Data Challenge 2018

Solution of the ***Structure-Dynamics paradox*** *in glass-forming liquids*

F.P. Landes, G. Biroli @ IPhT CEA & ENS Paris

Simons' Collaboration **"Cracking The Glass Problem"**

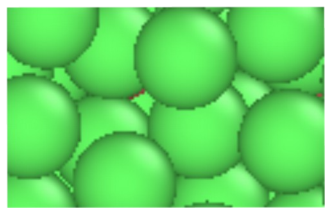
SIMONS FOUNDATION

The paradox of the formation of amorphous solids:

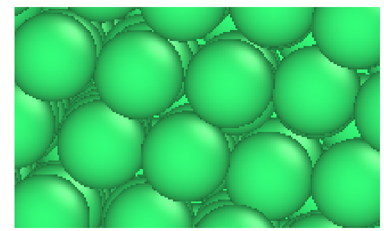
very different dynamics,

very similar structure

fluid

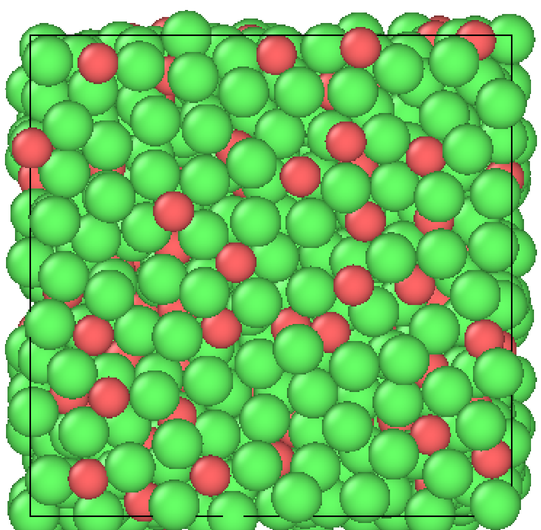


/

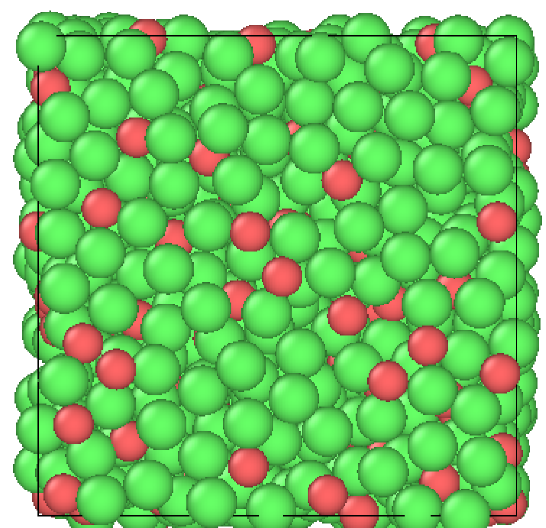


crystal

"glass"
(fluid)
high T



/



"glass"
(solid)
low T

Difficult to guess which is the **fluid/solid**

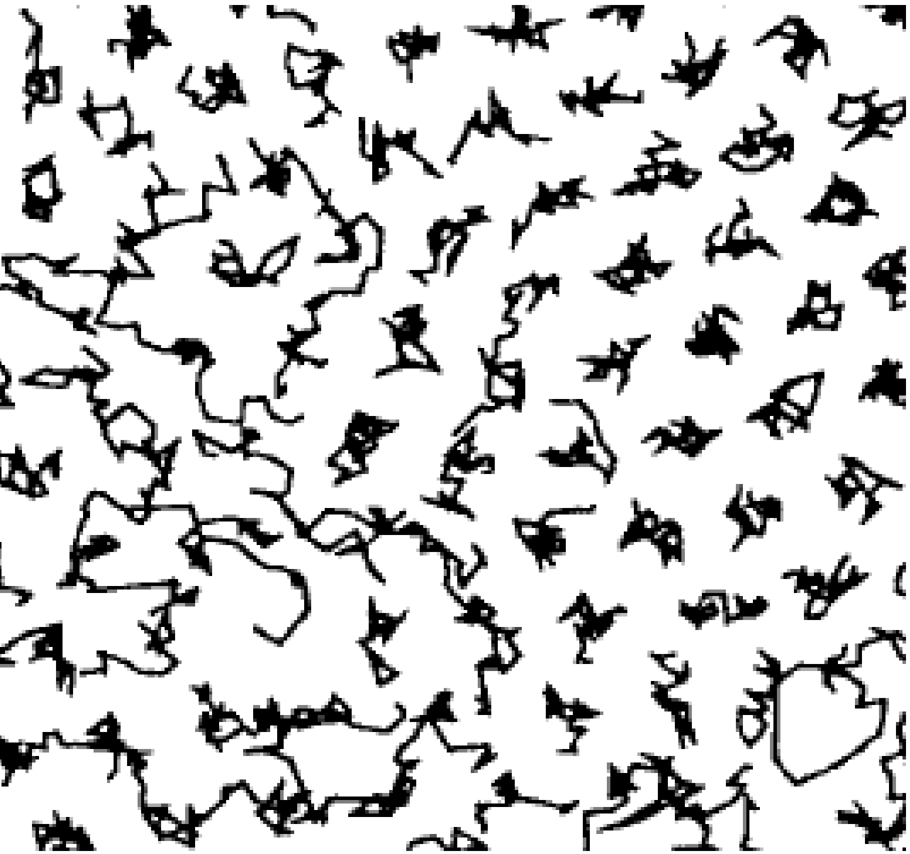
Understanding the glass transition

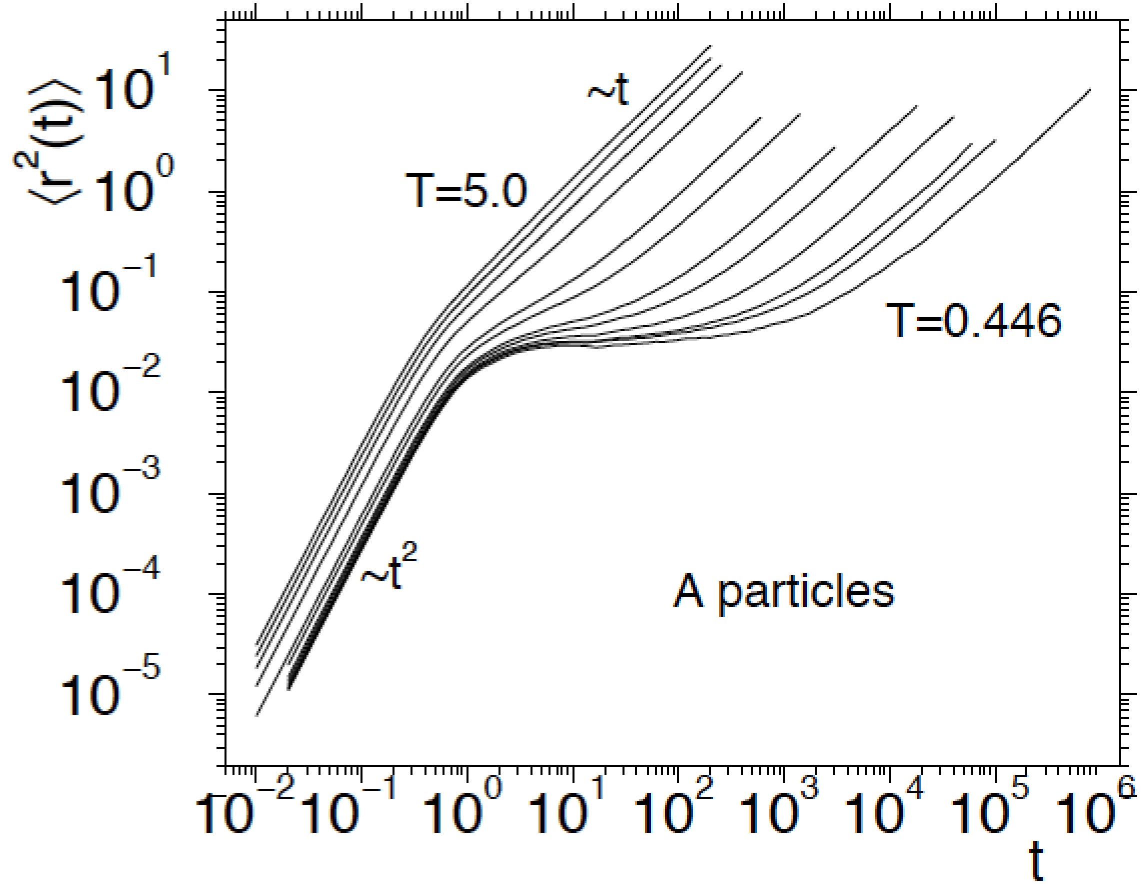
- Crucial problem in Physics (statistical physics, condensed matter)
- Several ideas* in the Physics literature, but the problem is still open
- **Can Machine Learning help solving it?***

* *Review of Modern Physics*, ***83***, *587, Berthier, Biroli, 2011*

* For pioneering ML approaches to glasses, see:
PNAS, 114(2), 263–267, Schoenholz, Cubuk, et. al.;
Nat. Phys., 12(5), 469–471, Schoenholz, Cubuk, Sussman, et.al.

Glasses: Microscopic Dynamics





Particles are **caged over long times**

Cage lifetime increases when T decreases

Mobility (the Label)

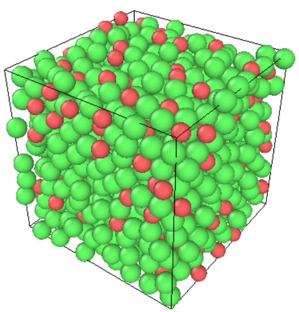
Input:

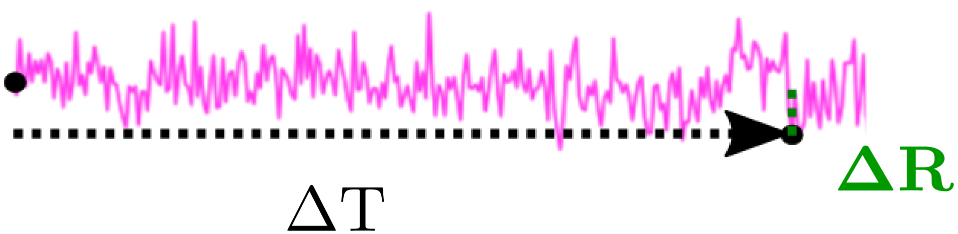
Initial positions and velocities $\rightarrow$

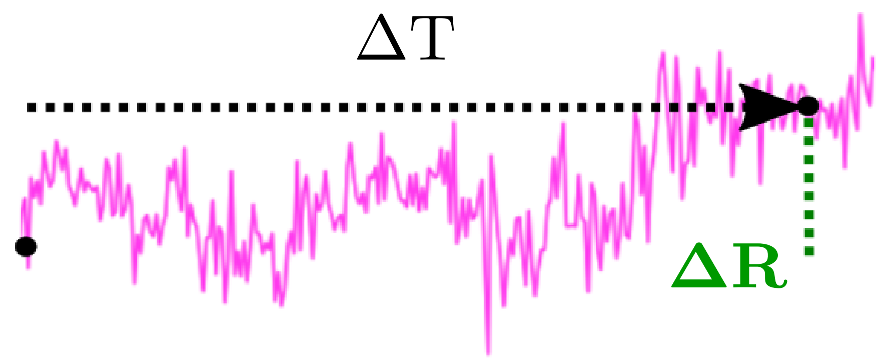
Output:

Mobile and Immobile particles

$(\mathbf{\Delta R}(\Delta \mathrm{T}) > \Delta \mathrm{R}^* = \mathrm{const.})$







$\mathbf{\Delta R} < \Delta \mathrm{R}^*$, **Immobile**

$\mathbf{\Delta R} > \Delta \mathrm{R}^*$, **Mobile**

(NVE integration, $\Delta \mathrm{T}$ quite large, a fraction of the α-relaxation time)

Data types

- **Inputs:**
  1000 **coordinates** (3D) and **velocities** (3D) (800 A+200 B)
  for the snapshot number i: x_i =
  {x1,y1,z1,x2,y2,z2,....,z1000,vx1,vy1,vz1,...vz1000}
  (6000 values per snapshot)

- **Outputs:**
  the **800 mobilities** of atoms of type A,
  y_i={s1,s2,...,s800} ; each s∈{0,1}
  y_i={which atoms moved between time i and i+T}

- n=1877 independent training snapshots (~1.5 M particles)
  n_t=510 independent Testing snapshots (~0.4 M particles)

Goal : Reformulation

- $\{\mathbf{x(t)},\mathbf{v(t)},\mathbf{V(r)}\}$ + [***Newton's equations***]
  $\rightarrow$ positions at $t+\Delta T$
- $\{\mathbf{x(t)},\mathbf{v(t)}$, **examples**$\}$ + [***Machine Learning***]
  $\rightarrow$ $\mathbf{y=f(x)}$ at $t+\Delta T$

$\rightarrow$ Can you "learn" **what matters** in the initial condition to predict important features at later times? (**without** knowledge of the Netwon equations nor the potential V(r), but **with training examples!**)

Scoring

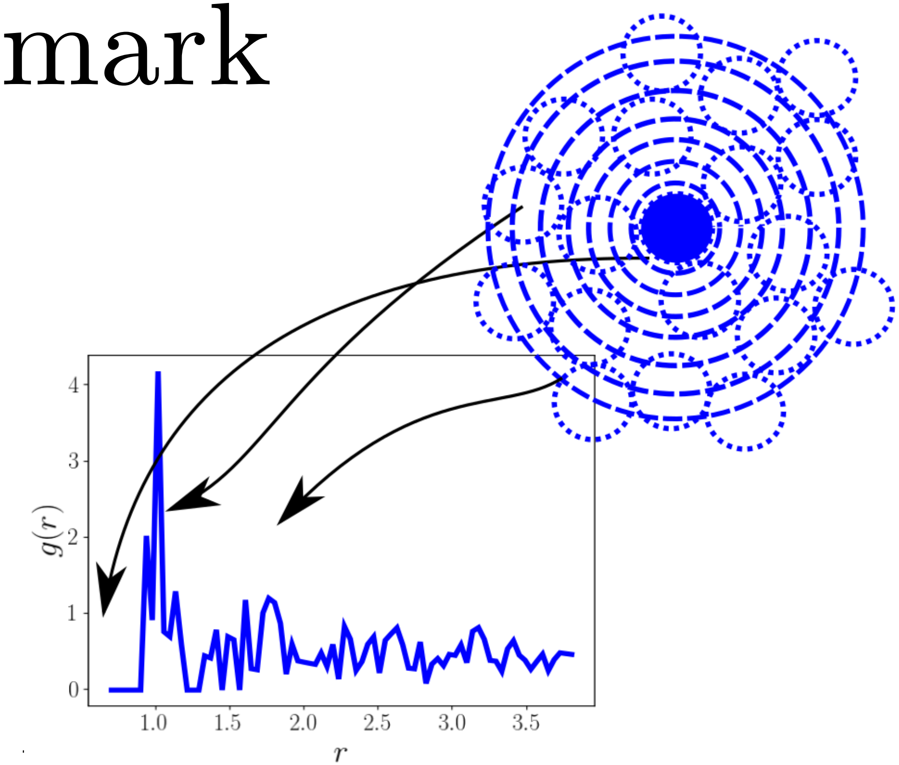
- Classes are approximately balanced

- **Score = Accuracy = 1 − Risk =**

$$\frac{\textit{number correctly classified}}{\textit{number tested}}$$

Our benchmark

- Simple **features***:
  - **Histograms of neighbors' density**, for AA and AB pairs separately
  - Not even use velocities



- Simple classification with SVM, linear Kernel
- Accuracy ~ 60% only !
- Newtons' equations: 100% accuracy, but we **learn nothing about what matters**

*inspired from: *PRL, 114(10), 108001, Schoenholz, Cubuk, et. al.; Nat. Phys., 12(5), 469–471, Schoenholz, Cubuk, Sussman, et.al.*

Suggestions, Ideas

- Use **more exhaustive features**?

  Higher order correlation functions (more than 2-point)

- Use **deeper networks**?
- Use the **velocities**?

Outcomes

- **High predictability** means **discovering the physical origin of dynamics** in glass-forming liquids, **by machine learning**
- Great progress in **one of the most studied physics problem** in statistical physics and condensed matter!
- Possible **high-profile academic publication**, **collaboration**, NYC conference participation (2019)

Thank you !

- check the **website** for more information:
`http://lptms.u-psud.fr/francois-landes/data-challenge/`

- A **python code** is available to take care of the
Periodic Boundary Conditions (PBC)