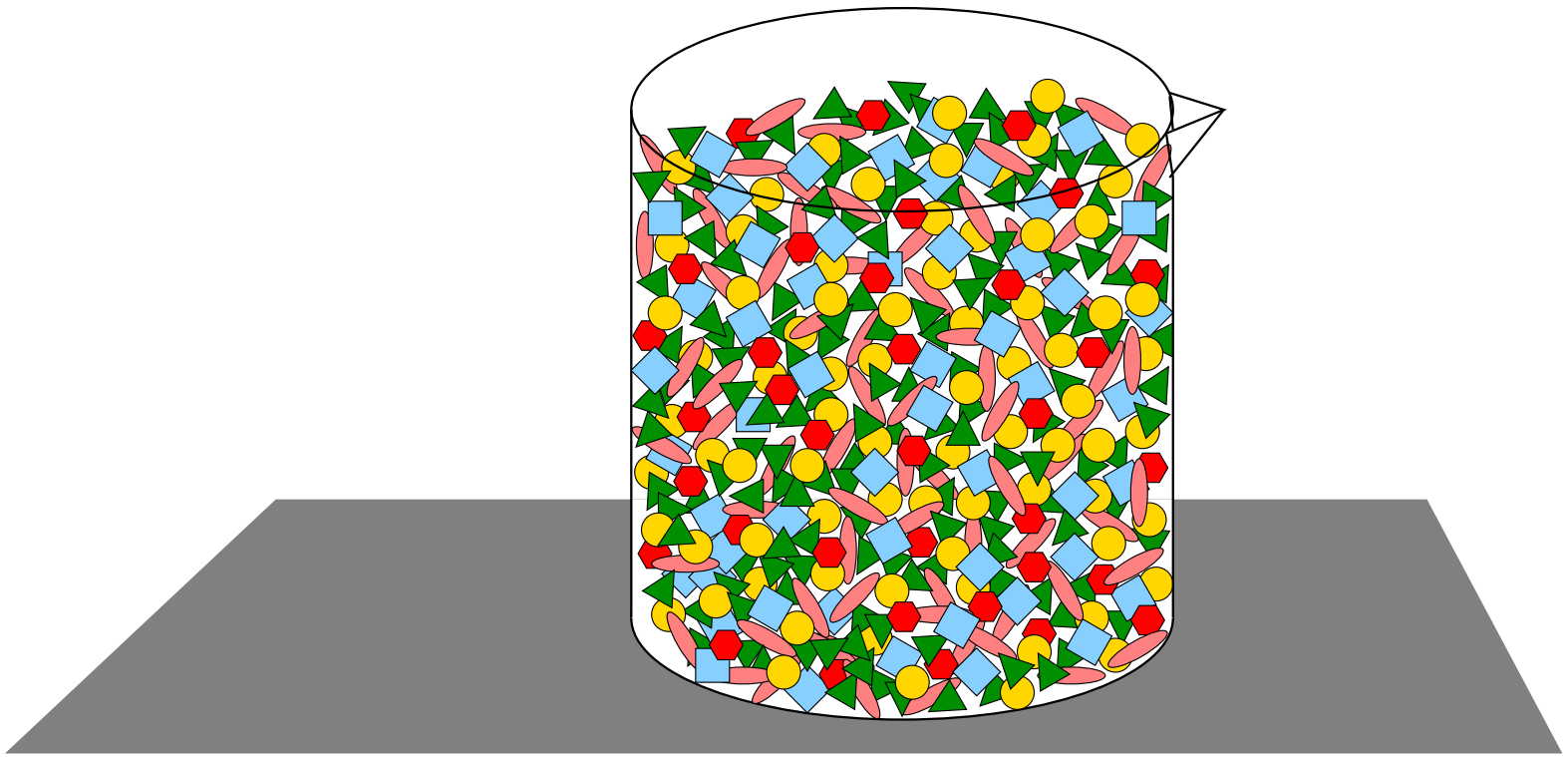
Categorised Counting Mediated by Blotting Membrane Systems for Particle-based Data Mining and Numerical Algorithms

Thomas Hinze[1,2] Konrad Grützmann[3] Benny Höckner[1]
Peter Sauer[1] Sikander Hayat[4]

[1]Brandenburg University of Technology Cottbus
Institute of Computer Science and Information and Media Technology
[2]Friedrich Schiller University Jena
[3]Helmholtz Centre for Environmental Research Leipzig
[4]Harvard Medical School Boston

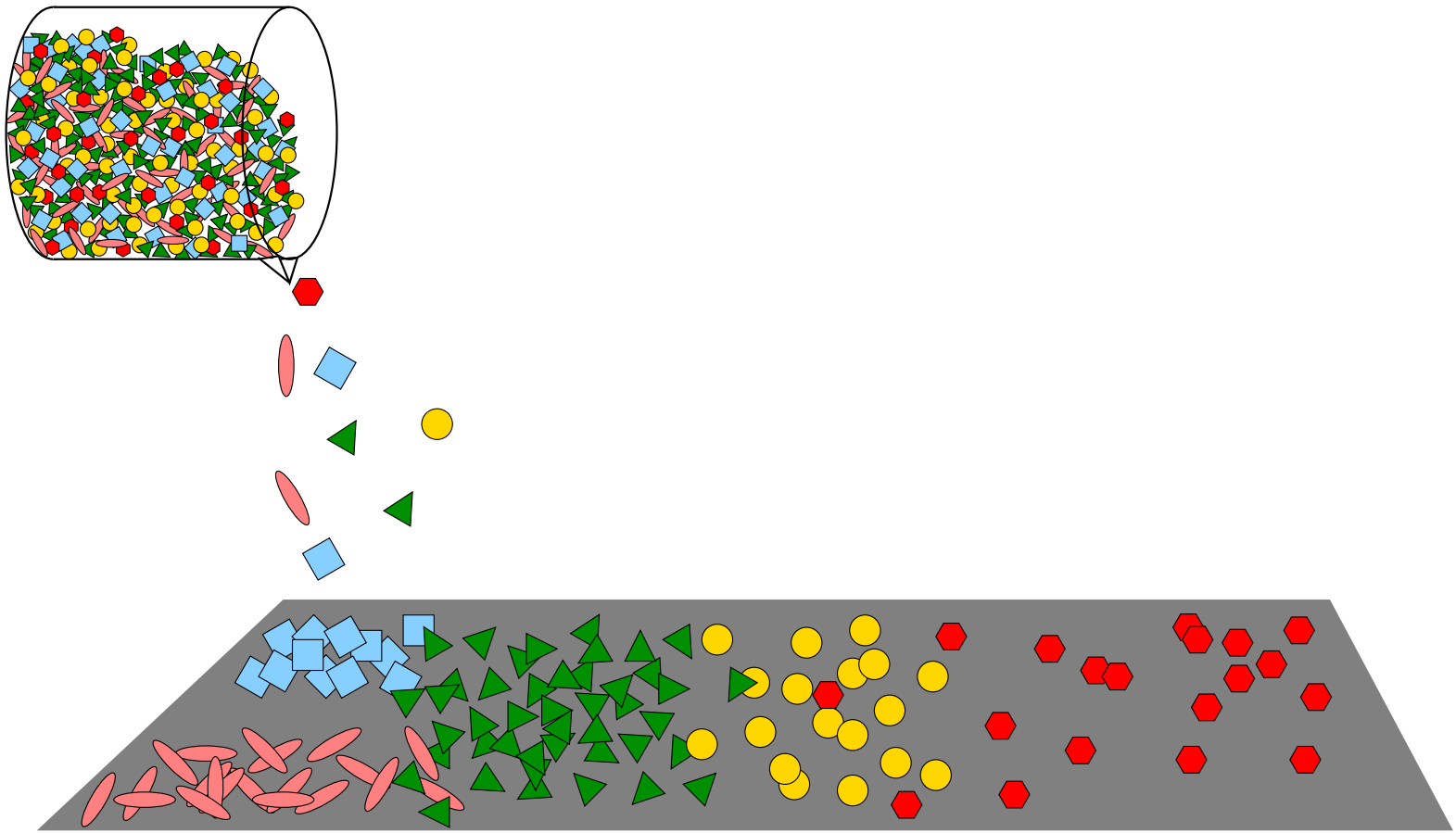
thomas.hinze@tu-cottbus.de

Blotting – Productive and Simple Principle (I)



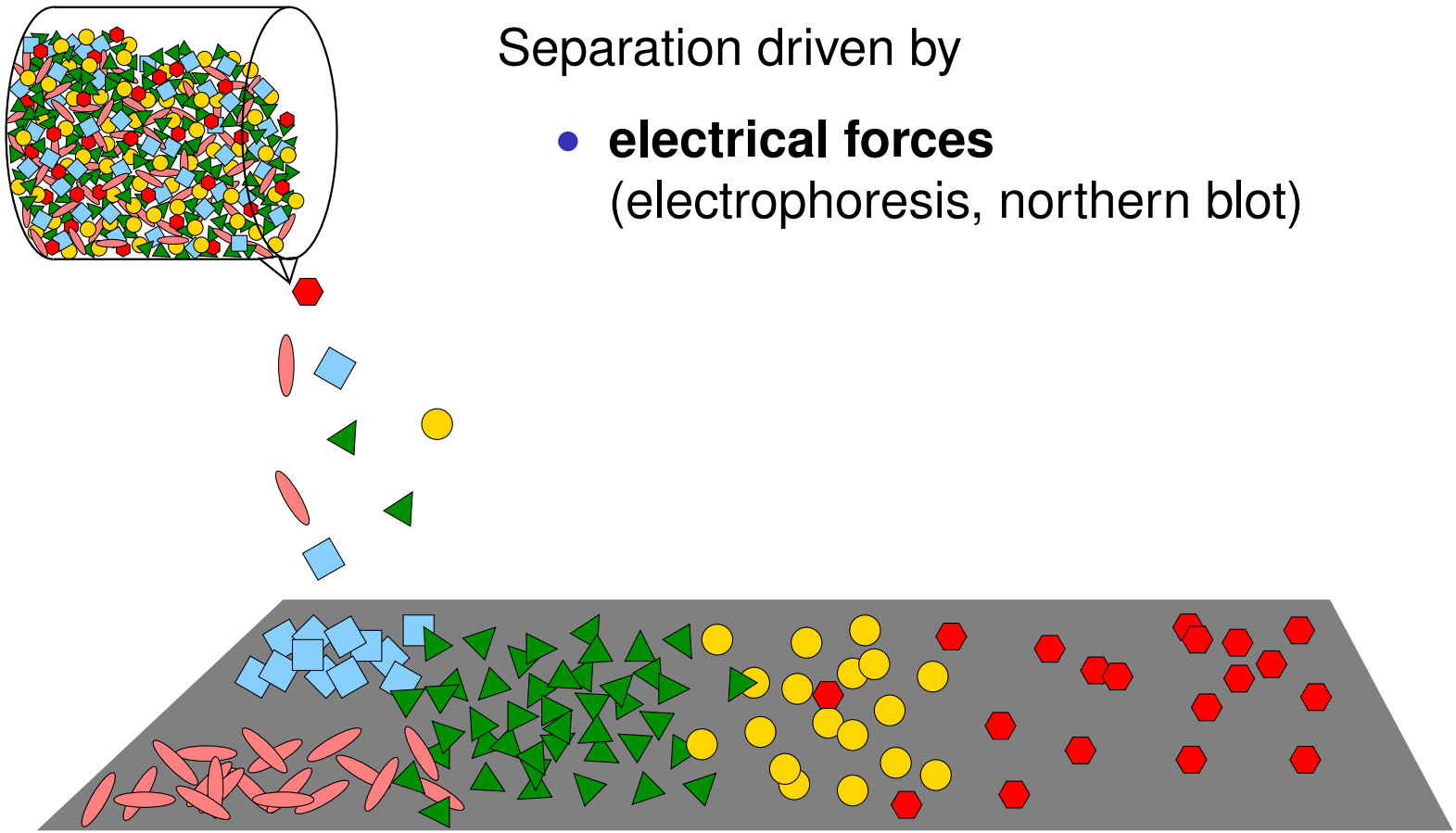
1. Mixture of particles like reactive or labelled molecules

Blotting – Productive and Simple Principle (II)



2. Spatial separation of particles on a grid according to molecular attributes

Blotting – Productive and Simple Principle (II)

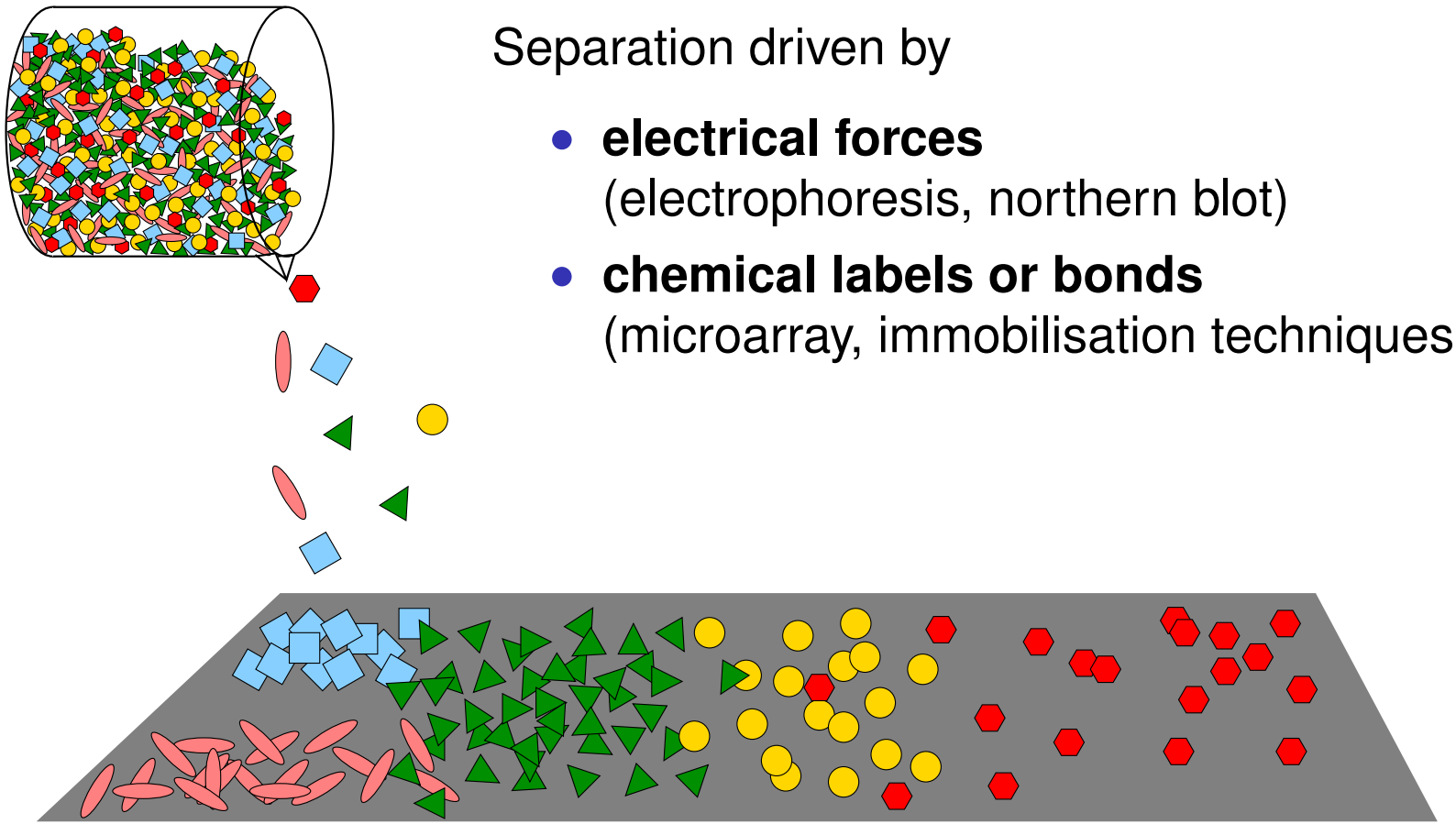


Separation driven by

- **electrical forces**
  (electrophoresis, northern blot)

2. Spatial separation of particles on a grid according to molecular attributes

Blotting – Productive and Simple Principle (II)

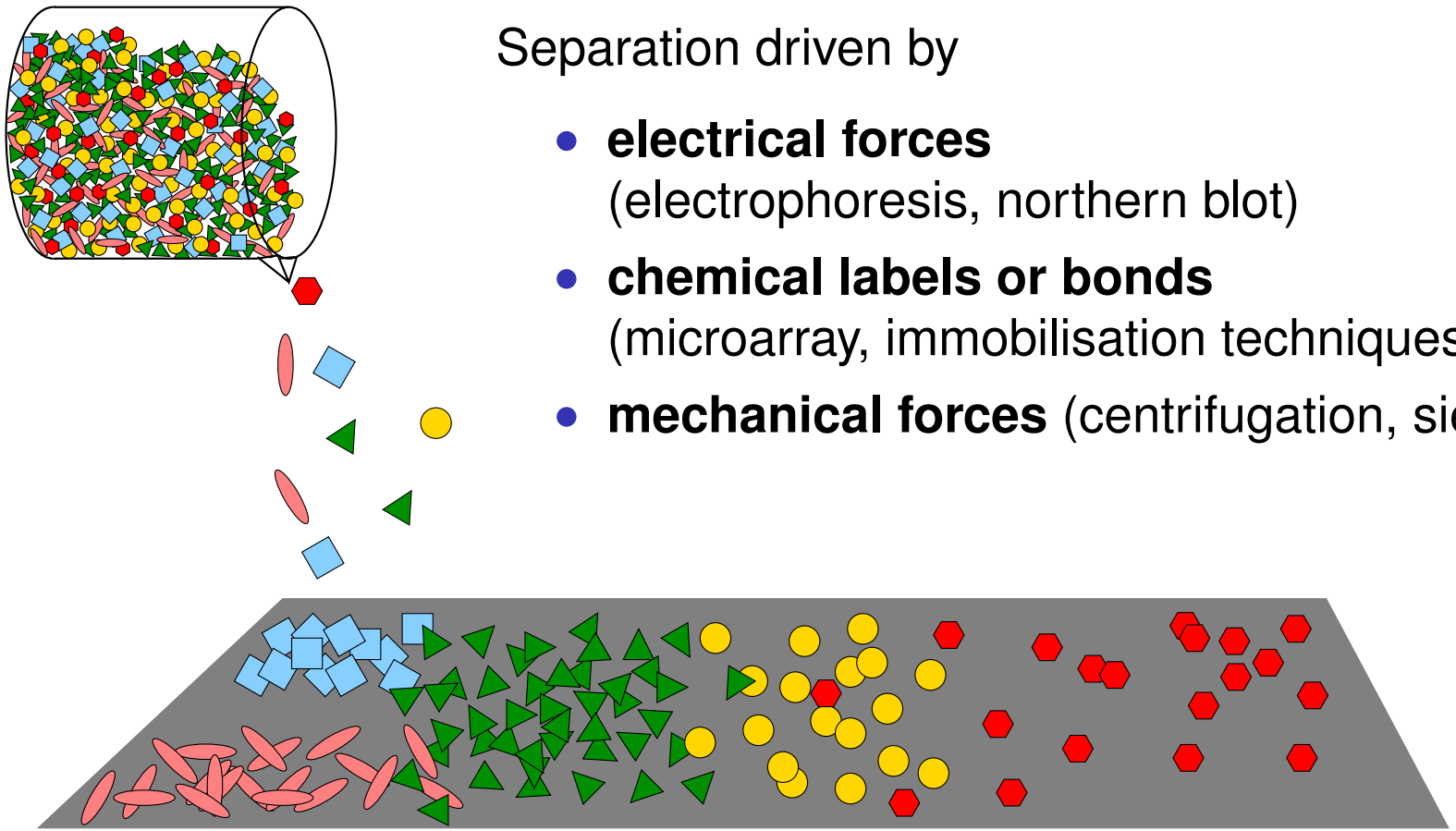


Separation driven by

- **electrical forces**
  (electrophoresis, northern blot)
- **chemical labels or bonds**
  (microarray, immobilisation techniques)

2. Spatial separation of particles on a grid according to molecular attributes

Blotting – Productive and Simple Principle (II)

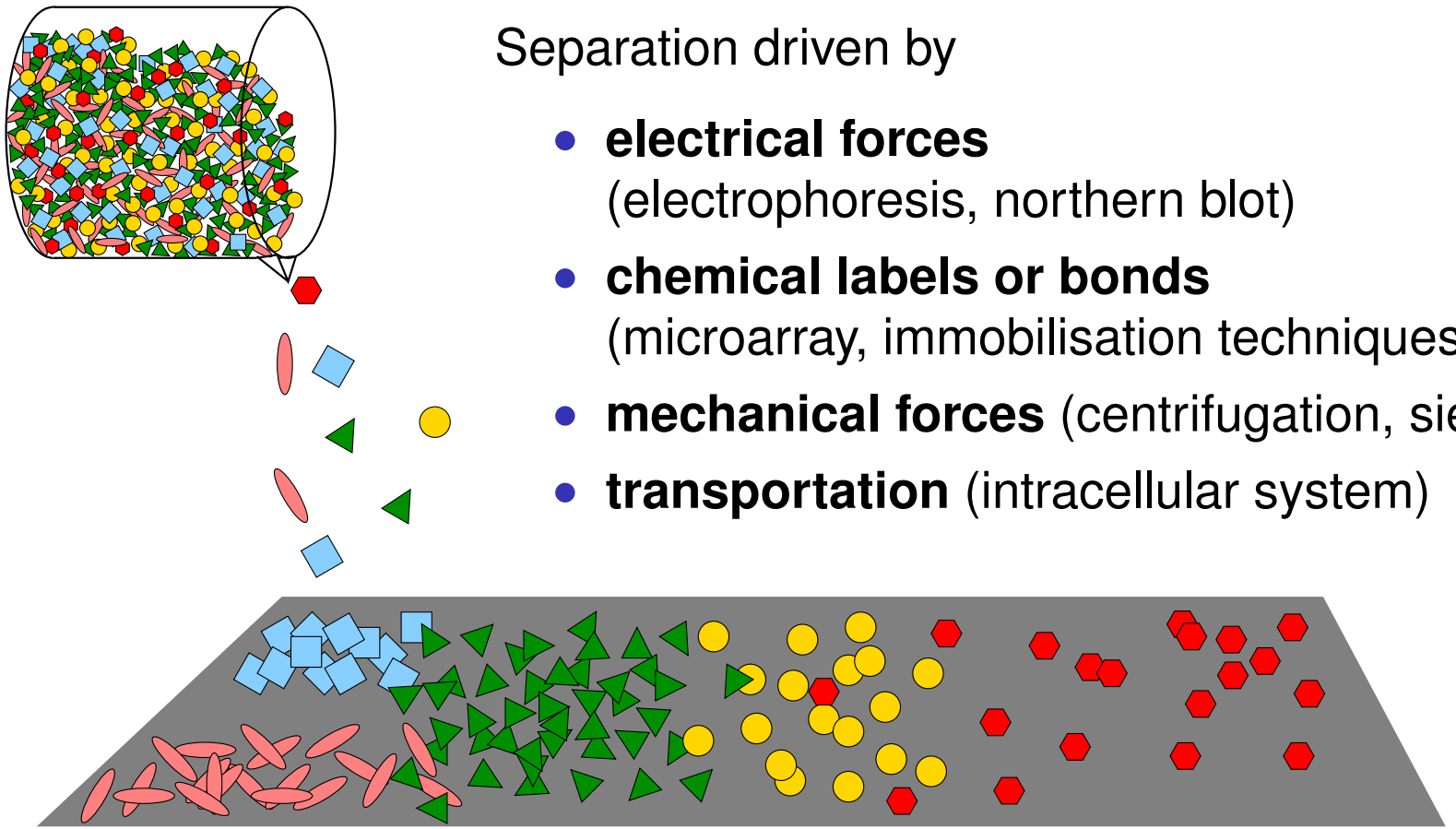


Separation driven by

- **electrical forces**
  (electrophoresis, northern blot)
- **chemical labels or bonds**
  (microarray, immobilisation techniques)
- **mechanical forces** (centrifugation, sieve)

2. Spatial separation of particles on a grid according to molecular attributes

Blotting – Productive and Simple Principle (II)

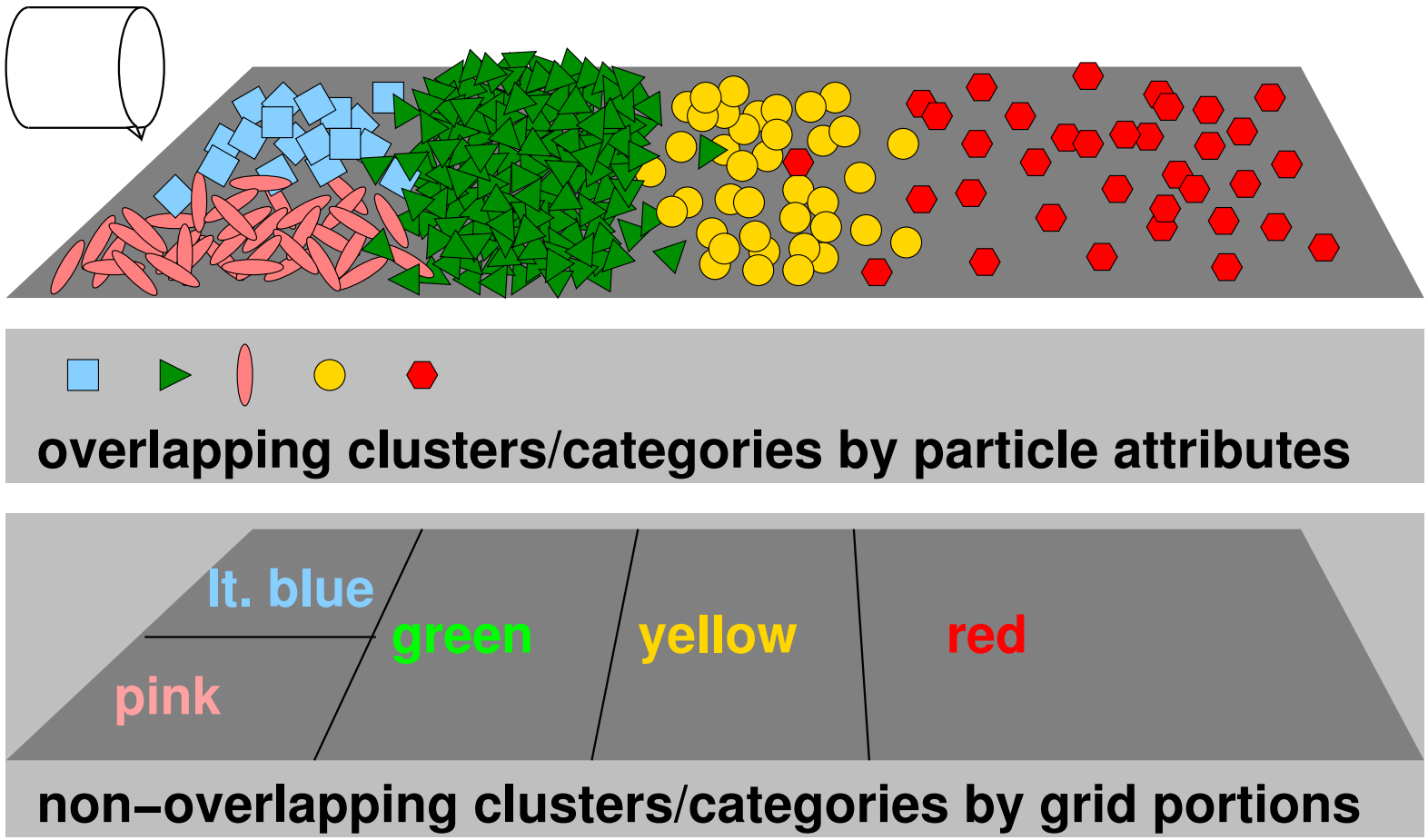


Separation driven by

- **electrical forces** (electrophoresis, northern blot)
- **chemical labels or bonds** (microarray, immobilisation techniques)
- **mechanical forces** (centrifugation, sieve)
- **transportation** (intracellular system)

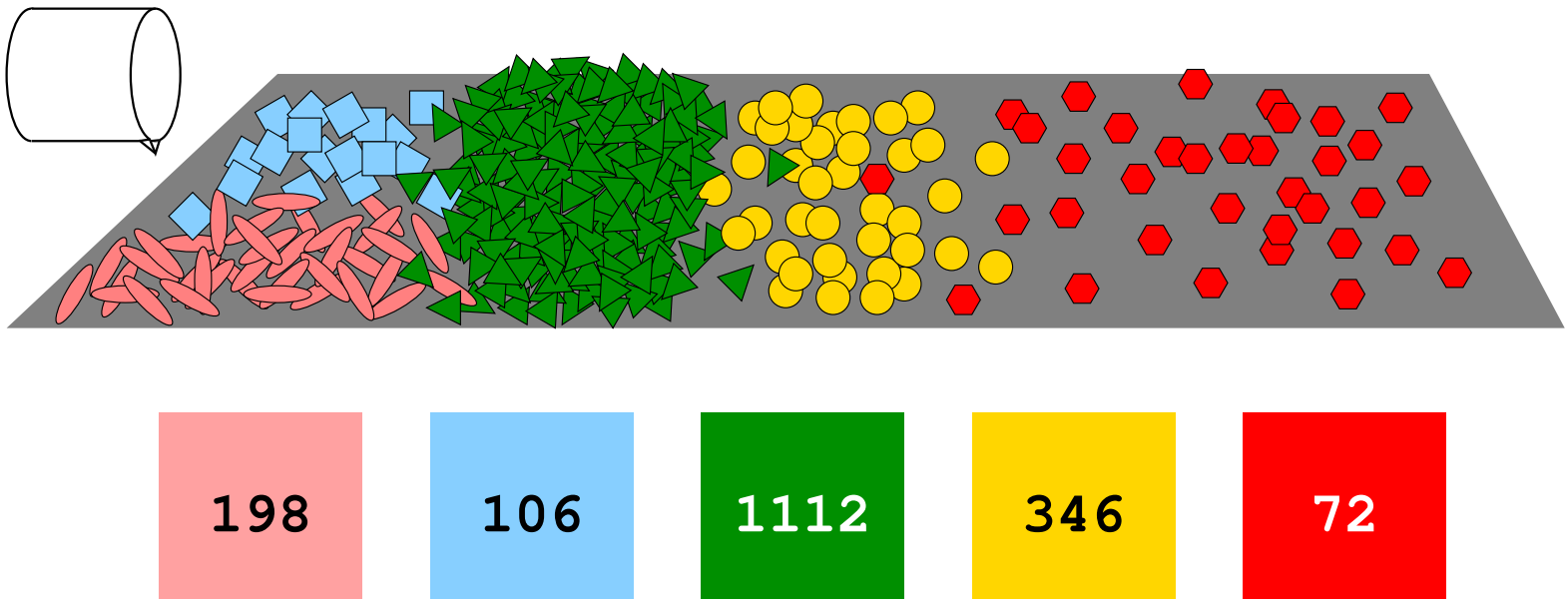
2. Spatial separation of particles on a grid according to molecular attributes

Blotting – Productive and Simple Principle (III)



3. Identification of particle *clusters* on the grid or *categories* of particles (overlapping or non-overlapping)

Blotting – Productive and Simple Principle (IV)



4. ***Counting*** **or** ***scoring*** **of particles within each cluster/category**

Blotting – Productive and Simple Principle (V)

maximum ratio:

green/red with 1112/72 approx. 15.4

5. Generate *response* resulting from numerical analysis coinciding with question(s) of interest

Blotting as Computation

- **Input:** grid coordinates of all particles under study
- **Output:** final response resulting from scores or counts

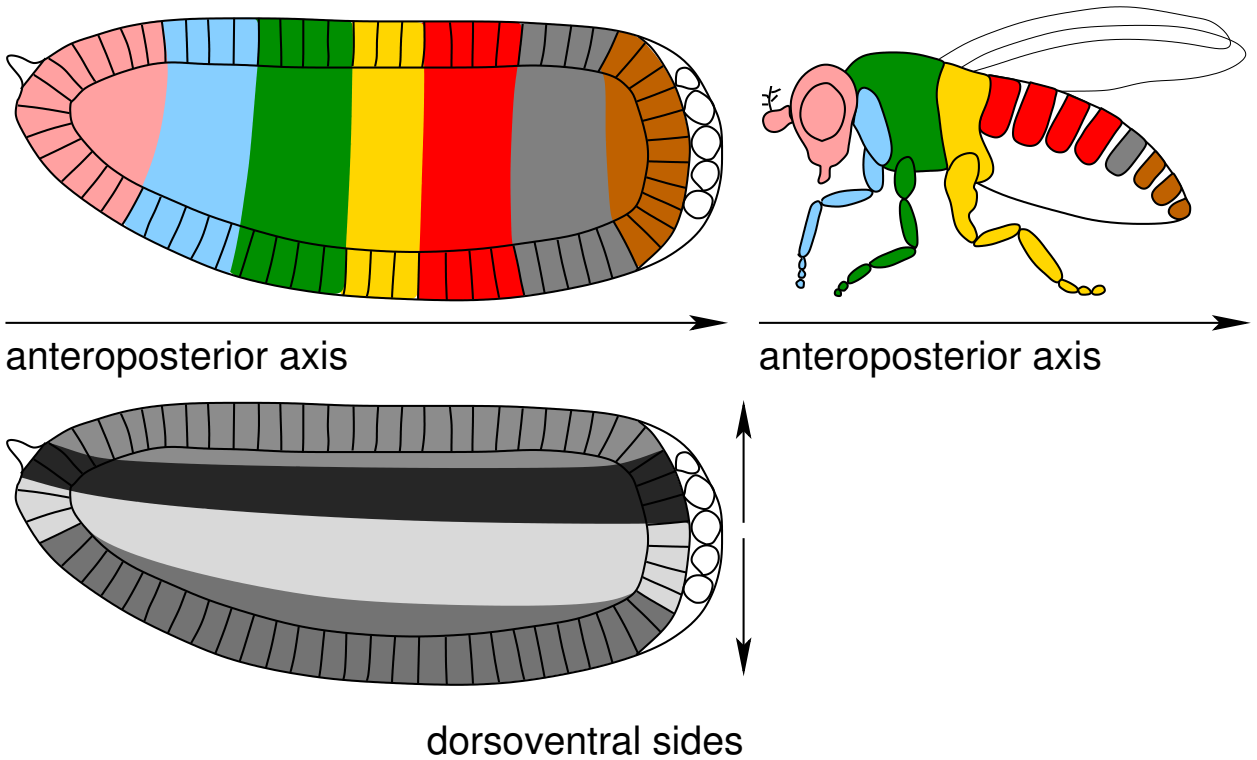
Blotting as Computation

- **Input:** grid coordinates of all particles under study
- **Output:** final response resulting from scores or counts

Utilisation

- Tremendous *data reduction* keeping essential information
- Support of *data mining* strategies for applications in bioinformatics, especially in image evaluation
- Tool for performing *unconventional computing*
- Experimental setup for *algorithmic design* inspired by placement of particles
- Promising aspect in *applications of membrane systems* and its underlying modelling formalism

An Example of Spatial Blotting in Nature



- Embryonic pattern in *drosophila melanogaster* forms a 7×4-grid
- 28 clusters with specific cytokine combinations
- Cell differentiation and proliferation during maturation

1. Motivation and Principle of Blotting
2. **Blotting Membrane Systems**
   - Definition
   - Toy Example: Approximation of Constant $\pi \approx 3.14$
3. Particle-based Numerical Integration
4. Electrophoresis: A Molecular Bucket Sort

Definition Blotting Membrane System $\boxed{\Pi}$

$$\boxed{\Pi} = (P, L, C, B_1, \ldots, B_{|C|}, S, R, \mathrm{r})$$

Particles

L ..... arbitrary set of available labels

Definition Blotting Membrane System $\boxed{\Pi}$

$$\boxed{\Pi} = (P, L, C, B_1, \ldots, B_{|C|}, S, R, \mathrm{r})$$

Particles

L ..... arbitrary set of available labels

$P \subset \mathbb{R} \times \mathbb{R} \times L$ ..... final set of particles, each of them specified by grid position and label

Definition Blotting Membrane System $\boxed{\Pi}$

$$\boxed{\Pi} = (P, L, C, B_1, \ldots, B_{|C|}, S, R, \mathrm{r})$$

Particles

L ..... arbitrary set of available labels

$P \subset \mathbb{R} \times \mathbb{R} \times L$ ..... final set of particles, each of them specified by grid position and label

Categories

C ..... arbitrary set of available categories either defined explicitly or obtained implicitly as result of a classification over P

Definition Blotting Membrane System $\boxed{\Pi}$

$$\boxed{\Pi} = (P, L, C, B_1, \ldots, B_{|C|}, S, R, \mathrm{r})$$

Particles

L ..... arbitrary set of available labels

$P \subset \mathbb{R} \times \mathbb{R} \times L$ ..... final set of particles, each of them specified by grid position and label

Categories

C ..... arbitrary set of available categories either defined explicitly or obtained implicitly as result of a classification over P

$B_1 \subseteq P$

$\vdots$ ..... entirety of blots, each of them specified by the accumulated particles

$B_{|C|} \subseteq P$

Definition Blotting Membrane System $\boxed{\Pi}$

$$\boxed{\Pi} = (P, L, C, B_1, \ldots, B_{|C|}, S, R, \mathrm{r})$$

Particles

L ..... arbitrary set of available labels

$P \subset \mathbb{R} \times \mathbb{R} \times L$ ..... final set of particles, each of them specified by grid position and label

Categories

C ..... arbitrary set of available categories either defined explicitly or obtained implicitly as result of a classification over P

$B_1 \subseteq P$

$\vdots$ ..... entirety of blots, each of them

$B_{|C|} \subseteq P$ specified by the accumulated particles

Scores and response

$S : C \longrightarrow \mathbb{N}$. multiset subsuming the score values over all categories

Definition Blotting Membrane System Π

$$\Pi = (P, L, C, B_1, \ldots, B_{|C|}, S, R, \mathrm{r})$$

Particles

L ..... arbitrary set of available labels

$P \subset \mathbb{R} \times \mathbb{R} \times L$ ..... final set of particles, each of them specified by grid position and label

Categories

C ..... arbitrary set of available categories either defined explicitly or obtained implicitly as result of a classification over P

$B_1 \subseteq P$

$\vdots$ ..... entirety of blots, each of them

$B_{|C|} \subseteq P$ specified by the accumulated particles

Scores and response

$S : C \longrightarrow \mathbb{N}$. multiset subsuming the score values over all categories

R ..... arbitrary set specifying the response domain

Definition Blotting Membrane System $\boxed{\Pi}$

$$\boxed{\Pi} = (P, L, C, B_1, \ldots, B_{|C|}, S, R, \mathrm{r})$$

Particles

L ..... arbitrary set of available labels

$P \subset \mathbb{R} \times \mathbb{R} \times L$ ..... final set of particles, each of them specified by grid position and label

Categories

C ..... arbitrary set of available categories either defined explicitly or obtained implicitly as result of a classification over P

$B_1 \subseteq P$

$\vdots$ ..... entirety of blots, each of them

$B_{|C|} \subseteq P$ specified by the accumulated particles

Scores and response

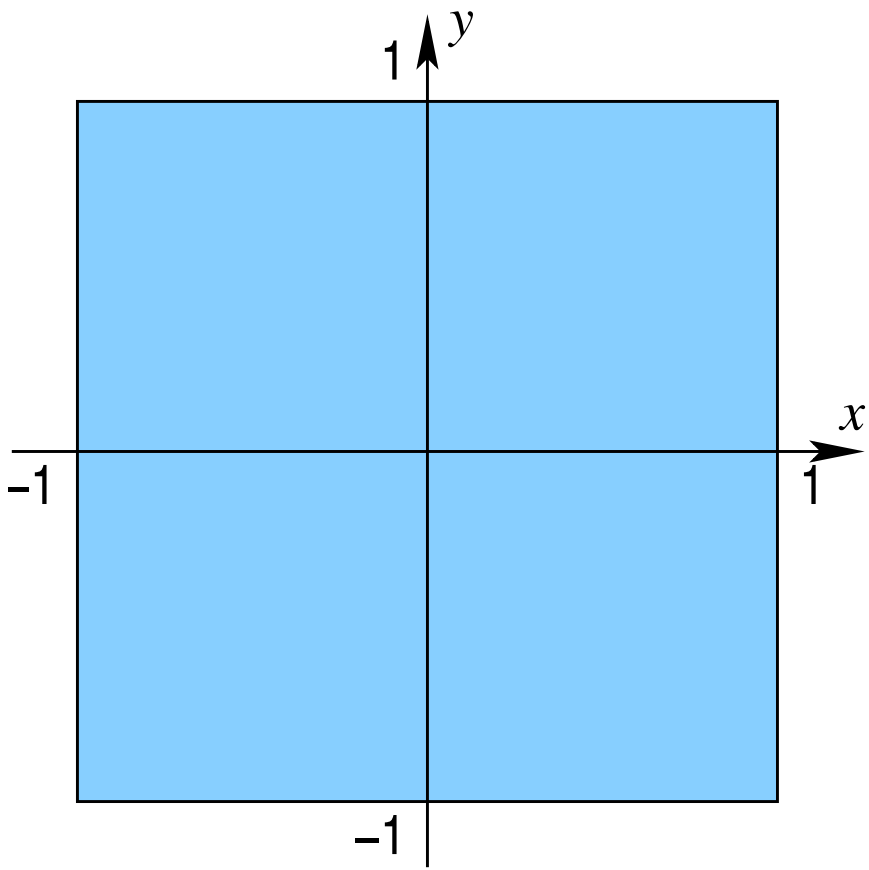
$S : C \longrightarrow \mathbb{N}$. multiset subsuming the score values over all categories

R ..... arbitrary set specifying the response domain

$\mathrm{r} : \mathbb{N}^{|C|} \longrightarrow R$ ..... response function

Toy Example: Approximation of Constant $\pi \approx 3.14$

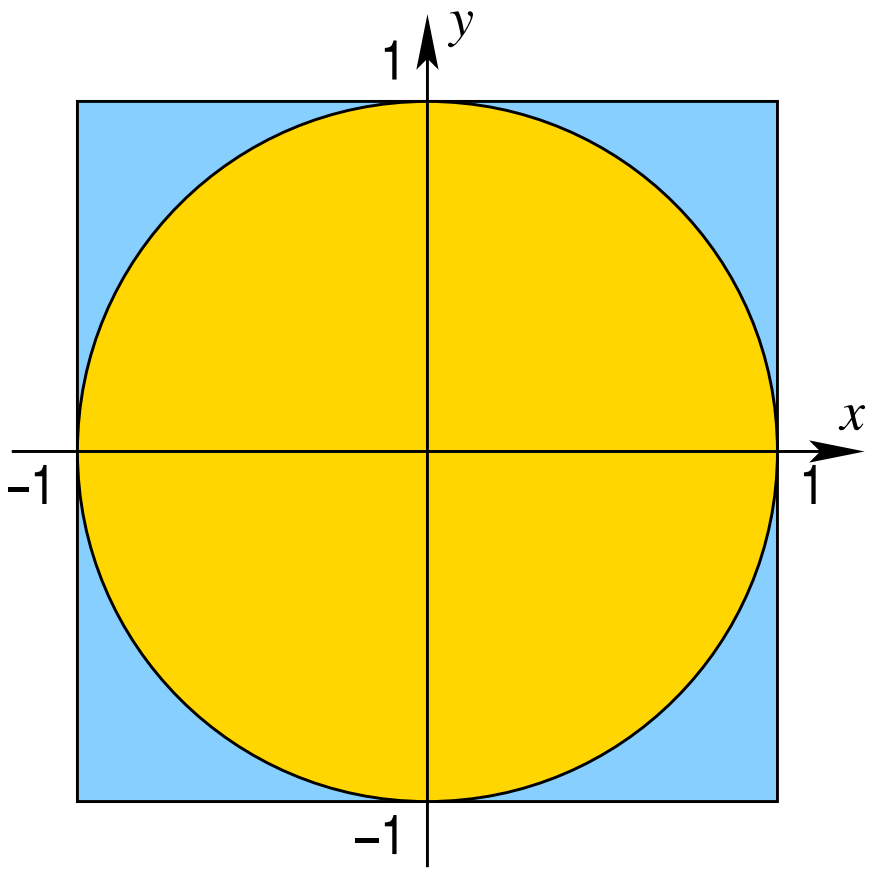
Idea



Choose a square-shaped underlying grid with
Cartesian coordinates, centered point of origin, unit length 1

Toy Example: Approximation of Constant $\pi \approx 3.14$

Idea

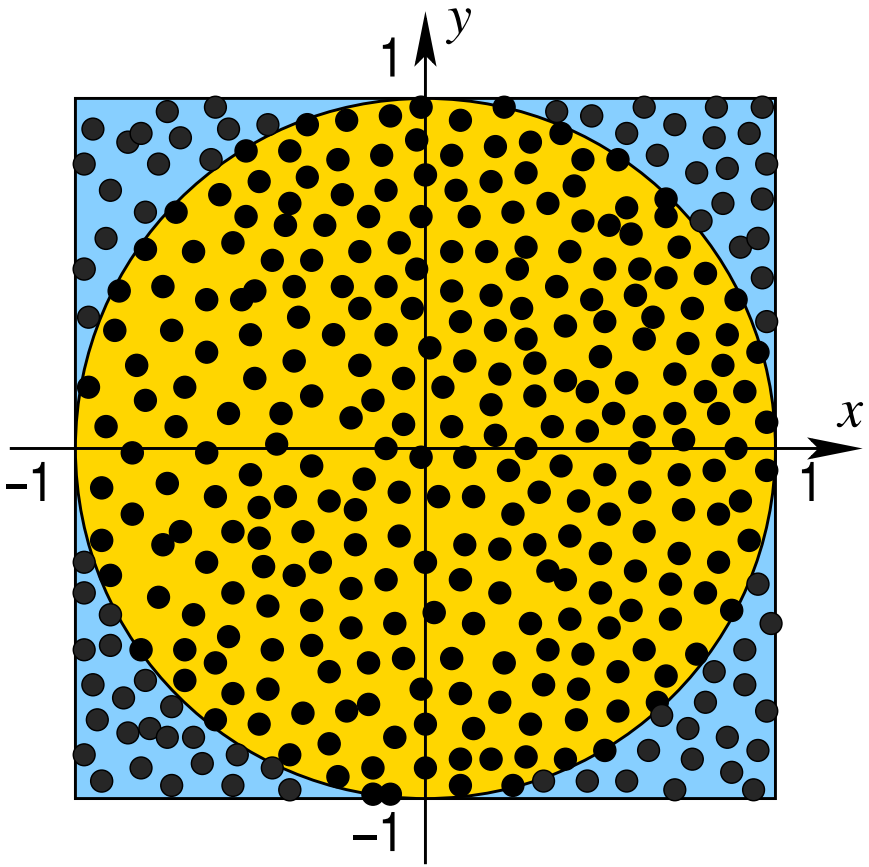


Inscribe a circle with radius 1

Circle and square form overlapping categories on the grid

Toy Example: Approximation of Constant $\pi \approx 3.14$

Idea

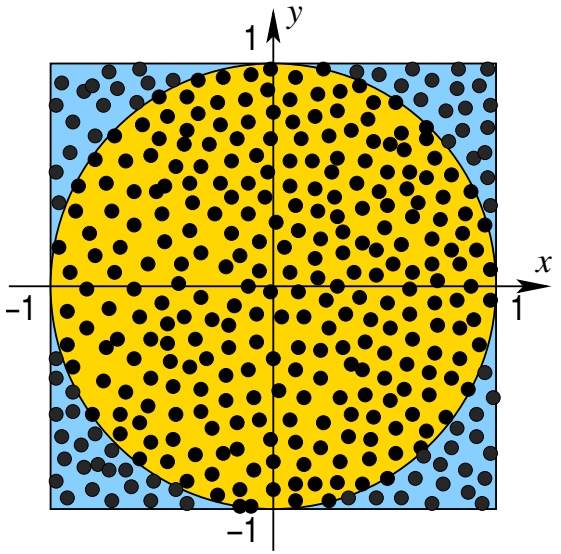


Place a huge number of **particles** on the grid randomly in spatial homogeneity

Toy Example: Approximation of Constant $\pi \approx 3.14$

Algorithmic design

- Circle with radius 1 covers area of $\pi = 3.14159265\ldots$
- Square constitutes 4 surface units on the grid
- Number of particles acts as measured value for circle area and square area
- As an approximation, we obtain:



$$\frac{\pi}{4} = \frac{\text{number of particles placed within the circle}}{\text{number of particles in total on the whole grid}}$$

Toy Example: Approximation of Constant $\pi \approx 3.14$

Blotting Membrane System

$$
\begin{aligned}
\boxed{\Pi} &= (P, L, C, B_1, \ldots, B_{|C|}, S, R, \mathrm{r}) \quad \text{with} \\
L &= \{l\} \\
P &= \{(0.70191, -0.21355, l), \ldots, (-0.45160, 0.52241, l)\} \\
C &= \{\odot, \boxdot\} \\
B_{\odot} &= \{(x, y, l) \mid (x, y, l) \in P \wedge x^2 + y^2 \leq 1\} \\
B_{\boxdot} &= \{(x, y, l) \mid (x, y, l) \in P \wedge |x| \leq 1 \wedge |y| \leq 1\} \\
S(c) &= |B_c| \; \forall c \in C \\
R &= \mathbb{R} \\
\mathrm{r}(S) &= 4 \cdot \frac{S(\odot)}{S(\boxdot)}
\end{aligned}
$$

Toy Example: Approximation of Constant $\pi \approx 3.14$

Results

$\lvert P \rvert$	$S(\odot)$	$S(\boxdot)$	rational approx. r of π
10,000	7,928	10,000	3.1712... (2 reliable digits)
1,000,000	785,502	1,000,000	3.1421... (3 reliable digits)
100,000,000	78,542,447	100,000,000	3.1417... (4 reliable digits)

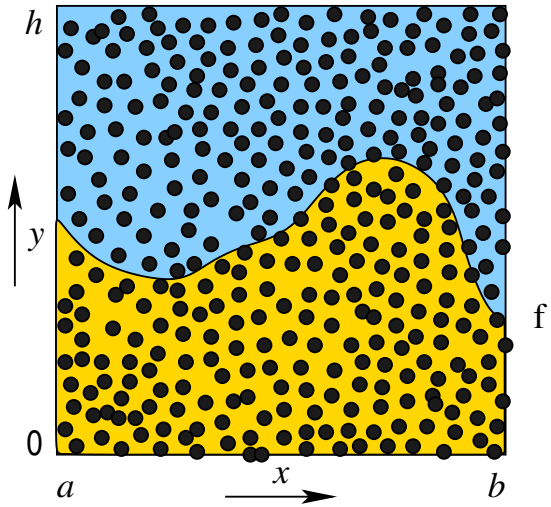
- Ascending number of particles $\rightarrow$ higher accuracy of the approximation
- 100-fold increase of the total particle number to obtain one additional reliable digit
- Slow convergence behaviour due to two-dimensional nature of experimental setting
- Numerical precision of particle coordinates needs to be adapted as well if needed

1. Motivation and Principle of Blotting
2. Blotting Membrane Systems
3. **Particle-based Numerical Integration**
   - Approximate Definite Integral
   - Periodical Cicada's Life Cycle
4. Electrophoresis: A Molecular Bucket Sort

Enhancing Previous Idea for Numerical Integration

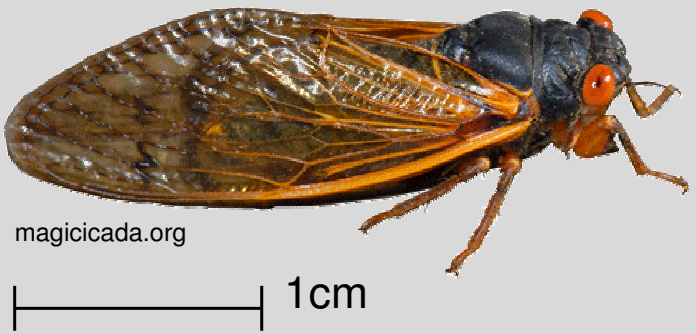
Algorithmic design

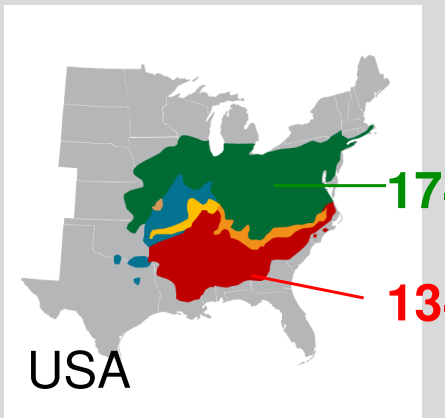
- Area below real-valued function $\mathrm{f} : \mathbb{R} \longrightarrow \mathbb{R}_+$ to be integrated numerically within range $[a, b]$
- Grid forms rectangle by height h and width $b - a$
- Number of particles acts as measured value for area
- As an approximation, we obtain:

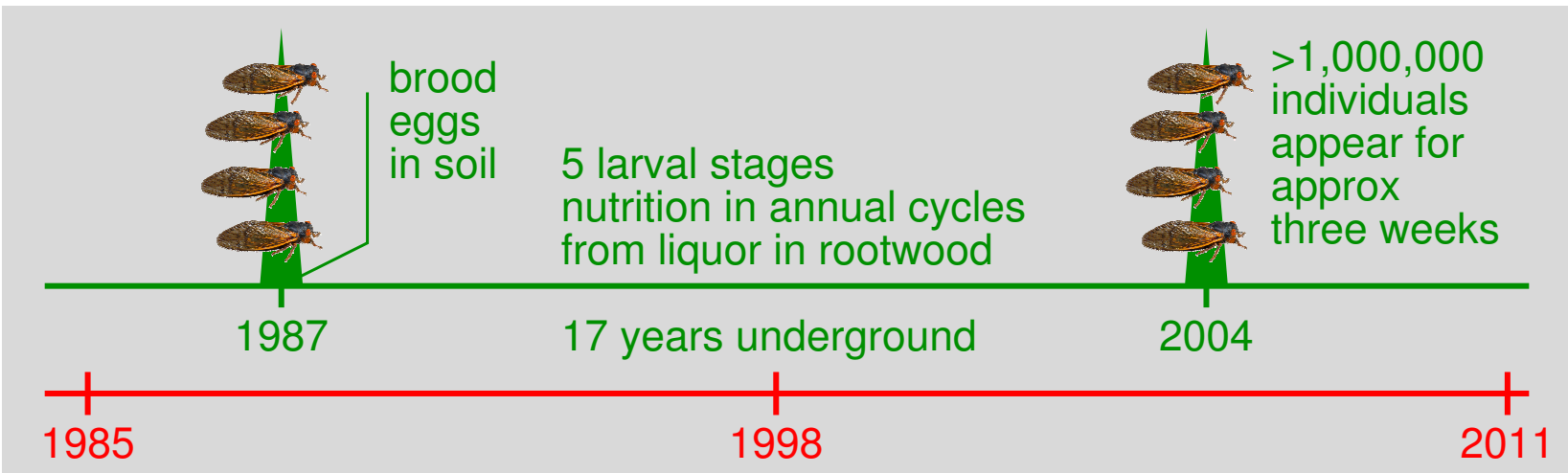


$$\frac{\int_a^b \mathrm{f}(x)\,dx}{h \cdot (b-a)} = \frac{\text{number of particles placed below the function course of f}}{\text{number of particles in total on the whole grid}}$$

17-years and 13-years Periodical Cicadas with Synchronous Life Cycle







How Do Cicadas Estimate Period of 17 or 13 Years?

- No external stimulus with natural period of 17 or 13 years known up to now

How Do Cicadas Estimate Period of 17 or 13 Years?

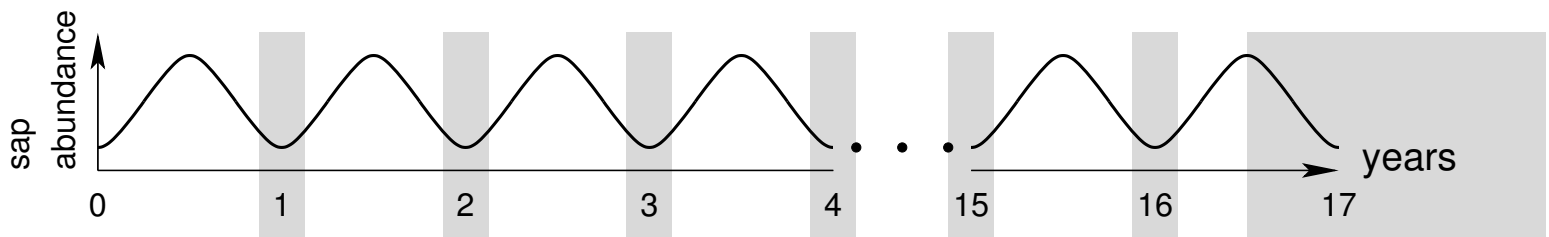
- No external stimulus with natural period of 17 or 13 years known up to now
- Molecular mechanism to precisely measure the passage of 17 (or 13) years?

How Do Cicadas Estimate Period of 17 or 13 Years?

- No external stimulus with natural period of 17 or 13 years known up to now
- Molecular mechanism to precisely measure the passage of 17 (or 13) years?
- How to keep mechanism simple and evolutionary adaptable to variety of time periods?

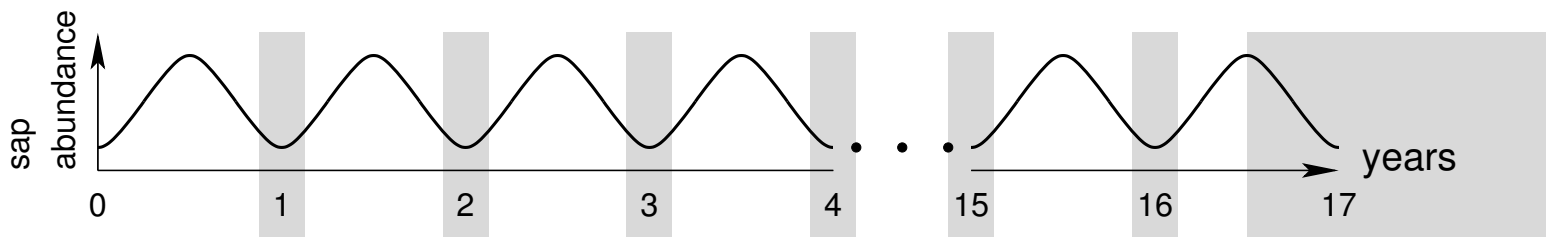
How Do Cicadas Estimate Period of 17 or 13 Years?

- No external stimulus with natural period of 17 or 13 years known up to now
- Molecular mechanism to precisely measure the passage of 17 (or 13) years?
- How to keep mechanism simple and evolutionary adaptable to variety of time periods?
- Possible annual stimulus: availability of sap in ambient root capillares



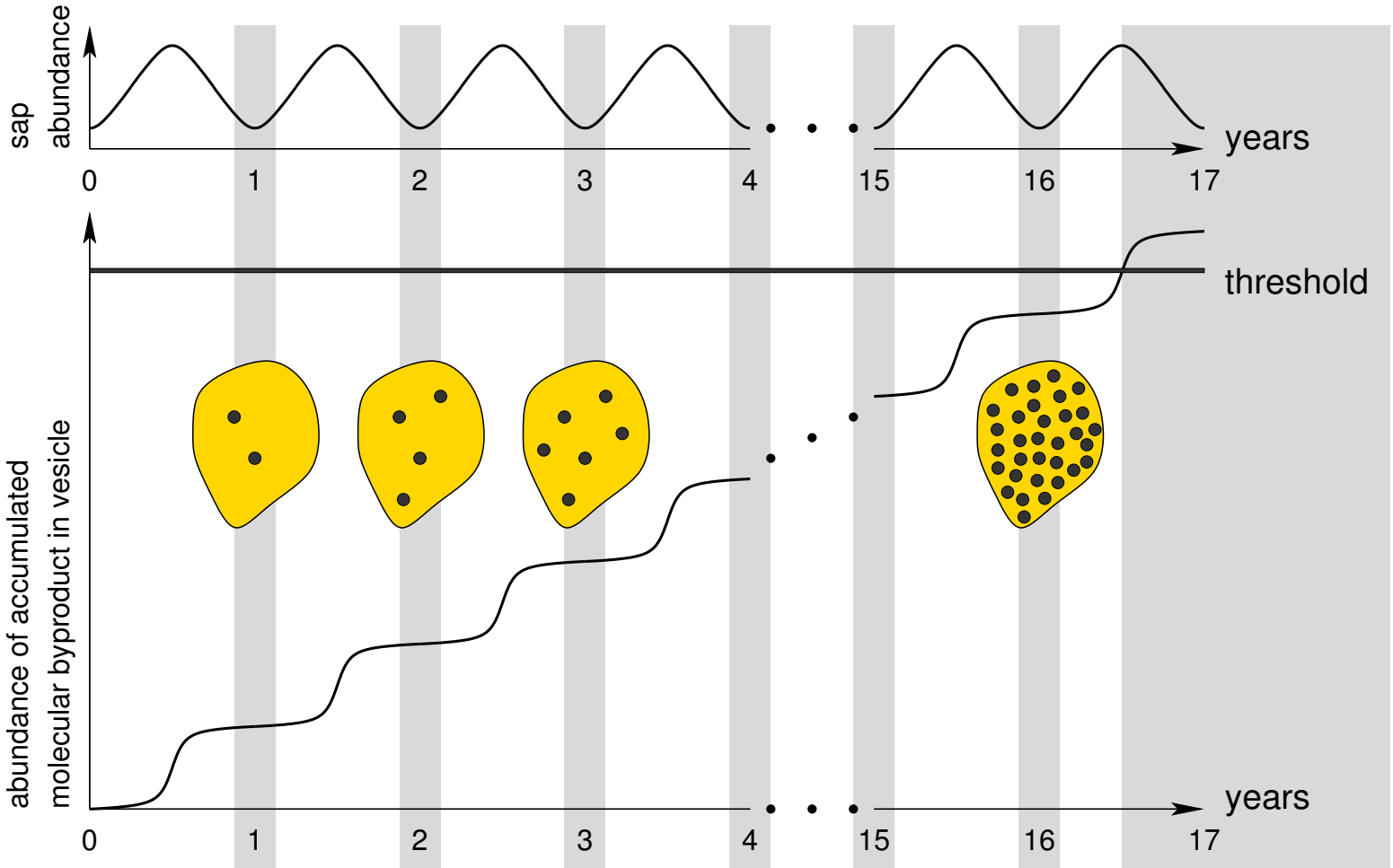
How Do Cicadas Estimate Period of 17 or 13 Years?

- No external stimulus with natural period of 17 or 13 years known up to now
- Molecular mechanism to precisely measure the passage of 17 (or 13) years?
- How to keep mechanism simple and evolutionary adaptable to variety of time periods?
- Possible annual stimulus: availability of sap in ambient root capillares



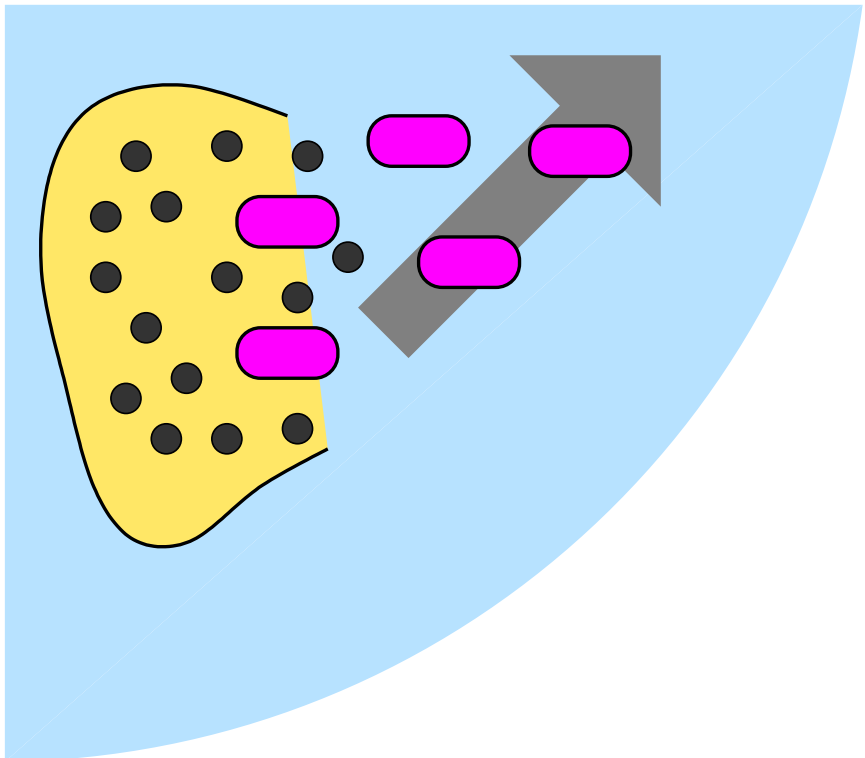
Speculative idea: chemical integrator equipped with a threshold for final alert

Accumulation of Byproduct Molecules in a Vesicle



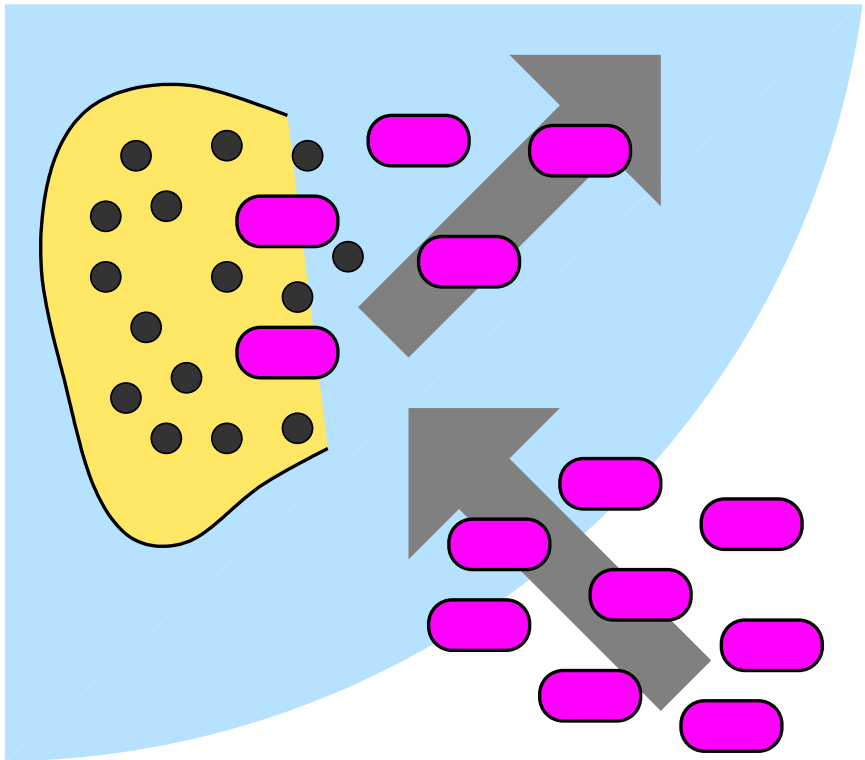
Byproduct molecules successively from metabolism in sap digestion

Threshold Exceeded: Release of Trigger Molecules



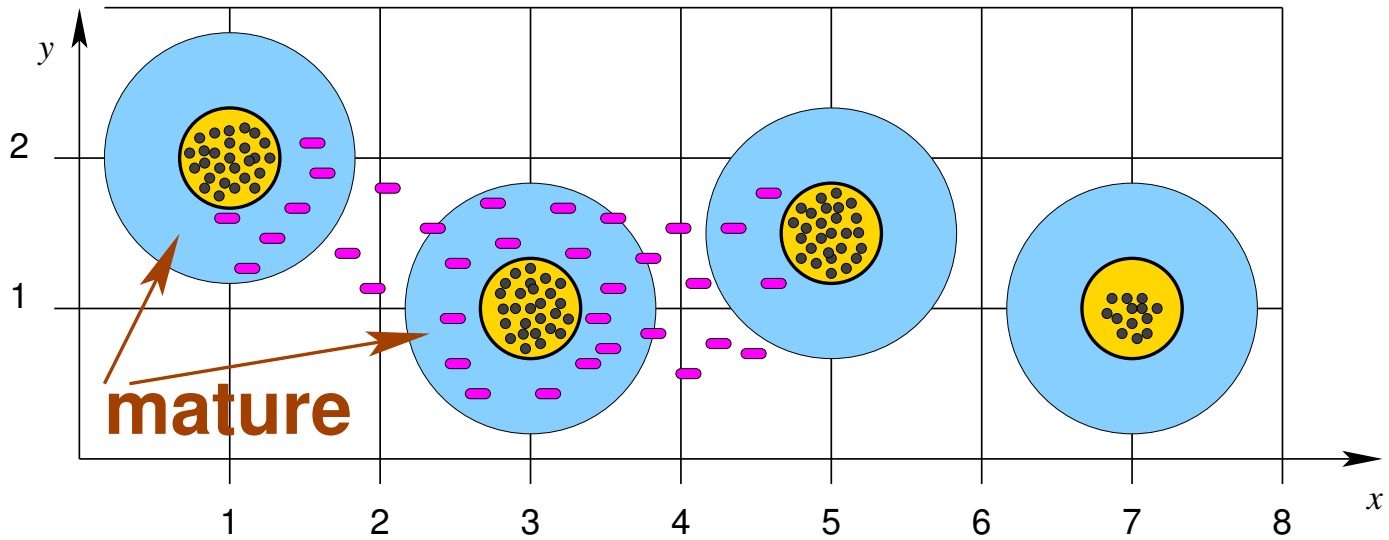
- Vesicle burst
- Generation and release of trigger molecules into environment
- Trigger molecules mark blots in the soil locating mature cicadas

Synchronisation Fine Tuning by Trigger Molecules



- Cicada also perceives trigger molecules from others
- Possible special form of *quorum sensing*
- Blotting membrane system identifies number of mature cicadas from an fluorescence image of the soil

Blotting Membrane System for Image Evaluation

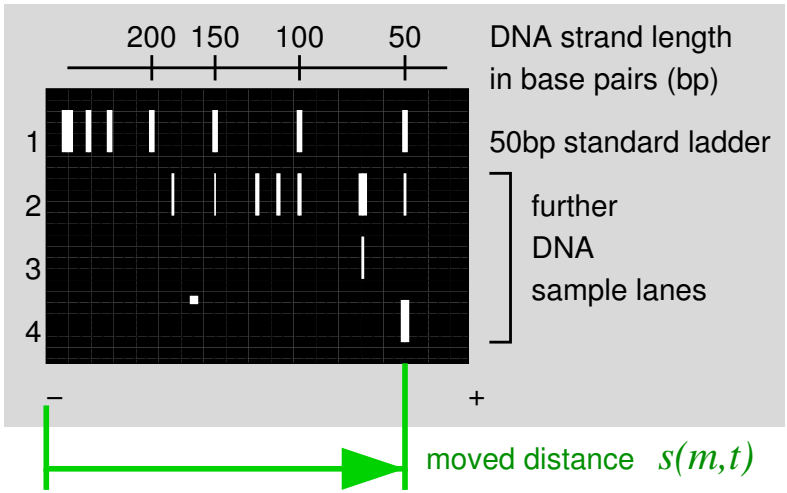


- Plan view of soil with fluorescent molecules taken as grid
- Identification of mature cicadas by corresponding blots
- Response: number of mature cicadas (here 2 out of 4)
- For detailed system's description please see paper

1. Motivation and Principle of Blotting
2. Blotting Membrane Systems
3. Particle-based Numerical Integration
4. **Electrophoresis: A Molecular Bucket Sort**
   - Principle and Gel Image Generation
   - Image Evaluation

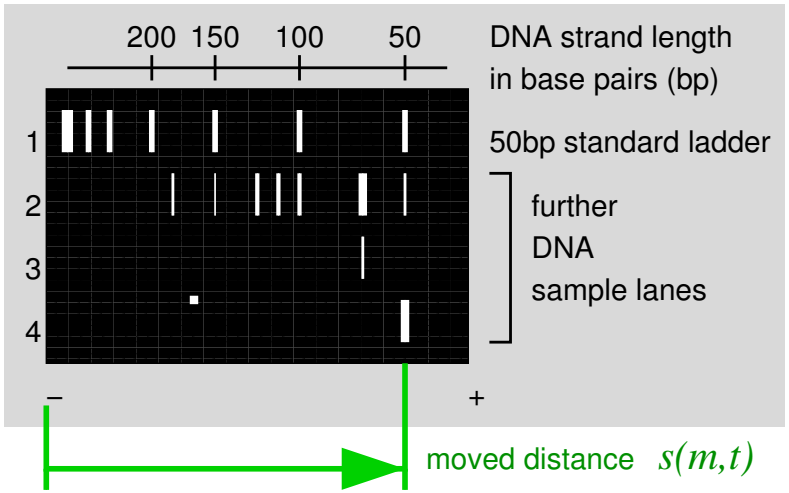
Electrophoresis: A Molecular Bucket Sort

Spatial Separation of Electrically Charged Molecules like DNA by Weight



Electrophoresis: A Molecular Bucket Sort

Spatial Separation of Electrically Charged Molecules like DNA by Weight



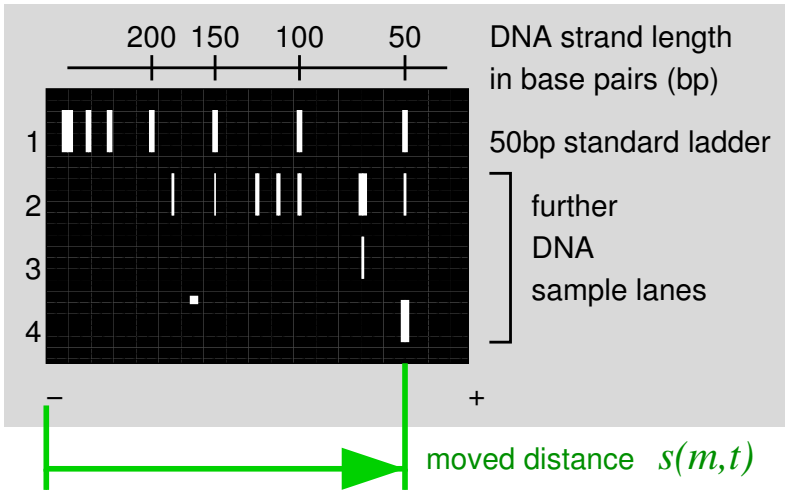
$$s(m,t) = G \cdot \frac{E}{\eta} \cdot \frac{1}{m^{\frac{1}{3}}} \cdot t$$

t . . . . . elapsed time

m . . . . . weight of individual charged molecule ($\sim$ DNA strand length)

Electrophoresis: A Molecular Bucket Sort

Spatial Separation of Electrically Charged Molecules like DNA by Weight



$$s(m,t) = G \cdot \frac{E}{\eta} \cdot \frac{1}{m^{\frac{1}{3}}} \cdot t$$

t . . . . . elapsed time

m . . . . . weight of individual charged molecule ($\sim$ DNA strand length)

Further parameters for gel running conditions

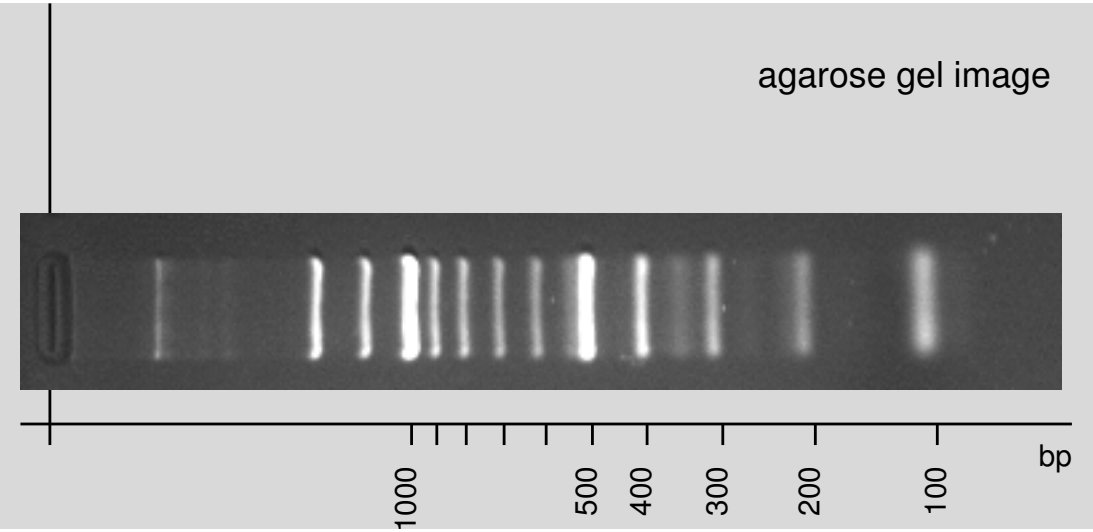
E . . . electrical field from DC voltage and distance between electrodes

η . . . . . viscosity, average pore size and density in gel

G . . . . . global zooming factor

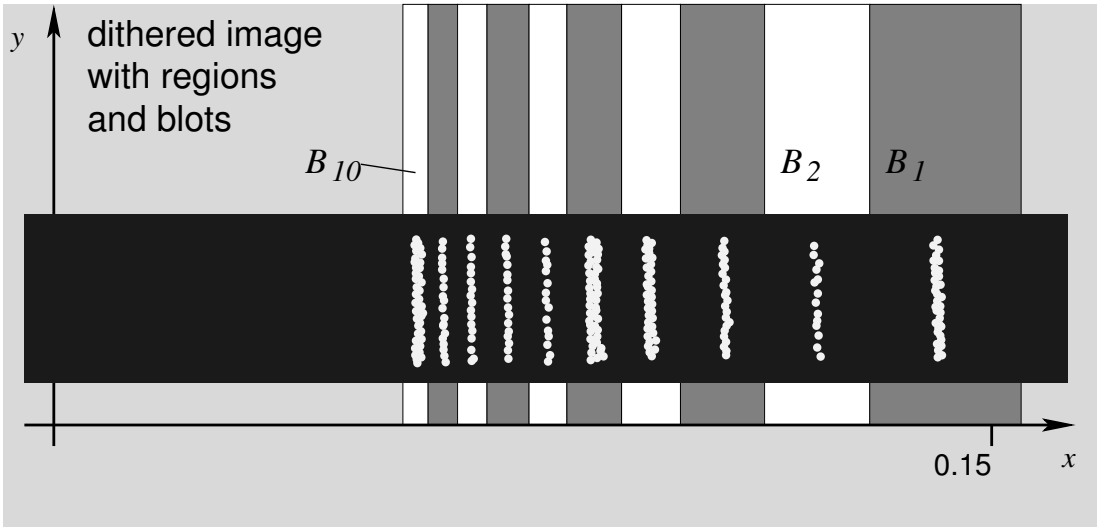
Function s derived from proximated parity between friction to be overcome by electrical force

An Electrophoresis Gel Image for Evaluation



- Gel image resulting from DNA separation by strand lengths
- Need of automatic image evaluation
- DNA strand lengths present in sample?
- Typical application scenario in bioinformatics
- Utilisation of a blotting membrane system

Image Evaluation using Blotting Membrane System



- *Dithering* of gel image produces variety of individual dots
- Each dot comes with grid coordinates (x, y)
- Definition of buckets corresponding to DNA bands (clusters)
- Each dot considered as a particle for categorised counting
- Blotting membrane system's response provides buckets filling

Take Home Message

Blotting membrane systems developed mainly to support image evaluation in bioinformatics and molecular biology

Further work

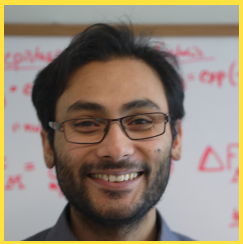
- System extension in order to capture *dynamics*
- Incorporation of distinct classification methods
- Exploit computational capacity
- Provide larger pool of advantageous algorithms

Acknowledgements to my Team Colleagues

Konrad Gruetzmann

Benny Hoeckner

Peter Sauer



Sikander Hayat