UNIVERSITY OF LIVERPOOL

LEVERHULME TRUST

The Leverhulme Research Centre for Functional Materials Design

Information-geometry based space group optimization algorithm

Miloslav Torda

Supervisors:

Prof Yannis Goulermas
Dr Vitaliy Kurlin

Prof Graeme M. Day
Dr Linjiang Chen

CCDC diamond Hartree Centre Science & Technology Facilities Council Imperial College London ICNS King Abdullah University of Science and Technology University of Southampton

Table of Contents

Molecular crystal structure prediction

CSP problem

- Prediction of stable crystal structures of solids from first principles given some pressure-temperature conditions.
- **Approaches**
  - Topological
  - Data-mining
  - Computational optimization

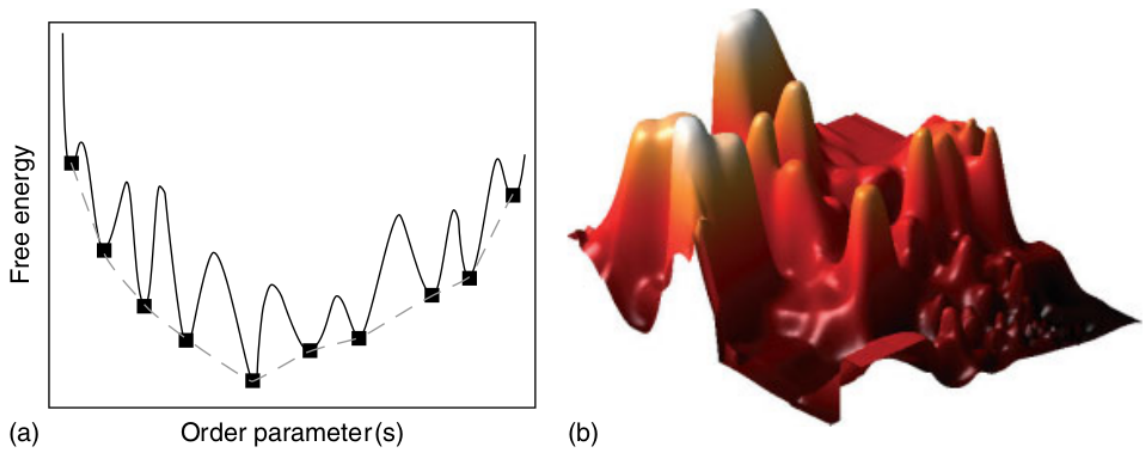
Molecular crystal structure prediction

Energy landscapes

- The stability of a molecular crystal is assessed by it's free energy

$$\mathbf{E} : \mathbf{X} \to \mathbb{R}$$

$\mathbf{X}$ is a configuration space of a molecular crystal



Energy landscape: (a) schematic illustration showing the full landscape (solid line) and reduced land- scape (dashed line interpolating local minima points), (b) 2D-projection of the reduced landscape of $\mathbf{Au_8Pd_4}$ showing all low-energy structures clustered in one region of configuration space.

A. R. Oganov ed., Modern Methods of Crystal Structure Prediction, 2011.

Molecular crystal structure prediction

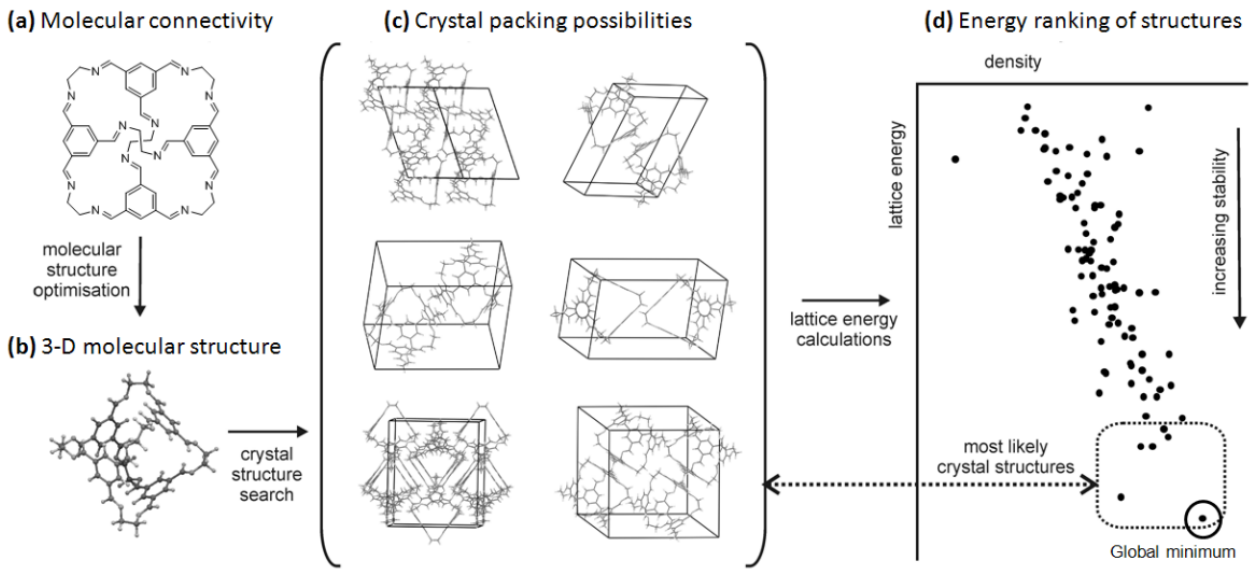
Molecular CSP work flow

In principle a "well defined" bounded constrained optimization problem

$$\hat{x} = \underset{x \in \mathbf{X}}{\operatorname{argmin}}\, \mathbf{E}(x)$$

subject to

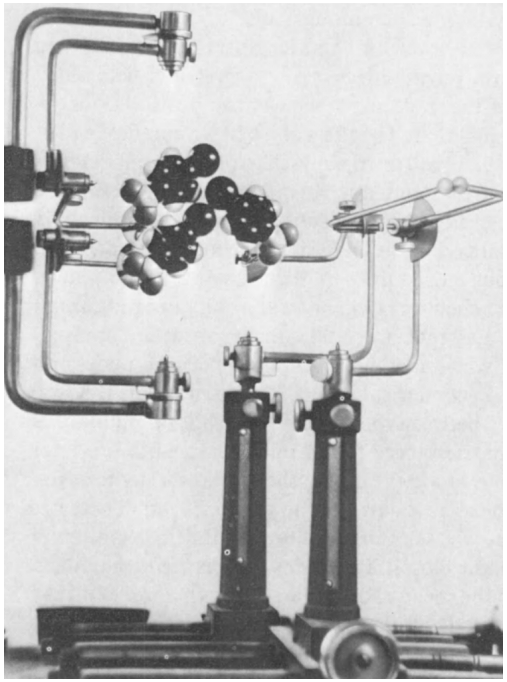
- periodic boundary conditions.
- fundamental region constraints.
- nonoverlap constraints.



G. M. Day and A. I. Cooper, Energy–Structure–Function Maps: Cartography for Materials Discovery, Advanced Materials, 2018, 30, 1704944.

Molecular crystal structure prediction

Structure seeker



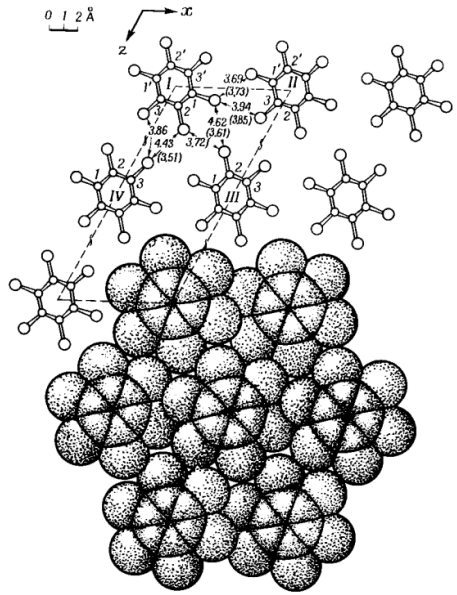


Fig. 3. Projection xOz of hexachlorobenzene structure.

A. I. Kitaigorodsky, Molecular crystals and Molecules, 1973, Academic Press.

- instead of minimizing **E(x)**:

$$\mathbf{E(x)} \rightarrow \mathbf{F(x)}$$

  F - fitness ($\mathbf{E}^2$, $|\mathbf{E}|$ etc)
- **Black box** means no additional information on **E(x)** besides fitness evaluation is required e.g. continuity, differentiability.
- **Define**:

$$g(\theta) := E[\mathbf{F}|\theta] = \int \mathbf{F}(\mathrm{x}) dP(\theta)$$

  $P(\theta)$ - probability measure
- **New formulation of the original problem:**

$$\tilde{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}}\, g(\theta).$$

- Family of probability distributions parametrized by $\theta \in \Theta \subseteq \mathbb{R}^n$:

$$S = \{P(\theta) = P(x;\theta) | \theta \in \Theta\}$$

  is called a **Statistical manifold**
- **Fisher information**:

$$F_{ij}(\theta) = -\int \frac{\partial^2 \ln P(\theta)}{\partial\theta_i\theta_j} dP(\theta)$$

- (S, F) has the structure of a Riemannian manifold.
  - F is the metric tensor.
- **Information-geometry** studies these objects.

- Gradient of $g(\boldsymbol{\theta})$ on (S, F) (Amari 1998):

$$\tilde{\nabla} g(\boldsymbol{\theta}) = F^{-1} \nabla g(\boldsymbol{\theta})$$

  - $\tilde{\nabla} = F^{-1}\nabla$ - Natural gradient
  - ∇ - Vanilla gradient
- IGO gradient flow on S:

$$\frac{d\theta}{dt} = \tilde{\nabla} g(\theta).$$

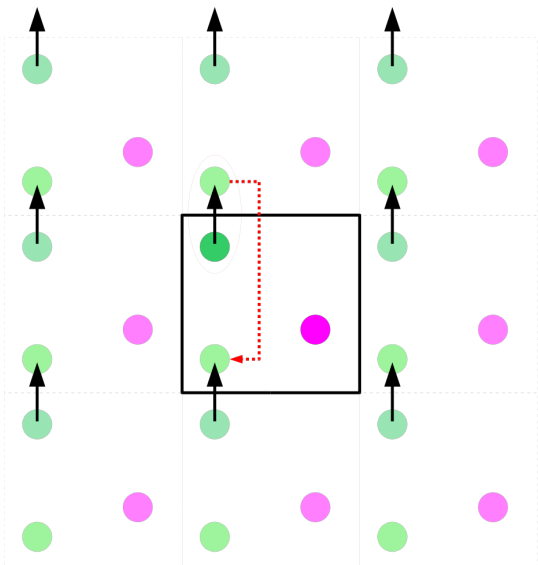
- IGO gradient ascent:

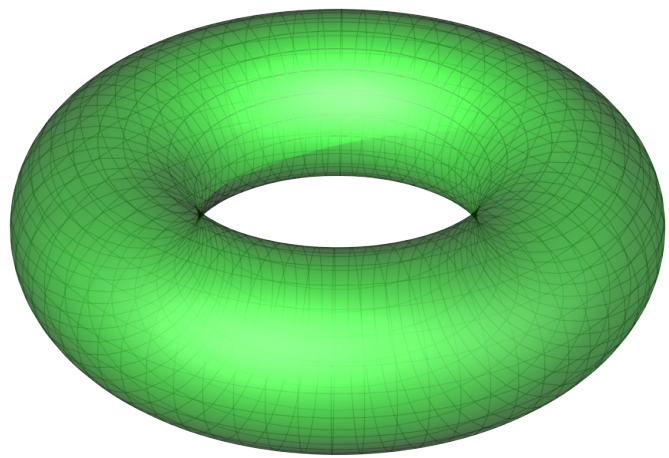
$$\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^{t} + \gamma \tilde{\nabla} g(\boldsymbol{\theta}).$$

  - γ - step size or learning rate
- Notable instance of IGO: Natural Evolution Strategies

1. Statistical manifold
2. Fitness function (Utility function)
3. Learning rate

- Periodic boundary conditions induce a quotient space $\mathbb{R}^n/\Lambda$
  - Λ is the group of a n-dimensional lattice
- $\mathbb{R}^n/\Lambda$ is topologically equivalent to a n-torus T^n.
- restricting S to probability distributions on n-torus we naturally get rid of the boundary constraints in bounded optimization.
  - n-torus is without boundary





wikipedia

Multivariate von Mises distribution (MvM)

- Probability distribution on a n-torus (Mardia et. al. 2008):

$$f(\boldsymbol{\theta}|\boldsymbol{\mu},\boldsymbol{\kappa},\Lambda) =$$

$$= \frac{1}{Z(\boldsymbol{\mu},\boldsymbol{\kappa},\Lambda)} \exp\left\{\boldsymbol{\kappa}^{\mathrm{T}} c(\boldsymbol{\theta}-\boldsymbol{\mu}) + \frac{1}{2} s(\boldsymbol{\theta}-\boldsymbol{\mu})^{\mathrm{T}} \Lambda s(\boldsymbol{\theta}-\boldsymbol{\mu})\right\}$$

$$c(\boldsymbol{\theta}-\boldsymbol{\mu}) = [cos(\theta_1-\mu_1),\ldots,cos(\theta_n-\mu_n)]^{\mathrm{T}}$$

$$s(\boldsymbol{\theta}-\boldsymbol{\mu}) = [sin(\theta_1-\mu_1),\ldots,sin(\theta_n-\mu_n)]^{\mathrm{T}}$$

$$-\pi \leq \theta_i \leq \pi,\ -\pi \leq \mu_i \leq \pi,\ 0 \leq \kappa_i,\ -\infty \leq \lambda_{ij} \leq \infty$$

$$\Lambda_{ij} = \lambda_{ij} = \lambda_{ji},\ i \neq j,\ ,\lambda ii = 0$$

- $Z(\boldsymbol{\mu},\boldsymbol{\kappa},\Lambda)$ - normalization constant
  - known only for the special case of bivariate von Mises distribution
- $\boldsymbol{\mu}$ - mean direction parameters
- $\boldsymbol{\kappa}$ - concentration parameters
- Λ - parameters for interactions between angles

Extended Multivariate von Mises distribution (EMvM)

- We extended MvM:

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

$$0 \leq \theta_i \leq 2\pi,\ 0 \leq \mu_i \leq 2\pi,\ 0 \leq \kappa_i,\ -\infty \leq d_{ij} \leq \infty$$
$$\mathrm{D}_{i,j} = d_{i,j} = d_{j,i},\ i,j = 1,\ldots,2n$$
$$d_{k,k} = -d_{k+n,k+n},\ k = 1,\ldots,n$$

- $S_{\mathrm{MvM}} \subset S_{\mathrm{EMvM}}$
  - since $\dim(S_{\mathrm{EMvM}}) = 2n(n+1) > \dim(S_{\mathrm{MvM}}) = \frac{n(n+3)}{2}$
  - MvM is a special case of EMvM
- Special cases:
  - $\boldsymbol{\kappa} = 0, \mathrm{D} = 0$ - Uniform distribution on n-torus
  - $\boldsymbol{\kappa} \to \infty, \mathrm{D} \to \infty$ - Dirac δ function.

hIGO

Monte Carlo estimation of the hIGO gradients

- Since the normalization constant Z in EMvM is not known we can not derive an explicit natural gradient $\tilde{\nabla} g(\boldsymbol{\theta})$ formula.
- Exponential family of probability distributions:

$$f(\boldsymbol{x}|\boldsymbol{\theta}) = h(\boldsymbol{x}) \exp\left\{\boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \boldsymbol{T}(\boldsymbol{x}) - A(\boldsymbol{\theta})\right\}$$

- IGO gradient flow for exponential family of distributions parametrization $\boldsymbol{\eta}$ (Ollivier et. al. 2017)

$$\frac{d\boldsymbol{\eta}}{dt} = \tilde{\nabla}_{\boldsymbol{\eta}} g = \mathrm{Cov}_{P_{\boldsymbol{\eta}}}(\boldsymbol{T}(\boldsymbol{x}), \boldsymbol{T}(\boldsymbol{x}))^{-1} \mathrm{Cov}_{P_{\boldsymbol{\eta}}}(\boldsymbol{T}(\boldsymbol{x}), \mathbf{F}(\boldsymbol{x})).$$

- Monte Carlo IGO gradient ascent update equations

$$\boldsymbol{\eta}^{t+1} = \boldsymbol{\eta}^{t} + \gamma \widehat{\mathrm{Cov}}_{P_{\boldsymbol{\eta}}}(\boldsymbol{T}(\boldsymbol{x}), \boldsymbol{T}(\boldsymbol{x}))^{-1} \widehat{\mathrm{Cov}}_{P_{\boldsymbol{\eta}}}(\boldsymbol{T}(\boldsymbol{x}), \mathbf{F}(\boldsymbol{x}))$$

  - $\widehat{\mathrm{Cov}}(\cdot)$ - Ledoit–Wolf covariance shrinkage estimator (Ledoit and Wolf, 2004).

- We try to map the EMvM parameters to the exponential family parametrization:

$$\boldsymbol{\eta} = \begin{bmatrix} \boldsymbol{\kappa} \odot c(\boldsymbol{\mu}) \\ \boldsymbol{\kappa} \odot s(\boldsymbol{\mu}) \\ \operatorname{vec}(\boldsymbol{D}) \end{bmatrix} \quad \boldsymbol{T} = \begin{bmatrix} c(\boldsymbol{\theta}) \\ s(\boldsymbol{\theta}) \\ \operatorname{vec}\left(\begin{bmatrix} c(\boldsymbol{\theta}) \\ s(\boldsymbol{\theta}) \end{bmatrix} \begin{bmatrix} c(\boldsymbol{\theta}) \\ s(\boldsymbol{\theta}) \end{bmatrix}^{\mathrm{T}} \right) \end{bmatrix}$$

  - $\operatorname{vec}(\cdot)$ - vectorization transformation
  - $\odot$ - Hadamard product
- The map $(\boldsymbol{\kappa}, \boldsymbol{\mu}, \boldsymbol{D}) \to \boldsymbol{\eta}$ is non-injective $(\kappa_i = 0,\ i = 1, \ldots, n)$
- We have new family of probability distributions on n-torus - eEMvM.
- $S_{\text{eEMvM}} = S_{\text{EMvM}},\ \boldsymbol{\kappa} \neq 0$

- We implement Gibbs sampling algorithm to generate samples from eEMvM.
- Gibbs sampler is used to generate samples from multivariate probability distribution when it's possible to generate samples from univariate conditional distribution.
  - Univariate eEMvM conditionals define an ergodic Markov chain with eEMvM as the stationary distribution.

Univariate eEMvM conditionals 1

$$f(\theta_k|\theta_1,\ldots,\theta_{k-1},\theta_{k+1},\ldots,\theta_n) \approx \exp\left\{\gamma_k^1\cos(\theta_k-\nu_k^1)+\gamma_k^2\cos(2(\theta_k-\nu_k^2))\right\}$$

$$\gamma_k^1 = \left\{\left(\eta_k^c+\sum_{\substack{i=1\\i\neq k}}^{n}(\eta_{ki}^{cc}+\eta_{ik}^{cc})\cos(\theta_i)+\sum_{\substack{i=1\\i\neq k}}^{n}(\eta_{ki}^{cs}+\eta_{ik}^{sc})\sin(\theta_i)\right)^2+\left(\eta_k^s+\sum_{\substack{i=1\\i\neq k}}^{n}(\eta_{ki}^{ss}+\eta_{ik}^{ss})\sin(\theta_i)+\sum_{\substack{i=1\\i\neq k}}^{n}(\eta_{ki}^{sc}+\eta_{ik}^{cs})\cos(\theta_i)\right)^2\right\}^{\frac{1}{2}}$$

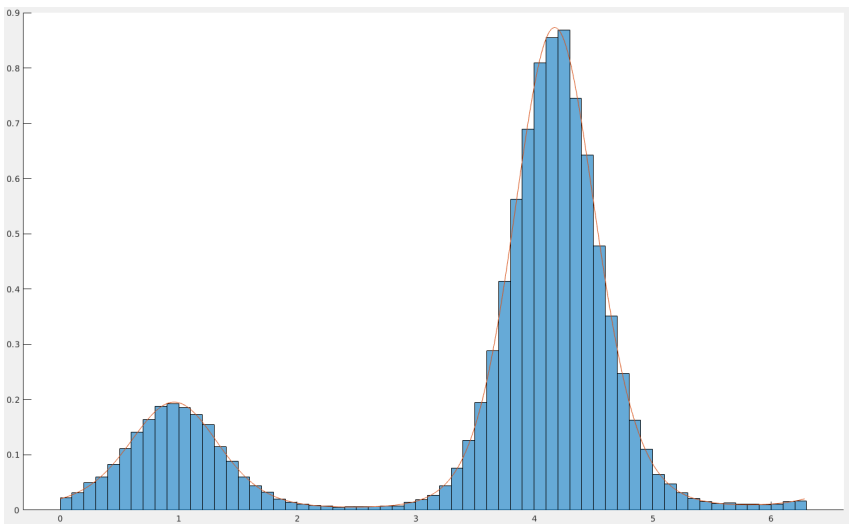
$$\nu_k^1 = \operatorname{atan2}\left(\frac{\eta_k^s+\sum_{\substack{i=1\\i\neq k}}^{n}(\eta_{ki}^{ss}+\eta_{ik}^{ss})\sin(\theta_i)+\sum_{\substack{i=1\\i\neq k}}^{n}(\eta_{ki}^{sc}+\eta_{ik}^{cs})\cos(\theta_i)}{\eta_k^c+\sum_{\substack{i=1\\i\neq k}}^{n}(\eta_{ki}^{cc}+\eta_{ik}^{cc})\cos(\theta_i)+\sum_{\substack{i=1\\i\neq k}}^{n}(\eta_{ki}^{cs}+\eta_{ik}^{sc})\sin(\theta_i)}\right)$$

$$\gamma_k^2 = \left\{\left(\eta_{kk}^{cc}-\eta_{kk}^{ss}\right)^2+(\eta_{kk}^{sc}+\eta_{kk}^{cs})^2\right\}^{\frac{1}{2}}$$

$$\nu_k^2 = \frac{1}{2}\operatorname{atan2}\left(\frac{\eta_{kk}^{sc}+\eta_{kk}^{cs}}{\eta_{kk}^{cc}-\eta_{kk}^{ss}}\right)$$

- η_k^c and η_k^s are the elements of $\boldsymbol{\eta}$ associated with $\kappa_k\cos(\mu_k)$ and $\kappa_k\sin(\mu_k)$
- η_{ij}^{cs}, η_{ij}^{sc}, η_{ij}^{cc}, η_{ij}^{ss} are the elements of $\boldsymbol{\eta}$ associated with the elements of the D^{cs}, D^{sc}, D^{cc}, D^{cc} submatrices of D.

- Univariate eEMvM conditionals are Generalized von Mises (GvM) distributed
  - Probability distribution on a unit circle
- To generate samples from $\mathrm{GvM}_2(\nu_k^1, \nu_k^2, \gamma_k^1, \gamma_k^2)$ we implement von Neumann rejection sampling algorithm (Gatto, 2008).

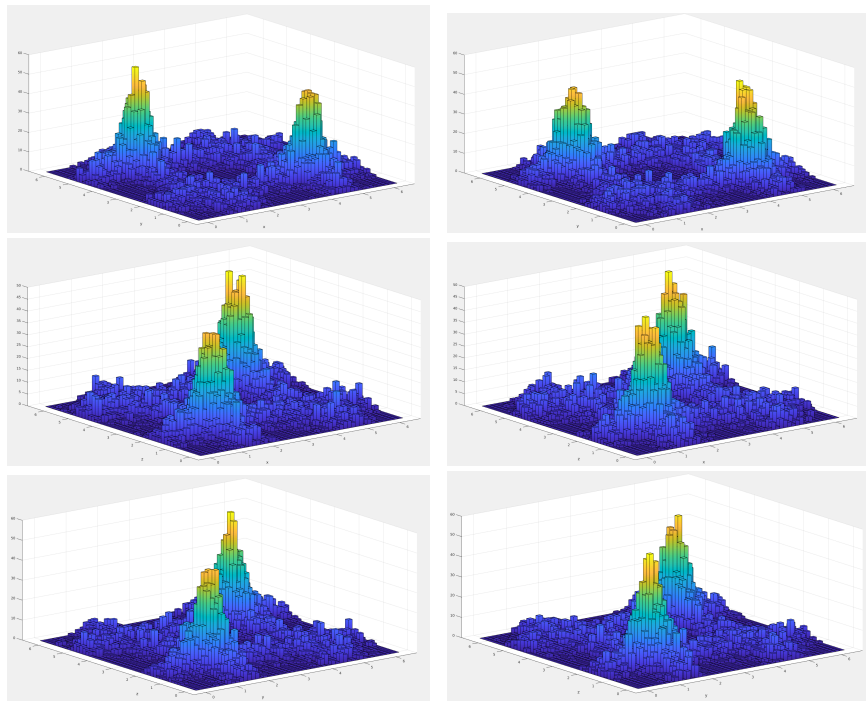


Histogram of 10000 $\mathrm{GvM}_2(4.5, 2, 0.8, 2)$ samples

hIGO

MvM rejection sampler and eEMvM Gibbs sampler comparison

- We implemented Mardia and Voss MvM rejection sampling algorithm (Mardia and Voss, 2012).
- Compared it to eEMvM Gibbs sampler for the trivariate case of MvM.



Histogram of 10000 MvM_3 samples. From top: projections along **z**, **y**, **x** axes. **Left**: MvM rejection sampling algorithm. **Right**: eEMvM Gibbs sampler.

Fitness function: Adaptive selection quantile

- (μ, λ)-ES truncation selection:

$$W_{\mathbf{E}}(\mathbf{x}) = \begin{cases} \frac{1}{\mu} & \text{if } P(\mathbf{E}(\mathbf{x}) \leq \mathbf{E}(\mathbf{X})) \geq 1 - \frac{1}{\mu} \\ 0 & \text{otherwise} \end{cases}$$

  - $1 - \frac{1}{\mu}$ - μ-quantile
- Angle α between two consecutive step directions $\triangle\boldsymbol{\eta}^t = \boldsymbol{\eta}^t - \boldsymbol{\eta}^{t-1}$ and $\triangle\boldsymbol{\eta}^{t-1} = \boldsymbol{\eta}^{t-1} - \boldsymbol{\eta}^{t-2}$ (Ollivier et. al. 2017):

$$\cos(\alpha^t) = \frac{(\triangle\boldsymbol{\eta}^t)^T F(\boldsymbol{\eta}^t) \triangle\boldsymbol{\eta}^{t-1}}{\sqrt{|(\triangle\boldsymbol{\eta}^t)^T F(\boldsymbol{\eta}^t) \triangle\boldsymbol{\eta}^t|} \sqrt{|(\triangle\boldsymbol{\eta}^{t-1})^T F(\boldsymbol{\eta}^t) \triangle\boldsymbol{\eta}^{t-1}|}}$$

- Adaptive selection quantile:

$$W_{\mathbf{E}}^t(\mathbf{x}) = \begin{cases} 1 & \text{if } P(\mathbf{E}(\mathbf{x}) > \mathbf{E}(\mathbf{X})) < \max\left[\frac{1}{q_0} \exp\left\{-\beta \sum_{i=3}^{t} \cos(\alpha^i)\right\}, \frac{1}{q_{\text{cap}}}\right] \\ 0 & \text{otherwise} \end{cases}$$

- Simulated annealing - $(1, 1)$-ES with time dependent selection pressure and constant mutation rate.

- To get a better control of the hIGO convergence we express the hIGO gradients in the original $\boldsymbol{\mu},\ \boldsymbol{\kappa},\ \mathrm{D}$ by solving:

$$\begin{bmatrix} \frac{d\eta_i^c}{dt_s} \\ \frac{d\eta_i^s}{dt} \end{bmatrix} = \begin{bmatrix} -\kappa_i \sin(\mu_i) & \cos(\mu_i) \\ \kappa_i \cos(\mu_i) & \sin(\mu_i) \end{bmatrix} \begin{bmatrix} \frac{d\mu_i}{dt} \\ \frac{d\kappa_i}{dt} \end{bmatrix},\ i = 1 \ldots n,\ \boldsymbol{\kappa} > 0.$$

- $\boldsymbol{\mu},\ \boldsymbol{\kappa}$ parametrization hIGO gradients:

$$\begin{bmatrix} \tilde{\nabla}_{\mu_i} g \\ \tilde{\nabla}_{\kappa_i} g \end{bmatrix} = \begin{bmatrix} \frac{d\mu_i}{dt} \\ \frac{d\kappa_i}{dt} \end{bmatrix} = \begin{bmatrix} -\frac{\sin(\mu_i)}{\kappa_i} & \frac{\cos(\mu_i)}{\kappa_i} \\ \cos(\mu_i) & \sin(\mu_i) \end{bmatrix} \begin{bmatrix} \frac{d\eta_i^c}{dt_s} \\ \frac{d\eta_i^s}{dt} \end{bmatrix},\ i = 1 \ldots n,\ \boldsymbol{\kappa} > 0$$

$$\tilde{\nabla}_{\mathrm{D}} g = \frac{d\boldsymbol{\eta}^{\mathrm{D}}}{dt}$$

- hIGO update equations with learning rates $\gamma_{\boldsymbol{\mu}},\ \gamma_{\boldsymbol{\kappa}},\ \gamma_{\mathrm{D}}$:

$$\boldsymbol{\mu}^{t+1} = \boldsymbol{\mu}^t + \gamma_{\boldsymbol{\mu}} \tilde{\nabla}_{\boldsymbol{\mu}} g$$
$$\boldsymbol{\kappa}^{t+1} = \boldsymbol{\kappa}^t + \gamma_{\boldsymbol{\kappa}} \tilde{\nabla}_{\boldsymbol{\kappa}} g$$
$$\mathrm{D}^{t+1} = \mathrm{D}^t + \gamma_{\mathrm{D}} \tilde{\nabla}_{\mathrm{D}} g$$

- Modification of Silva–Almeida adaptive learning rates (Silva and Almeida, 1990)

$$\gamma_{\mu_i}^t = \begin{cases} \min\left\{c_{up}\gamma_{\mu_i}^{t-1},\ \gamma_{\mu_i}^0\right\} & \text{if} \quad \operatorname{sgn}\left(\tilde{\nabla}_{\mu_i} g^t\right) = \operatorname{sgn}\left(\tilde{\nabla}_{\mu_i} g^{t-1}\right) \\ c_{down}\gamma_{\mu_i}^{t-1} & \text{if} \quad \operatorname{sgn}\left(\tilde{\nabla}_{\mu_i} g^t\right) \neq \operatorname{sgn}\left(\tilde{\nabla}_{\mu_i} g^{t-1}\right) \end{cases}$$

$$\gamma_{\kappa_i}^t = \begin{cases} \min\left\{c_{up}\gamma_{\kappa_i}^{t-1},\ \gamma_{\kappa_i}^0\right\} & \text{if} \quad \operatorname{sgn}\left(\tilde{\nabla}_{\kappa_i} g^t\right) = \operatorname{sgn}\left(\tilde{\nabla}_{\kappa_i} g^{t-1}\right) \\ c_{down}\gamma_{\kappa_i}^{t-1} & \text{if} \quad \operatorname{sgn}\left(\tilde{\nabla}_{\kappa_i} g^t\right) \neq \operatorname{sgn}\left(\tilde{\nabla}_{\kappa_i} g^{t-1}\right) \end{cases} \qquad i, j = 1 \ldots n$$

$$\gamma_{d_{ij}}^t = \begin{cases} \min\left\{c_{up}\gamma_{d_{ij}}^{t-1},\ \gamma_{d_{ij}}^0\right\} & \text{if} \quad \operatorname{sgn}\left(\tilde{\nabla}_{d_{ij}} g^t\right) = \operatorname{sgn}\left(\tilde{\nabla}_{d_{ij}} g^{t-1}\right) \\ c_{down}\gamma_{d_{ij}}^{t-1} & \text{if} \quad \operatorname{sgn}\left(\tilde{\nabla}_{d_{ij}} g^t\right) \neq \operatorname{sgn}\left(\tilde{\nabla}_{d_{ij}} g^{t-1}\right) \end{cases}$$

  - sgn(·) - signum function
  - $c_{up} > 1,\ c_{down} < 1$ - are real positive constants.
- Momentum constants $\alpha_{\boldsymbol{\mu}},\ \alpha_{\boldsymbol{\kappa}},\ \alpha_{\mathrm{D}}$:

$$\begin{aligned} m_{\mu_i}^t &= \tilde{\nabla}_{\mu_i} g^t + \alpha_{\boldsymbol{\mu}} m_{\mu_i}^{t-1} \\ m_{\kappa_i}^t &= \tilde{\nabla}_{\kappa_i} g^t + \alpha_{\boldsymbol{\kappa}} m_{\kappa_i}^{t-1} \\ m_{d_{ij}}^t &= \tilde{\nabla}_{d_{ij}} g^t + \alpha_{\mathrm{D}} m_{d_{ij}}^{t-1} \end{aligned} \quad , \ i, j = 1 \ldots n$$

- The final form of the hIGO update equations:

$$\begin{aligned} \mu_i^{t+1} &= \mu_i^t + \gamma_{\mu_i}^t m_{\mu_i}^t \\ \kappa_i^{t+1} &= \kappa_i^t + \gamma_{\kappa_i}^t m_{\kappa_i}^t \\ d_{ij}^{t+1} &= d_{ij}^t + \gamma_{d_{ij}}^t m_{d_{ij}}^t \end{aligned} \quad , \ i, j = 1 \ldots n.$$

- EMvM is defined on the quotient space space:

$$\prod_{i=1}^{n} [0; 2\pi] \Big/ \sim$$

  - $\sim$: $(x_1, x_2, \ldots, x_n) \sim (x_1 + 2\pi, x_2, \ldots, x_n) \sim (x_1, x_2 + 2\pi, \ldots, x_n) \sim (x_1, x_2, \ldots, x_n + 2\pi)$
- In order to evaluate the objective function **E** we define a map from $[0; 2\pi]^n$ to the optimization boundary constraints.
  - $y_i : [0; 2\pi) \to [l_i; u_i)$

$$y_i = \begin{cases} l_i + \frac{\theta_i}{2\pi}(u_i - l_i) & \text{if} \quad \theta_i \text{ is periodic} & \theta_i \in [0; 2\pi) \\ l_i + \frac{\theta_i}{\pi}(u_i - l_i) & \text{if} \quad \theta_i \text{ is non-periodic} & \theta_i \in [0; \pi) \\ 2u_i - l_i - \frac{\theta_i}{\pi}(u_i - l_i) & \text{if} \quad \theta_i \text{ is non-periodic} & \theta_i \in [\pi; 2\pi) \end{cases}$$

$$i = 1, \ldots, n$$

  - u_i, l_i - ith variable's upper and lower bound.
  - θ_i non-periodic - y_i is non-injective

- We use penalty function based on feasibility (Deb, 2000):
  1. Any feasible solution is better than any unfeasible solution.
  2. Between two feasible solutions, the one with better objective function is preferred.
  3. Between two infeasible solutions, the one with lower constraint violation is preferred.
- Penalty function for the constraints $h_j(\mathrm{x}) \leq 0,\ j = 1, \ldots, K$:

$$\mathbf{P}(\mathrm{x}) = \begin{cases} \mathbf{E}(\mathrm{x}) & \text{if} \quad h_j(\mathrm{x}) \leq 0\ \forall\ j = 1, \ldots, K \\ \mathbf{E}_{\max} + \sum_{i \in I} h_i(\mathrm{x}) & \text{otherwise} \quad I = \{j : h_j(\mathrm{x}) > 0,\ j = 1, \ldots, K\} \end{cases}$$

  - $\mathbf{E}_{\max} = \max\{\mathbf{E}(\mathrm{x}) \mid \mathrm{x} :\ h_j(\mathrm{x}) \leq 0\ \forall\ j = 1, \ldots, K\}$
- No additional information about the constraints besides constraint violation evaluation is needed.
- Objective function is not evaluated for unfeasible solutions.

hIGO

Parallelizing computations

- We parallelize eEMvM sampling, objective function evaluations and constraint violation evaluations:
- Every worker:
  1. Generates $\frac{\lambda}{T}$ candidate solutions $x \sim \text{eEMvM}(\boldsymbol{\eta})$.
  2. Evaluates constraints violation for all solutions.
  3. Evaluates objective function for solutions that meet the constraint requirements.
  4. Passes the results for processing.

  - λ - number of samples used in hIGO.
  - T - number of workers available.
- Bottleneck is that the algorithm needs to wait until all workers are done with computations in order to proceed.

Proof of concept: densest plane group packings

Space group packing

- **Space group packing**:

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

$$\text{int}\,(g_i K) \cap \text{int}\,(g_j K) = \emptyset$$

$$\forall\ g_i,\ g_j \in G$$

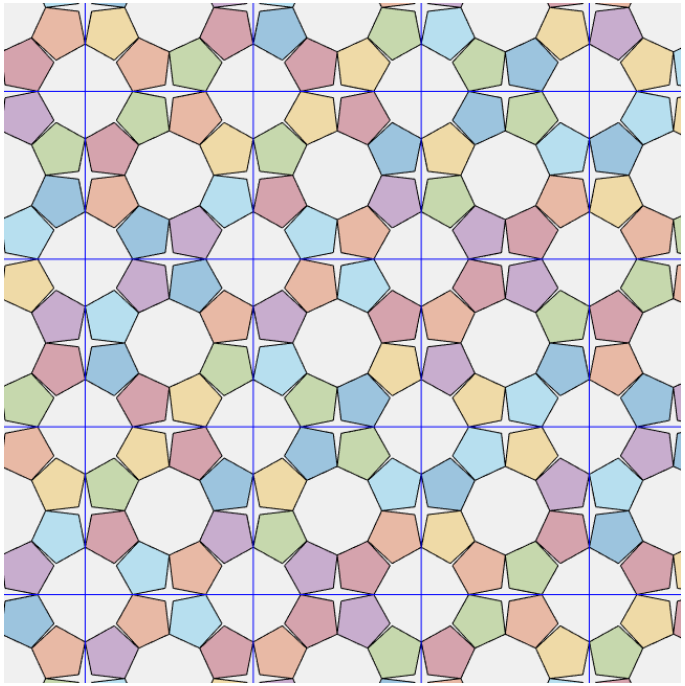
$$g_i \neq g_j$$

  - G - a space group
  - $K \subset \mathbb{R}^n$ - closed
  - $\mathcal{K}_G$ - G-set

- Periodic packing: G is a lattice group.
- K is the *fundamental region* if $\mathcal{K}_G = \mathbb{R}^n$ (tiling) and K is the subset of *unit cell*.

Proof of concept: densest plane group packings

p4mm packing of pentagons



- G - p4mm
- K - regular pentagon

Proof of concept: densest plane group packings

Space group problem

- **Space group packing density**:

$$\rho\left(\mathcal{K}_G\right) = \frac{N\text{vol}(K)}{\text{vol}(\overline{\Lambda})}$$

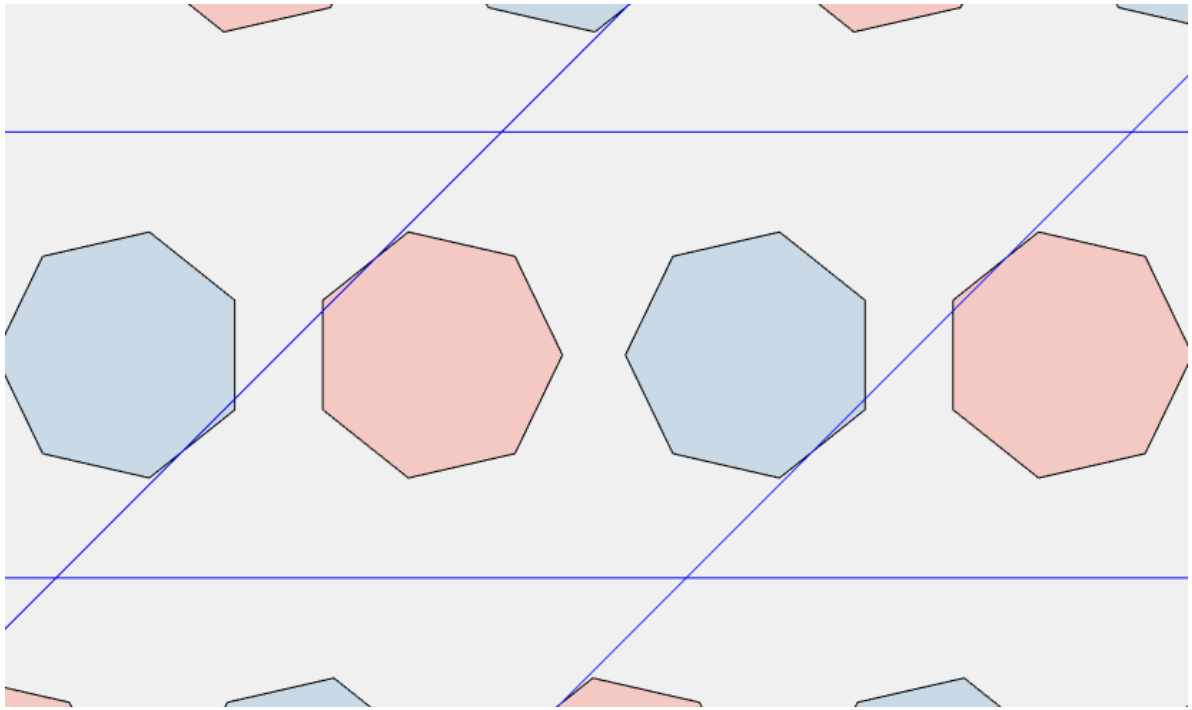
  - N - number of symmetry operations in G
  - $\overline{\Lambda}$ - unit cell
- **Space group packing problem**:

$$\mathcal{K}_{\text{max}} = \underset{\mathcal{K}_G}{\text{argmin}}\, \text{vol}(\overline{\Lambda})$$

- non-linear bounded constrained optimization problem
  - $\text{vol}(\overline{\Lambda})$ - determinant of the unit cell vectors
  - bounds given by the space group's fundamental region
  - fundamental region linear constraints
  - non-overlap constraints
    - minimum euclidean distance between "all" objects in the packing
    - continuous but not differentiable function.

Proof of concept: densest plane group packings

Packing regular heptagons in p2



- p2 is also called a *double lattice*

Packing regular heptagons in p2

Setting the hIGO parameters

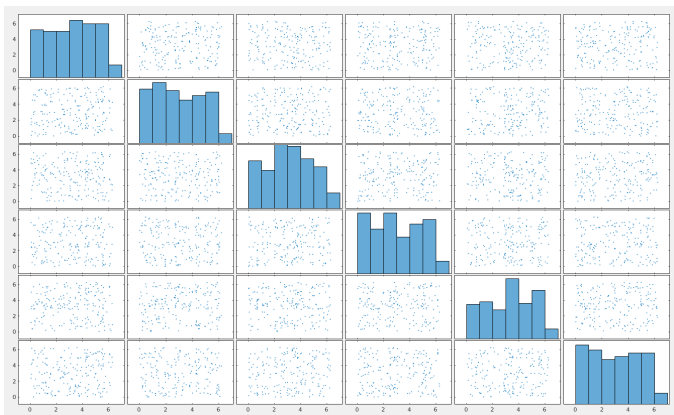
$\gamma^0_{\boldsymbol{\mu}}$	0.1
$\gamma^0_{\boldsymbol{\kappa}}$	0.6
γ^0_{D}	0.075
c_{up}	1.01
c_{down}	0.99
$\alpha_{\boldsymbol{\mu}}$	0.5
$\alpha_{\boldsymbol{\kappa}}$	0.6
α_{D}	0.4
q_0	4
q_{cap}	10
β	0.001
Number of samples	200
Number of iterations	2000

- Has to be determined experimentally.
- Seems robust:
  - Same parameters worked on all tested plane groups and shapes later on.
- Reduce the number of parameters or metalearning???

Packing regular heptagons in p2

Initial setup

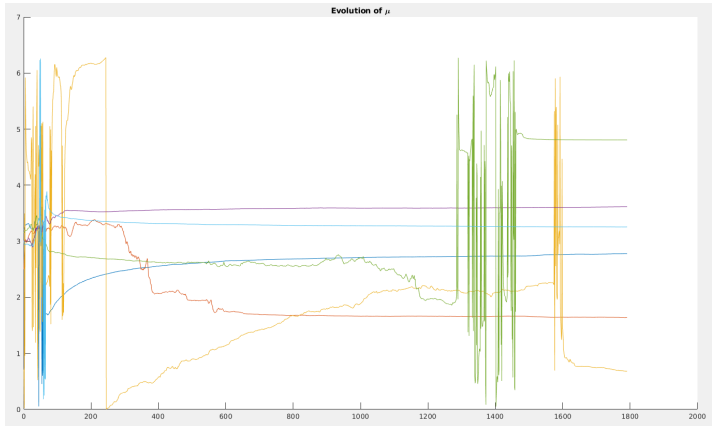
- 6 variables to optimize:
  - 3 periodic:
    1. $x \in [0; \frac{1}{2}]$ coordinate of centre of heptagon in the fundamental region
    2. $y \in [0; 1]$ coordinate of centre of heptagon in the fundamental region
    3. $\omega \in [0; 2\pi]$ - rotation angle of heptagon in the fundamental region
  - 3 aperiodic:
    1. $a \in [0; 4]$ - size of size of unit cell vector **a**
    2. $b \in [0; 4]$ - size of size of unit cell vector **b**
    3. $\alpha \in [0; \frac{\pi}{2}]$ - angle between **a** and **b**
- all the eEMvM$_6$ parameters are set to 0 - *uniform distribution* on 6-torus.
  - maximum entropy distribution on 6-torus with no specified constraints



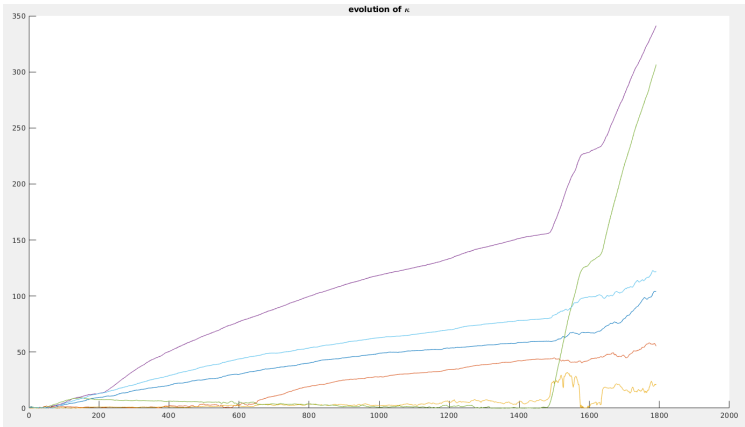
Plot matrix of the initial 200 eEMvM$_6$ samples

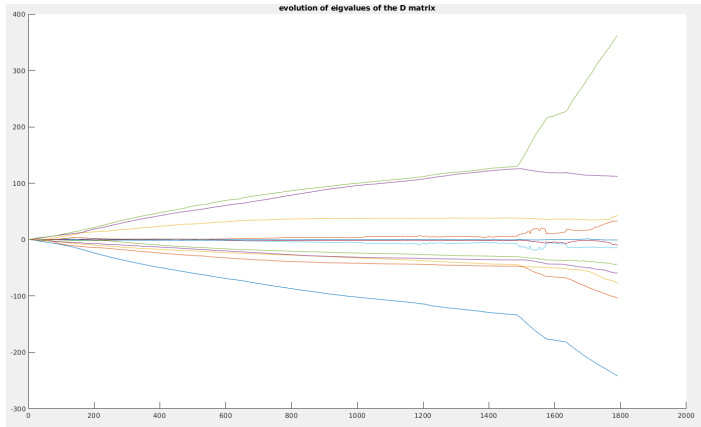
Packing regular heptagons in p2

hIGO EMvM$_6$ trajectories



μ trajectories

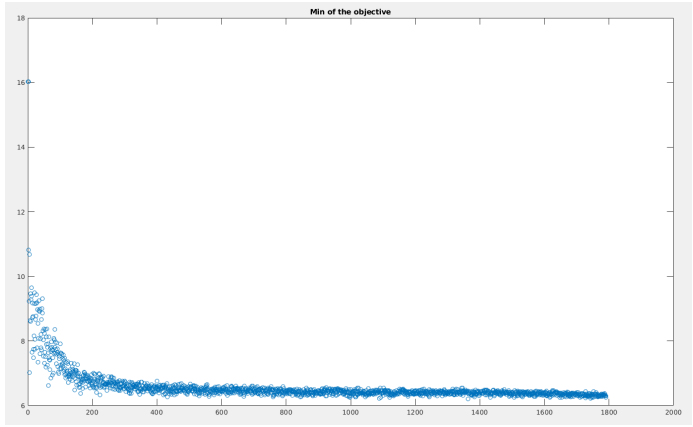




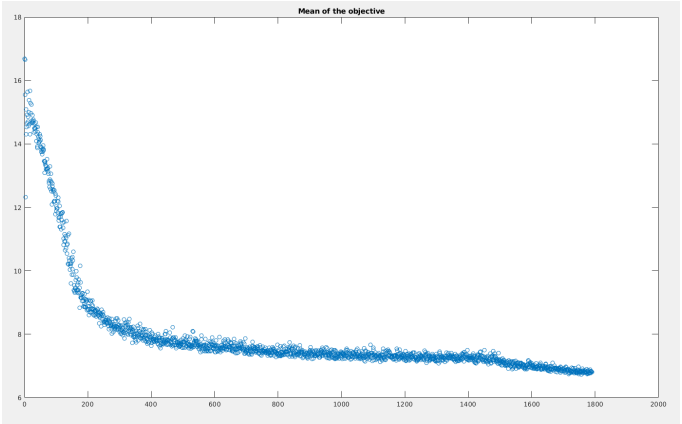
Left: κ trajectories. **Right:** D eigenvalues trajectories

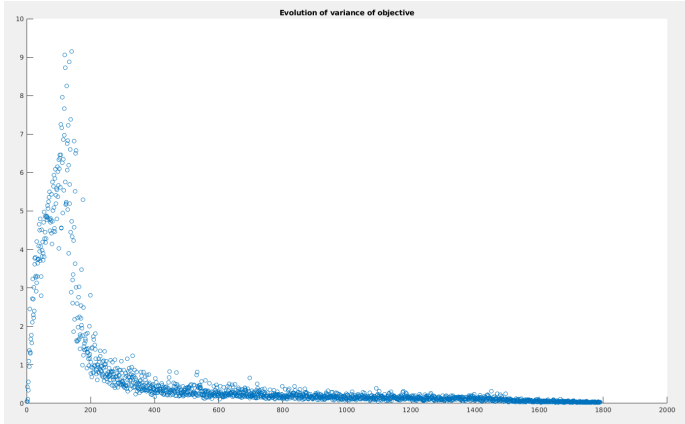
Packing regular heptagons in p2

Evolution of the objective's summary statistics



Evolution of minimum

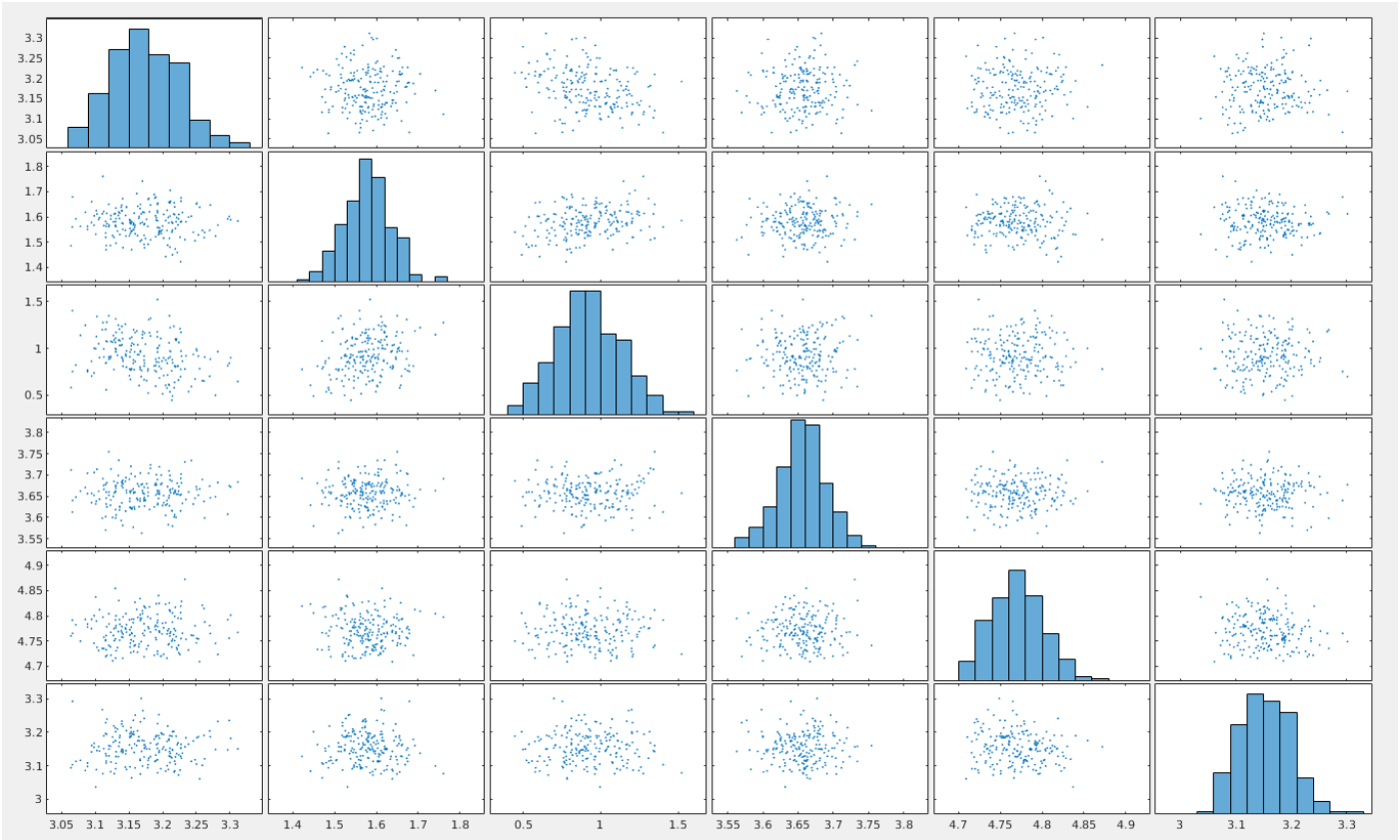




Left: Evolution of average. **Right:** Evolution of variance

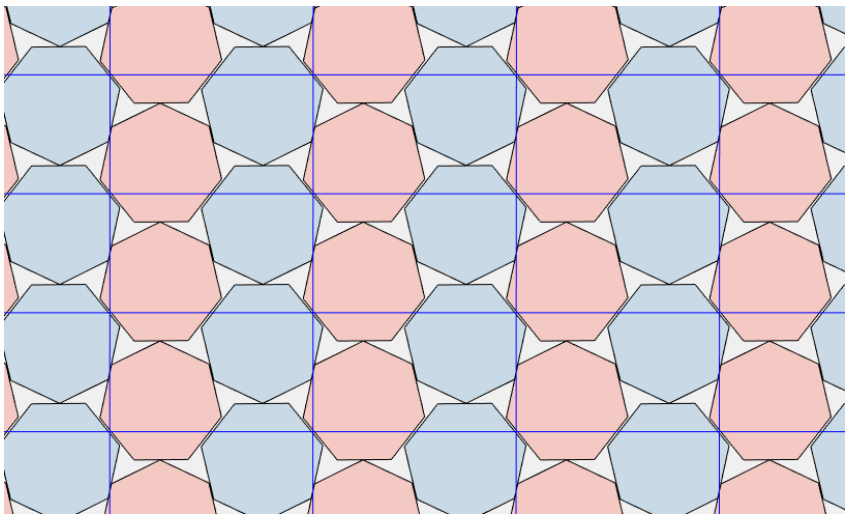
Packing regular heptagons in P2

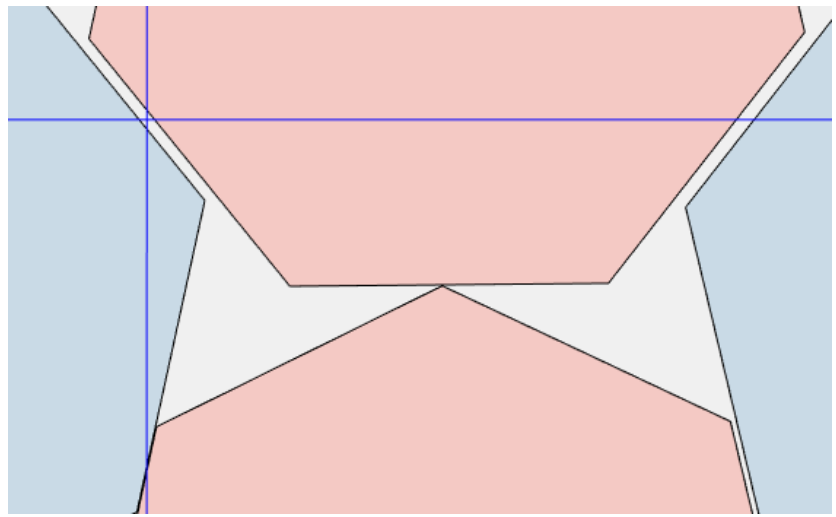
End probability distribution



Packing regular heptagons in P2

Best solution found during the run





$\rho = 0.88194645326253$

Solution refining

hIGO microscope (hIGOmicro)

- $\mathbf{x}^{\text{best}}$ solution found so far.
- We create an ϵ^t neighbourhood of $\mathbf{x}^{\text{best}}$:

$$\epsilon_i^t = \left(\frac{1}{c_\epsilon}\right)^t (u_i - l_i), \quad i = 1, \ldots, n$$

  - $c_\epsilon > 1$
  - $\mathbf{u}$ - upper bounds
  - $\mathbf{l}$ - lower bounds
- We crate new upper $\mathbf{u}_\epsilon$ and lower bounds $\mathbf{l}_\epsilon$:

$$\begin{aligned} u_i^\epsilon &= \min\left(x^{\text{best}} + \epsilon_i^t, u_i\right) \\ l_i^\epsilon &= \max\left(x^{\text{best}} - \epsilon_i^t, l_i\right) \end{aligned}, \quad i = 1, \ldots, n$$

- Run hIGO with the new bounds $\mathbf{u}_\epsilon$, $\mathbf{l}_\epsilon$
  - All optimization variables are now aperiodic.
  - We iterate this process until desired precision is achieved ($t = 1, 2, \ldots$).

Proof of concept: densest plane group packings

Densest double lattice packing of regular heptagons

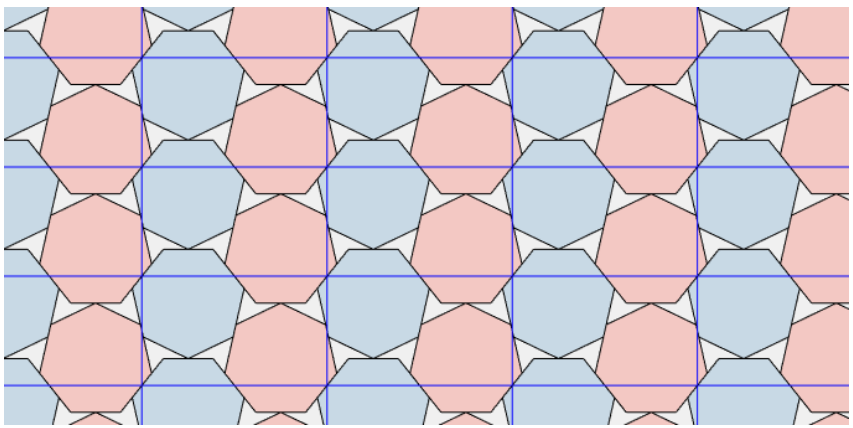
- Densest double lattice of regular heptagons (Kuperberg and Kuperberg, 1990):

$$\rho_{\text{opt}} = \frac{2}{97}\left(-111 + 492\cos\left(\frac{\pi}{7}\right) - 356\cos^2\left(\frac{\pi}{7}\right)\right) =$$

$$= 0.89269068612650\ldots$$

	ρ	$\rho_{\text{opt}} - \rho$
hIGOmicro	0.892690680150324	$5.9762\ 10^{-9}$

- Optimal solution lies on the boundary: $\alpha = \frac{\pi}{2}$.
- Well defined problem?
  - At least 2 global minima.



Proof of concept: densest plane group packings

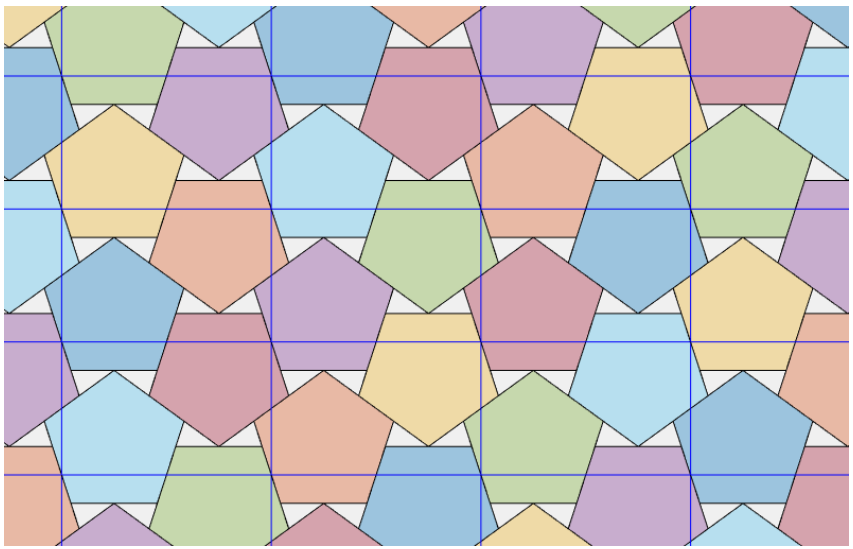
Densest double lattice packing of regular pentagons

- Densest packing of regular pentagons (Hales and Kusner, 2016):

$$\rho_{\text{opt}} = \frac{5-\sqrt{5}}{3} = 0.9213106741667\ldots$$

	ρ	$\rho_{\text{opt}} - \rho$
ASC	0.921301	$1.0\ 10^{-5}$
hIGOmicro	0.921310671743827	$2.4229\ 10^{-9}$

Table: Atkinson S., Jiao, Y., Torquato S., Maximally dense packings of two-dimensional convex and concave noncircular particles, Phys. Rev. E, 2012.



Proof of concept: densest plane group packings

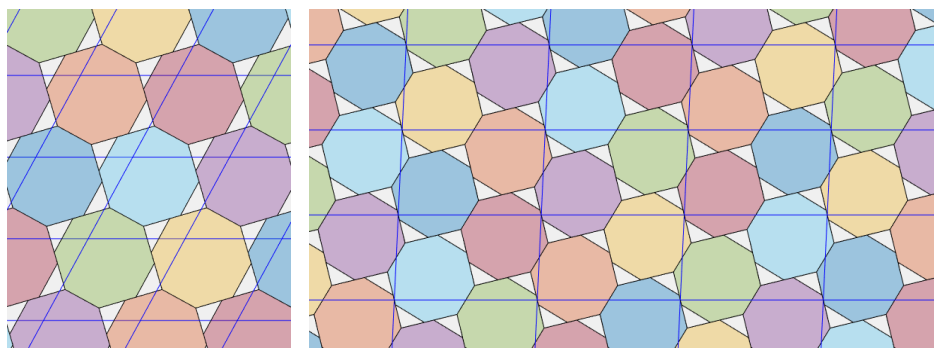
Densest packing of regular octagons

- Densest lattice packing of regular octagons:

$$\rho_{\text{opt}} = \frac{4+4\sqrt{2}}{5+4\sqrt{2}} = 0.90616367864394\ldots$$

	ρ	$\rho_{\text{opt}} - \rho$
ASC	0.906144	$2.0\ 10^{-5}$
hIGOmicro p1	0.906163653094707	$2.5549\ 10^{-8}$
hIGOmicro p2	0.906163111422202	$5.6722\ 10^{-7}$

Table: Atkinson S., Jiao, Y., Torquato S., Maximally dense packings of two-dimensional convex and concave noncircular particles, Phys. Rev. E, 2012.

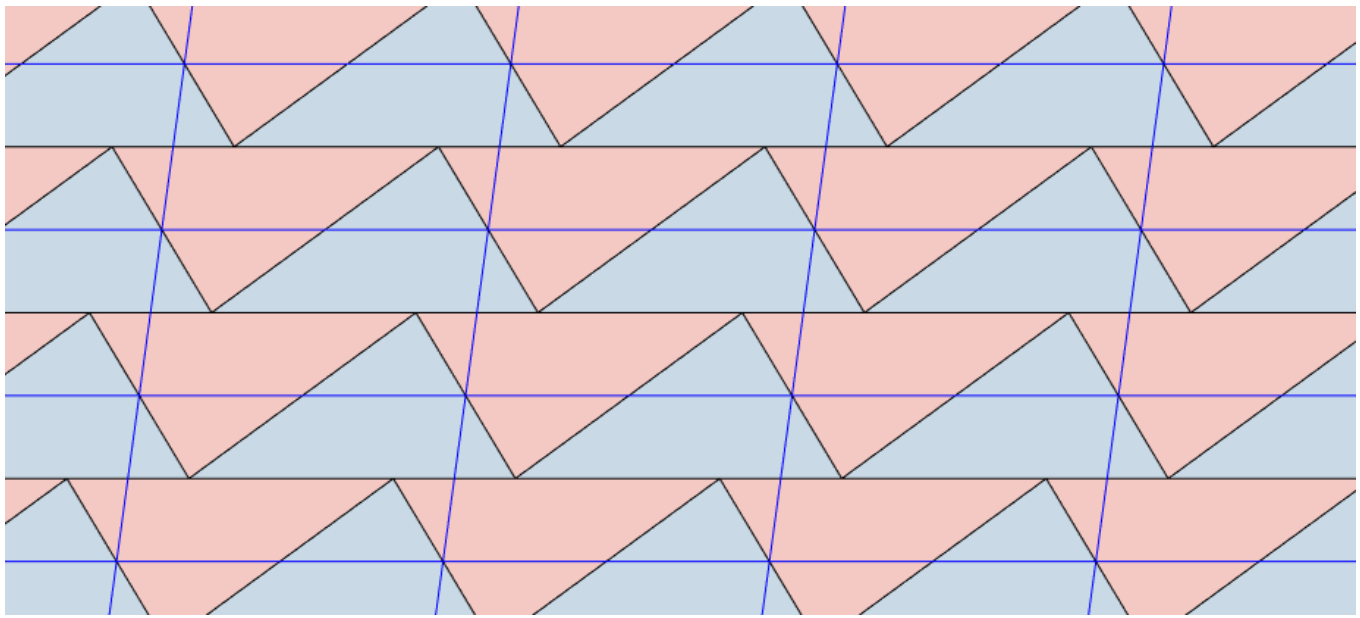


Left: p1 packing. **Right:** p2 packing

Proof of concept: densest plane group packings

Densest double lattice packing of scalene triangles

	ρ	$\rho_{\text{opt}} - \rho$
hIGOmicro	0.999999961042656	$3.8957\ 10^{-8}$



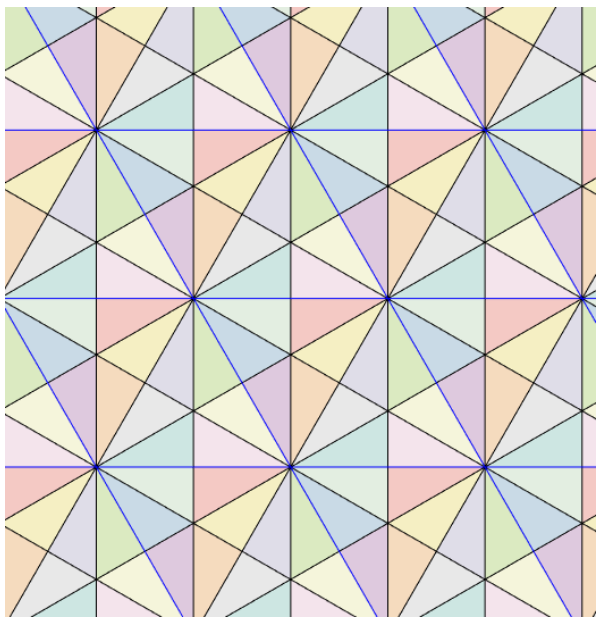
Proof of concept: densest plane group packings

Densest p6mm packing of 30-60-90 triangle

- Kisrhombille tiling

	ρ	$\rho_{\text{opt}} - \rho$
hIGOmicro	0.999999998714671	$1.2853\ 10^{-9}$

- 30-60-90 triangle defines the fundamental region for p6mm.



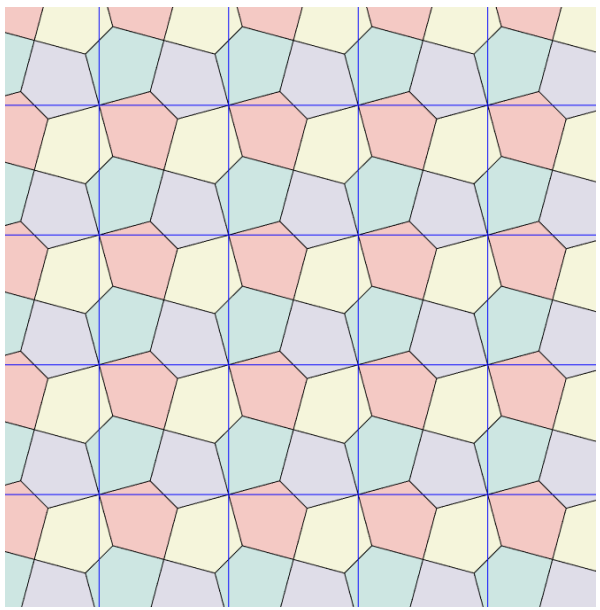
Proof of concept: densest plane group packings

Densest p4 packing of irregular pentagons

- Cairo tiling

	ρ	$\rho_{\text{opt}} - \rho$
hIGOmicro	0.999999963353343	$3.6647\ 10^{-8}$

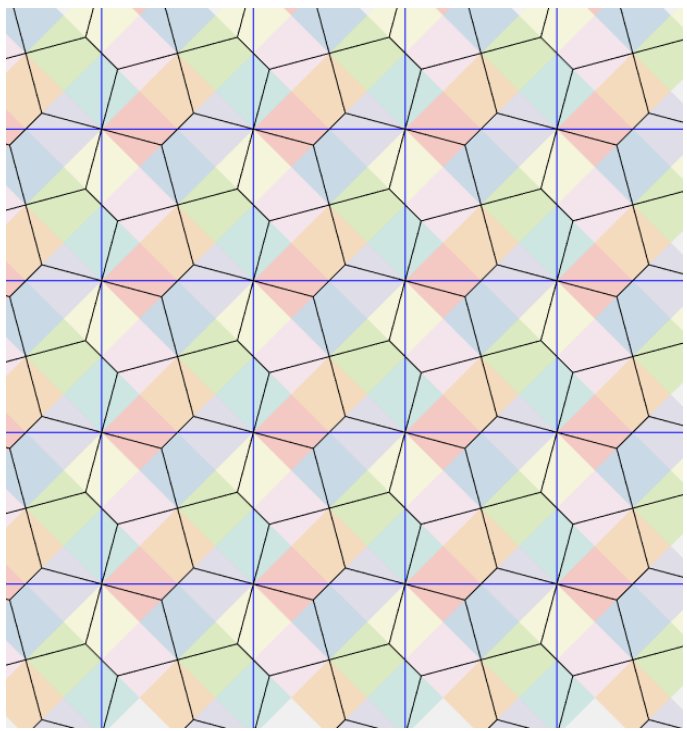
- but Cairo tiling has p4g symmetry.



Densest p4g packing of squares

- Cairo tiling

	ρ	$\rho_{\text{opt}} - \rho$
hIGOmicro	0.999999987587046	$1.2413\ 10^{-8}$



THANK YOU

Appendix

Crystal

Crystal

1. A crystal is a solid that has long-range positional order.
2. Long-range positional order can be inferred from the existence of Bragg peaks in the Fourier spectrum of the solid.

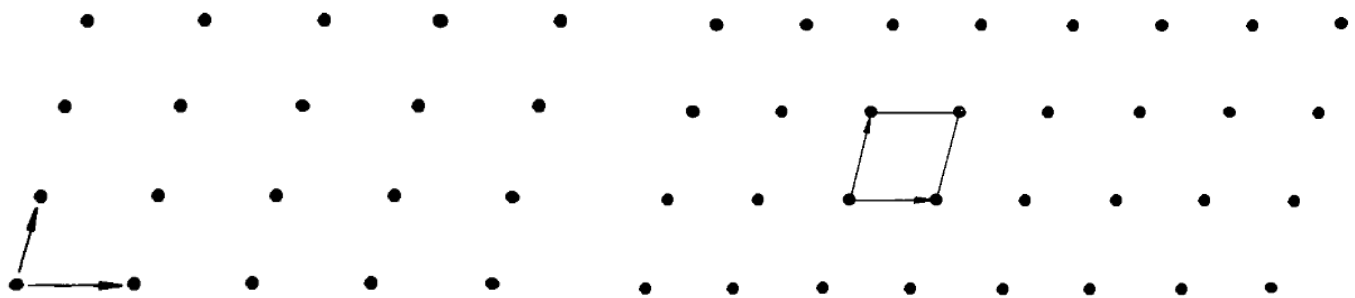
Periodic crystal

- A structure built of one or a few different kinds of discrete units, arranged in more or less modular fashion.

Appendix

Lattice and unit cell

- The module is a space filling polyhedron that generates a lattice.
- The set $\Lambda = \{u_1 \mathrm{a}_1 + u_2 \mathrm{a}_2 + \cdots + u_n \mathrm{a}_\mathrm{n} \mid u_i \in \mathbb{Z}\} \subset \mathbb{R}^n$ where $\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_\mathrm{n}$ are n linearly independent vectors is called a **lattice**.
- $\overline{\Lambda} = \{v_1 \mathrm{a}_1 + v_2 \mathrm{a}_2 + \cdots + v_n \mathrm{a}_\mathrm{n} \mid v_i \in [0, 1]\}$ is called a **unit cell**.
- Lattice points are orbits of a translational symmetry group (lattice group).

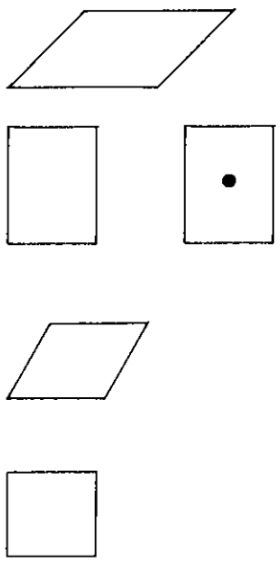


(a) A plane lattice and (b) a corresponding unit cell.

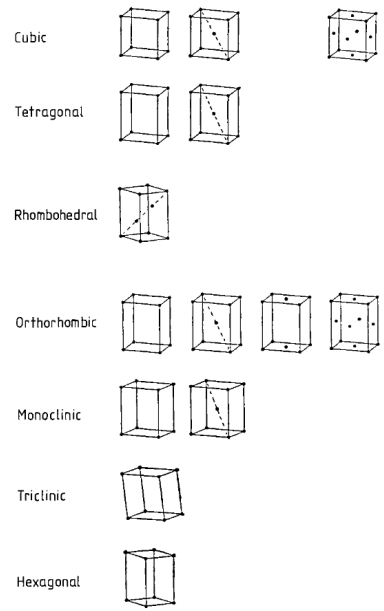
Appendix

Crystal system

- Crystal systems are assigned to lattices according to their stabilizer groups.



Four crystal systems in a plane.

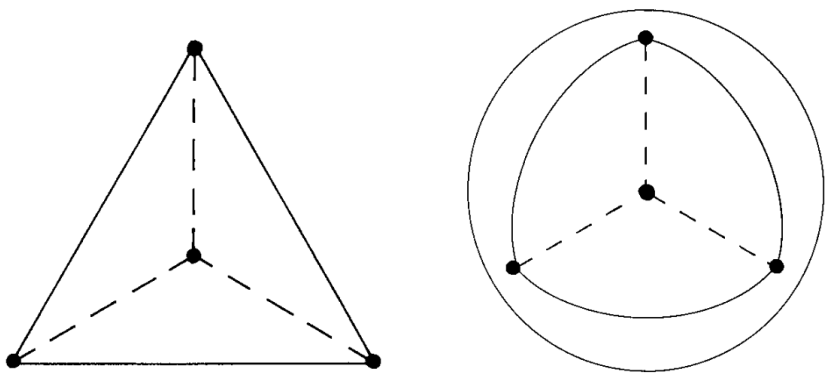


7 crystal systems in 3D. Classes of conjugate finite subgroups of $O(3)$.

Appendix

Point group

- **Point group** is a finite subgroup of the group of all symmetries of a sphere.
- The orthogonal group: $O(n) = \{M \in GL(n) : M^T = M^{-1}\}$.

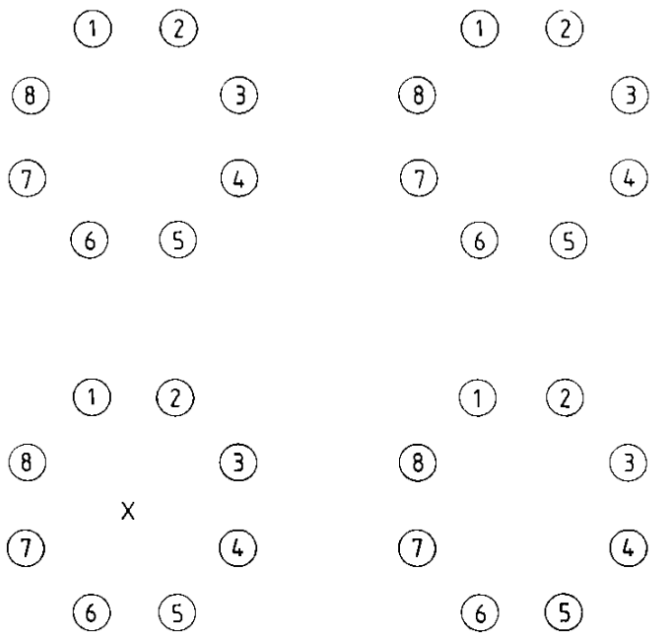


The points of an orbit of a group of symmetries of a tetrahedron and their projection on a sphere.

Appendix

Space group

- **Space group** is a lattice group 'extended' by a point group.
- Every orbit of a space group is a union of finite number of congruent lattices.
- Each of the lattices is an orbit for the translation group.
- The symmetry operations associated with the point group permutes the lattices.



An orbit of p4m; each of the eight lattices has been assigned a number.

Appendix

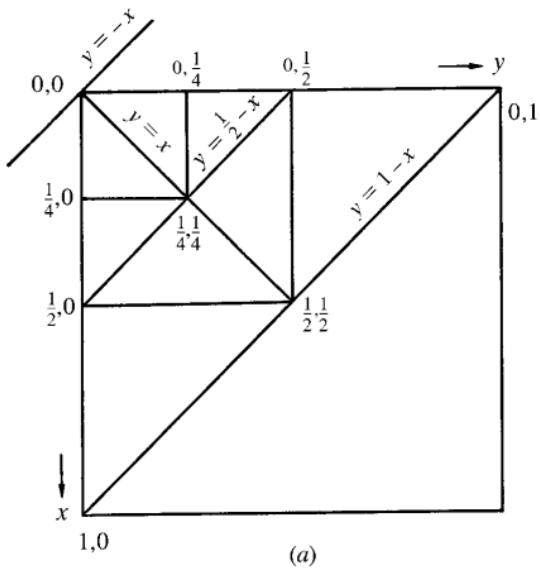
Fundamental region

$$S = \bigcup_{g \in G} gF$$

$$\text{int}\,(g_i F) \cap \text{int}\,(g_j F) = \emptyset$$

$$\forall\ g_i,\ g_j \in G;\quad g_i \neq g_j$$

- G - a group
- S - topological G-set
- $F \subset S$ - a group



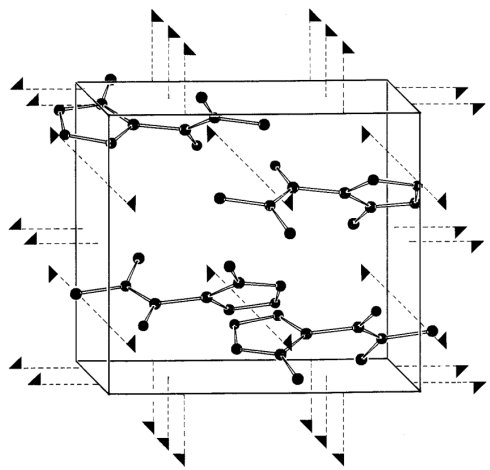
Boundary planes of asymmetric units occurring in the space group tables.

Hahn T. Ed., International Tables for Crystallography Volume A: Space-Group Symmetry, 2005.

Appendix

Molecular crystal model

- Space group
- Position and rotation of a molecule in the fundamental region
- Size of the unit cell



C. Giacovazzo et al., Fundamentals of Crystallography, 2011.