Binary and multiple stellar systems from data analysis of the space observatory GAIA

Petr Závada and Karel Píška

FZU Institute of Physics of the Czech Academy of Sciences

Seminar of Division of Elementary Particle Physics 30/11/2023

Outline

- **A few words about the Gaia observatory**
- **Analysis of astrometric data - inspired by heavy-ions physics**
- **Wide binaries and multiple bound systems**
  - **Physical motivation**
  - **Selection algorithm**
  - **Results of analysis**
- **What next?**
- **Summary and conclusion**

Details can be found in Zavada P. and Píška K.:

[1] A statistical analysis of two-dimensional patterns and its application to astrometry, 2018 A&A 614 A137

[2] Statistical Analysis of Binary Stars from the Gaia Catalog Data Release 2, 2020 AJ 159 33

[3] Catalog of Wide Binary, Trinary and Quaternary Candidates from the Gaia Data Release 2 (Region |b| > 25°), 2022 *AJ* 163 33

A few words about Gaia

Gaia is a space observatory of the European Space Agency (ESA), launched in 2013 and expected to operate until 2025.

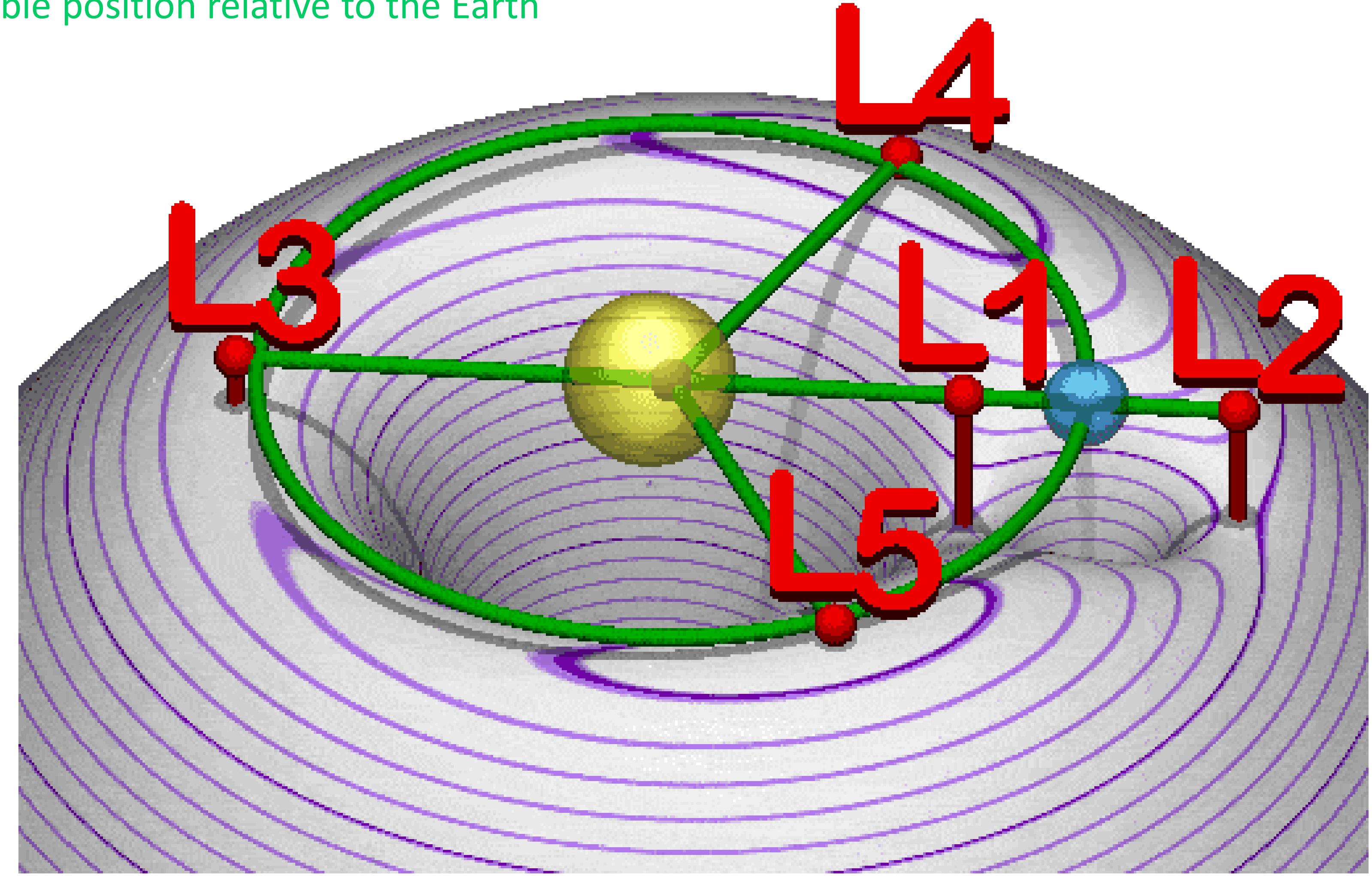
The spacecraft is designed to measure the positions, distances, motions and other parameters of stars with unprecedented precision.

[5] Gaia Collaboration, T. Prusti, J. H. J. de Bruijne, A. G. A. Brown, A. Vallenari, C. Babusiaux, C. A. L. Bailer-Jones, U. Bastian, M. Biermann, D. W. Evans and et al. (2016b) The Gaia mission. A&A 595, pp. A1.
[6] Gaia Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, F. Mignard, R. Drimmel, C. Babusiaux, C. A. L. Bailer-Jones, U. Bastian and et al. (2016a) Gaia Data Release 1. Summary of the astrometric, photometric, and survey properties. A&A 595, pp. A2.
[7] Gaia Collaboration, F. Arenou et al. Gaia Data Release 1. Catalogue validation. A&A 599 (2017), pp. A50.

Gaia is located at L2, 1.5 million kilometres from Earth. (Webb telescope is located also at the L2 - at a safe distance)

L_i – Lagrange points
stable position relative to the Earth

Three-body **problem**: $m_{Gaia} \ll m_{Sun}, m_{Earth}$

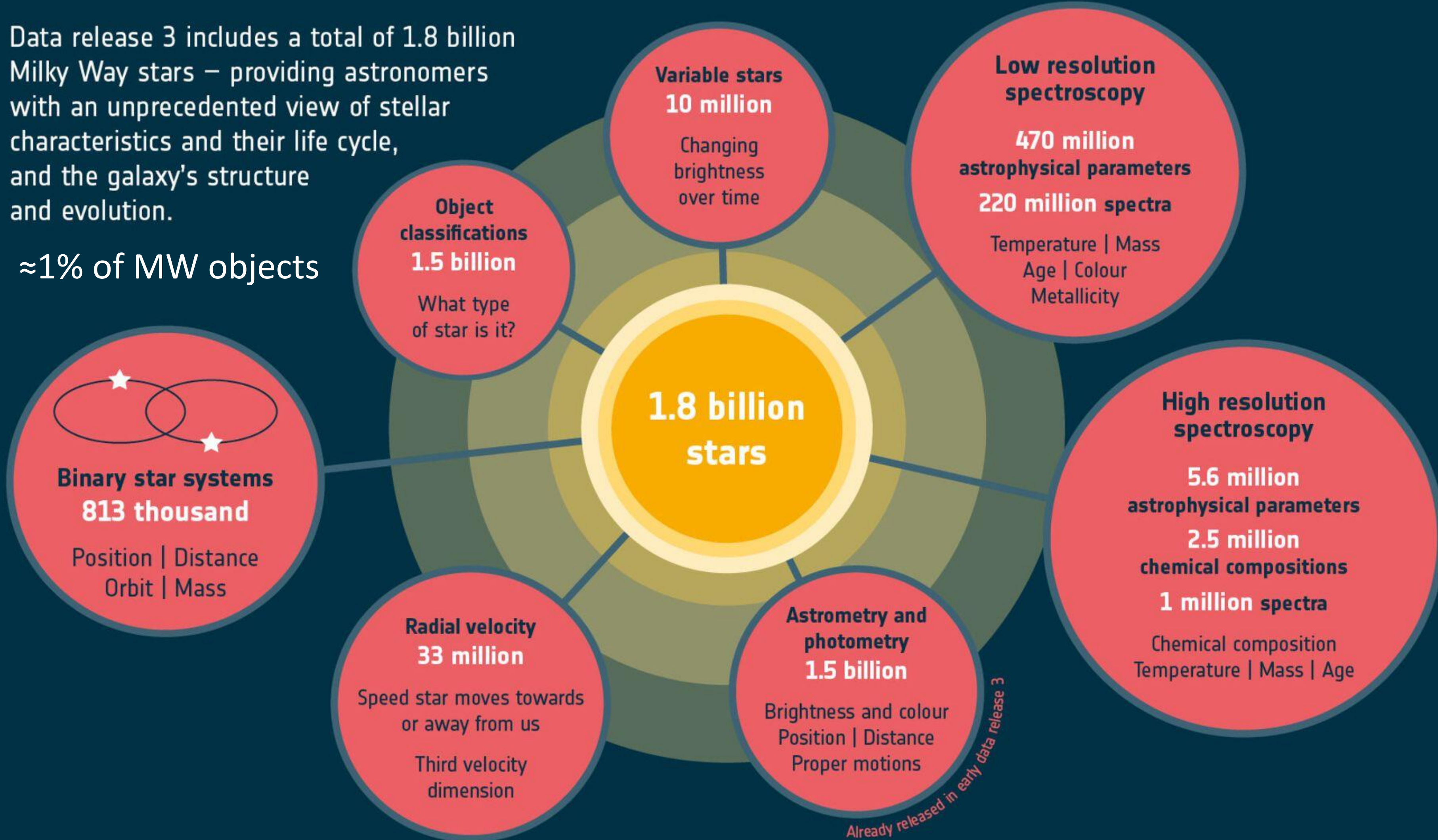


Gravitational forces of Sun and Earth and centrifugal force are compensated

MILKY WAY STARS - DATA MEASURED BY GAIA

Data release 3 includes a total of 1.8 billion Milky Way stars – providing astronomers with an unprecedented view of stellar characteristics and their life cycle, and the galaxy's structure and evolution.

≈1% of MW objects



Unique data are free!

Galactic reference frame, SUN sits in its origin

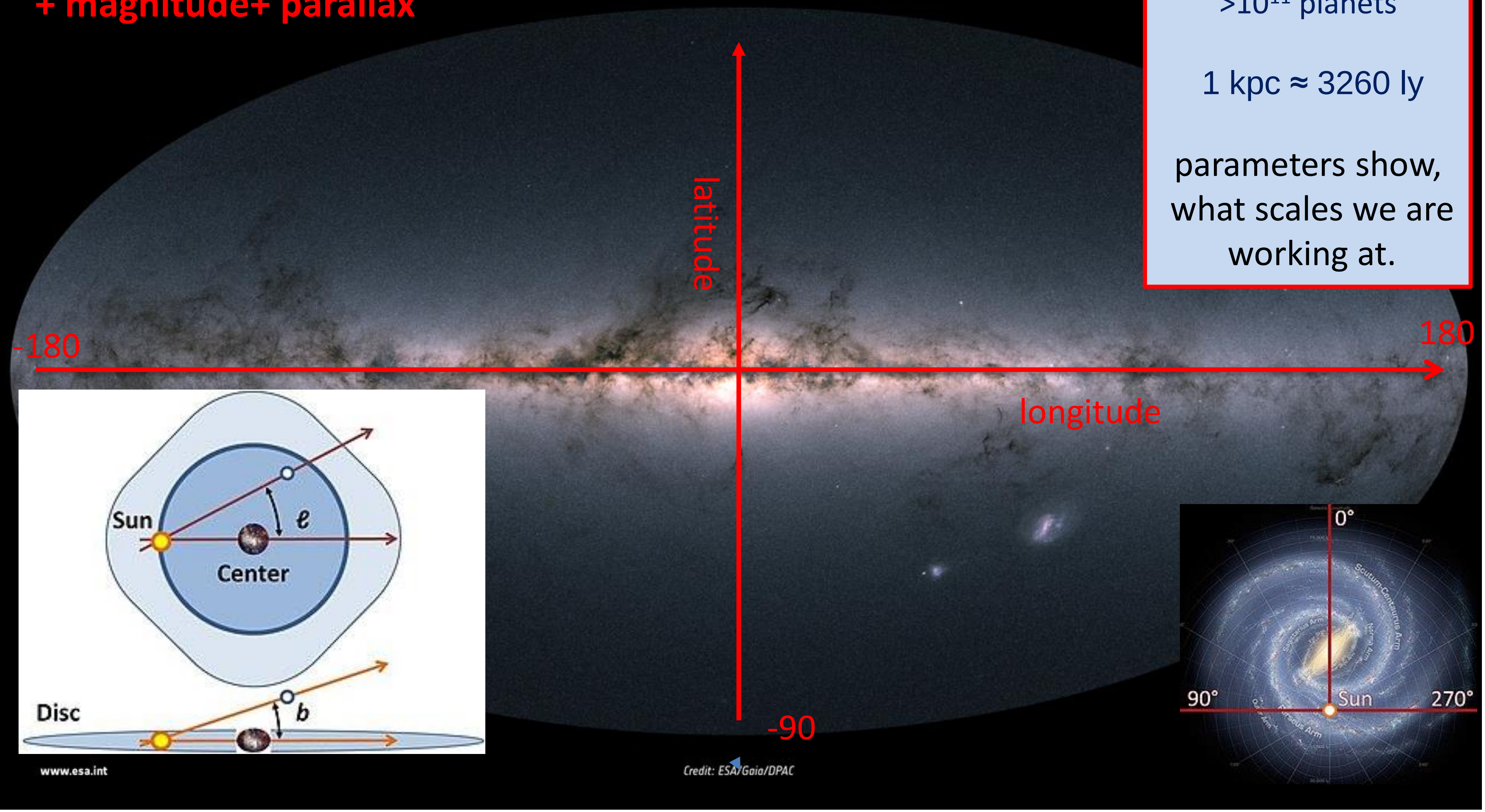
We deal with: angular positions (2) +proper motion (2) + magnitude+ parallax

Milky Way Disc

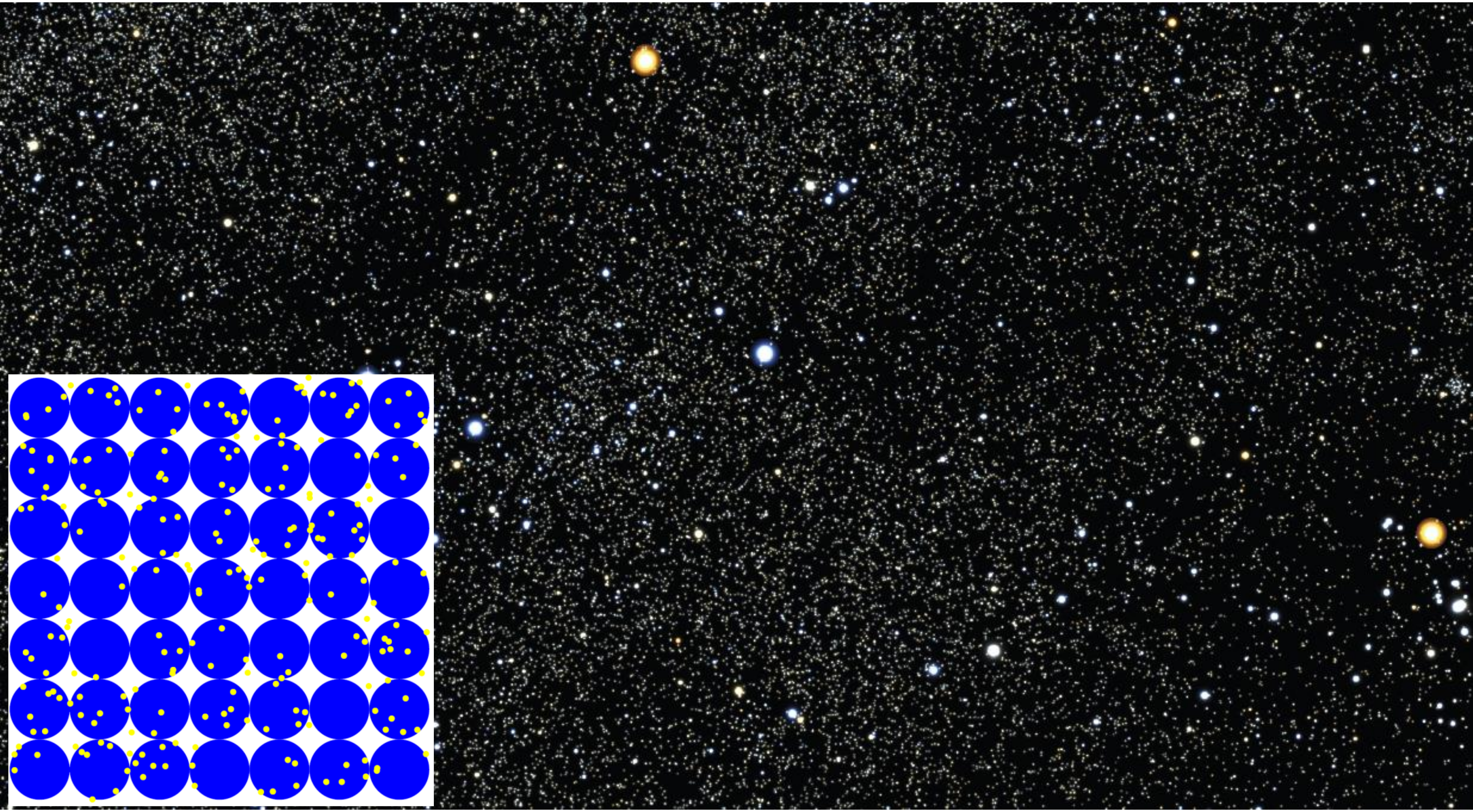
$R \approx 14$kpc, $\Delta z \approx 0.3$kpc
$V \approx 600$km/s, $R_S \approx 8$kpc
1-4x10^{11} stars
$>10^{11}$ planets

1 kpc ≈ 3260 ly

parameters show, what scales we are working at.



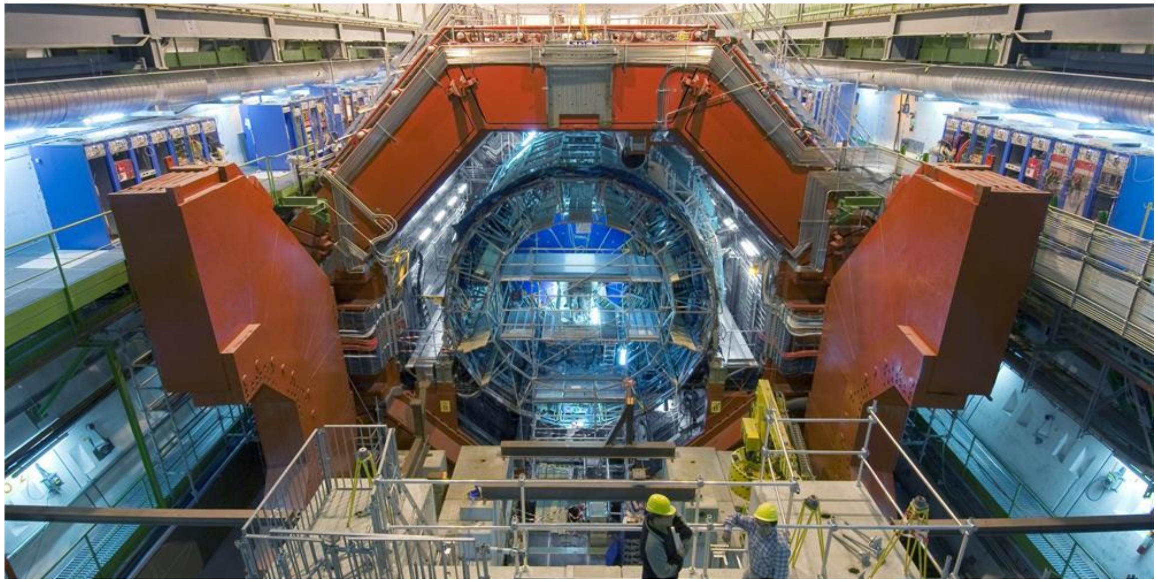
… in better resolution:

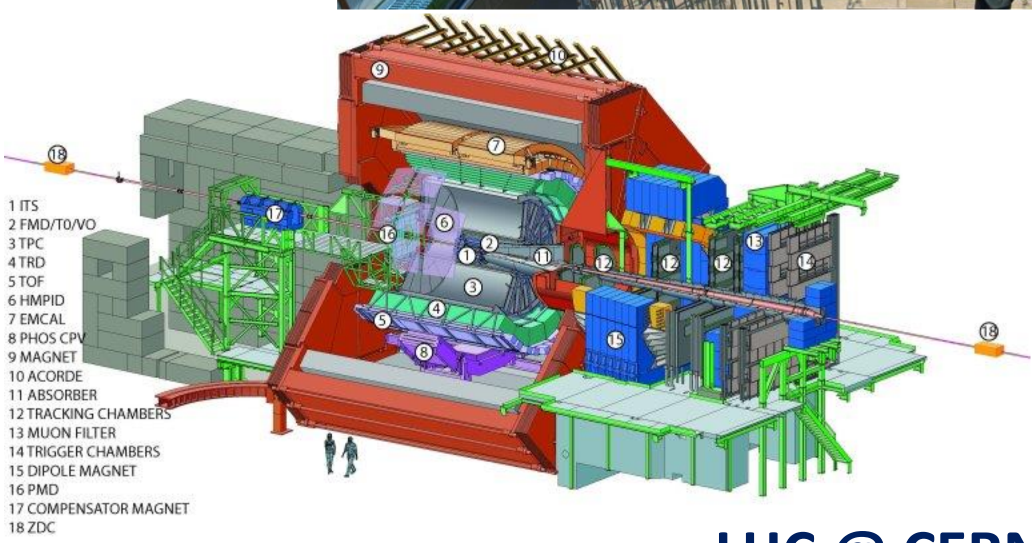


Our research is based on a statistical analysis of the patterns within the circles covering the sky. The method of analysis is motivated by methods known from HI physics (ALICE).

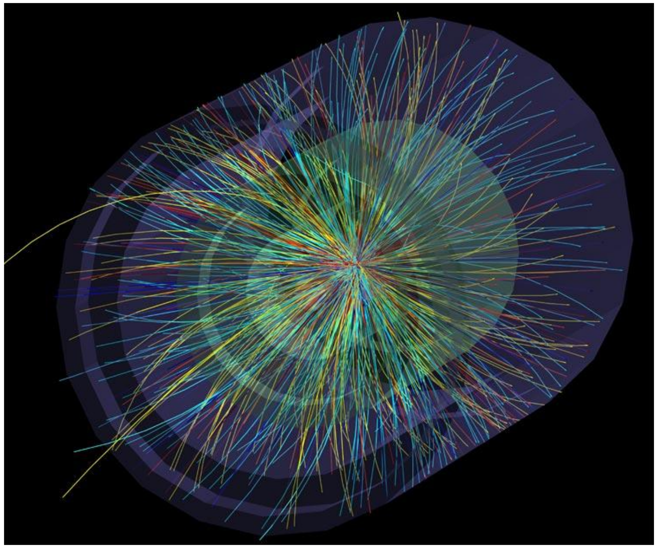
ALICE – A Large Ion Collider Experiment

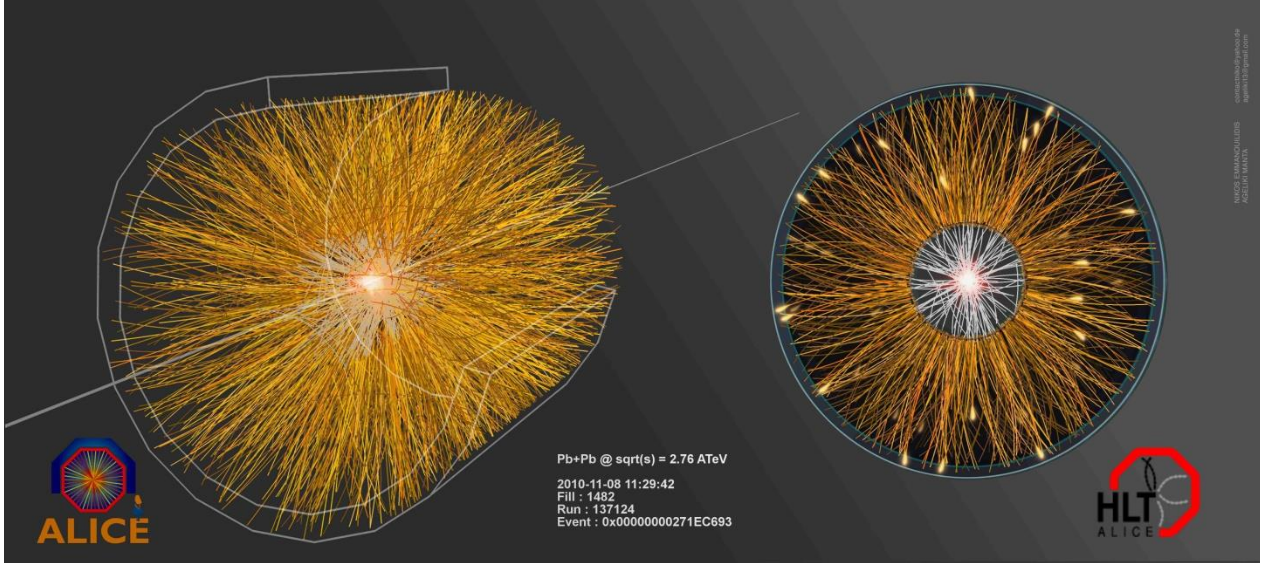
provides data on HI collisions





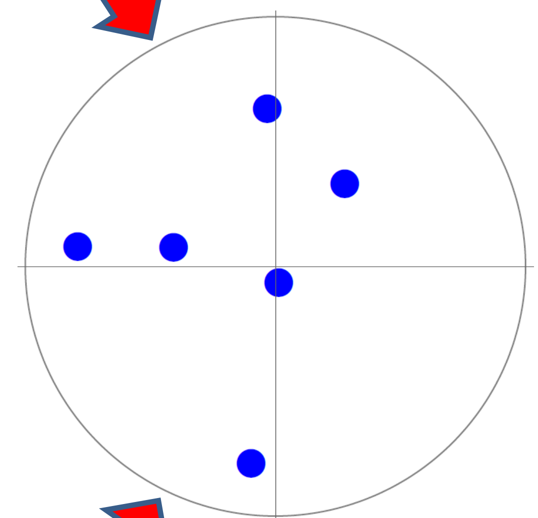
LHC @ CERN

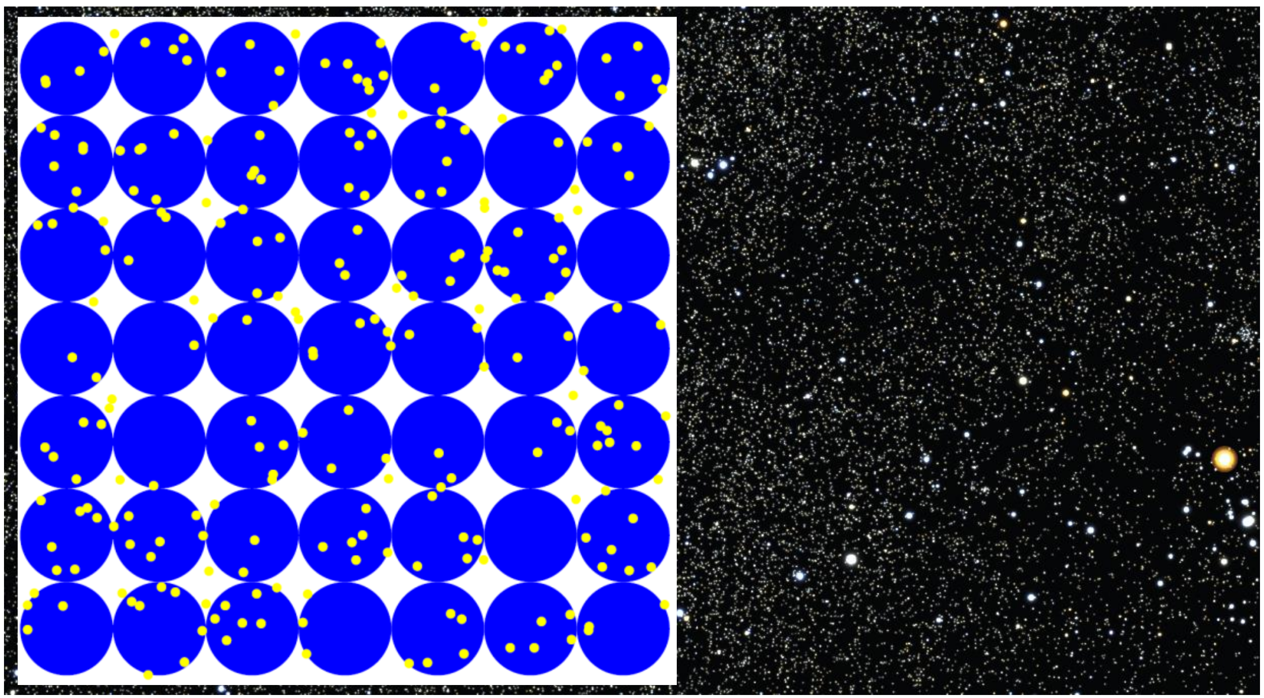




In HI we study momentum distribution in the transverse plane.

We study patterns like this:

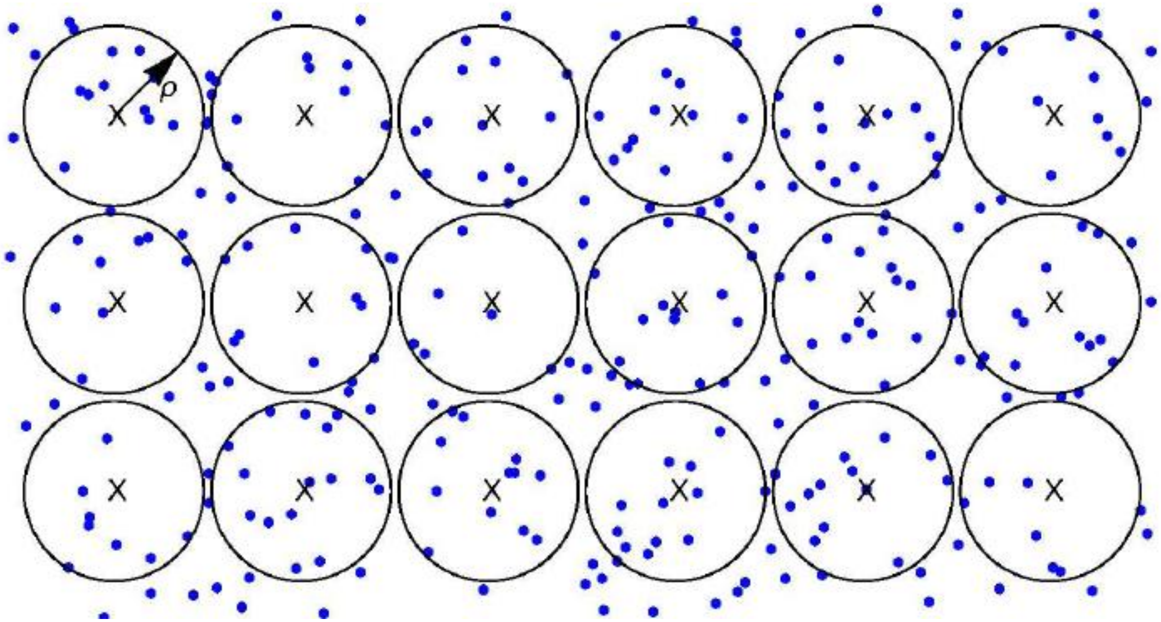


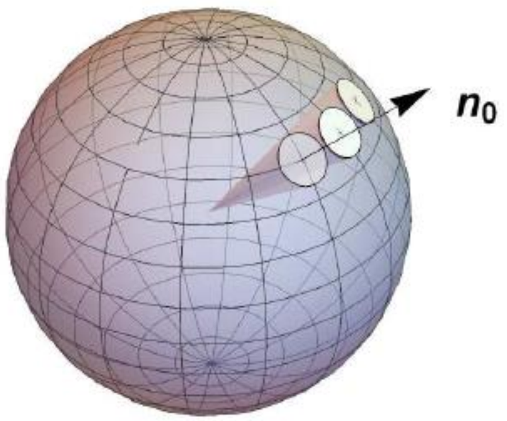


patterns generated in HI are very similar to patterns in the sky. We will try to apply HI methods to analysis of the sky patterns - after some modification.

Analysis

We analyze distribution of random points (stars) inside the circles (we call them events). Deviations from uniform distributions may indicate an interesting physics. <u>How to define and find these deviations?</u>





Do we observe more pairs of close stars than random statistics allow?

Inspiration by HI: Fourier analysis

a very useful tool for discussing azimuthal correlations and asymmetries in the transverse plane

In HI we work with this form of Fourier decomposition: (v_n,Ψ_n are free parameters)

$$P(\varphi) = \frac{1}{2\pi}\left(1 + 2\sum_{n=1}^{\infty} v_n \cos\left[n\left(\varphi - \Psi_n\right)\right]\right)$$

If $\langle f(\varphi)\rangle \equiv \int_{-\pi}^{\pi} P(\varphi) f(\varphi)\, d\varphi$

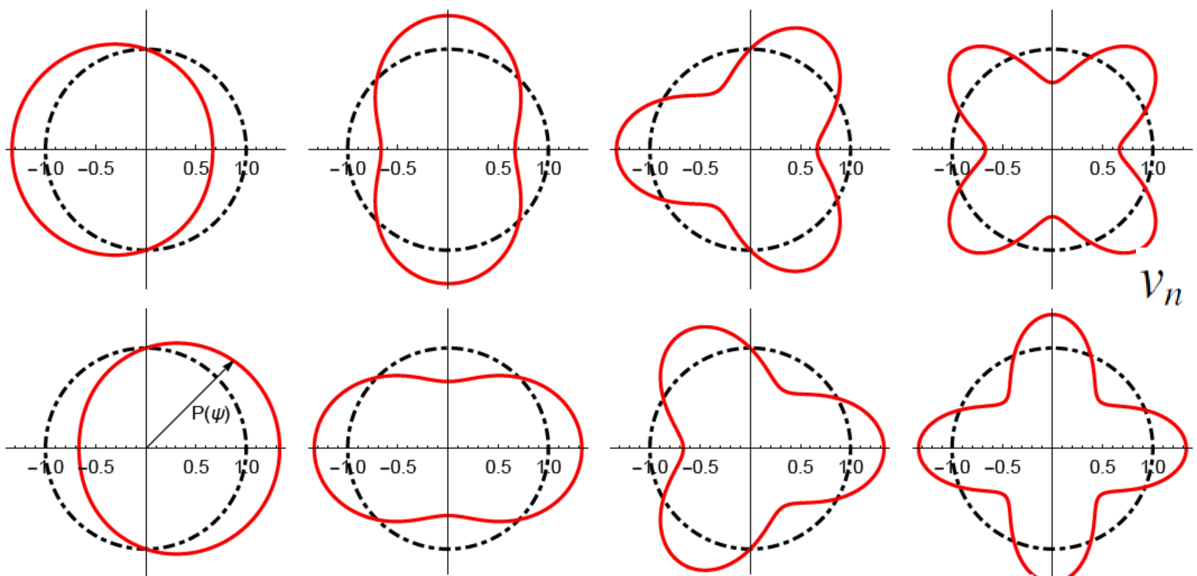
then for any n:

$$v_n = \langle \cos\left[n\left(\varphi - \Psi_n\right)\right]\rangle$$

$$\tan\left(n\Psi_n\right) = \frac{\langle \sin(n\varphi)\rangle}{\langle \cos(n\varphi)\rangle}.$$

Decomposition n=1,2,3,4:

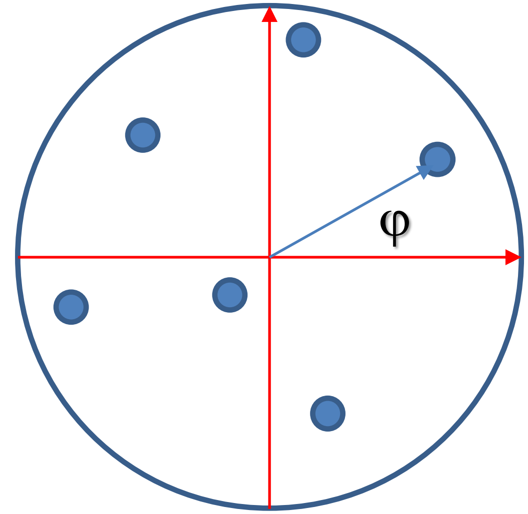
The individual terms represent different kinds of azimuthal asymmetry.
e.g.
directed flow,
elliptic flow,…



$v_n = \mp 1/6$

$$P(\psi) = 1 + 2v_n \cos(n\psi), \qquad \psi = \varphi - \Psi_n$$

Finite patterns, event-by-event



For a finite set $\{\varphi_1 \ldots \varphi_M\}\,; \qquad -\pi < \varphi_i < \pi$

of multiplicity M

we replace $$\langle f(\varphi)\rangle \equiv \int_{-\pi}^{\pi} P(\varphi) f(\varphi)\, d\varphi$$

by $$\langle f(\varphi)\rangle_M \equiv \frac{1}{M}\sum_{k=1}^{M} f(\varphi_k)$$

then for any *n*:

$$v_n(M) = \langle \cos[n(\varphi - \Psi_n)]\rangle_M$$

$$\tan(n\Psi_n(M)) = \frac{\langle \sin(n\varphi)\rangle_M}{\langle \cos(n\varphi)\rangle_M}.$$

$\Longrightarrow$ **Average over events:** $$\left\langle v_n^2(M)\right\rangle = \frac{1}{M}\left[1 + \frac{2}{M}\sum_{1\le k<l\le M}\left\langle \cos(n\varphi_k^j - n\varphi_l^j)\right\rangle\right]$$

For $M \to \infty \qquad \langle f(\varphi)\rangle_M \to \langle f(\varphi)\rangle\,, \qquad v_n(M) \to v_n, \qquad \Psi_n(M) \to \Psi_n$

For uniform distribution:

$$M\left\langle v_n^2(M)\right\rangle = 1$$

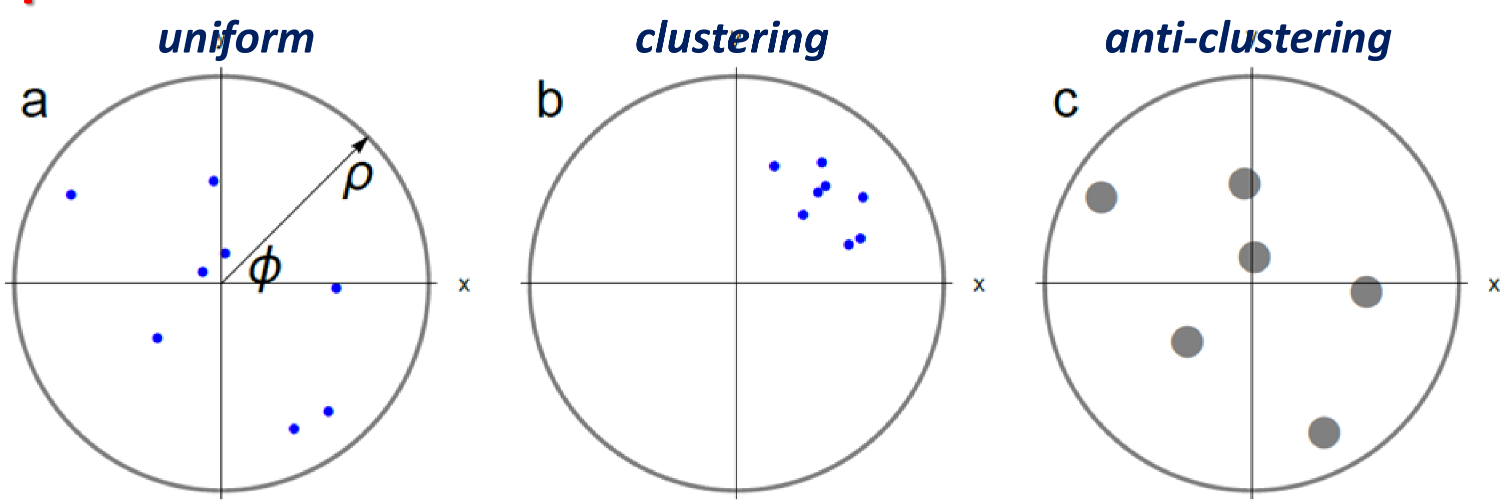
Classification of the event sets

Definition of characteristic functions:

$$\Theta_n(M) = M\left\langle v_n^2(M)\right\rangle = \frac{M}{N_M}\sum_{k=1}^{N_M} v_{n,k}^2(M)$$

where N_M is the number of events of multiplicity M, involve <u>important information about character of patterns (events)</u>

Examples:



$\Theta_n(M) = 1$ $\Theta_n(M) > 1$ $\Theta_n(M) < 1$

Toy examples - simulation

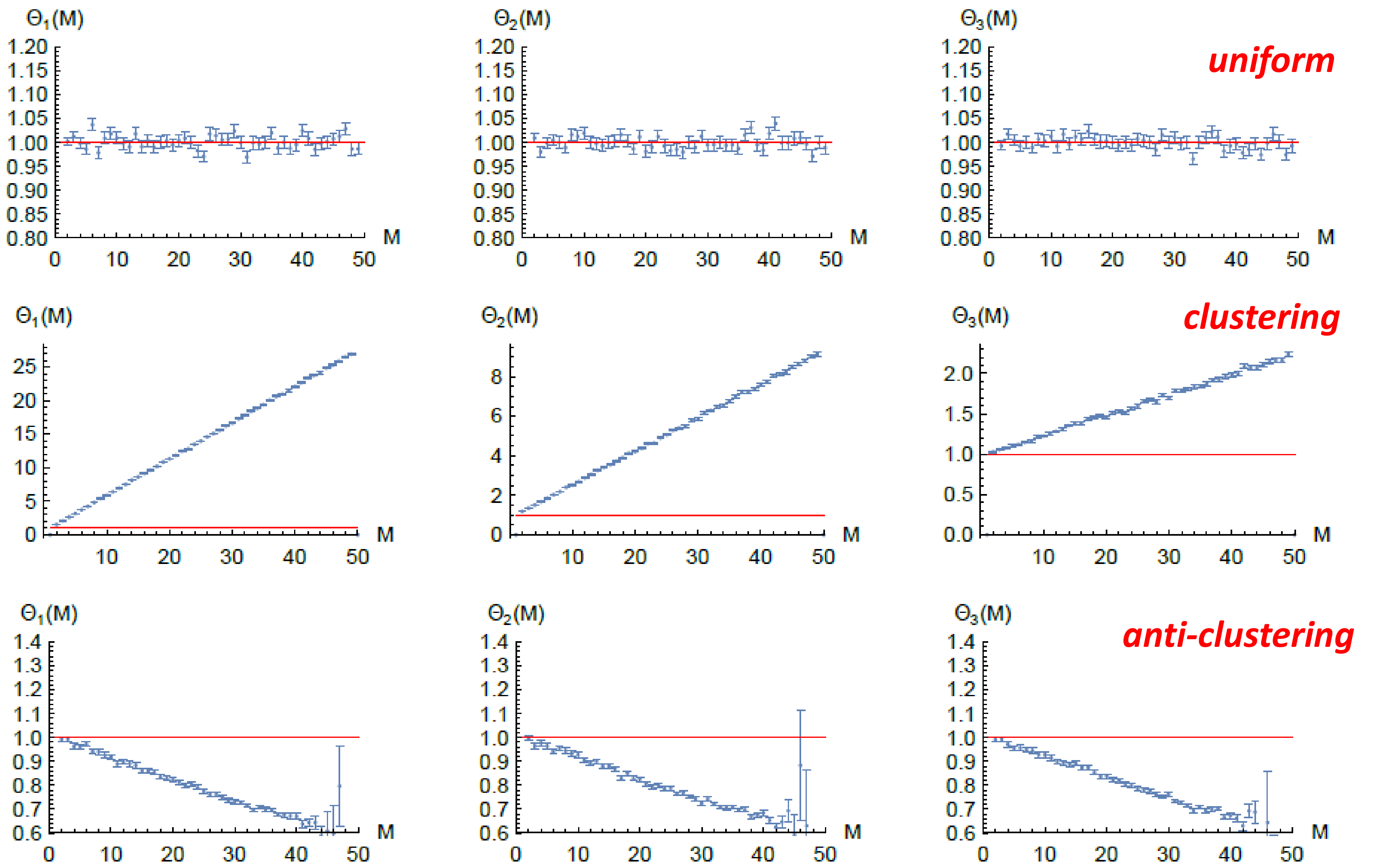
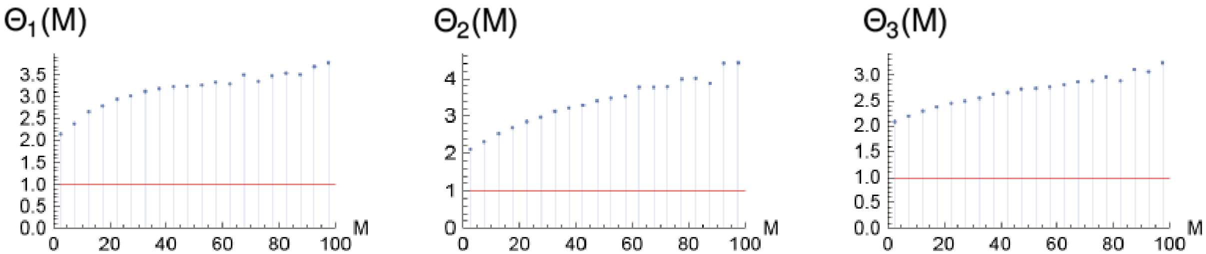
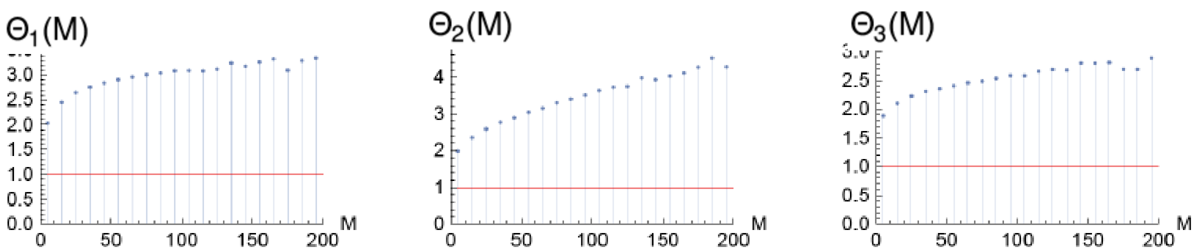


Fig. 4. The functions $\Theta_n(M)$, $n = 1, 2, 3$ for Monte-Carlo events in the scenario of uniform distribution (upper panels). The remaining represent clustering $\lambda = 1/3$ (middle panels) and anti-clustering scenario $\lambda = 1/20$ (lower panels). The each multiplicity bin is generated by $N_M = 6000$ events. The red lines correspond to the expected dependence for uniform distribution, $\Theta_n(M) = 1$ (Eq. (34)).

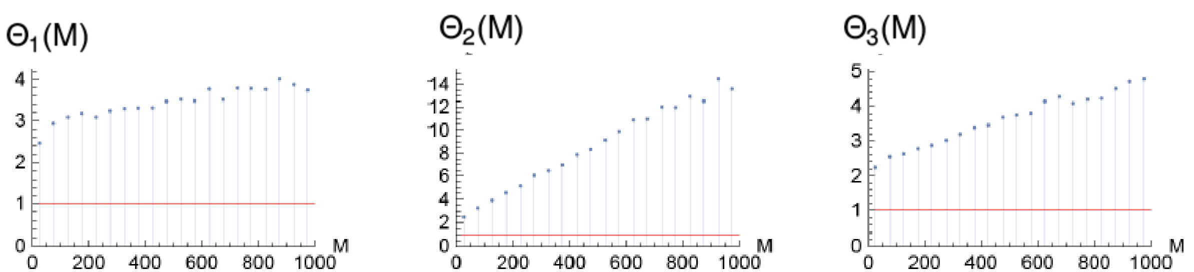
Data from ALICE



pp data, LHC10e, run 127712, $p_T > 0.3GeV/c, |\eta| \leq 0.7.$



pPb data, LHC13b, run 195351, $p_T > 0.3GeV/c, |\eta| \leq 0.7.$



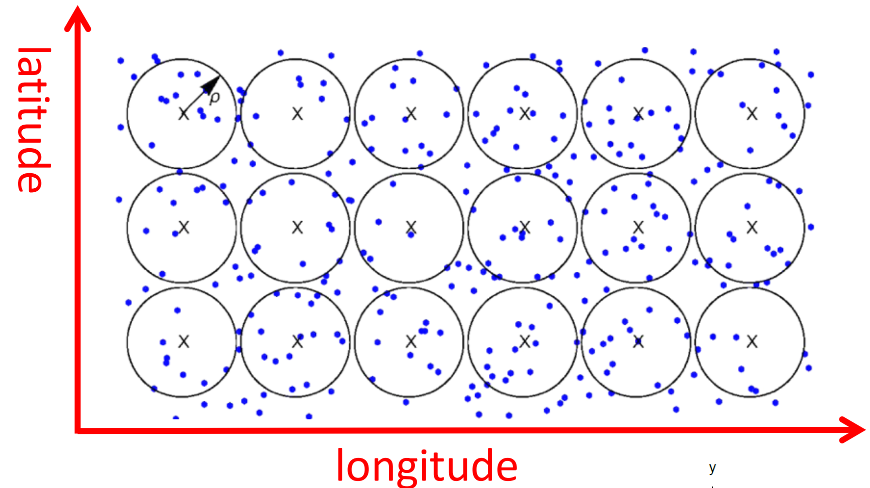
PbPb data, LHC10h, run 137544, $p_T > 0.3GeV/c, |\eta| \leq 0.7.$

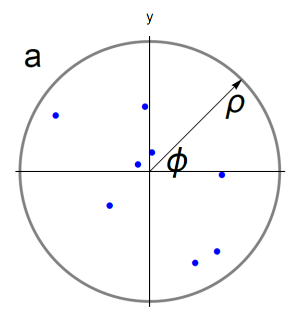
Clustering in momentum space = azimuthal asymmetry

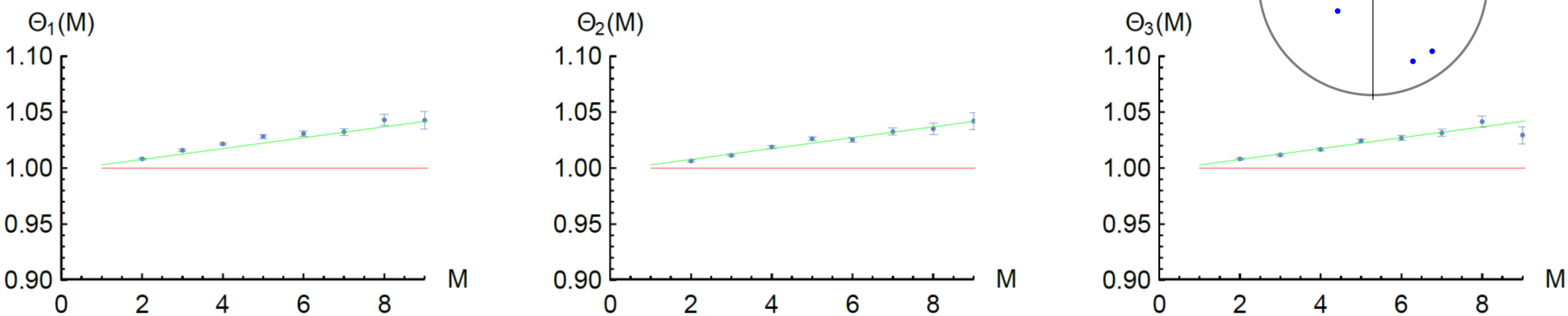
Data from GAIA

The characteristic functions clearly indicate the presence of multiple star systems, mostly binaries.

We will try to study them in more detail.



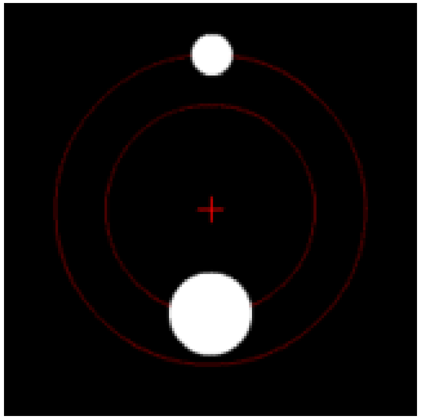




Clustering, now in stellar field (transverse projection)

Wide binaries and multiple bound systems

- A binary star is a system of two stars that are gravitationally bound and orbiting each other.



- More than half of all stars are binaries
- Orbital periods varies: hours – thousands years
- Multiple systems are much less probable
- Gaia astrometric data allow to identify wide binaries separated > 0.5'' (treshold of resolution) – this class will be subject of our analysis

Physical motivation (wide binary)

- Can provide deeper insight into the formation and evolution of galaxies
- May serve as a probe of galactic gravitational potential
- May provide data on the presence of dark matter in the galaxy

(see references in [3])

Input data

Table 1
Analyzed Regions $\boldsymbol{R}_{1,2}$ in the DR2 Catalog, where ρ_2 is the Angular Radius of the Events, $\langle L \rangle$, $\langle M \rangle$ are Average Distance and Event Multiplicity, and N_e is the Total Number of Events

	2D Region: $l \times b(\deg^2)$	$\rho_2(as)$	$\langle L \rangle$(pc)	$\langle M \rangle$	N_e	N_s	N_{tot}
R_1	$\langle -180, 180 \rangle \times \langle \pm 45, \pm 90 \rangle$	72	1807	3.86	5,887,737	22,083,670	30,643,238
R_2	$\langle -180, 180 \rangle \times \langle \pm 25, \pm 45 \rangle$	72	2134	7.86	6,985,043	53,043,629	80, 653, 496

Note. Only sources with positive parallax and in distance <15,000 pc are taken into account and only events $2 \leqslant M \leqslant 25$ are accepted for present analysis. N_s is number of sources after these cuts, and N_{tot} is total number of sources with positive parallax in $\boldsymbol{R}_{1,2}$.

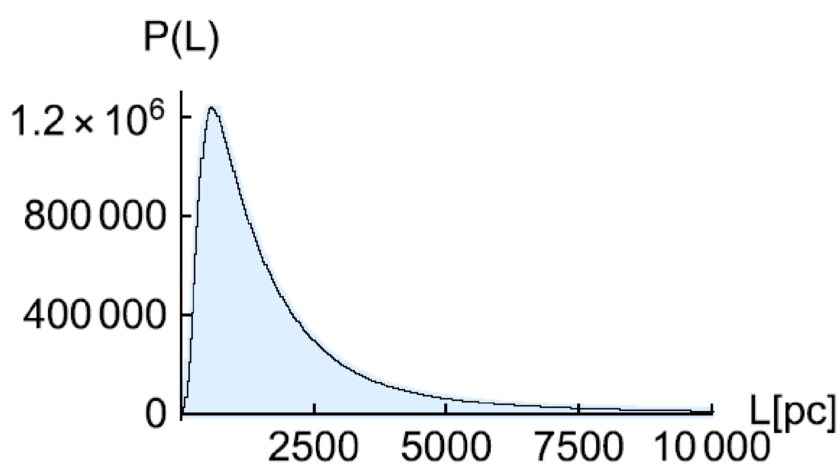
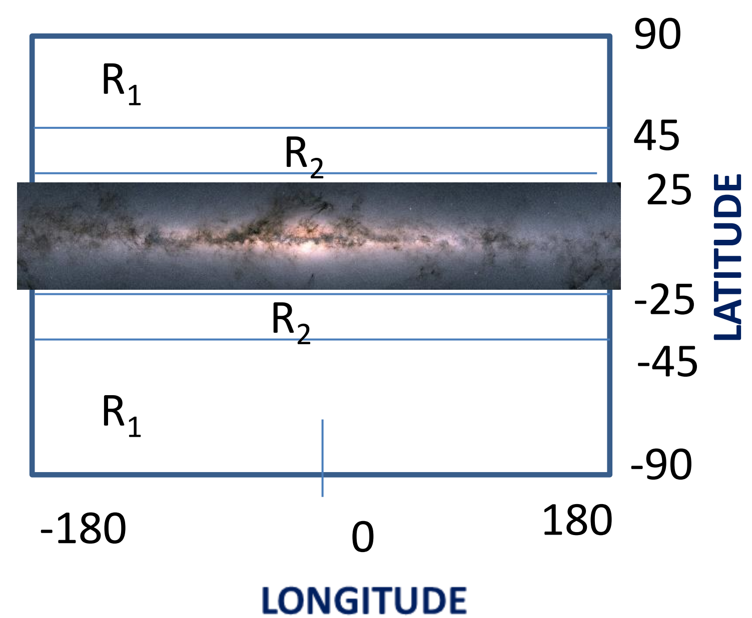
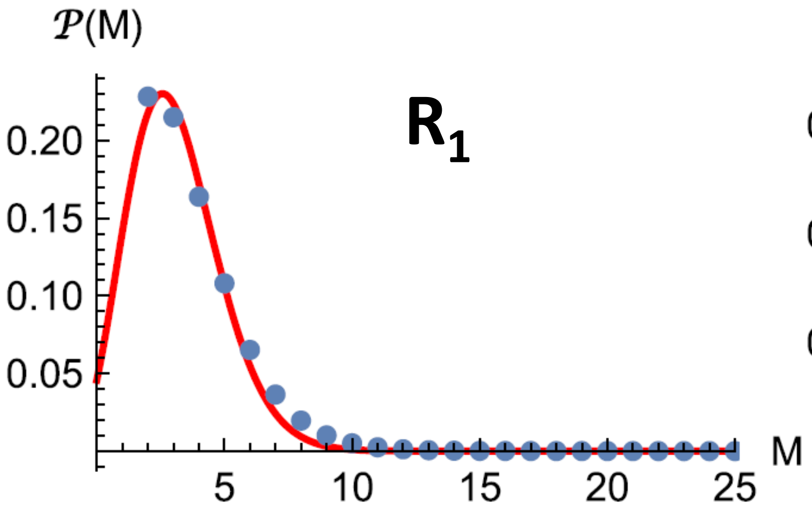


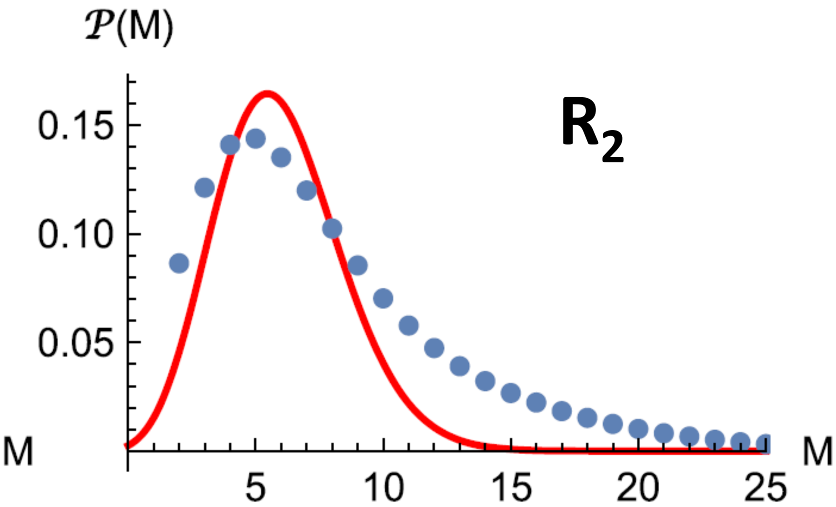
Figure 2. Distribution of distances of stars of the region $\boldsymbol{R}$. Binning: 30 pc.

$$\boldsymbol{R} = \boldsymbol{R}_1 \bigcup \boldsymbol{R}_2$$



Multiplicity distribution



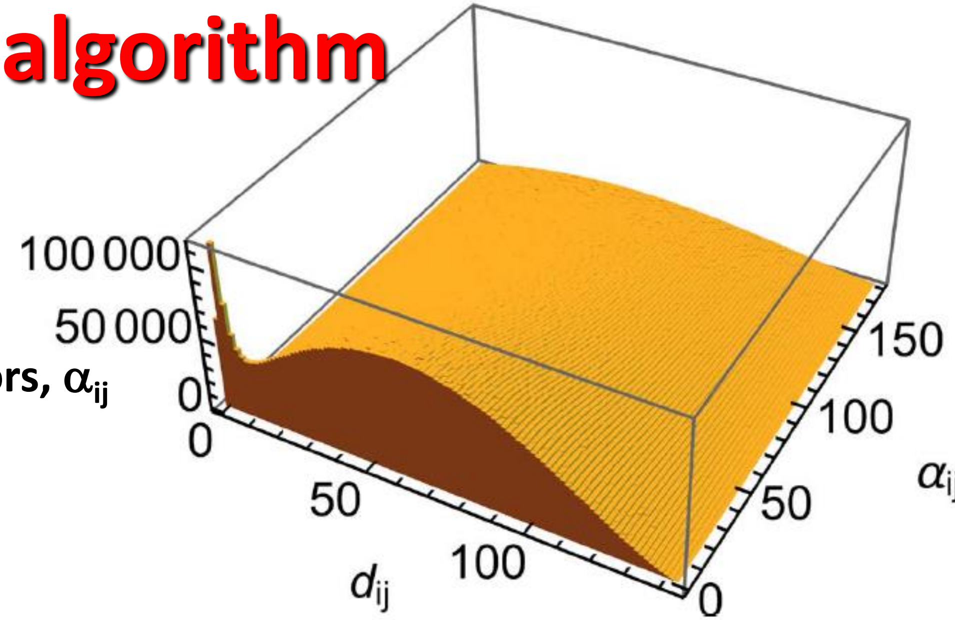


$$\mathcal{P}(M) = \frac{\lambda^M}{M!}\exp(-\lambda); \qquad \lambda = \langle M \rangle$$

R_1: sparse region, almost constant density, nearly Poisson ditribution

R_2: dense region, fluctuating density – great open clusters, imprint of Galactic spiral pattern…

Selection algorithm

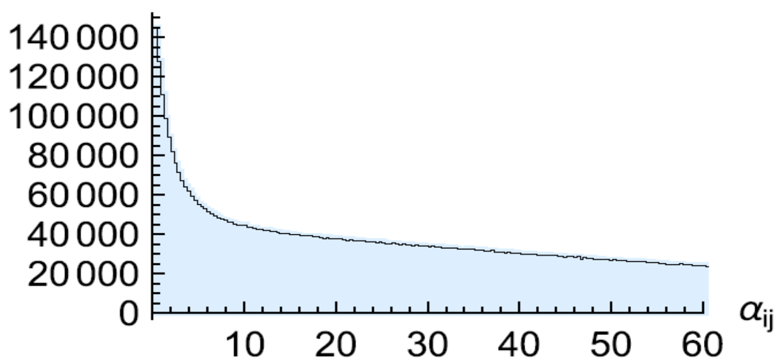


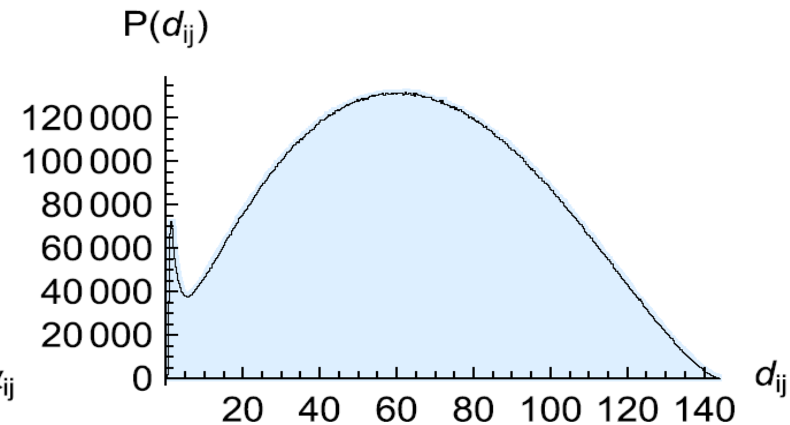
For binaries we expect:

- **Small (angular) separation, d_{ij} (angle between position vectors)**
- **Small angle between proper motion vectors, α_{ij} (common motion of bound stars)**

u **– proper motion, 2D angular velocity**

$$\alpha_{ij} = \arccos \frac{\boldsymbol{u}_i \cdot \boldsymbol{u}_j}{u_i u_j}$$





peak region defined by the cuts: $d_{ij} \leqslant 15\ as, \qquad \alpha_{ij} \leqslant 15^\circ$

Candidates are defined by these cuts

Background

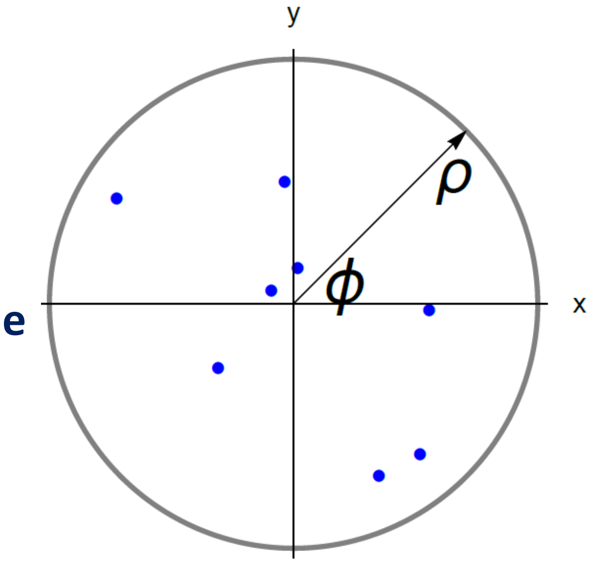
$$x_{ij} = |x_j - x_i|; \qquad y_{ij} = |y_j - y_i|;$$

$$d_{ij} = \sqrt{x_{ij}^2 + y_{ij}^2}; \qquad i, j = 1, 2, \ldots M,$$

2

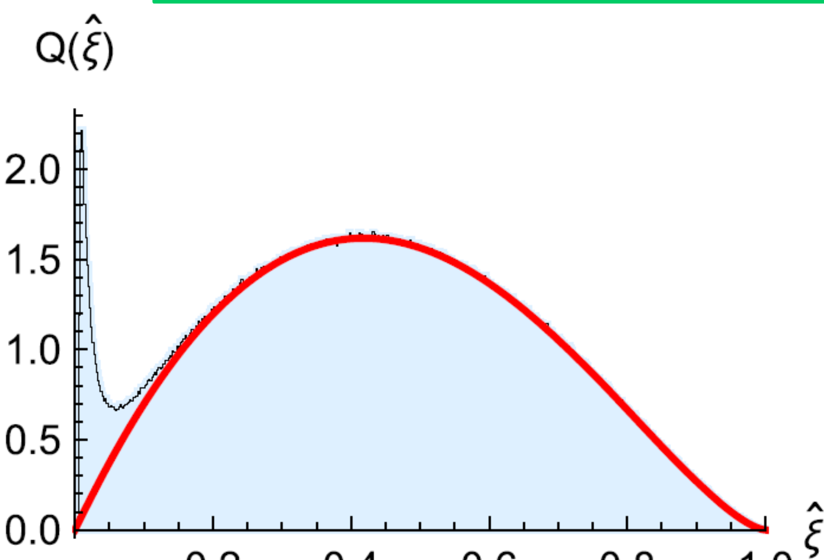
$$\hat{\xi} = \frac{d_{ij}}{2\rho_2}; \quad 0 \leqslant \hat{\xi} \leqslant 1,$$

BG is calculated as the distribution of separations of random points inside the circle (all pairs enter distribution).



$$q(\hat{\xi}) = \frac{16\hat{\xi}}{\pi}(\arccos\hat{\xi} - \hat{\xi}\sqrt{1 - \hat{\xi}^2})$$

see [1]



Random separations have universal distribution, does not depend on M.

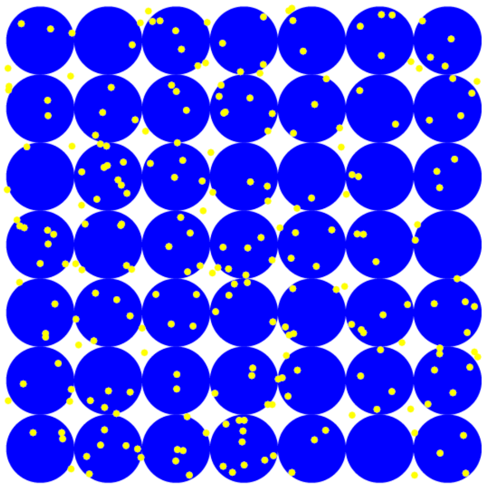
The red curve *q* precisely defines the background, the area of the peak above the curve represents the number of binaries.

Quality ratio: $\beta = n_P/(n_P+n_B)$ can be increased by additional cuts.

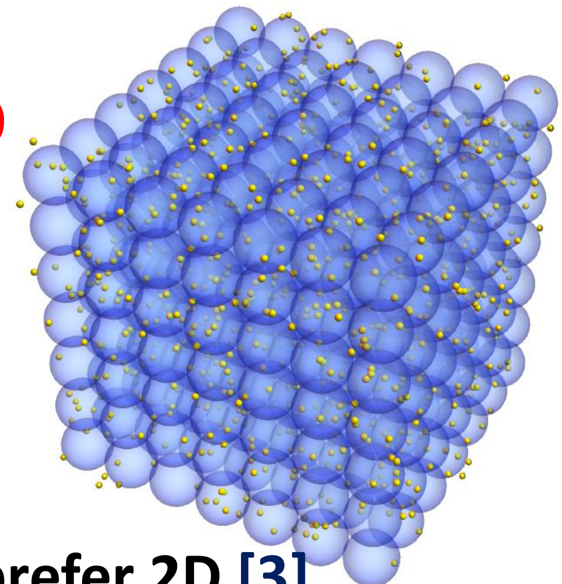
Remark:

We worked also with 3D grid (see [2]):

2D

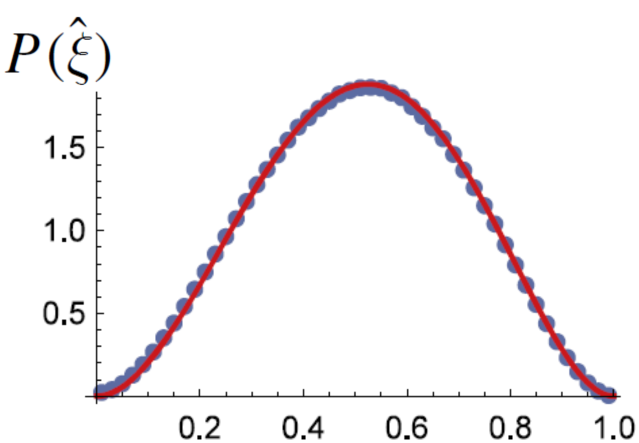


3D



3D turned out to be less effective, we prefer 2D [3]

Random separations in 3D: $P(\hat{\xi}) = 12\hat{\xi}^2(2 - 3\hat{\xi} + \hat{\xi}^3)$



$$\hat{\xi} = \frac{D_{ij}}{2\rho_3}$$

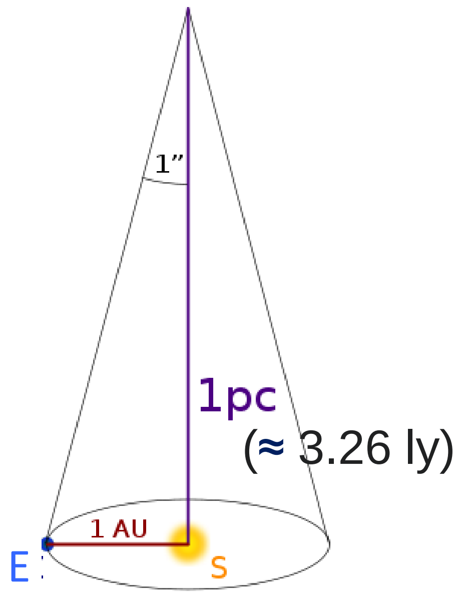
Disadvantage: due to the low accuracy of the radial separation, many true pairs exceed the diameter of the event ball. Such pairs are lost.

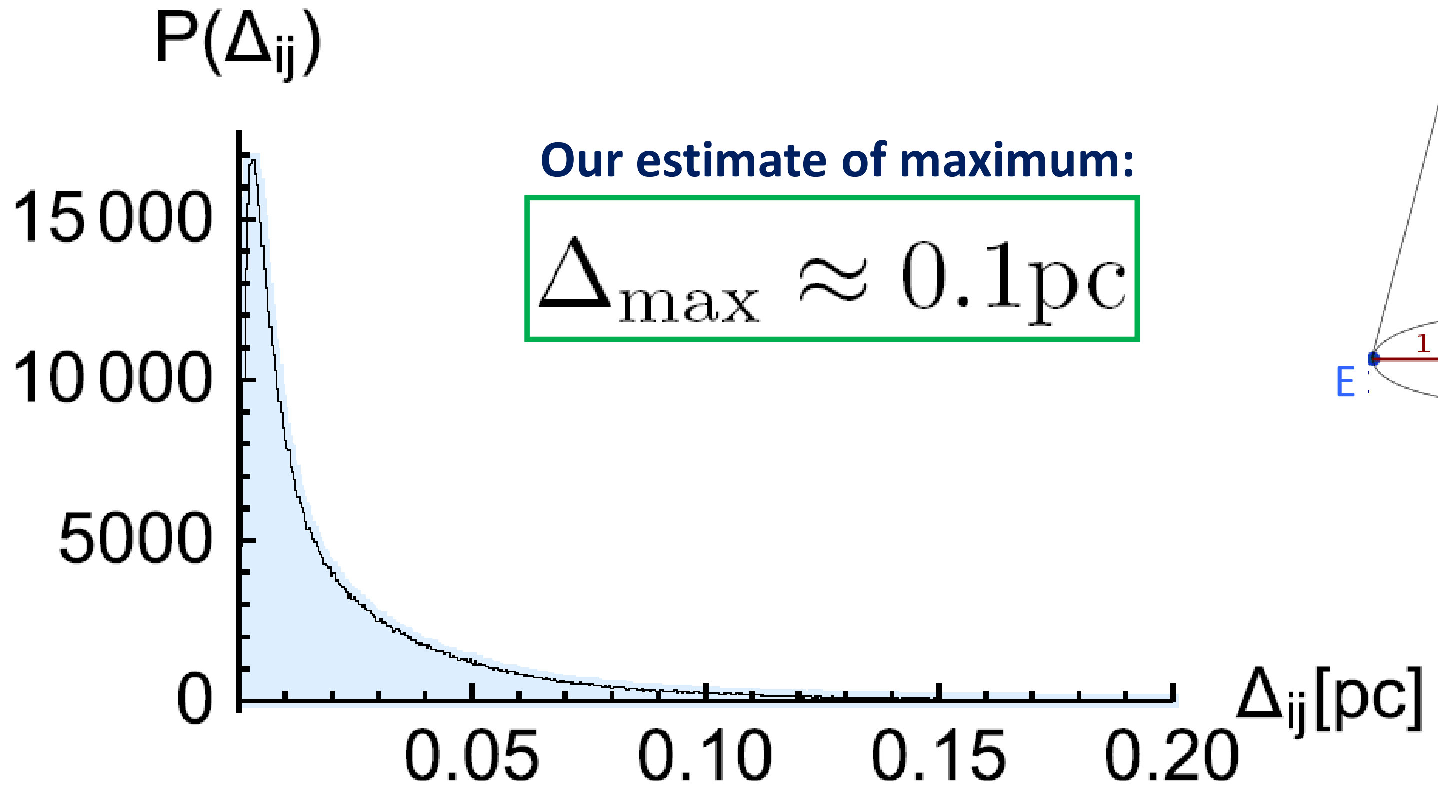
Results on wide binaries

1. Projected absolute separation (binary peak region)

Since we know parallax, we can calculate distance L
If d_{ij} = angular separation, then absolute separation of both stars:

$$\Delta_{\mathrm{ij}} = d_{\mathrm{ij}} \frac{L_i + L_j}{2}$$





Our estimate of maximum:

$$\Delta_{\mathrm{max}} \approx 0.1\mathrm{pc}$$

2. Periods and masses of binaries

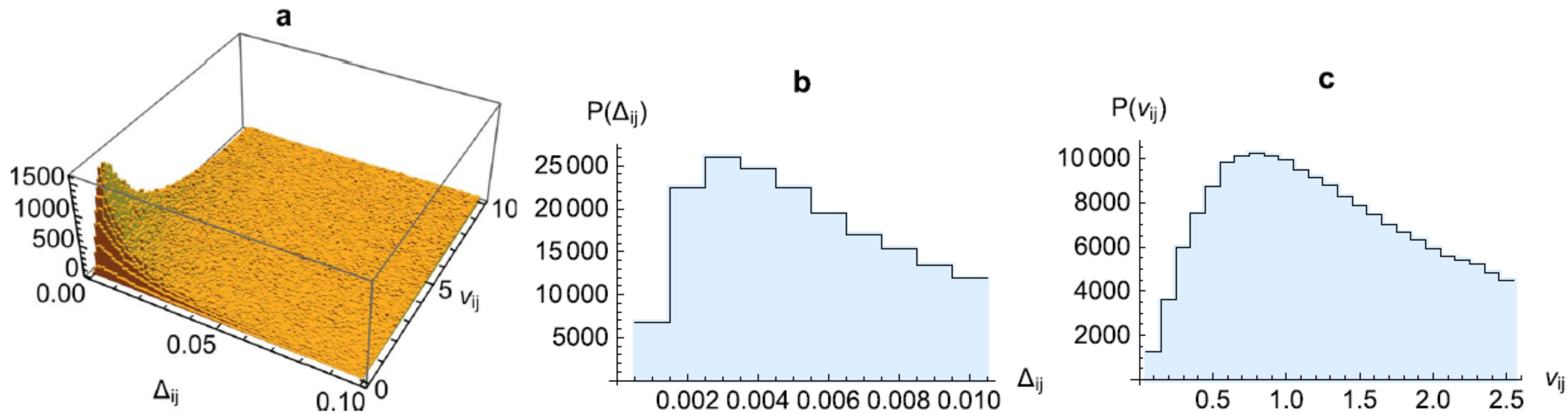


Figure 14. (a): correlation of the transverse separation Δ_{ij} with the transverse velocity v_{ij} of orbital motion in region $\boldsymbol{R}$. (b) and (c): distributions of Δ_{ij} and v_{ij} in domain (37). Units: Δ_{ij}[pc], v_{ij}[kms^{-1}]. Binning: 0.001 pc $\times$ 0.1 km s^{-1}, 0.001 pc, 0.1 km s^{-1}.

Orbital velocity estimation (projection):

$$v_{\rm ij} = |\boldsymbol{u}_i - \boldsymbol{u}_j|\frac{L_i + L_j}{2}$$

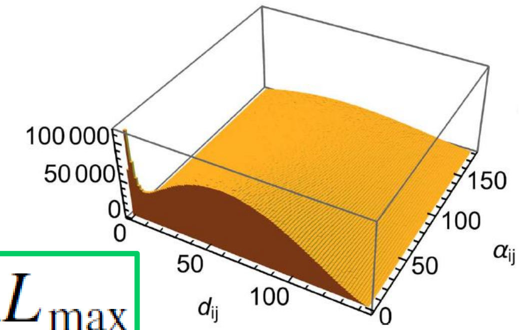
With the use of Kepler's law of periods

$$T_g = 2\pi\sqrt{\frac{a^3}{GM_{\rm tot}}}$$

we obtain approximate average [2,3]

$$\langle T \rangle \approx 4.2 \times 10^4 {\rm y}, \qquad \langle M_{\rm tot} \rangle \approx 0.65 M_\odot$$

3. Catalog of wide binaries



1. step: selection of pairs is defined the condition :

$$d \leqslant 15\ as, \qquad \alpha \leqslant 15^{\circ}, \qquad \Delta L \leqslant \Delta L_{max}$$

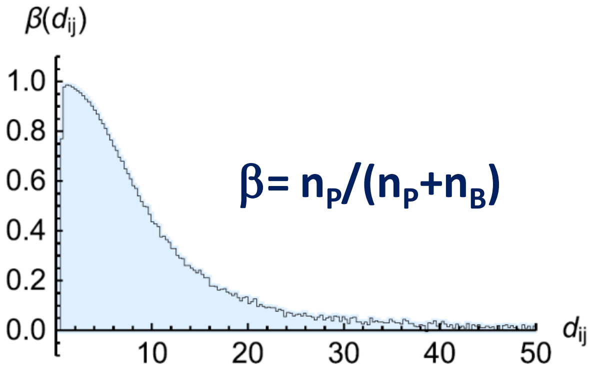
R_1:

$$\Delta L \leqslant \Delta L_{max} = 500 \text{ pc}$$

Cuts on Radial Separation in Galactic Longitude Subregions of Region $\boldsymbol{R}_2$

l[deg]	$\langle -30, +30 \rangle$	$\langle \pm 30, \pm 90 \rangle$	$\langle 90, 270 \rangle$
ΔL_{max} [pc]	50	100	400

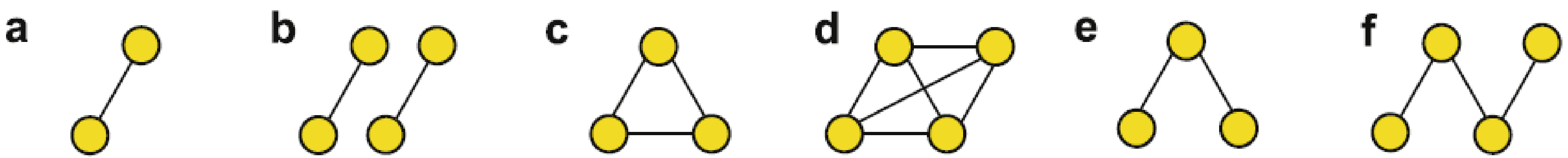
d, α define peak region,
ΔL reduces background. Calculation of corresponding background for this selection gives quality ratio ⟨β⟩ > 0,75



Too strict cuts generate a cleaner sample of binaries (higher ⟨β⟩), but more binaries are excluded. And vice versa, too soft cut preserves more binaries, but at the price of the higher background (lower ⟨β⟩).

2. step - wide Trinaries and Quaternaries:

Condition for binary (pair with a line) is satisfied by any pair in the bound system:

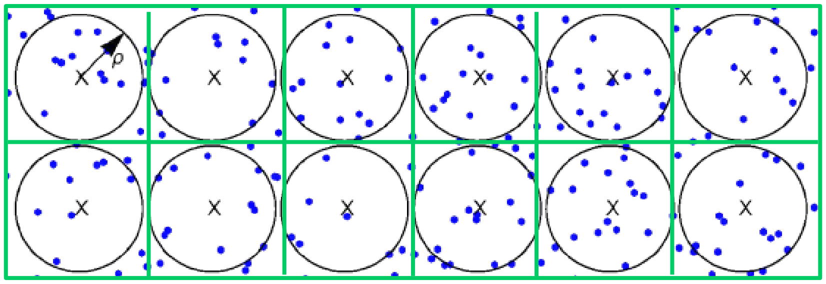


1 b 2 b 1 t 1 q ambiguous - rejected

More detailed procedure including evaluation of random background is described in [3].

3. step – from circle events to squares

We used squares instead of circles (full field coverage) for the final selection of candidates. Circle events were important for background calculation. Background level does not depend on event shape.



The resulting statistics (from 75 127 299 analysed stars)

m	N_m^A	β	n_m^A	Δn_m^A
2	900, 842	0.733	660, 317	696
3	5282	0.923	4875	67
4	30	1.	30	6

Number of multiple systems decreases with m:

$$\frac{n_3^A}{n_2^A} \approx 0.7\%, \qquad \frac{n_4^A}{n_3^A} \approx 0.6\%$$

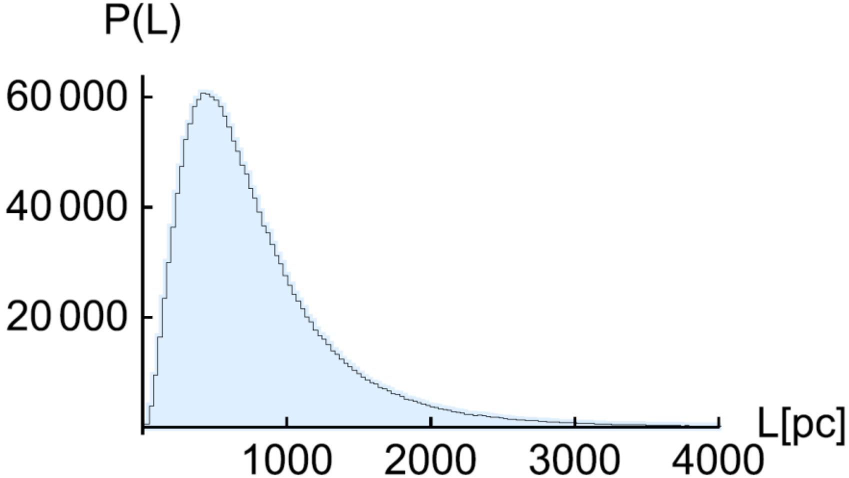


Figure 19. Distances of all candidate sources. Binning: 30 pc.

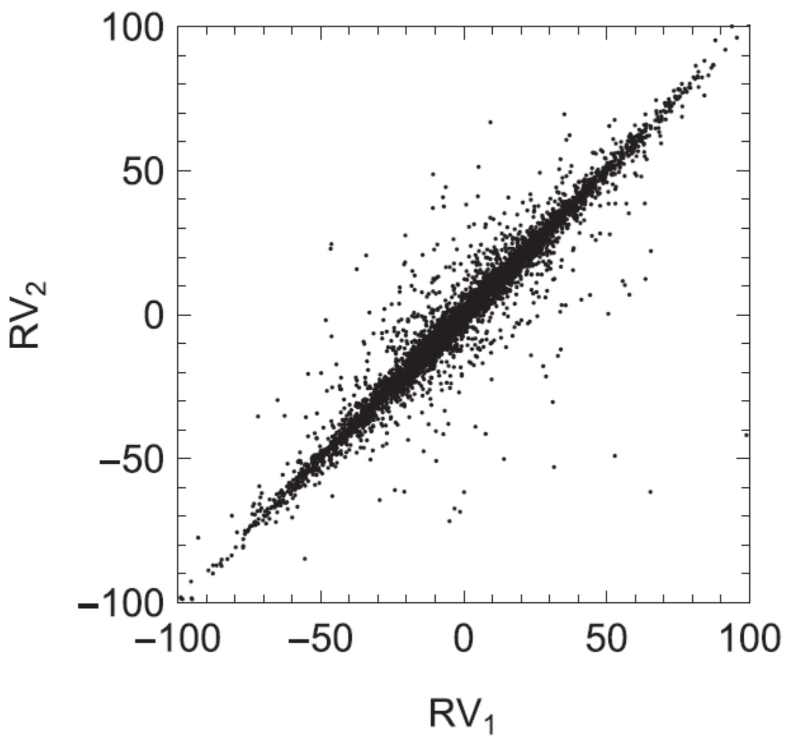


Figure 20. Correlation of radial velocities for 6469 pairs. Unit: [km s^{-1}].

Gaia data involve limited number of stars of known radial velocity. Binaries should have similar RV. Points (RV_1, RV_2) in one line prove it.

Comparison with others

Numbers of the Binary Candidates in the Compared Catalogs

	$N_{\rm tot}$	$N_{b>25}$	Reference/DR
A1	900,842	900,842	this paper/DR2
A2	1,256,400	496,888	El-Badry et al. (2021)/EDR3
A3	93,898	55,319	Hartman & Lépine (2020)/DR2
A4	80,560	40,107	Zavada & Píška (2020)/DR2
A5	3055	381	Jiménez-Esteban et al. (2019)/DR2
A6	9977	5546	Sapozhnikov et al. (2020)/DR2

- **A critical comparison of the catalogues is described in more detail in [3]. Our A1 is comparable to A2, and we have 2x more candidates in the region|b|>25. Our algorithm involves accurate background estimation, which implies a strong background in the |b|<25 region - especially for distant sources.**
- **Our catalogs A1, A4 are available at https://www.fzu.cz/~piska/Catalogue/**
- **For practical use, we have also created a merged catalog A1+A3+A4+A5+A6.**
- **We plan our A1 reprocess with next data release DR4, in the enlarged region, and using the further optimized algorithm.**

What next?

- We continue to work with Gaia data, our interest is focused on MW kinematics.
- We work with angular velocities (proper motion) – projection of 3D motion on celestial sphere.
- Despite of reduced information on motion we are able to reconstruct important parameters, like V_{sun} , or rotational curves: V(R), V(|Z|)…
 (see PZ, KP arXiv: 2308.11060)

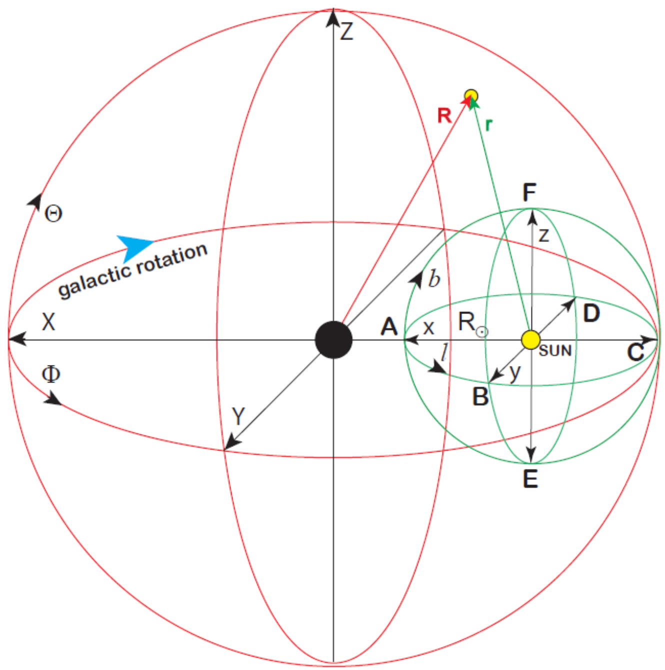


FIG. 1: Galactic (green) and Galactocentric (red) reference frames

Summary and conclusion

- Inspired by particle physics, we developed a statistical method for analysis of 2D & 3D patterns. The method can detect subtle deviations from random distributions, like a tendency to (anti-) clustering.

- This methodology has been applied to the detection of wide binaries and multiple star systems in Gaia data (|b|>25deg). Main results:

- Separation of binaries is limited roughly by $\Delta_{max} \approx 0.15$ pc.

- For wide binaries in Gaia data (d>0.5'') we estimate mean values

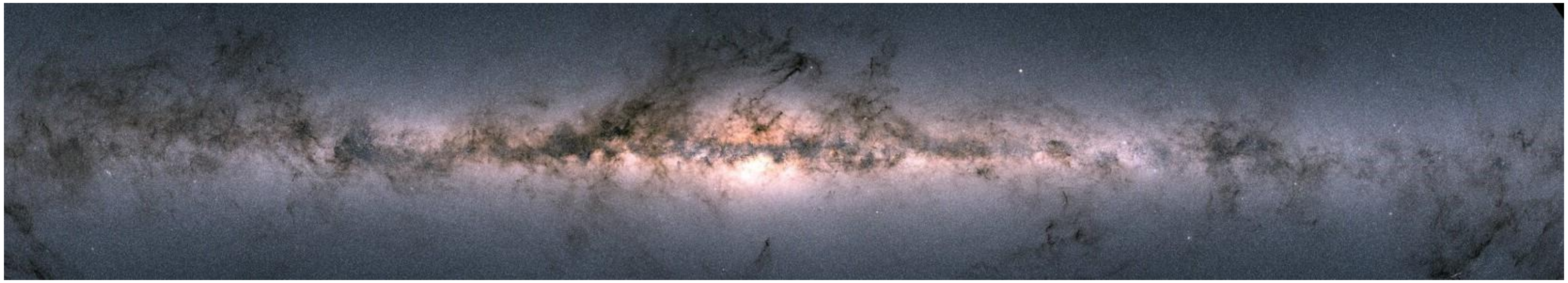
$$\langle T \rangle \approx 4.2 \times 10^4 \text{ y}, \qquad \langle M_{tot} \rangle \approx 0.65\, M_{Sun}$$

- We have created extensive catalogs of wide binary, trinary and quaternary candidates.

- We have started statistical analysis of MW kinematics.

Thank you for your attention!

Buckup slides



Step back to HI

In addition to the MW research, we plan to more thoroughly test the usefulness of the characteristic functions method in the ALICE HI data (my student V.Macháček).

Event parameters

Conditions for generating events:

- The radius ρ_2 must be significantly larger than a typical angular separation of true binary. At the same time, it must be so small that the distribution of stars within the event can be considered random and uniform.

- Events with too high multiplicity, M, in which various dense structures may dominate, are excluded from processing. To improve the quality ratio, additional cuts can be imposed (e.g. radial separation).

- The circular shape of the events is chosen due to the accurate formula for calculating a random background. Another shape would lead to a more complex function depending on other shape parameters (triangle, square, orientation ...).