Latest pgSphere developments

Markus Nullmeier

Zentrum für Astronomie der Universität Heidelberg
Astronomisches Rechen-Institut

`mnullmei@ari.uni.heidelberg.de`
`https://github.com/mnullmei`

pgSphere: infrastucture for the VO

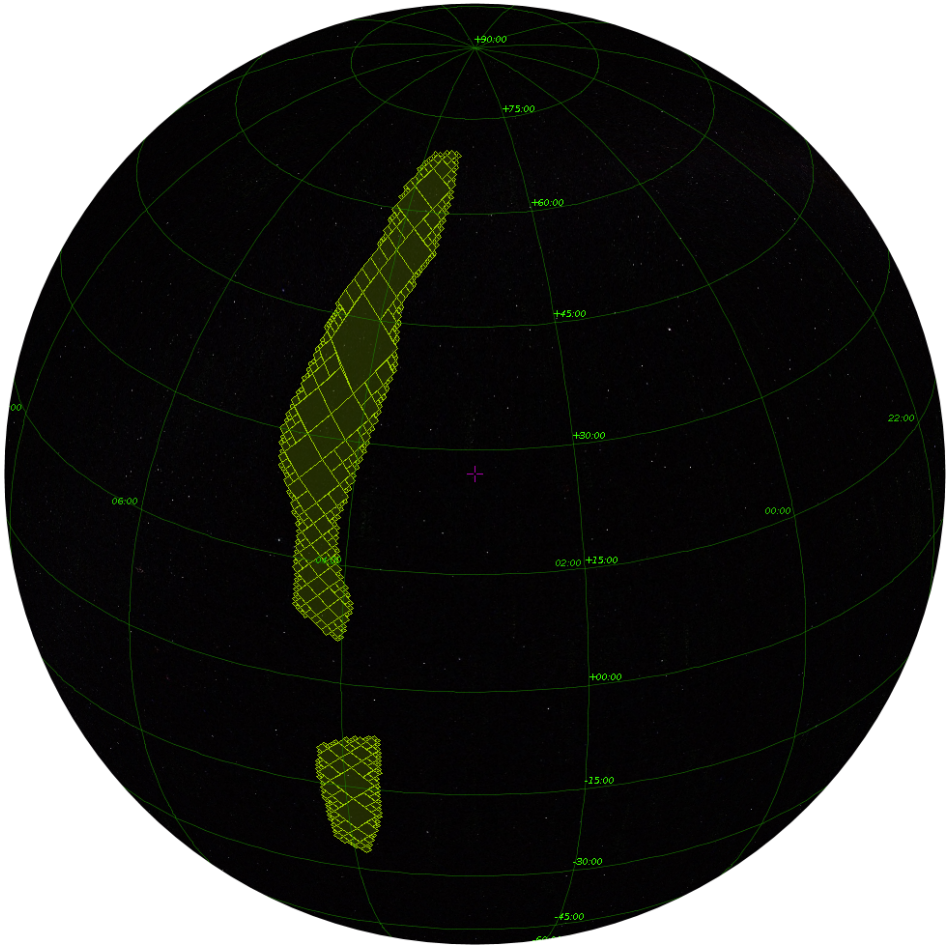
- Many VO data centres use the PostgreSQL RDBMS
  - VizieR, Simbad at CDS
  - ESAC at ESO
  - CADC
  - GAVO at ARI
  - several others, plus centres that are migrating ...
- pgSphere useful for
  - Custom PostgreSQL spherical data types
  - Indexing (fast queries)

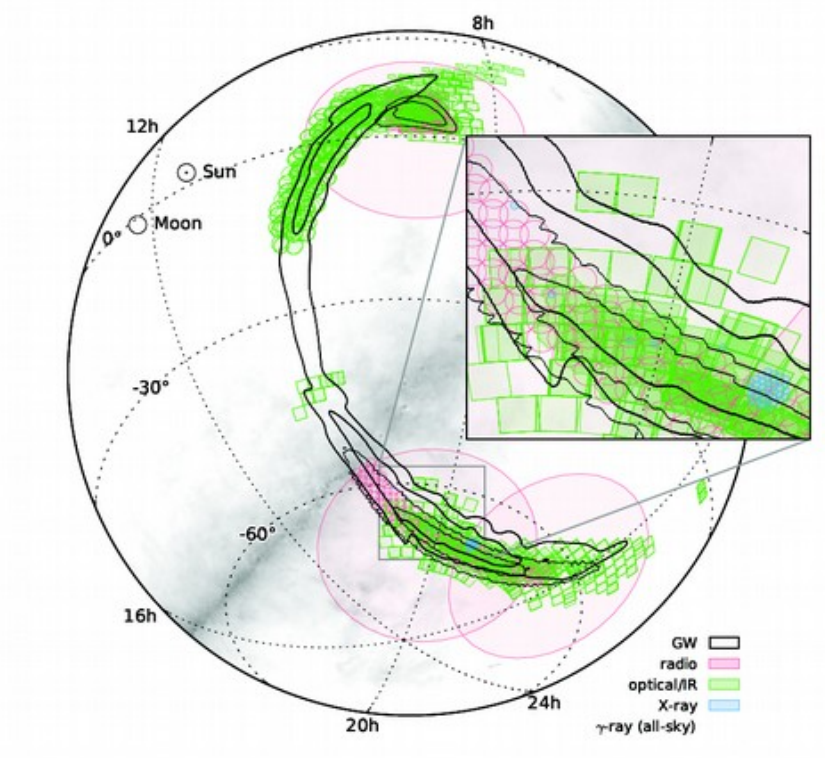
pgSphere update: overview

- New features and WIP
  - MOCs
  - OUZO: indexing for MOCs
  - BRIN indexing
  - Efficient crossmatch
  - PgSphere packages for Linux distributions
  - Official release
- Future projects for pgSphere
  - Integration with the JIT acceleration of PostgreSQL 11
  - Faster indexing in 2D (now: 3D)
  - Optimal read-only indexing (maybe GSOC)

MOCs: since 2017 / & WIP

To describe arbitrary sky regions such as those of gravitational wave events, we need something else:

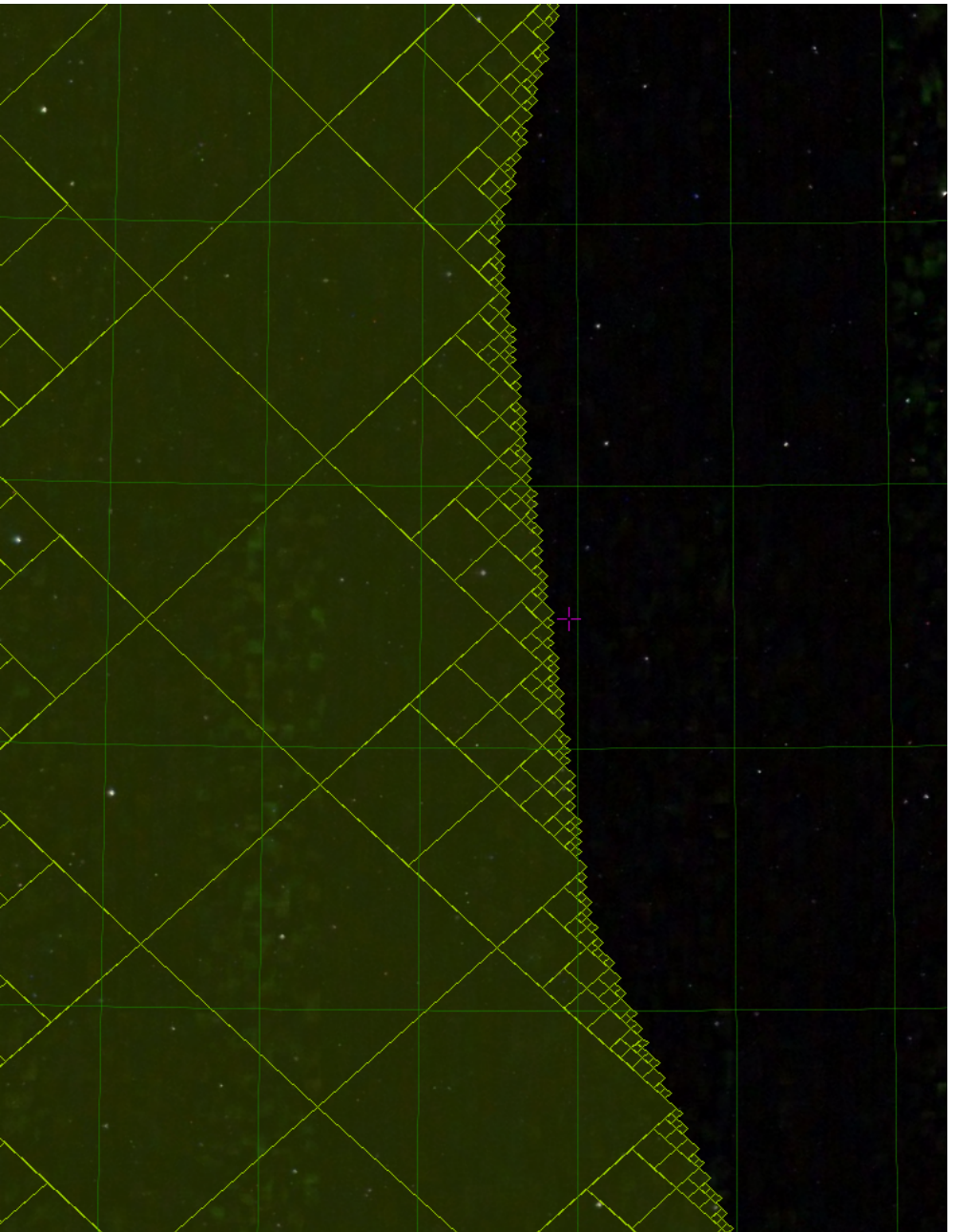




MOC: Multi-Order Coverage

= set of Healpix sphere elements of different orders

MOC internals



MOC: Multi-Order Coverage

= set of Healpix sphere elements (diamond-shaped) of different orders

1 diamond element
= 1 integer interval

1 MOC object
= 1 list of intervals

```
{[2, 6) [17, 30) [33, 40)
[123, 124) [332, 438), ...}
```

OUZO: indexing for MOCs

- **“inverted index” for the constituent intervals**
- **Thanks to https://github.com/postgrespro/rum**

sorted intervals	sets of pointers to MOCs
[17, 30)	{ moc7, moc11 }
[843, 2577)	{ moc2, **moc108**, moc109 }
[5756, 9433)	{ **moc108**, moc, moc1103 }
…	…

BRIN indexing

- Small indexes for big tables
- Originally a PostGIS project
- Thanks to Guiseppe Broccolo

Efficient crossmatch

- Problems with crossmatch (spatial join)
  - Everybody translates ADQL to SQL
  - Then, only the index of one table is used
  - … more often than not, the wrong index
- Solutions
  - Use both indexes at the same time (WIP, Alexander Korotkov)
  - Custom spatial joins for pgSphere (Dmitry Ivanov, crossmatch-cnode branch)

pgSphere packages for Linux distributions

- Why?
  - Saves DBA precious time
  - Fewer bugs for users
  - Free quality assurance!
  - Prerequisite for data centre tools for everybody...
- Which ones?
  - Debian → Ubuntu, PGDG: thanks to Ole Streicher, Christoph Berg
  - Fedora → CentOS, etc., PGDG: thanks to Christian Dersch

WIP: official pgSphere release

- Thorny problem: recover from old bugs in existing installations
  - Old PostgreSQL syntax (thanks: Pat Dowler, Alexander Korotkov)
  - Incomplete system tables (thanks: Markus Demleitner)
  - (inevitable) proliferation and use of development code
- HEALPix problems
  - Rather unportable official library
  - Official library has got wrong licence (GPL, not LGPL)
  - Changing to BSD-licenced library to have MOC here

Integration with the JIT acceleration of PostgreSQL 11

- Moore's Law is dead, everybody does JIT nowadays
  - Finally, also in PostgreSQL
  - pgSphere needs to adapt
  - Funding??
- More options for pgSphere (read: ADQL) speedups
  - Parallel queries
  - Database clustering (PostgreSQL WIP)

Faster indexing in 2D (now: 3D)

- **Announced at ADASS 2106**
  - **Simple idea, but:**
  - **Devil in the details**
  - **Potential synergy with GIS community**

Optimal read-only indexing

- VO use case is mosty read-only
  - Potential of huge ADQL speedups
  - Hope for GSOC student, thanky to Andrey Borodin

pgSphere update: overview

- New features and WIP
  - MOCs
  - OUZO: indexing for MOCs
  - BRIN indexing
  - Efficient crossmatch
  - PgSphere packages for Linux distributions
  - Official release
- Future projects for pgSphere
  - Integration with the JIT acceleration of PostgreSQL 11
  - Faster indexing in 2D (now: 3D)
  - Optimal read-only indexing (maybe GSOC)