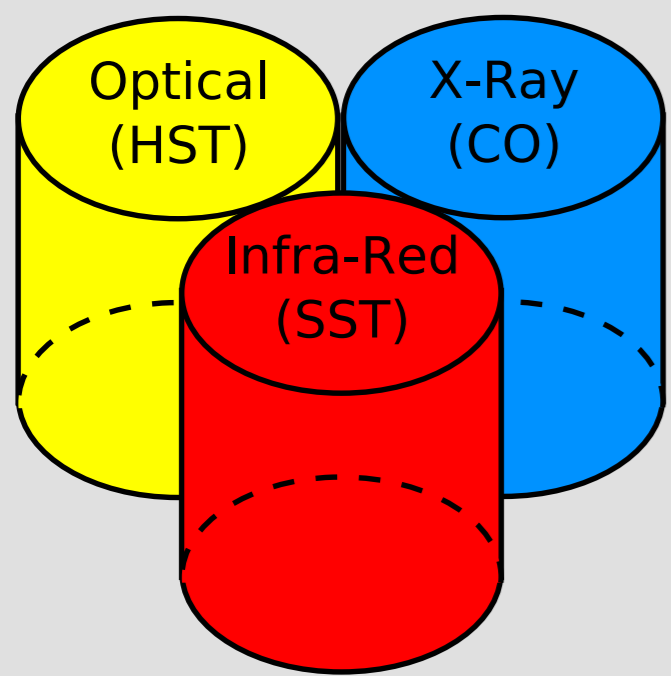


Histogram-based P2P Main Memory Database for Locality-Aware Data

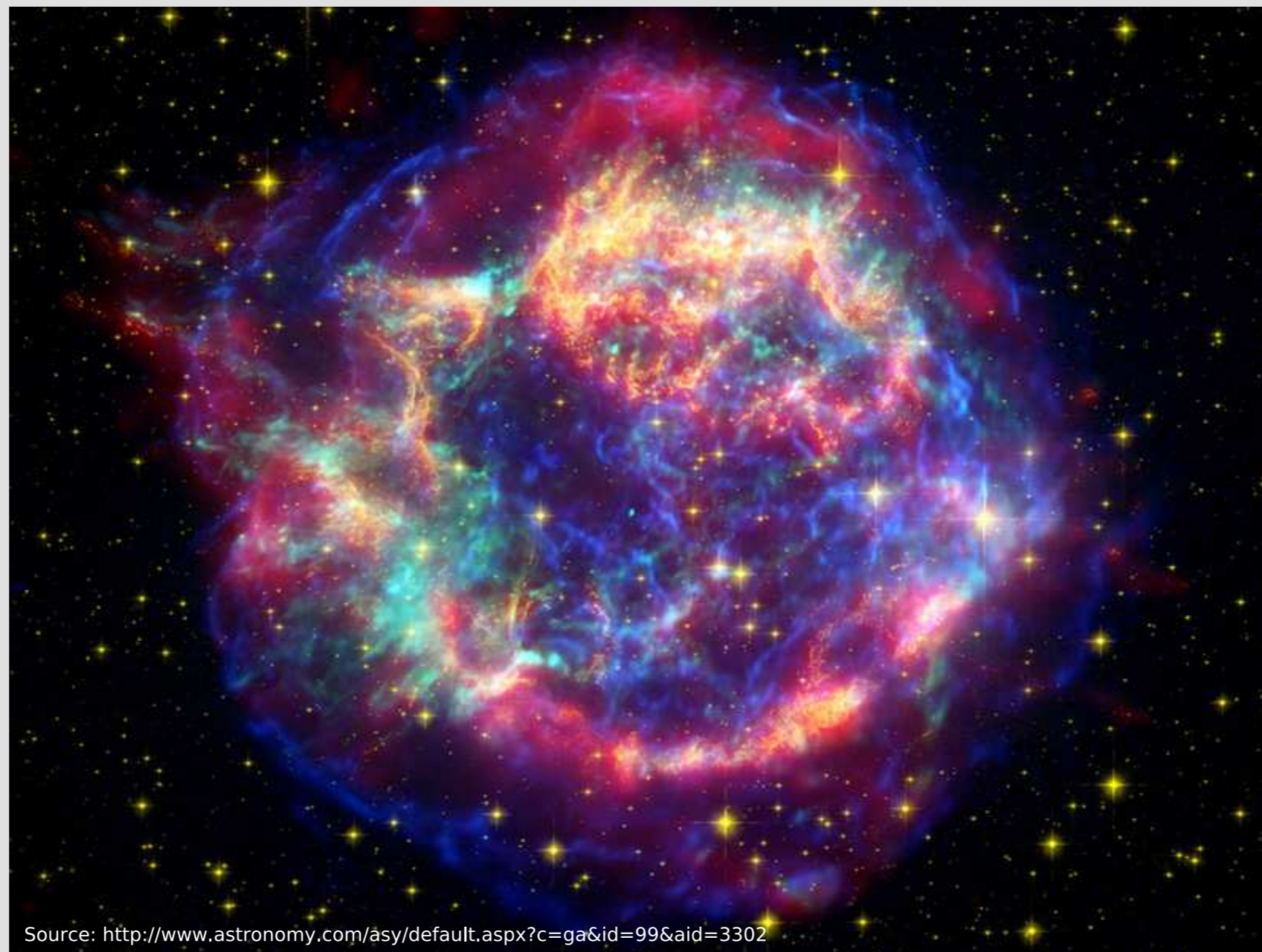
Tobias Scholl
Lehrstuhl Informatik III: Datenbanksysteme

Technische Universität München - Fakultät für Informatik

Supernova Remnant "Cassiopeia A"



- Hubble Space Telescope (HST)
- Chandra Observatory (CO)
- Spitzer Space Telescope (SST)



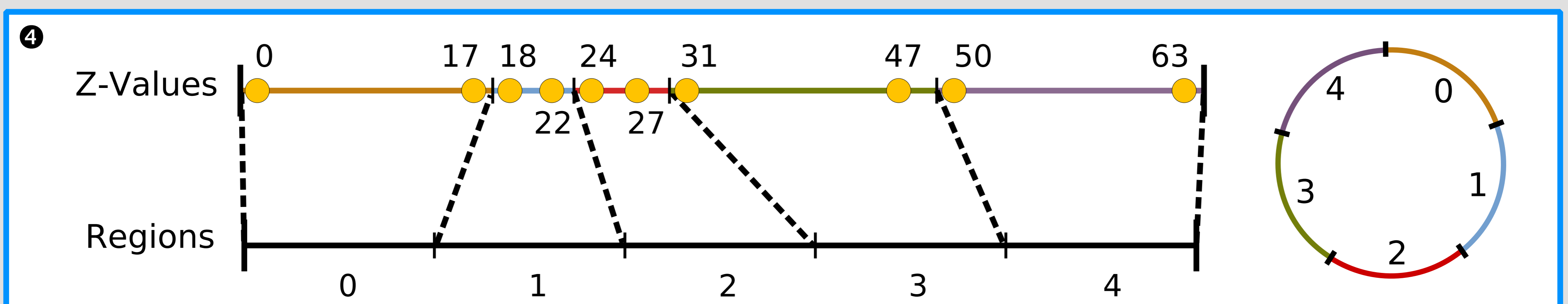
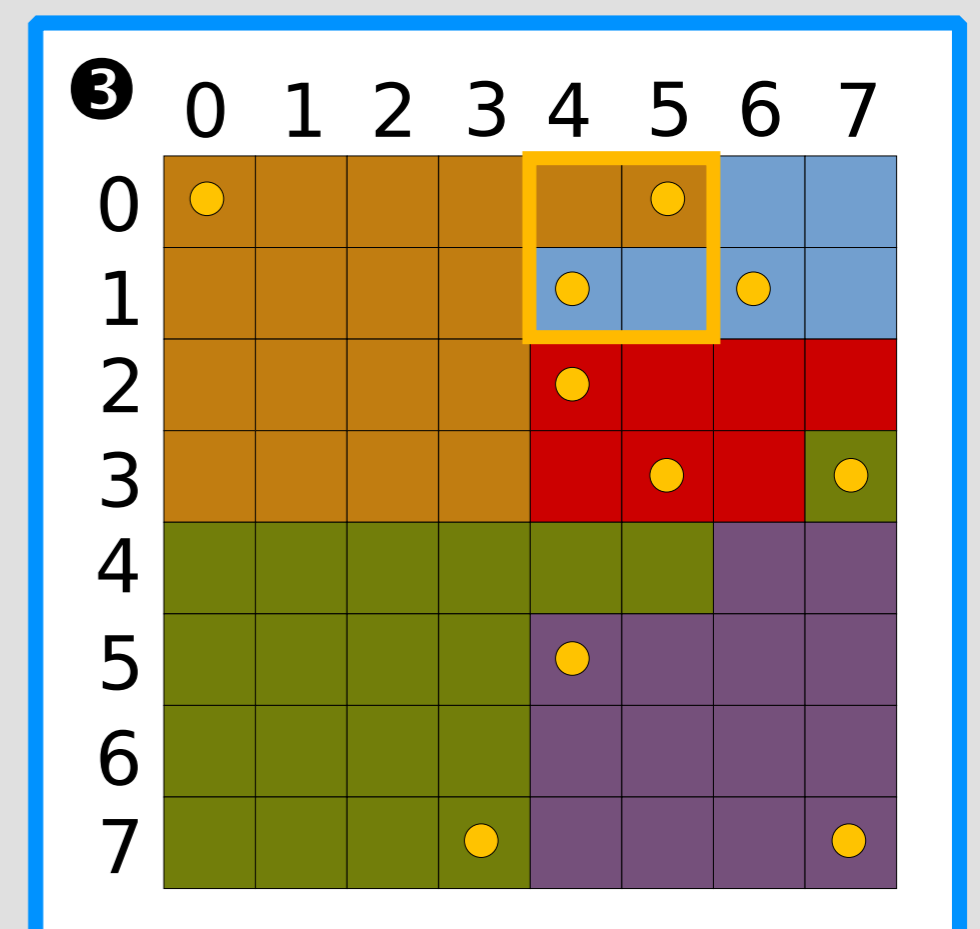
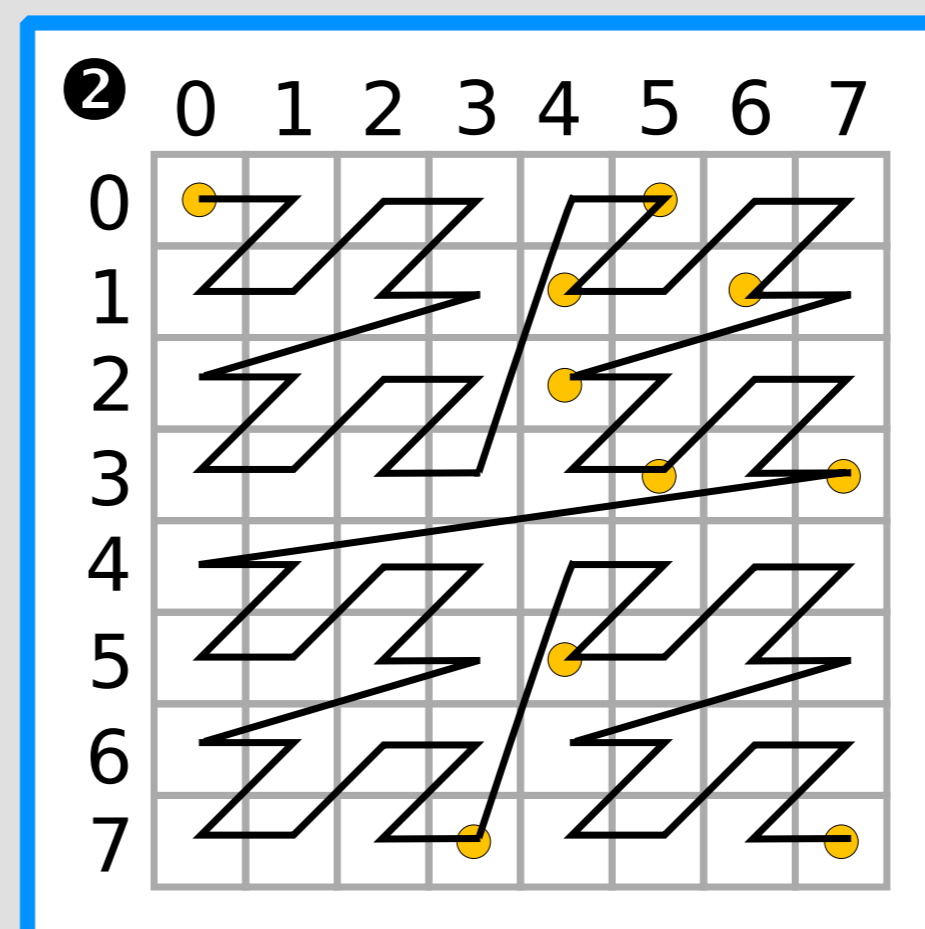
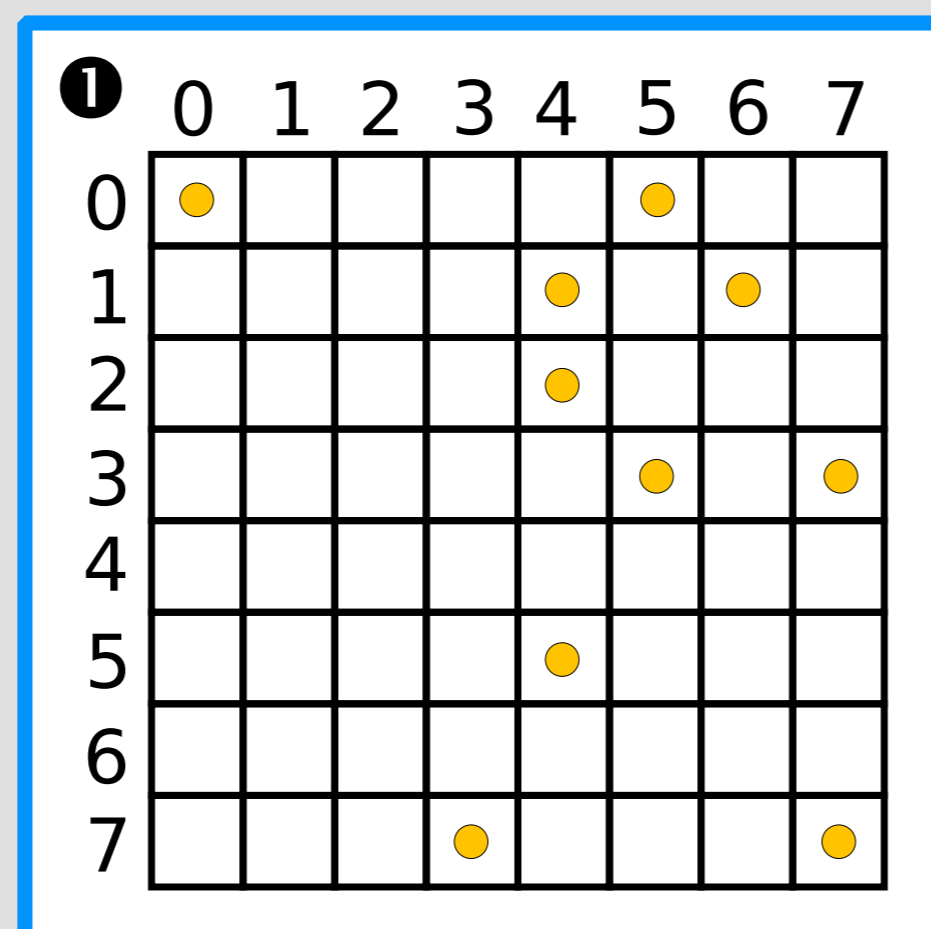
Source: <http://www.astronomy.com/asy/default.aspx?c=ga&id=99&aid=3302>

The Challenge

- Globally distributed archives
- Key research: correlation of archives
- Skewed data
- Region-based queries

"Distribute by Region — not by Archive"

- 1 Extract training set
- 2 Apply space-filling curve
- 3 Define regions
- 4 Distribute regions on keyspace of DHT system

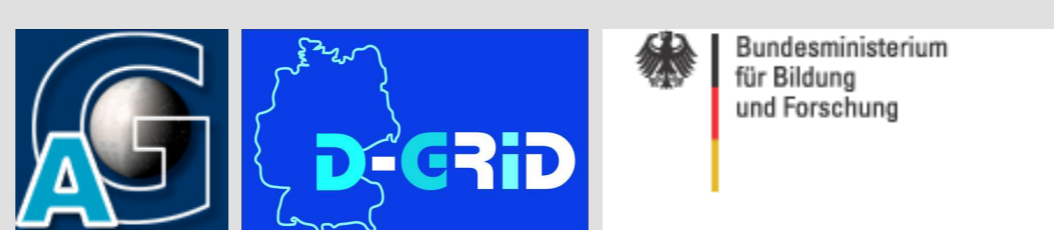


System Characteristics

- Highly distributed information management
⇒ Distributed hashtable (DHT) P2P architecture
- High performance query processing
⇒ Main memory database
- Semantic clustering and range queries
⇒ Equi-depth histograms, Preprocessing

Current Status

- Prototype implementation
 - Common API
 - Pastry
- Main memory database
 - HSQLDB
 - Evaluation, benchmarks
- Close cooperation with astrophysics partners in AstroGrid-D project



Research Agenda

- Distributed query processing
 - Coordination
 - Optimization
- Histogram techniques
- Training set extraction
- Query load and replication
- Scalability trade-off
 - Small-scale LAN
 - Large-scale WAN
- Persistent and streaming data