Integrating Multilevel Graph Partitioning with Hierarchical Set Oriented Methods for the Analysis of Dynamical Systems

Kathrin Padberg and Robert Preis

Paderborn Institute for Scientific Computation

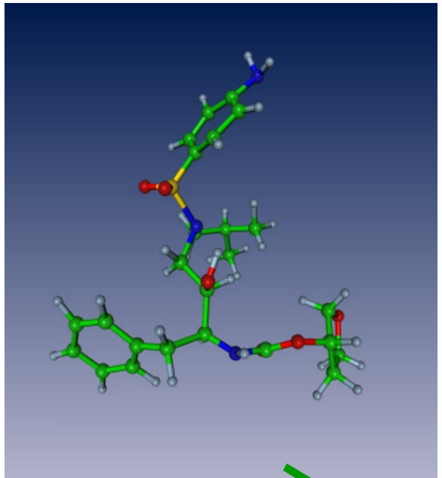
University of Paderborn, Germany

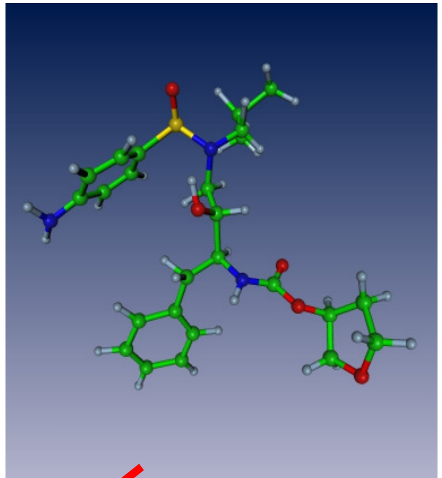
Contents

- Motivation: Biomolecular Systems
- Problem Formulation
  1. Continuous System
  2. Boxes / Discretized System
  3. Graph Partitioning
- Multilevel Structures
  - Hierarchical Set Oriented Approach
  - Multilevel Graph Partitioning Paradigm
  - Integration of these Multilevel Systems
- Experiments
- DONEs, TODOs and Outlook

Motivation: Biomolecular Systems

Data by Ch. Schütte (TU Berlin)



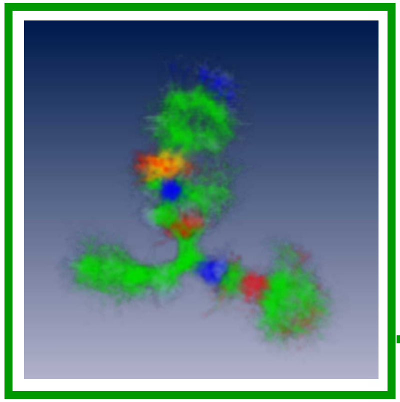


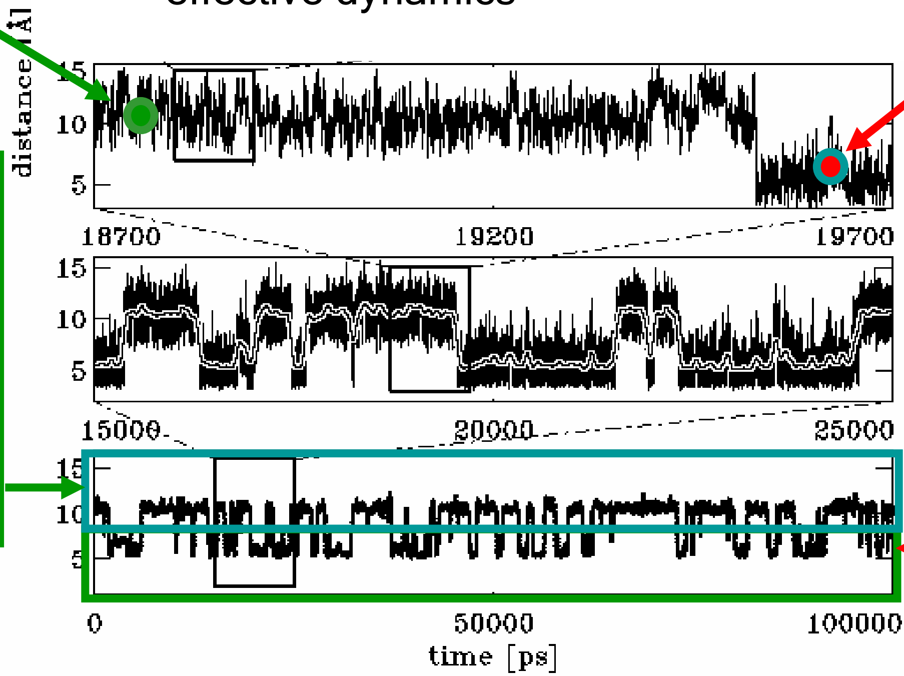
multiscale nature

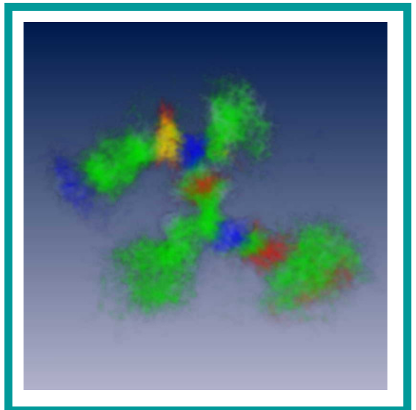
+ fast scale (oscillations around almost invariant global states)
+ slow scale (conformational dynamics)

aim: identification of

+ conformations/almost invariant sets & transition probabilities
+ essential degrees of freedom & effective dynamics







A Dynamical System

A map $f\colon X \to X$ on a compact subset $X \subset R^n$ defines a discrete dynamical system

$$x_{k+1} = f(x_k), \quad k = 0,1,2,\ldots$$

The data is usually given as a time series.

Goal:
Divide the state space into almost invariant sets in which trajectories stay for a long period of time before they enter other parts of the state space.

Problem 1: Continuous

Let m be the Lebesgue measure and for a set $S \subset X$ let

$$\rho(S) = \frac{m(f^{-1}(S) \cap S)}{m(S)}$$

be the transition probability that the set maps into itself.

Problem 1 (Continuous):

For some fixed $p \in N$ find a collection of pairwise disjoint sets $S = \{S_1, \ldots, S_p\}$ with $\bigcup_{1 \le i \le p} S_i = X$ and $m(S_i) > 0$ such that

$$\boxed{\rho(S) := \frac{1}{p} \sum_{i=1}^{p} \rho(S_i) \to \max}$$

Problem 2: Boxes

Discretize the state space by a box covering $B = \{B_1, \ldots, B_n\}$ such that

$$X = \bigcup_{i=1}^{b} B_i \quad \text{and} \quad m(B_i \cap B_j) = 0 \quad \text{for } i \neq j$$

The result is a transition matrix between the boxes with

$$P_B = (p_{ij}) \text{ where } \quad p_{ij} = \frac{m(f^{-1}(B_i) \cap B_j)}{m(B_j)}, \quad 1 \leq i,j \leq b$$

We obtain the natural invariant measure μ as the eigenvector to the eigenvalue 1 of P_B.

<u>Problem 2 (Boxes):</u>

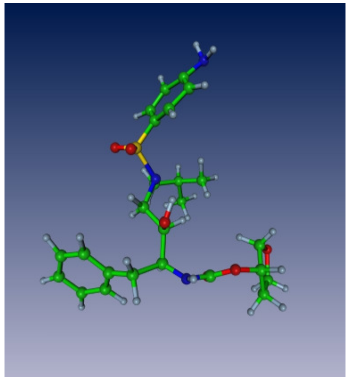
For some fixed $p \in N$ find a collection of pairwise disjoint sets

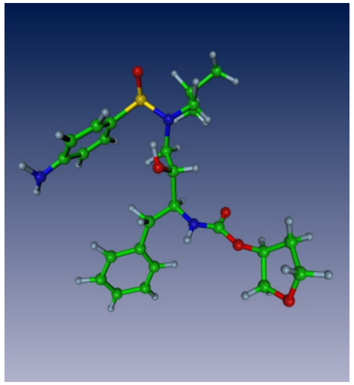
$S = \{S_1, \ldots, S_p\}$ with $\bigcup_{1 \leq i \leq p} S_i = B$ and $\mu(S_i) > 0$ such that

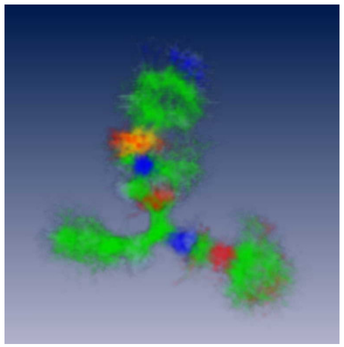
$$\boxed{\rho(S) := \frac{1}{p} \sum_{k=1}^{p} \rho(S_k) = \frac{1}{p} \sum_{k=1}^{p} \frac{\sum_{B_i, B_j \subset S_k} p_{ij} \cdot \mu(B_j)}{\sum_{B_j \in S_k} \mu(B_j)} \rightarrow \max}$$

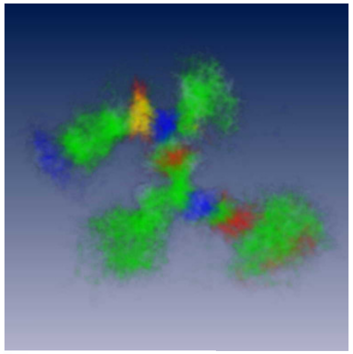
Almost Invariant Sets

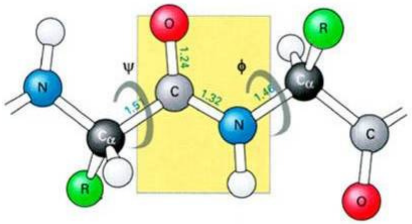
Example: Molecular Dynamics - Pentane

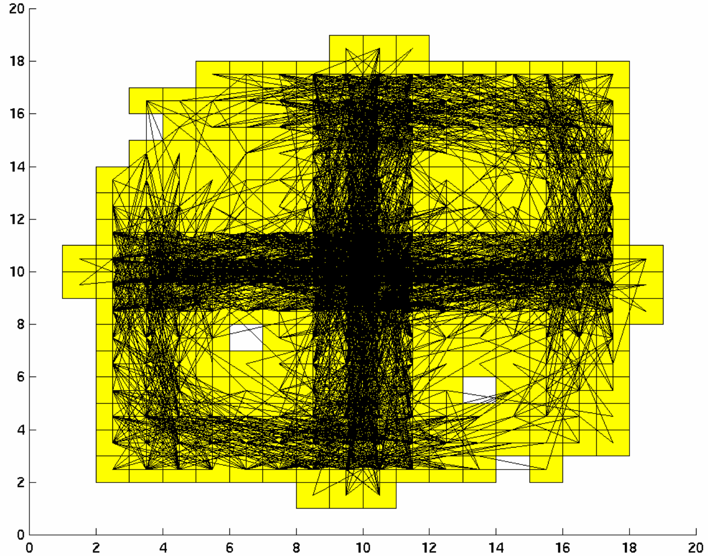




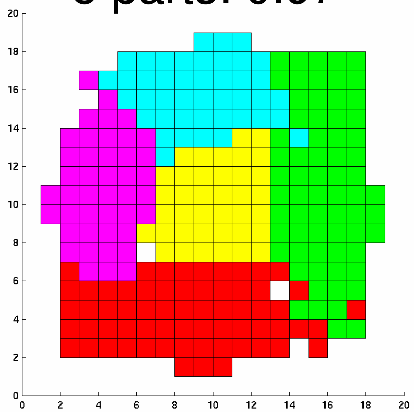




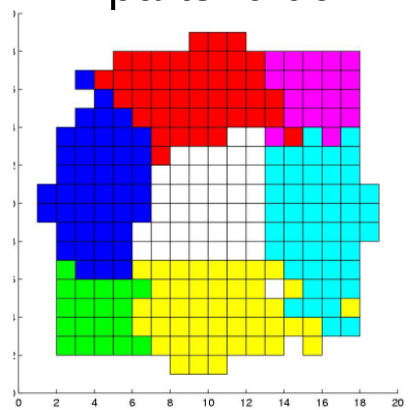




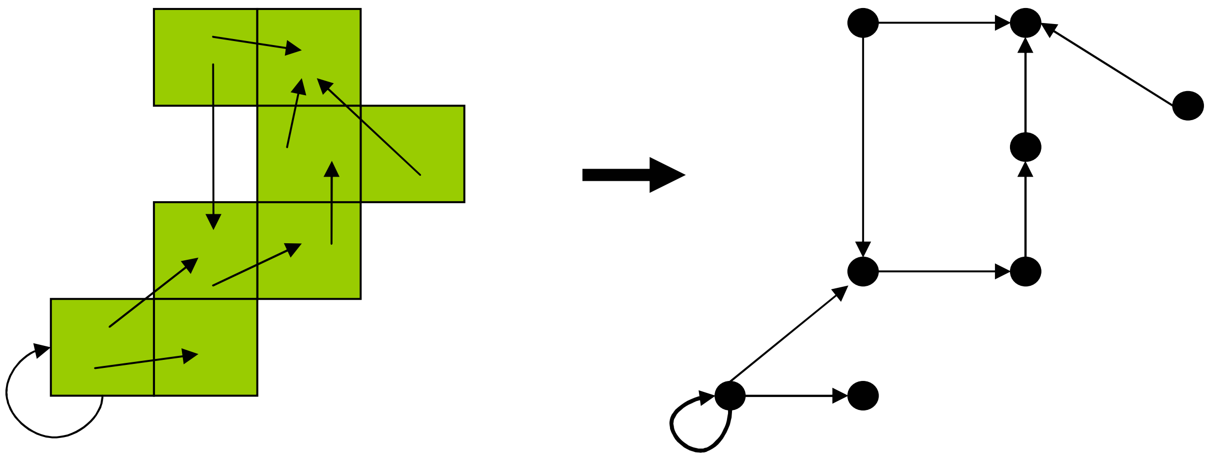
5 parts: 0.97



7 parts: 0.96



Graph Based Approach: The Transition Graph
(Froyland-Dellnitz 01, Dellnitz-P. 02)



- Boxes are vertices
- Coarse dynamics represented by edges
- Edge weights:

$$w(i, j) = \mu(B_i) \cdot P_{ij} = \mu(B_i \cap f^{-1}(B_j))$$

- To compute almost invariant sets use graph partitioning algorithms

Problem 3: Graph

Problem 2 can be formulated as a graph partitioning problem:

Let $G = (V,E)$ be a graph with

- vertex set $V = B$,

- directed edge set $E = \{(B_1,B_2) \in B \times B:\ f(B_1) \cap B_2 \neq 0\}$,

- vertex weights $vw(B_i) = \mu(B_i)$ and

- edge weights $ew((B_i,B_j)) = \mu(B_i)p_{ji}$.

Problem 3 (Graph):

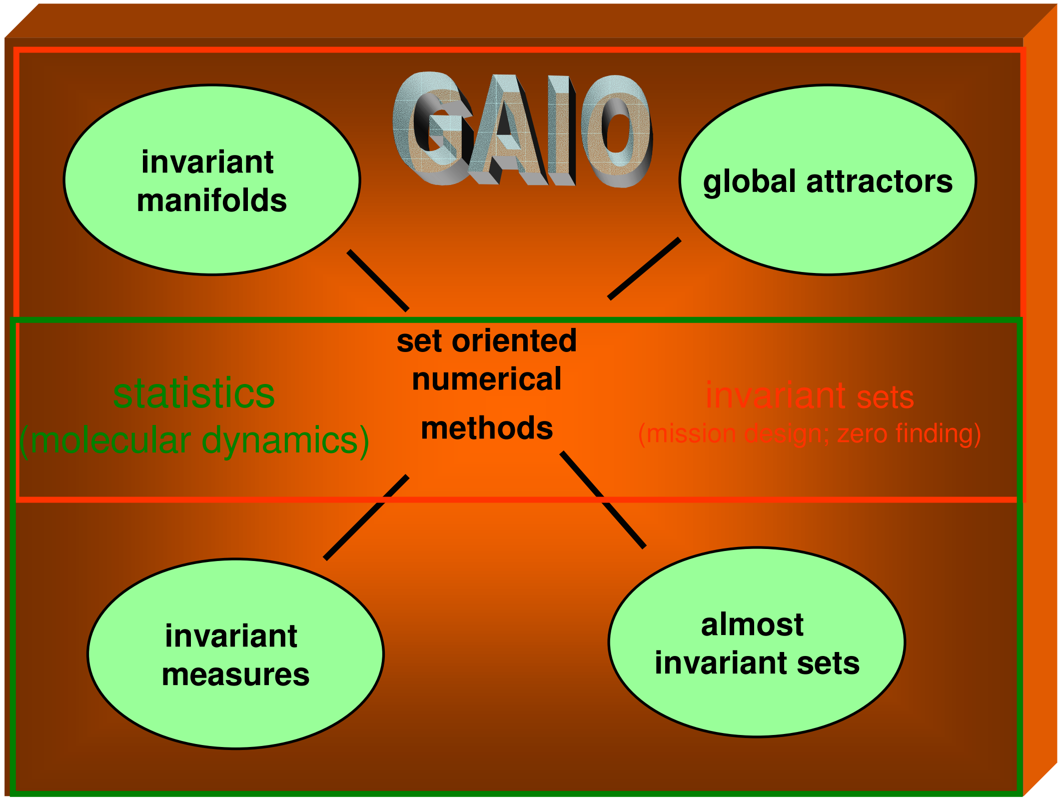
For some fixed $p \in N$ find a collection of pairwise disjoint sets

$S = \{S_1,\ldots,S_p\}$ with $\bigcup_{1 \le i \le p} S_i = V$ and $vw(S_i) > 0$ such that

$$\boxed{\rho(S) := C_{\text{int}}(S) = \frac{1}{p}\sum_{i=1}^{p}\frac{\sum_{(v,w)\in E;\ v,w \in S_i} ew(\{v,w\})}{\sum_{v\in S_i} vw(v)} \to \max}$$

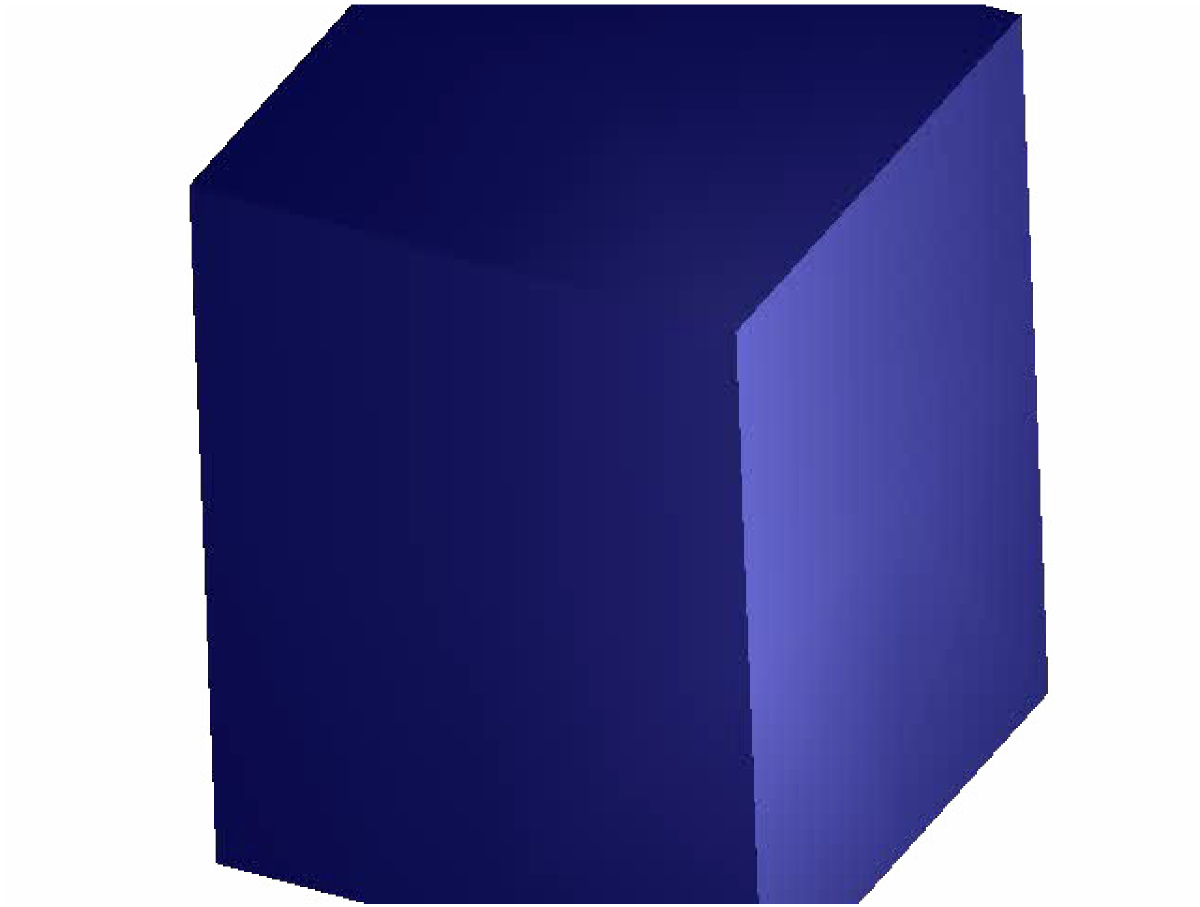
Toolboxes

GAIO (Dellnitz et al.):
Global Analysis of Invariant Objects



GADS (P.):
Graph Algorithms for Dynamical Systems (PARTY,…)

Set Oriented Approximation of Global Attractors



Hierarchical Set Oriented Approach

INPUT: initial box B_0 and number of levels l

FOR $k:=1$ TO l DO
- subdivide boxes of B_{k-1} to obtain box covering B_k
- select only boxes B from B_k for which $f^{-1}(B) \cap C \neq \emptyset$ for some other box C

OUTPUT: B_l

Proposition [Dellnitz-Hohmann 1997]:

Let A_Q be the global attractor of Q and $Q_k = \bigcup_{B \in B_k} B.$ Then

$$\lim_{k \to \infty} h(A_Q, Q_k) = 0$$

with $h(\cdot,\cdot)$ being the Hausdorff distance.

Realization of the Selection Step

We have to check whether $f^{-1}(B_i) \cap B_j \neq \emptyset \quad \forall\, i, j.$

Use test points:

$$\boxed{f(y) \notin B_i \ \text{ for all test points } \ y \in B_j \ ?}$$

Standard choice of test points:

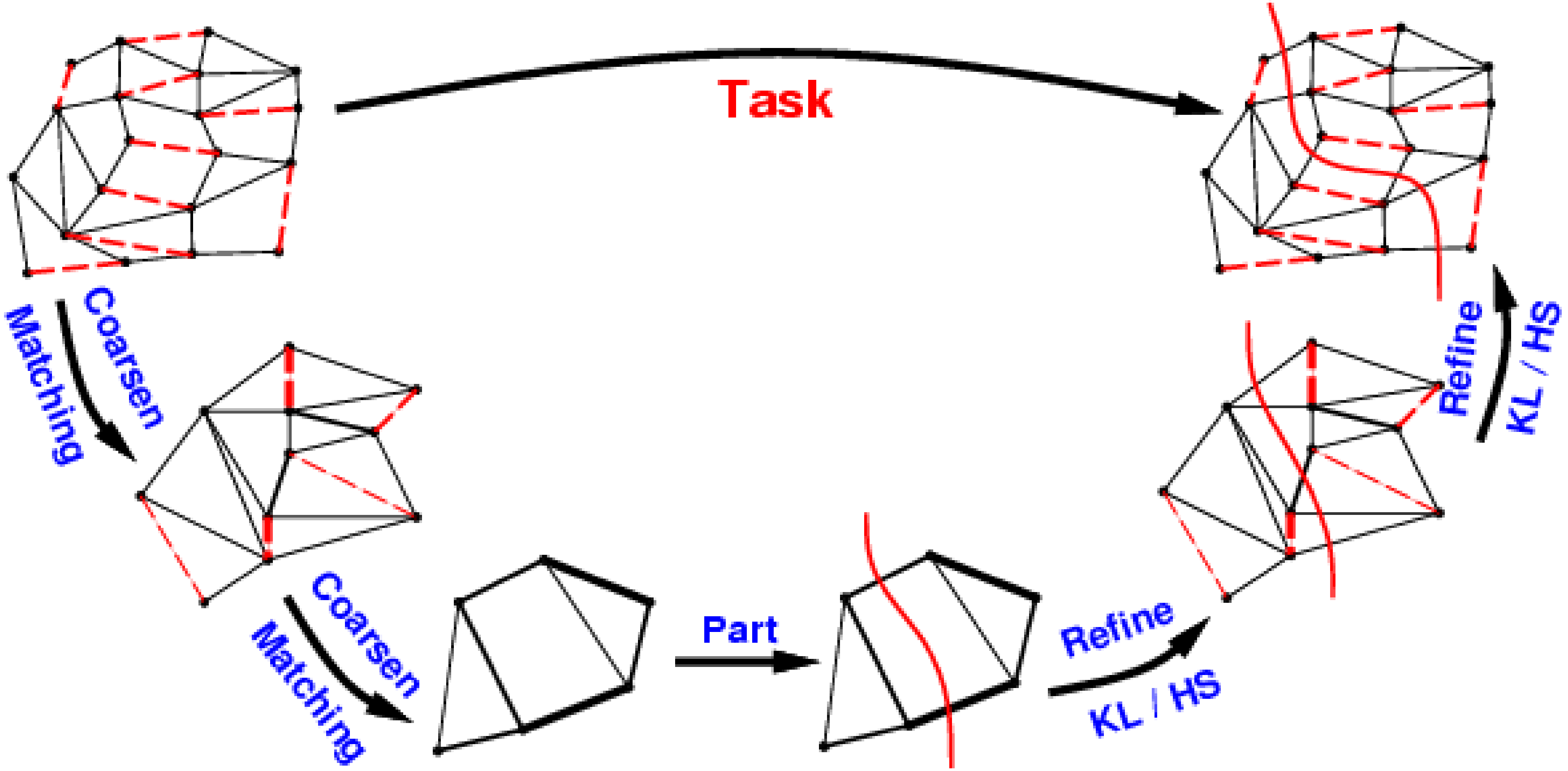
- For low dimensions:
  equidistant distribution, e.g. on the boundaries of boxes.
- For higher dimensions:
  stochastic distribution inside the boxes.

The Standard Approach

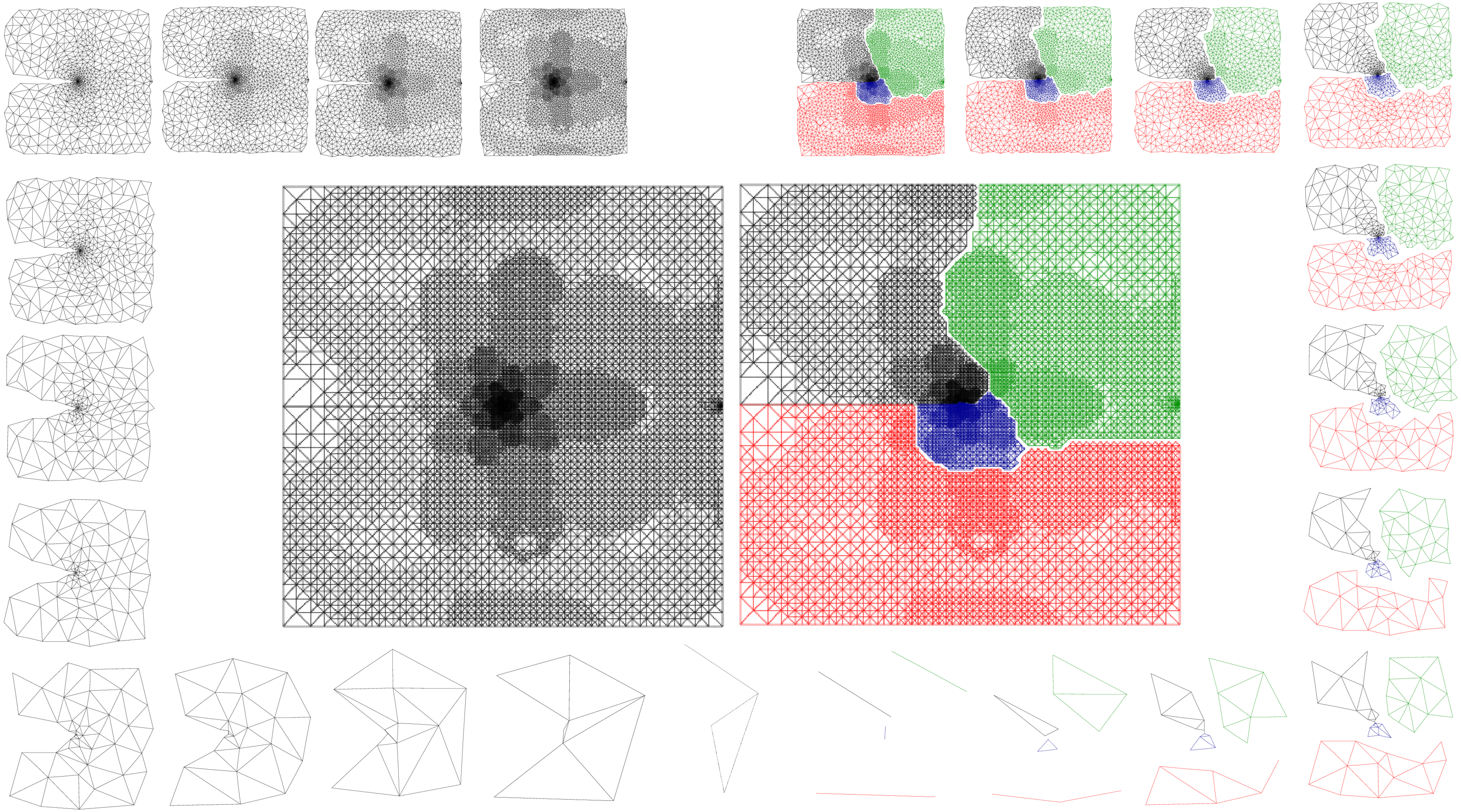
1. *Hierarchical Set Oriented Phase*
    1. *Discretize the space through several levels of boxes*
    2. *Calculate the transition matrix*

2. *Use Graph Partitioning Methods on the finest box level*

Multilevel Graph Partitioning Paradigm



Multilevel Graph Partitioning



UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

PaSCo
Paderborn Institute for
Scientific Computation

Multilevel Graph Partitioning

INPUT: G_0 , number of parts p and number of levels L

$k=0$

WHILE $V_k > L$ DO

- coarse vertices in G_k to construct a smaller Graph G_{k+1}
- k:=k+1

Compute partition S_k of V_k into p parts

WHILE $k > 0$ DO

- $k:=k-1$
- project S_{k+1} to a partition S_k of G_k
- locally optimize S_k with respect to C_{int}

OUTPUT: S_0

The Standard Approach

1. *Hierarchical Set Oriented Phase*
   - a *Discretize the state space through several levels of boxes*
   - b *Calculate the transition matrix*

2. *Multilevel Graph Partitioning Phase*
   - a *Translate the problem into a graph partitioning problem*
   - b *Use Multilevel Graph Partitioning Methods*

1.a and 2.b exhibit multilevel structures!

The Integrated Approach

INPUT: initial box B_0, the number of parts p and the number of levels l

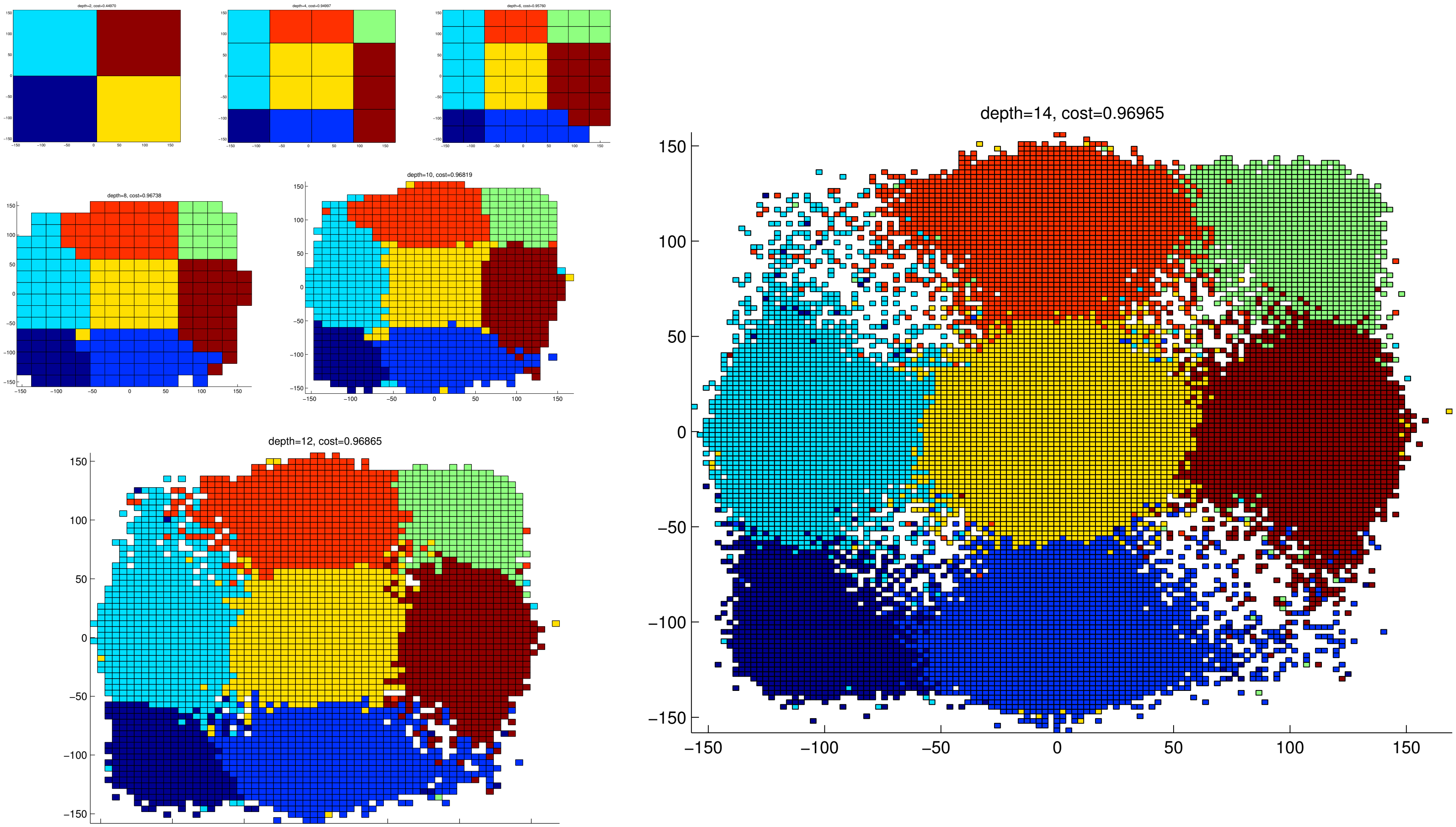
$S_0 = \{B_0\}$

FOR $k:=1$ TO l DO

1) subdivide boxes of B_{k-1} and select boxes to obtain box covering B_k
2) compute transition matrix P_k
3) project S_{k-1} to a partition S_k of B_k
4) locally optimize S_k with respect to C_{int}

OUTPUT: B_l , S_l

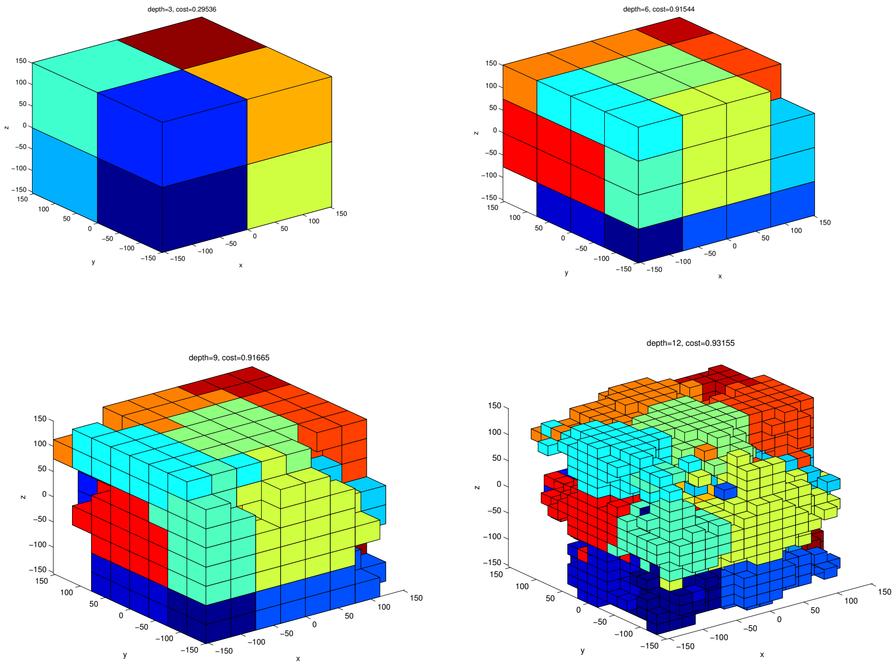
Example: Pentane (7 parts)

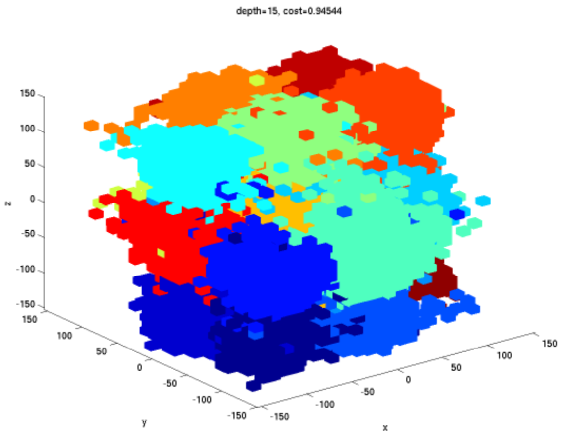


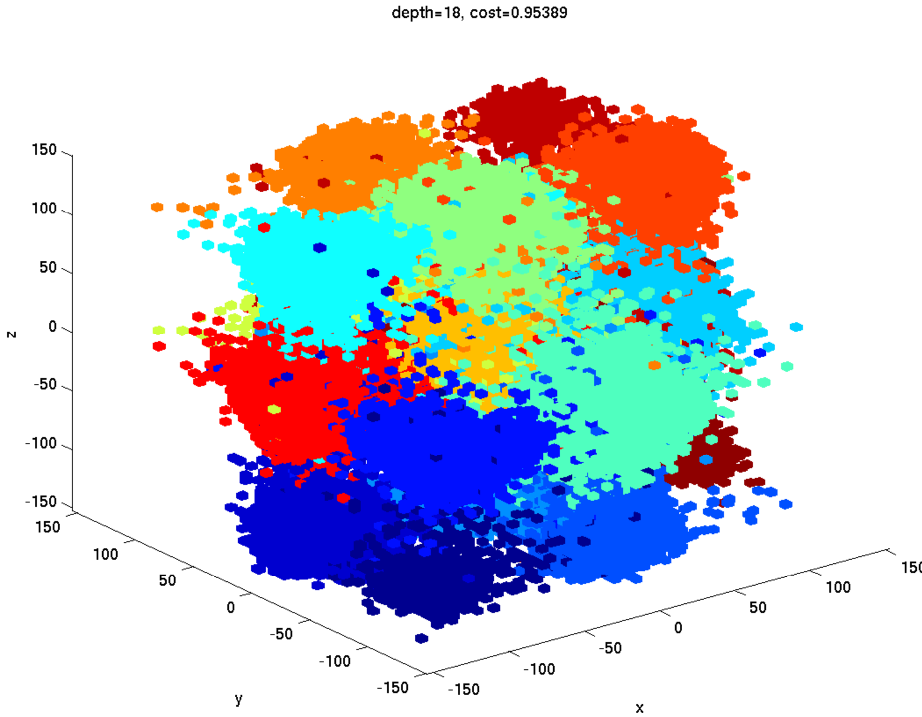
Internal Costs

Box level	No. Of boxes	C_{int} (proj.)	C_{int} (local)
2	4	.450	.450
4	16	.450	.950
6	63	.950	.958
8	230	.958	.967
10	851	.967	.968
12	3124	.968	.969
14	10624	.969	.969

Example: Hexane (17 parts)



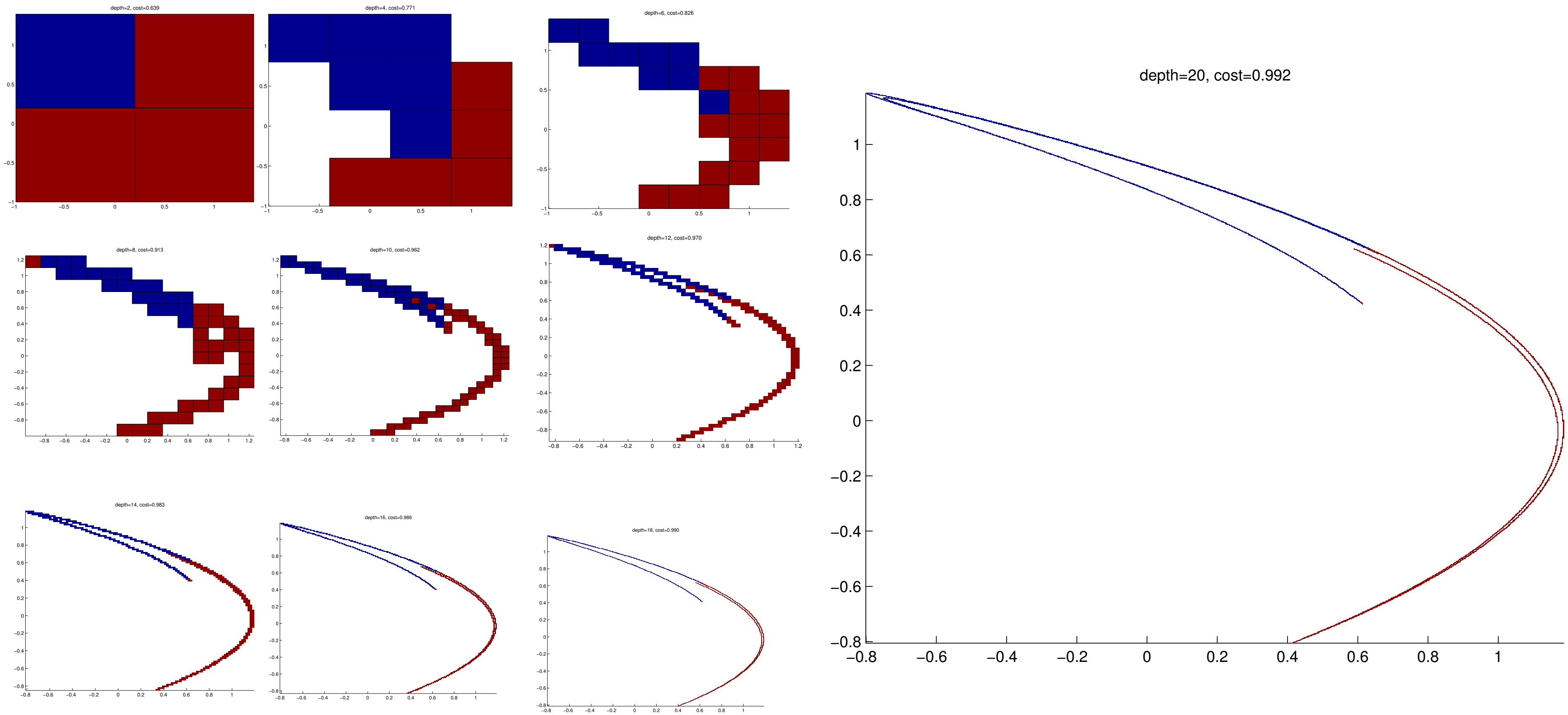




Internal Costs

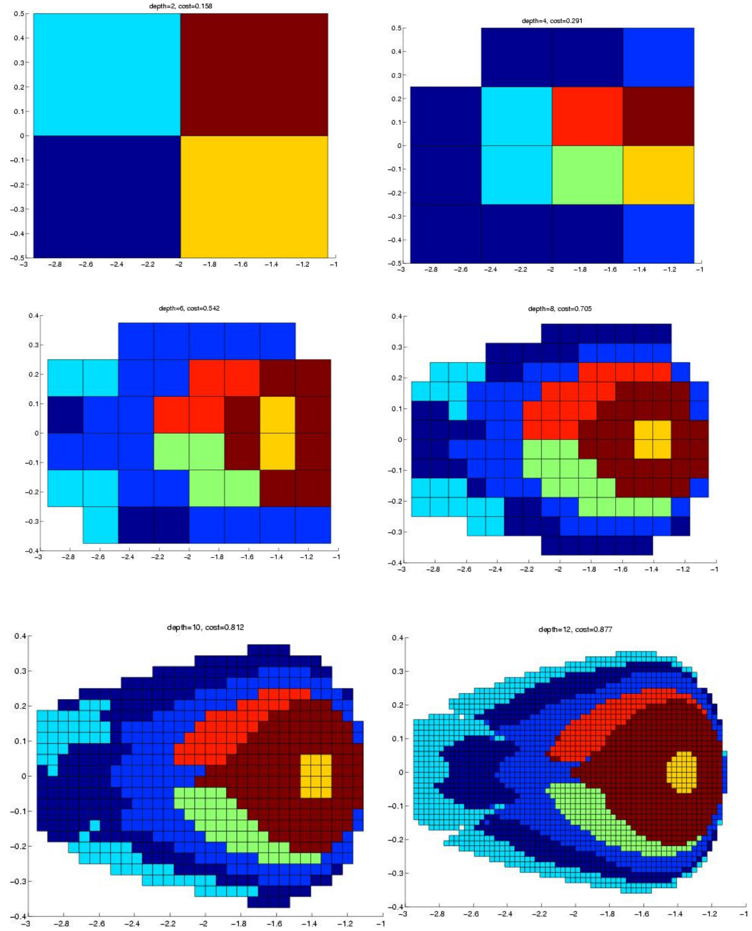
Box level	No. Of boxes	C_{int} (proj.)	C_{int} (local)
3	8	.295	.295
6	62	.295	.915
9	426	.915	.917
12	2262	.917	.932
15	10568	.932	.945
18	42537	.945	.954

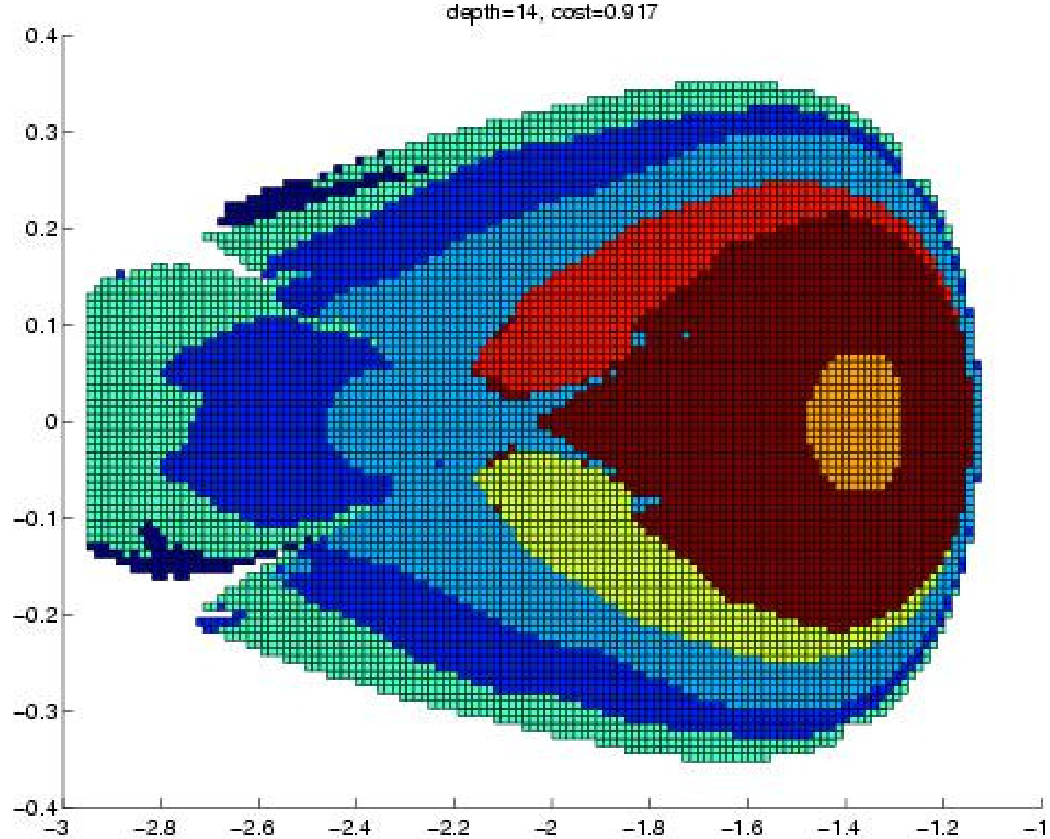
Example: Hénon Map (2 parts)



Example: PCRTB (7 parts)

Poincare Surface in the Planar Circular Restricted Three Body Problem of Sun, Jupiter and a particle





DONEs - TODOs

<u>DONEs</u>

- Interlock of multilevel mechanisms
- Application to applications

<u>TODOs</u>

- More/better interlock?
- Right number of parts p?
- Paths between almost invariant sets

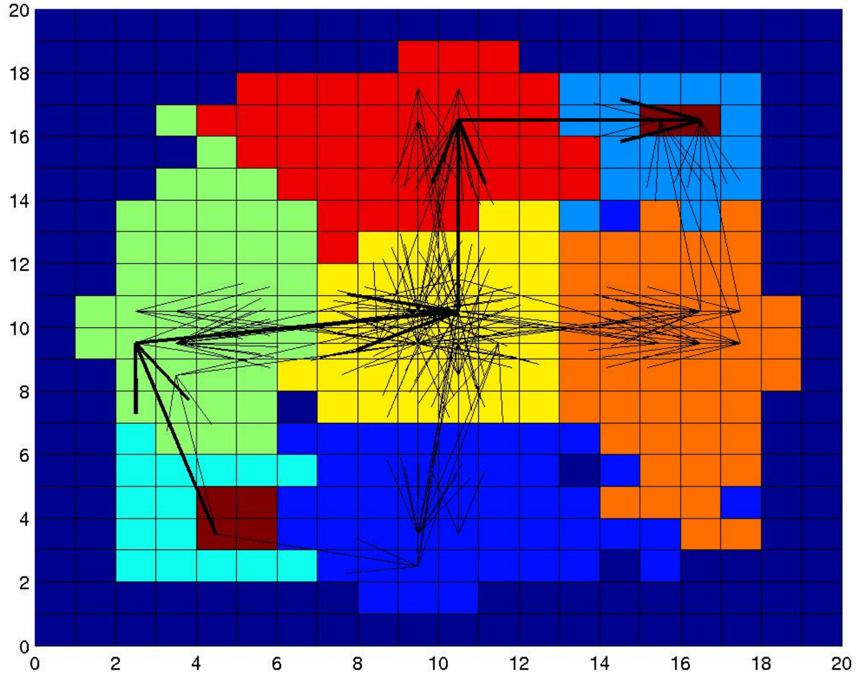
Outlook: Calculation of Transition Paths (Dominant Paths Between Almost Invariant Sets)

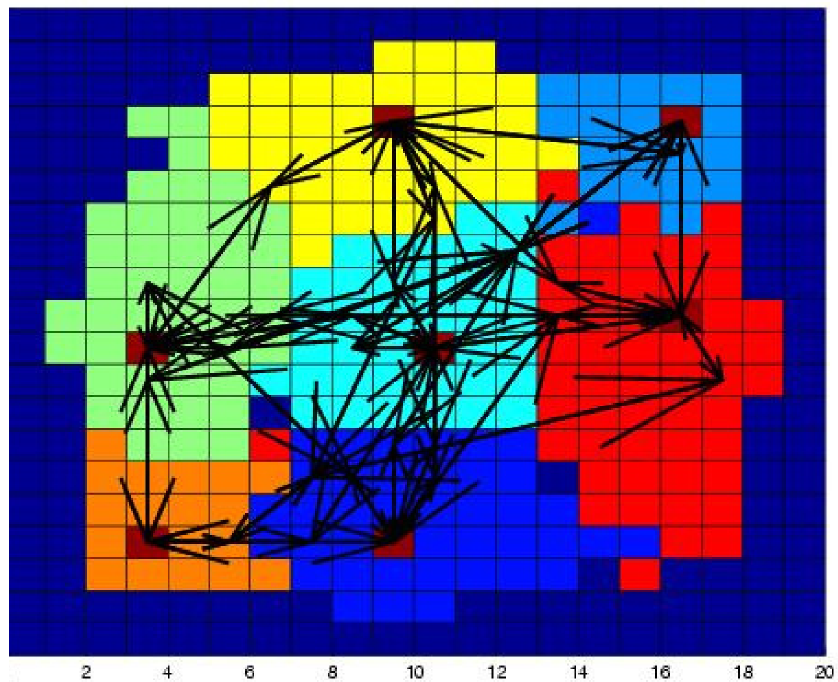
Shortest Paths Generalizations:

- Sets of sources/destinations
- All (1+e)-shortest paths

Strategy:

1. Generate boxes and graph
2. Calculate partition
3. Calculate centers of parts
4. Calculate shortest paths





UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

PaSCo
Paderborn Institute for Scientific Computation

Thank you for your attention.

The story about continuous people and discrete people
is like in a marriage:

you are either the man or the woman.

There are problems in life which are better addressed only by men or only by women, but for the most important problems in human life they have to

work together!