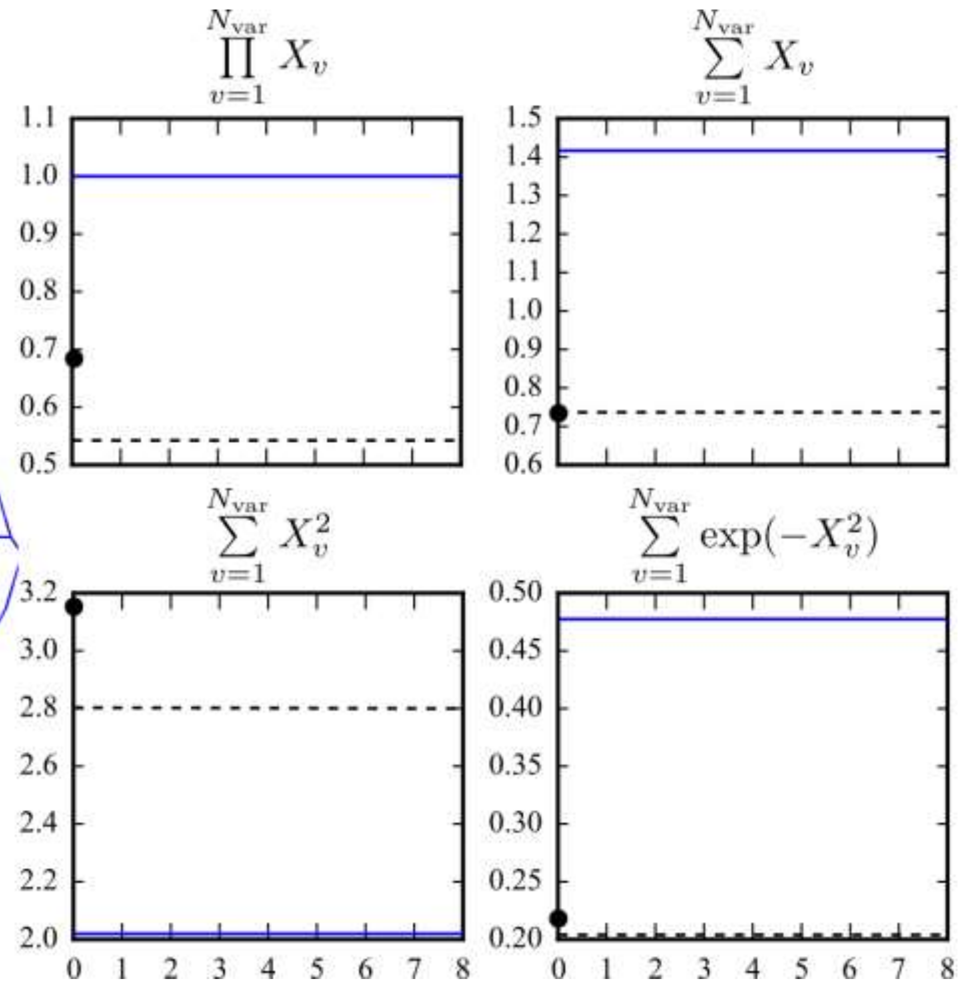


Conclusions

- *clipped* Voronoi tessellation (a tessellation limited to the design domain) inapplicable (presence of boundaries)
- *periodic* Voronoi tessellation slightly improves the integration if the location of sampling points is not optimized (e.g. crude MCS). However, the minor improvement does not seem to outweigh the additional effort spend on the evaluation of the volumes of the regions and the tessellation.

Future work

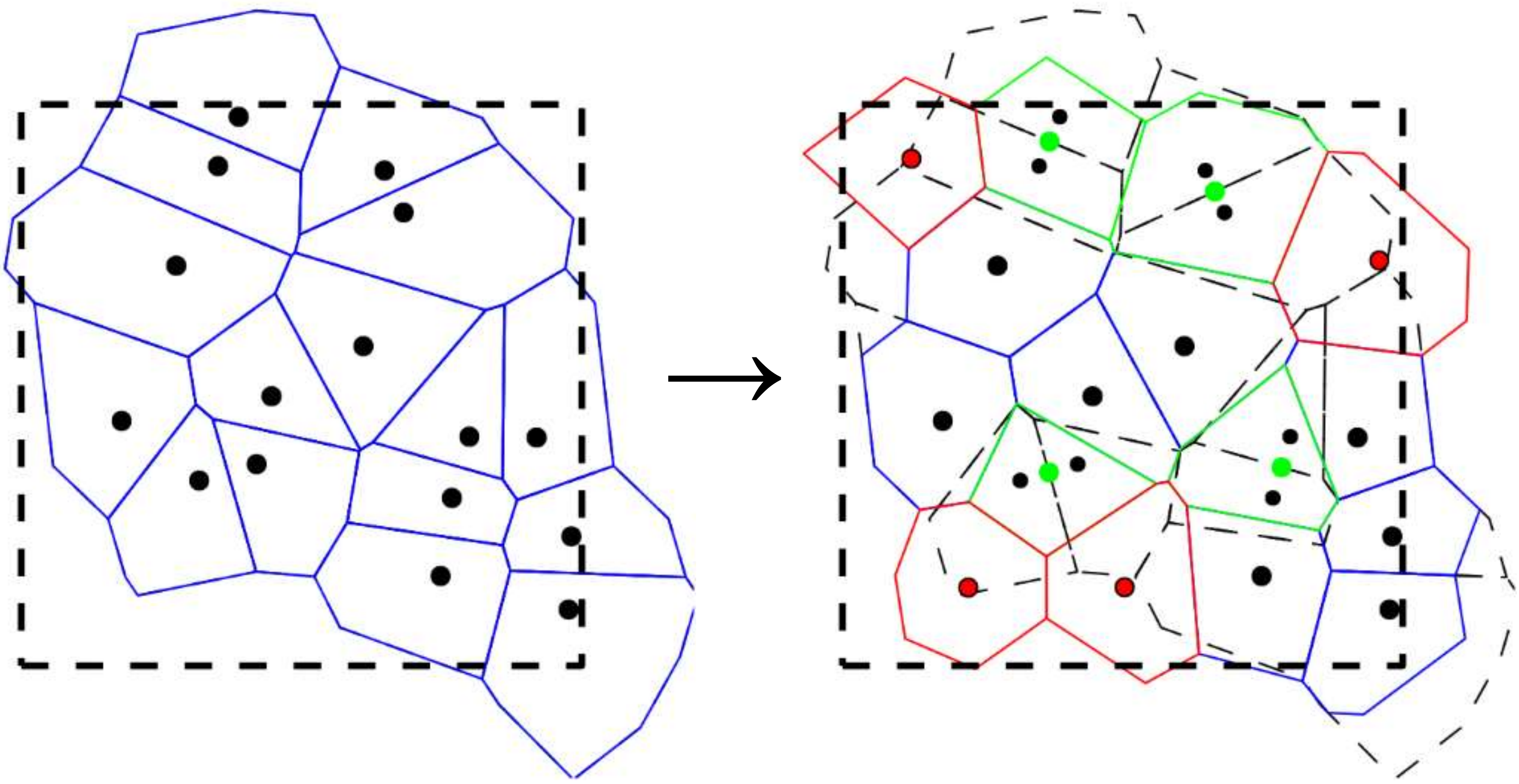




Weighted averages

- Grouping
  (for clustered points)
- Dummy points (unvisited regions)

Future work



$$\mathrm{E}[S\,[g\,(\boldsymbol{X})]] \approx \sum_{i=1}^{N_{\mathrm{sim}}} S\,[g(\boldsymbol{x}_i)] \cdot w_i$$

Future work

Lorenzo Rimoldini
Weighted skewness and kurtosis unbiased by sample size and Gaussian uncertainties, *Astronomy and Computing* 5 (2014) 1–8

$$V_p = \sum_{i=1}^{N_{\text{sim}}} w_i^p$$

unweighted forms for central moments → sample-size bias-corrected weighted mom's

$$M_2 = \frac{n}{n-1} m_2 = K_2$$

$$M_3 = \frac{n^2}{(n-1)(n-2)} m_3 = K_3$$

$$M_4 = \frac{n(n^2-2n+3)}{(n-1)(n-2)(n-3)} m_4 - \frac{3n(2n-3)}{(n-1)(n-2)(n-3)} m_2^2$$

$$K_4 = \frac{n^2(n+1)}{(n-1)(n-2)(n-3)} m_4 - \frac{3n^2}{(n-2)(n-3)} m_2^2$$

$$M_2 = \frac{V_1^2}{V_1^2 - V_2} m_2 = K_2$$

$$M_3 = \frac{V_1^3}{V_1^3 - 3V_1V_2 + 2V_3} m_3 = K_3$$

$$M_4 = \frac{V_1^2(V_1^4 - 3V_1^2V_2 + 2V_1V_3 + 3V_2^2 - 3V_4)}{(V_1^2 - V_2)(V_1^4 - 6V_1^2V_2 + 8V_1V_3 + 3V_2^2 - 6V_4)} m_4 - \frac{3V_1^2(2V_1^2V_2 - 2V_1V_3 - 3V_2^2 + 3V_4)}{(V_1^2 - V_2)(V_1^4 - 6V_1^2V_2 + 8V_1V_3 + 3V_2^2 - 6V_4)} m_2^2$$

$$K_4 = \frac{V_1^2(V_1^4 - 4V_1V_3 + 3V_2^2)}{(V_1^2 - V_2)(V_1^4 - 6V_1^2V_2 + 8V_1V_3 + 3V_2^2 - 6V_4)} m_4 - \frac{3V_1^2(V_1^4 - 2V_1^2V_2 + 4V_1V_3 - 3V_2^2)}{(V_1^2 - V_2)(V_1^4 - 6V_1^2V_2 + 8V_1V_3 + 3V_2^2 - 6V_4)} m_2^2$$