

The Leverhulme Research Centre  
for Functional Materials Design

# Maximally Dense Crystallographic Symmetry Group Packings through Entropic Trust Region

An Information Geometric Perspective

Miloslav Torda

Theory Cluster Meeting

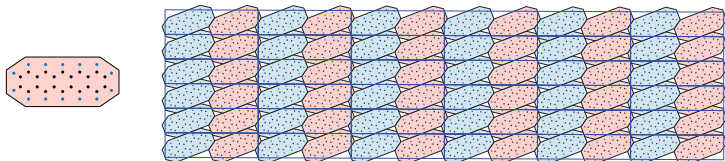
October 12, 2023

- M. TORDA, J. Y. GOULERMAS, R. PÚČEK, AND V. KURLIN, *Entropic Trust Region for Densest Crystallographic Symmetry Group Packings*, SIAM Journal on Scientific Computing, 45.4 (2023), pp. B493-B522.

# Crystal Structure Prediction (CSP)

## motivation

- An approach to accelerate Molecular CSP solvers:
  - Energy minimization  $\rightarrow$  Geometric packing density maximization



(Left) A geometric representation of pentacene as the convex hull of the atomic positions of the molecule with an offset given by hydrogen's van der Waals radius of  $1.09\text{\AA}$ . The dots symbolize atomic positions of (blue) hydrogen and (black) carbon. (Right) Visualization of the ETRPA output configuration of the densest  $p2$ -packing of the pentacene representation with density of  $0.9533821$  and resembles the configuration of single layer pentacene thin-film on graphite surface<sup>1</sup>, and via MD simulation of self-assembly of a disordered system of pentacene molecules on a graphene surface driven by the minimization of molecule-molecule interactions using the Lennard-Jones potential<sup>2</sup>.

<sup>1</sup>Chen W. et. al. (2008) Molecular orientation transition of organic thin films on graphite: the effect of intermolecular electrostatic and interfacial dispersion forces, Chemical Communications, pp. 4276–4278.

<sup>2</sup>Zhao Y. et. al. (2015) Molecular self-assembly on two-dimensional atomic crystals: insights from molecular dynamics simulations, The Journal of Physical Chemistry Letters, 6, pp. 4518–4524.

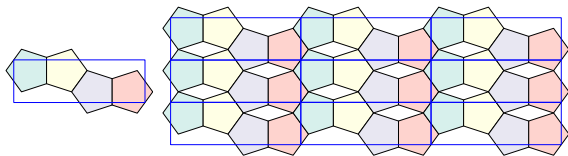
# Crystallographic Symmetry Group (CSG) packing

- **Configuration space:** Crystallographic Symmetry Group (CSG)
  - Discrete group of isometries of  $\mathbb{R}^n$  containing a lattice subgroup
- CSG packing

$$\mathcal{K}_G = \bigcup_{g \in G} gK,$$

$$\text{int}(g_i K) \cap \text{int}(g_j K) = \emptyset, \quad \forall g_i, g_j \in G, \quad g_i \neq g_j$$

- $G$  - CSG
- $K$  - Compact subset of  $\mathbb{R}^n$



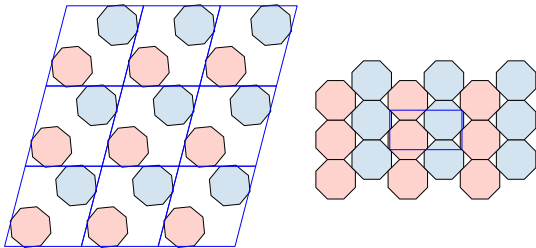
The 2D periodic structure with the  $p2mg$  plane group symmetry where  $K$  is a regular convex pentagon with the packing density of approximately 0.8541019. (Left) A single primitive cell. (Right) 9 primitive cells. The blue parallelogram denotes the primitive cell of the respective configuration. Colors represent symmetry operations modulo lattice translations.

# CSG packing problem

- CSG packing problem

$$\mathcal{K}_{\max} = \operatorname{argmax}_{\mathcal{K}_G: G \in \mathcal{G}} \rho(\mathcal{K}_G), \quad \mathcal{G} = \{H | H \cong G\}.$$

- $\rho(\mathcal{K}_G) = \frac{N \operatorname{area}(K)}{\det(\mathbf{U})}$  - 2D packing density.
  - $\mathbf{U}$  - Unit cell.
  - $N$  - number of symmetry operation modulo lattice translations.



CSG packings where  $\mathcal{G}$  of type  $p2$  and  $K$  is a regular octagon. (Left) packing with density  $\rho(\mathcal{K}_{p2}) \cong 0.413705$  and (right) optimal packing with density  $\rho(\mathcal{K}_{p2}) = \frac{4+4\sqrt{2}}{5+4\sqrt{2}} \cong 0.90616$ <sup>1</sup>. The blue parallelogram denotes the primitive cell of the respective configuration. Colors represent symmetry operations modulo lattice translations.

<sup>1</sup>Rogers, C. A. (1951). The closest packing of convex two-dimensional domains. Acta Mathematica, 86(1), 309-321.

# Entropic Trust Region Packing Algorithm

- Stochastic relaxation

$$\tilde{\theta} = \operatorname{argmax}_{\theta \in \Theta} J(\theta),$$

- $J(\theta) := E[\mathbf{F}|\theta] = \int_{\mathbf{x} \in \mathcal{X}} \mathbf{F}(\mathbf{x}) dP(\theta)$ 
  - $\mathbf{F}$  - Fitness of the packing density
  - $dP(\theta)$  - Probability measure from a parametric family  
 $S = \{dP(\theta) \mid \theta \in \Theta \subseteq \mathbb{R}^n\}$
  - $\mathcal{X}$  - Configuration space
- Non-euclidean trust region method
  - Trust region defined by the Kullback-Leibler Divergence from  $dP(\theta)$  to  $dP(\theta + \delta\theta)$ :  $D_{KL}(P_\theta \parallel P_{\theta+\delta\theta}) = \int_{\mathcal{X}} \ln \left( \frac{dP(\theta)}{dP(\theta+\delta\theta)} \right) dP(\theta)$ .
  - Update equations:  $\theta^{t+1} = \theta^t + \Delta^t \frac{\tilde{\nabla} J(\theta^t)}{\|\tilde{\nabla} J(\theta^t)\|_{\mathcal{I}_{\theta^t}}}$ 
    - $\tilde{\nabla} J(\theta) = \mathcal{I}_\theta^{-1} \nabla_\theta J(\theta)$  - Natural gradient<sup>1</sup> of the expected fitness  $J(\theta)$
    - $\mathcal{I}_\theta$  - Fisher metric tensor:  $\mathcal{I}_{\theta ij} = \int_{\mathcal{X}} \frac{\partial \ln(p(\theta))}{\partial \theta_i} \frac{\partial \ln(p(\theta))}{\partial \theta_j} dP(\theta)$ ;  
 $p(\theta) = \frac{dP(\theta)}{d\nu}$  is the Radon-Nikodym derivative of  $P(\theta)$  with respect to some reference measure  $\nu$  defined on  $\mathcal{X}$ .
    - $\Delta^t$  - Step size

---

<sup>1</sup>Amari S. I. (1998), Natural gradient works efficiently in learning. Neural Computation, 10, pp. 251–276.

# Extended Multivariate von Mises (EMvM) probability distribution

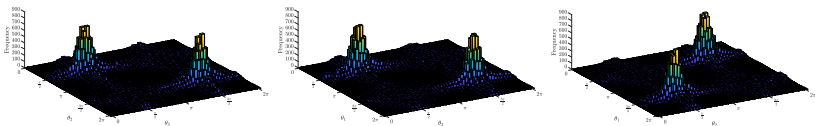
- The lattice subgroup  $L$  of a Crystallographic Symmetry Group induces quotient space  $\mathbb{R}^n/L \approx \mathcal{T}^n$
- Extended Multivariate von Mises<sup>1</sup> (EMvM) distribution:

$$f(\theta|\mu, \kappa, \mathbf{D}) = \frac{1}{Z(\mu, \kappa, \mathbf{D})} \exp \left\{ \kappa^T c(\theta - \mu) + \frac{1}{2} \left[ \begin{array}{c} c(\theta - \mu) \\ s(\theta - \mu) \end{array} \right]^T \mathbf{D} \left[ \begin{array}{c} c(\theta - \mu) \\ s(\theta - \mu) \end{array} \right] \right\}$$

$$c(\theta - \mu) = [\cos(\theta_1 - \mu_1), \dots, \cos(\theta_n - \mu_n)]^T \quad 0 \leq \theta_i \leq 2\pi, \quad 0 \leq \mu_i \leq 2\pi, \quad \mathbf{D} = \begin{bmatrix} \mathbf{D}^{CC} & \mathbf{D}^{CS} \\ \mathbf{D}^{CS \top} & \mathbf{D}^{SS} \end{bmatrix}$$

$$s(\theta - \mu) = [\sin(\theta_1 - \mu_1), \dots, \sin(\theta_n - \mu_n)]^T \quad 0 \leq \kappa_i$$

- $\mathbf{D}$  -  $2n \times 2n$  real-valued symmetric matrix with diagonal elements of  $\mathbf{D}^{CC}$ ,  $\mathbf{D}^{SS}$  and  $\mathbf{D}^{CS}$  set to zero.
  - Controls cosine-cosine, sine-sine and cosine-sine interactions.



Histograms of projections along **(left)**  $\theta_1$ , **(middle)**  $\theta_2$  and **(right)**  $\theta_3$  coordinates of 100000 samples from the trivariate EMvM.

<sup>1</sup>Mardia K. V. et. al. (2008). A multivariate von mises distribution with applications to bioinformatics. Canadian Journal of Statistics, 36, pp. 99–109.

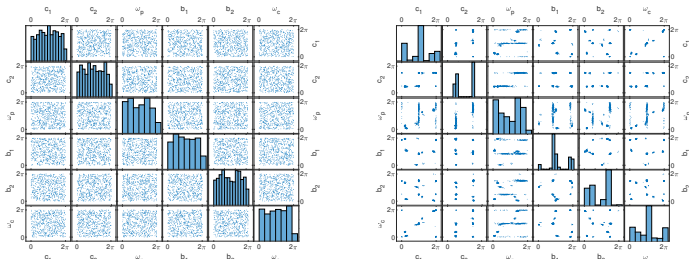
# Exponential reformulation of the EMvM

- Exponential family Extended Multivariate von Mises distribution

$$f(\boldsymbol{\theta} | \boldsymbol{\eta}, \mathbf{E}) = \exp \left\{ \begin{bmatrix} c(\boldsymbol{\theta}) \\ s(\boldsymbol{\theta}) \end{bmatrix}^T \boldsymbol{\eta} + \text{vec} \left( \begin{bmatrix} c(\boldsymbol{\theta}) \\ s(\boldsymbol{\theta}) \end{bmatrix} \begin{bmatrix} c(\boldsymbol{\theta}) \\ s(\boldsymbol{\theta}) \end{bmatrix}^T \right)^T \text{vec}(\mathbf{E}) - \psi(\boldsymbol{\eta}, \mathbf{E}) \right\}$$

$$c(\boldsymbol{\theta}) = [\cos(\theta_1), \dots, \cos(\theta_n)]^T \quad 0 \leq \theta_j \leq 2\pi, \quad \boldsymbol{\eta} = \begin{bmatrix} \kappa \odot c(\boldsymbol{\mu}) \\ \kappa \odot s(\boldsymbol{\mu}) \end{bmatrix}, \quad \mathbf{E} = \begin{bmatrix} \mathbf{E}^{CC} & \mathbf{E}^{CS} \\ \mathbf{E}^{SC} & \mathbf{E}^{SS} \end{bmatrix}, \quad \begin{array}{l} 0 \leq \mu_j \leq 2\pi \\ 0 < \kappa_j \end{array}$$

- $\mathbf{E}$  -  $2n \times 2n$  real-valued symmetric matrix with diagonal elements of  $\mathbf{E}^{CC}$ ,  $\mathbf{E}^{SS}$  and  $\mathbf{E}^{CS}$  set to zero.
- $\dim(S_{EMvM}) = 2n^2 > \frac{n(n+3)}{2} = \dim(S_{MvM})$

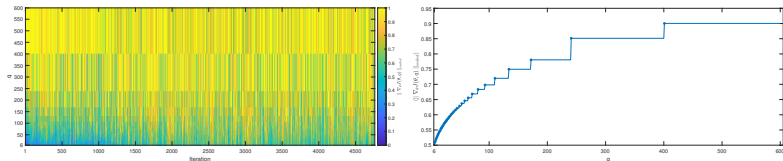


600 realizations of the Extended Multivariate von Mises distribution defined on an  $T^6$  from a single run of the ETRPA. **(Left)** Initial distribution and **(Right)** output distribution.



# Truncation selection transformation of the fitness function $\mathbf{F}$

- $\tilde{\mathbf{F}} := q \mathbf{1}_{\mathbf{F} \geq \mathbf{F}_{1-\frac{1}{q}}^{\tilde{\theta}}}(\mathbf{x}) = \begin{cases} q & \text{if } \mathbf{F}(\mathbf{x}) \geq \mathbf{F}_{1-\frac{1}{q}}^{\tilde{\theta}} \\ 0 & \text{otherwise} \end{cases}$ 
  - $\mathbf{F}_{1-\frac{1}{q}}^{\tilde{\theta}}$  is the  $P_{\tilde{\theta}}$ 's  $(q-1)$ th  $q$ -quantile of the fitness  $\mathbf{F}$ .
- $S^e := \{dP^e(\theta) = \exp\{\theta^T \mathbf{t}(\mathbf{x}) - \psi(\theta)\} dP_0(\mathbf{x})\}$ ,
  - Reference measure  $dP_0(\mathbf{x})$ .
  - The log-partition function  $\psi(\theta) = \ln \int \exp\{\theta^T \mathbf{t}(\mathbf{x})\} dP_0(\mathbf{x})$ .
- $\nabla_{\theta} J(\theta) |_{\theta=\tilde{\theta}} = \mu_{\mathbf{F}_{1-\frac{1}{q}}^{\tilde{\theta}}} - \mu$ ,
  - $\mu_{\mathbf{F}_{1-\frac{1}{q}}^{\tilde{\theta}}} = \int_{\mathbf{F}_{1-\frac{1}{q}}^{\tilde{\theta}}}^{\infty} \mathbf{F}^{-1}(y) \exp\{\tilde{\theta}^T \mathbf{t}(\mathbf{F}^{-1}(y)) - \psi(\tilde{\theta}) + \ln(q_t)\} dP_0 \circ \mathbf{F}^{-1}(y)$ .
  - $\mu = \int \mathbf{x} \exp\{\tilde{\theta}^T \mathbf{t}(\mathbf{x}) - \psi(\tilde{\theta})\} dP_0(\mathbf{x})$ .



Influence of  $q$  on (Left) the scaled expected fitness gradient  $\|\nabla_{\theta} J(\theta^t, q)\|_{\text{scaled}}$  and (Right) scaled expected fitness gradient averaged through all iterations ( $\langle \|\nabla_{\theta} J(\theta, q)\|_{\text{scaled}} \rangle$ ).

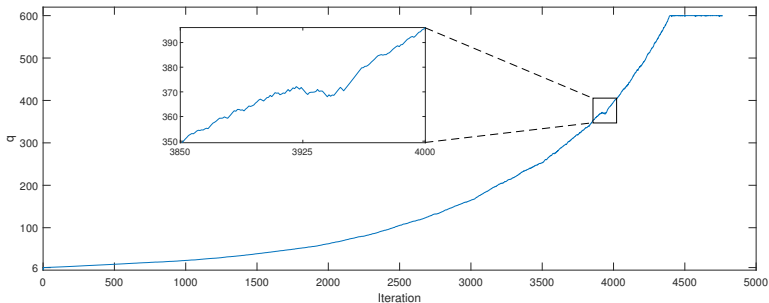
## Adaptive selection quantile

- $q$  as an equivalent to the annealing temperature control parameter

$$q_{t+1} = q_t \exp\{\beta \cos(\alpha^t)\}$$

$$\cos(\alpha^t) = \frac{\langle \Delta\theta^t, \Delta\theta^{t-1} \rangle_{\mathcal{I}_{\theta^{t-1}}}}{\|\Delta\theta^t\|_{\mathcal{I}_{\theta^{t-1}}} \|\Delta\theta^{t-1}\|_{\mathcal{I}_{\theta^{t-1}}}}, \quad \Delta\theta^t = \theta^t - \theta^{t-1},$$
$$\Delta\theta^{t-1} = \theta^{t-1} - \theta^{t-2},$$

- $\langle \cdot, \cdot \rangle_{\mathcal{I}_{\theta^{t-1}}}$  - scalar product with respect to  $\mathcal{I}_{\theta^{t-1}}$ .
- $\Delta\theta^t, \Delta\theta^{t-1} \in \mathcal{T}_{\theta^{t-1}}\mathcal{S}$ .



Evolution of the adaptive selection quantile  $q$ .

## Entropic proximal maximization for exponential families

- Euclidean fitness gradient  $\nabla_{\theta} J(\theta) |_{\theta=\tilde{\theta}} = \mu_{F_{1-\frac{1}{q^*}}} - \mu$ , for a fixed  $q^*$ , determines proximal map maximization<sup>1</sup>

$$\theta^{t+1} = \operatorname{argmax}_{\theta \in \Theta} \left\{ \theta^{\top} \nabla_{\theta} J(\theta^t) - \frac{1}{\epsilon} D_{KL}(P_{\theta^t} \parallel P_{\theta}) \right\},$$

- Natural gradient updates in dual coordinate systems via the Legendre transforms  $\mu = \nabla_{\theta} \psi(\theta)$  and  $\theta = \nabla_{\mu} \phi(\mu)$ <sup>2</sup>:

$$\theta^{t+1} = \theta^t + \epsilon^t \mathcal{I}_{\theta^t}^{-1} \nabla_{\theta} J(\theta^t), \quad \mu^{t+1} = \mu^t + \epsilon^t \mathcal{I}_{\mu^t}^{-1} \nabla_{\mu} J(\mu^t),$$

- ETRPA:  $\epsilon^t = \frac{\Delta^t}{\sqrt{\nabla_{\theta} J(\theta^t)^{\top} (\mathcal{I}_{\theta^t})^{-1} \nabla_{\theta} J(\theta^t)}}$ ,

---

<sup>1</sup>Beck A. and Teboulle M. (2003) Mirror descent and nonlinear projected subgradient methods for convex optimization, Operations Research Letters, 31, pp. 167–175.

<sup>2</sup>Raskutti G. and Mukherjee S. (2015). The information geometry of mirror descent, IEEE Transactions on Information Theory, 61, pp. 1451–1457.

# Information-geometry of ETRPA

- ETRPA as a solution of a maximin problem

$$\max_{P_{\theta^{q^*}} \in S^{q^*}} \min_{P_{\theta^1} \in S^1} D_{KL}(P_{\theta^{q^*}} \parallel P_{\theta^1}); S^{q^*}, S^1 \subset S.$$

- $S = \bigcup_{q \in [1; \infty)} S^q$ ,  $S^q = \{dP^q(\theta) \mid \theta \in \Theta^q \subset \mathbb{R}^n, \}$

$$dP^q(\theta) = \begin{cases} \exp \{ \theta^T \mathbf{t}(\mathbf{F}^{-1}(y)) - \psi(\theta) + \ln(q) \} dP_0 \circ \mathbf{F}^{-1}(y) & y \geq \mathbf{F}_{1-\frac{1}{q}}^\theta \\ 0 & \text{otherwise} \end{cases}$$

- Fixing  $P_{\theta^{q^*}}$  minimizing with respect to  $\theta^1$ , and fixing  $P_{\theta^1}$  maximizing with respect to  $\theta^q$

$$\theta^{1^{t+1}} = \theta^{1^t} + \epsilon \mathcal{I}_{\theta^{1^t}}^{-1}(\mu^{q^*t} - \mu^{1^t}), \quad \mu^{q^*t+1} = \mu^{q^*t} + \epsilon \mathcal{I}_{\mu^{q^*t}}^{-1}(\theta^{q^*t} - \theta^{1^t}).$$

- Considering  $S^{q^*}$  and  $S^1$  as embeddings of  $S^e$  in  $S$  we recover  $S^e$  from  $S$

- For  $\theta^1 \in S^1$  the  $\theta$ -coordinates of  $S^e$ :  $[\theta_i^1]_{i=1, \dots, n} \rightarrow \theta$ .

- For  $\mu^{q^*} \in S^{q^*}$  the  $\mu$ -coordinates of  $S^e$ :

$$\mu = \frac{1}{q^*} [\mu_i^{q^*}]_{i=1, \dots, n} + \frac{q^* - 1}{q^*} \mu_{\mathbf{F}^C}^{1-\frac{1}{q^*}},$$

$$\mu_{\mathbf{F}^C}^{1-\frac{1}{q^*}} = \int_{-\infty}^{\mathbf{F}_{1-\frac{1}{q^*}}^\theta} \mathbf{F}^{-1}(y) \exp \left\{ \theta^T \mathbf{t}(\mathbf{F}^{-1}(y)) - \psi(\theta) + \ln \left( \frac{q^*}{q^* - 1} \right) \right\} dP_0 \circ \mathbf{F}^{-1}(y)$$

## Maximization of stochastic dependence

- $\max_{P_{\theta^{q^*}} \in S^{q^*}} \min_{P_{\theta^1} \in S^1} D_{KL}(P_{\theta^{q^*}} \parallel P_{\theta^1})$
- For a fixed  $P_{\theta^{q^*}} \in S^{q^*}$

$$\hat{\theta}^1 = \operatorname{argmin}_{P_{\theta^1} \in S^1} D_{KL}(P_{\theta^{q^*}} \parallel P_{\theta^1})$$

- $P_{\hat{\theta}^1} \in S^1$  - maximum entropy estimate of  $P_{\theta^{q^*}}$
- $D_{KL}(P_{\theta^{q^*}} \parallel P_{\hat{\theta}^1}) = D_{KL}(P_{\theta^{q^*}} \parallel P_{\hat{\theta}_s^1}) - D_{KL}(P_{\hat{\theta}^1} \parallel P_{\hat{\theta}_s^1})$ 
  - $P_{\hat{\theta}_s^1} \in S^1$  - split model corresponding to  $P_{\hat{\theta}^1}$
- For a fixed  $P_{\hat{\theta}^1}$ :

$$\max_{P_{\theta^{q^*}} \in S^{q^*}} D_{KL}(P_{\theta^{q^*}} \parallel P_{\hat{\theta}^1}) \Leftrightarrow \max_{P_{\theta^{q^*}} \in S^{q^*}} D_{KL}(P_{\theta^{q^*}} \parallel P_{\hat{\theta}_s^1})$$

- Special case of multi-information/total correlation<sup>1</sup>:

$$D_{KL}(\cdot \parallel P_{\hat{\theta}_s^1})$$

- A measure of stochastic dependence in complex systems.

---

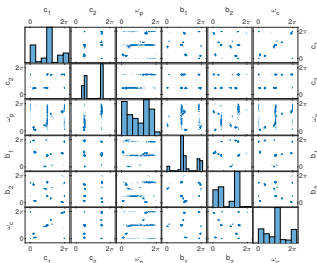
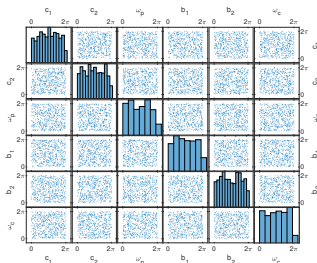
<sup>1</sup>Studený M. and Vejnarová J. (1998). The multiinformation function as a tool for measuring stochastic dependence, in Learning in Graphical Models. pp. 261–297.

# Maximization of stochastic dependence

- Multi-information as generalization<sup>1</sup> of the Infomax principle<sup>2</sup>:

$$\max_{\{f \in \mathcal{F} | O=f(I)\}} MI(I; O)$$

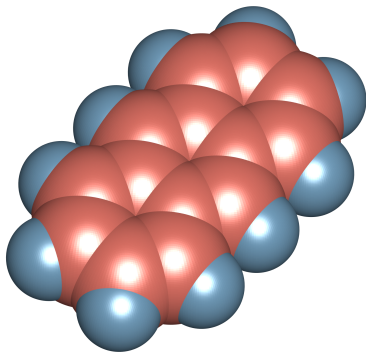
- A rule for training artificial neural networks.
- ETRPA moves towards maximizing stochastic dependence among elements of the extended multivariate von Mises distributed random vector.
  - Learns the optimization landscape given by the CSG packing problem.



<sup>1</sup>Ay N. and Knauf A. (2006). Maximizing multi-information, *Kybernetika*, 42, pp. 517–538.

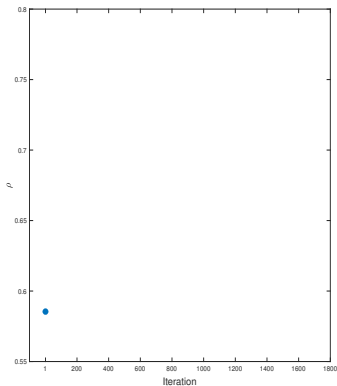
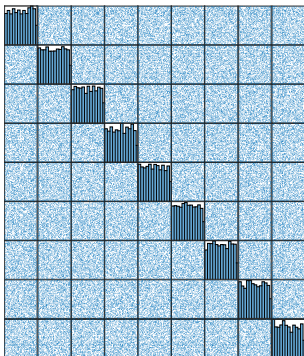
<sup>2</sup>Linsker R. (1997). A local learning rule that enables information maximization for arbitrary input distributions, *Neural Computation*, 9, pp. 1661–1665.

Van der Waals sphere model molecule  
packing  
Anthracene



# Van der Waals sphere model molecule packing

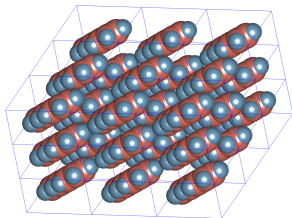
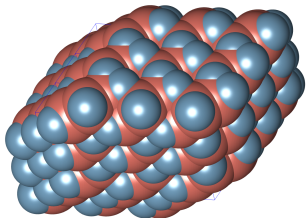
*P1* packing of ANTHRACENE





# Van der Waals sphere model molecule packing

Overlap constraints



(**Left**) Unfeasible  $P1$  configuration and (**right**) corresponding feasible configuration.

# Van der Waals sphere model molecule packing

Some Riemannian gymnastics

- Recall the update equations:  $\theta^{t+1} = \theta^t + \Delta^t \frac{\tilde{\nabla} J(\theta^t)}{\|\tilde{\nabla} J(\theta^t)\|_{\mathcal{I}_{\theta^t}}}$ 
  - $\tilde{\nabla} J(\theta) = \mathcal{I}_{\theta}^{-1} \nabla_{\theta} J(\theta)$
- Dually flat structure of Exponential families given by the Legendre transforms:  $\mu = \nabla_{\theta} \psi(\theta)$  and  $\theta = \nabla_{\mu} \phi(\mu)$ 
  - $\mu = \mathbf{E}[\mathbf{X} \mid f(x|\theta)]$

- Adaptive selection quantile gradient:

$$\nabla_{\theta} J(\theta) |_{\theta=\bar{\theta}} = \mu_{F_{1-\frac{1}{q^*}}} - \mu \in \mathcal{T}_{\mu} \Leftrightarrow \mathcal{I}_{\theta}^{-1} \left( \mu_{F_{1-\frac{1}{q^*}}} - \mu \right) \in \mathcal{T}_{\theta}$$

- $\mathcal{I}_{\theta^t}^{-1} \left( \mu_{\text{rep}}^t - \mu^t \right) \in \mathcal{T}_{\theta^t}$

- $\mathcal{I}_{\theta_{\text{rep}}^t}^{-1} \left( \mu_{\text{rep} \rightarrow \text{sel}}^t - \mu_{\text{rep}}^t \right) \in \mathcal{T}_{\theta_{\text{rep}}^t}$

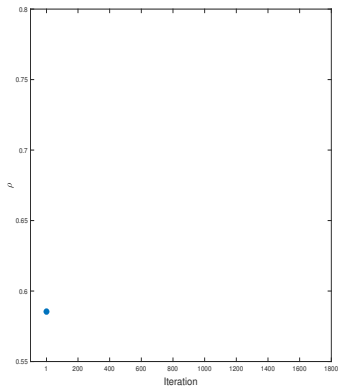
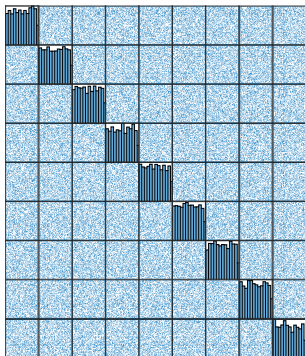
- $\mathcal{I}_{\theta_{\text{rep} \rightarrow \text{sel}}^t}^{-1} \left( \mu_{\text{rep} \rightarrow \text{sel} F_{1-\frac{1}{q^*}}}^t - \mu_{\text{rep} \rightarrow \text{sel}}^t \right) \in \mathcal{T}_{\theta_{\text{rep} \rightarrow \text{sel}}^t}$

- $\tilde{\nabla} J(\theta^t) =$

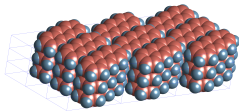
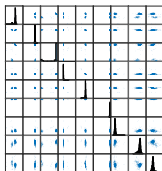
$$\mathcal{I}_{\theta_{\text{rep} \rightarrow \text{sel}}^t}^{-1} \left( \mu_{\text{rep} \rightarrow \text{sel} F_{1-\frac{1}{q^*}}}^t - \mu_{\text{rep} \rightarrow \text{sel}}^t \right) + \mathcal{I}_{\theta_{\text{rep}}^t}^{-1} \left( \mu_{\text{rep} \rightarrow \text{sel}}^t - \mu_{\text{rep}}^t \right) + \mathcal{I}_{\theta^t}^{-1} \left( \mu_{\text{rep}}^t - \mu^t \right)$$

# Van der Waals sphere model molecule packing

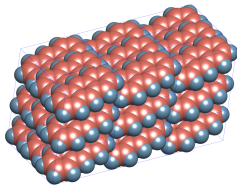
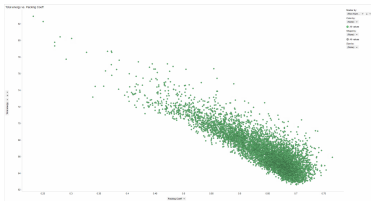
*P1* packing of ANTHRACENE



# Van der Waals sphere model molecule packing



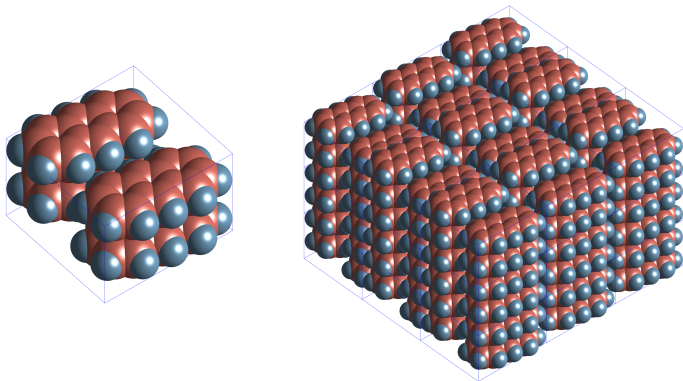
(Left) *P1* configurations with packing density higher than 0.7 and (right) highest density *P1* configuration with density of 0.7845.



(Left) Energy-density landscape from Anthracene *P1* CSP and (right) lowest energy *P1* configuration with density of 0.7006.

# Van der Waals sphere model molecule packing

*P21C* configuration with packing density of 0.7507



# THANK YOU



[milotorda.net](http://milotorda.net)