

The background of the slide is a vibrant, blue-toned visualization of a cosmic web simulation. It features a complex network of dark blue filaments and clusters, with numerous bright yellow and orange points representing galaxies or star-forming regions. The overall effect is a dynamic and deep-space aesthetic.

# **Data Mining of the MultiDark Simulation**

Adrian Partl  
eAstronomy team, AIP Potsdam  
Astrostatistics and Data Mining Workshop  
2.6.2011, La Palma, Spain

# Collaboration



- The MultiDark Database team:
  - Multimessenger Approach for Dark Matter Detection (MULTIDARK)
    - DB Coordinator: Francisco Prada, IAA-CSIC Spain
  - eAstronomy at the Leibniz-Institut für Astrophysik Potsdam (AIP), Germany
    - Harry Enke, Kristin Riebe, Adrian Partl, Arman Khalatyan
  - Stefan Gottlöber, AIP
  - Anatoly Klypin, NMSU, USA
  - Jaime Forrero, AIP
  - Sebastian Trujillo, NMSU, USA
  - supported by: GAVO, Millennium DB Gerard Lemson, MPA, Germany



# Simulations



<i><math>\Lambda</math>CDM Model</i>	$H_0 = 71 \pm 4 \text{ km s}^{-1}$
	$\Omega_b = 0.29 \pm 0.07$
	$\Omega_c = 0.047 \pm 0.006$
	$\sigma_8 = 0.9 \pm 0.1$
	$t_0 = 13.4 \pm 0.3 \text{ Gyr}$

Cosmological Model

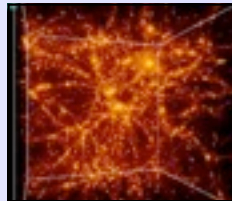
the model:  
initial conditions

# Simulations



$H_0$	$= 71 \pm 4 \text{ km s}^{-1}$
$\Omega_b$	$= 0.29 \pm 0.07$
$\Omega_c$	$= 0.047 \pm 0.006$
$\sigma_8$	$= 0.9 \pm 0.1$
$t_0$	$= 13.4 \pm 0.3 \text{ Gyr}$

Cosmological Model



simulation snapshots



computer cluster

the model:  
initial conditions



run a large  
simulation

pre-analyse

store data

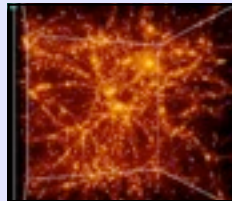
# Simulations



**$\Lambda$ CDM Model**

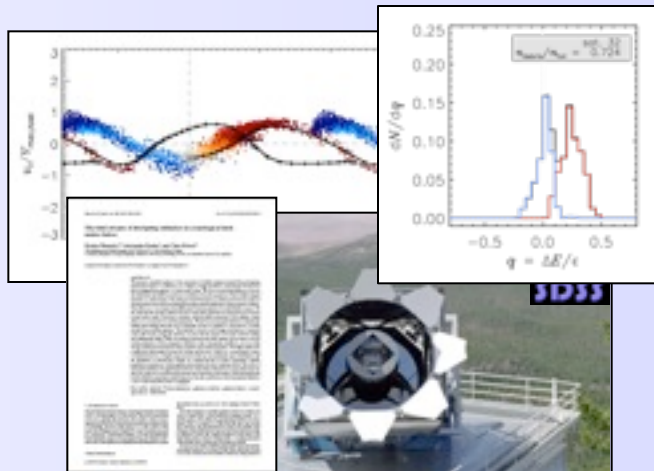
$H_0$	$= 71 \pm 4 \text{ km s}^{-1}$
$\Omega_b$	$= 0.29 \pm 0.07$
$\Omega_c$	$= 0.047 \pm 0.006$
$\sigma_8$	$= 0.9 \pm 0.1$
$t_0$	$= 13.4 \pm 0.3 \text{ Gyr}$

Cosmological Model



simulation snapshots

computer cluster



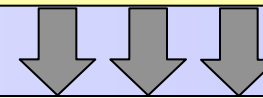
the model:  
initial conditions



run a large  
simulation

pre-analyse

store data



analyse

compare with  
observations

publish results



# Simulations



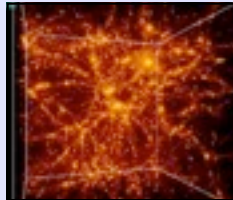
**$\Lambda$ CDM Model**

$H_0$	$= 71 \pm 4 \text{ km s}^{-1}$
$\Omega_b$	$= 0.29 \pm 0.07$
$\Omega_c$	$= 0.047 \pm 0.006$
$\sigma_8$	$= 0.9 \pm 0.1$
$t_0$	$= 13.4 \pm 0.3 \text{ Gyr}$

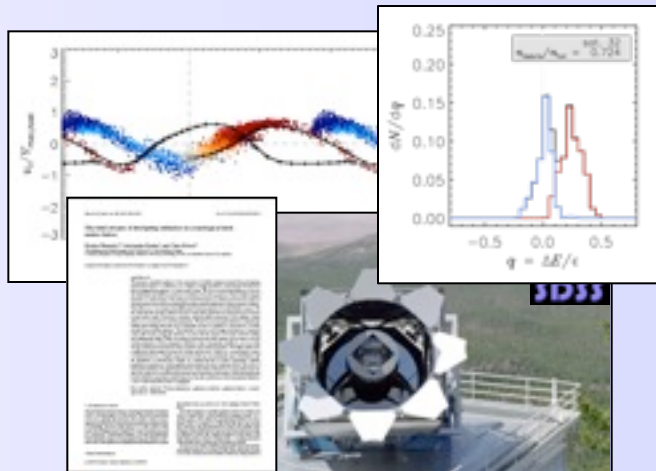
Cosmological Model



simulation snapshots



computer cluster



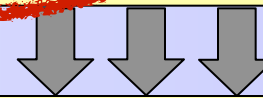
the model:  
initial conditions



run a large  
simulation

pre-analyse

store data



analyse

compare with  
observations

publish results

# Simulations



$H_0$	$= 71 \pm 4 \text{ km s}^{-1}$
$\Omega_m$	$= 0.29 \pm 0.07$
$\Omega_b$	$= 0.047 \pm 0.006$
$\sigma_8$	$= 0.9 \pm 0.1$
$t_0$	$= 13.4 \pm 0.3 \text{ Gyr}$

Cosmological Model

the model:  
initial conditions

run a large  
simulation

pre-analyse

store data

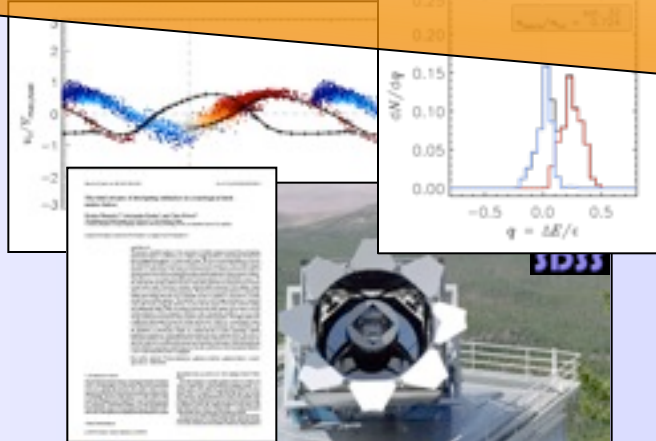
analyse

compare with  
observations

publish results

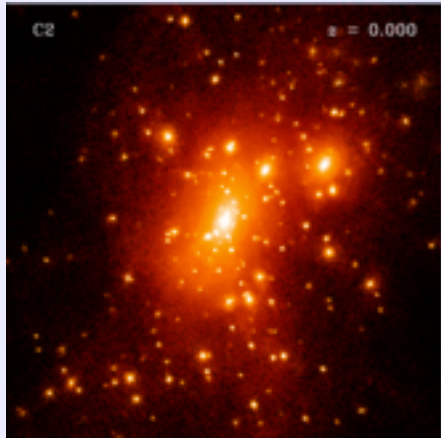


## MultiDark Database



# Pre-analysis: Halo Catalogues

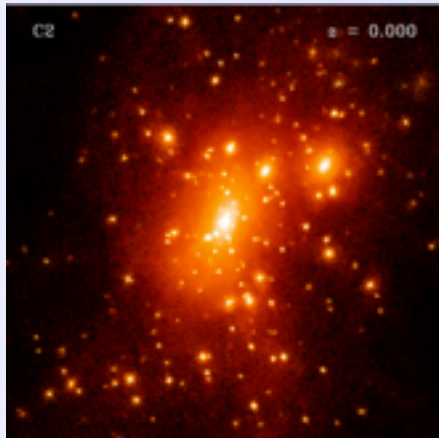
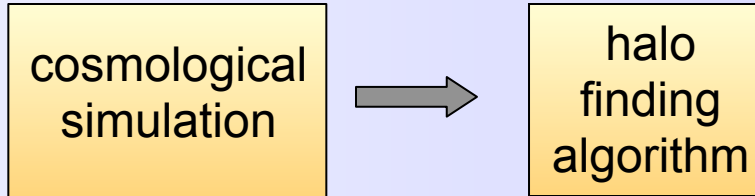
cosmological  
simulation



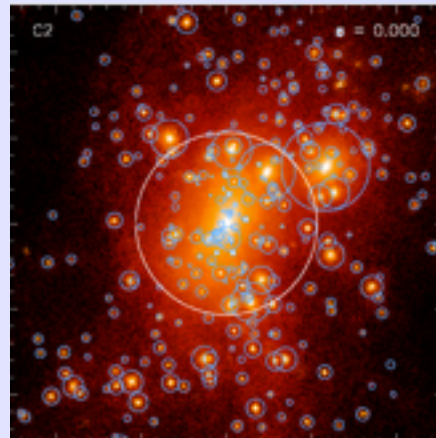
snapshot,  
x-y projection



# Pre-analysis: Halo Catalogues

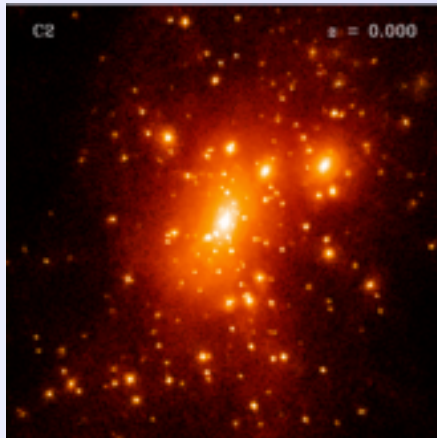


snapshot,  
x-y projection

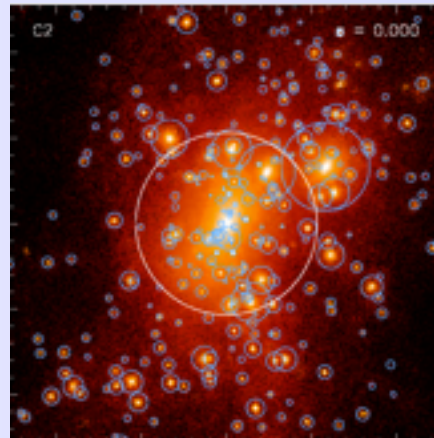


circles mark halos,  
self-bound structures;  
possible locations for  
galaxies

# Pre-analysis: Halo Catalogues



snapshot,  
x-y projection



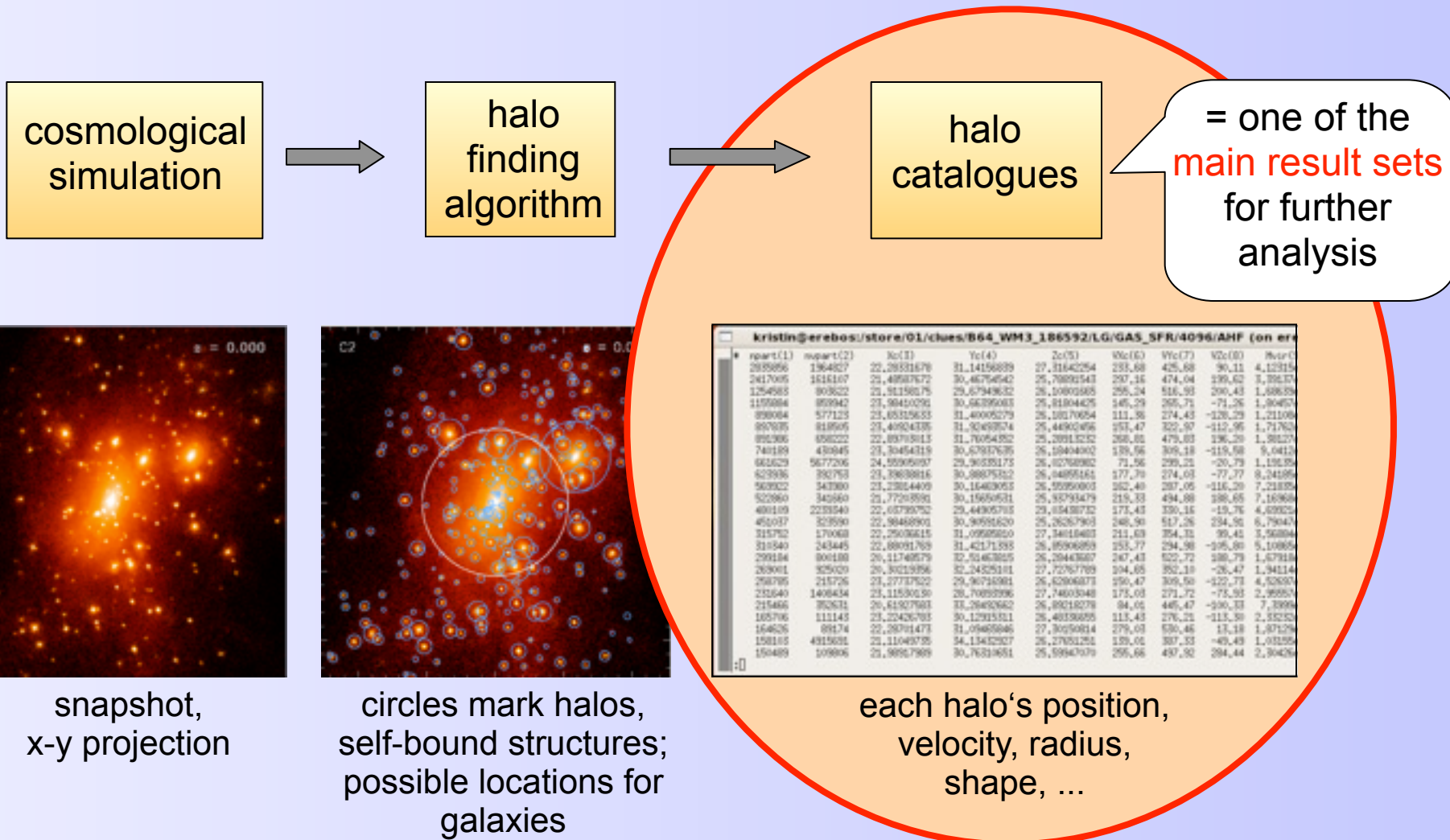
circles mark halos,  
self-bound structures;  
possible locations for  
galaxies

kristin@erebos:/store/01/ches/B64\_WM3\_186592/LG/GAS\_SFR/4096/AHF (on ere

idpart(1)	idpart(2)	Xc(3)	Yc(4)	Zc(5)	Wc(6)	Vc(7)	Wc(8)	Mvir(9)
2858856	1964827	22.28331678	31.14158839	27.31642254	233.68	425.68	90.11	4.12315
2617005	1616107	21.40587672	30.46756542	25.70801543	297.16	474.04	139.62	1.39637
1254988	803622	21.91158175	29.67943632	26.10803585	256.24	516.93	200.43	1.68639
1155864	859942	23.98410291	30.66395603	26.81804425	145.29	265.71	-71.36	1.90457
898064	577123	23.65315633	31.40005279	26.10370654	111.36	274.43	-128.29	1.21108
850705	818505	23.40924335	31.32493574	25.44932456	151.47	322.97	-112.96	1.71762
855206	958222	22.89793613	31.7054352	25.20315252	268.01	479.03	136.29	1.98227
740189	439945	23.30454319	30.67937636	26.10404862	139.56	305.18	-119.58	9.0413
603629	5677206	24.59395987	29.36333173	26.02768882	71.56	230.21	-20.73	1.15635
623936	392763	23.39838816	30.98875312	26.44895161	177.79	274.63	-77.77	8.34185
563922	343800	23.23824499	30.16463953	26.59595803	262.40	287.05	-116.20	7.21835
522860	342660	23.77598591	30.15650531	25.83791479	219.33	494.88	188.65	7.16368
400109	2233540	22.63795752	29.44305703	29.63438732	173.43	330.16	-19.76	4.69923
452037	323590	22.58488901	30.90981620	25.26267963	246.90	517.26	234.50	6.79047
315752	170068	22.35036615	31.09589820	27.34018403	211.63	354.31	99.41	3.56804
318340	243445	22.88091769	31.42171535	26.85906889	253.77	294.98	-105.80	5.10865
293184	800180	26.11740579	32.51463825	26.20443607	267.43	522.72	188.79	1.67918
263001	325029	26.30219356	32.24525181	27.72767789	104.65	352.10	-26.47	1.94114
258705	215726	23.27737522	29.96714581	26.62806873	150.47	309.50	-122.71	4.52857
212540	1408434	23.11530150	28.70893396	27.74603048	175.03	271.72	-73.93	2.95957
215466	782631	26.61927583	33.28432642	26.89218278	94.61	445.47	-89.33	7.3999
183796	111143	23.22426783	30.12333311	26.40336685	111.43	276.22	-113.30	2.53232
164626	89474	22.38791473	31.09465846	27.30750814	279.03	630.46	13.18	1.87129
158185	4915631	22.11049378	34.13432327	26.27851251	139.01	387.33	-49.49	1.93195
150489	109806	21.98967989	30.76318651	25.59943079	255.66	437.92	284.44	2.30426

each halo's position,  
velocity, radius,  
shape, ...

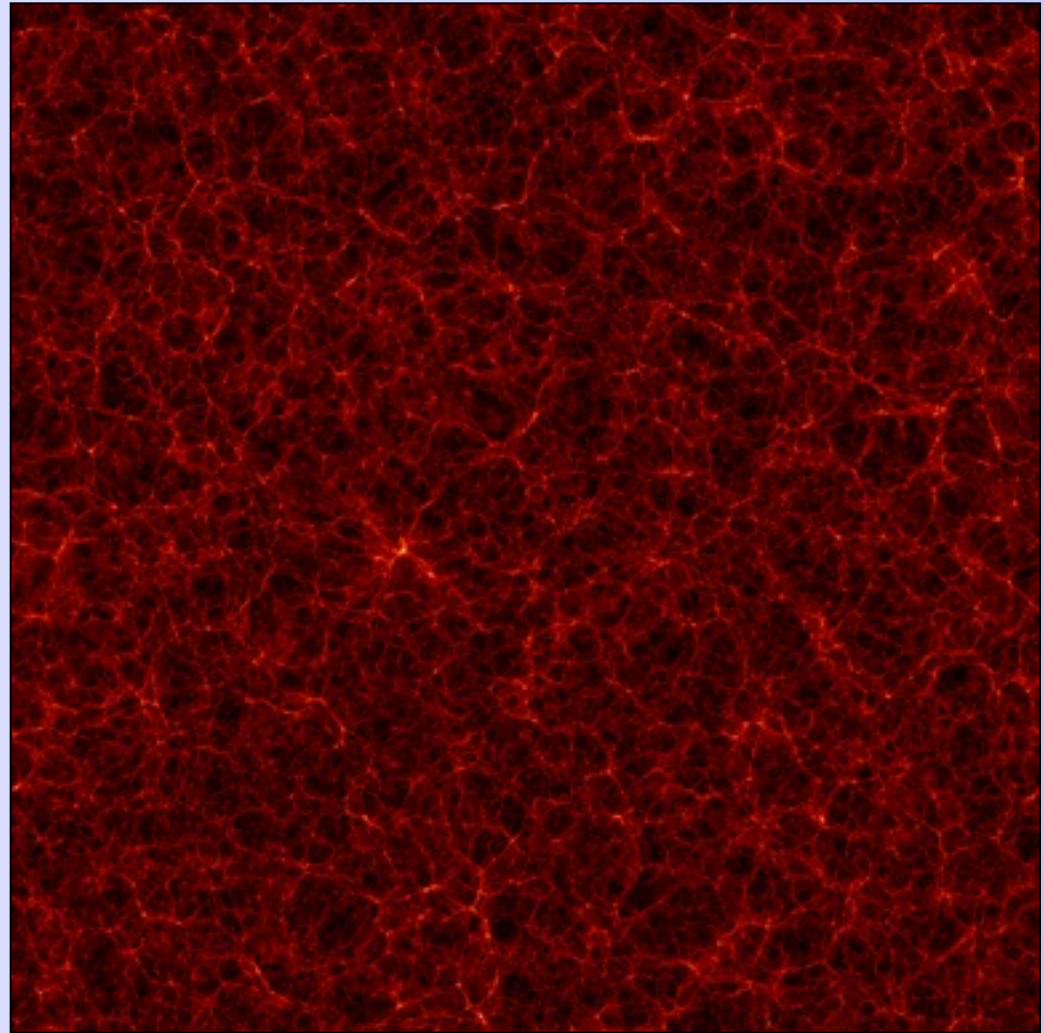
# Pre-analysis: Halo Catalogues



# MultiDark Simulations in DR1

- WMAP5 cosmology
- MultiDark Run 1 (MDR1)
  - $(1 \text{ Gpc/h})^3$  box
  - $2048^3$  particles  
(~ 8.6 billion)
  - mass resolution  
 $8.7 \cdot 10^9 M_{\text{sun}}/\text{h}$

box larger than  
Millennium simulation



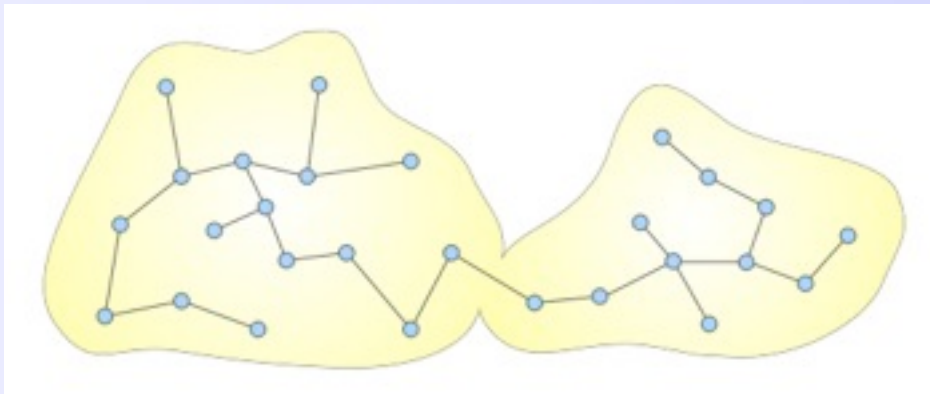
MDR1, 1 Gpc/h



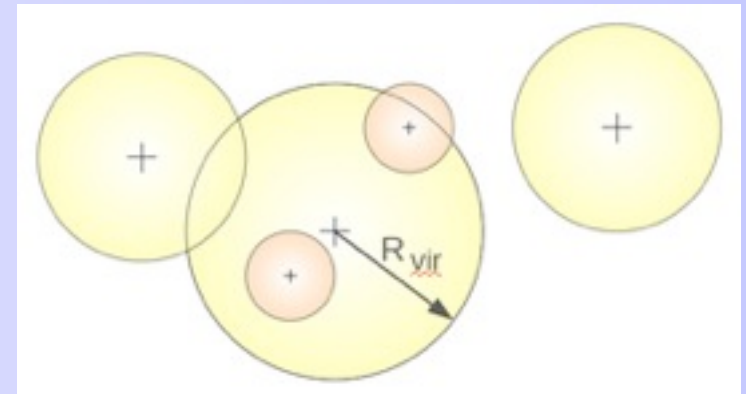
# MultiDark Simulation Database

- 43 snapshots between  $10.34 < z < 0.0$
- Halo catalogs:
  - BDM halos: for 2 density criteria: V:  $360\rho_{\text{back}}$  W:  $200\rho_{\text{crit}}$
  - FOF halos: 6 different linking lengths (0.2, 0.17, 0.085, ...)
- Halo profiles:
  - BDM halos with more than 100 particles
  - Profiles cover up to  $2 R_{\text{vir}}$

FOF

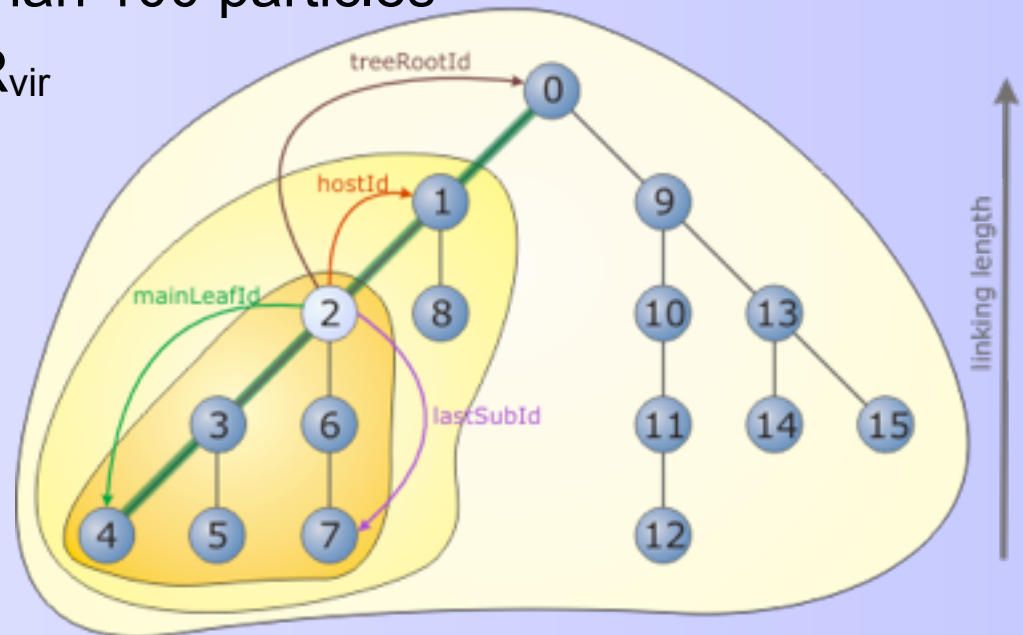


BDM

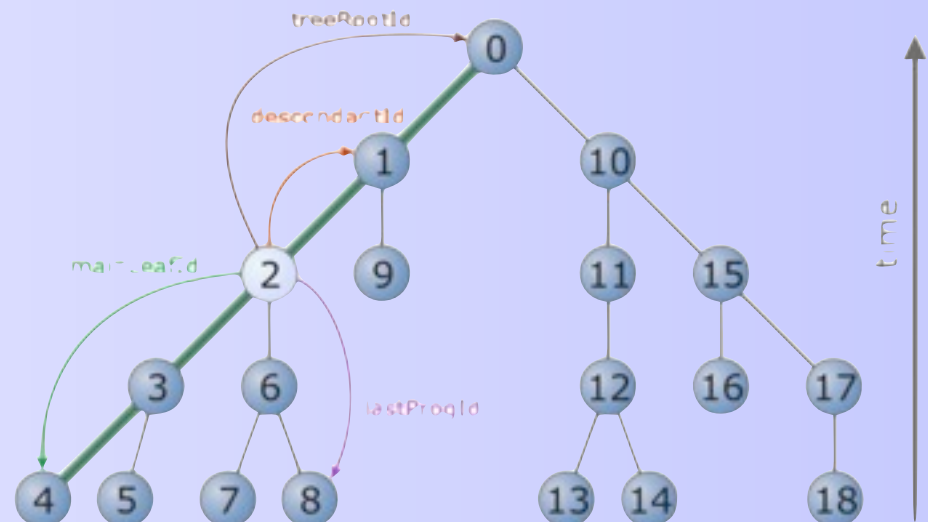




- 43 snapshots between  $10.34 < z < 0.0$
- Halo catalogs:
  - BDM halos: for 2 density criteria: V:  $360\rho_{\text{back}}$  W:  $200\rho_{\text{crit}}$
  - FOF halos: 6 different linking lengths (0.2, 0.17, 0.085, ...)
- Halo profiles:
  - BDM halos with more than 100 particles
  - Profiles cover up to  $2 R_{\text{vir}}$
- FOF substructure tree

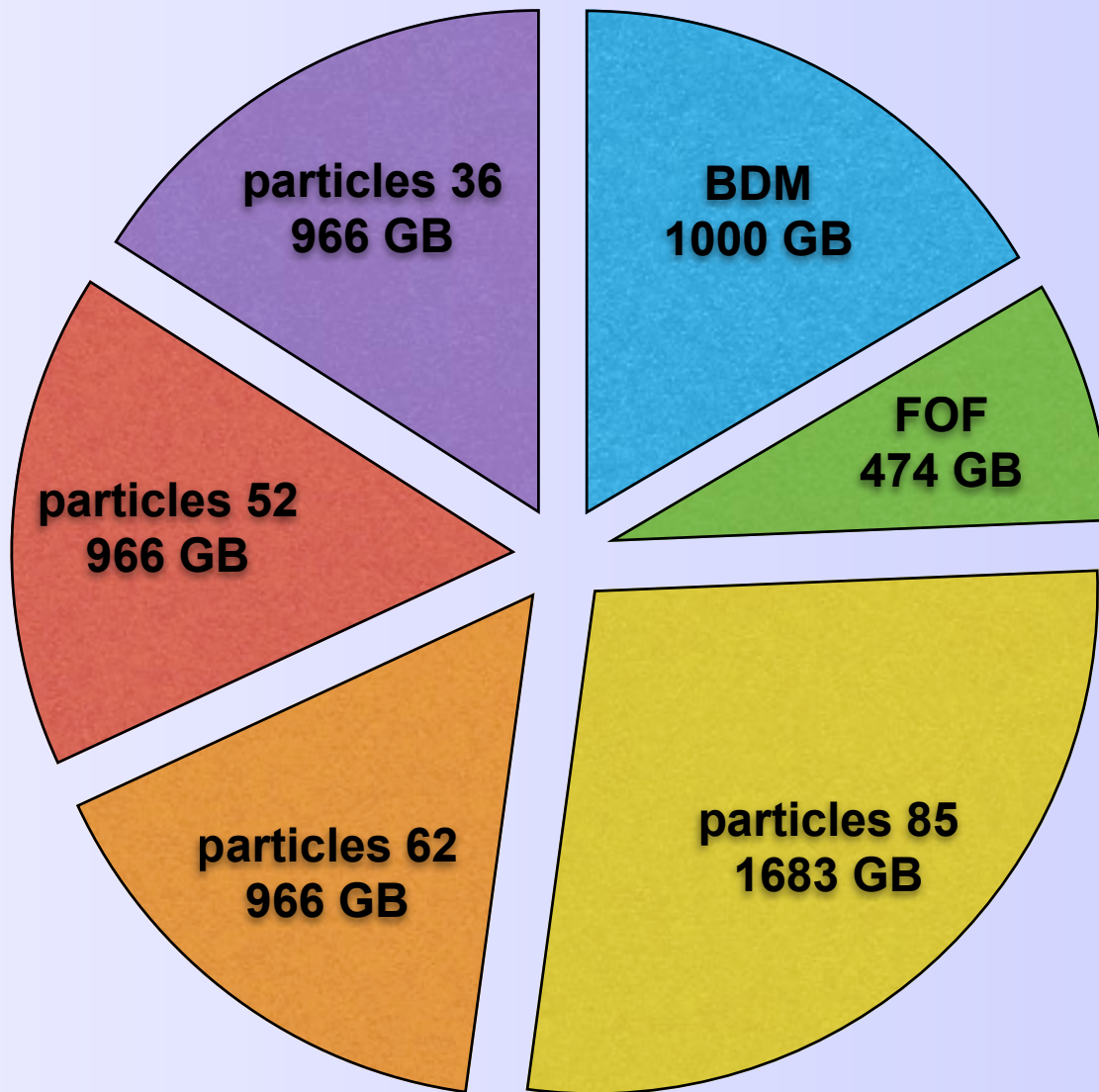


- 43 snapshots between  $10.34 < z < 0.0$
- Halo catalogs:
  - BDM halos: for 2 density criteria: V:  $360\rho_{\text{back}}$  W:  $200\rho_{\text{crit}}$
  - FOF halos: 6 different linking lengths (0.2, 0.17, 0.085, ...)
- Halo profiles:
  - BDM halos with more than 100 particles
  - Profiles cover up to  $2 R_{\text{vir}}$
- FOF substructure tree
- FOF merger tree



- 43 snapshots between  $10.34 < z < 0.0$
- Halo catalogs:
  - BDM halos: for 2 density criteria: V:  $360\rho_{\text{back}}$  W:  $200\rho_{\text{crit}}$
  - FOF halos: 6 different linking lengths (0.2, 0.17, 0.085, ...)
- Halo profiles:
  - BDM halos with more than 100 particles
  - Profiles cover up to  $2 R_{\text{vir}}$
- FOF substructure tree
- FOF merger tree
- Full particle data of 4 snapshots:
  - $z = 0, 0.5, 1, 2.9$  online
  - more snapshots in future data releases
- Link FOF halos at  $z=0$  with particle information

# Amount of Data



Total data size:  
6.05 TB  
Total row count:  
 $4.87 \cdot 10^{10}$

# Web Interface to Database

- interactive access through web interface: [www.multidark.org](http://www.multidark.org)
  - access to data via standard SQL queries
  - history of previous queries; demo queries
- scripted access for IDL, R, Topcat: [wget.multidark.org/MyDB](http://wget.multidark.org/MyDB)  
private DB space (registered users)
- extensive documentation
- extensive collection of useful queries
- support by AIP eAstronomy team
  - Kristin Riebe
  - Adrian Partl



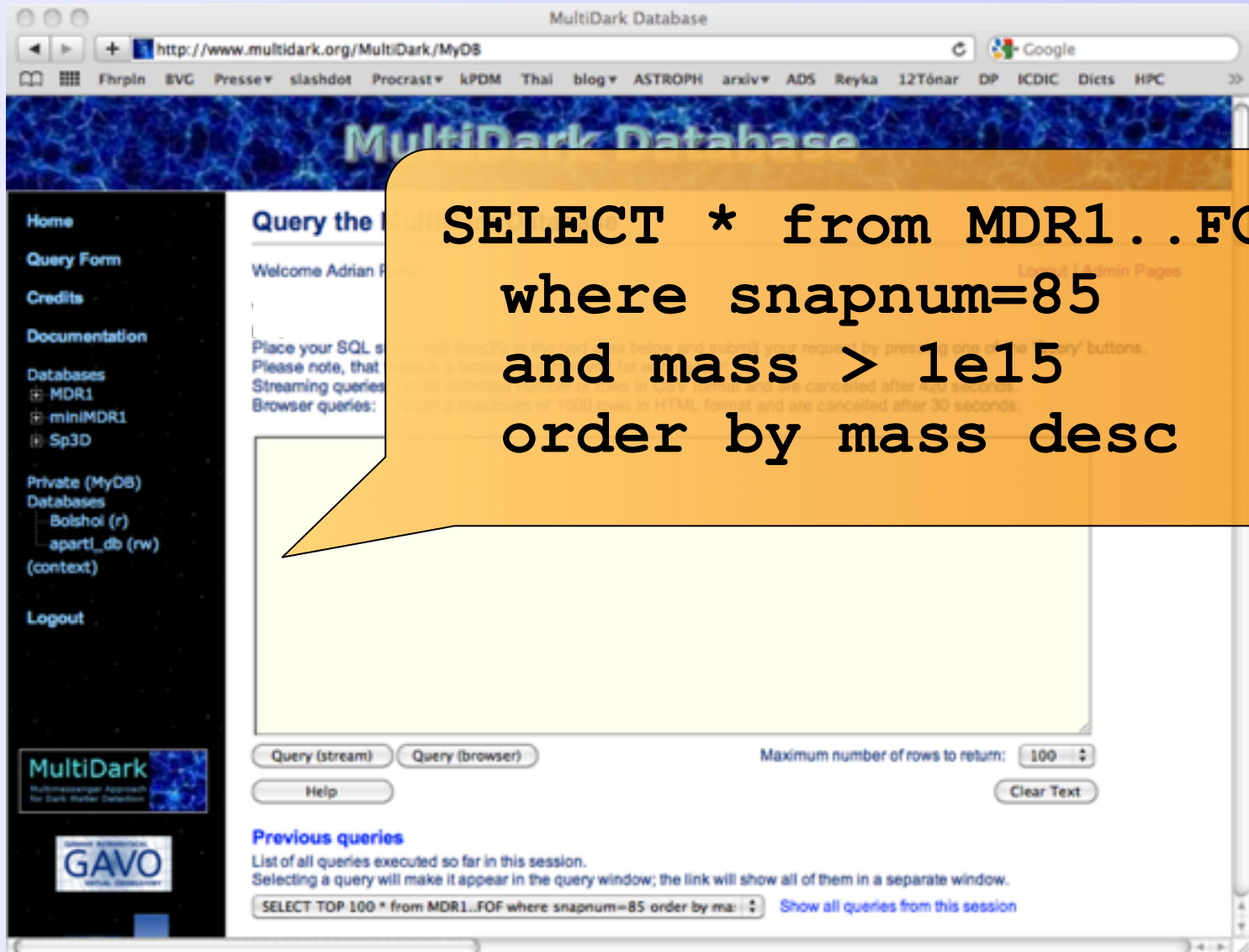


# Determine most massive halos



The screenshot shows the MultiDark Database web interface in a browser window. The address bar shows the URL `http://www.multidark.org/MultiDark/MyDB`. The page has a dark blue header with the "MultiDark Database" title. A left sidebar contains navigation links: Home, Query Form, Credits, Documentation, Databases (MDR1, miniMDR1, Sp3D), Private (MyDB) Databases (Bolshoi (r), apartl\_db (rw), context), and Logout. The main content area is titled "Query the MultiDark Database" and includes a welcome message "Welcome Adrian Partl" with a "Logout | Admin Pages" link. It provides instructions on using SQL queries and includes a large text input area. Below the input area are buttons for "Query (stream)", "Query (browser)", and "Help", along with a "Maximum number of rows to return" dropdown set to 100 and a "Clear Text" button. A "Previous queries" section lists executed queries, showing a query: `SELECT TOP 100 * from MDR1..FOF where snapnum=85 order by ma:` with a link to "Show all queries from this session".

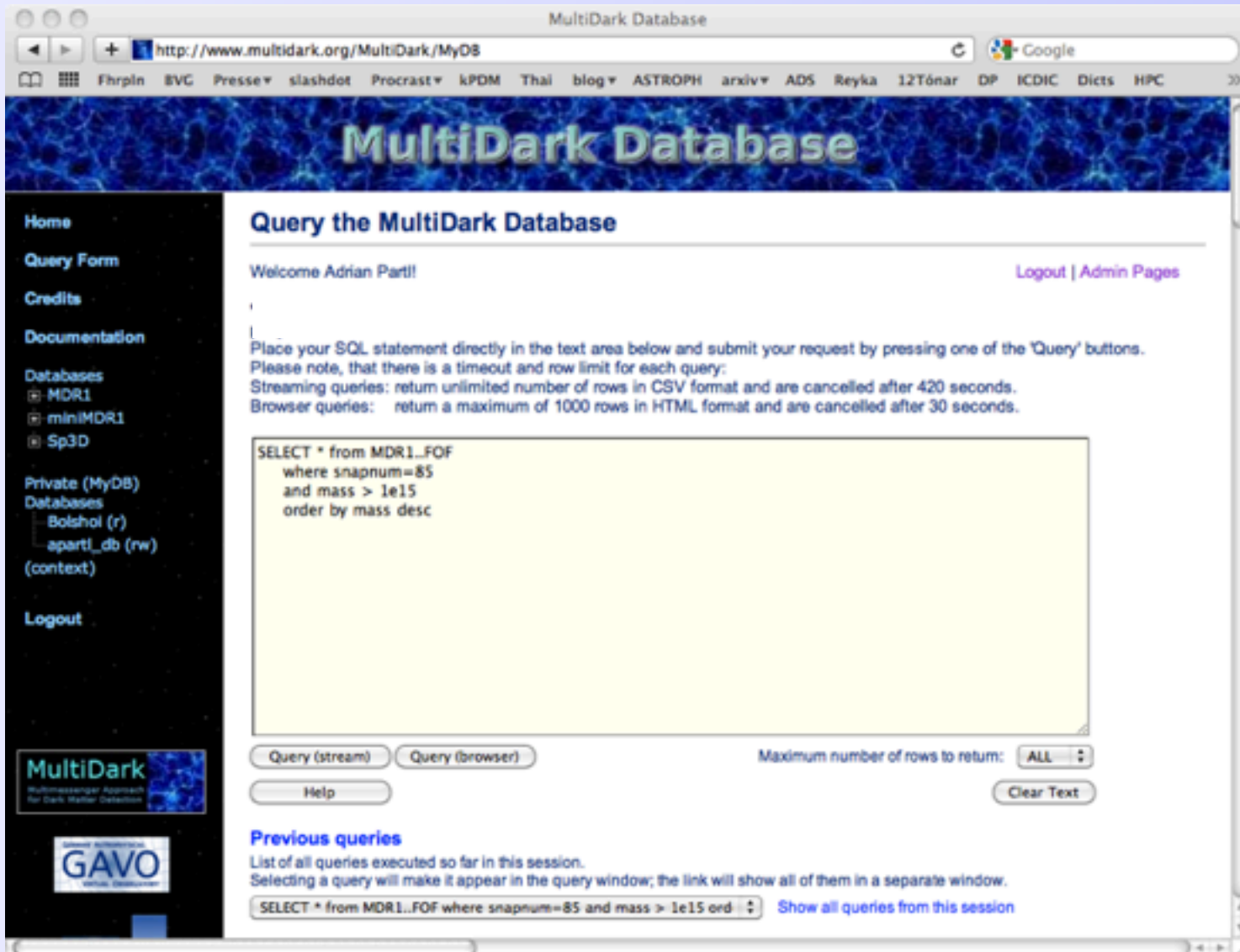
# Determine most massive halos



The screenshot shows the MultiDark Database web interface. A large orange speech bubble overlay contains the SQL query: `SELECT * from MDR1..FOF where snapnum=85 and mass > 1e15 order by mass desc`. The interface includes a navigation menu on the left with links like Home, Query Form, Credits, Documentation, Databases (MDR1, miniMDR1, Sp3D), Private (MyDB) Databases (Bolshoi (r), apartl\_db (nw), context), and Logout. The main content area has a 'Query the' section with a welcome message and a large text input field for SQL queries. Below the input field are buttons for 'Query (stream)', 'Query (browser)', and 'Help'. A 'Maximum number of rows to return' dropdown is set to 100, with a 'Clear Text' button. The 'Previous queries' section shows a list of queries executed, with the first one being 'SELECT TOP 100 \* from MDR1..FOF where snapnum=85 order by ma:'. A link 'Show all queries from this session' is also present.

```
SELECT * from MDR1..FOF
where snapnum=85
and mass > 1e15
order by mass desc
```

# Determine most massive halos



The screenshot shows the MultiDark Database web interface. The browser address bar displays `http://www.multidark.org/MultiDark/MyDB`. The page title is "MultiDark Database". The left sidebar contains navigation links: Home, Query Form, Credits, Documentation, Databases (MDR1, miniMDR1, Sp3D), Private (MyDB) Databases (Bolshoi (r), apartl\_db (rw) (context)), and Logout. The main content area is titled "Query the MultiDark Database" and includes a welcome message for Adrian Partl, a "Logout | Admin Pages" link, and instructions for using the query form. A text area contains the SQL query: `SELECT * from MDR1..FOF where snapnum=85 and mass > 1e15 order by mass desc`. Below the text area are buttons for "Query (stream)", "Query (browser)", and "Help". A dropdown menu for "Maximum number of rows to return:" is set to "ALL". A "Clear Text" button is also present. The "Previous queries" section shows a list of queries executed in the session, with a link to "Show all queries from this session".

# Determine most massive halos

MultiDark Database

http://www.multidark.org/MultiDark/MyDB

Create View Drop Table Create Index

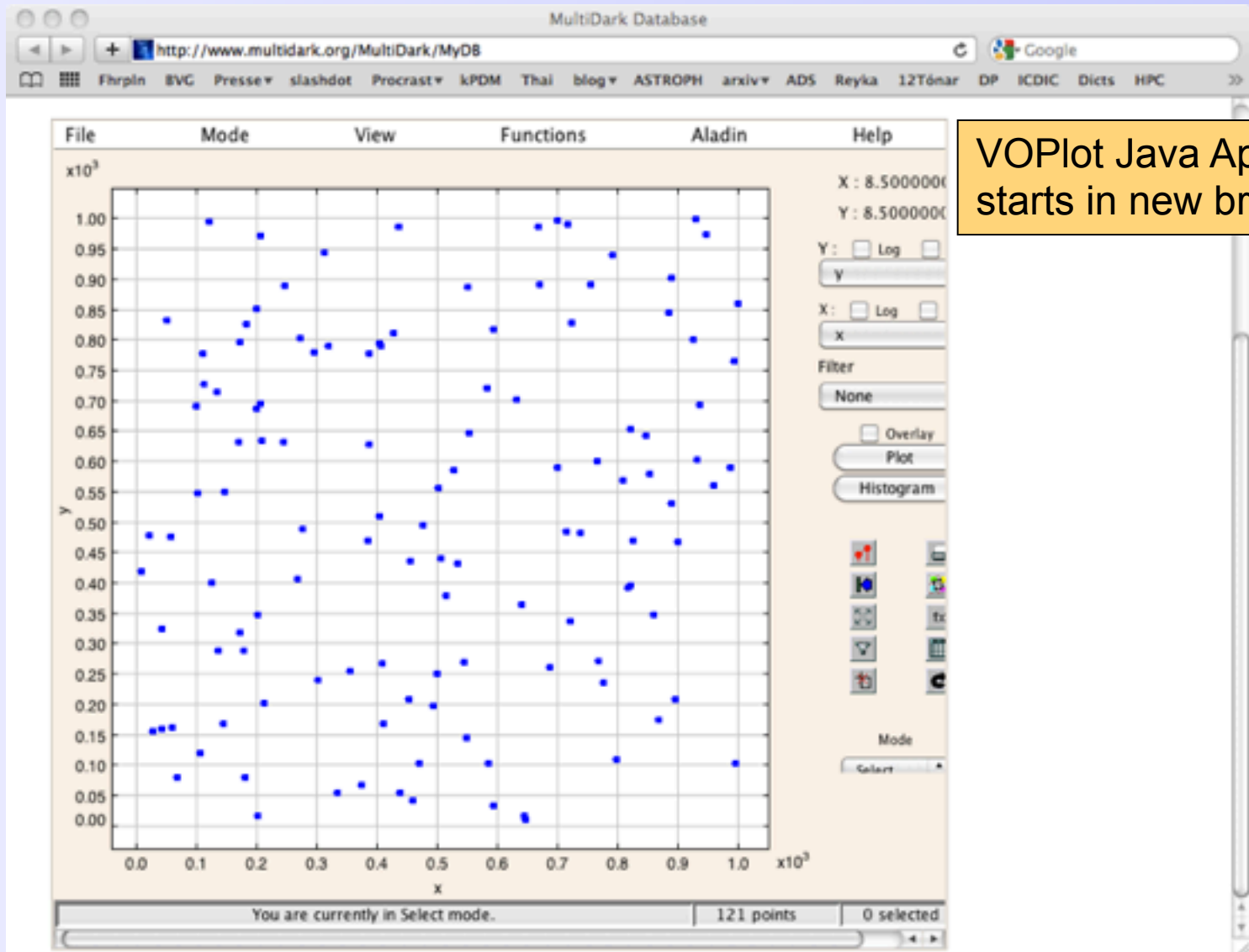
Query time (in millisec) = 2640  
Number of rows retrieved from database = 121

Select an output format and press the "Get table" button to open a new window with your data in the chosen table format.  
Click "Plot" to start the VOPlot Java Applet in your browser.

CSV Get table Plot

fofid	snapnum	level	NinFile	x	y	z	vx	vy	vz	np	mass	size	disp	disp_v	delta
85000000000	85	0	0	384.7144	468.4821	873.0618	-115.12	-294.5	87.51	310057	2.70400699E15	2.5719	2.606	2000.0	505.5
85000000001	85	0	1	548.8059	144.013	895.2401	-113.73	167.99	458.66	302086	2.63449187E15	2.3652	2.014	2200.0	633.5
85000000002	85	0	2	946.1964	973.1105	268.3232	-81.65	-38.23	-116.38	296022	2.58160767E15	2.4821	2.601	1900.0	537.0
85000000003	85	0	3	895.8632	208.5117	112.89	318.04	-44.99	-0.29	278745	2.43093538E15	2.6043	3.748	1700.0	437.6
85000000004	85	0	4	544.7999	269.1872	124.5663	-249.74	209.61	-354.03	240991	2.10168252E15	2.4074	3.2	1700.0	479.0
85000000005	85	0	5	593.6349	34.1149	299.8866	216.7	-45.37	116.06	233073	2.03262965E15	2.2214	2.3	1900.0	589.9
85000000006	85	0	6	666.5557	986.1232	816.0831	-233.39	-110.96	-306.8	231450	2.01847545E15	2.3544	2.507	1900.0	491.8
85000000007	85	0	7	514.3323	379.8545	887.7273	-65.69	-60.15	80.62	228979	1.99692585E15	2.0904	1.756	2000.0	695.7
85000000008	85	0	8	699.2684	590.2208	941.3842	254.51	-239.83	-155.44	222498	1.94040502E15	2.2641	1.748	1900.0	531.8
85000000009	85	0	9	737.6971	482.4798	972.2432	116.68	40.94	-253.1	212893	1.85663987E15	1.9481	1.249	2200.0	799.3
85000000010	85	0	10	714.5986	483.9961	516.4453	238.65	129.22	474.56	208725	1.82029076E15	1.9228	1.188	2300.0	815.0
85000000011	85	0	11	110.2769	776.6714	398.6211	-175.38	-239.79	20.37	204627	1.78455207E15	2.3503	1.746	1900.0	437.0
85000000012	85	0	12	271.8358	803.4151	142.9995	375.99	45.81	259.78	202339	1.76459832E15	2.1327	1.538	2100.0	578.7
85000000013	85	0	13	276.5656	488.5082	548.7596	232.59	45.74	-240.68	199983	1.74405173E15	2.0452	1.357	2100.0	648.7
85000000014	85	0	14	120.8181	994.9	482.4189	178.65	180.07	-117.96	193788	1.69002507E15	2.0936	1.41	1900.0	585.9
85000000015	85	0	15	889.7112	901.3621	201.6839	-171.24	16.19	-48.88	191968	1.67415288E15	2.0757	1.383	2000.0	595.6
85000000016	85	0	16	643.7282	16.3073	310.9569	-258.13	-82.61	68.22	191705	1.67185924E15	2.1691	1.59	1900.0	521.1
85000000017	85	0	17	475.6505	493.9411	453.5785	-363.38	-194.31	58.2	185131	1.61452746E15	2.0351	1.268	2300.0	609.5

# Quick check: Plot distribution



VOPlot Java Applet starts in new browser tab



# Limitations and Challenges

# Limitations and Challenges

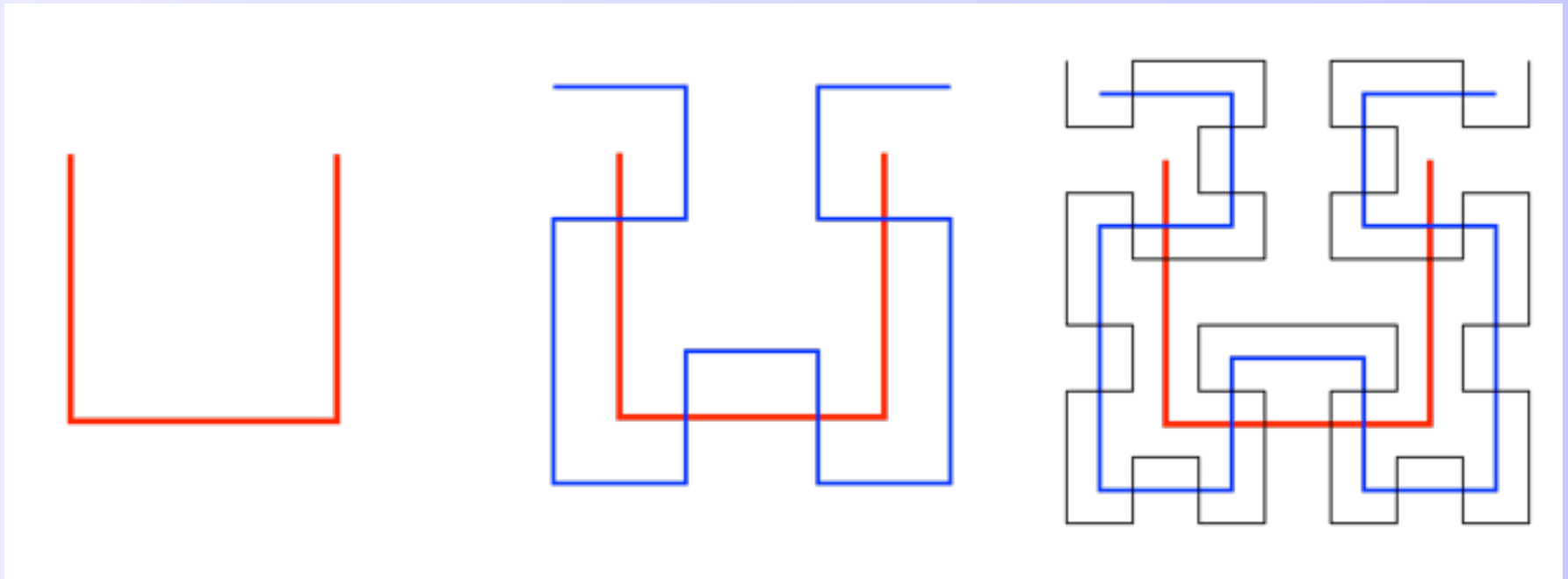
- User side:
  - Complicated queries with joins can take hours:  
We cannot create an index for every problem one wants to solve.
  - Spatial queries are only feasible in simple cases  
e.g. cone search / neighbour search
  - Access to particle data rather slow

# Limitations and Challenges

- User side:
  - Complicated queries with joins can take hours:  
We cannot create an index for every problem one wants to solve.
  - Spatial queries are only feasible in simple cases  
e.g. cone search / neighbour search
  - Access to particle data rather slow
- Server side:
  - Long ingest times
  - Load balancing
  - Performance limitations of commercial DBs reached

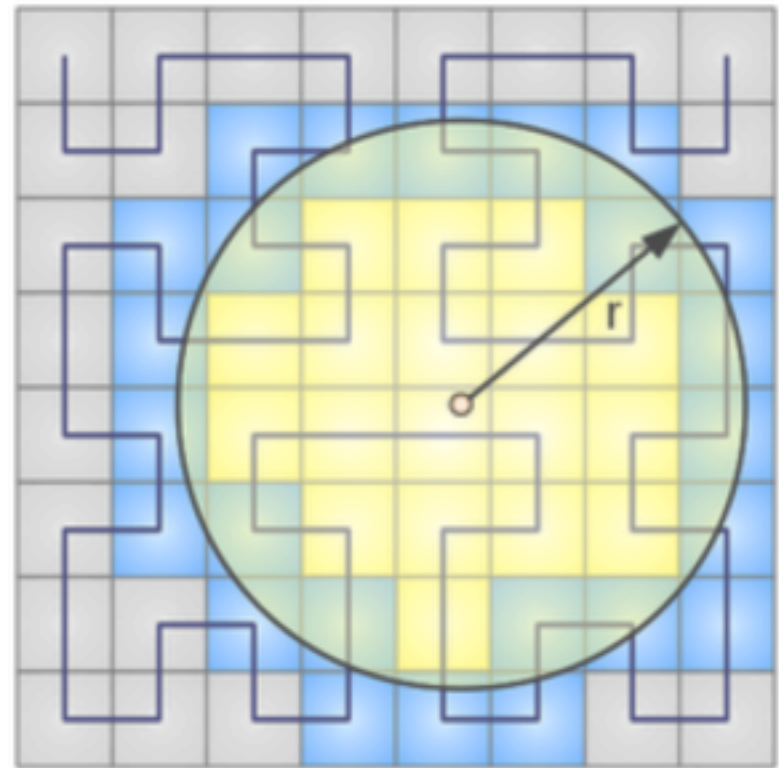
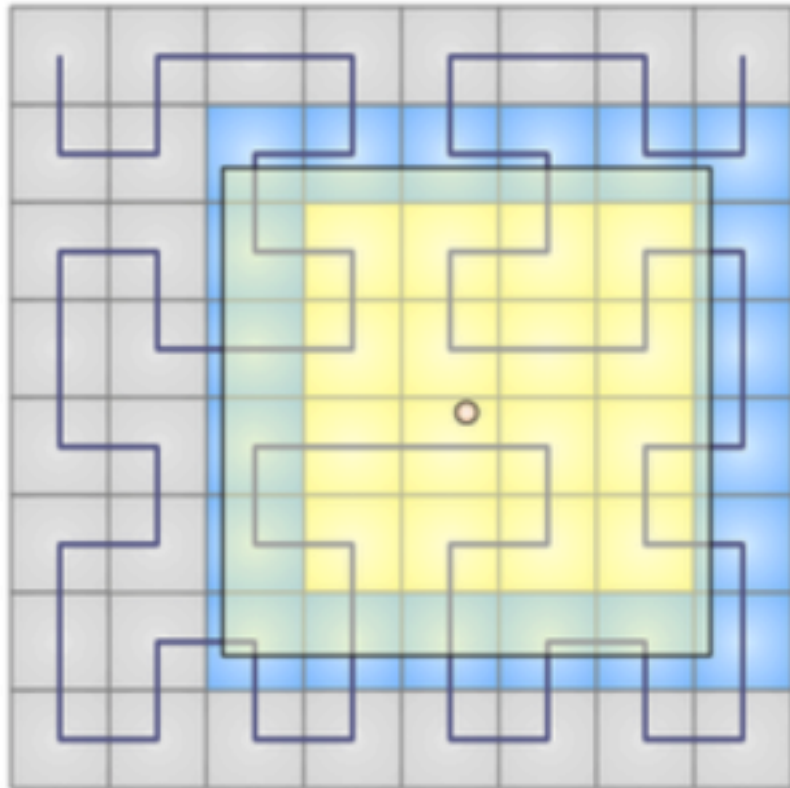
# Spatial queries - current implementation

- Millenium DB Spatial3D library
  - Based on Peano-Hilbert Curve
- Mapping 3D to 1D
  - 1 index lookup faster than a lookup in 3 indexes



Wikipedia

# Spatial queries - current implementation





# Possible solution

DBMS

Plugin architecture

data access

spatial index

other index

GADGET / ART / ... Simulation files

**Harvest the results of the MultiDark  
simulations!**

**Register today!**

**<http://www.multidark.org>**

**Support: [apartl@aip.de](mailto:apartl@aip.de)**