The architecture of planetary systems - a machine learning approach

Yann ALIBERT

ISPM - 28/10/2022

The National Centres of Competence in Research (NCCR) are a research instrument of the Swiss National Science Foundation

UNIVERSITÄT BERN
CSH CENTER FOR SPACE AND HABITABILITY

1- Planet formation and the Bern model

2- Models *versus* observations

3- Planetary internal structure and Deep Learning

4- Correlation in planetary systems with random forrest: finding a second Earth

UNIVERSITÄT BERN

CSH CENTER FOR SPACE AND HABITABILITY

1- Planet formation and the Bern model

2- Models *versus* observations

3- Planetary internal structure and Deep Learning

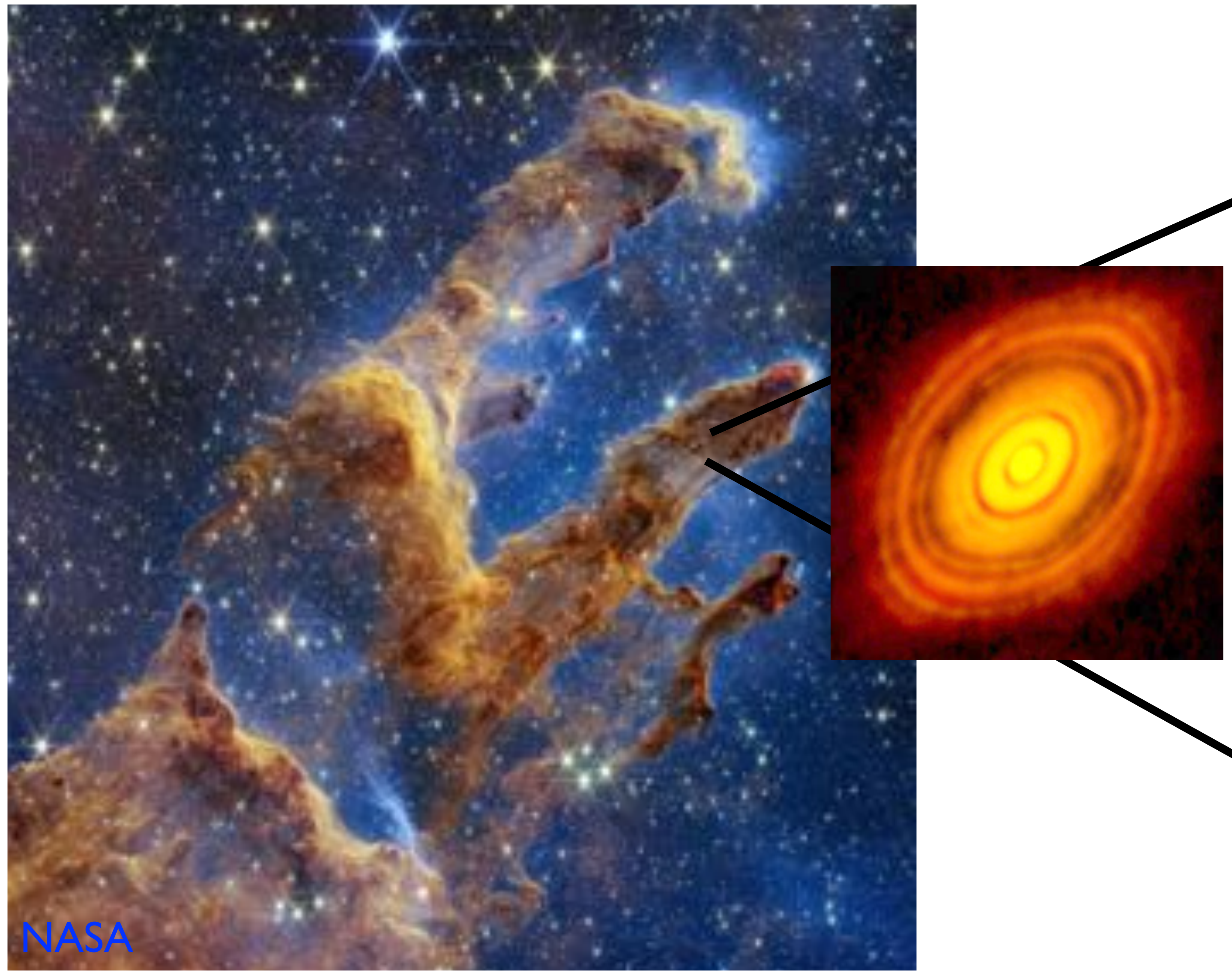
4- Correlation in planetary systems with random forrest: finding a second Earth

From giant molecular clouds to planetary systems

UNIVERSITÄT BERN
CSH CENTER FOR SPACE AND HABITABILITY

Giant Molecular Cloud

Light years ~10^5 AU



Protoplanetary disk

ALMA

~100s AU

99% gas   1% solids

$T_{\rm disk} < 10^7 {\rm yr}$

Solar System

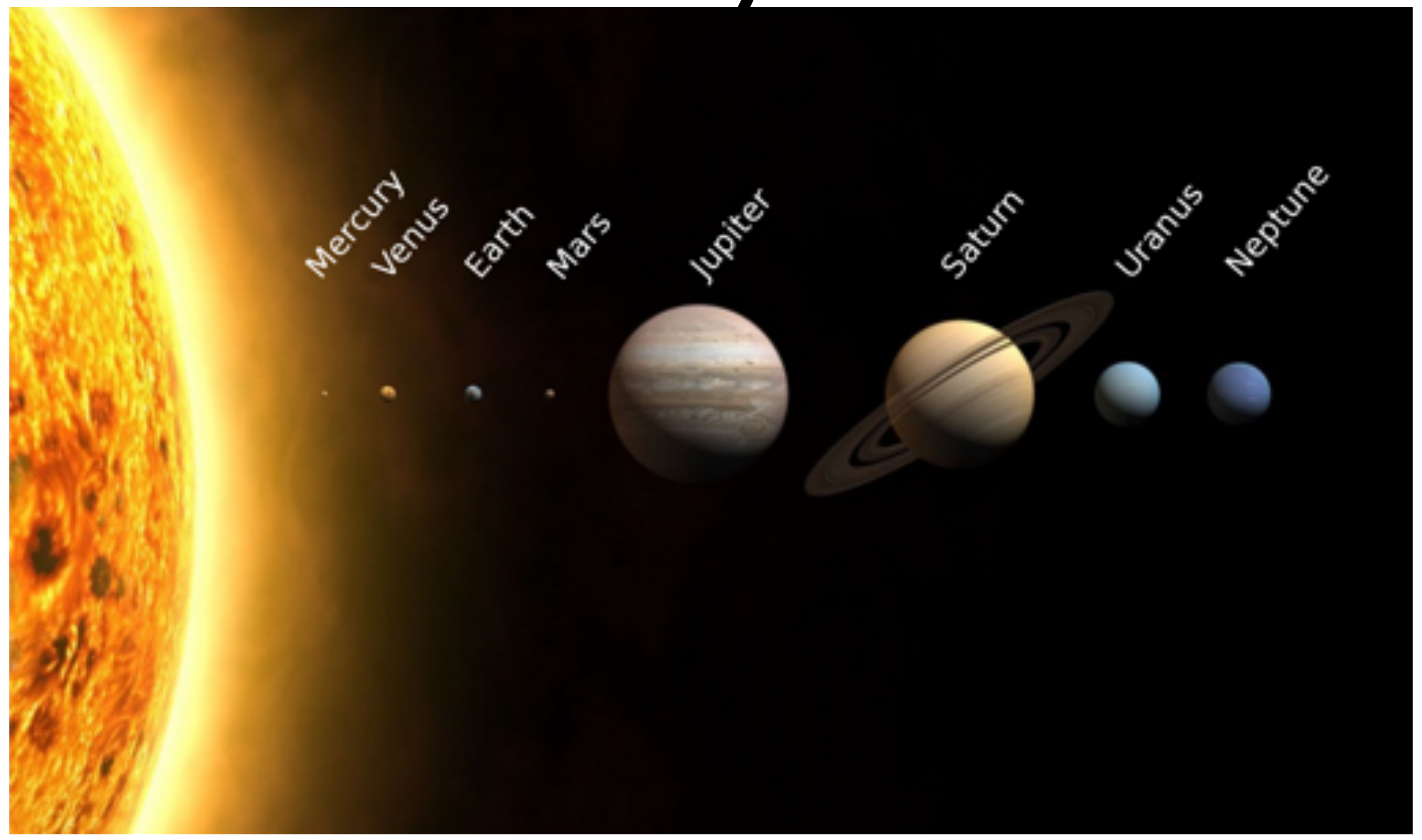


Image by WP - Planets2008.jpg, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=45708230

Exoplanets

The core accretion model (1)

UNIVERSITÄT BERN

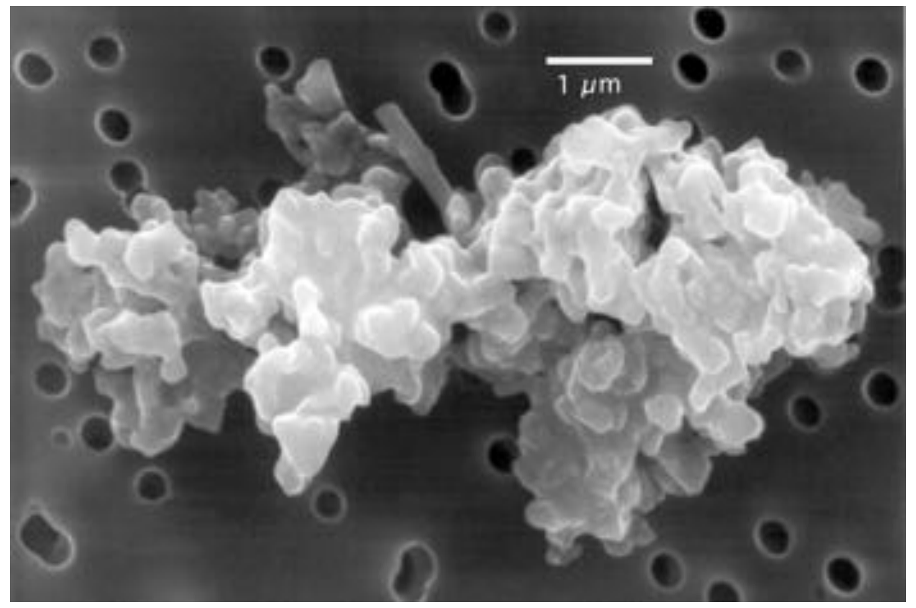
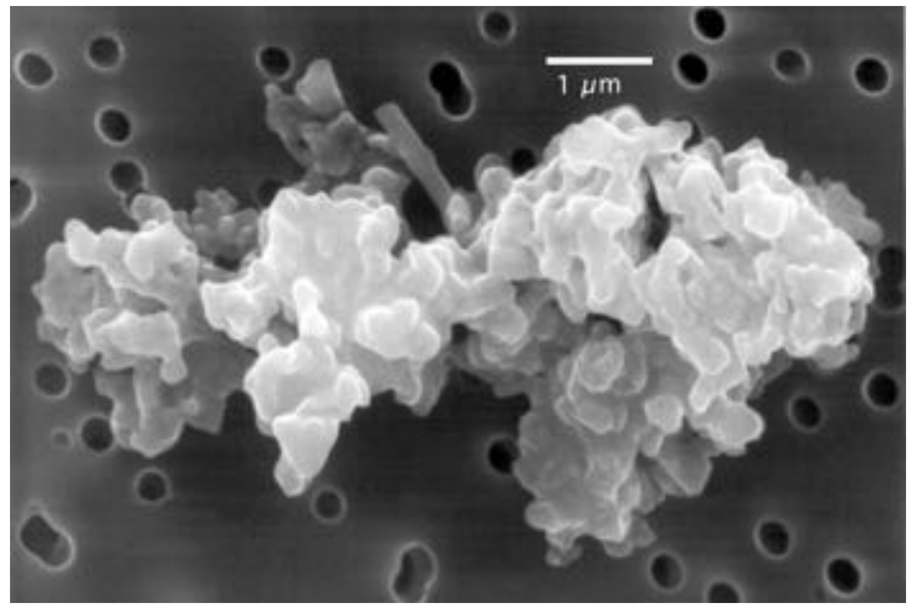
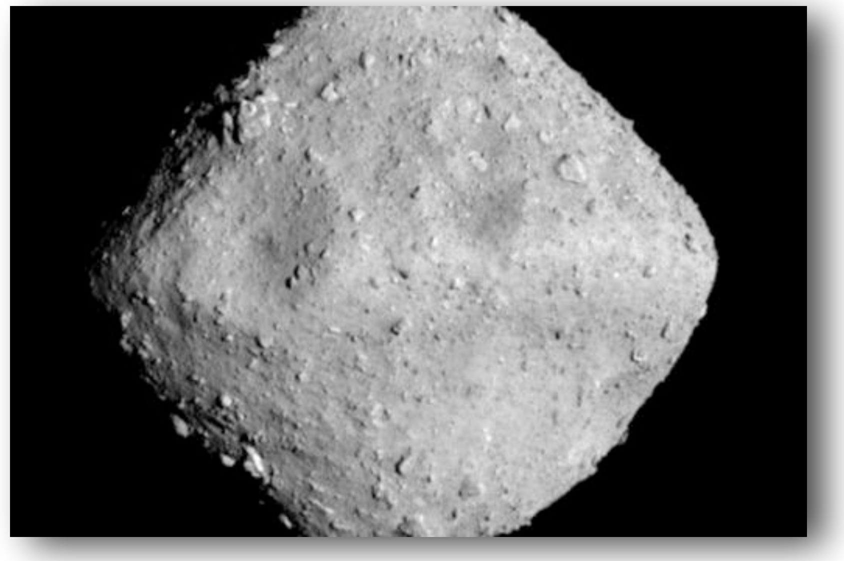
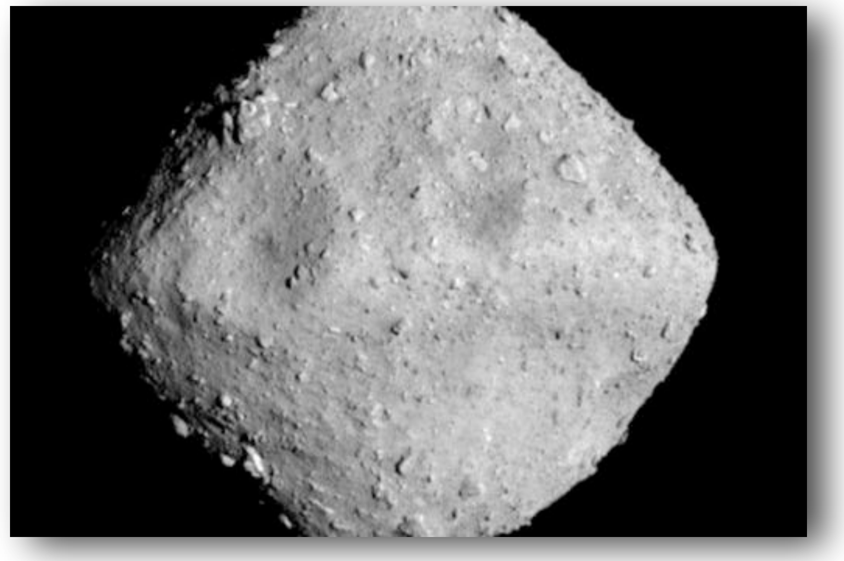
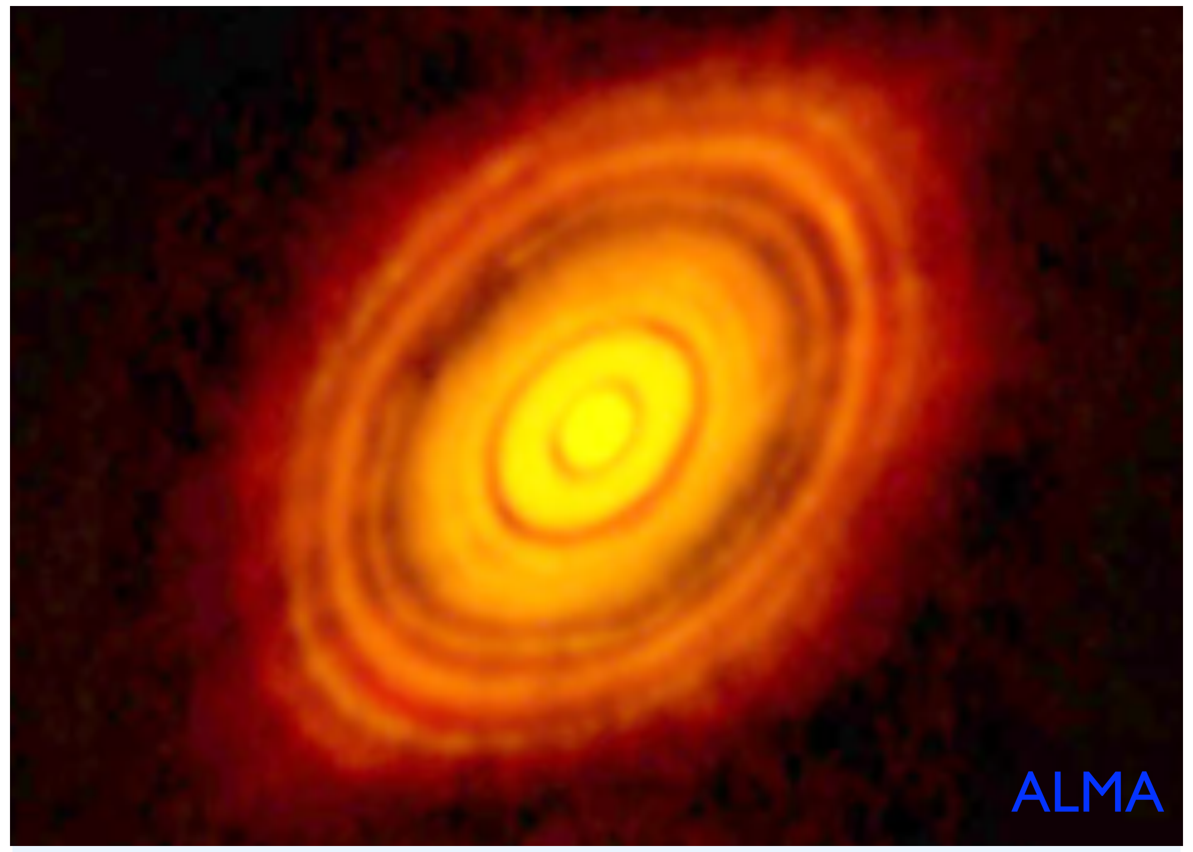
Dust		Pebbles		Planetesimals		Planet embryo
 + 	Collisional ⇒ growth	 + 	Streaming ⇒ instability	 + 	Collisional ⇒ growth	
Micrometer		Centimeter		100m-100km		1000 km

Image credit:
Dust: By The original uploader was Amara at English Wikipedia. - Transferred from en.wikipedia to Commons by Common Good using CommonsHelper., CC BY 1.0, https://commons.wikimedia.org/w/index.php?curid=6267225
Pebble: Eco outdoor
Ryugu: JAXA

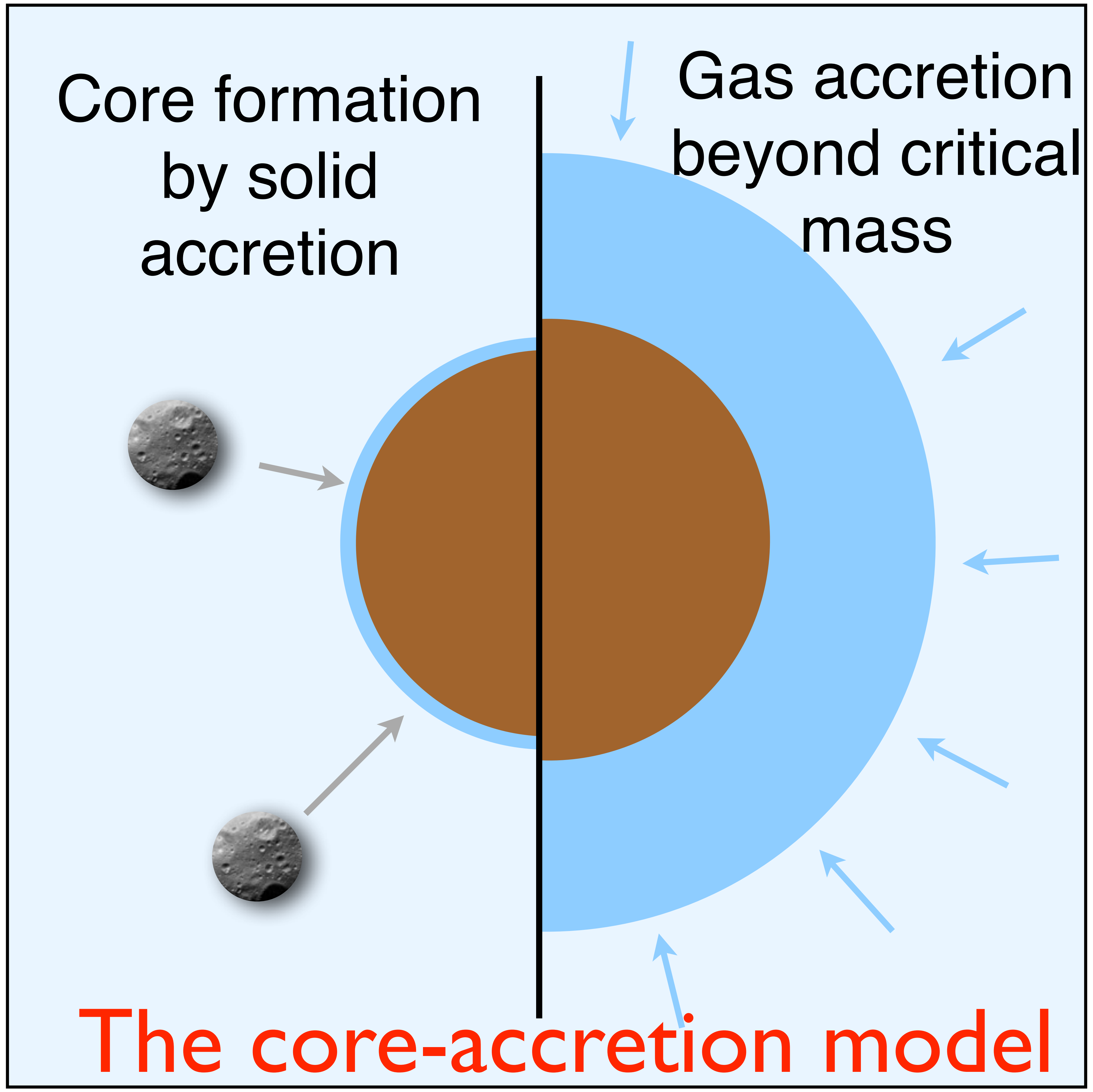
The core accretion model

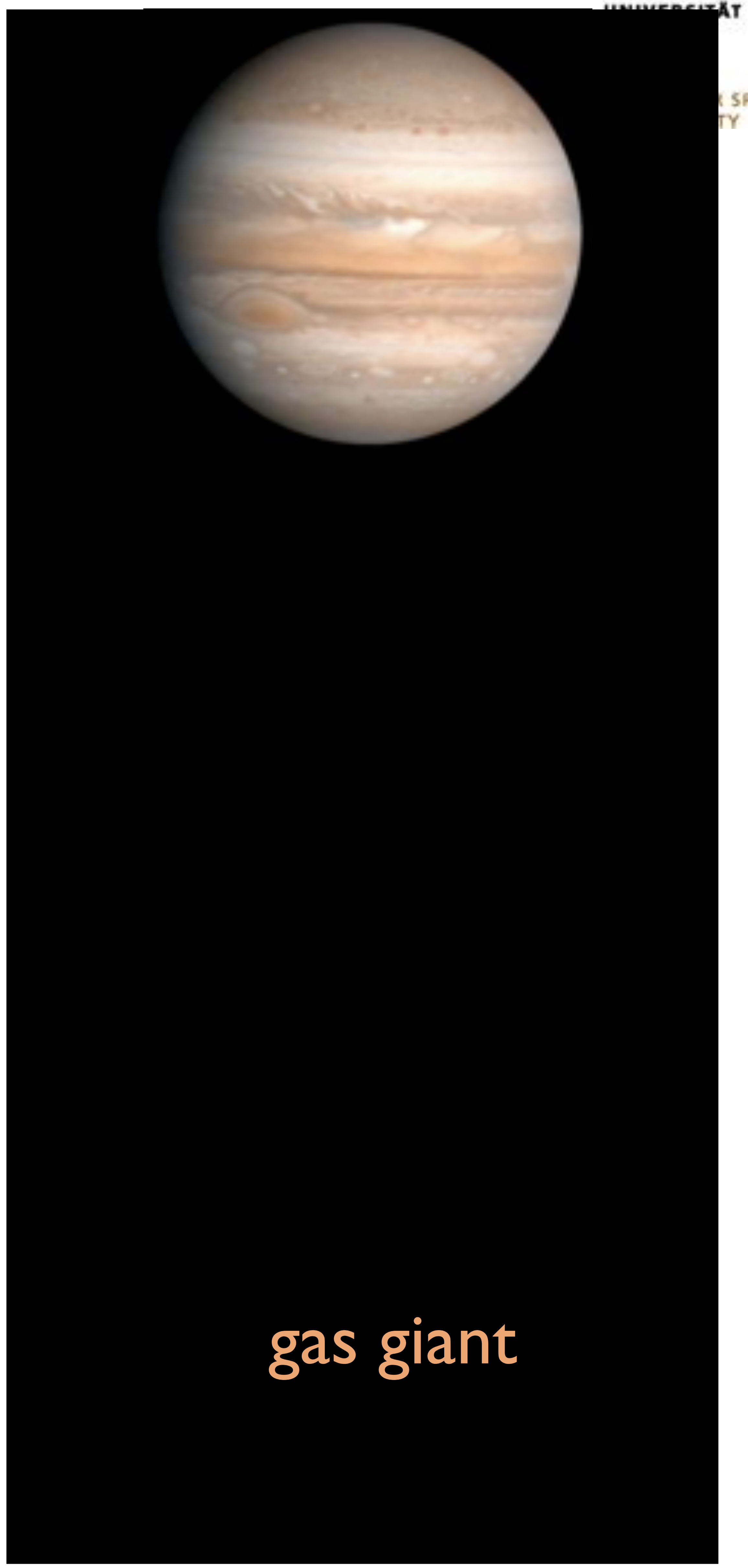


~100s AU

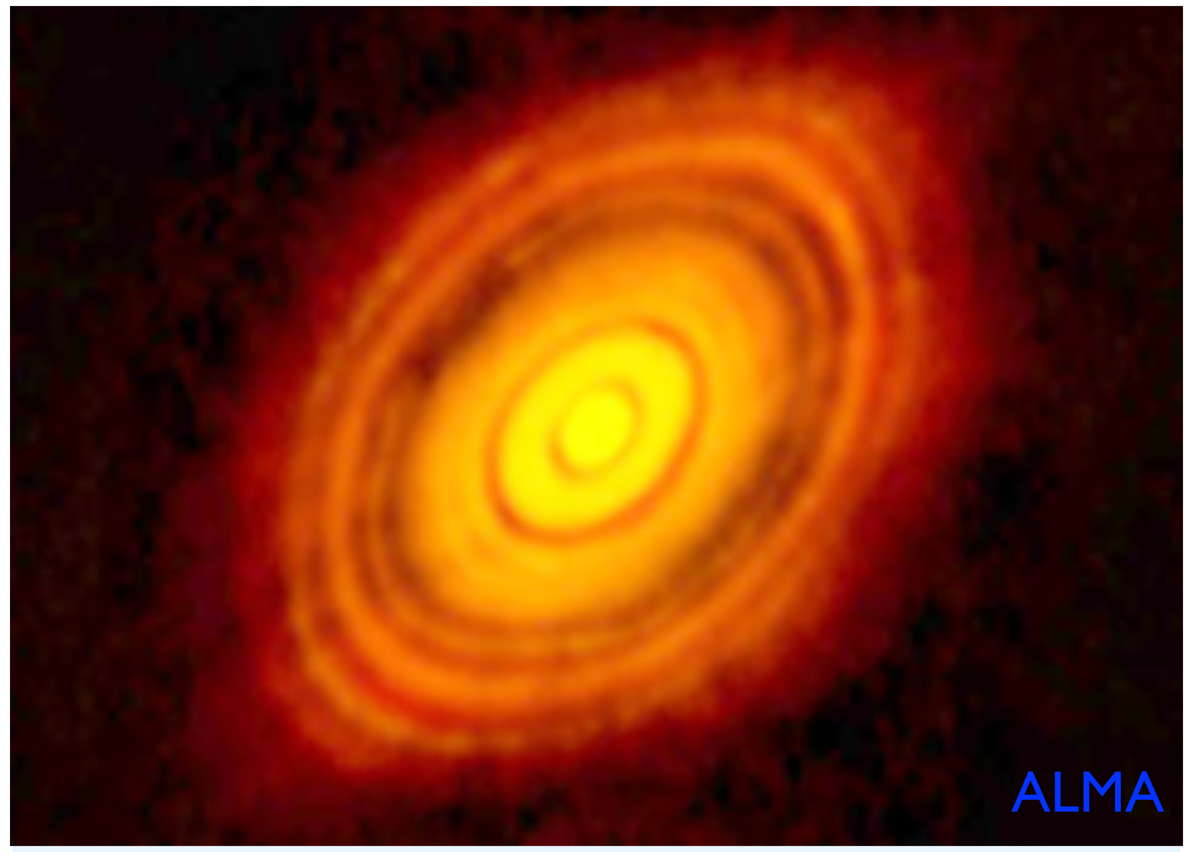
99% gas 1% solids

$T_{\rm disk} < 10^7 \rm yr$





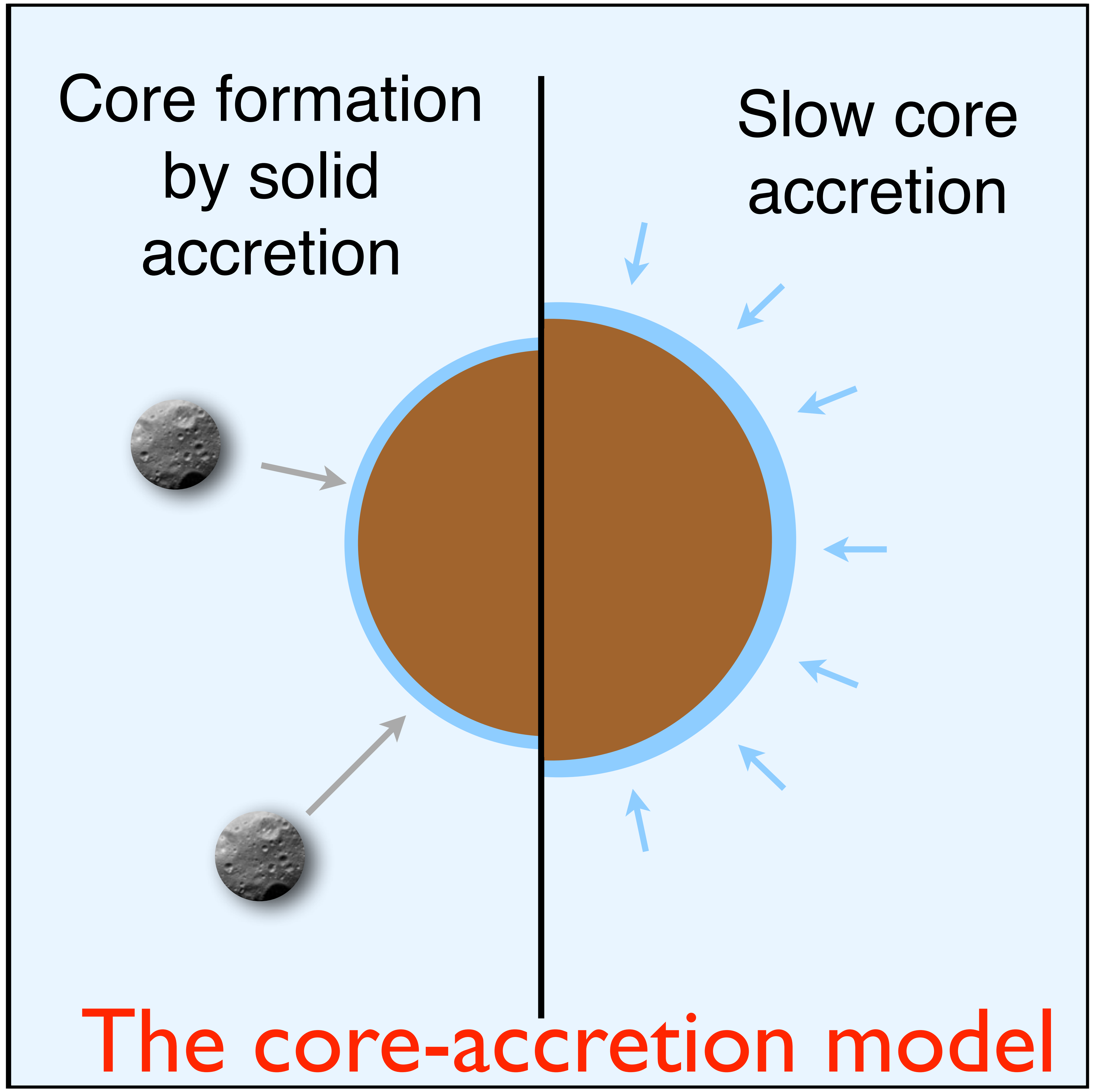
The core accretion model

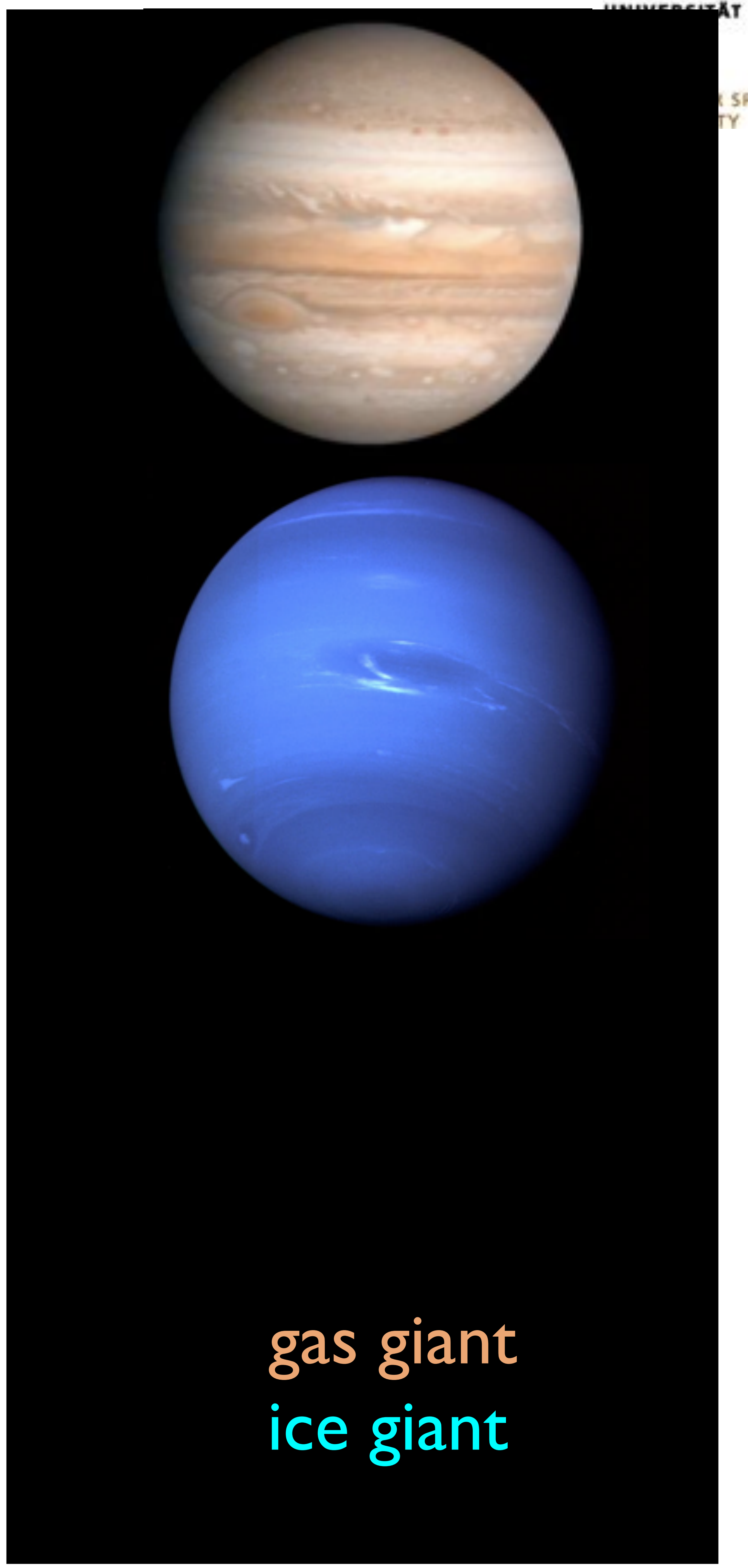


~100s AU

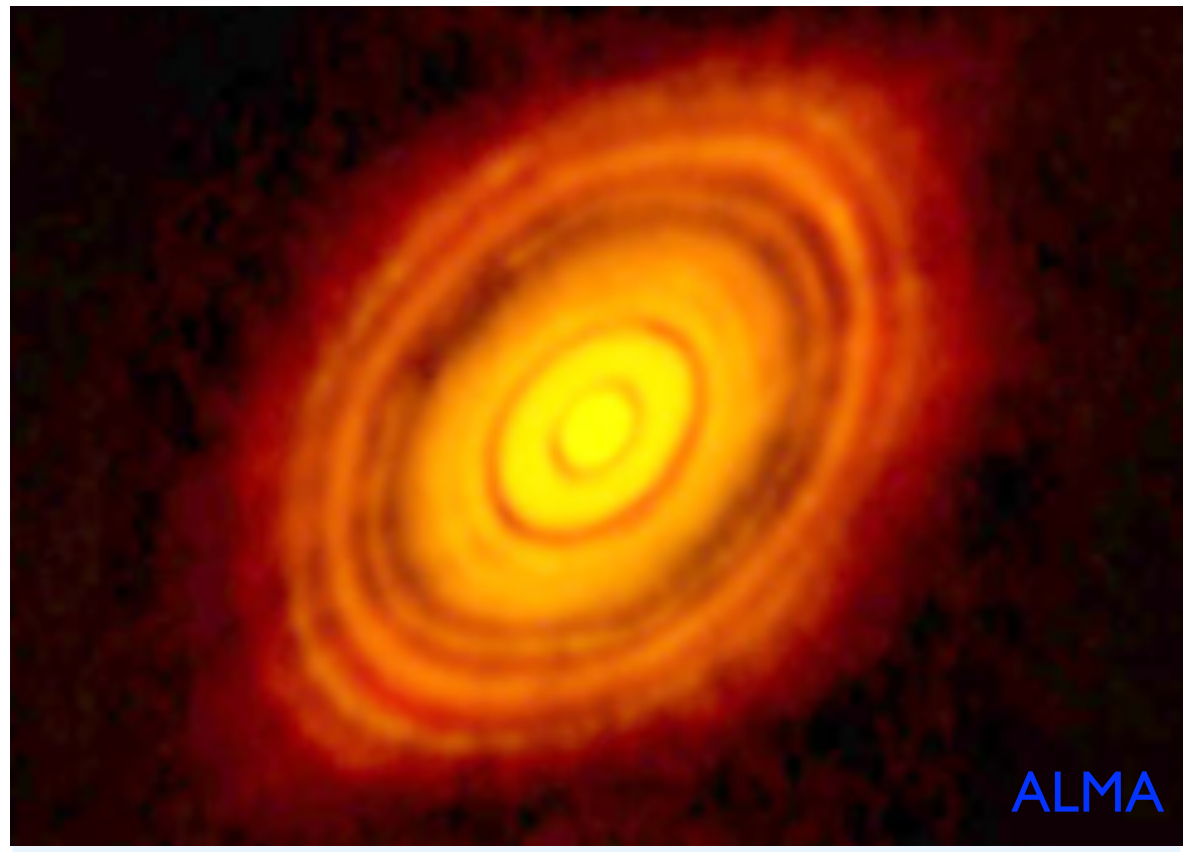
99% gas 1% solids

$T_{\rm disk} < 10^7 {\rm yr}$





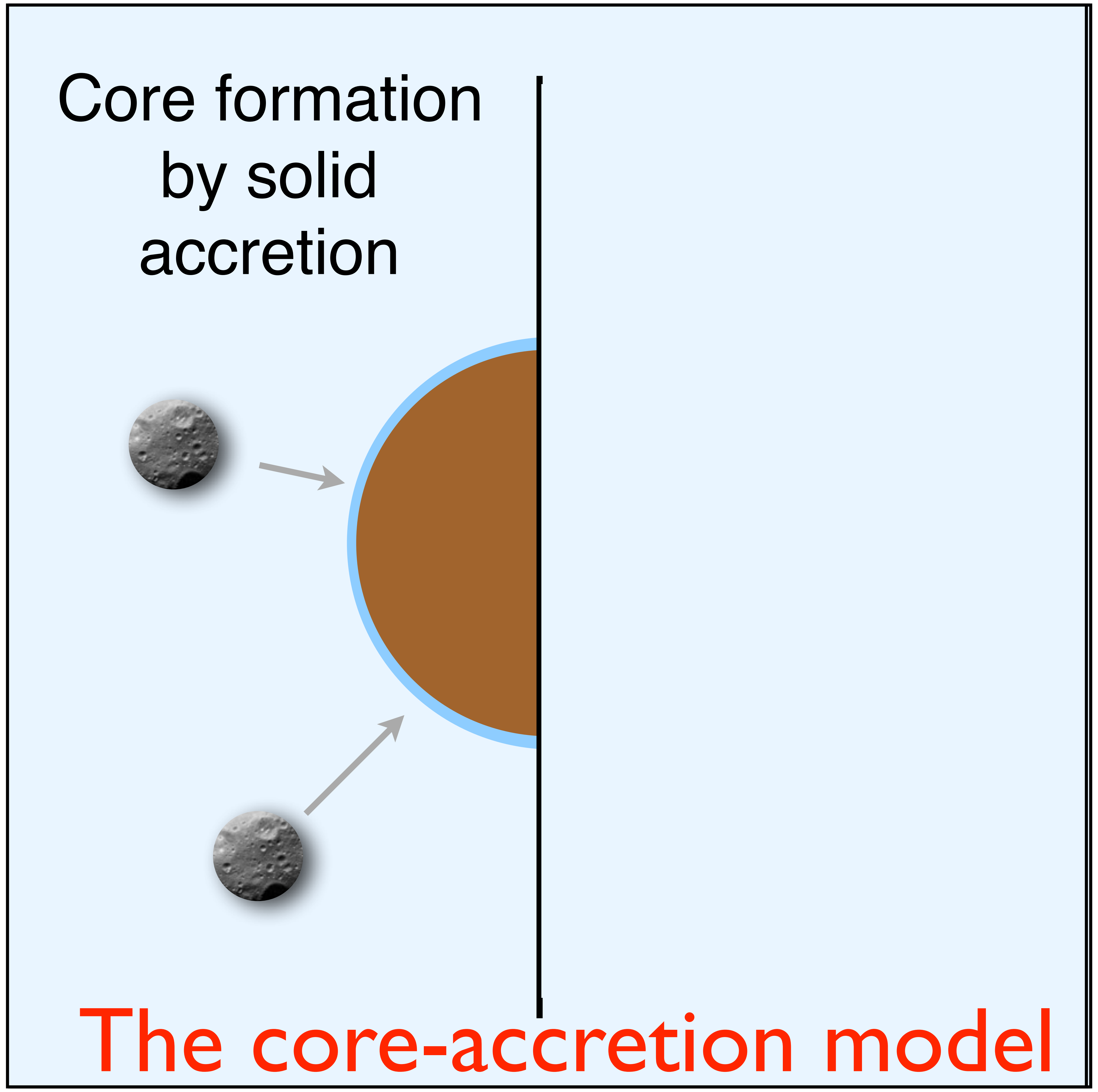
The core accretion model

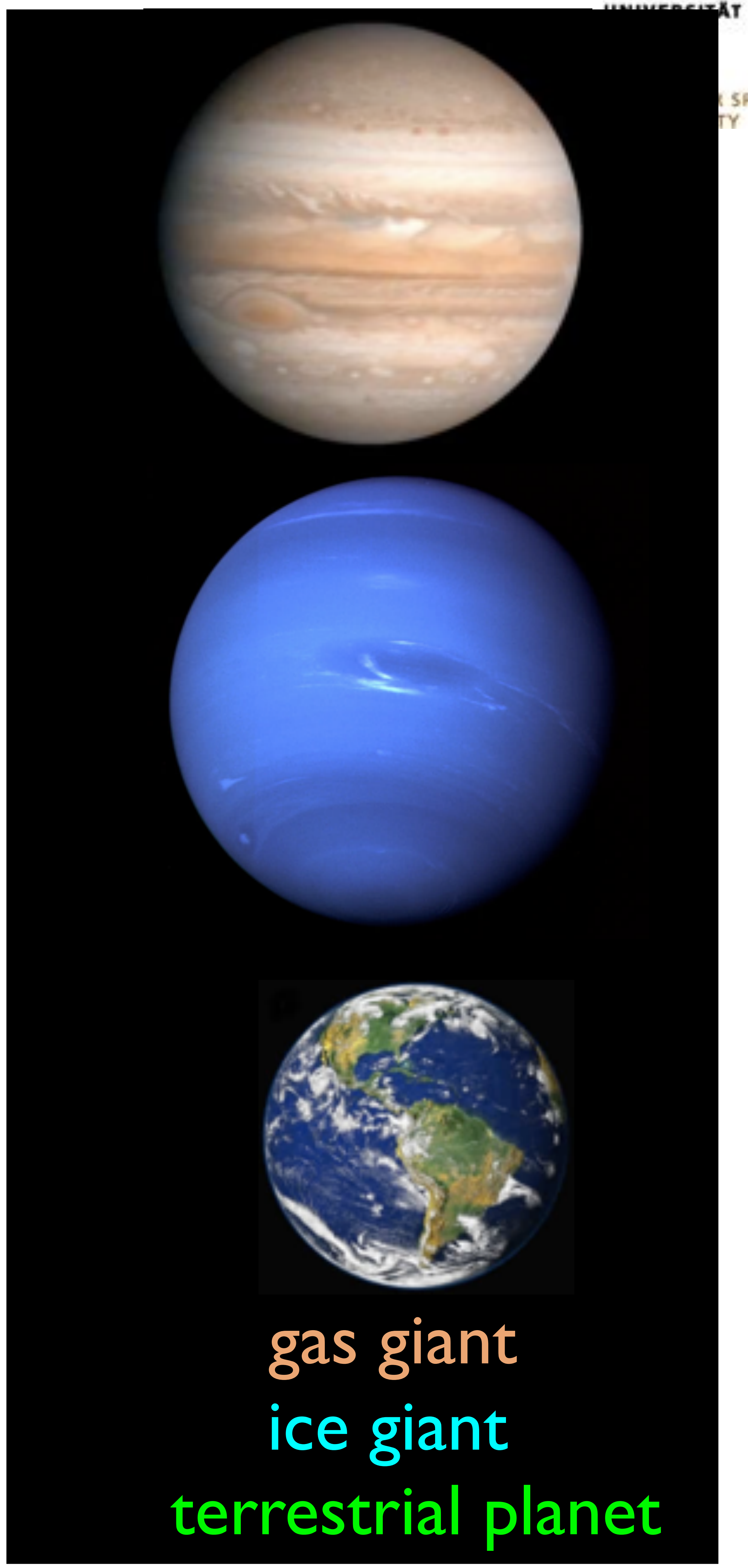


~100s AU

99% gas   1% solids

$T_{\rm disk} < 10^7 {\rm yr}$

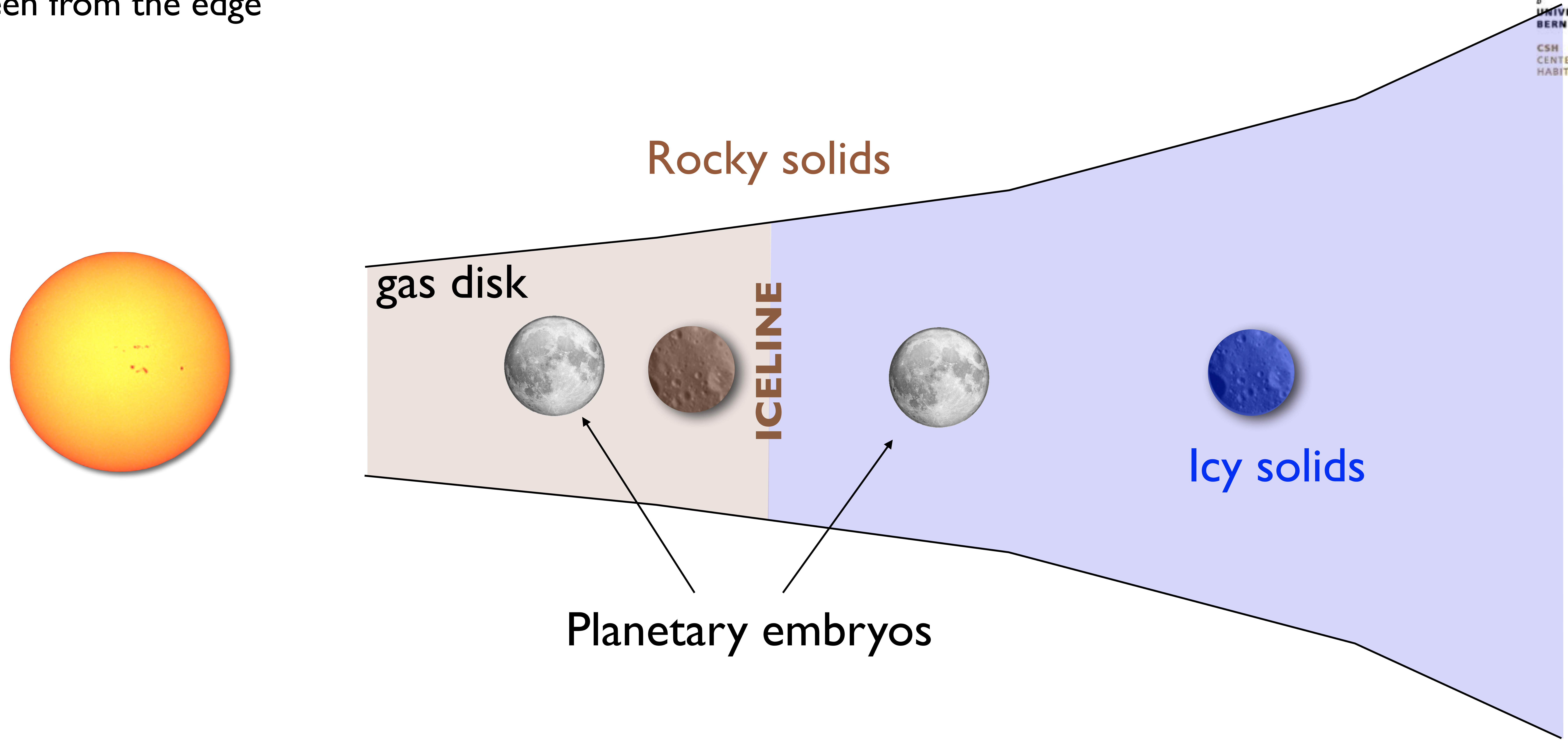




Jupiter: NASA/JPL/USGS
Neptune: NASA/JPL
Earth: NASA

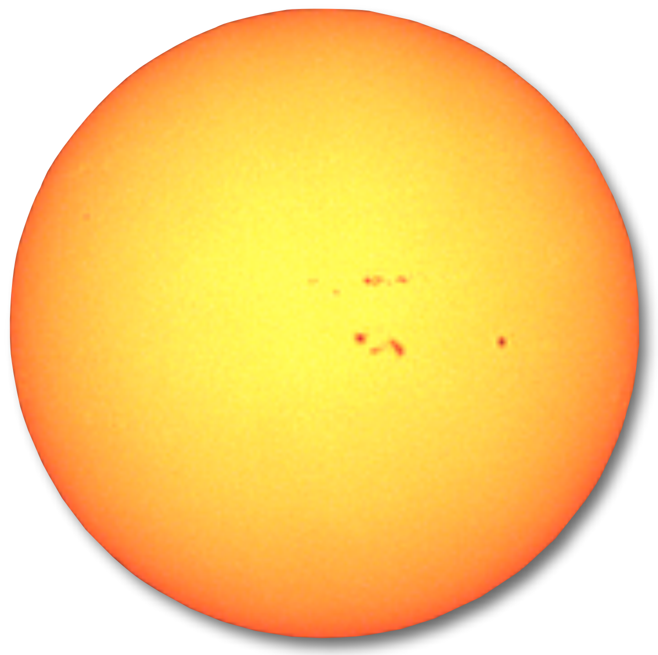
The Bern model

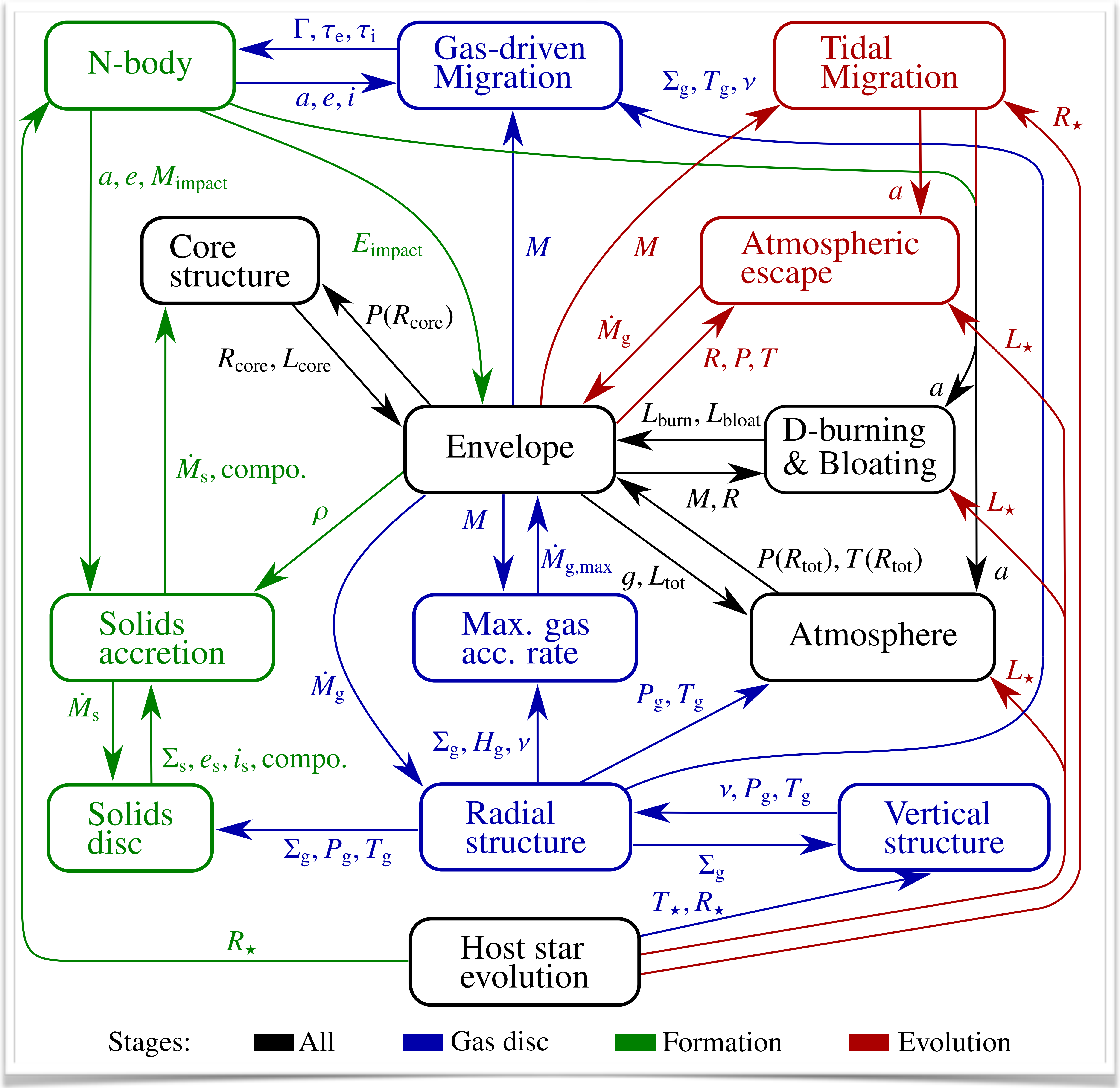
Seen from the edge

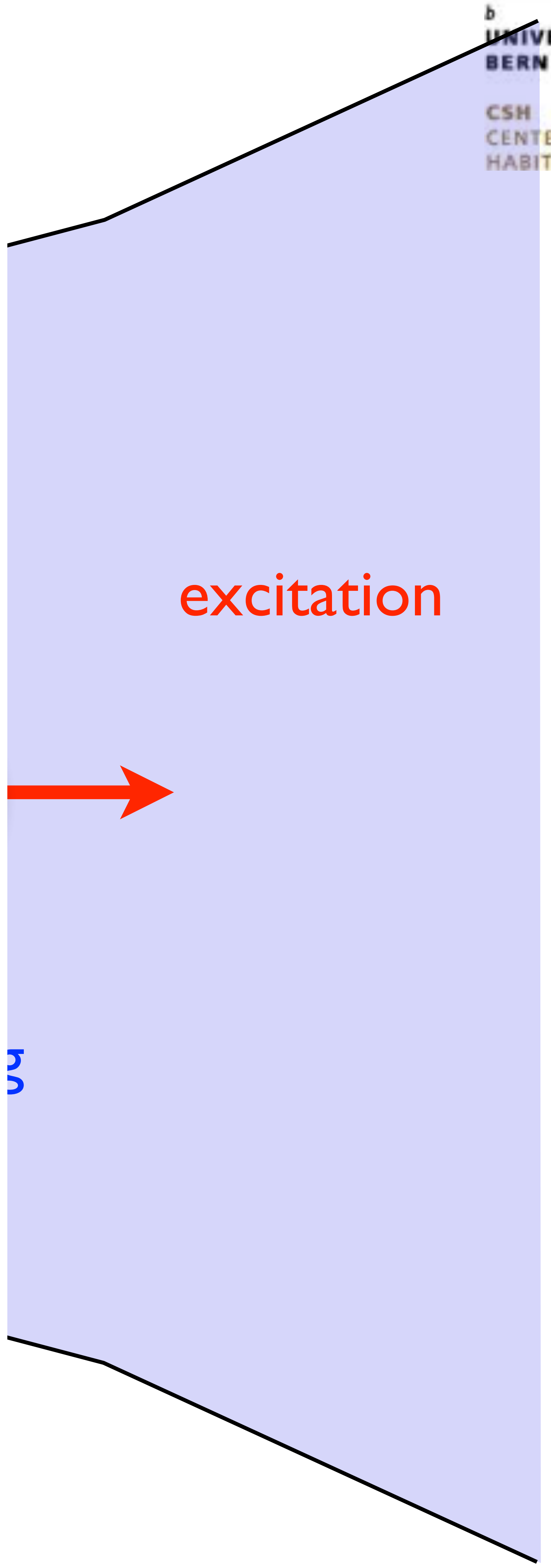


protoplanetary disk = gas disk + solids

The Bern model







The Bern model - planetary population

UNIVERSITÄT BERN
CSH CENTER FOR SPACE AND HABITABILITY

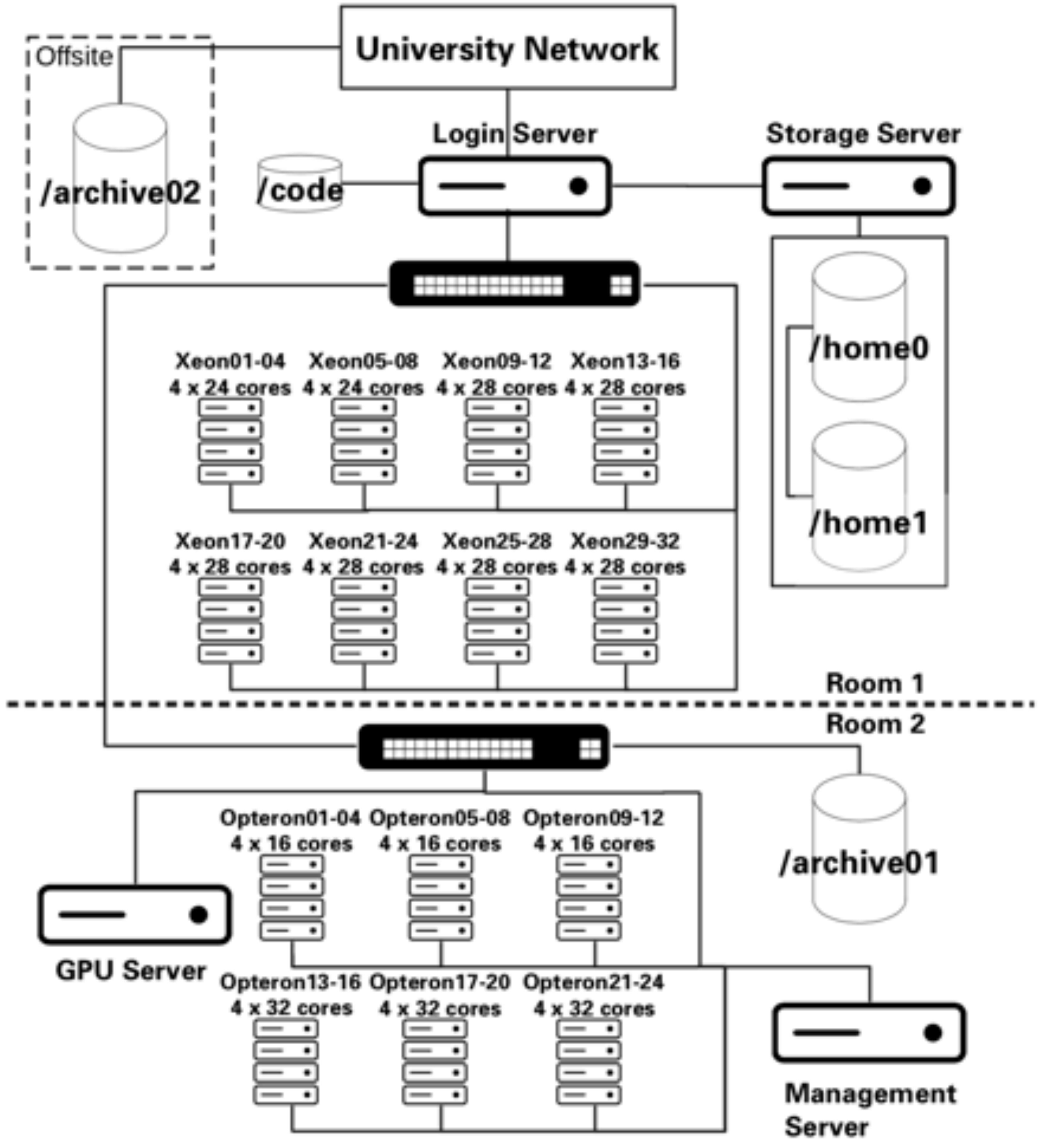
1000 systems

Starting with 100 planetary embryos

All stars are solar

Disk properties follow observations

~1 Million CPU hours

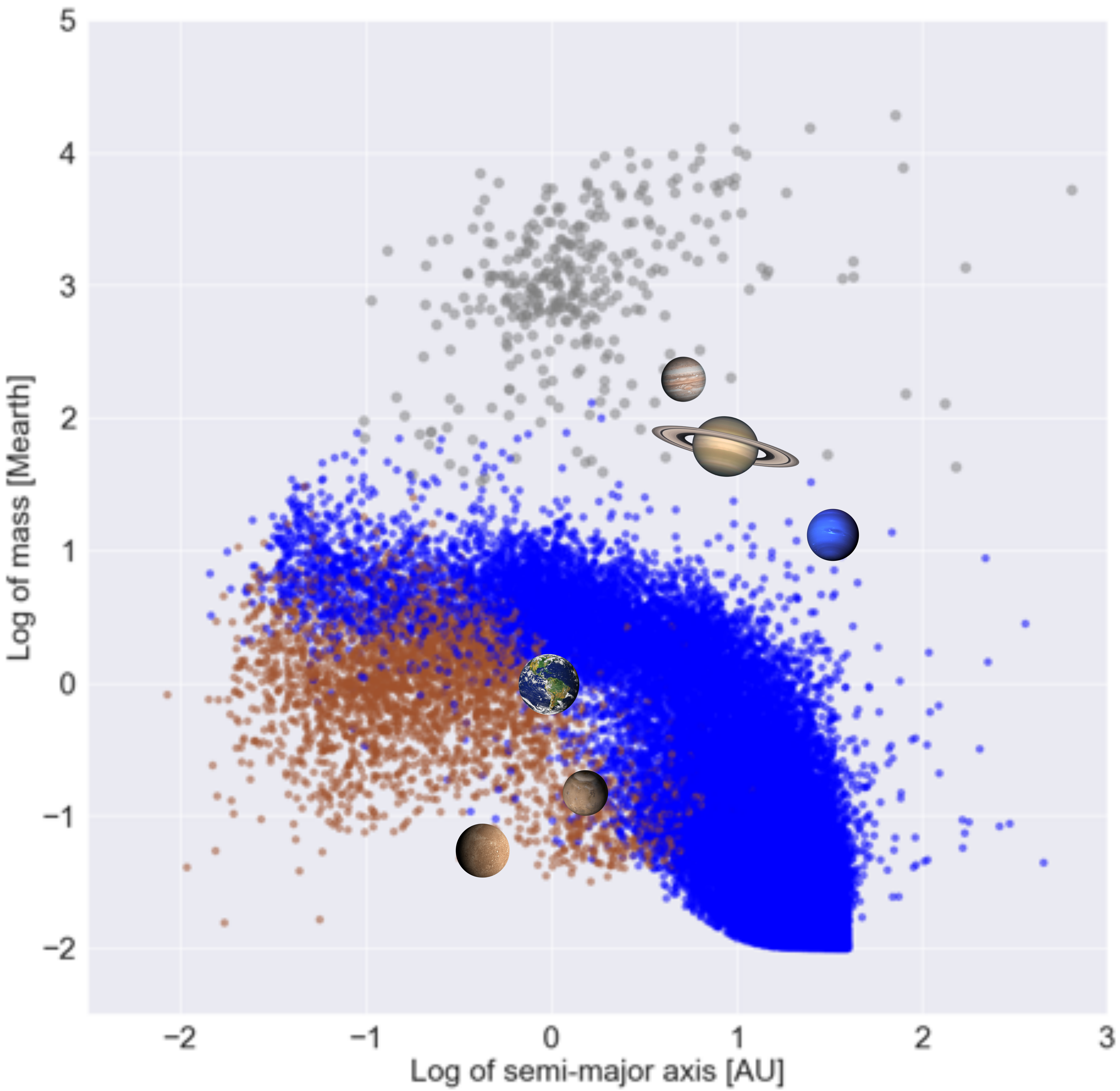


J. Haldemann, PhD Thesis

Adapted from Emsenhuber et al., 2021

The Bern model - planetary population

UNIVERSITÄT BERN
CSH CENTER FOR SPACE AND HABITABILITY



Gas giant

More than 50 wt% in gas

Rocky planet

Less than 50 wt% in gas

Less than 1 wt% in volatiles

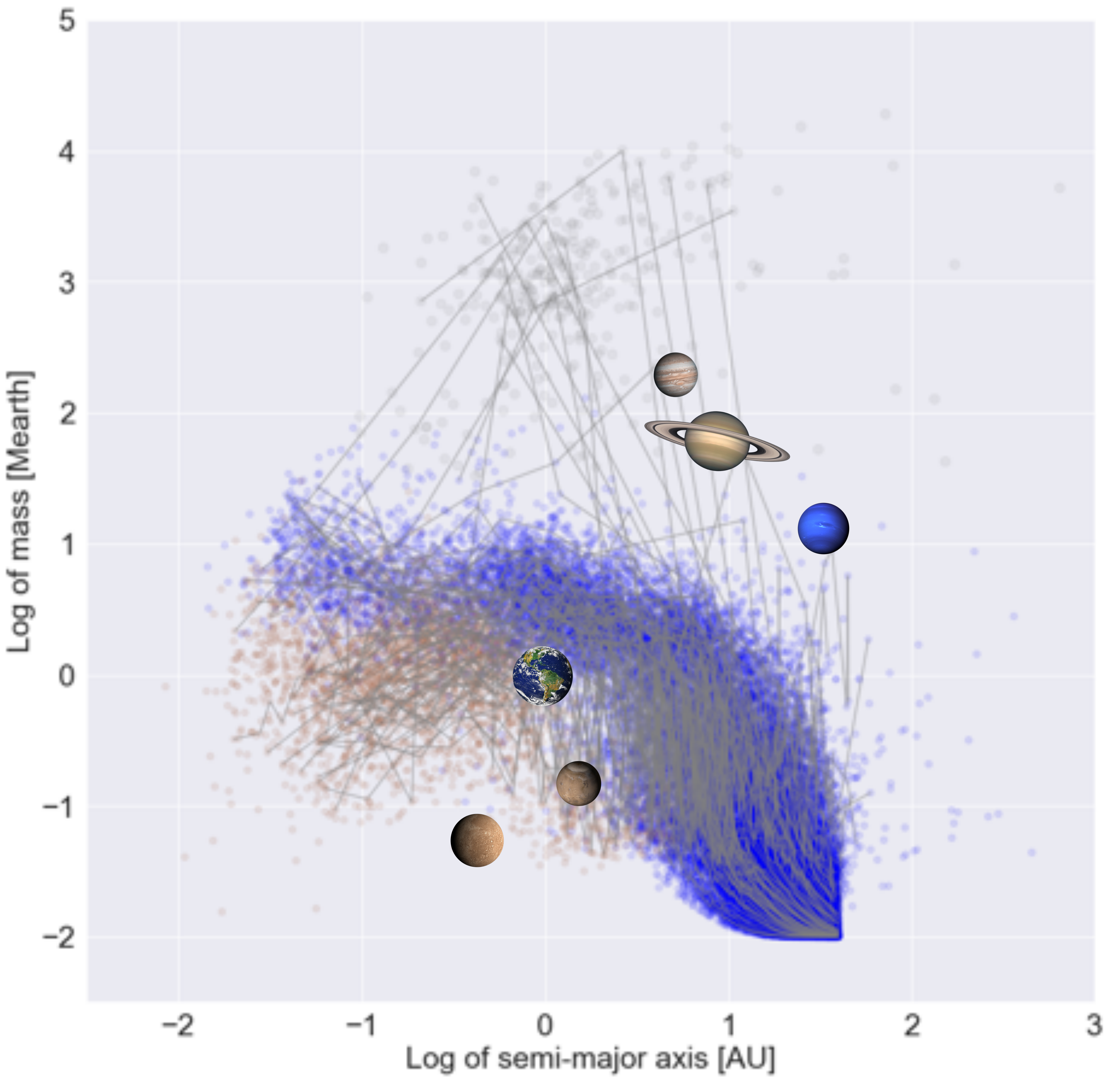
Icy planet

Less than 50 wt% in gas

More than 1 wt% in volatiles

Adapted from Emsenhuber et al., 2021

The Bern model - planetary systems



Gas giant

More than 50 wt% in gas

Rocky planet

Less than 50 wt% in gas

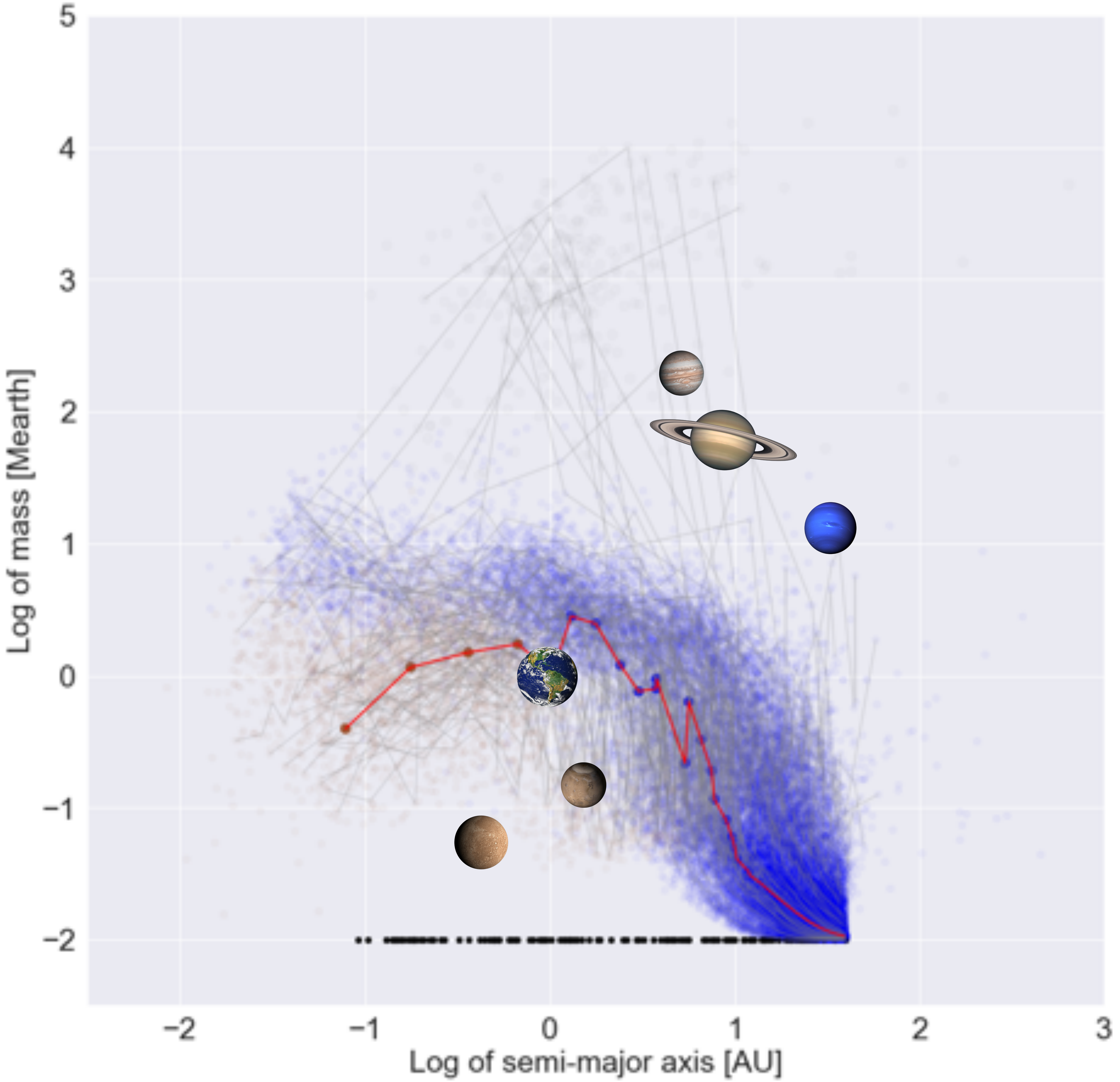
Less than 1 wt% in volatiles

Icy planet

Less than 50 wt% in gas

More than 1 wt% in volatiles

The Bern model - planetary systems



Gas giant

More than 50 wt% in gas

Rocky planet

Less than 50 wt% in gas

Less than 1 wt% in volatiles

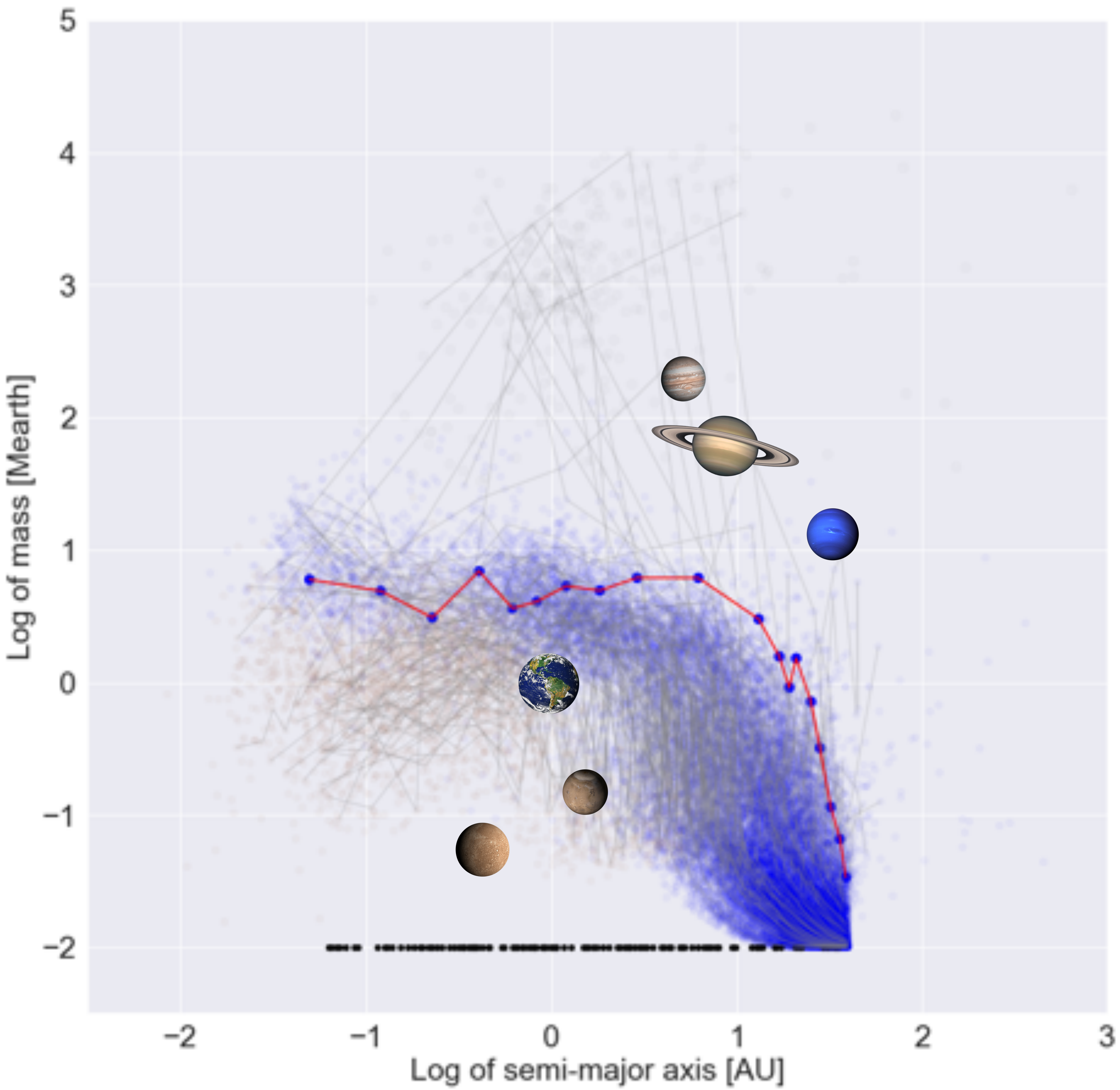
Icy planet

Less than 50 wt% in gas

More than 1 wt% in volatiles

The Bern model - planetary systems

UNIVERSITÄT BERN
CSH CENTER FOR SPACE AND HABITABILITY



Gas giant

More than 50 wt% in gas

Rocky planet

Less than 50 wt% in gas

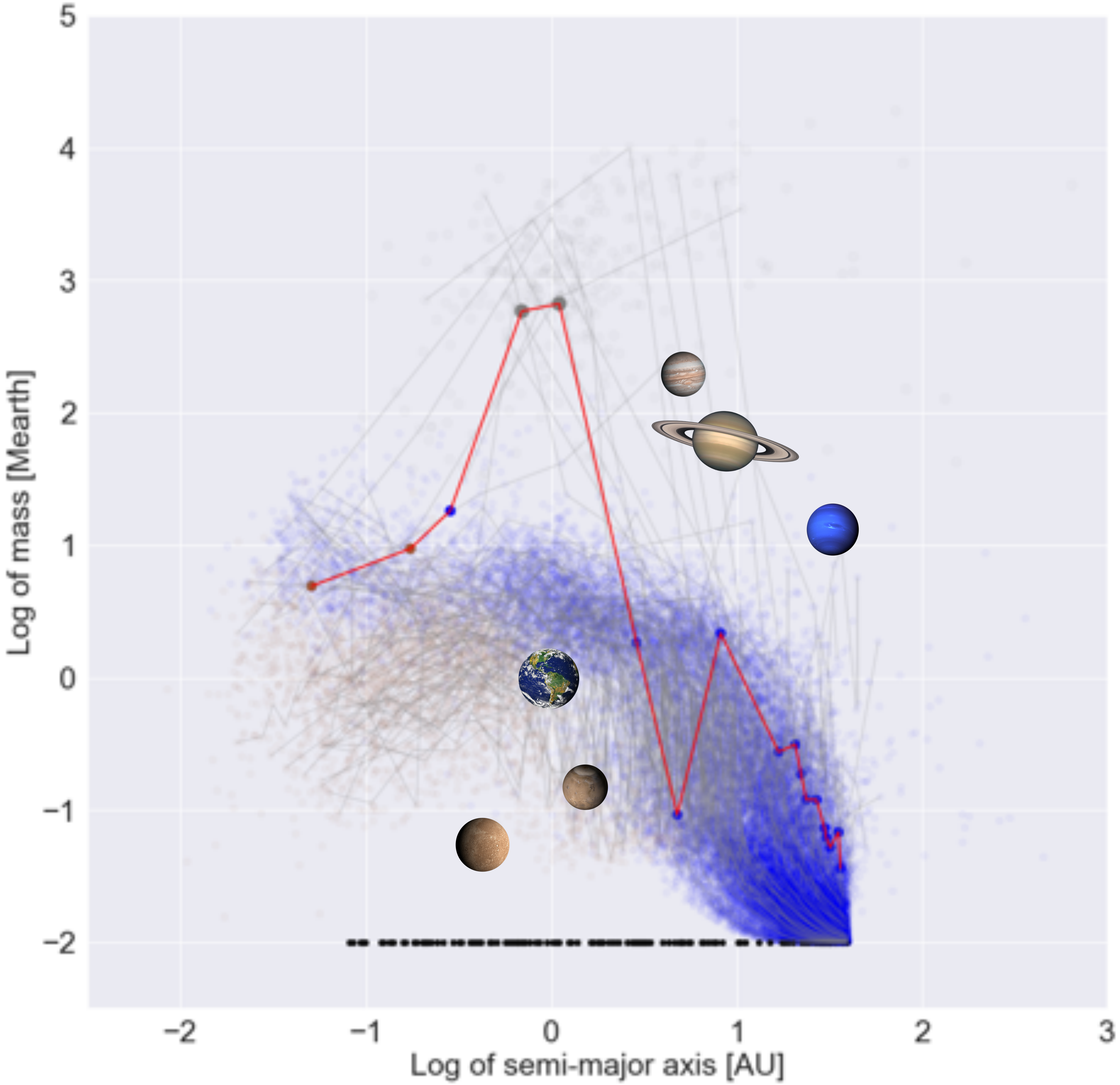
Less than 1 wt% in volatiles

Icy planet

Less than 50 wt% in gas

More than 1 wt% in volatiles

The Bern model - planetary systems



Gas giant

More than 50 wt% in gas

Rocky planet

Less than 50 wt% in gas

Less than 1 wt% in volatiles

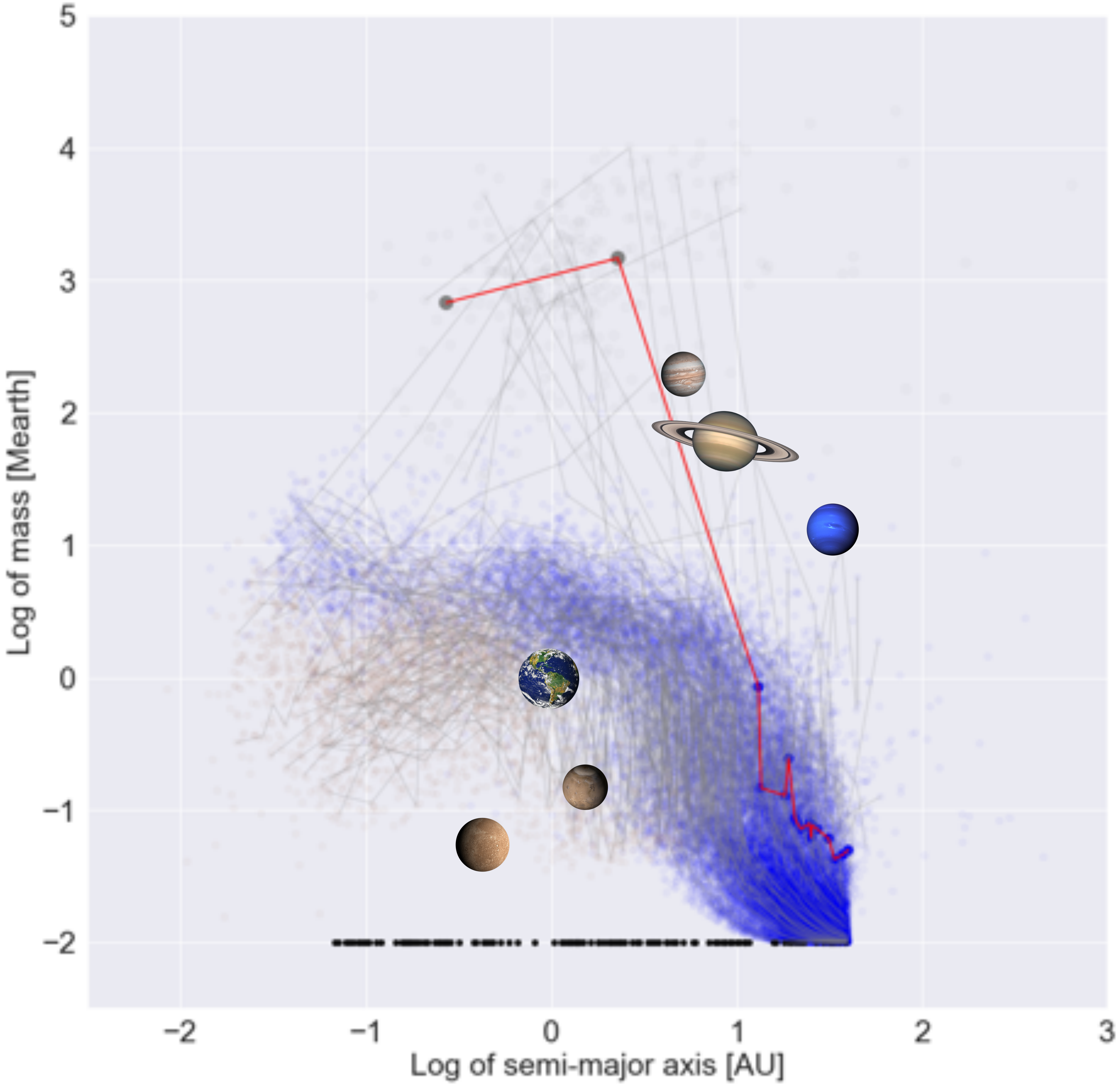
Icy planet

Less than 50 wt% in gas

More than 1 wt% in volatiles

The Bern model - planetary systems

UNIVERSITÄT BERN
CSH CENTER FOR SPACE AND HABITABILITY



Gas giant

More than 50 wt% in gas

Rocky planet

Less than 50 wt% in gas

Less than 1 wt% in volatiles

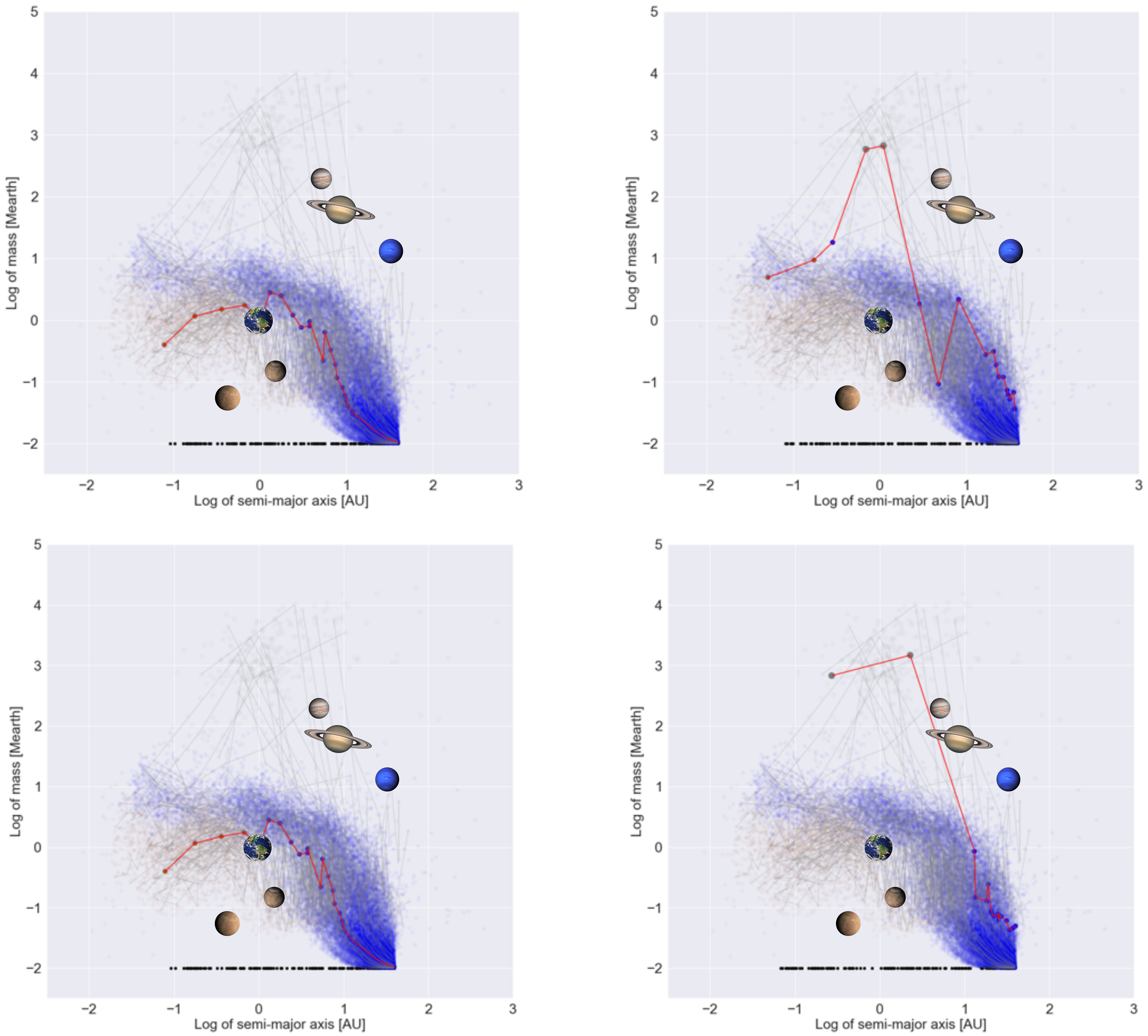
Icy planet

Less than 50 wt% in gas

More than 1 wt% in volatiles

The Bern model - planetary systems

UNIVERSITÄT BERN
CSH CENTER FOR SPACE AND HABITABILITY



Which of the four systems looks more similar to the solar system?

1- Planet formation and the Bern model

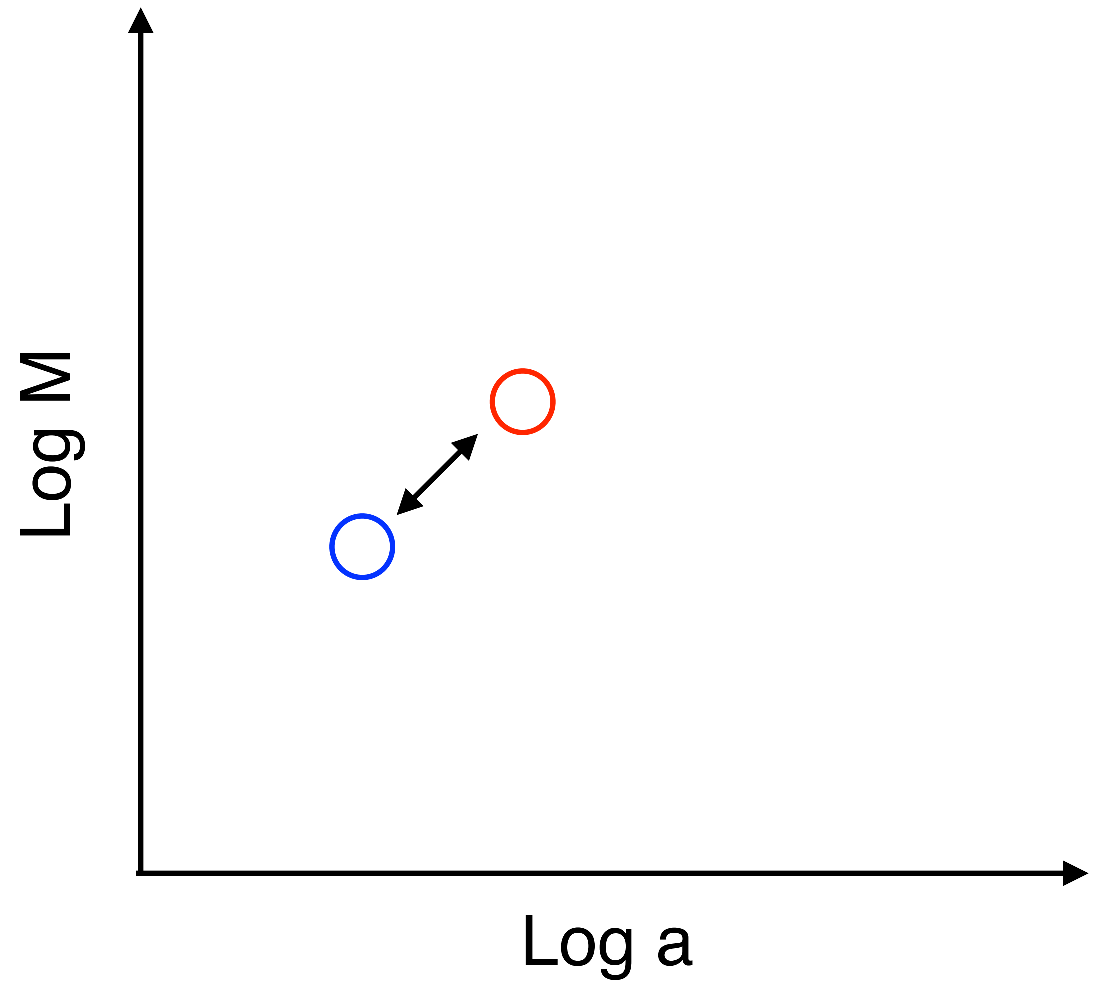
2- Models *versus* observations

3- Planetary internal structure and Deep Learning

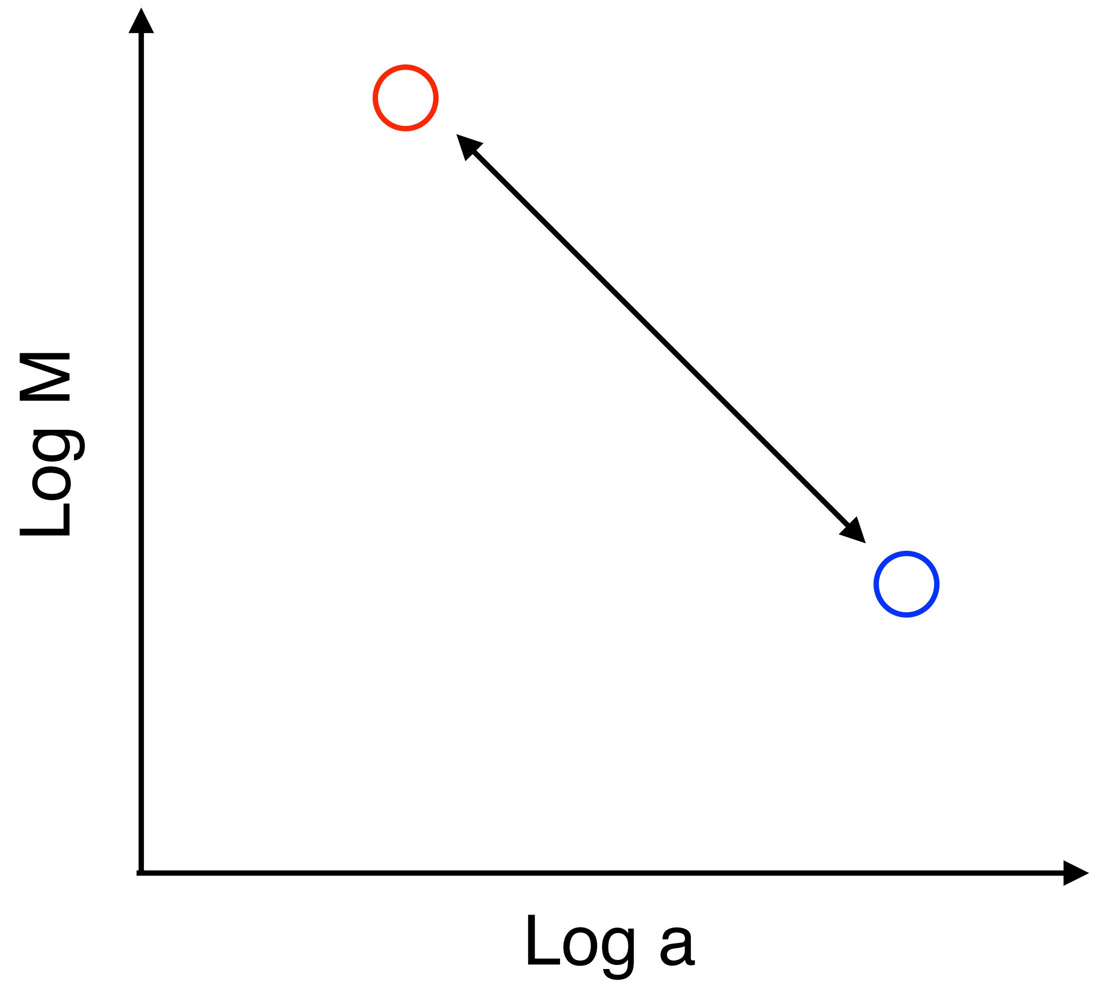
4- Correlation in planetary systems with random forrest: finding a second Earth

Distance between 2 systems - 1-planet case

In the case N=1, the easiest concept of distance is the Euclidian distance in the m-a space (we will concentrate on the case where only m and a are relevant to define a planet - an over simplistic view - but generalisation is easy).



Very similar

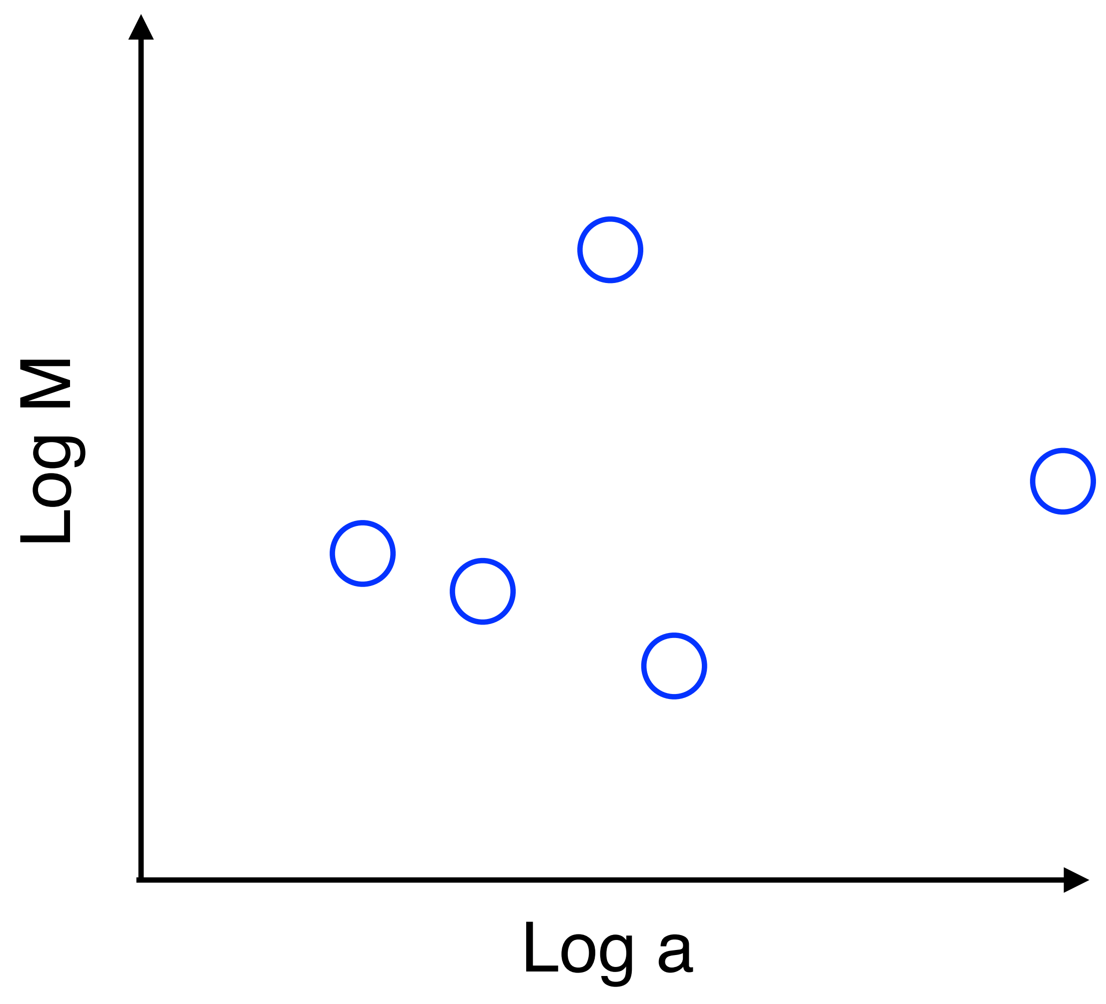


Very dissimilar

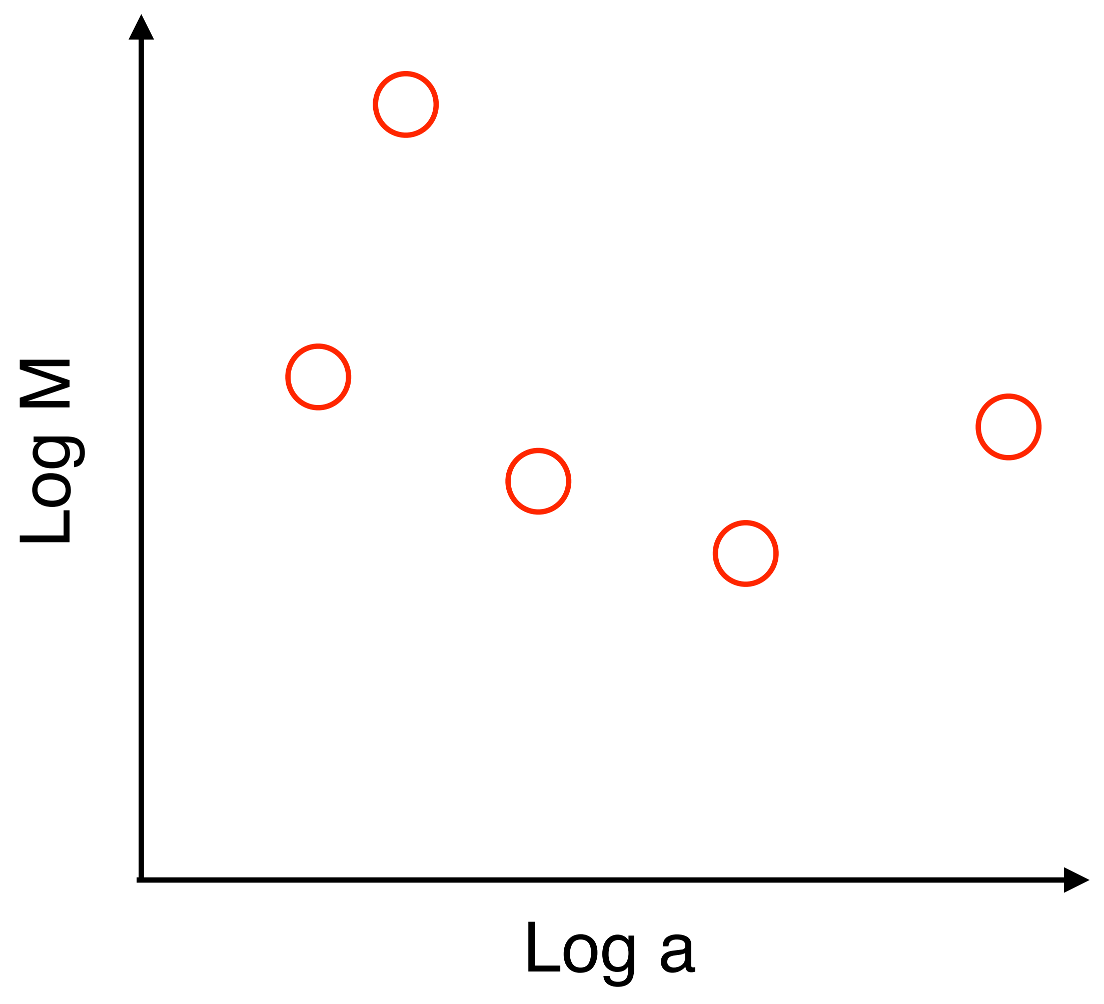
As this distance is an Euclidian distance, it ***is*** a distance mathematically speaking!

Distance between 2 systems - N-planet case

In the case of two N-planet systems, with the same N for both, we would like to define the distance as ‘the sum of distances between one planet of one side and another planet of the other side’.



System 1



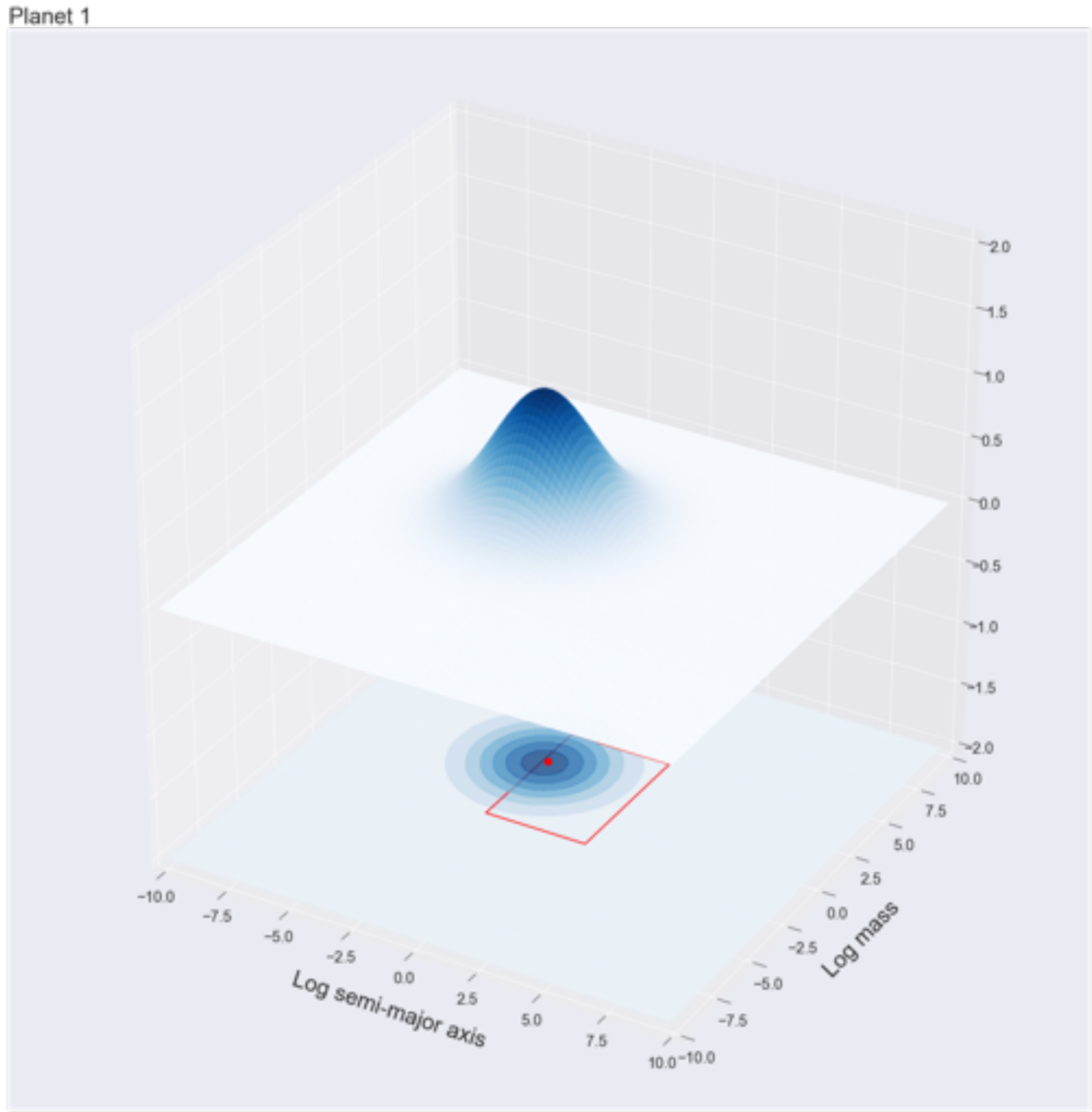
System 2

-> many different possible pairs possible -> how can we choose?

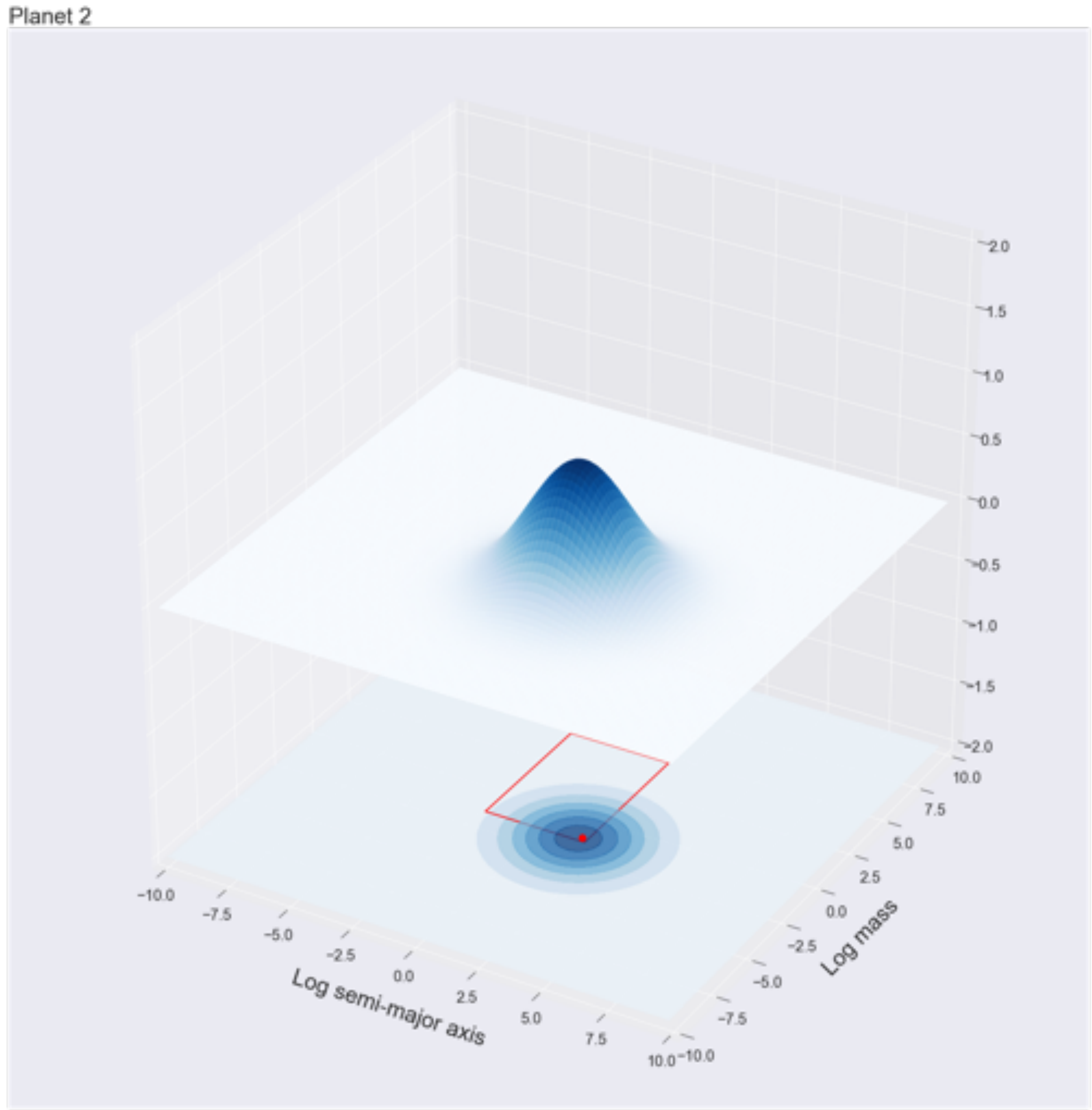
Distance between 2 systems - N-planet case

We first 'spread' planets in the Log(a) - Log(M) space.

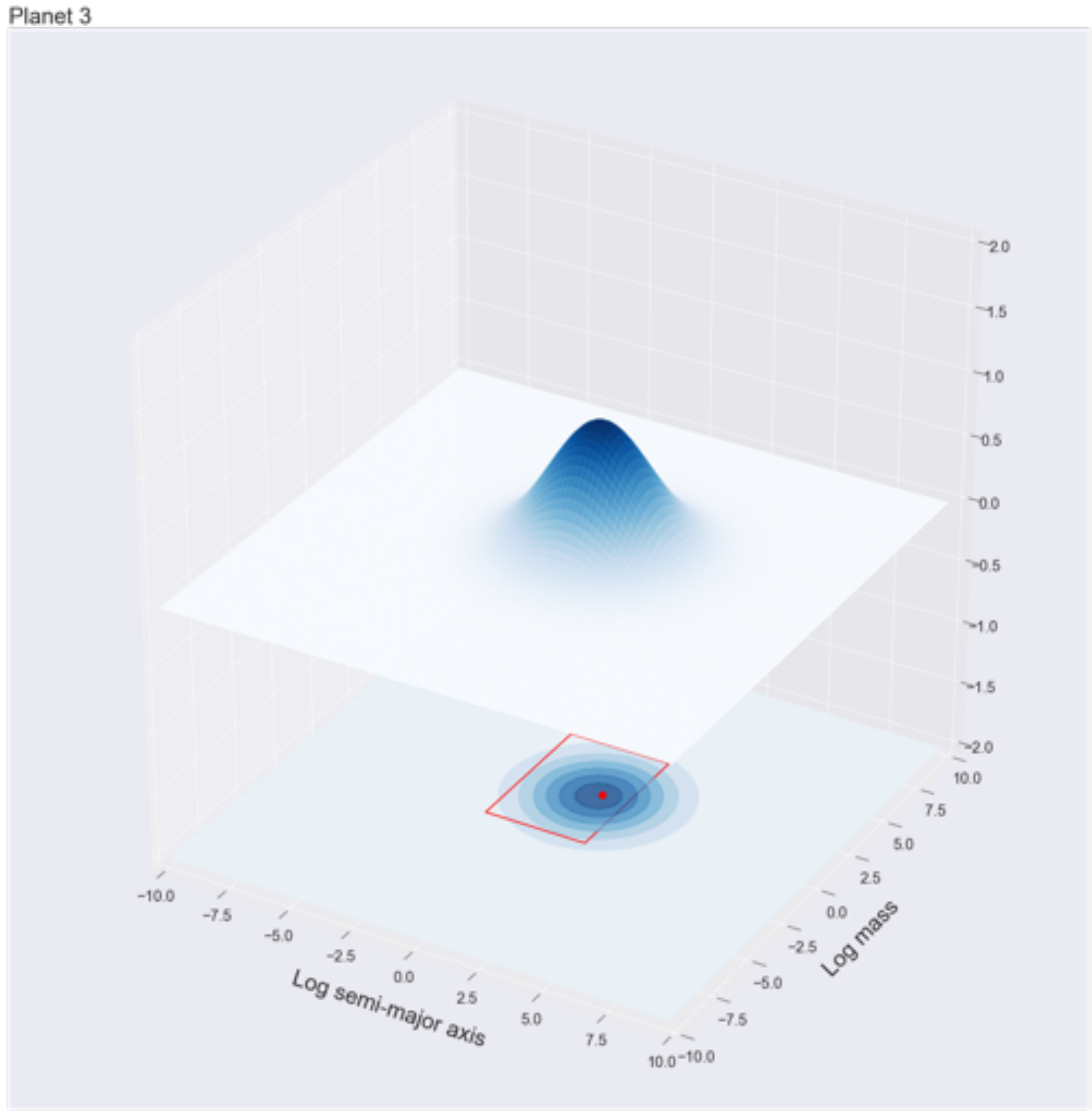
$$f_p(M,a) = \exp\left(-\left(\frac{\log M - \log M_p}{2\sigma_m}\right)^2 - \left(\frac{\log a - \log a_p}{2\sigma_a}\right)^2\right)$$



+

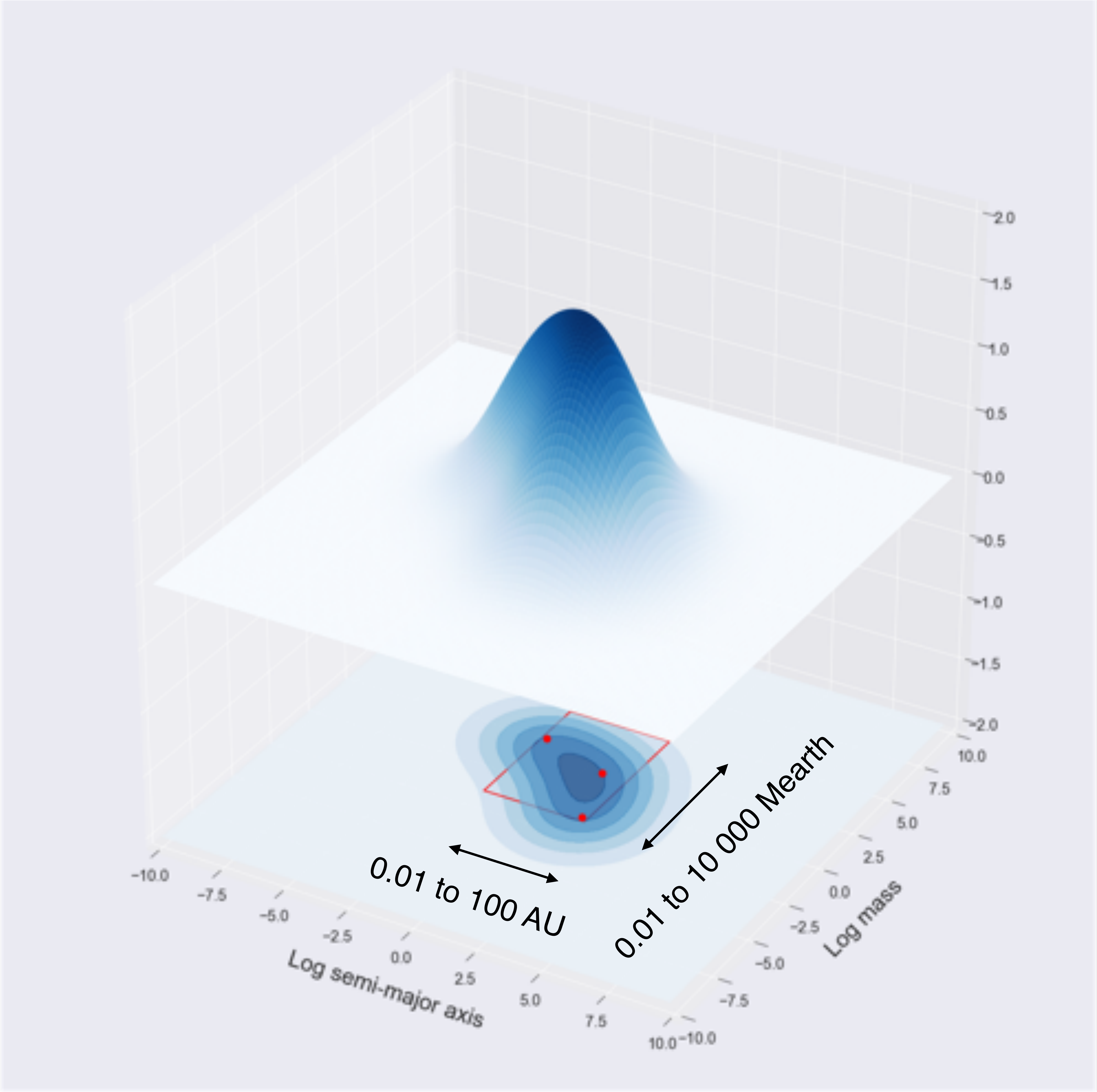


+



=

Distance between 2 systems - N-planet case



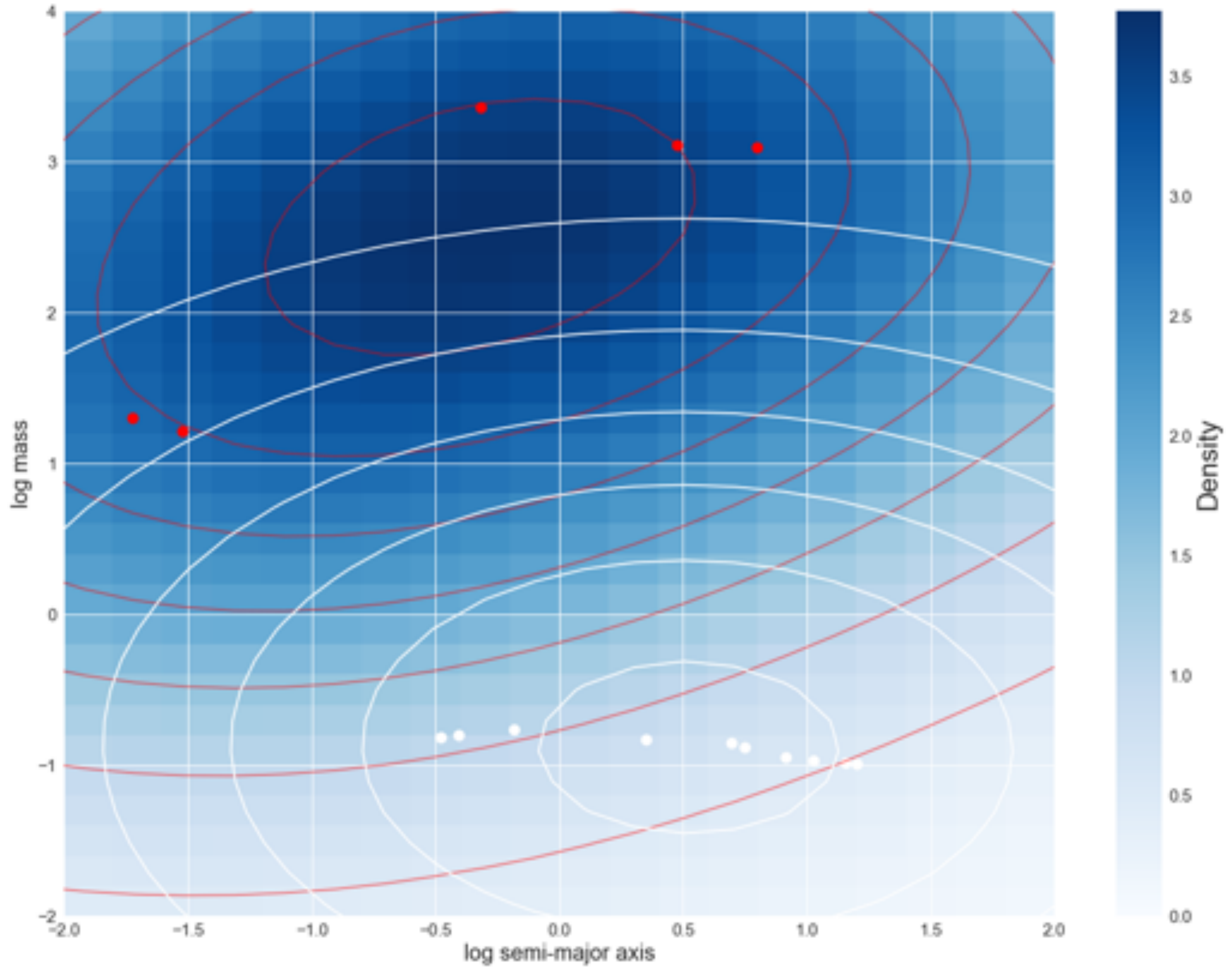
$$\psi_s(M,a) = \sum_{p \in s} f_p(M,a)$$

UNIVERSITÄT BERN
CSH CENTER FOR SPACE AND HABITABILITY

Distance between 2 systems - N-planet case

The square of the distance between two planetary systems is then

$$d(s_1, s_2) = \sqrt{\int \int (\psi_{s_1} - \psi_{s_2})^2 \, d\log M \, d\log a}$$



from 20-D to 2-D using T-SNE

UNIVERSITÄT BERN
CSH CENTER FOR SPACE AND HABITABILITY

We want to project our 200-D data points on 2D, while keeping the maximum of information. For this we minimise the KL divergence:

$$D(p||q) = \sum_{i,j \in S} p_{i,j} \log \left(\frac{p_{i,j}}{q_{i,j}} \right)$$

p: similarity in *system space (200D)*

q: similarity in *projection space (2D)*

-> measures the gain in information we get by using the *p* distribution (the real high-D one) instead of the *q* distribution (the low-D one).

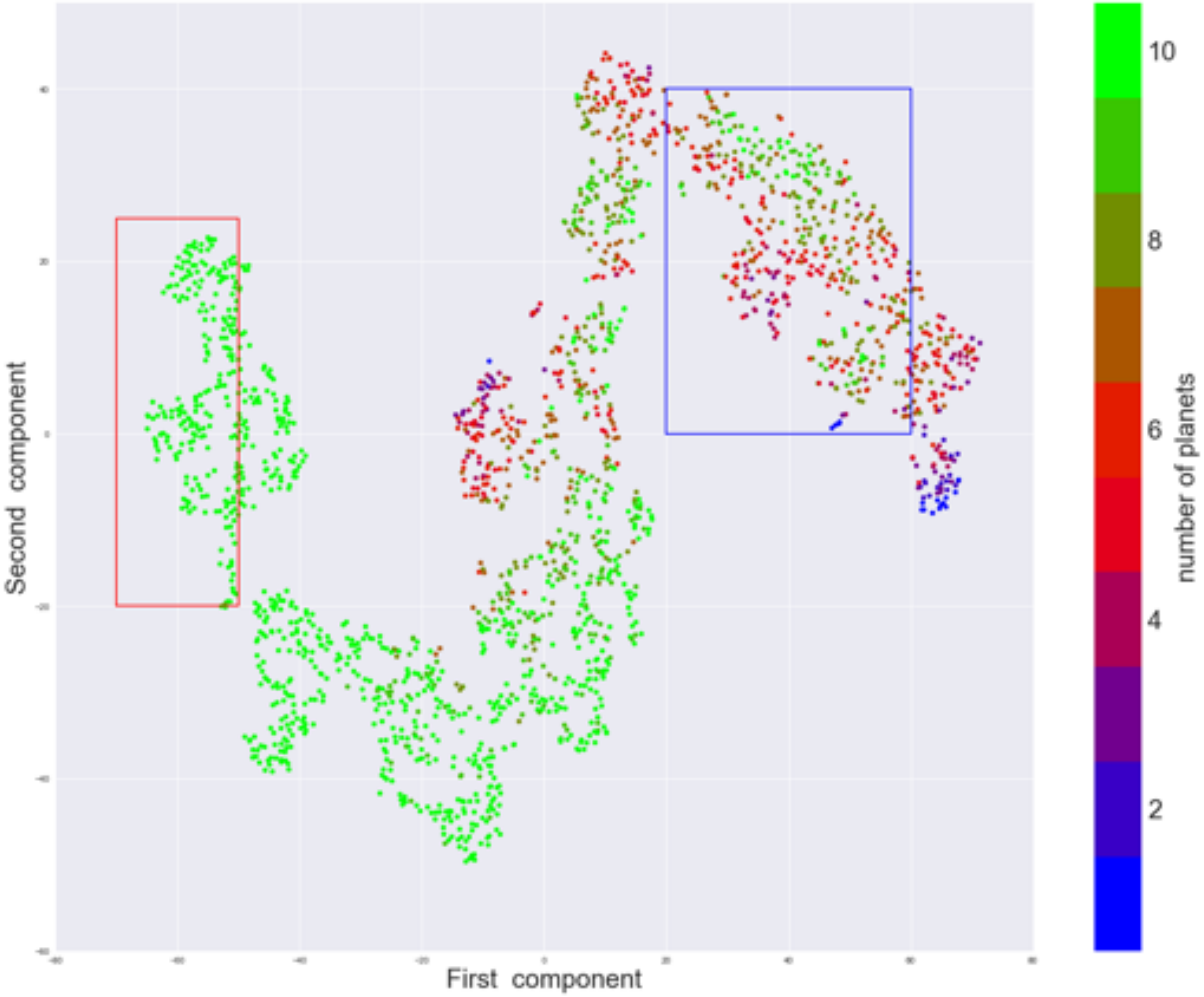
If p is large, q *needs to be large.* If p si small, q *can be whatever*

-> the local structure is preserved (close-in points), not the global one

If q is small, p is small: systems fare away in the 2D space are dis-similar in the 200-D space.

!! this requires the distance to be a distance !!

T-SNE - reference population

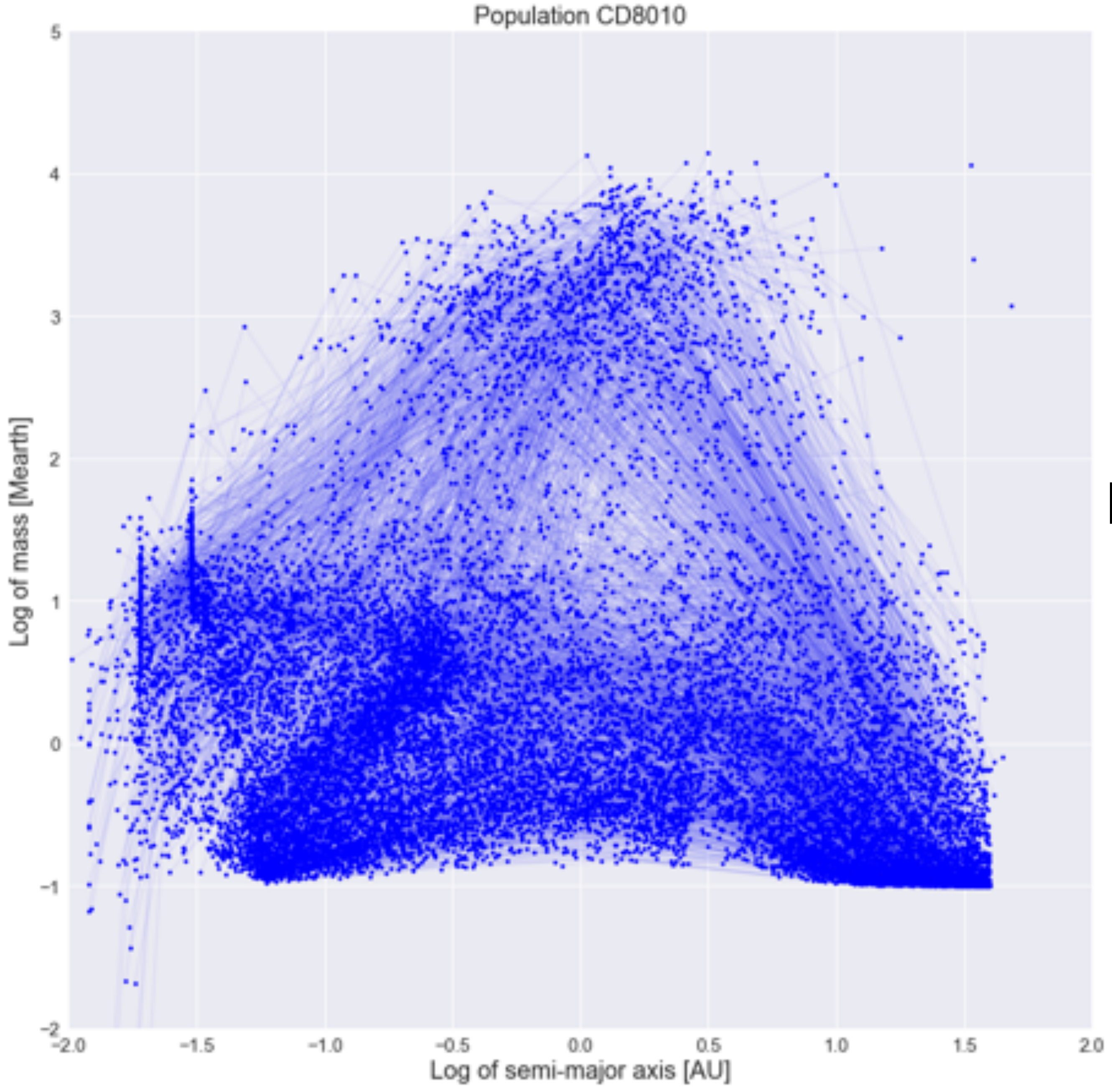


Alibert 2019

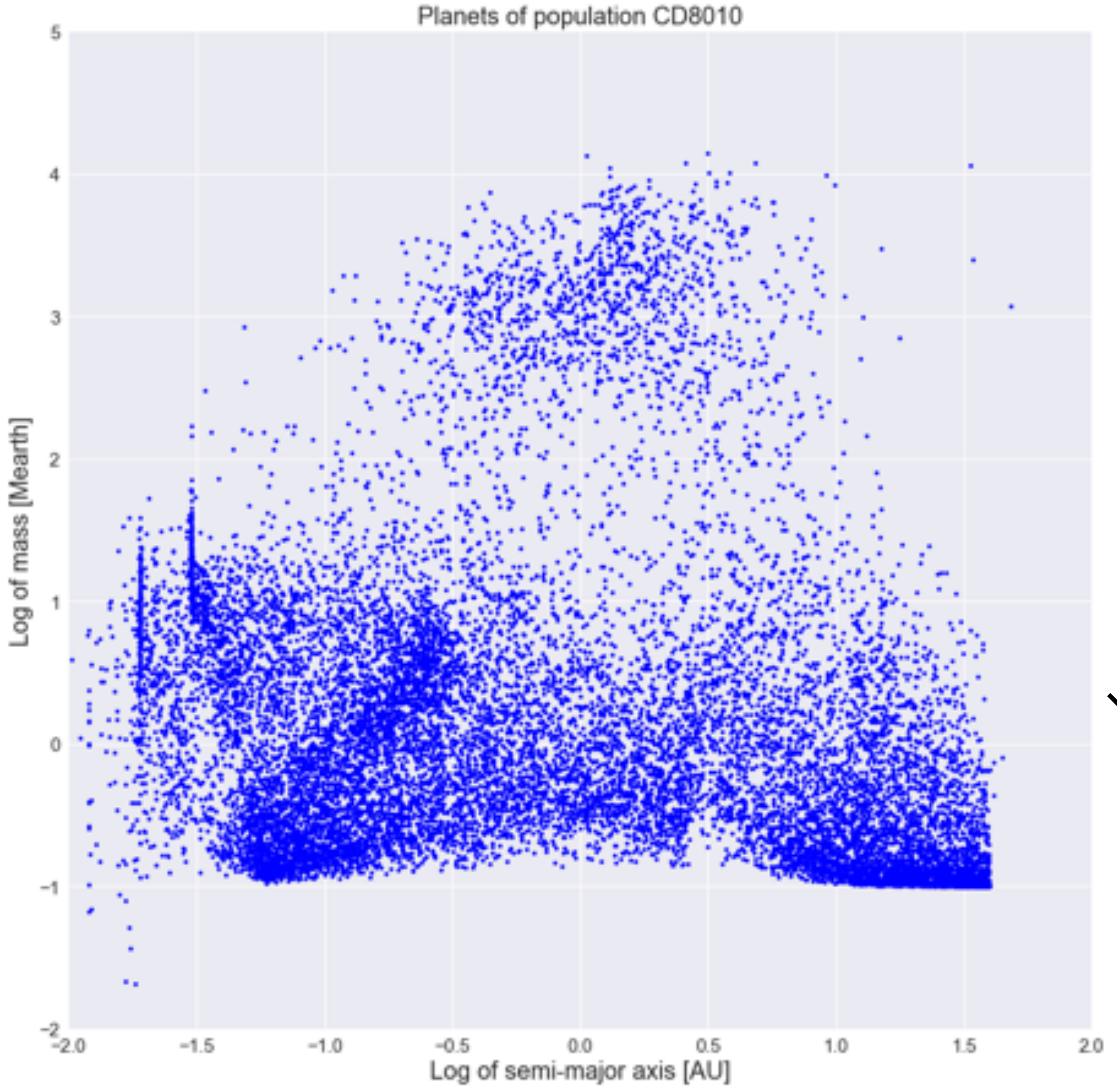
comparing with a population that has same planets but different architecture

UNIVERSITÄT BERN

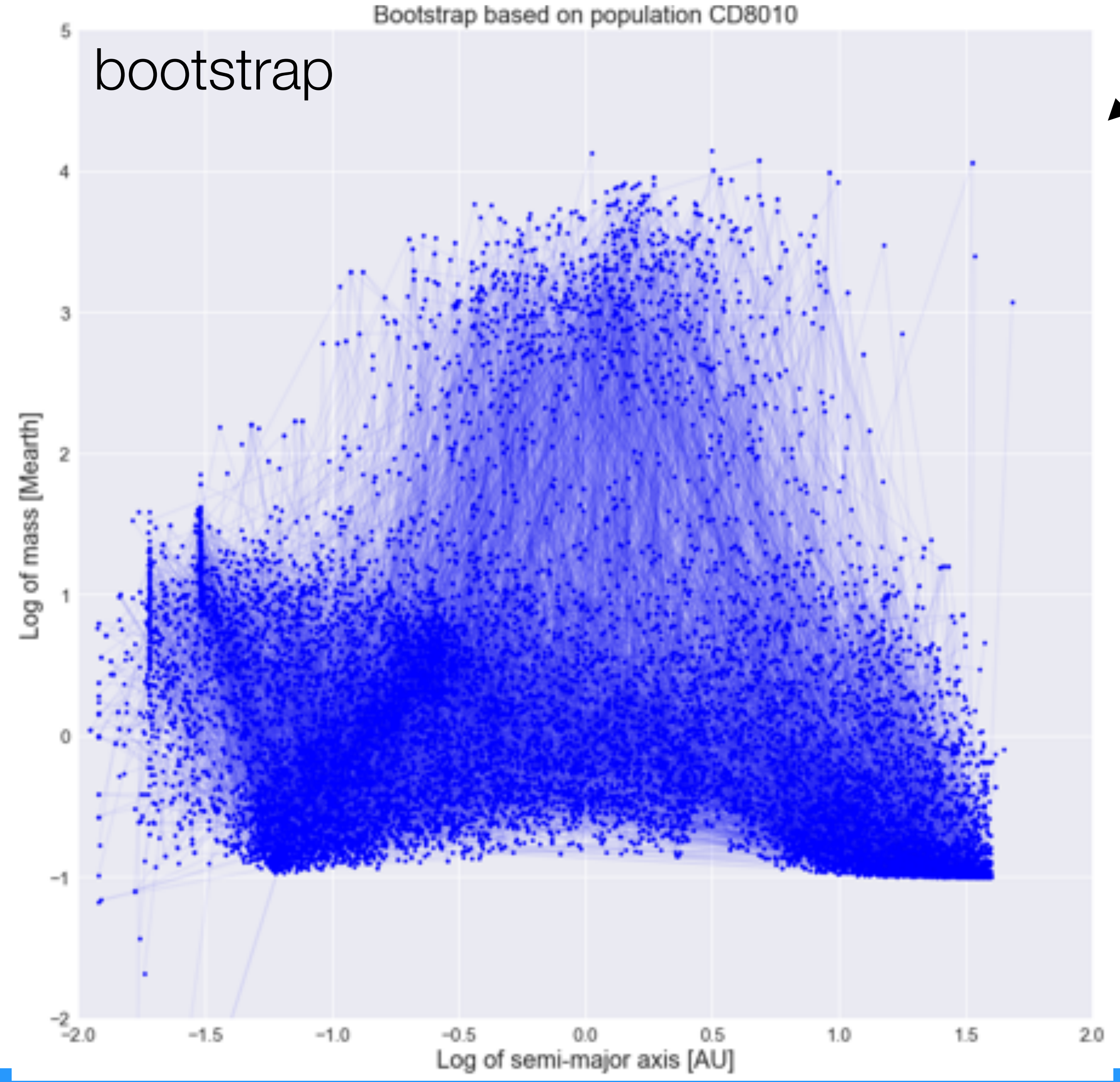
CSH CENTER FOR SPACE AND HABITABILITY



break the systems



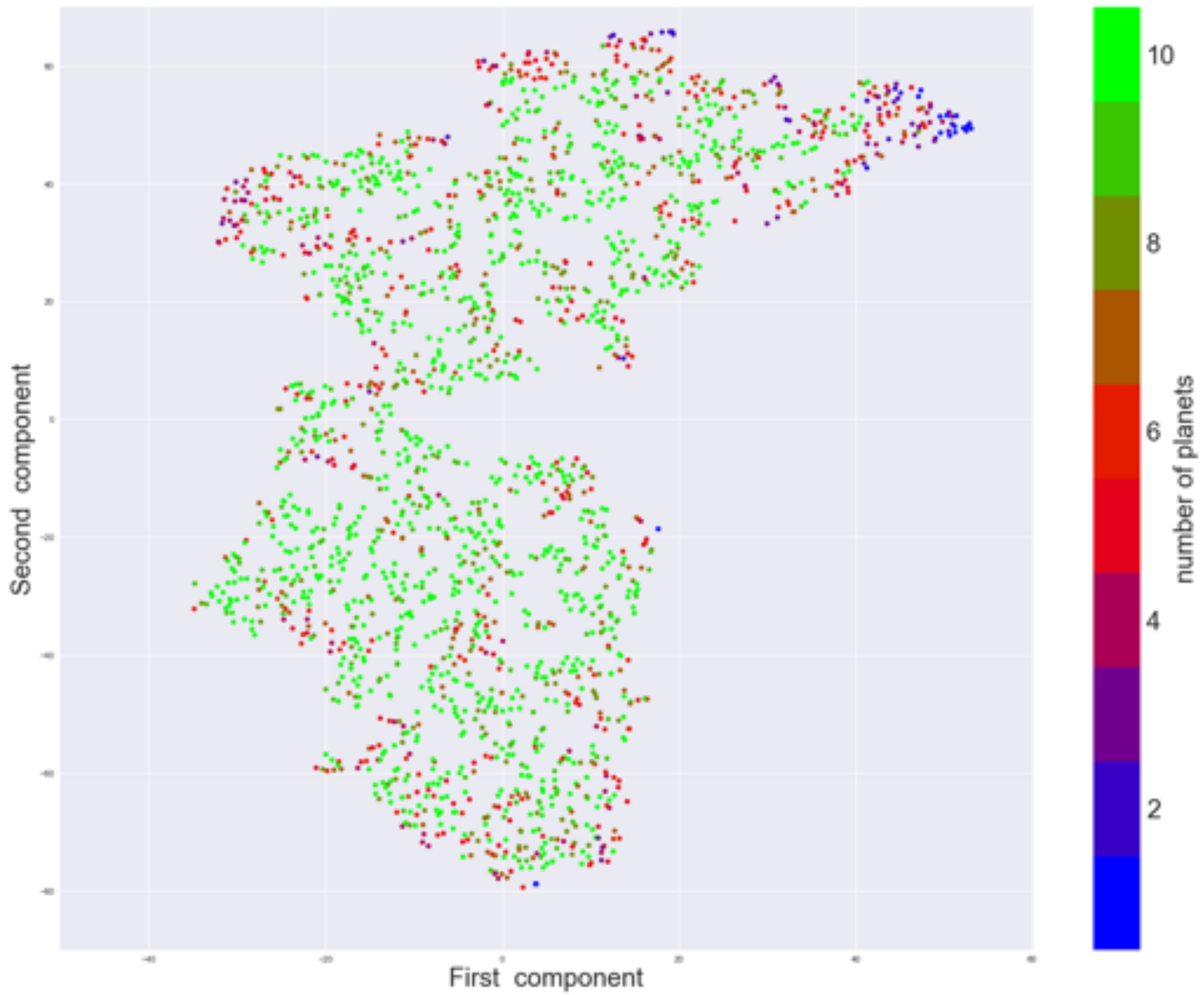
reconstruct new systems



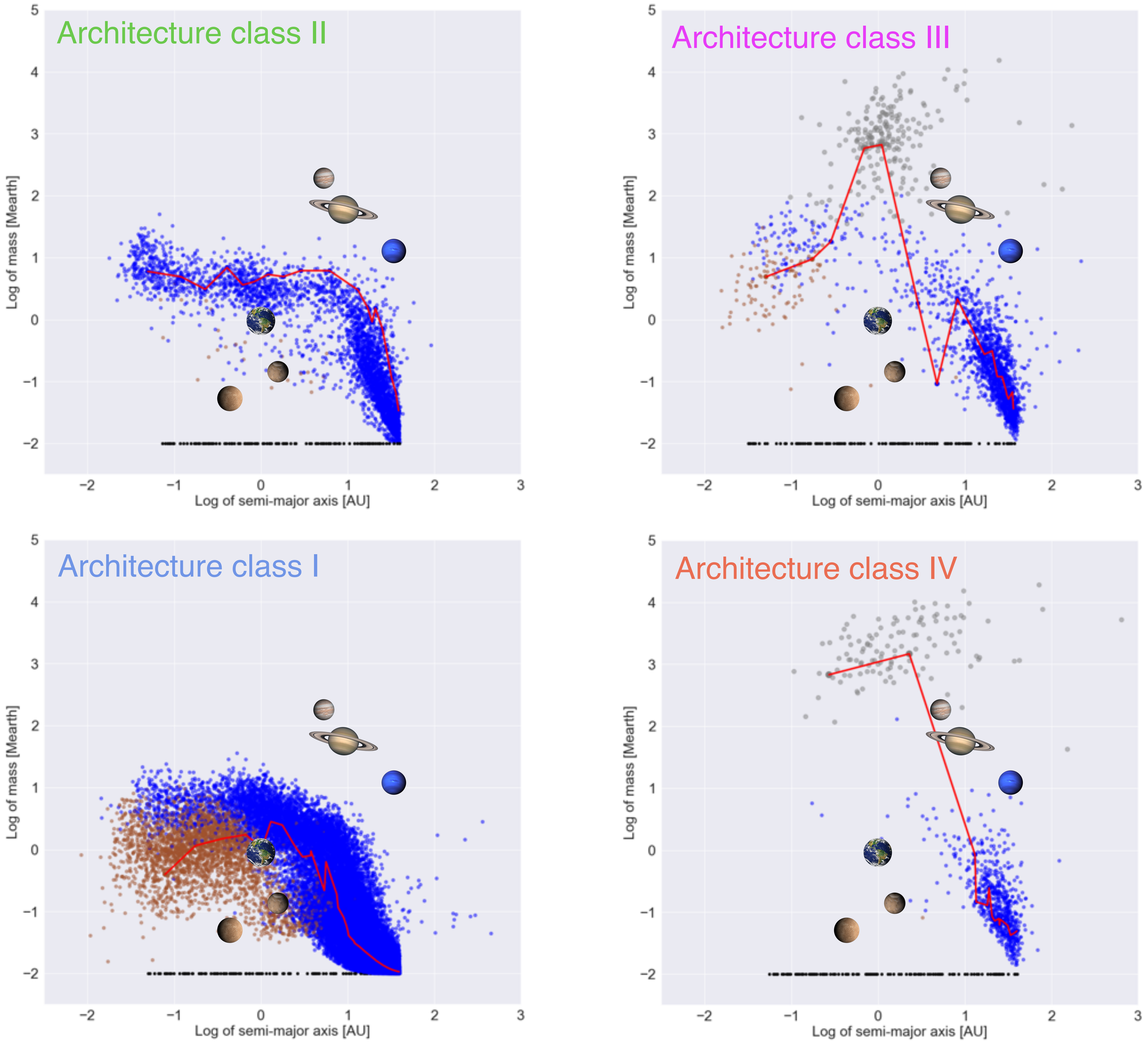
distribution of final number of planets in each system is kept unchanged

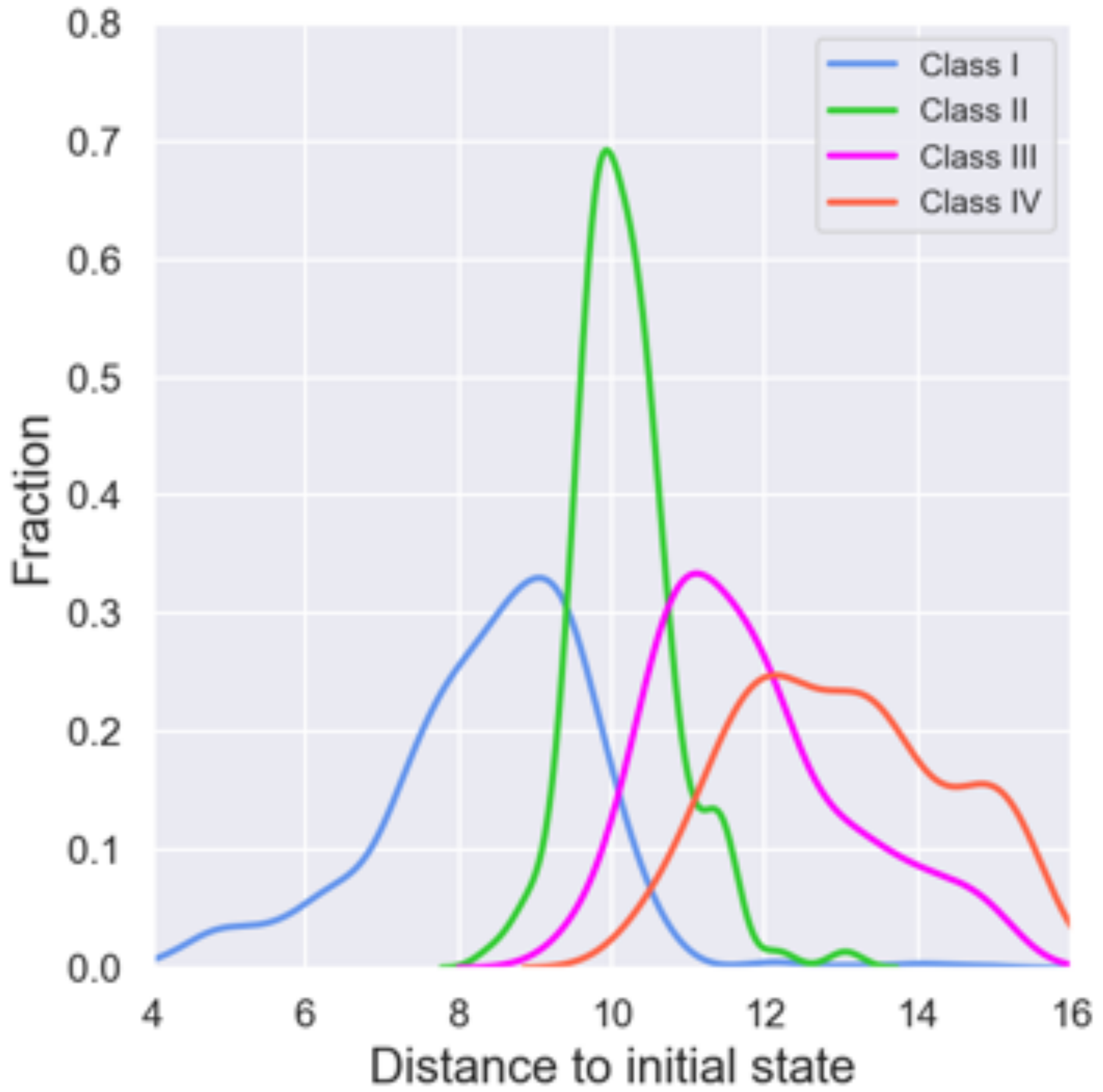
T-SNE - non physical population

Physical systems are intrinsically difference from un-physical systems built with the *same* planets -> simulations lead to different *architectures.*



Four classes of architecture





Alibert, 2019

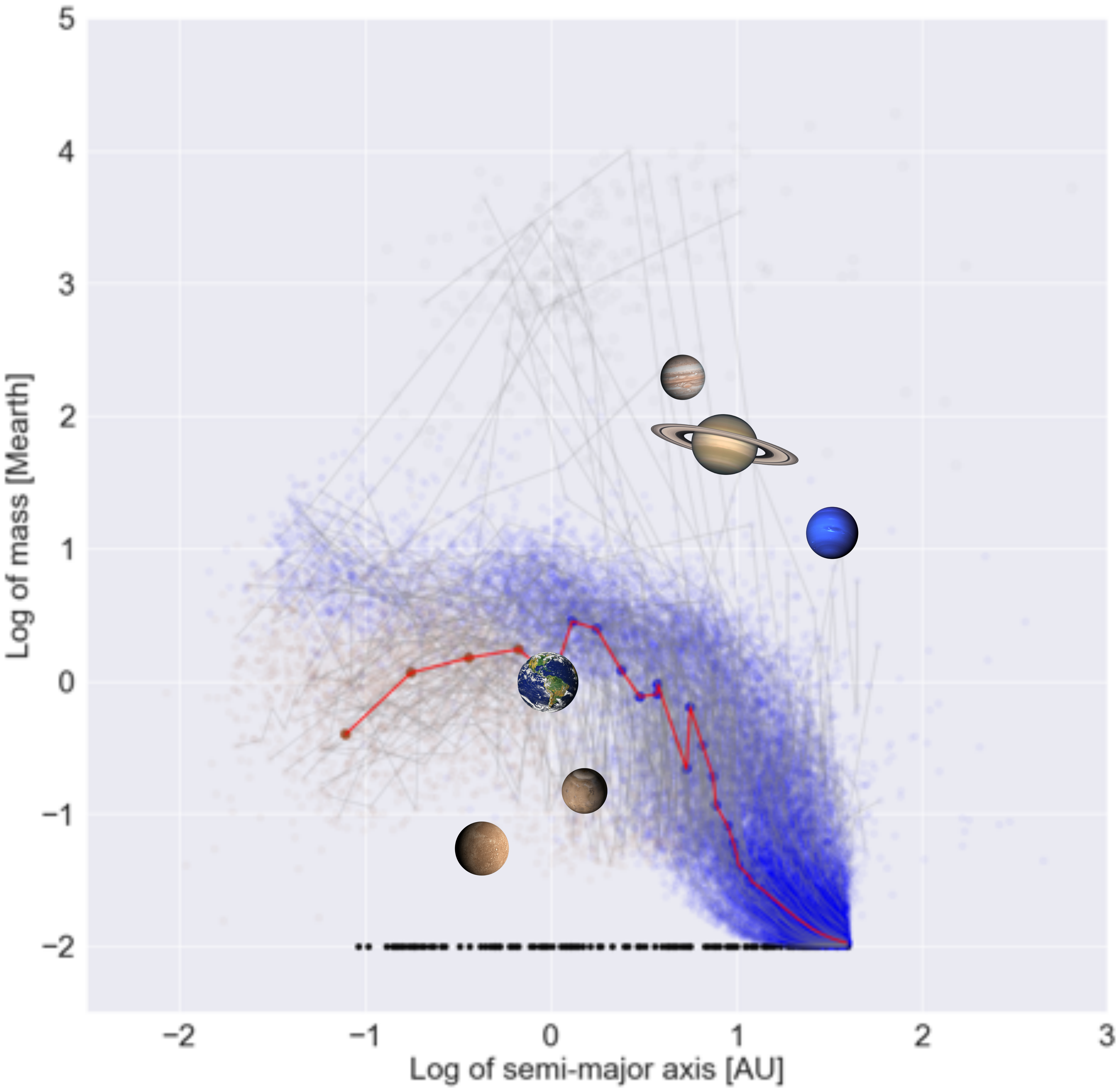
1- Planet formation and the Bern model

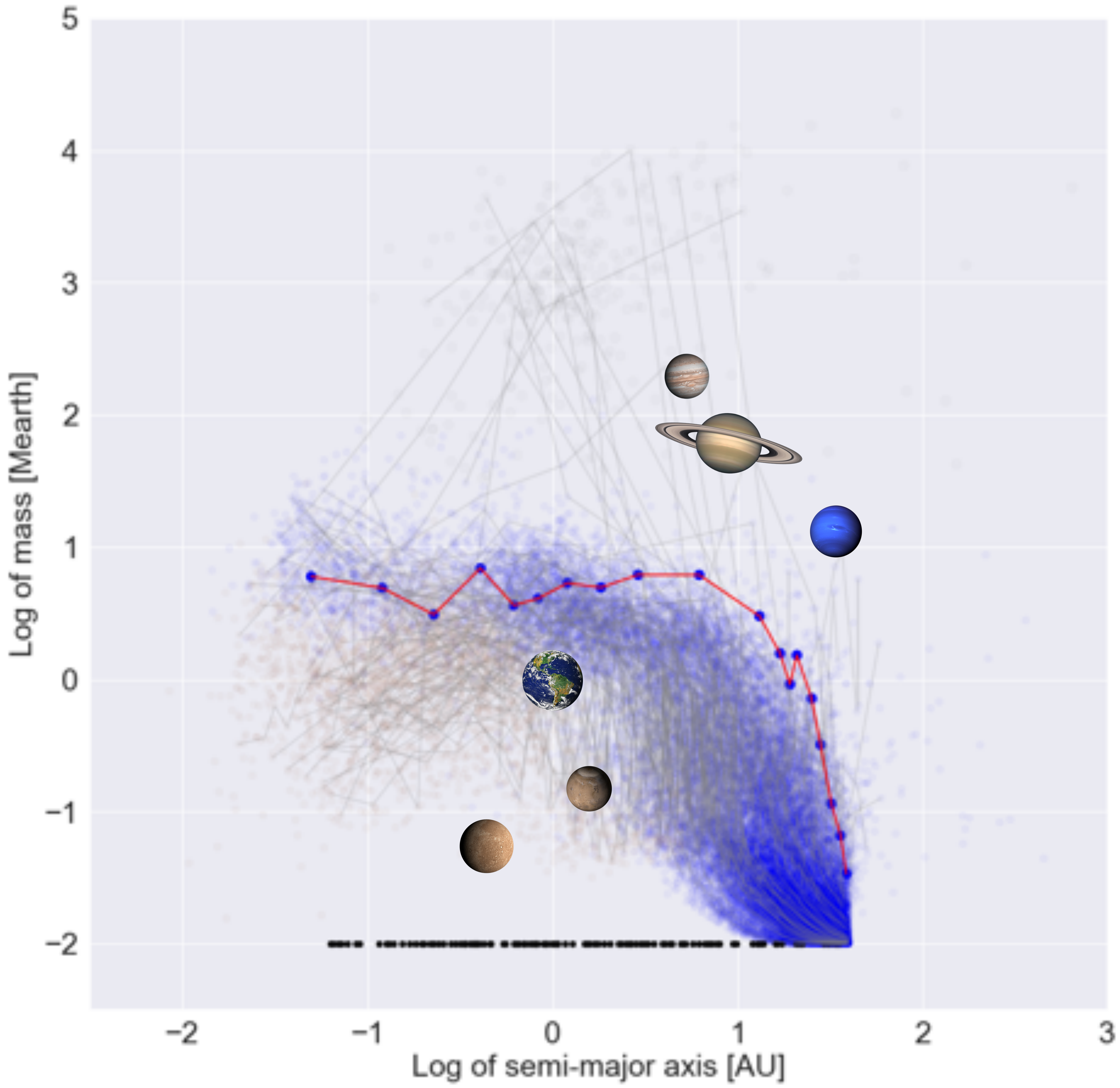
2- Models *versus* observations

3- Planetary internal structure and Deep Learning

4- Correlation in planetary systems with random forrest: finding a second Earth

Architecture *versus* composition

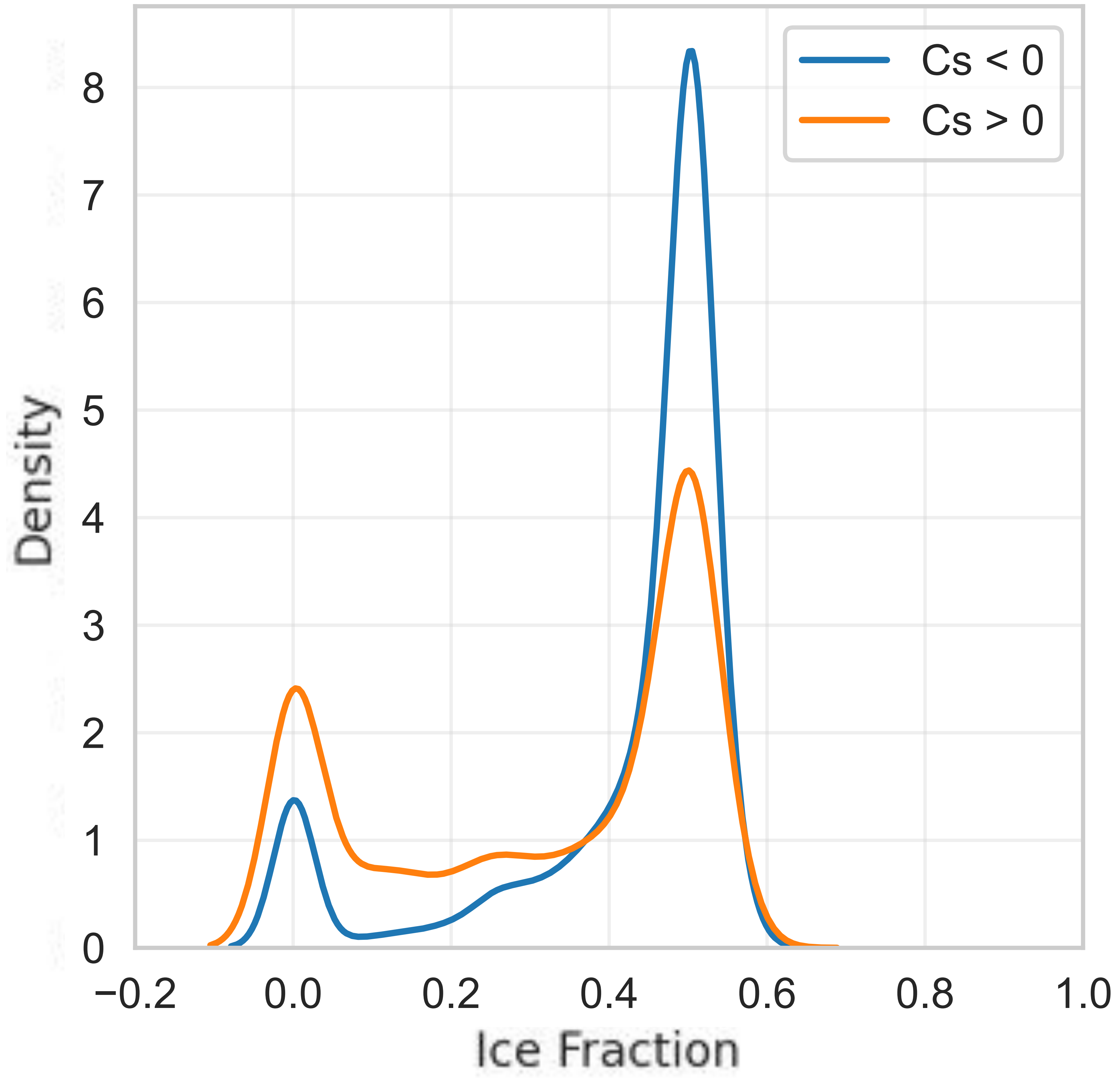




Architecture *versus* composition

UNIVERSITÄT BERN

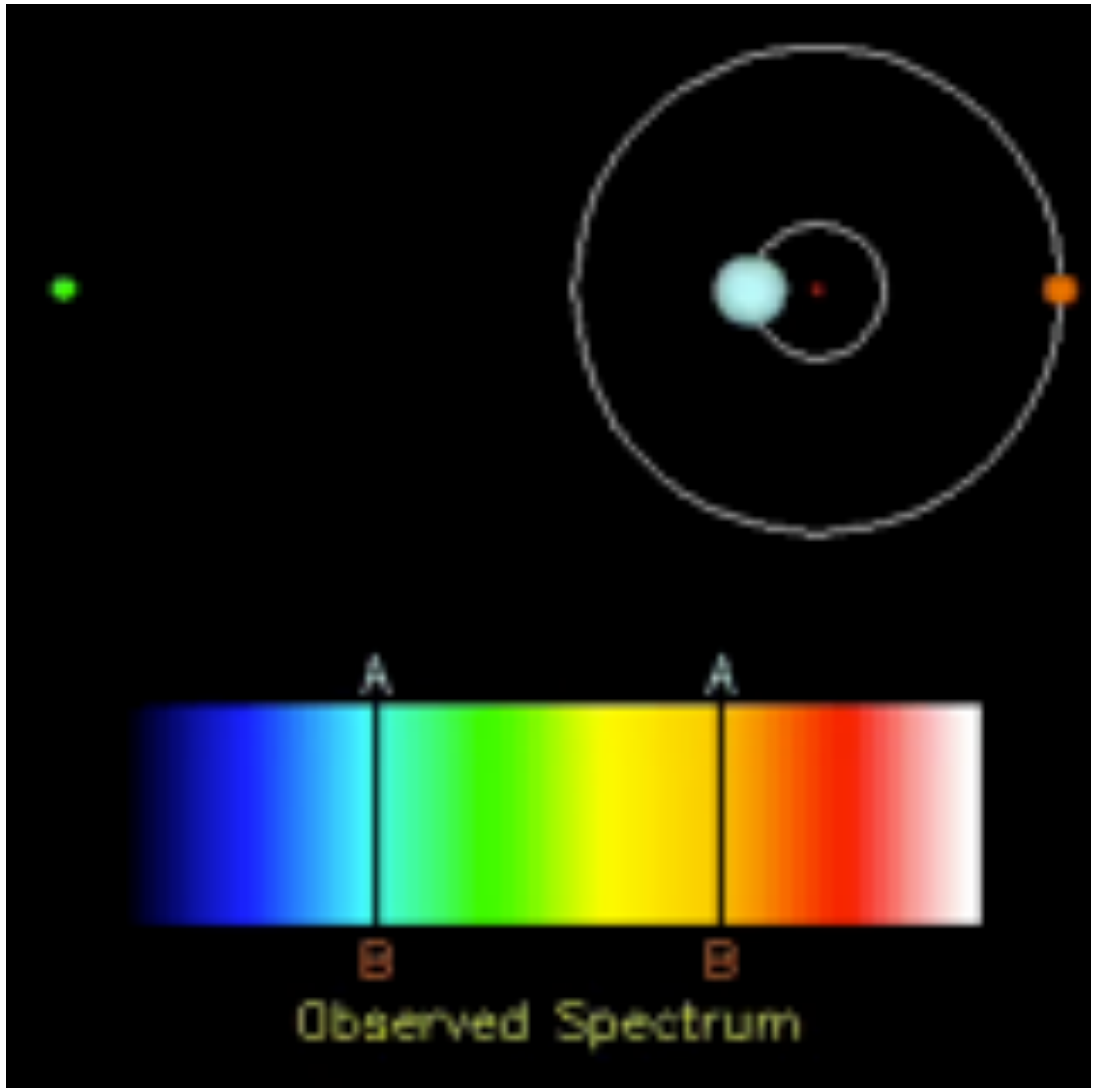
CSH CENTER FOR SPACE AND HABITABILITY

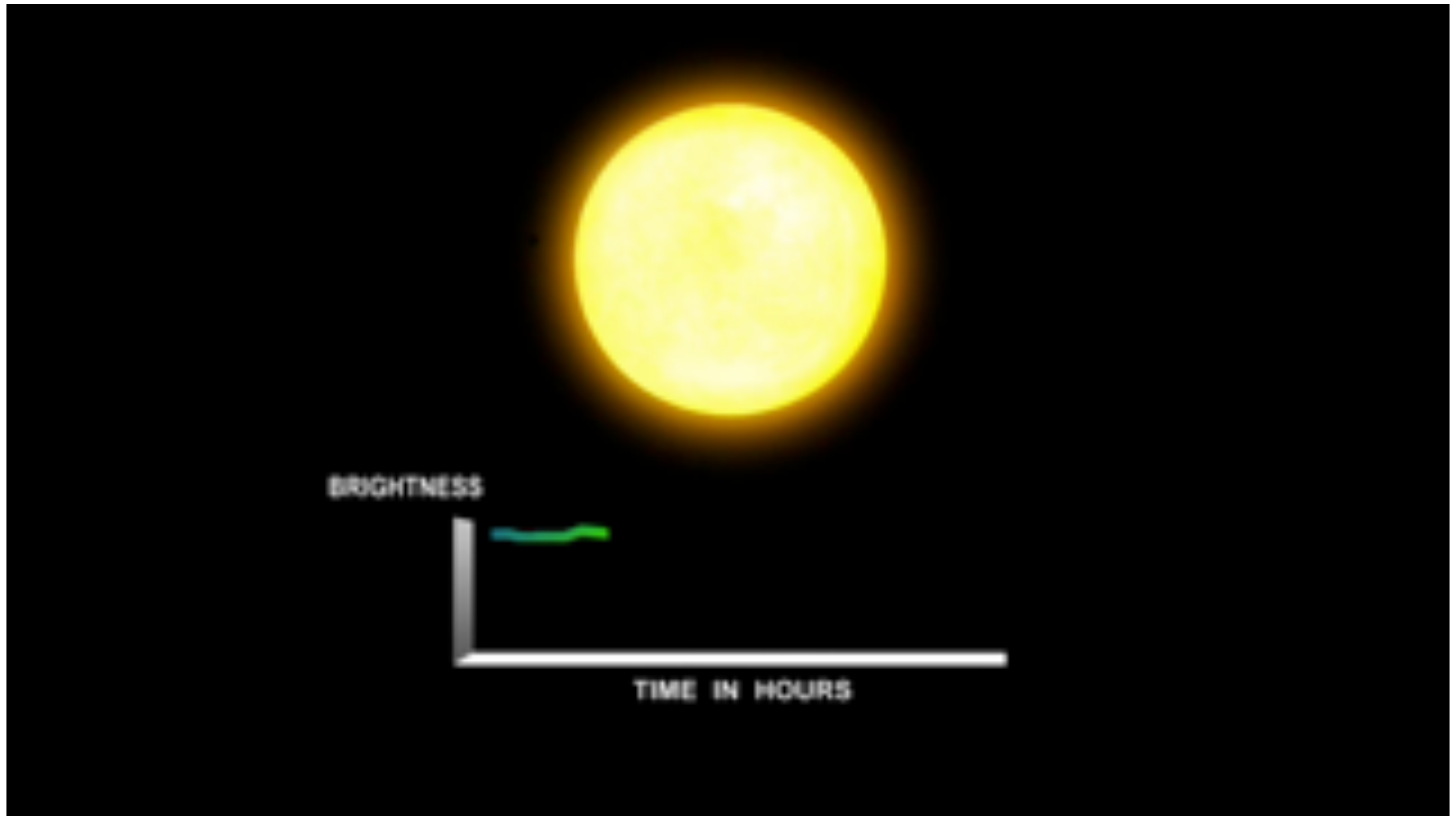


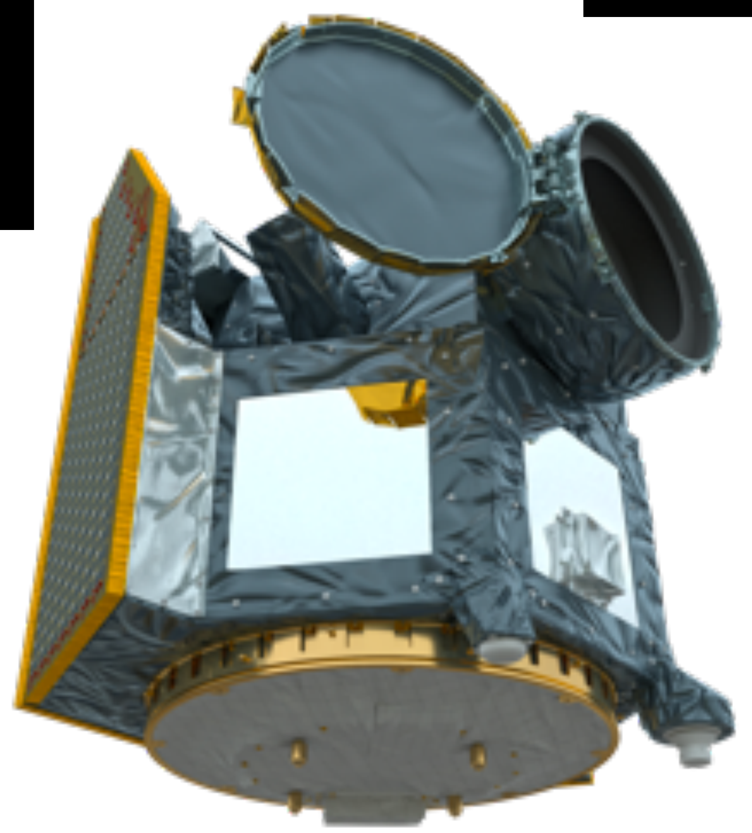
how can we observationally determine the ice fraction?

Observing planets

UNIVERSITÄT BERN
CSH CENTER FOR SPACE AND HABITABILITY



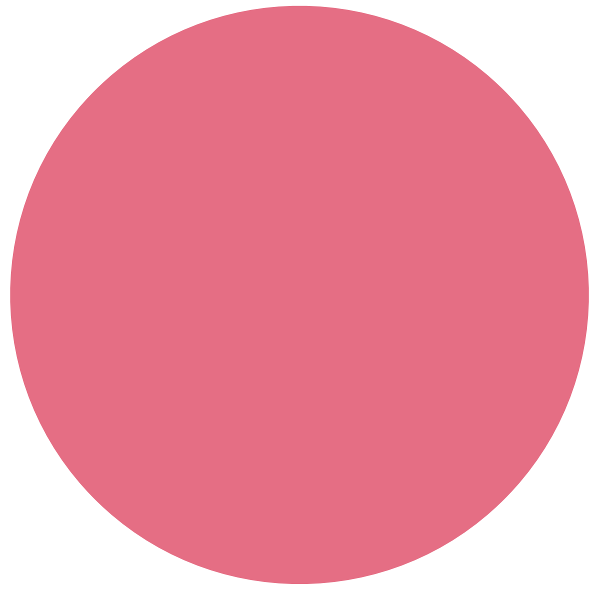




M_p/M_{star}

$(R_p/R_{star})^2$

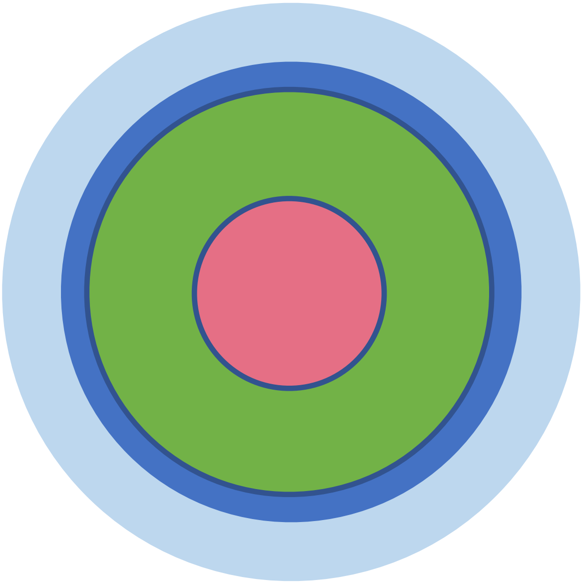
Internal structure and composition



Observations

- Mass
- Radius

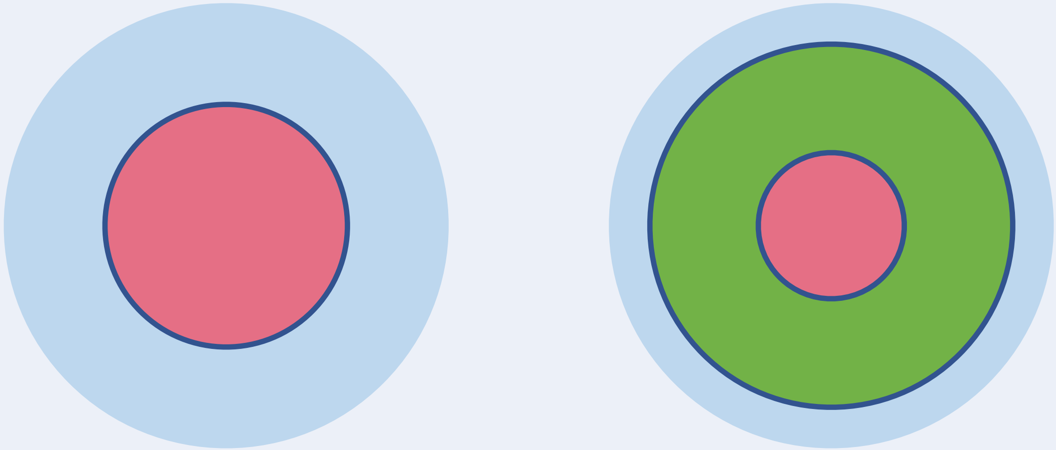
Internal Structure



Problem:

Inherent degeneracy - multiple structures can lead to the same mean density

Example:



UNIVERSITÄT BERN

CSH
CENTER FOR SPACE AND HABITABILITY

Calculate conditional probability of model parameters (x) given the observations (y)

Bayes' Theorem:

$$p(x|y) = \frac{p(y|x)\,p(x)}{p(y)}$$

$p(x\|y)$	posterior
$p(y\|x)$	likelihood
$p(x)$	prior
$p(y)$	marginal likelihood (Bayes integral)

Bayesian notion of probability:
$p(x)$ - degree of belief that x will occur

Internal structure and composition

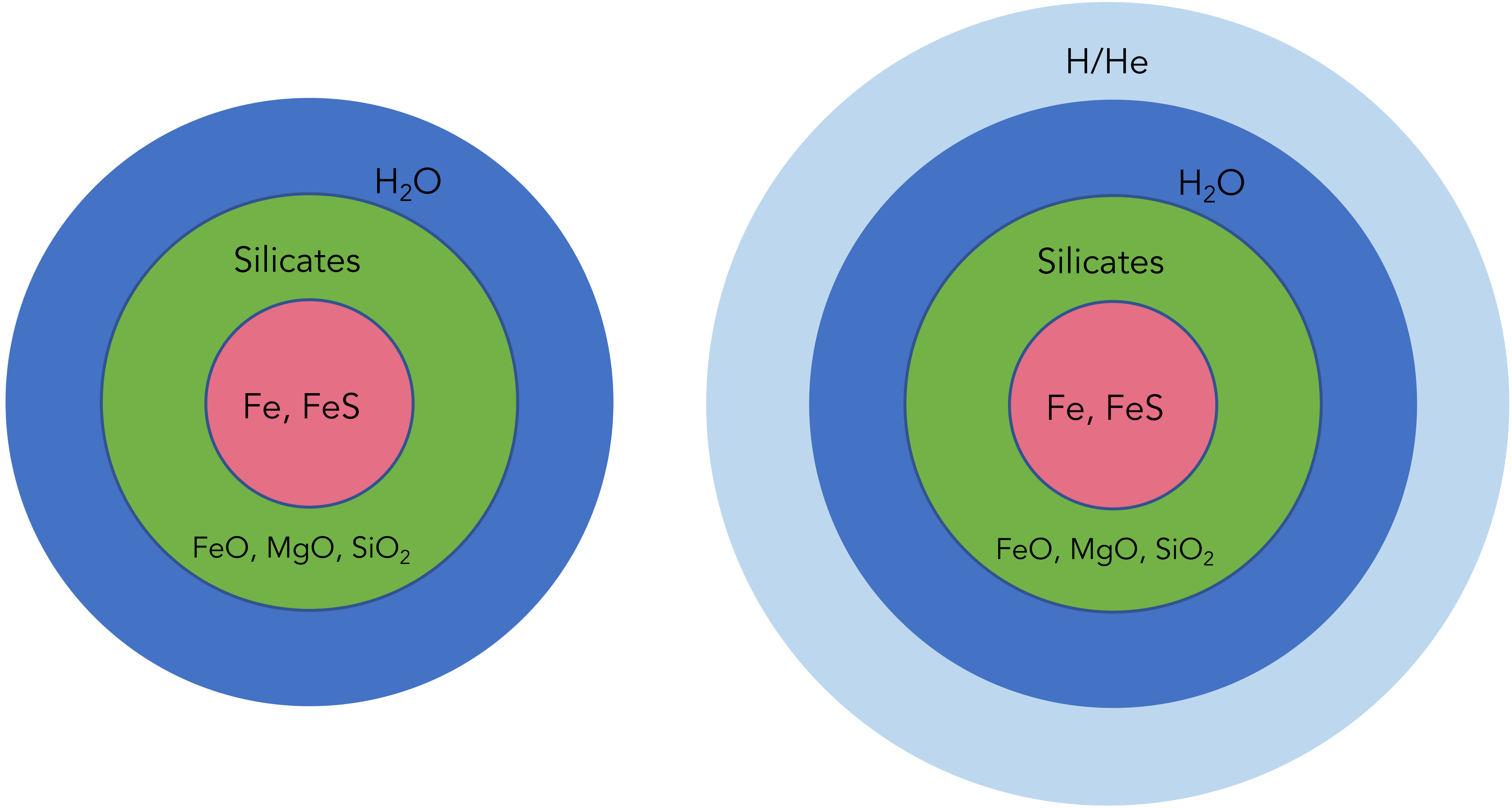
UNIVERSITÄT BERN
CSH CENTER FOR SPACE AND HABITABILITY

Input parameters:
mass
luminosity
equilibrium temperature
layer mass fractions
core composition
mantle composition

Model solves the planetary structure equations (Kippenhahn et al. 2012)

Output parameters:
transit radius
layer thicknesses

Spherically symmetric 1D model with fully differentiated layers:



Internal structure and composition

UNIVERSITÄT BERN

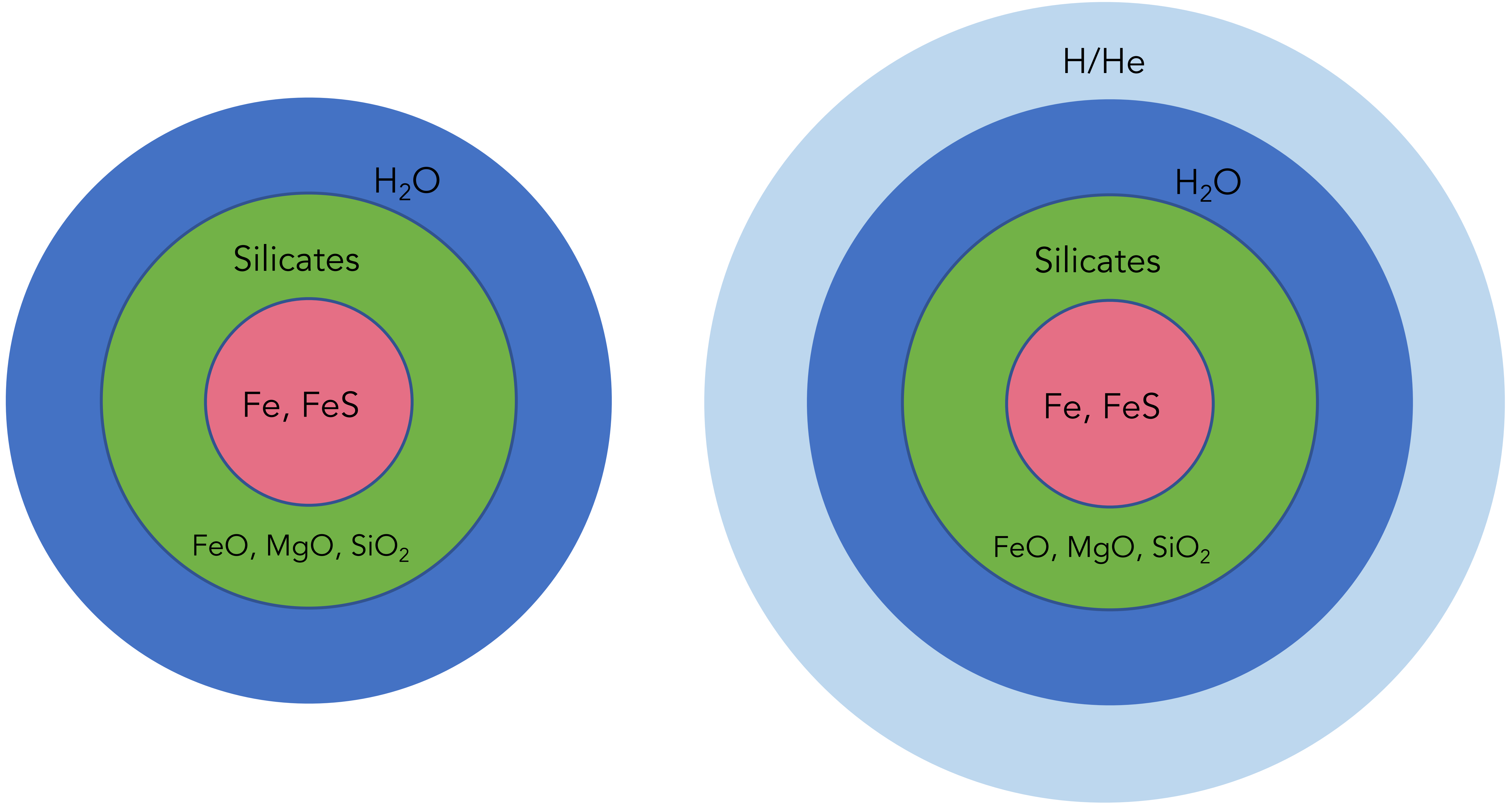
CSH CENTER FOR SPACE AND HABITABILITY

Input parameters:
mass
luminosity
equilibrium temperature
layer mass fractions
core composition
mantle composition

Model solves the planetary structure equations (Kippenhahn et al. 2012)

Output parameters:
transit radius
layer thicknesses

Spherically symmetric 1D model with fully differentiated layers:



Problem:

Solving planetary structure equations takes computation time

(around 2 hours for 1000 structures)

Sampling the posterior distribution

1- MCMC approach

Method: random walk inside parameter space, biased to explore ideal part of it for the given problem

Problems:

- difficult to know when MCMC method has actually converged to final distribution
- in case of multimodal posterior distribution, potentially only one mode is found

2- Brut force approach

Method: explore the whole parameter space

Problems:

- even more models to compute

Sampling the posterior distribution

UNIVERSITÄT BERN
CSH CENTER FOR SPACE AND HABITABILITY

Problem with both MCMC and brut force approach:

Measurements of observable quantities taken with respect to stellar quantities
high correlation between planets in multiplanetary systems

Modelling n planets in a system simultaneously means
- dimension of parameter space is multiplied by n
- necessary number of sampling points (and therefore computation time) increases drastically

Internal structure and composition

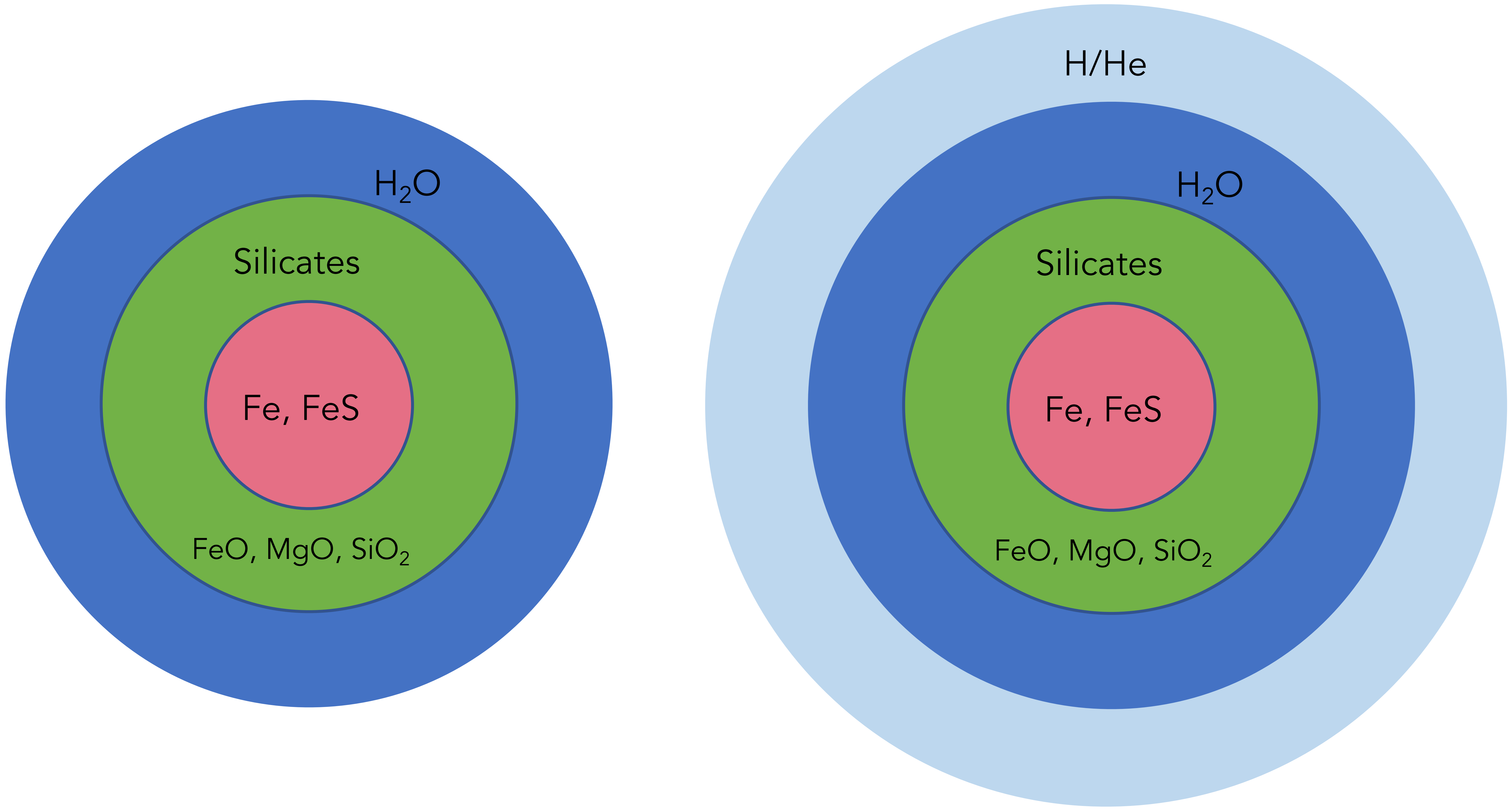
UNIVERSITÄT BERN
CSH CENTER FOR SPACE AND HABITABILITY

Input parameters:
mass
luminosity
equilibrium temperature
layer mass fractions
core composition
mantle composition

Model solves the planetary structure equations (Kippenhahn et al. 2012)

Output parameters:
transit radius
layer thicknesses

Spherically symmetric 1D model with fully differentiated layers:



Problem:
Solving planetary structure equations takes computation time
(around 2 hours for 1000 structures)

Internal structure and composition

UNIVERSITÄT BERN
CSH
CENTER FOR SPACE AND HABITABILITY

Input parameters:
mass
luminosity
equilibrium temperature
layer mass fractions
core composition
mantle composition

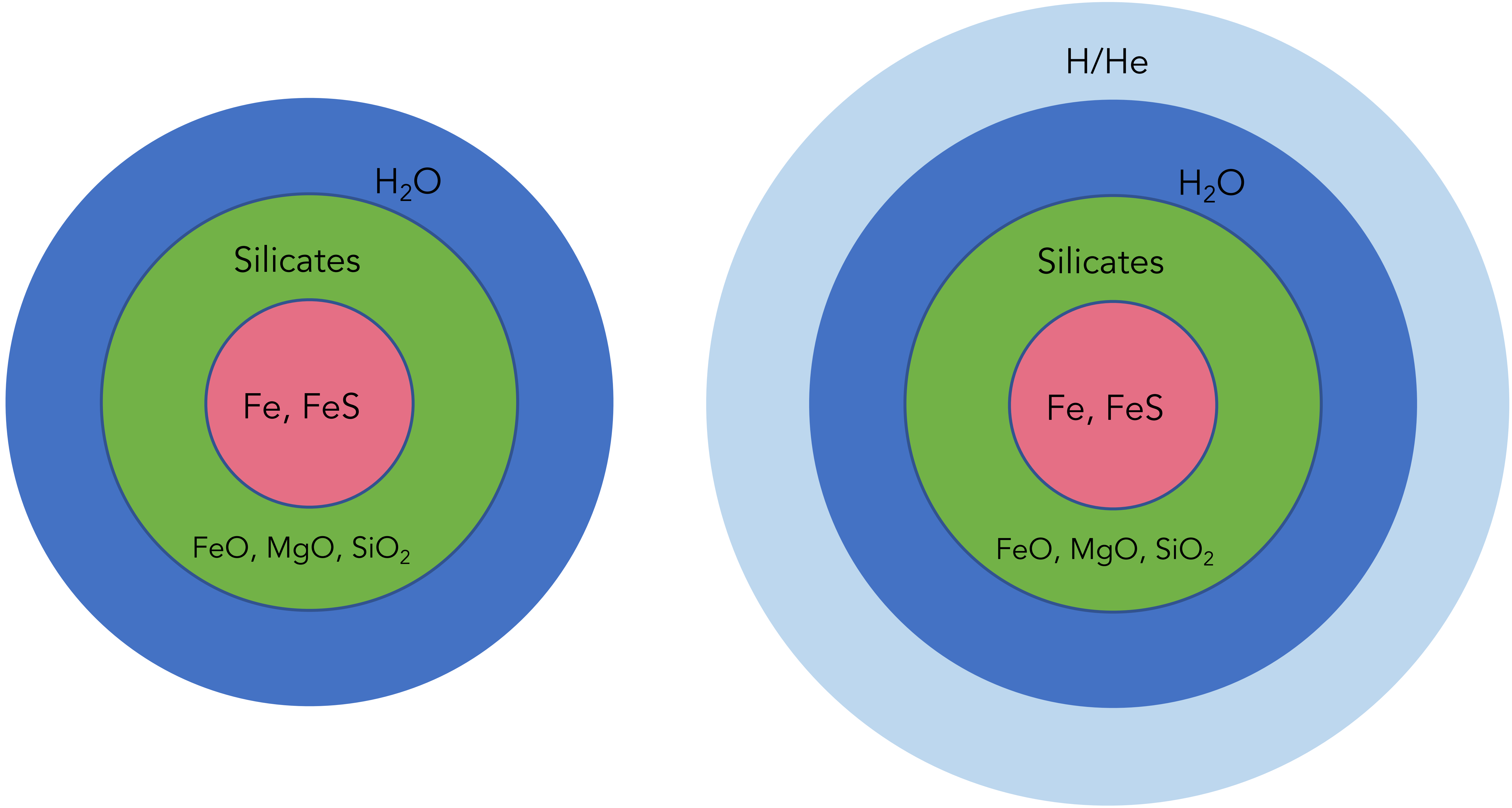
↓

Deep Neural Network (DNN)

↓

Output parameters:
transit radius
layer thicknesses

Spherically symmetric 1D model
with fully differentiated layers:



➡ **Train DNN based on internal structure model**

Training the DNN

UNIVERSITÄT BERN

CSH CENTER FOR SPACE AND HABITABILITY

Generate database of 37 million data points

M_{tot} [M_{earth}]	T_{eq} [K]	L_{int} [erg/s]	w_{core}	w_{water}	w_{atmo}	$x_{S,c}$	$x_{FeO,m}$	$x_{SiO2,m}$	$R_{transit}$ [R_{earth}]	R_{core} [R_{earth}]	R_{mantle} [R_{earth}]	R_{water} [R_{earth}]	R_{atmo} [R_{earth}]
1.645	1548.2	1.35E20	0.135	0.324	0.021	0.031	0.249	0.158	1.272	0.149	0.452	0.501	0.204
…	…	…	…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…

split into:

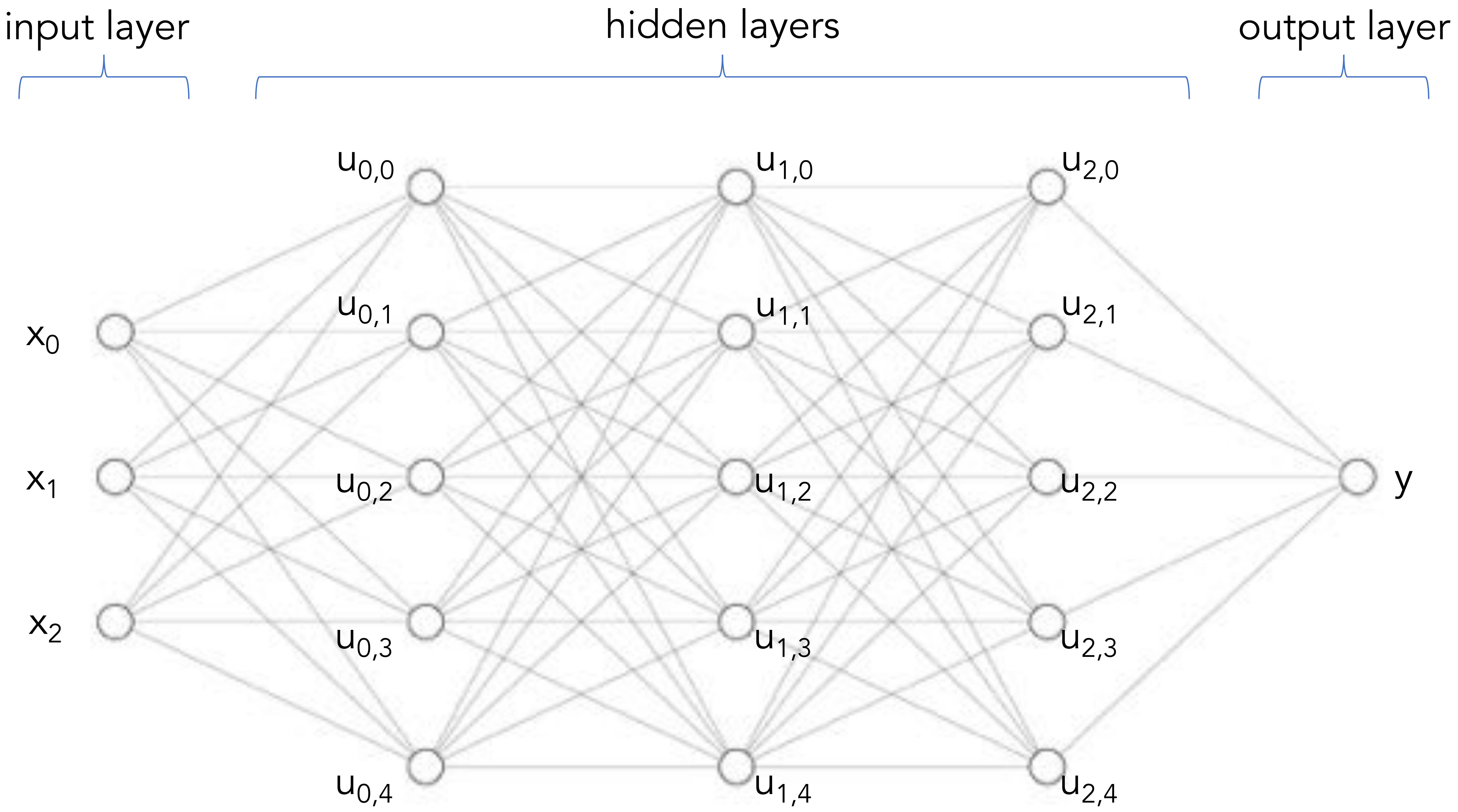
Training Data
20 million data points

Validation Data
2 million data points

Test Data
15 million data points

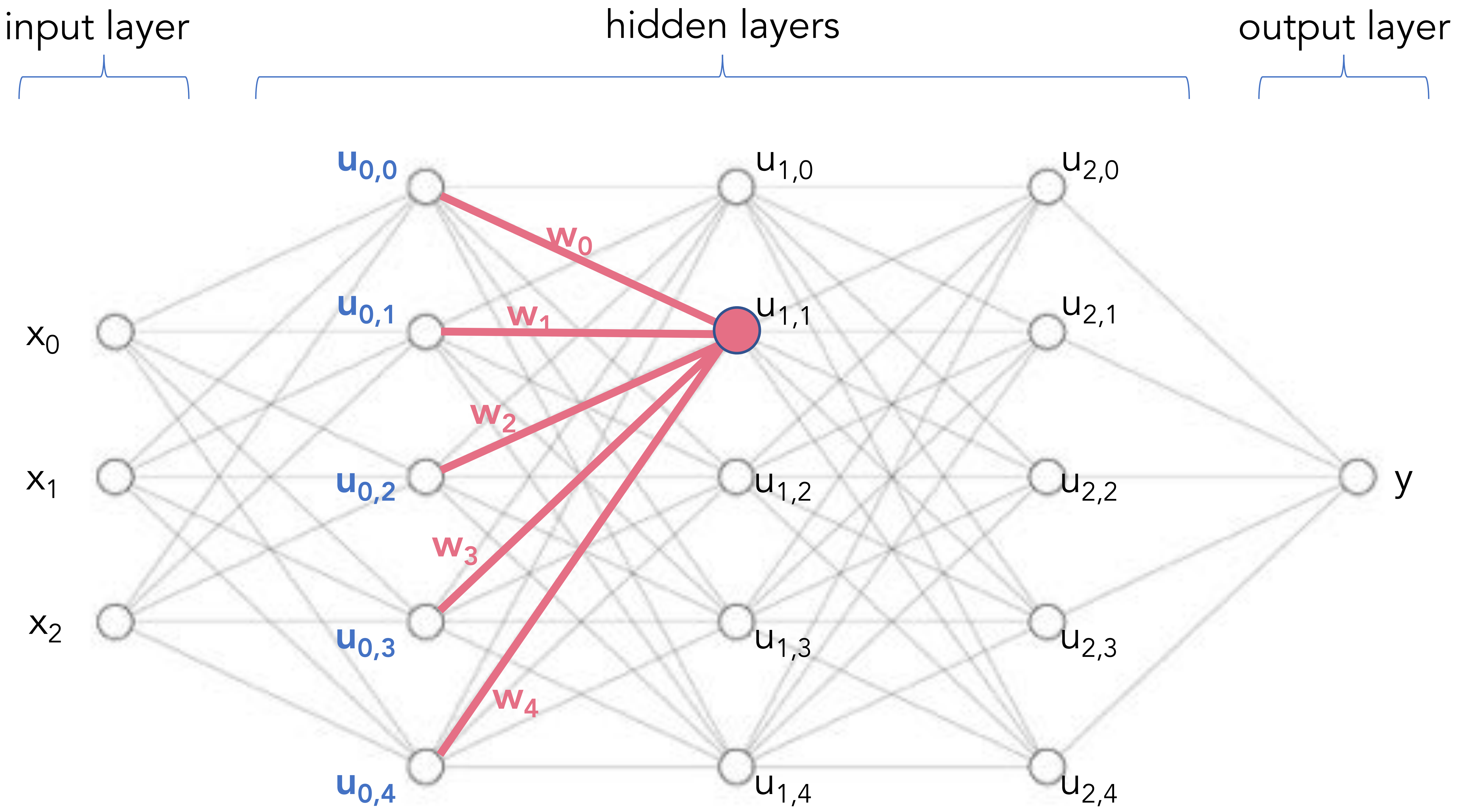
Training the DNN

UNIVERSITÄT BERN
CSH CENTER FOR SPACE AND HABITABILITY



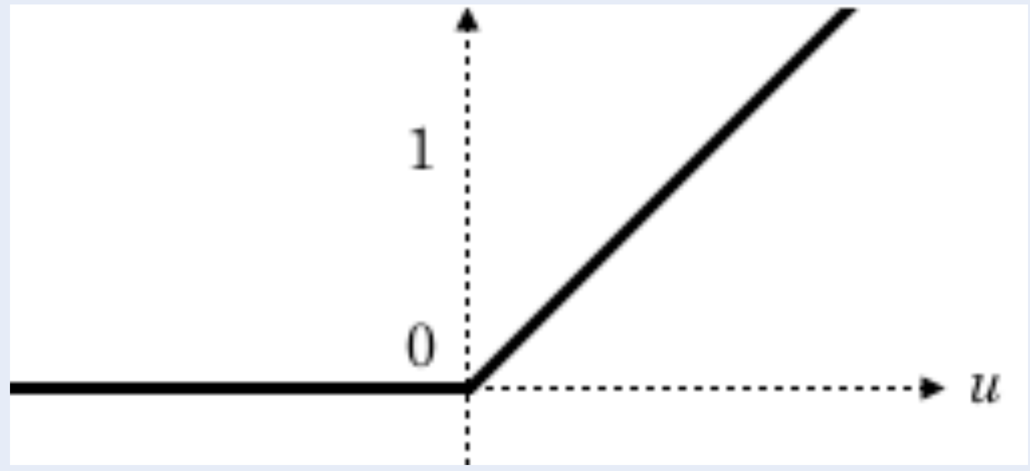
Training the DNN

UNIVERSITÄT BERN
CSH CENTER FOR SPACE AND HABITABILITY

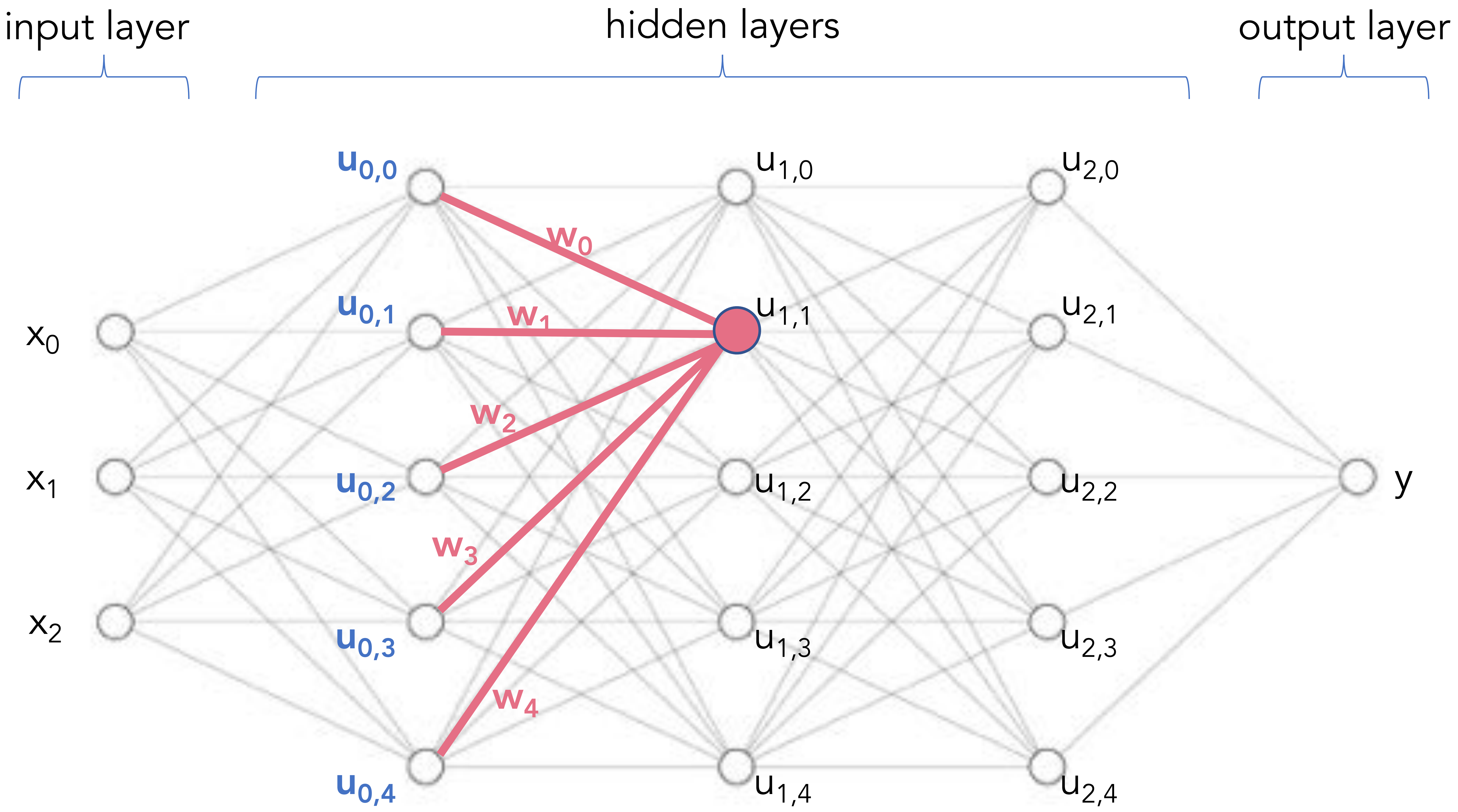


Calculate output:

Activation function is used to allow for nonlinearity, e.g. ReLu:



Training the DNN



Training DNN:

Define loss function, e.g. MSE:

Find good set of weights by searching for a local minimum of

Training the DNN

UNIVERSITÄT BERN

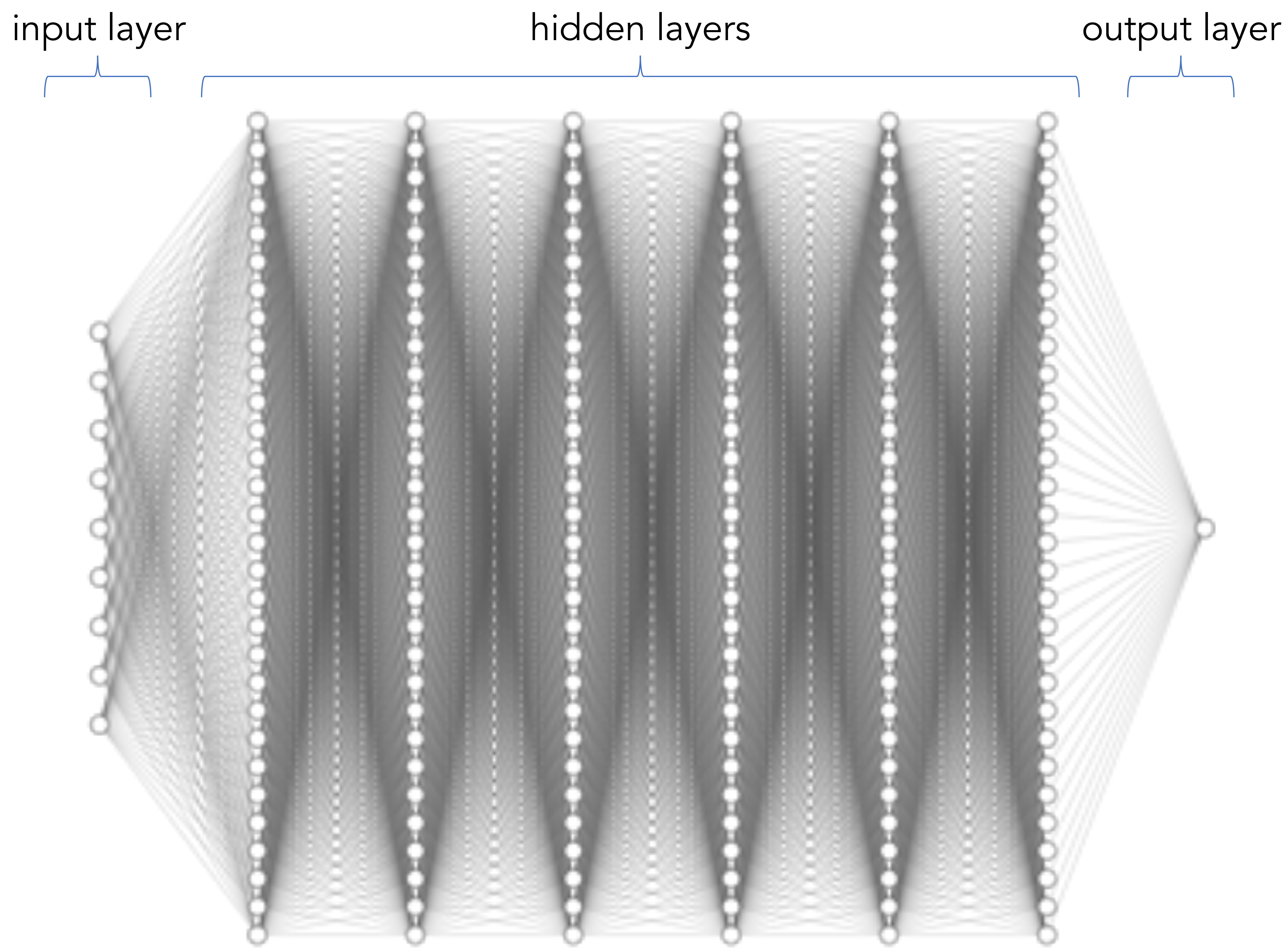
CSH CENTER FOR SPACE AND HABITABILITY

6 hidden layers

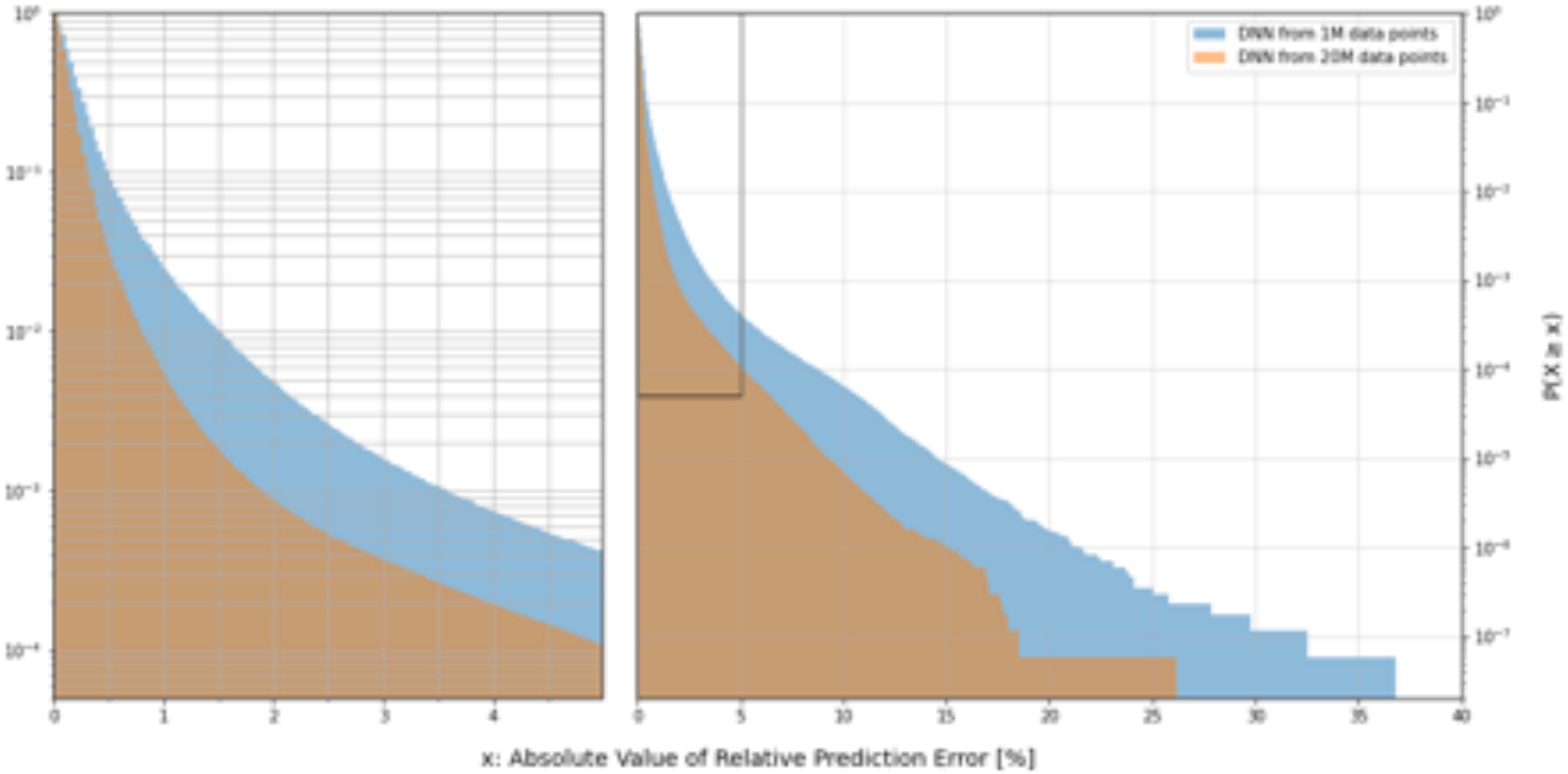
2048 units per hidden layer

Input parameters:
mass
luminosity
equilibrium temperature
layer mass fractions
core composition
mantle composition

Output parameters:
transit radius
future project: layer thicknesses



DNN performances



Data generation = 1 week

DNN performances

1. Choose priors of internal structure parameters

2. Sample from entire parameter space according to priors

3. For each sampled structure, use internal structure model to calculate radius

4. Keep the structure based on likelihood (probability of observation given parameters)

→ posterior distribution of internal structure parameters

Advantages:

- relatively low computation time (~1-3 hours)
- no problem of local maxima
- no problem of convergence

Sampling the posterior distribution

UNIVERSITÄT BERN

CSH CENTER FOR SPACE AND HABITABILITY

Problem with both MCMC and brut force approach:

Measurements of observable quantities taken with respect to stellar quantities
high correlation between planets in multiplanetary systems

Modelling n planets in a system simultaneously means
- dimension of parameter space is multiplied by n
- necessary number of sampling points (and therefore computation time) increases drastically

Sampling the posterior distribution

u^b
UNIVERSITÄT BERN
CSH
CENTER FOR SPACE AND HABITABILITY

1. Sample from stellar parameter space

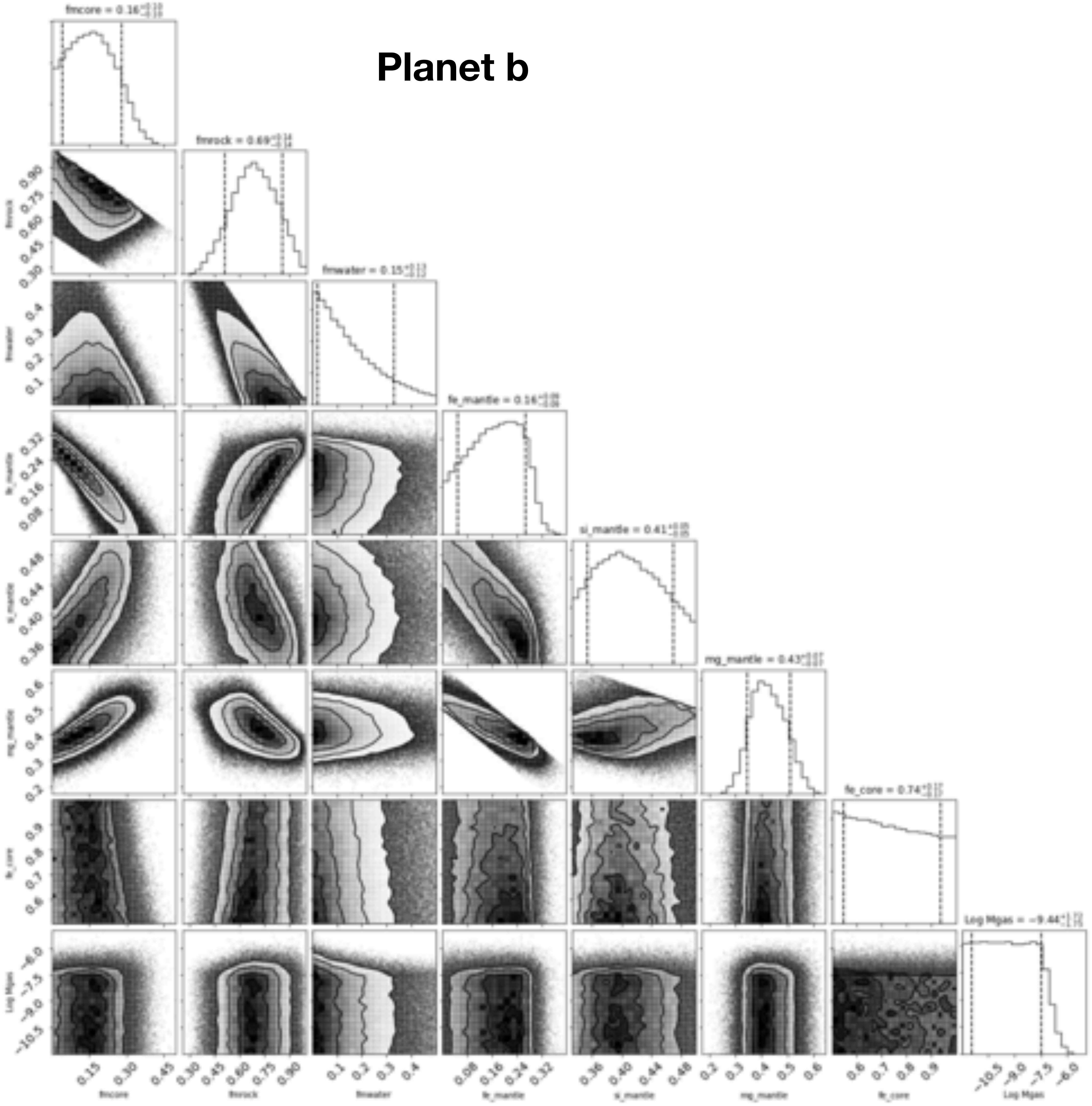
2. Create planets for each generated star:
   For each planet in the system, sample from priors of internal structure parameters

3. Use DNN to calculate radii of planets

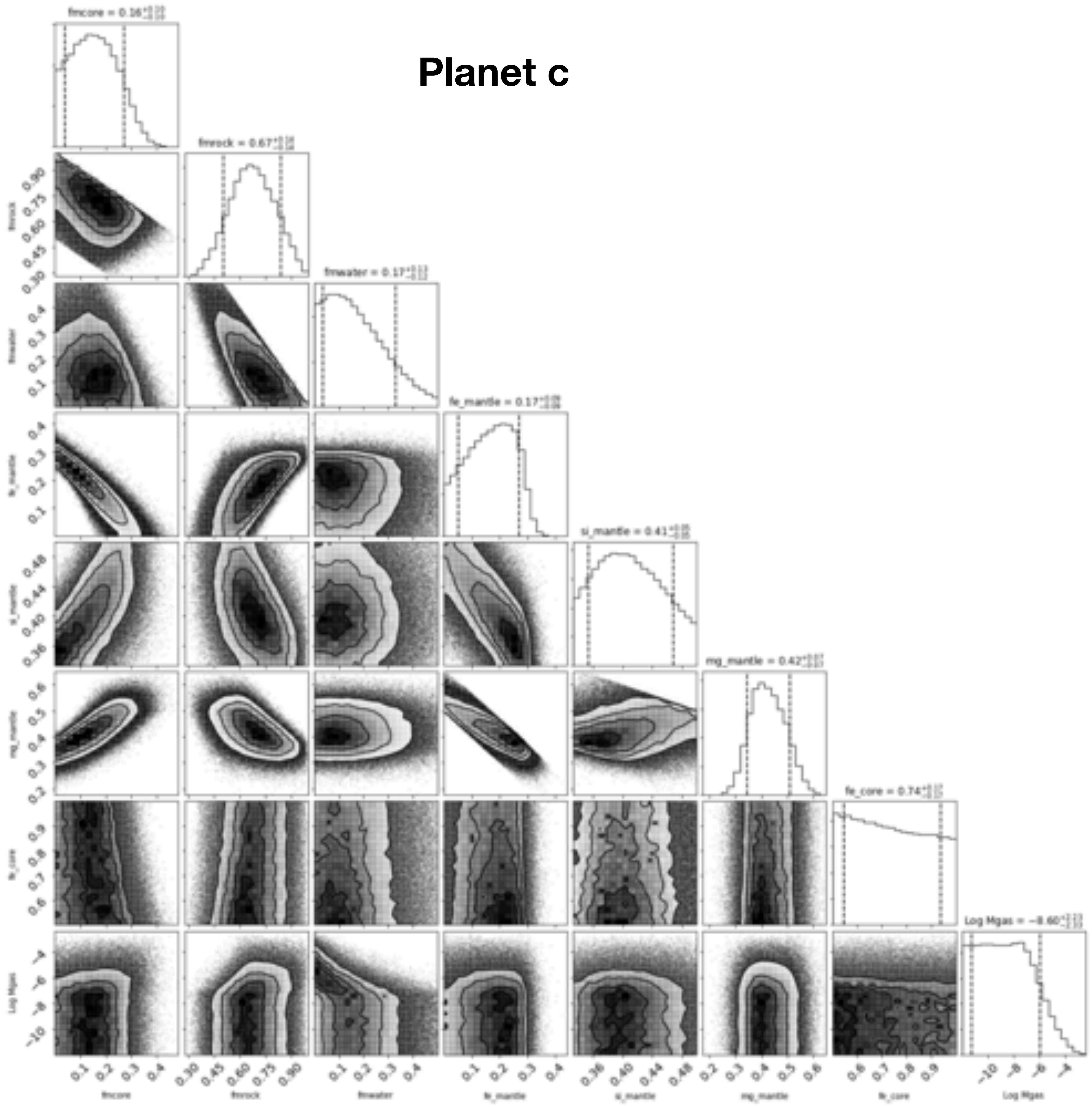
4. Calculate transit depth and keep the system based on likelihood of planets:

Exemple of TOI178

Planet b



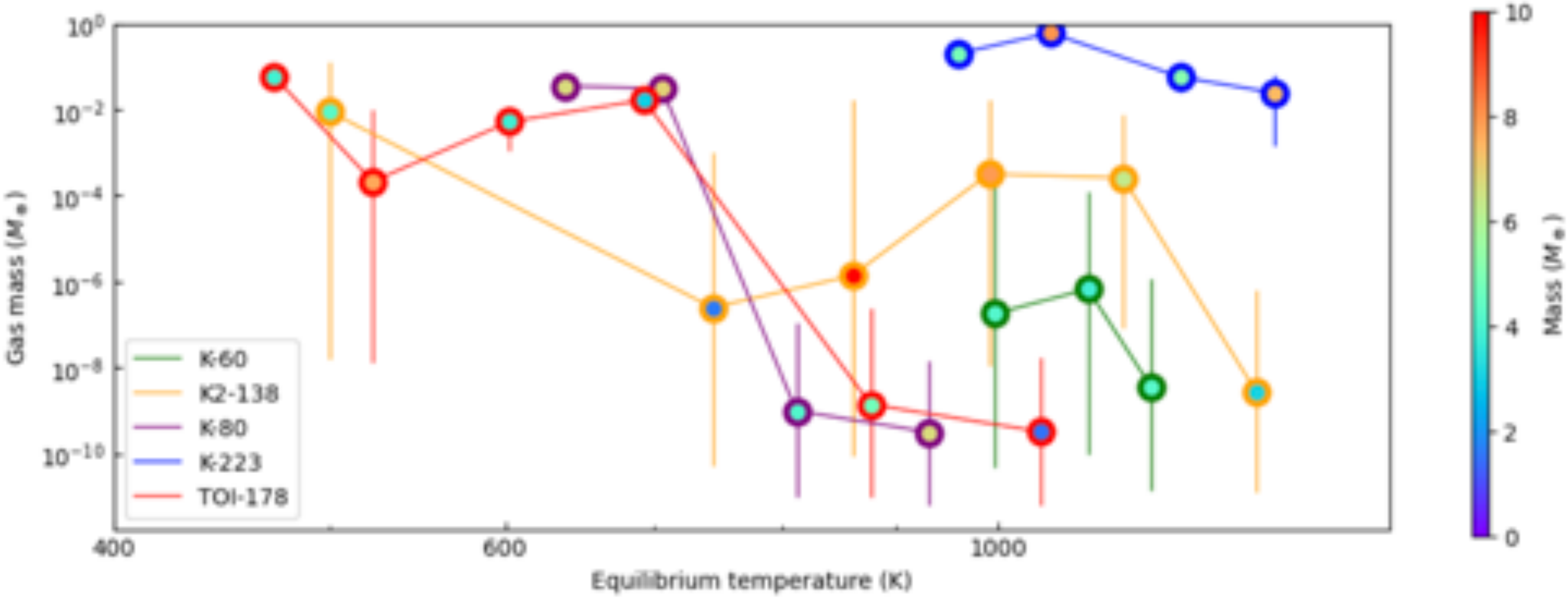
Planet c



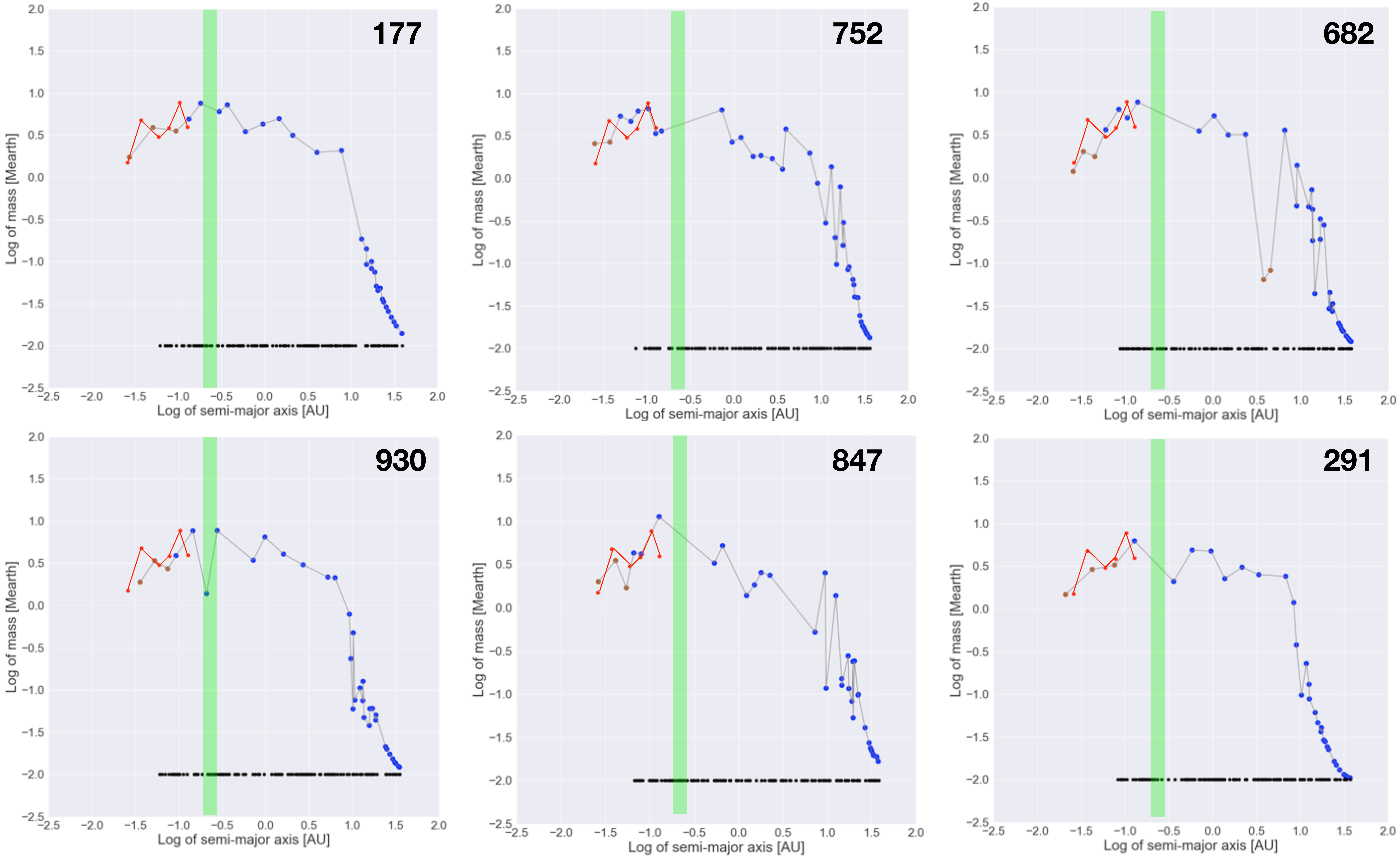
TOI178

UNIVERSITÄT BERN

CSH CENTER FOR SPACE AND HABITABILITY

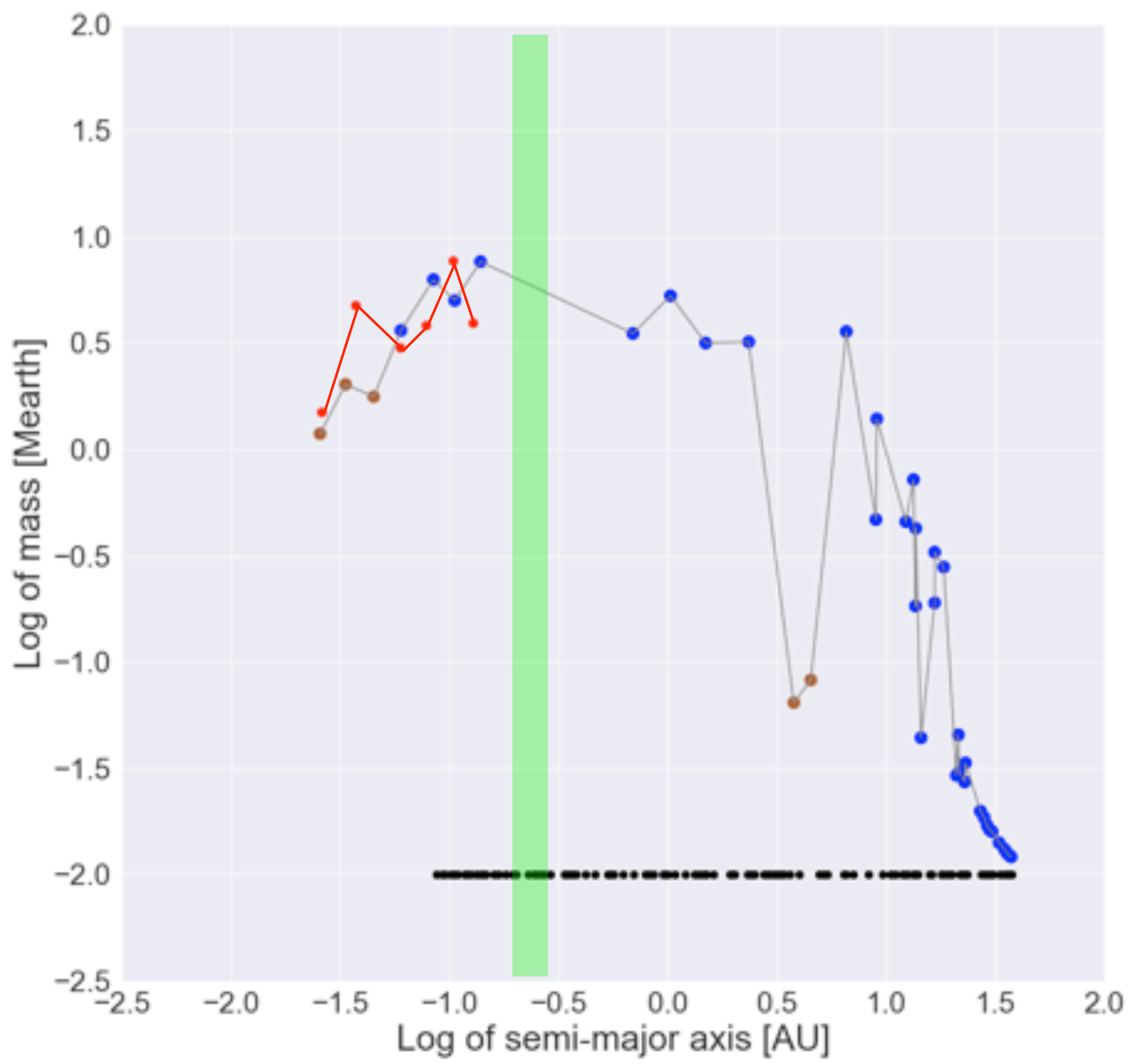


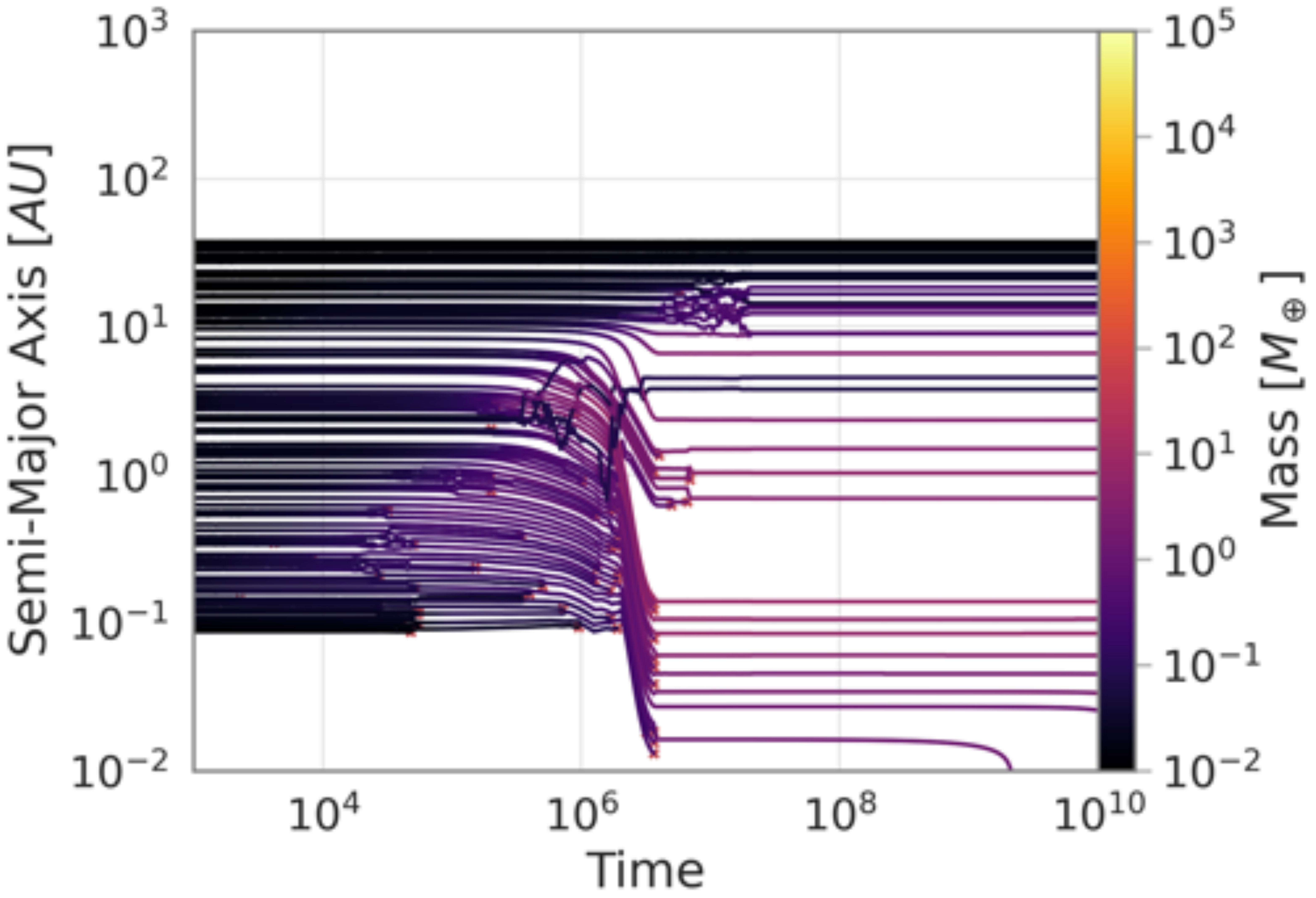
Exemple of TOI178



UNIVERSITÄT BERN
CSH CENTER FOR SPACE AND HABITABILITY

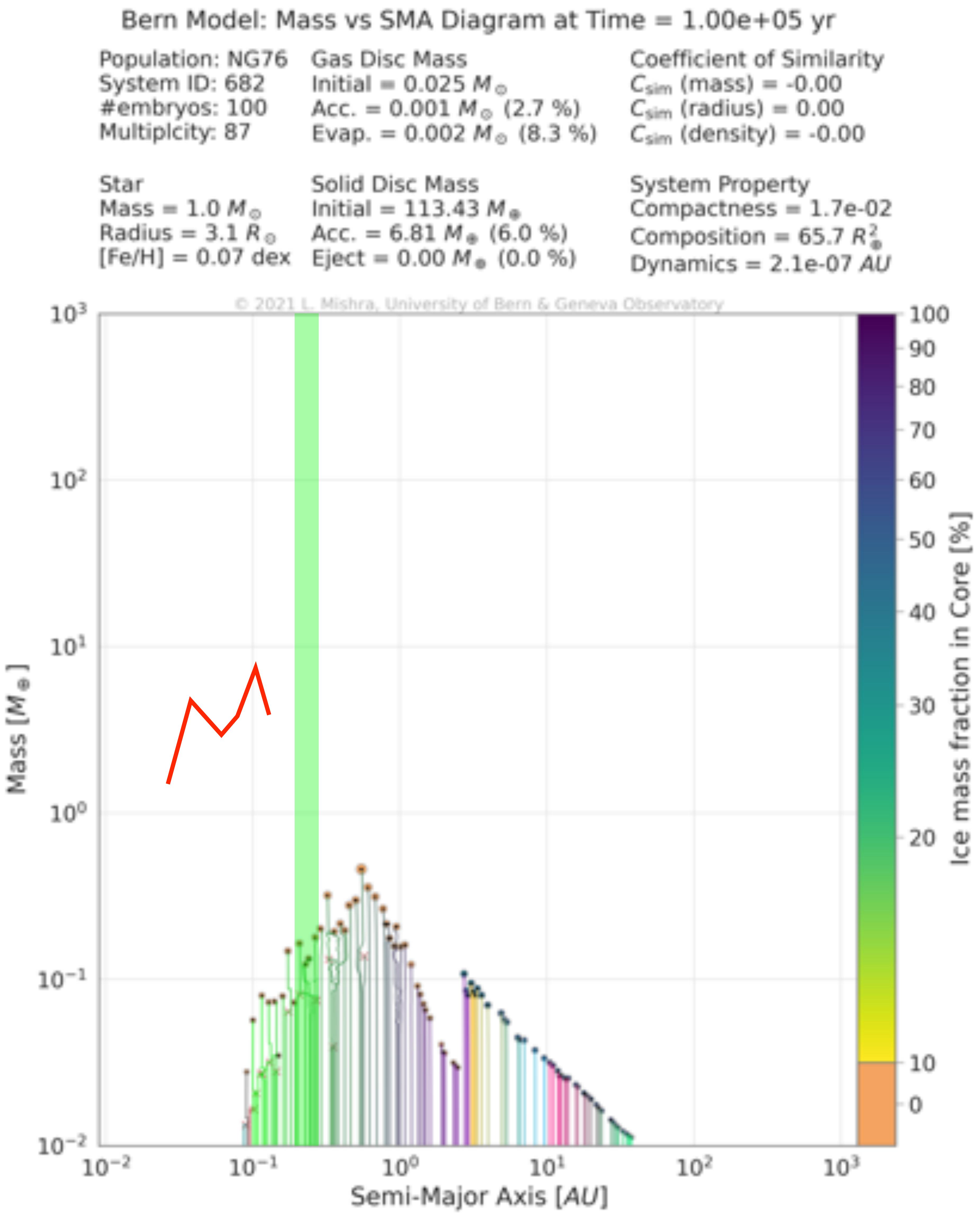
Exemple of TOI178



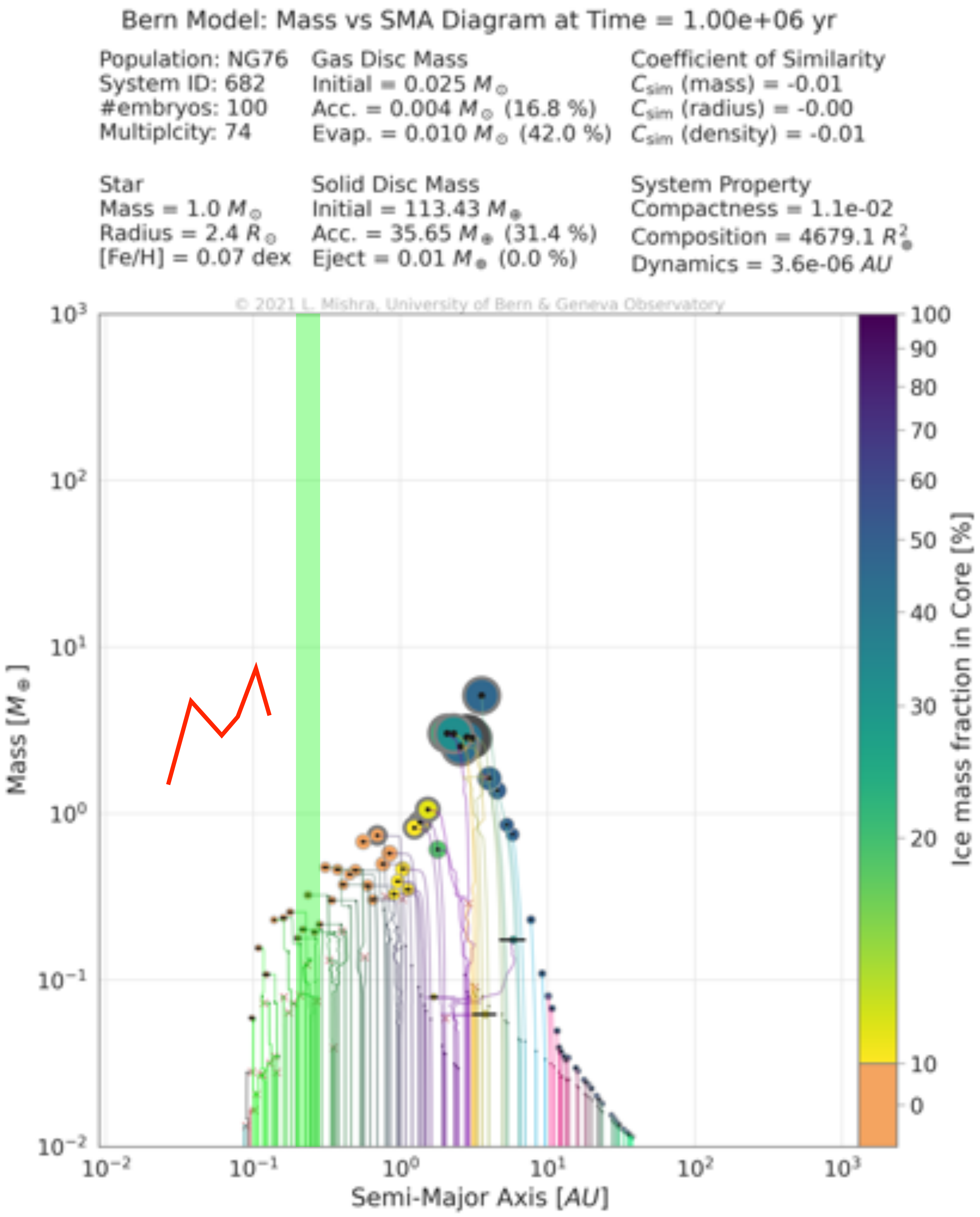


Strong density variation from planet to planet in the inner system: from 1 to 5

System 682

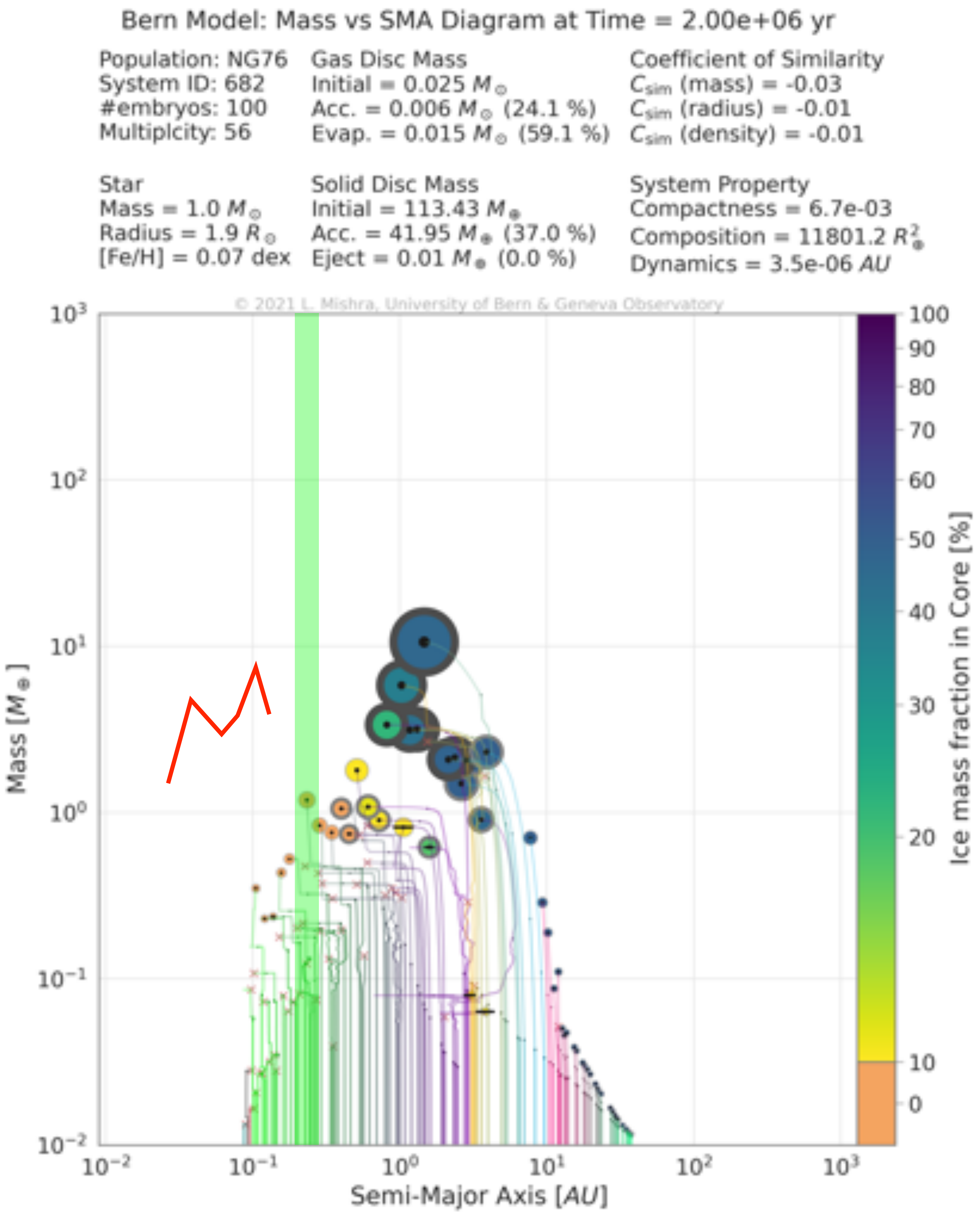


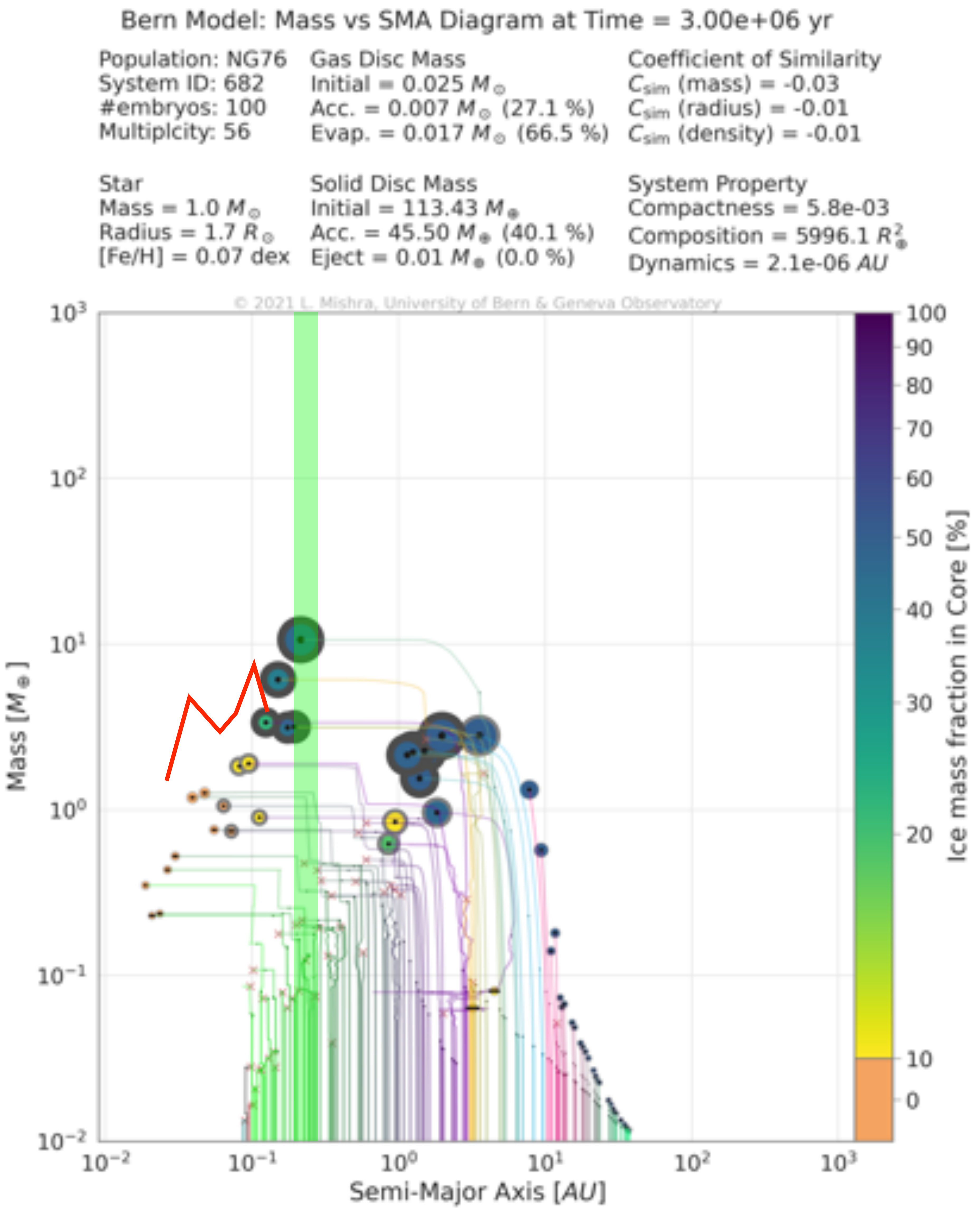
Bern Model: Mass vs SMA Diagram at Time = 1.00e+05 yr
Population: NG76
System ID: 682
#embryos: 100
Multiplcity: 87
Gas Disc Mass
Initial = 0.025 $M_\odot$
Acc. = 0.001 $M_\odot$ (2.7 %)
Evap. = 0.002 $M_\odot$ (8.3 %)
Coefficient of Similarity
C_{sim} (mass) = -0.00
C_{sim} (radius) = 0.00
C_{sim} (density) = -0.00
Star
Mass = 1.0 $M_\odot$
Radius = 3.1 $R_\odot$
[Fe/H] = 0.07 dex
Solid Disc Mass
Initial = 113.43 $M_\oplus$
Acc. = 6.81 $M_\oplus$ (6.0 %)
Eject = 0.00 $M_\oplus$ (0.0 %)
System Property
Compactness = 1.7e-02
Composition = 65.7 $R_\oplus^2$
Dynamics = 2.1e-07 AU
© 2021 L. Mishra, University of Bern & Geneva Observatory
Mass [$M_\oplus$]
Semi-Major Axis [AU]
Ice mass fraction in Core [%]



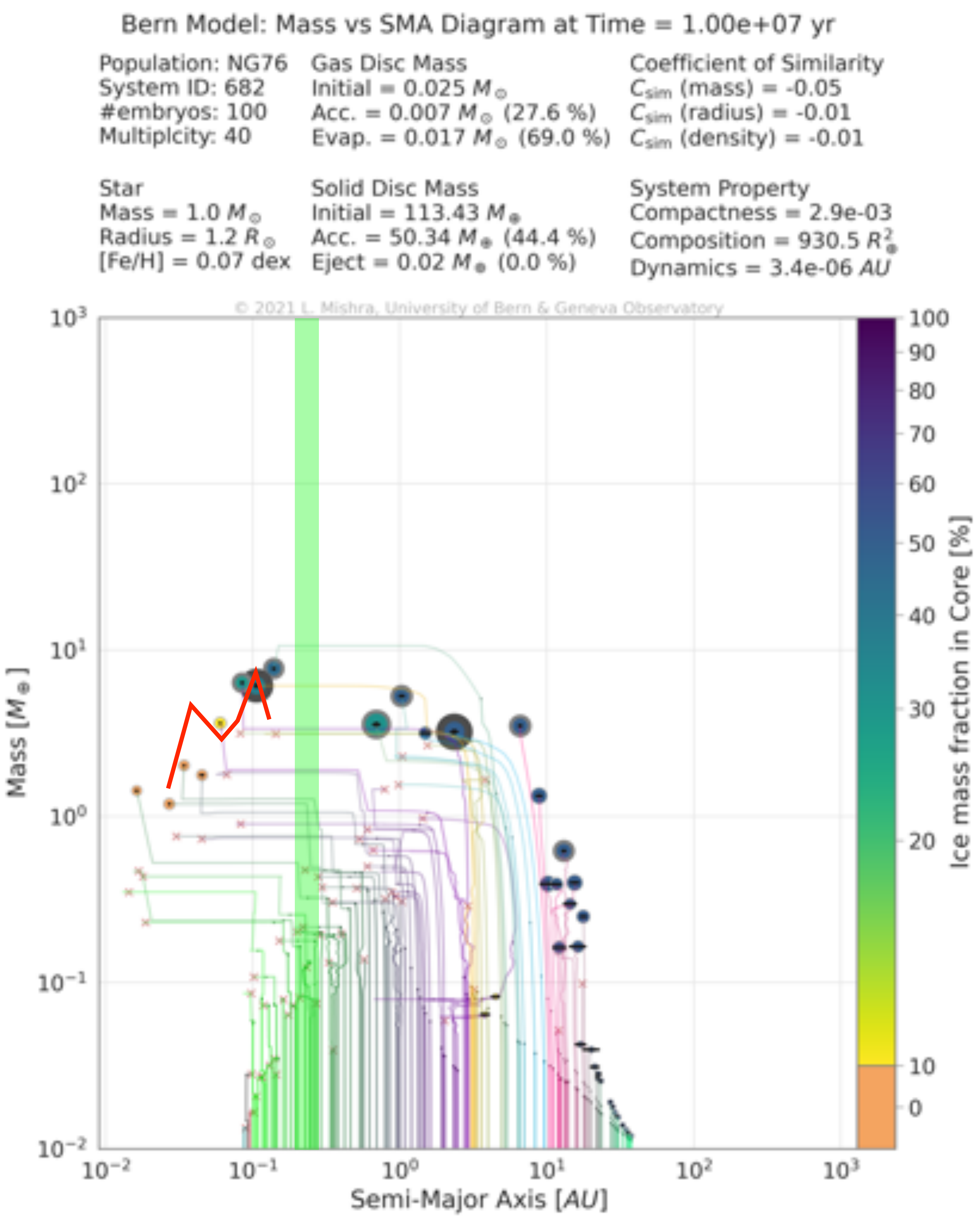
Bern Model: Mass vs SMA Diagram at Time = 1.00e+06 yr
Population: NG76
System ID: 682
#embryos: 100
Multiplcity: 74
Gas Disc Mass
Initial = 0.025 $M_\odot$
Acc. = 0.004 $M_\odot$ (16.8 %)
Evap. = 0.010 $M_\odot$ (42.0 %)
Coefficient of Similarity
C_{sim} (mass) = -0.01
C_{sim} (radius) = -0.00
C_{sim} (density) = -0.01
Star
Mass = 1.0 $M_\odot$
Radius = 2.4 $R_\odot$
[Fe/H] = 0.07 dex
Solid Disc Mass
Initial = 113.43 $M_\oplus$
Acc. = 35.65 $M_\oplus$ (31.4 %)
Eject = 0.01 $M_\oplus$ (0.0 %)
System Property
Compactness = 1.1e-02
Composition = 4679.1 $R_\oplus^2$
Dynamics = 3.6e-06 AU
© 2021 L. Mishra, University of Bern & Geneva Observatory
Mass [$M_\oplus$]
Semi-Major Axis [AU]
Ice mass fraction in Core [%]

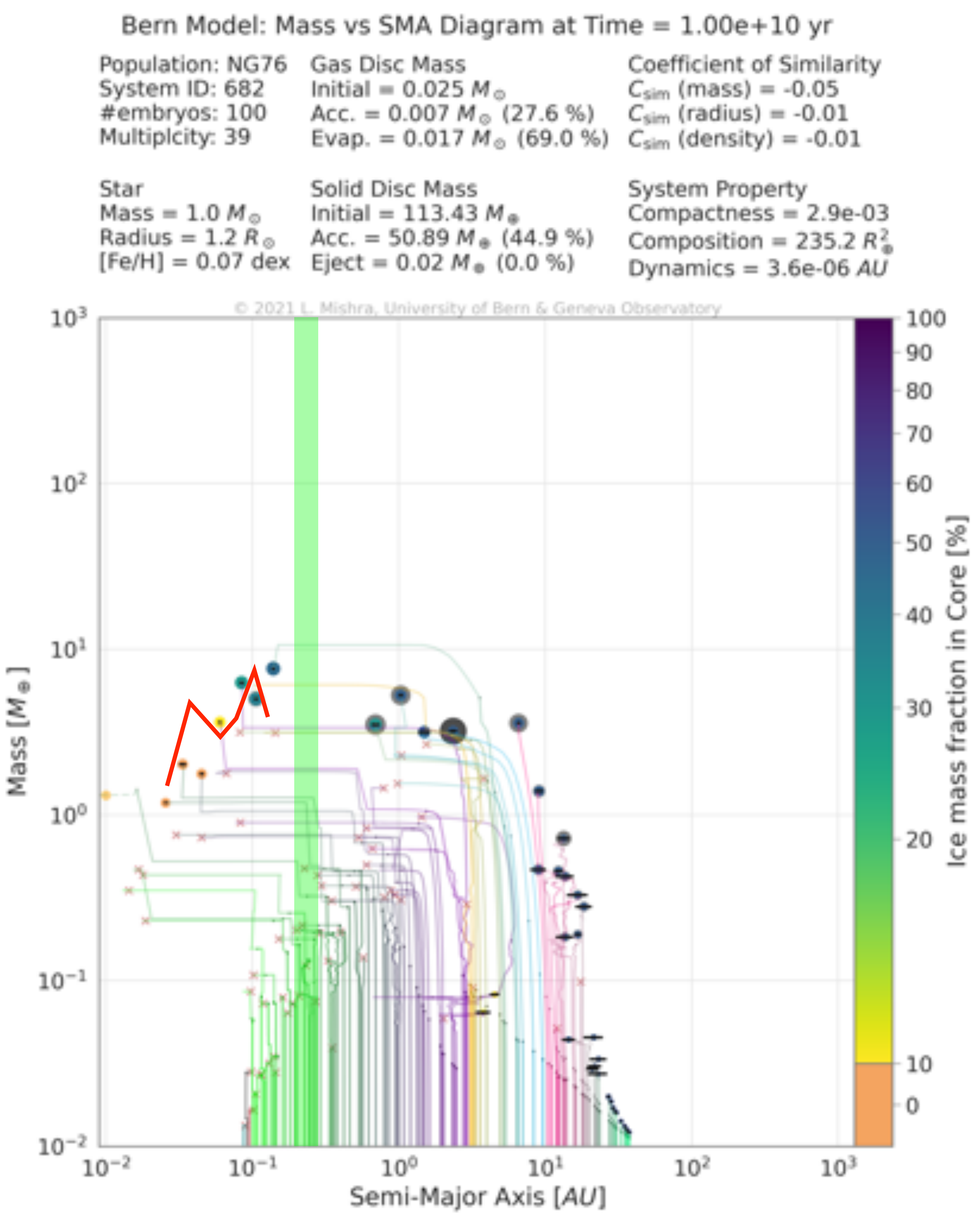
System 682


Bern Model: Mass vs SMA Diagram at Time = 2.00e+06 yr
Population: NG76
System ID: 682
#embryos: 100
Multiplcity: 56
Gas Disc Mass
Initial = 0.025 $M_\odot$
Acc. = 0.006 $M_\odot$ (24.1 %)
Evap. = 0.015 $M_\odot$ (59.1 %)
Coefficient of Similarity
C_{sim} (mass) = -0.03
C_{sim} (radius) = -0.01
C_{sim} (density) = -0.01
Star
Mass = 1.0 $M_\odot$
Radius = 1.9 $R_\odot$
[Fe/H] = 0.07 dex
Solid Disc Mass
Initial = 113.43 $M_\oplus$
Acc. = 41.95 $M_\oplus$ (37.0 %)
Eject = 0.01 $M_\oplus$ (0.0 %)
System Property
Compactness = 6.7e-03
Composition = 11801.2 $R_\oplus^2$
Dynamics = 3.5e-06 AU
© 2021 L. Mishra, University of Bern & Geneva Observatory
Mass [$M_\oplus$]
Semi-Major Axis [AU]
Ice mass fraction in Core [%]


Bern Model: Mass vs SMA Diagram at Time = 3.00e+06 yr
Population: NG76
System ID: 682
#embryos: 100
Multiplcity: 56
Gas Disc Mass
Initial = 0.025 $M_\odot$
Acc. = 0.007 $M_\odot$ (27.1 %)
Evap. = 0.017 $M_\odot$ (66.5 %)
Coefficient of Similarity
C_{sim} (mass) = -0.03
C_{sim} (radius) = -0.01
C_{sim} (density) = -0.01
Star
Mass = 1.0 $M_\odot$
Radius = 1.7 $R_\odot$
[Fe/H] = 0.07 dex
Solid Disc Mass
Initial = 113.43 $M_\oplus$
Acc. = 45.50 $M_\oplus$ (40.1 %)
Eject = 0.01 $M_\oplus$ (0.0 %)
System Property
Compactness = 5.8e-03
Composition = 5996.1 $R_\oplus^2$
Dynamics = 2.1e-06 AU
© 2021 L. Mishra, University of Bern & Geneva Observatory
Mass [$M_\oplus$]
Semi-Major Axis [AU]
Ice mass fraction in Core [%]

System 682


Bern Model: Mass vs SMA Diagram at Time = 1.00e+07 yr
Population: NG76
System ID: 682
#embryos: 100
Multiplcity: 40
Gas Disc Mass
Initial = 0.025 $M_\odot$
Acc. = 0.007 $M_\odot$ (27.6 %)
Evap. = 0.017 $M_\odot$ (69.0 %)
Coefficient of Similarity
C_{sim} (mass) = -0.05
C_{sim} (radius) = -0.01
C_{sim} (density) = -0.01
Star
Mass = 1.0 $M_\odot$
Radius = 1.2 $R_\odot$
[Fe/H] = 0.07 dex
Solid Disc Mass
Initial = 113.43 $M_\oplus$
Acc. = 50.34 $M_\oplus$ (44.4 %)
Eject = 0.02 $M_\oplus$ (0.0 %)
System Property
Compactness = 2.9e-03
Composition = 930.5 $R_\oplus^2$
Dynamics = 3.4e-06 AU
© 2021 L. Mishra, University of Bern & Geneva Observatory
Mass [$M_\oplus$]
Semi-Major Axis [AU]
Ice mass fraction in Core [%]


Bern Model: Mass vs SMA Diagram at Time = 1.00e+10 yr
Population: NG76
System ID: 682
#embryos: 100
Multiplcity: 39
Gas Disc Mass
Initial = 0.025 $M_\odot$
Acc. = 0.007 $M_\odot$ (27.6 %)
Evap. = 0.017 $M_\odot$ (69.0 %)
Coefficient of Similarity
C_{sim} (mass) = -0.05
C_{sim} (radius) = -0.01
C_{sim} (density) = -0.01
Star
Mass = 1.0 $M_\odot$
Radius = 1.2 $R_\odot$
[Fe/H] = 0.07 dex
Solid Disc Mass
Initial = 113.43 $M_\oplus$
Acc. = 50.89 $M_\oplus$ (44.9 %)
Eject = 0.02 $M_\oplus$ (0.0 %)
System Property
Compactness = 2.9e-03
Composition = 235.2 $R_\oplus^2$
Dynamics = 3.6e-06 AU
© 2021 L. Mishra, University of Bern & Geneva Observatory
Mass [$M_\oplus$]
Semi-Major Axis [AU]
Ice mass fraction in Core [%]

UNIVERSITÄT BERN
CSH CENTER FOR SPACE AND HABITABILITY

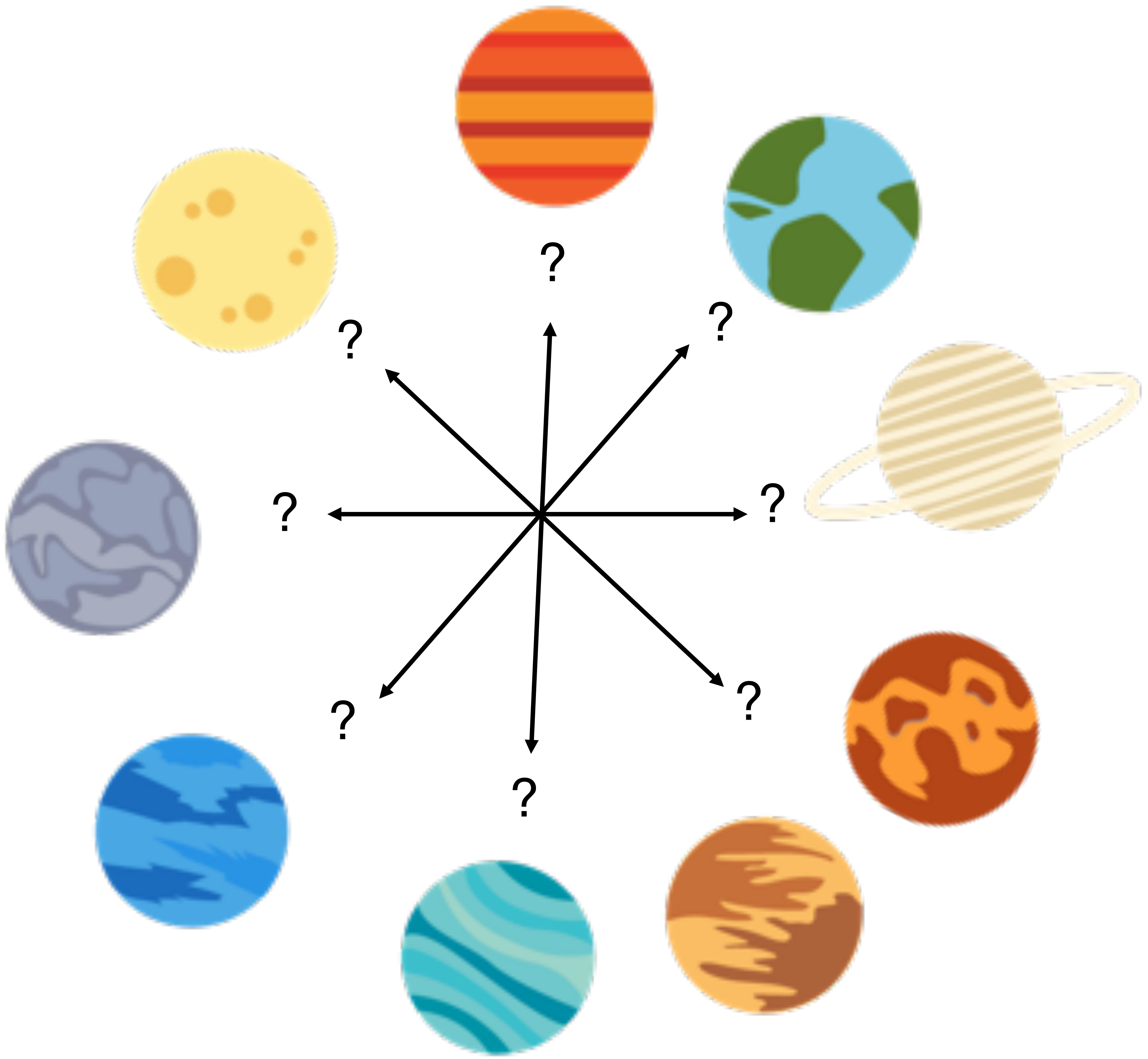
1- Planet formation and the Bern model

2- Models *versus* observations

3- Planetary internal structure and Deep Learning

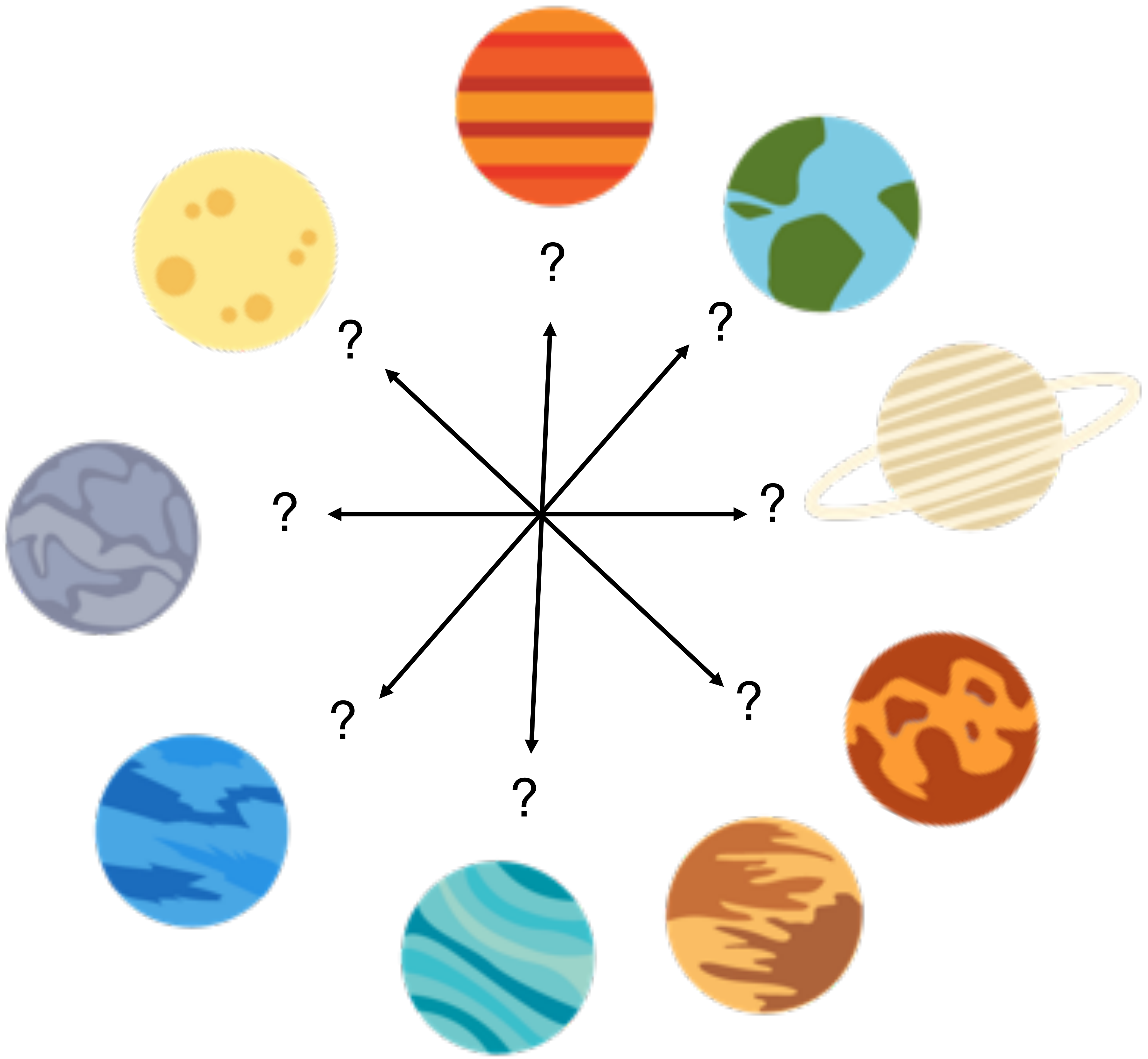
4- Correlation in planetary systems with random forrest: finding a second Earth

Correlations in planetary systems



Correlations in planetary systems

u^b
UNIVERSITÄT BERN
CSH CENTER FOR SPACE AND HABITABILITY

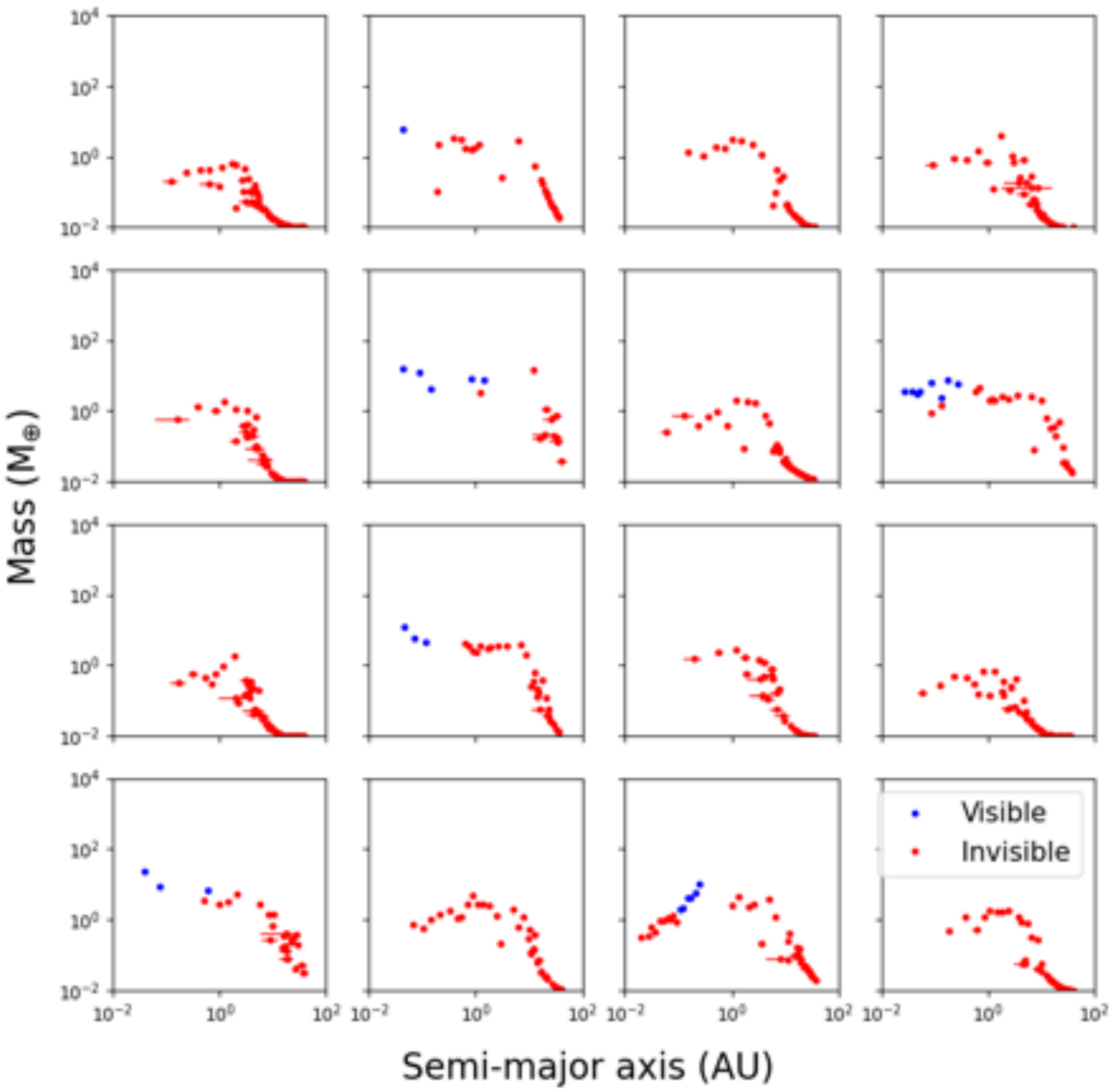


Formation and evolution of planetary systems ?

Properties
of
hidden planets ?

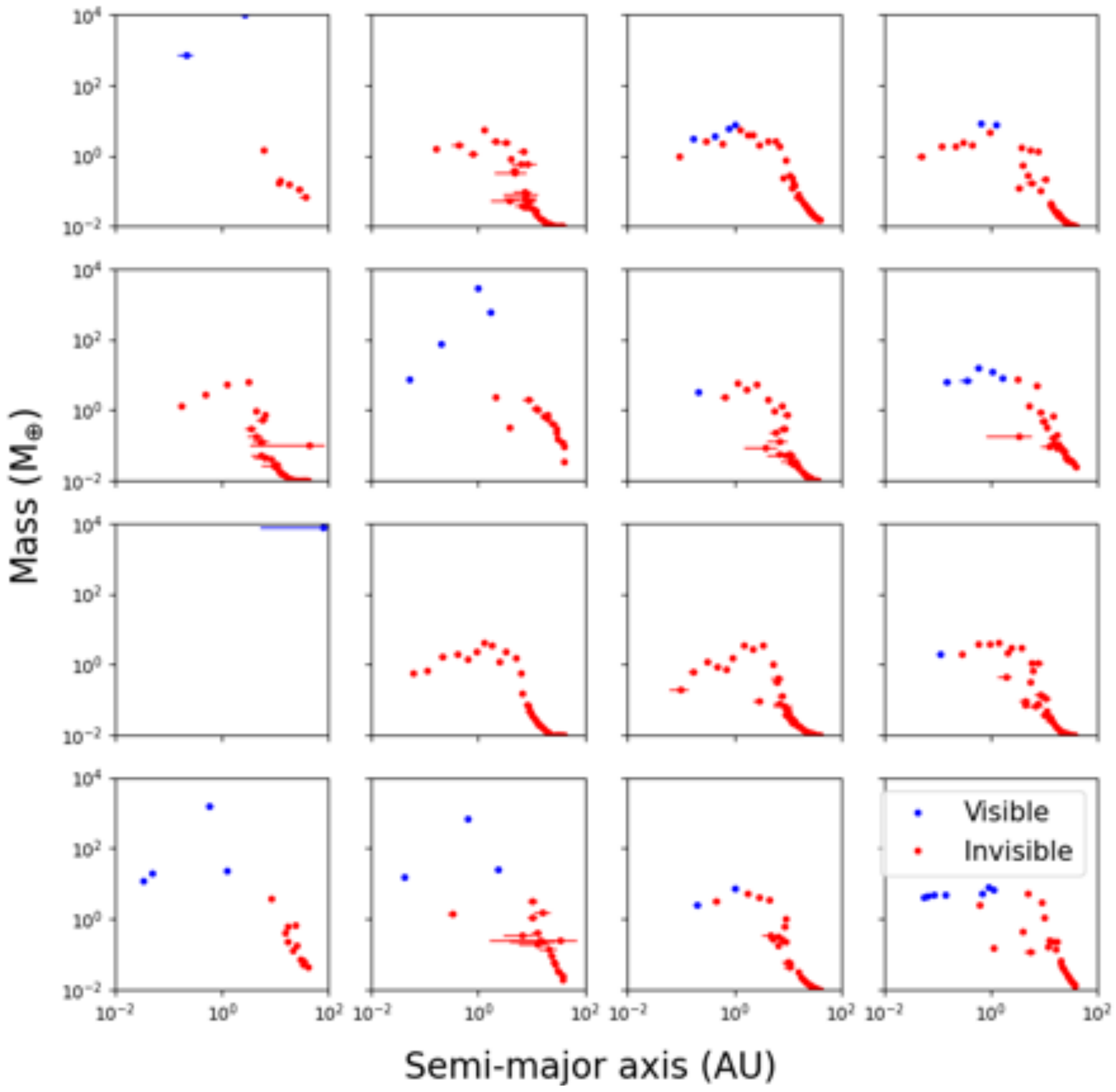
Systems with Earth-like planet

UNIVERSITÄT BERN
CSH CENTER FOR SPACE AND HABITABILITY



- Systems look regular
- Almost no Giant planet (>100 $M_\oplus$)
- Some Neptunes (~ 10 $M_\oplus$)
- Few visible planets
- The innermost planet is often smaller than 1 Mearth

Systems without Earth-like planet



16 systems from population NG76 without any Earth-like planet

- Systems look less regular
- More giant planets (> 100 M⊕)
- Presence of a Neptune is common (> 10 M⊕)
- More visible planets
- The mass of the innermost planet can be high

Random forrest classifier

- The number of visible planets
- The number of giant planets inside 5 AU
- The innermost planet’s mass

- Metallicity of the star
- Mass of the star

Random forrest classifier

1- 1000 trees, all slightly different

2- Final decision depends on fraction of positive/negative trees (50%, 70%...)

3- Performances depend on this threshold

≥ 50%	Precision score	0.73
	Recall score	0.87
	Accuracy score	0.74
≥ 70%	Precision score	0.83
	Recall score	0.44
	Accuracy score	0.61
≥ 90%	Precision score	0.86
	Recall score	0.05
	Accuracy score	0.44

Random forrest classifier

1- 1000 trees, all slightly different

2- Final decision depends on fraction of positive/negative trees (50%, 70%...)

3- Performances depend on this threshold

50%	Precision score	0.73
	Recall score	0.87
70%	Precision score	0.83
	Recall score	0.44

<- more false positive

<- more false negative

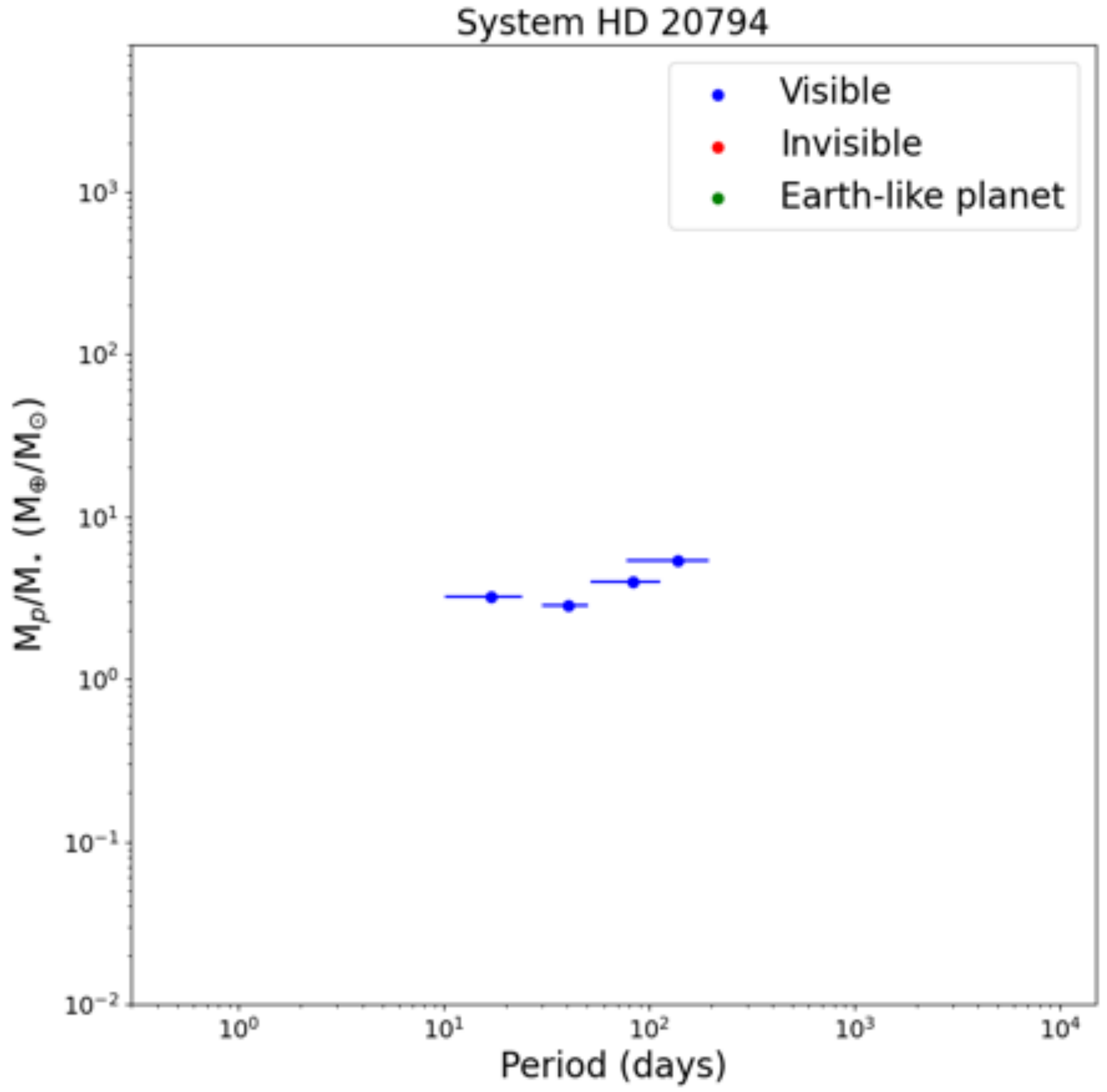
4- 3 systems identify > 80% threshold

Random forrest classifier

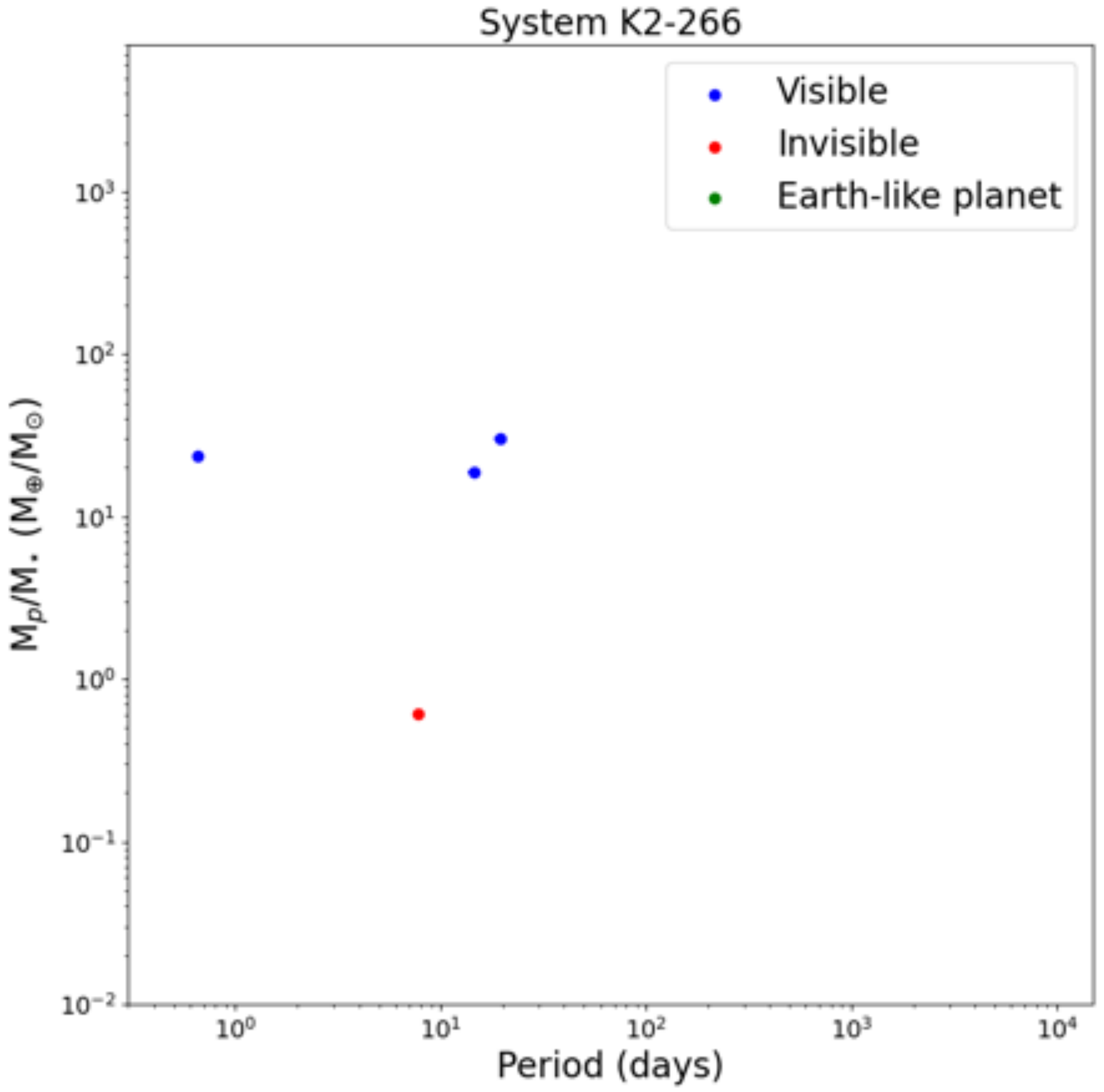
UNIVERSITÄT BERN

CSH CENTER FOR SPACE AND HABITABILITY

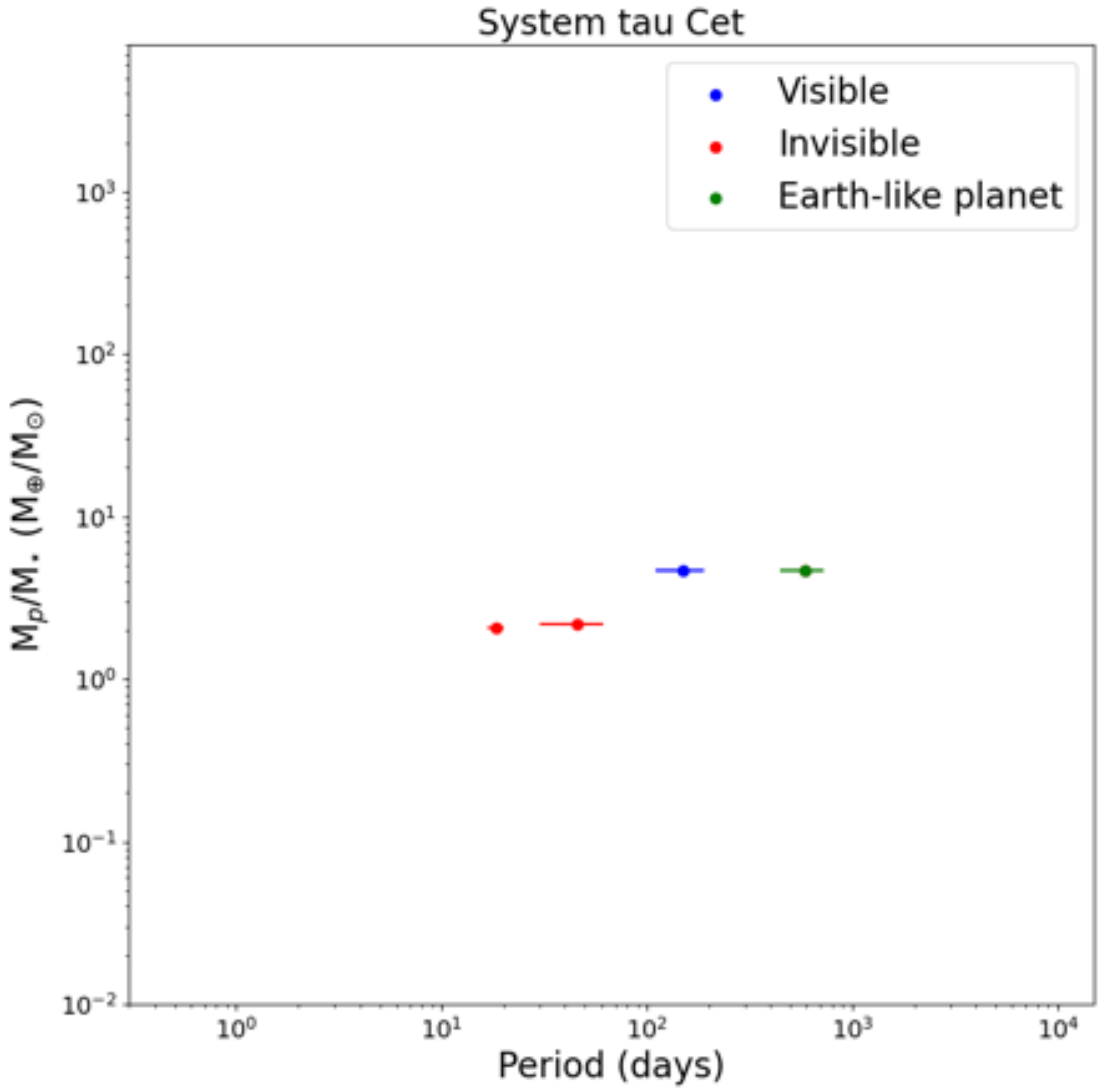
0.94 Msol



0.69 Msol



0.92 Msol



One system has an 'Earth-like' planet (hidden to the classifier), for the two others, observations are ongoing...

Conclusion

UNIVERSITÄT BERN

CSH CENTER FOR SPACE AND HABITABILITY

1- Planetary system formation is a complicated non-linear process

-> simulations rely on HPC computing

-> analysis of simulation results require specific tools

2- Turning observations into planetary composition is degenerate

-> DNN-based tool (trained on simulations) increases dramatically the search efficiency

3- Models can be used to know where to search Earth-twins

-> random forrest classifiers will be used to optimze future observations

UNIVERSITÄT BERN

CSH CENTER FOR SPACE AND HABITABILITY

Question?