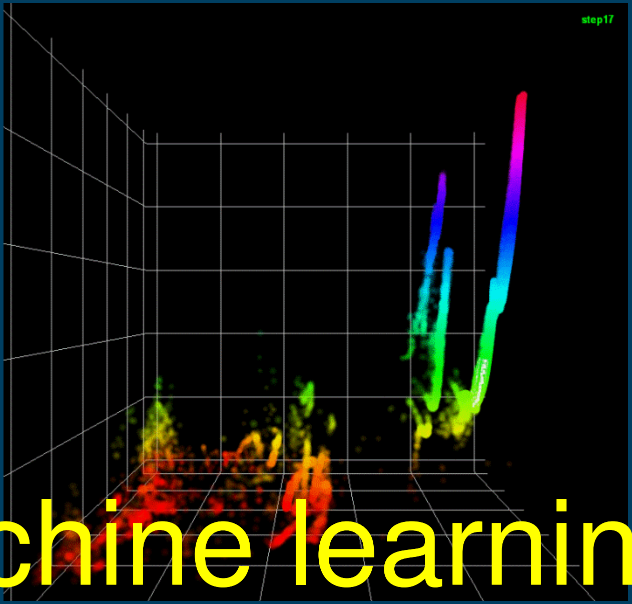


Machine learning on accelerator simulation data

Daniela M. Ushizima
Lawrence Berkeley National Laboratory,

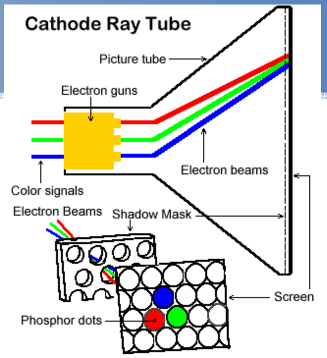
Participants**:** E. Wes Bethel, Prabhat, Oliver Rubel, Gunther H. Weber,
Cameron G.R. Geddes, Estelle Cormier-Michel,
Bernd Hamann, Peter Messmer, Hans Hagen

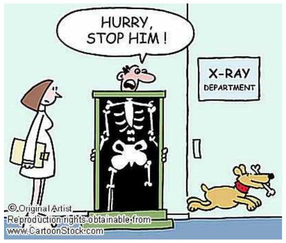
Overview

- **Background:** manual exploration of particles subjected to acceleration, given space and energy variables.
- **Goal:** use machine learning to automate detection of compact (*highest energy*) group of particles in simulations;
- **Material:** millions of particles in plasma under electro-magnetic field;
- **Contribution:** automated framework to select highly accelerated particles exhibiting spatial coherence:
  - Bunches of electrons per time step
  - Lifetime diagram representation
  - Fuzzy clustering to detect high density hypervolume

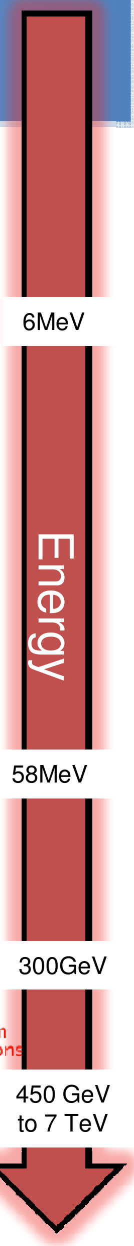
1. Why is it important?

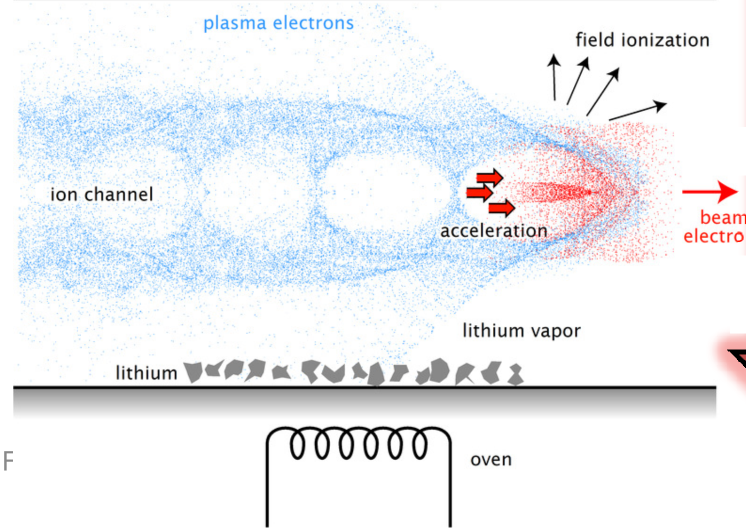
- Particle accelerators:
  - Low energy
  - High energy: Slac, LEP, LHC
- Plasma acceleration:
  - LWFA: compact source of high-energy electron beams and radiation (3m to 5cm);
  - new technology;
  - applications:
    - proton therapy (cancer),
    - material characterization,
    - radiation-driven chemistry,
    - high-energy particle physics.





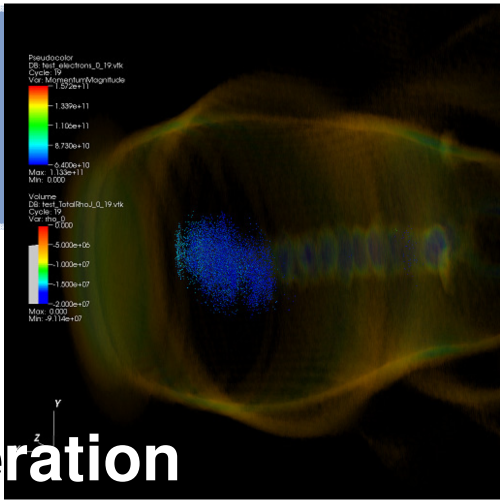
©Original Artist
Reproduction rights obtainable from
www.CartoonStock.com





http://en.wikipedia.org/wiki/Plasma_acceleration

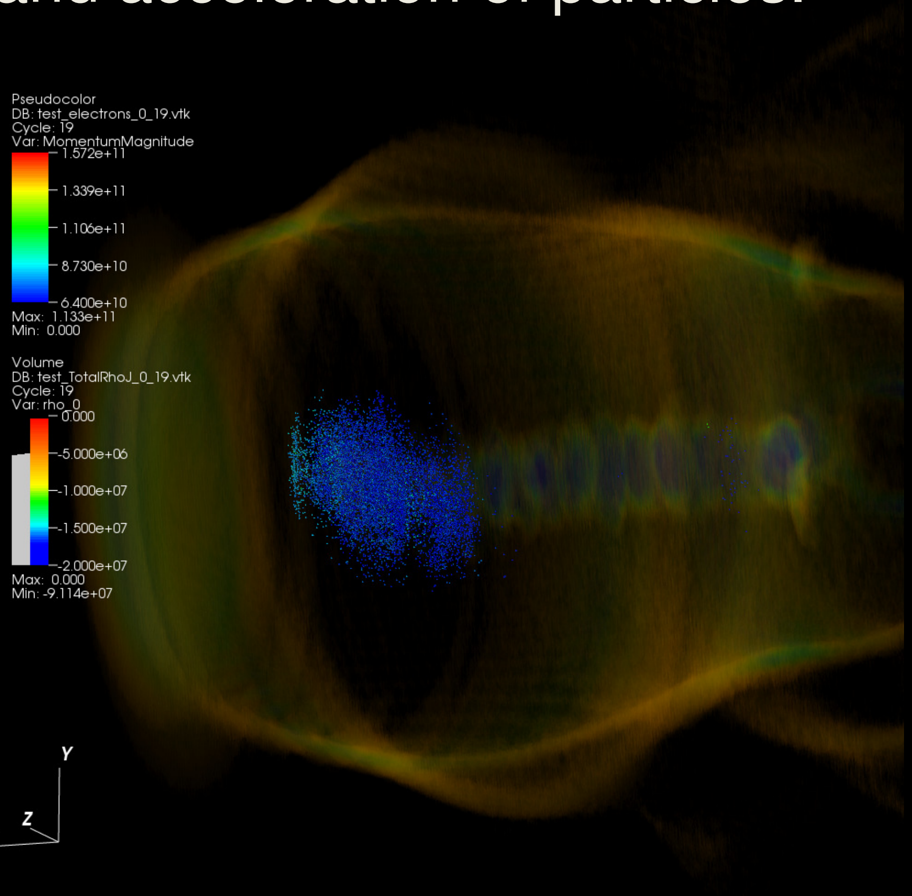
2. The physics of surfing



Human surfing	**Particle acceleration**
1. High waves	1. Plasma wakefield
2. Surfer	2. Particles
3. Surfer paddles (or jet ski) to match the speed of the wave if want to ride with it	3. Particles in “the right spot” of the wakefield can be accelerated to very high energies
4. Surfer drop down the front of the wave, so the gravitational potential energy it gains is converted into kinetic energy.	4. Particles surf the wake due to the electric field of the plasma, particles can be self-trapped and accelerated*;*

3. Particle acceleration

- Phenomenon of interest: trapping and acceleration of particles.



1. When the particles “catch the wave”, the electrons are deflected, pulled back to the center and pile up;
2. Particles are accelerated by the electric field of the plasma wave (wake);
3. After outrunning the wave they form a monoenergetic electron bunch;

4. Variables under investigation in LWFA

Dataset	Particles (10^6)	Timesteps	Total Size (GB)
A	0.4	37	1.3
B	1.6	35	4.5
C	0.4	37	1.3
D	3.2	45	11

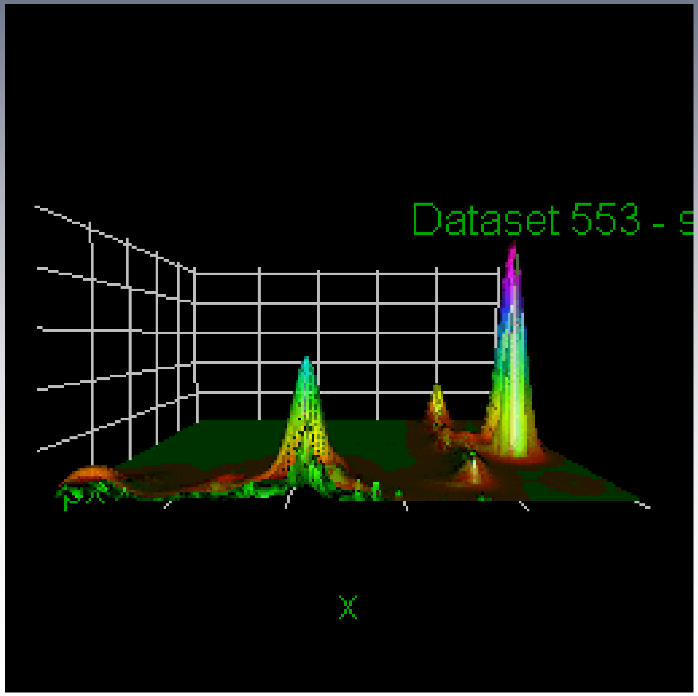
- Manual exploration on large datasets
- … to be larger –> 3D simulation ()
- Variables:
  - Spatial: x,y (m)
  - Momentum: px,py (MeV/C)

Cluster: how to *organize* observed data into meaningful structures?

- Tryon (1939): encompasses a number of different methods for grouping objects of similar kind into categories;
- BC TRY system for multidimensional analysis;
- Cluster analysis: sort different objects into groups, such that the degree of association between two objects is <u>maximal if they belong to the same group</u> and minimal otherwise.
- Useful to discover structures in data without providing an explanation/interpretation.
- Cluster analysis simply discovers structures in data without explaining why they exist.

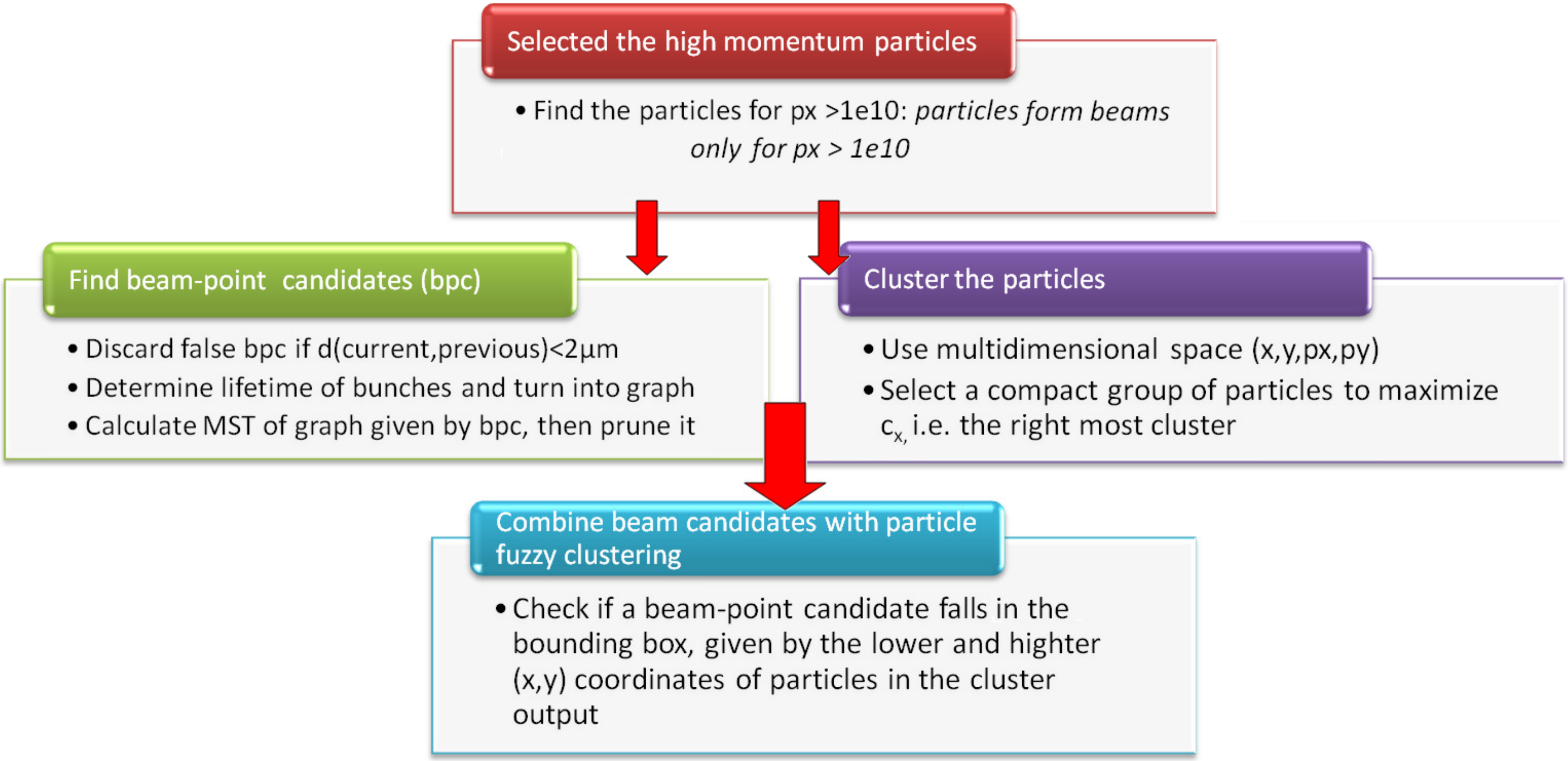
5. Software: data & tools

- Vorpal 2D simulation data:
  - Particle-in-cell simulation;
  - Parameters selected by the physicists;
- R-project statistical framework:
  - Hdf5 format reader;
  - Machine learning libraries;
  - Visualization libraries;
  - Tinn: http://www.sciviews.org/Tinn-R/



Pipeline

Proposed algorithm for beam detection



Selected the high momentum particles

- Find the particles for px >1e10: *particles form beams only for px > 1e10*

Find beam-point candidates (bpc)

- Discard false bpc if d(current,previous)<2µm
- Determine lifetime of bunches and turn into graph
- Calculate MST of graph given by bpc, then prune it

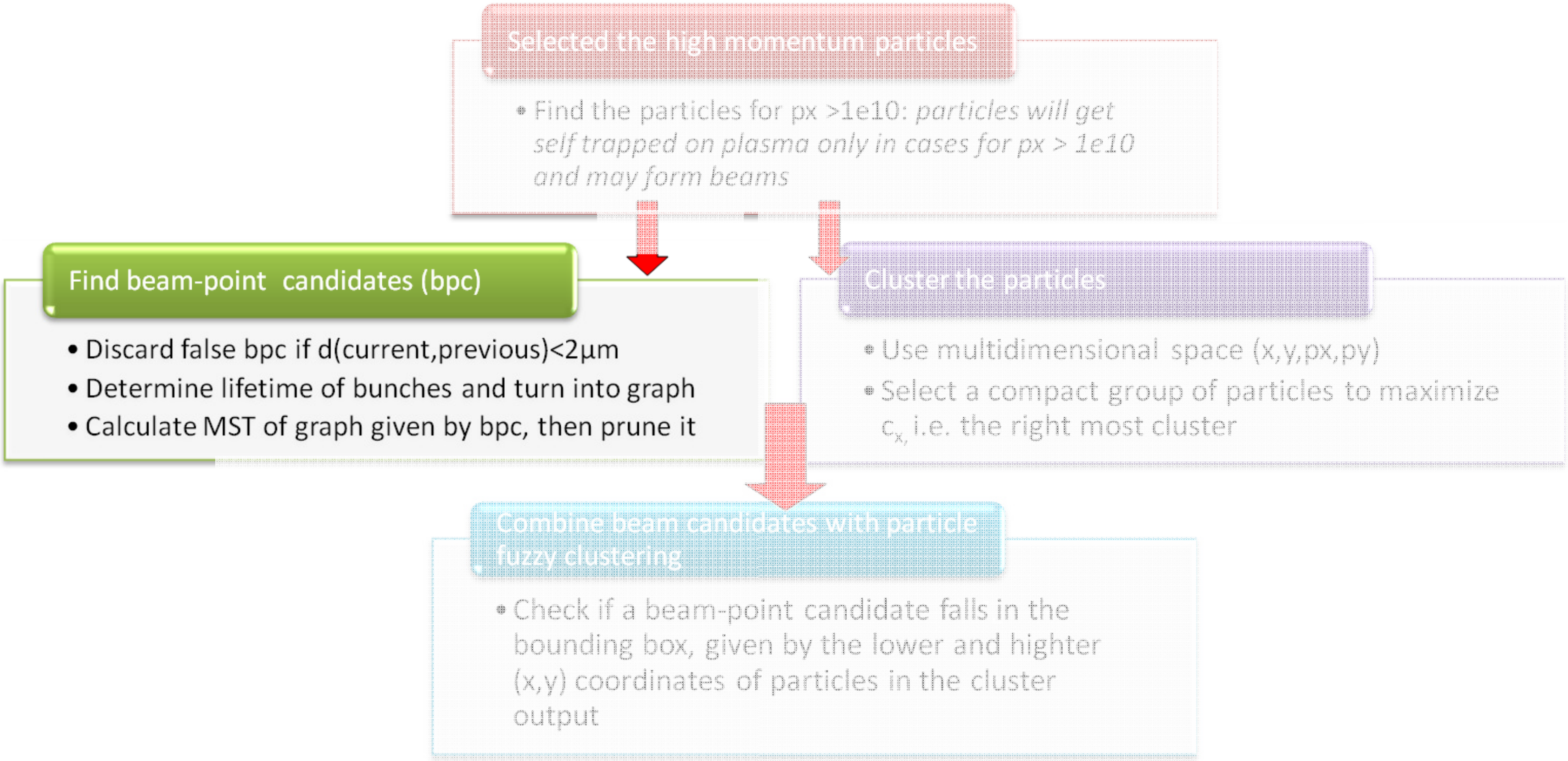
Cluster the particles

- Use multidimensional space (x,y,px,py)
- Select a compact group of particles to maximize $c_{x,}$ i.e. the right most cluster

Combine beam candidates with particle fuzzy clustering

- Check if a beam-point candidate falls in the bounding box, given by the lower and highter (x,y) coordinates of particles in the cluster output

1. Find beam-point candidate



Selected the high momentum particles

- Find the particles for px >1e10: *particles will get self trapped on plasma only in cases for px > 1e10 and may form beams*

Find beam-point candidates (bpc)

- Discard false bpc if d(current,previous)<2µm
- Determine lifetime of bunches and turn into graph
- Calculate MST of graph given by bpc, then prune it

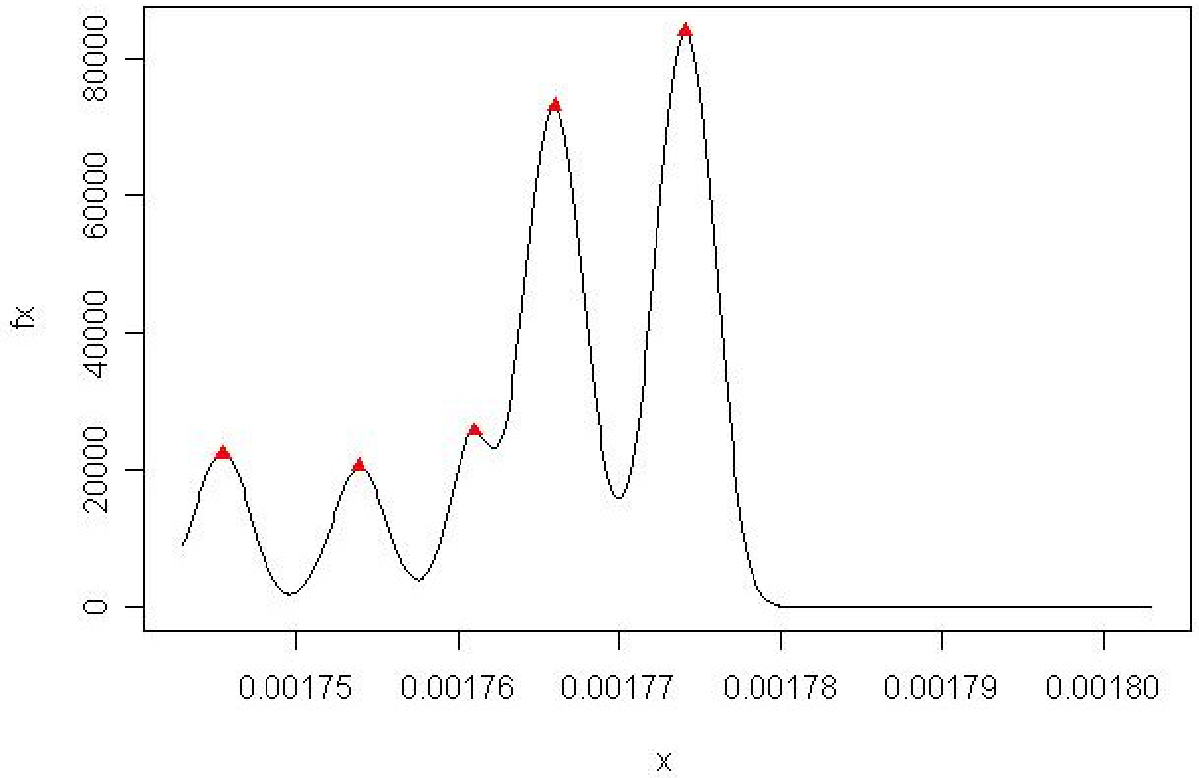
Cluster the particles

- Use multidimensional space (x,y,px,py)
- Select a compact group of particles to maximize c_x, i.e. the right most cluster

Combine beam candidates with particle fuzzy clustering

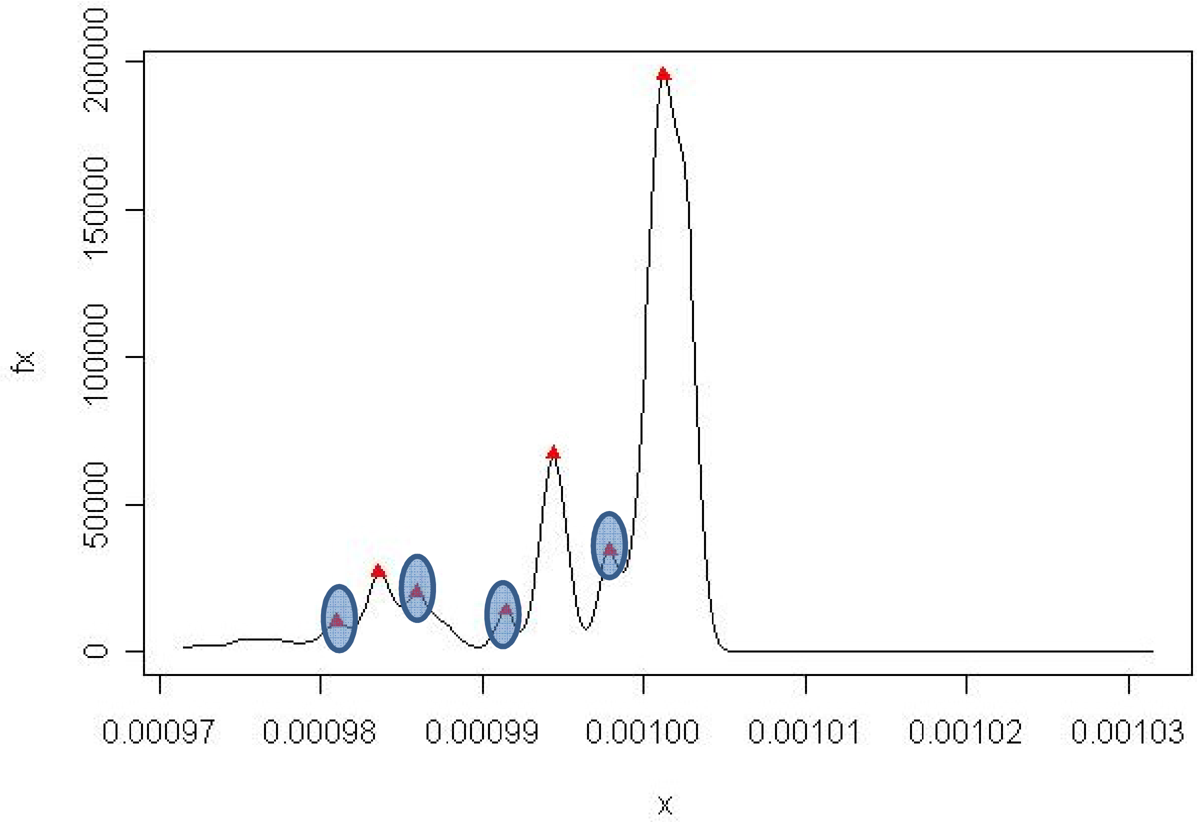
- Check if a beam-point candidate falls in the bounding box, given by the lower and highter (x,y) coordinates of particles in the cluster output

1.1. Find beam center candidates



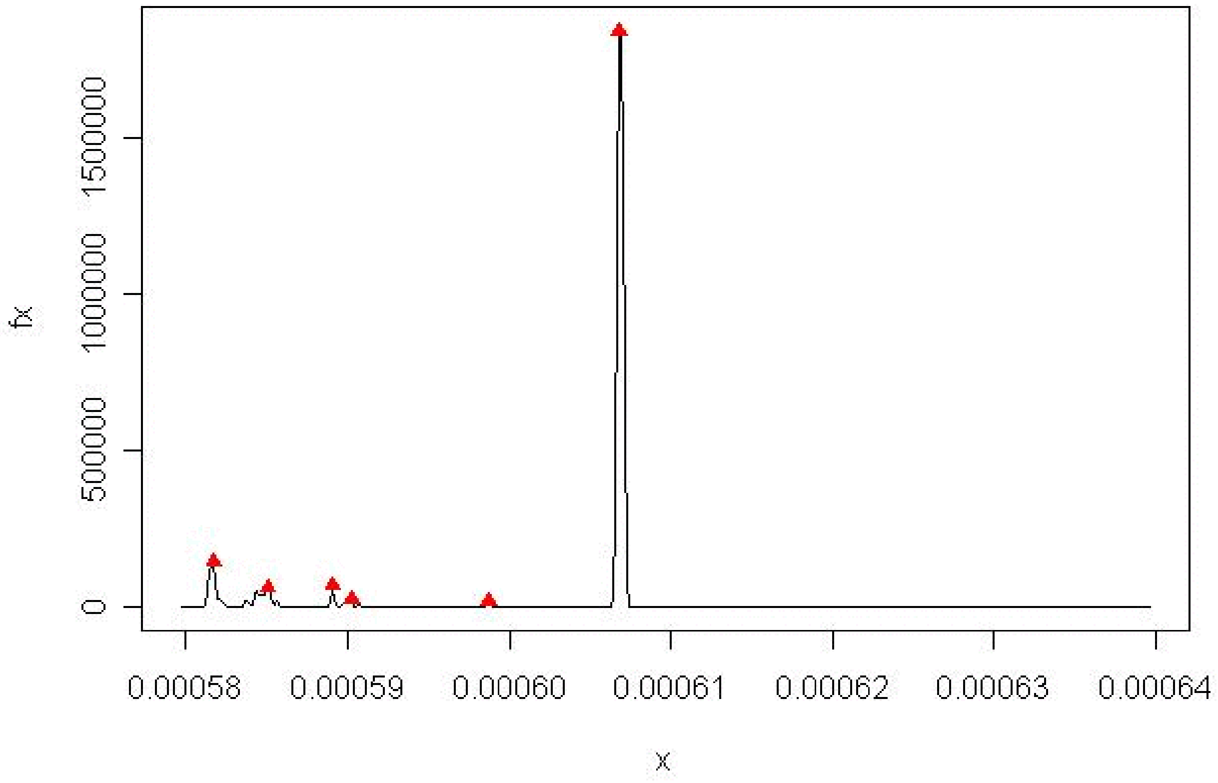
1. *Kernel smoothing*
2. *Zero crossing on the df/dx for f(x) from positive to negative*
3. *Assumption: peak>1st quartile of estimated pdf*

1.2. Decreasing false candidates

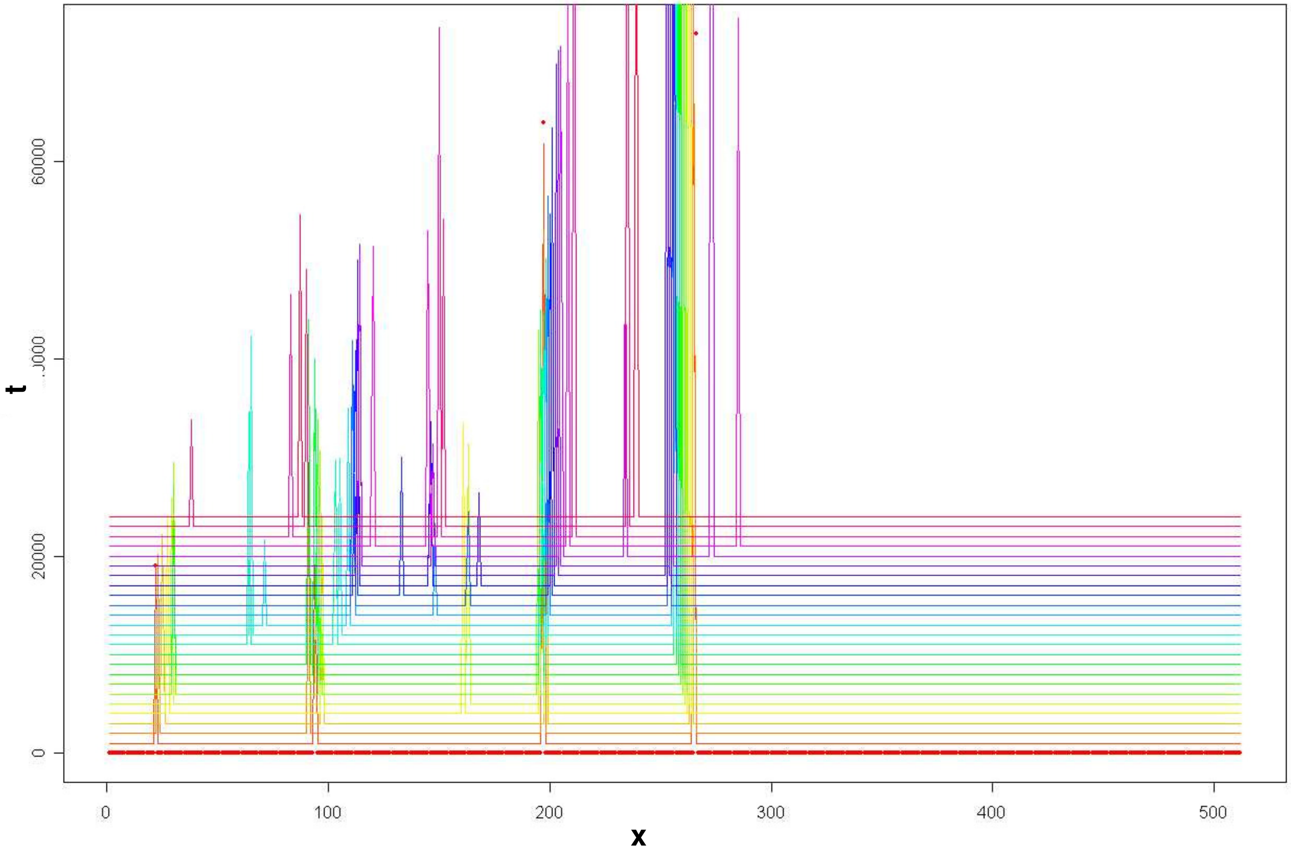


- Tolerance = peaks must be 2 microns apart, otherwise retains the most representative.

1.2.1. For all the time steps…



1.3. Stack of h(x,t)



Dataset A

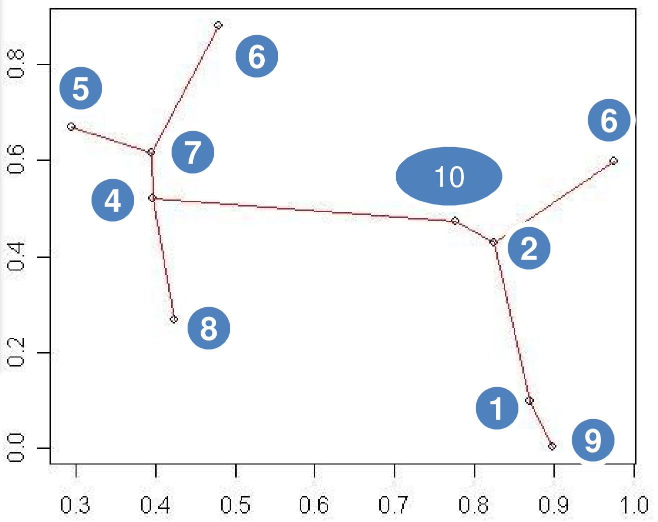
1.4. How to check for pairwise correlation?

- Calculate the incidence matrix
- MST
- Example:

```
Position of my nodes (x,t)
        [,1]        [,2]
[1,] 0.8687145 0.099536987
[2,] 0.8239852 0.427755543
[3,] 0.4784496 0.880953808
[4,] 0.3966939 0.520083743
[5,] 0.2954766 0.668804865
[6,] 0.9744532 0.599013541
[7,] 0.3948746 0.615403768
[8,] 0.4232120 0.267684618
[9,] 0.8975742 0.003440116
[10,] 0.7767071 0.473417633
```

→ Dist →

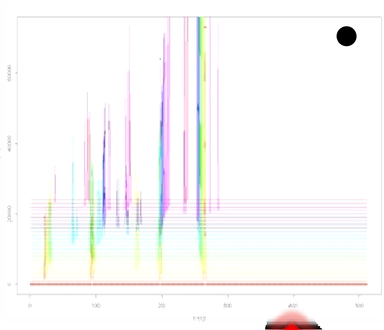
	1	2	3	4	5	6	7	8	9	10
1	0	1	0	0	0	0	0	0	1	0
2	1	0	0	0	0	1	0	0	0	1
3	0	0	0	0	0	0	1	0	0	0
4	0	0	0	0	0	0	1	1	0	1
5	0	0	0	0	0	0	1	0	0	0
6	0	1	0	0	0	0	0	0	0	0
7	0	0	1	1	1	0	0	0	0	0
8	0	0	0	1	0	0	0	0	0	0
9	1	0	0	0	0	0	0	0	0	0
10	0	1	0	1	0	0	0	0	0	0



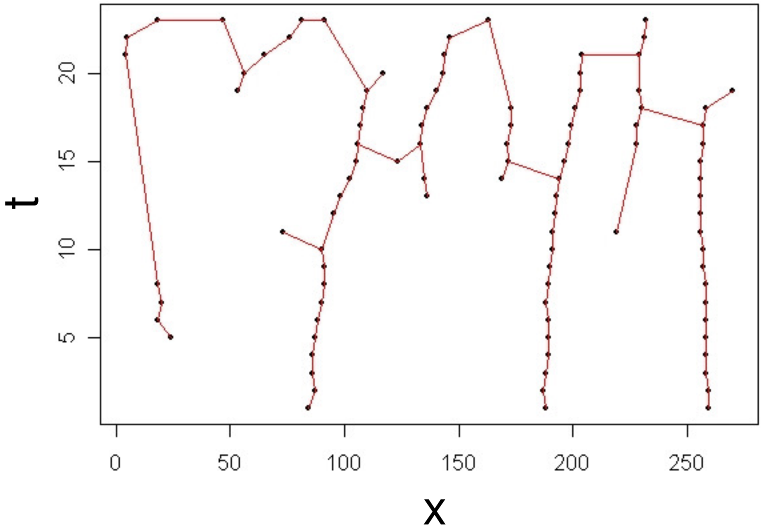
1.5. Hypothesis

- **“Beam life time is given by a pruned MST”**
- Short branches indicate beam consistency along times steps
- Pruning process:
  - find minimum cost subgraph G, such as subgraph:
    - Minimize distance between nodes;
    - Disconnect peaks in the same time step;
  - Orphan nodes are deleted.

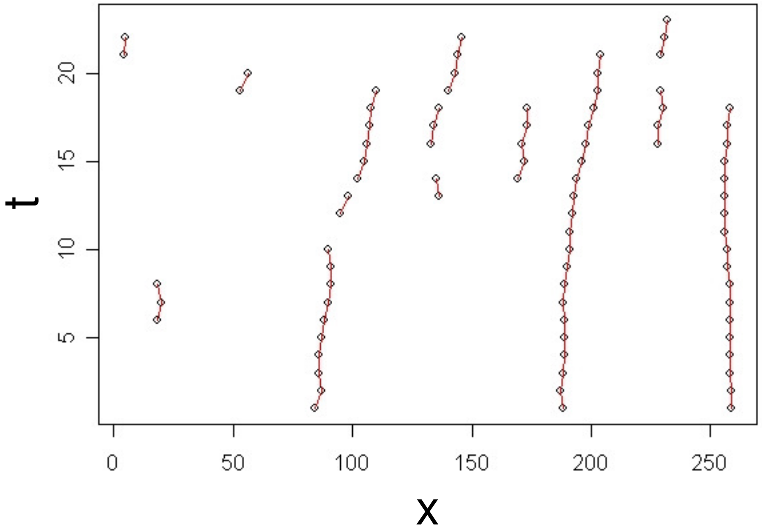
1.6. From graph to beam



- Lifetime representation: particle history as a pruned MST with likely branches and connected nodes as beam-point candidates.

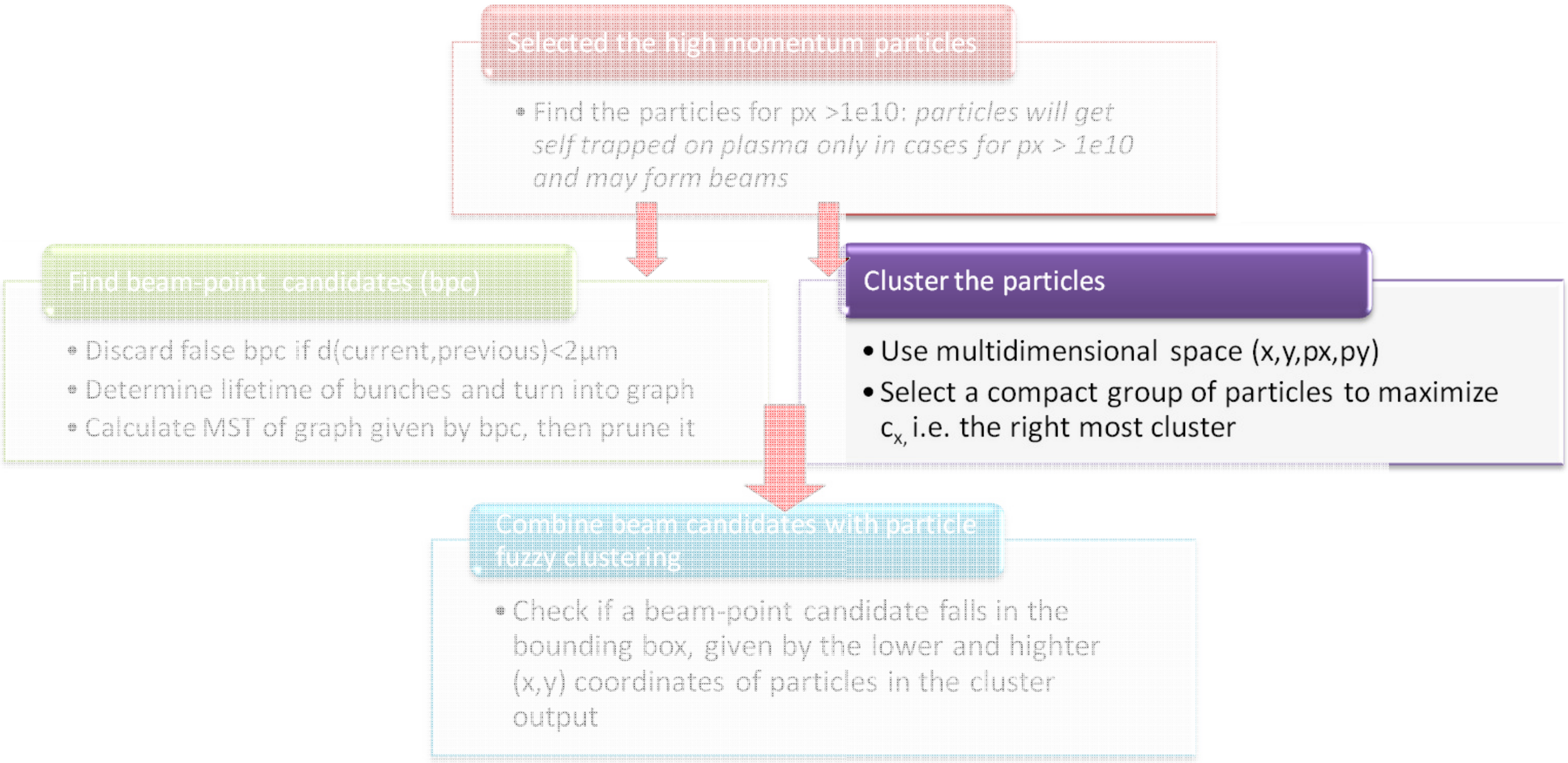


MST of particle density peaks as nodes: t from bottom to top



Candidates after prune1: dist

2.Cluster the particles



Selected the high momentum particles

- Find the particles for px >1e10: *particles will get self trapped on plasma only in cases for px > 1e10 and may form beams*

Find beam-point candidates (bpc)

- Discard false bpc if d(current,previous)<2µm
- Determine lifetime of bunches and turn into graph
- Calculate MST of graph given by bpc, then prune it

Cluster the particles

- Use multidimensional space (x,y,px,py)
- Select a compact group of particles to maximize $c_{x,}$ i.e. the right most cluster

Combine beam candidates with particle fuzzy clustering

- Check if a beam-point candidate falls in the bounding box, given by the lower and highter (x,y) coordinates of particles in the cluster output

2.1. Cluster analysis to LWFA

- How similar are the particles for each beam-point candidate?
  - – imprecise;
  - – absence of sharply defined criteria of class membership;
- Algorithm requirements:
  - – one beam formation, high *px*, compact in *x,y,px,py;*
  - – varying degrees of membership.

2.2.Fuzzy clustering

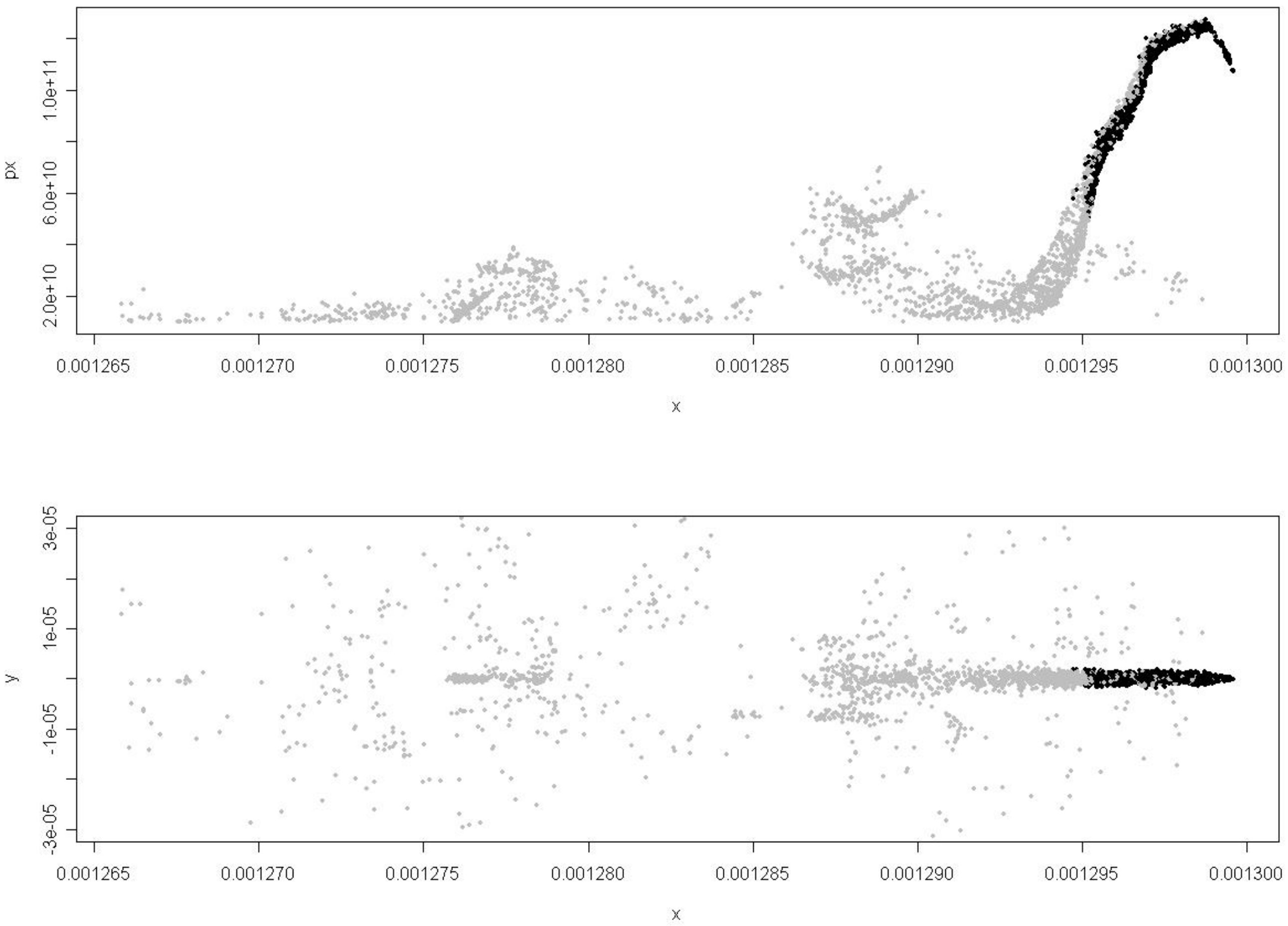
- Condition relaxation using fuzzy clustering
- Objective function: *argmin*[F(D,X,P,U)]
- Parameters:
  - Fuzzifier: 2
  - Number of clusters: 2
  - tolerance for convergence (relative convergence of the fit criterion) = 1e-10
  - Degree of membership = 70%
  - Squared Euclidean distance to place progressively greater weight on objects that are further apart.

p_i *in* R^4,
u_{ij} *in* $R^4 x R^4$

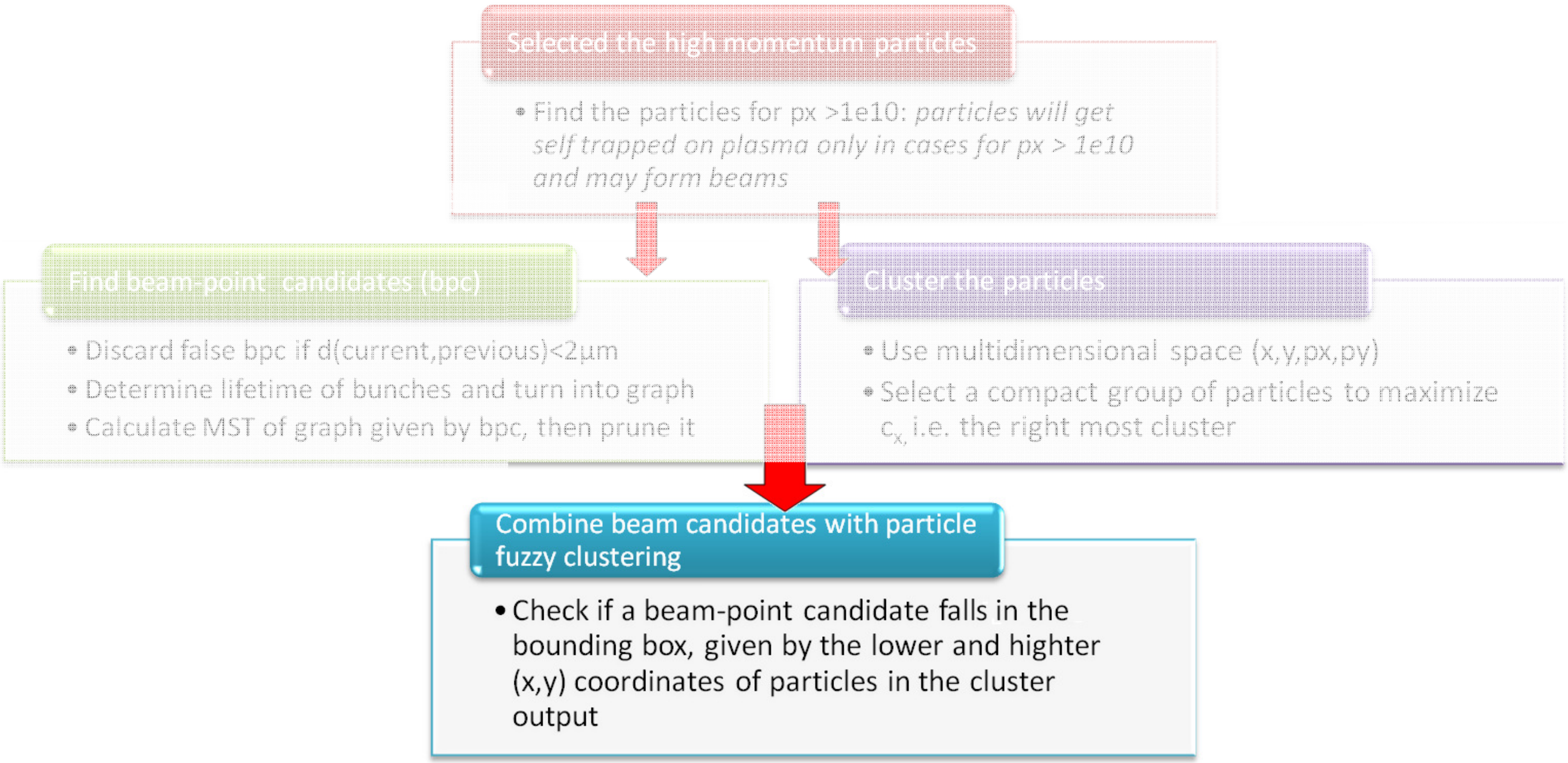
D = distance,
X = particle data,
P = prototypes,
U = partition matrix

$$F = \sum_{v=1}^{k} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} u_{iv}^{m} u_{jv}^{m} d(i,j)}{2 \sum_{j=1}^{n} u_{jv}^{m}}$$

2.3. Particles: beam or not-beam



3. Fusion probes with groups



Selected the high momentum particles

- Find the particles for px >1e10: *particles will get self trapped on plasma only in cases for px > 1e10 and may form beams*

Find beam-point candidates (bpc)

- Discard false bpc if d(current,previous)<2µm
- Determine lifetime of bunches and turn into graph
- Calculate MST of graph given by bpc, then prune it

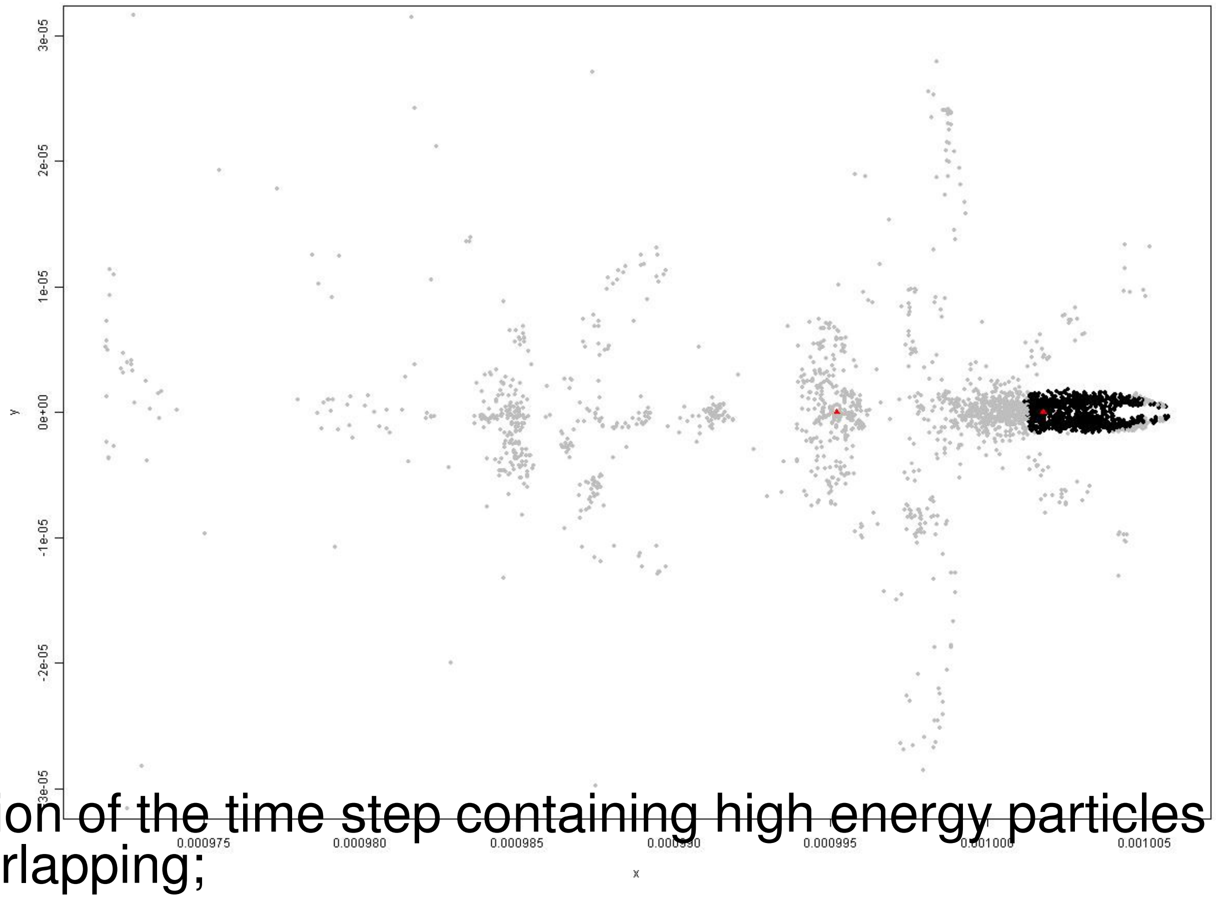
Cluster the particles

- Use multidimensional space (x,y,px,py)
- Select a compact group of particles to maximize $c_{x,}$ i.e. the right most cluster

Combine beam candidates with particle fuzzy clustering

- Check if a beam-point candidate falls in the bounding box, given by the lower and highter (x,y) coordinates of particles in the cluster output

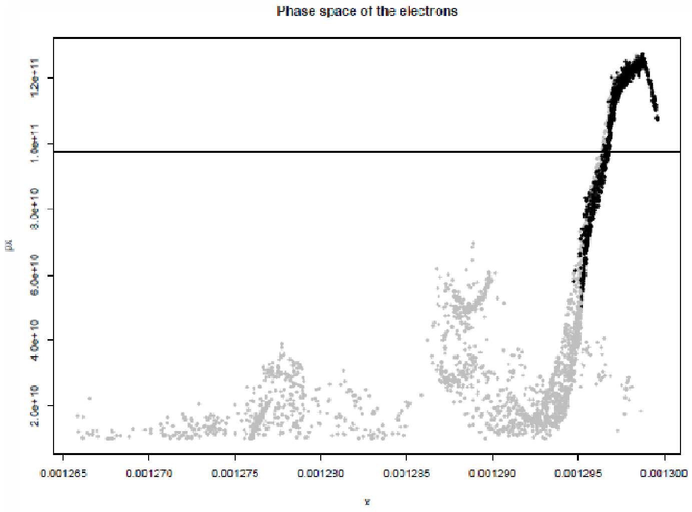
3.1. Combine beam candidates with particle fuzzy clustering



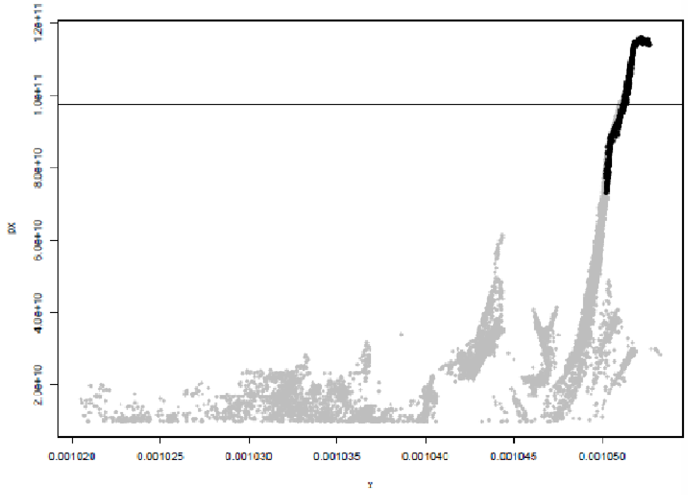
- Detection of the time step containing high energy particles by checking for overlapping;
- Estimation of a beam containing particles that behave similarly, according to their spatial coordinates and energy attributes

DU1

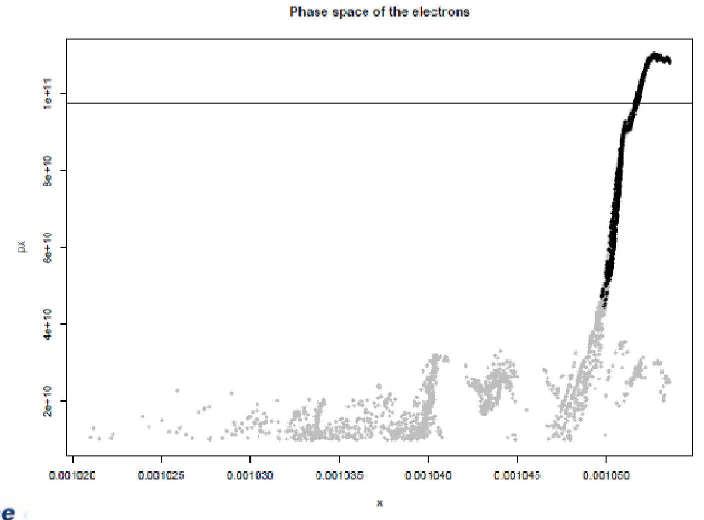
3.3. Steps with high energy particles



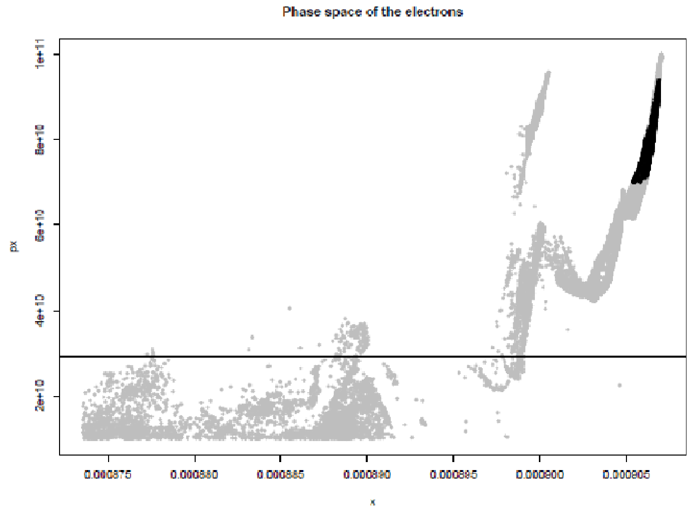
(a) Dataset A



(b) Dataset B



(c) Dataset C



(d) Dataset D

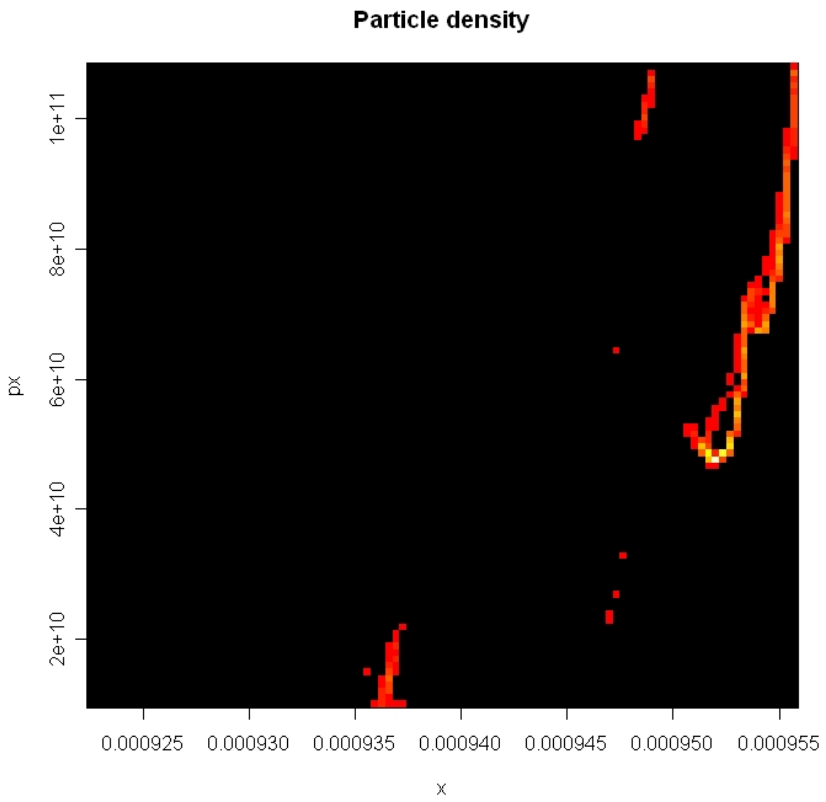
Slide 25

DU1 Daniela Ushizima, 11/24/2008

4. Contributions

- A method to identify and track density patterns in particle acceleration data:
  - MST representation and pruning to recover high density peaks: lifetime diagram of high-density bunches of electrons;
  - Fusion of probes with groups: beam-point candidates and fuzzy clustering to segment the beam particles;
- Results:
  - Four different datasets illustrate our experimental results by comparing to a manual selection by experts;
  - Detection of high energy particles given space-energy parameters;
  - Limitations: low quality beams are not formed by the highest energy particles (may not be detected); multiple beams.

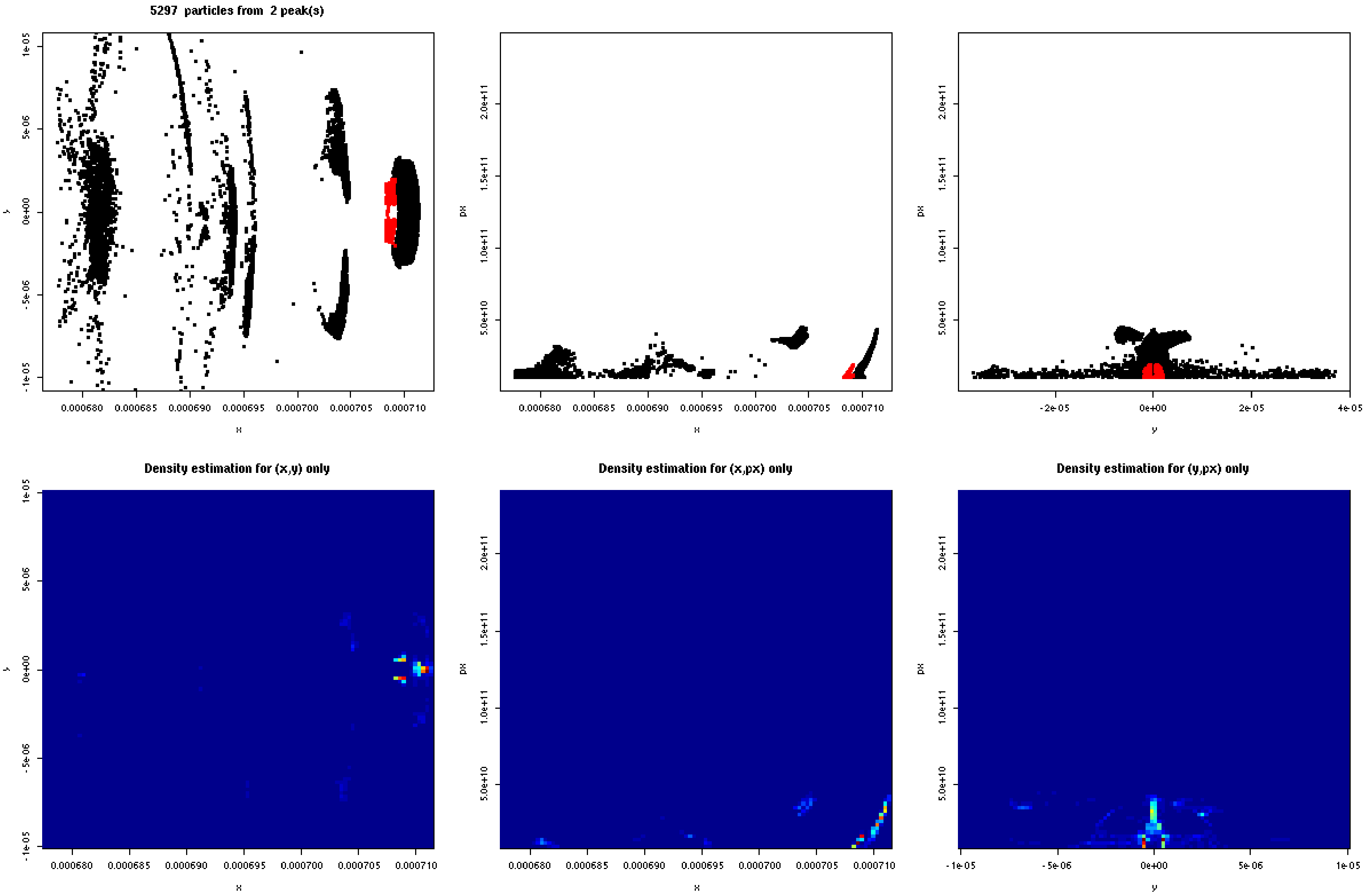
5. Future developments



Dataset D

- Statistics using Parzen windows: spatial and energy components for density estimation analysis;
- Data reduction using geometrical methods (Math - LBL);
- Beam quality characterization based on intra/inter cluster measurements (intra-beam scattering;
- Domain decomposition (time steps) for parallelization (ORNL).
- Particle tracking along time steps.

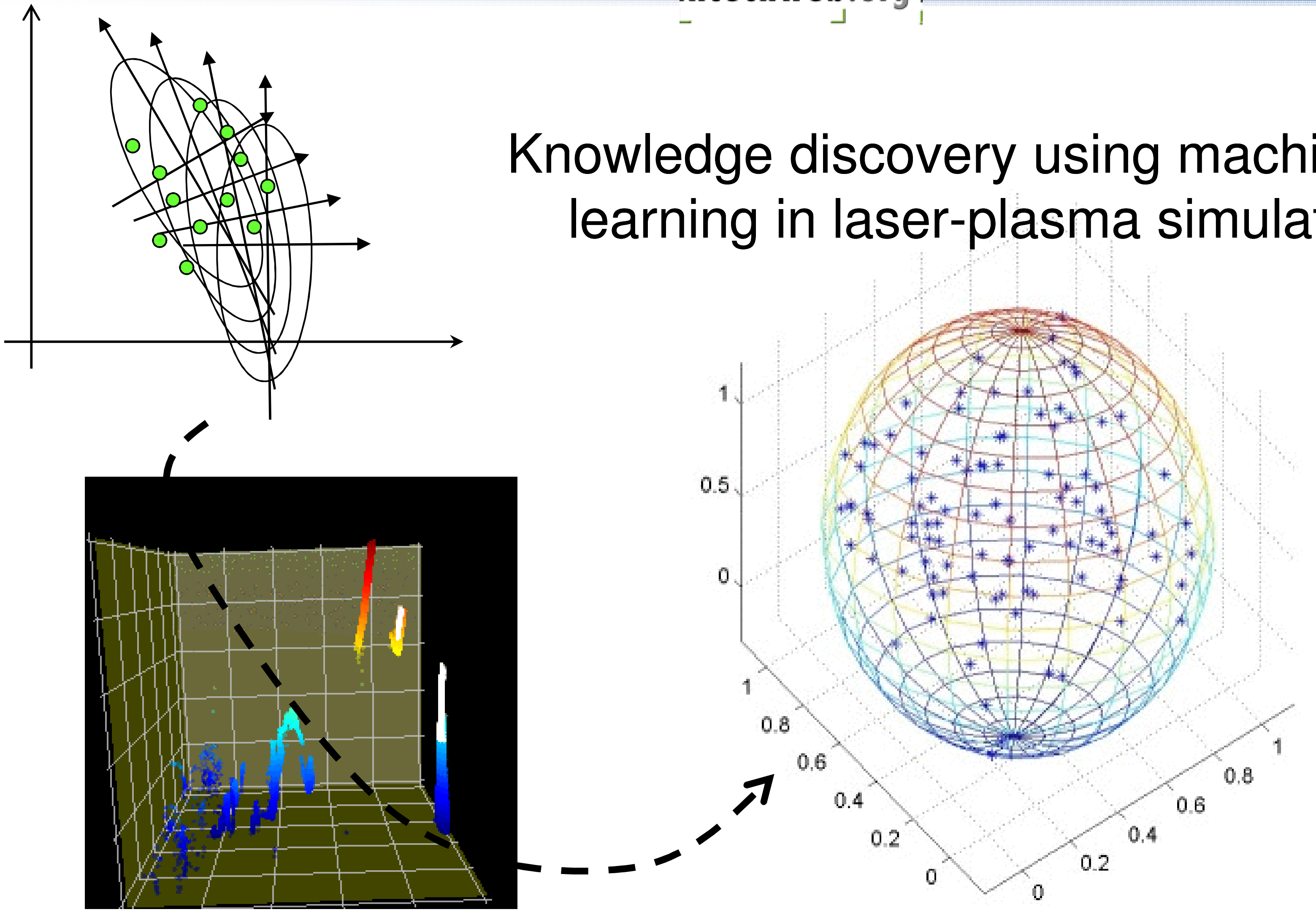
What to say about the bubble effect? =>minimum volume enclosing ellipsoid



To appear…

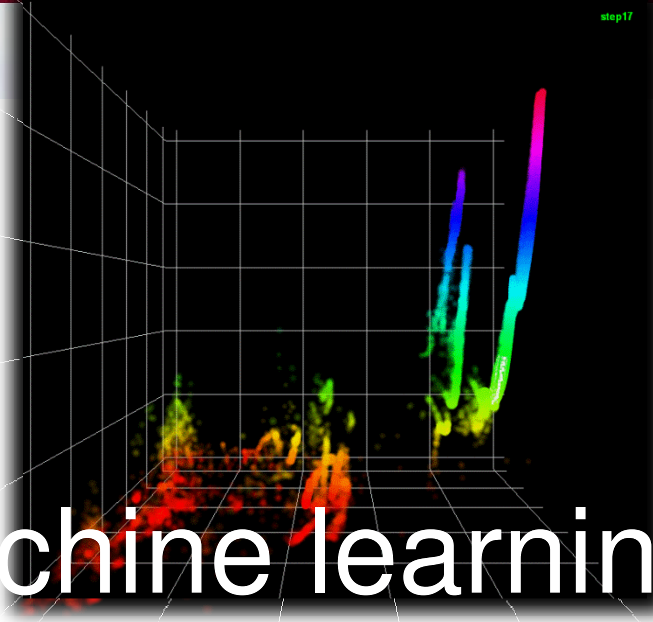
intechweb.org

Knowledge discovery using machine learning in laser-plasma simulations



6.References

- Daniela Ushizima, Oliver Rübel, Prabhat, Gunther Weber, E. Wes Bethel, Cecilia Aragon, Cameron Geddes, Estelle Cormier-Michel, Bernd Hamann, Peter Messmer, Hans Hagen. "Automated Analysis for Detecting Beams in Laser Wakefield Simulations". 2008 Seventh International Conference on Machine Learning and Applications, Proc. of IEEE ICMLA'08, 2008.
- Geddes et al, "High-quality electron beams from a laser wakefield accelerator using plasma- channel guiding," Nature 2004.
- Malka et al, “Principles and applications of compact laser–plasma accelerators”, *Nature Physics* **4**, 447 – 453, 2008.
- M. J. Crawley, “The R Book”. John Wiley and Sons, Ltd, 2007.
- L. Kaufman and P. Rousseeuw, “Finding Groups in Data: An Introduction to Cluster Analysis” John Wiley and Sons, Ltd,1990.
- E. Kreyszig, “Advanced Engineering Mathematics”, John Wiley and Sons, Ltd, 2006.
- K.-D. Liss, A. Bartels, A. Schreyer, H. Clemens, “High energy X-rays: A tool for advanced bulk investigations in materials science and physics”, Textures and Microstructures, 2003.
- Research of Advanced Accelerators at KERI http://caa.keri.re.kr/research.htm
- http://vis.lbl.gov/Events/SC06/Incite7/index.html



http://vis.lbl.gov/~daniela

Machine learning on accelerator simulation data

Daniela M. Ushizima
Lawrence Berkeley National Laboratory,

Participants**:** Oliver Rubel, Prabhat, Gunther H. Weber, E. Wes Bethel, Cecilia R. Aragon, Cameron G.R. Geddes, Estelle Cormier-Michel, Bernd Hamann, Peter Messmer, Hans Hagen