博士毕业论文答辩

论文名称：FAST 数据处理系统可视化分析与自动化管理技术研究

报告人：钱旭冉

专业：天文技术与方法

指导老师：朱明（国家天文台）

合作指导老师：Alyssa Goodman（CfA）

Michelle Borkin（NEU）

中国科学院大学

University of Chinese Academy of Sciences

Credit: mg1.mydrivers.com

论文提纲

1. 研究背景与意义
2. 可视化分析模块的设计与实现
3. 批量任务管理工具的开发
4. FAST数据产品的三维展示
5. 总结与展望

论文提纲

1. 研究背景与意义
2. 可视化分析模块的设计与实现
3. 批量任务管理工具的开发
4. FAST数据产品的三维展示
5. 总结与展望

FAST工程概况

FAST发展历程：

- 1994年4月，启动贵州选址工作；
- 2007年7月，FAST项目正式立项；
- 2016年9月，FAST工程竣工；
- 2017年8月，FAST首次新发现两颗脉冲星；
- 2018年4月，发现毫秒级脉冲星并得到国际认证。

FAST工程三大自主创新：

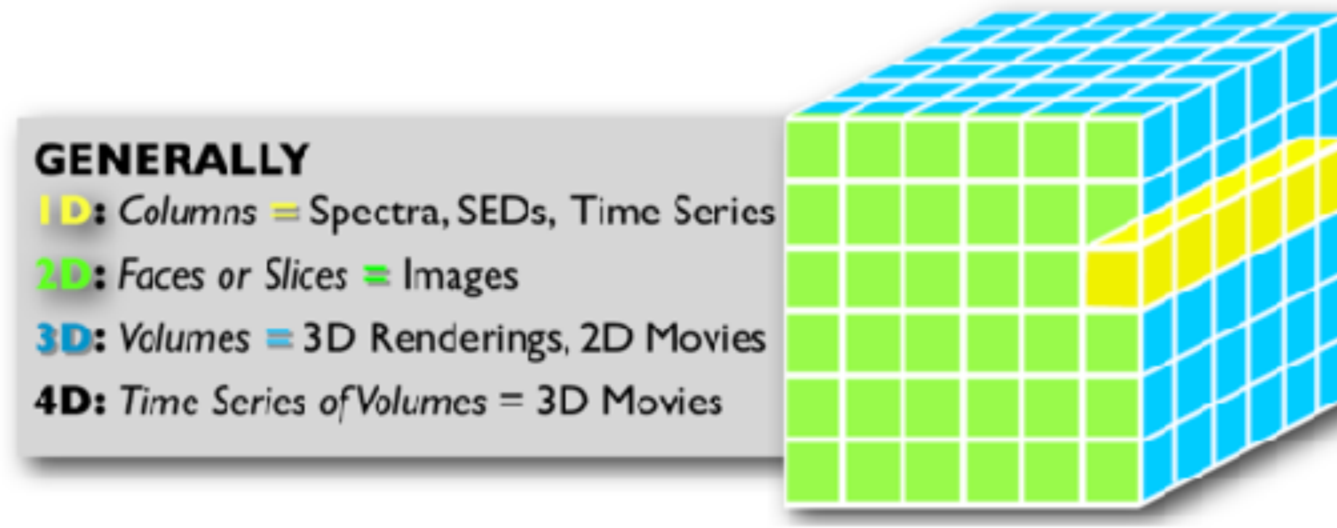
贵州天然喀斯特洼地台址；轻型索拖动机构和并联机器人实现接收机高精度定位与跟踪；主动反射面。

FAST主要科学目标：

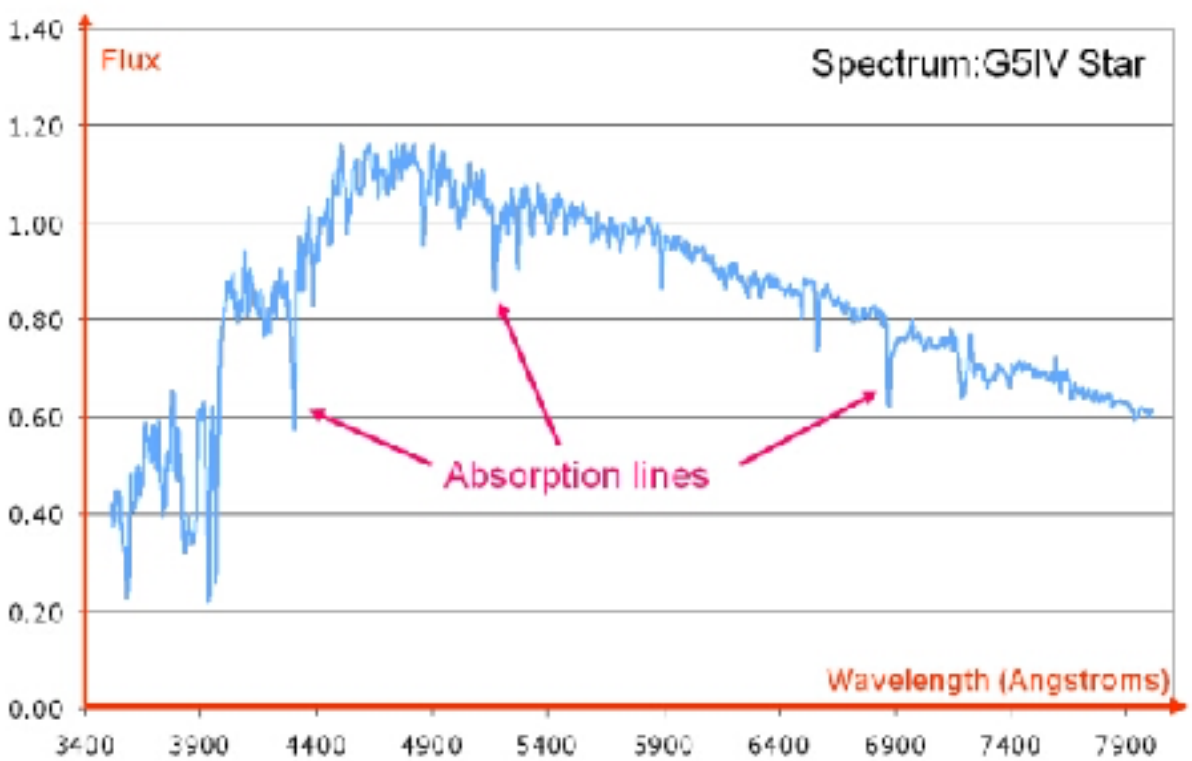
河内河外中性氢巡天，脉冲星观测

FAST数据产品

三维Cube

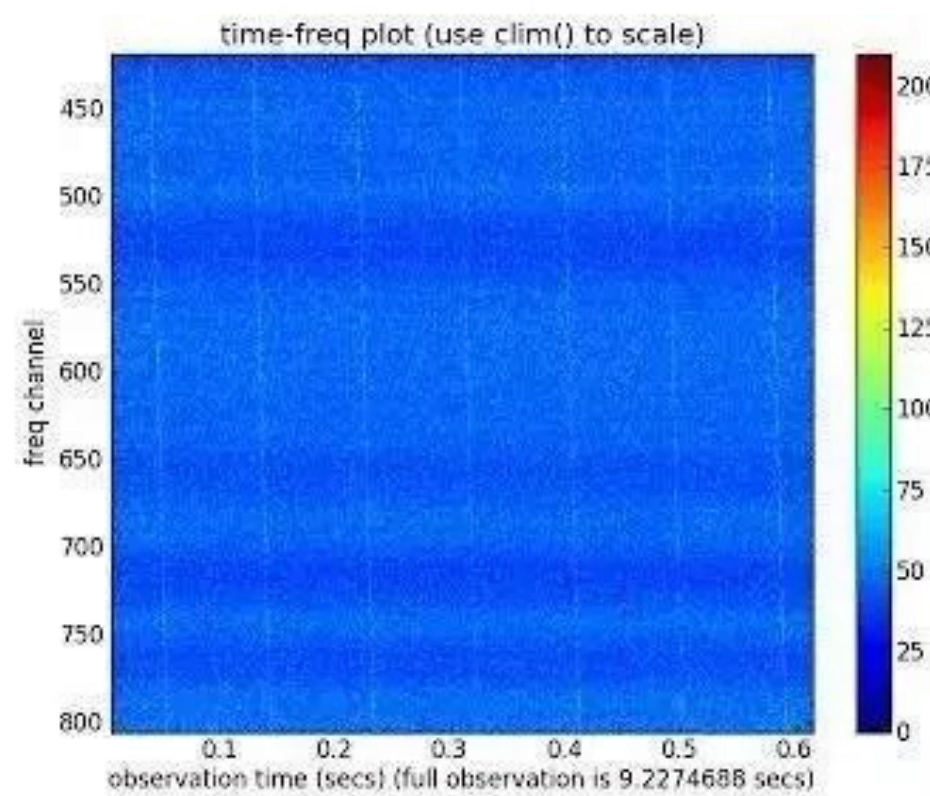


连续谱



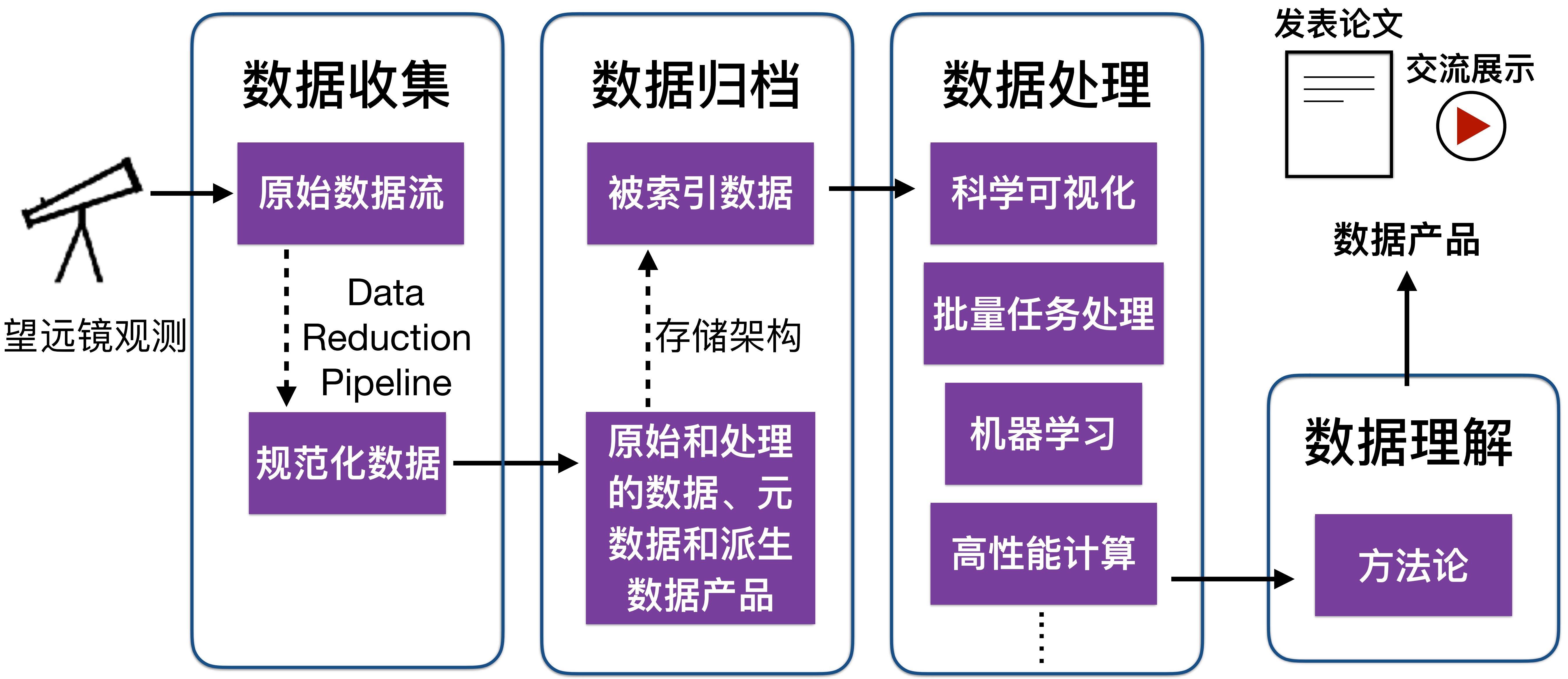
source: Santos et al. 2002

脉冲星时域信号

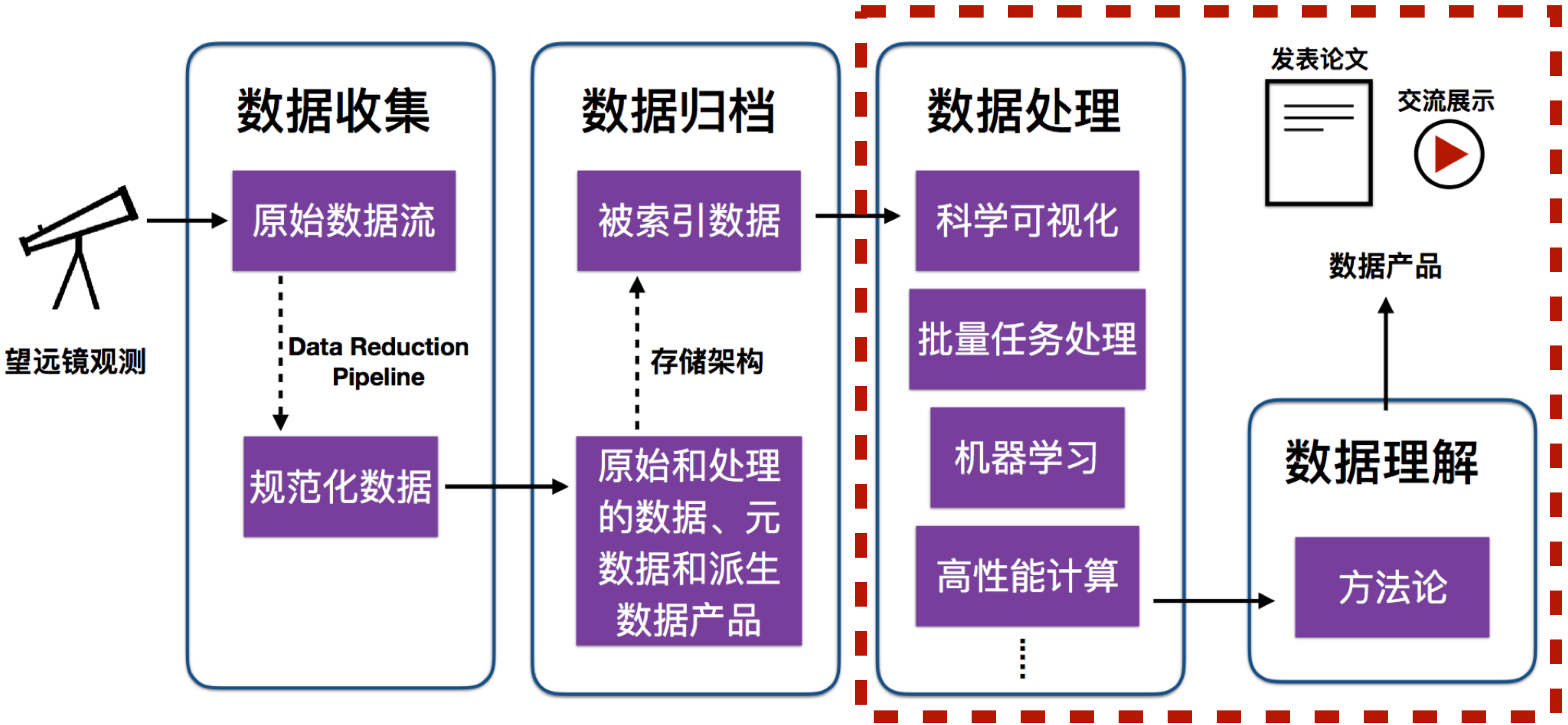


source: http://public.ska.ac.za/kat-7/kat-7-data-reduction/pulsars-and-transients

FAST数据处理系统概念框架图



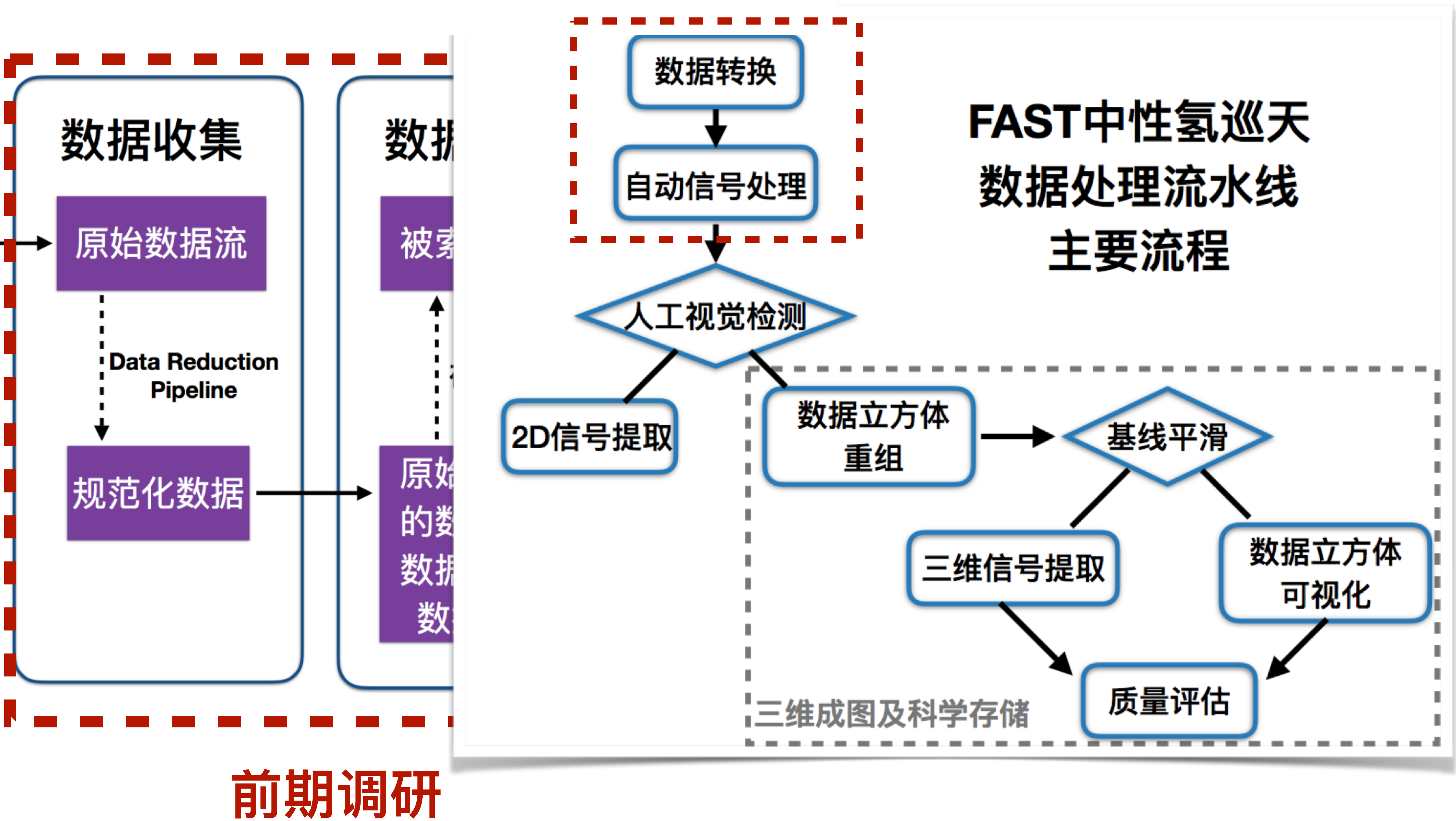
FAST数据处理系统概念框架图

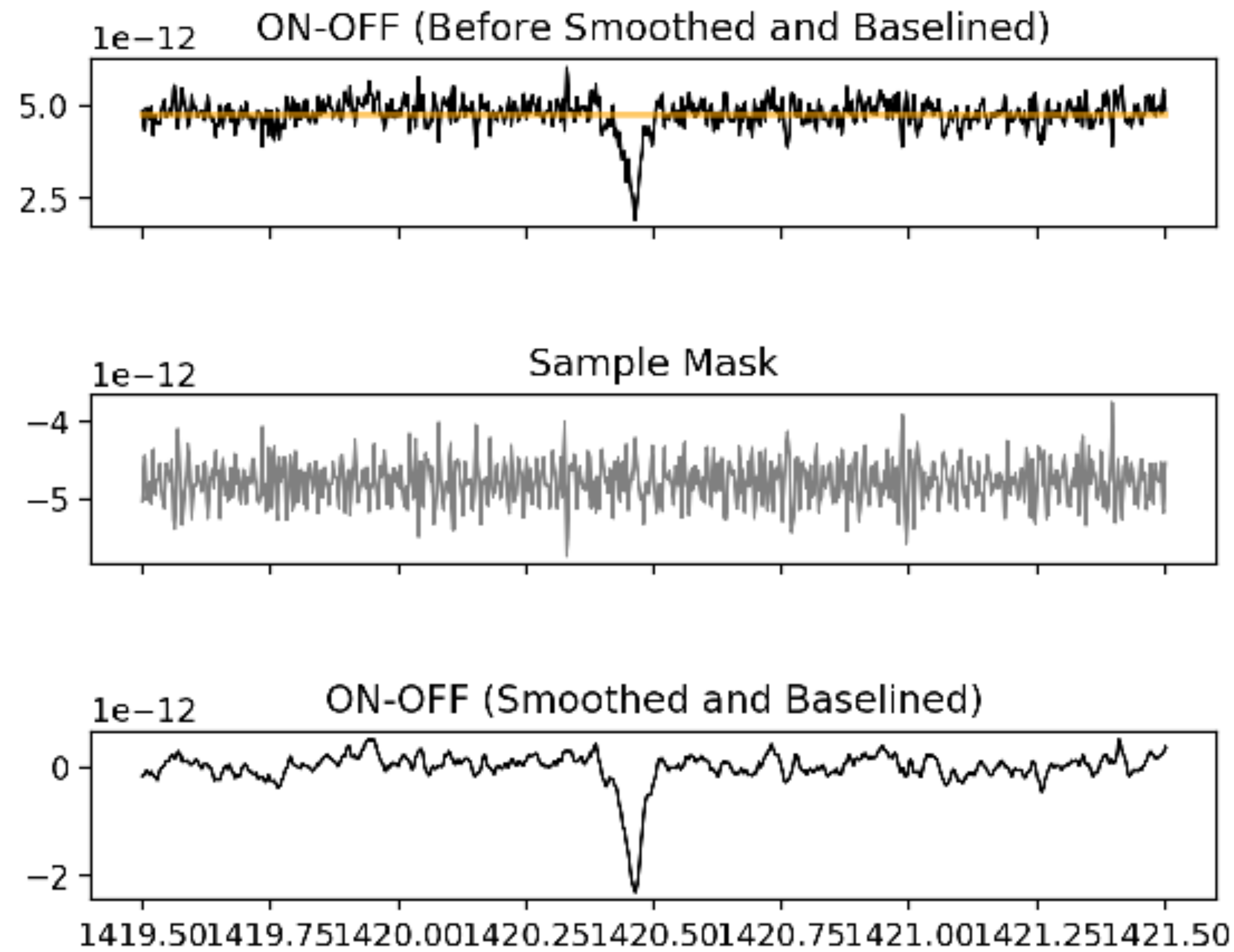


- 可视化分析模块的设计与实现
- 批量任务管理工具的开发
- FAST数据产品的三维展示

重点实现 - 结合背景优势

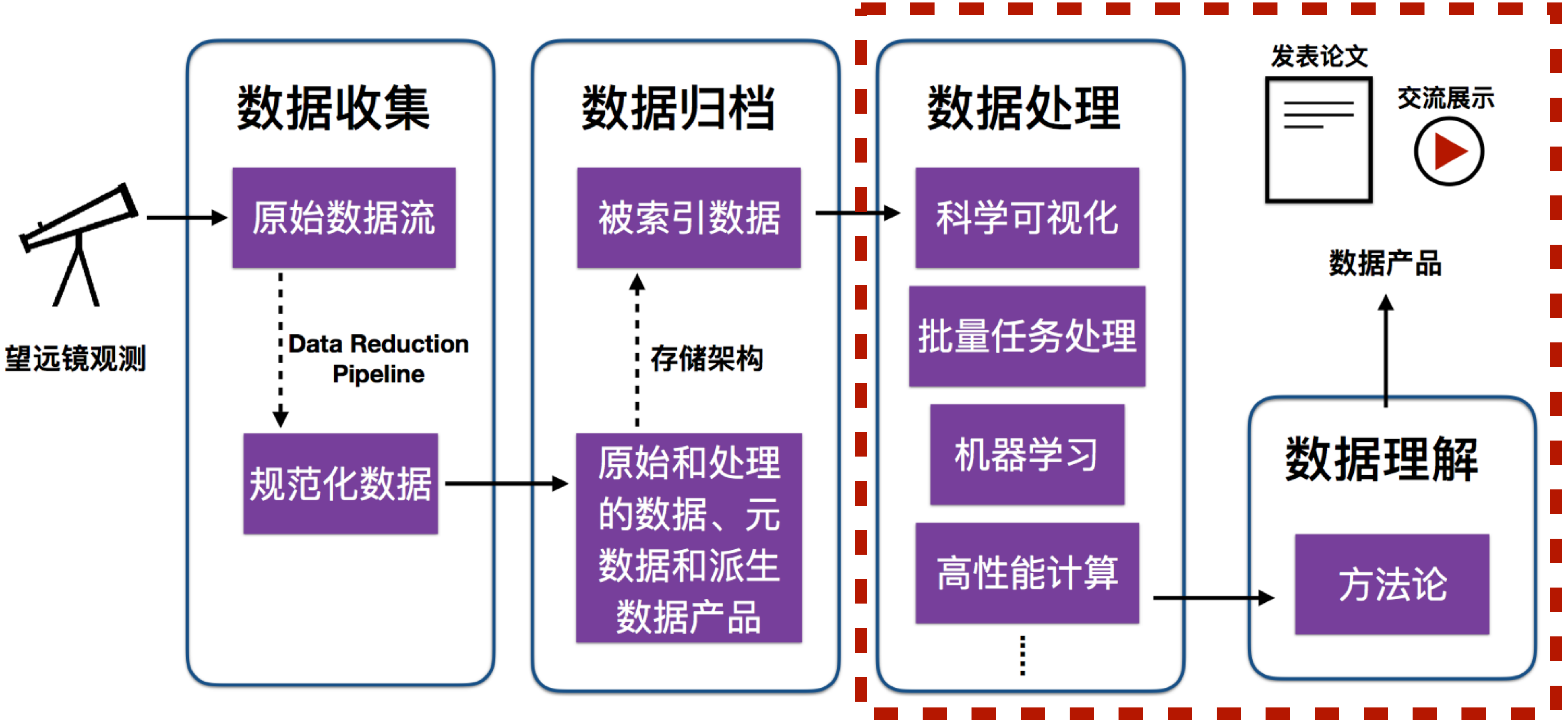
FAST数据处理系统概念框架图






```
➜ test_data git:(master) ✗ fastp
Please choose your observation mo
 [1 - Drifting 2 - Tracking]1
Please choose your instrument option
 [1 - Spectrometer 2 - Crane]1
Load data from local or remote? [l/r]l
Please type folder name: 20171105_UGC00613_HI_hql
('self.freq is', ['1419.5', '1421.5'])
Please type the smoothing width of the boxcaraverage: [10]
Baseline the result? [y/n] default as yes
Reading data by time for [2] Sessions
Loading session 1 :  [||||||||||||||||||||||||||||....]
```

FAST数据处理系统概念框架图



- 可视化分析模块的设计与实现
- 批量任务管理工具的开发
- FAST数据产品的三维展示

重点实现 - 结合背景优势

论文提纲

1. 研究背景与意义
2. 可视化分析模块的设计与实现
3. 批量任务管理工具的开发
4. FAST数据产品的三维展示
5. 总结与展望

研究目的

不同观测数据

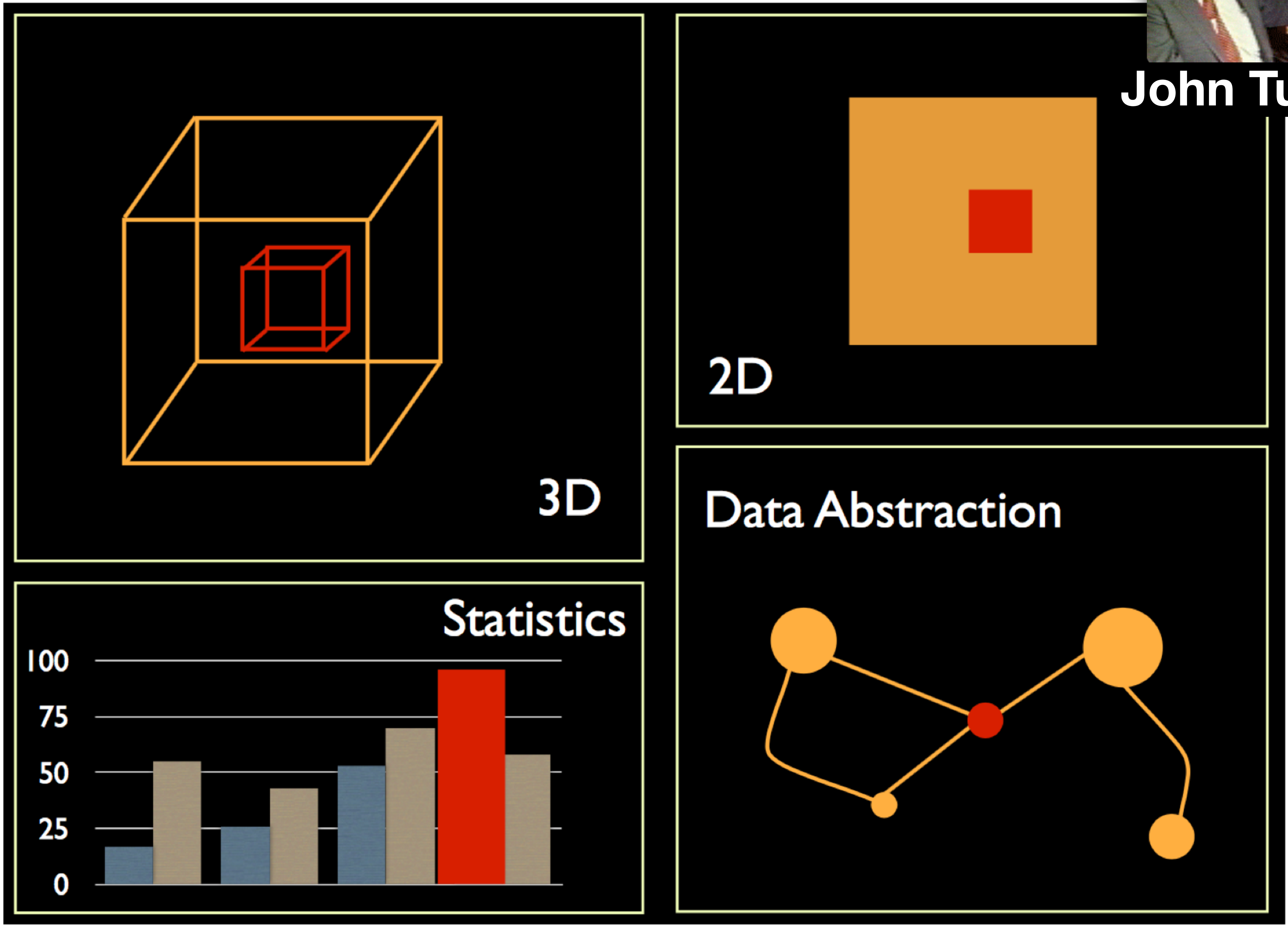
不同数据类型

观测	波段	连续谱	3D Cube	源列表	谱线
RAMPS*	1 cm		★		
CORNISH*	60 mm	★		★	
HOPS	12 mm		★		★
GRS	3 mm		★		
MALT90	3 mm				★
THRUMMS	3 mm		★		
Dame CO	2.6 mm		★		
BGPS	1 mm	★		★	★
CHIMPS	1 mm		★		
COHRS	1 mm		★		
ATLASGAL	870 μm	★		★	
JCMT*	850 μm	★		★	
HIGAL*	70-500 μm	★			
MIPSGAL	24, 70 μm	★			
WISE	3.4, 4.6, 12, 22.0 μm	★			
GLIMPSE	3.6, 4.5, 5.8, 8.0 μm	★			
UKIDSS-GPS*	1.3, 1.6, 2.2 μm	★			

- 数据结构复杂性
- 数据之间关联性
- 数据高维度

milkyway3d.org

可视化分析与视图互联

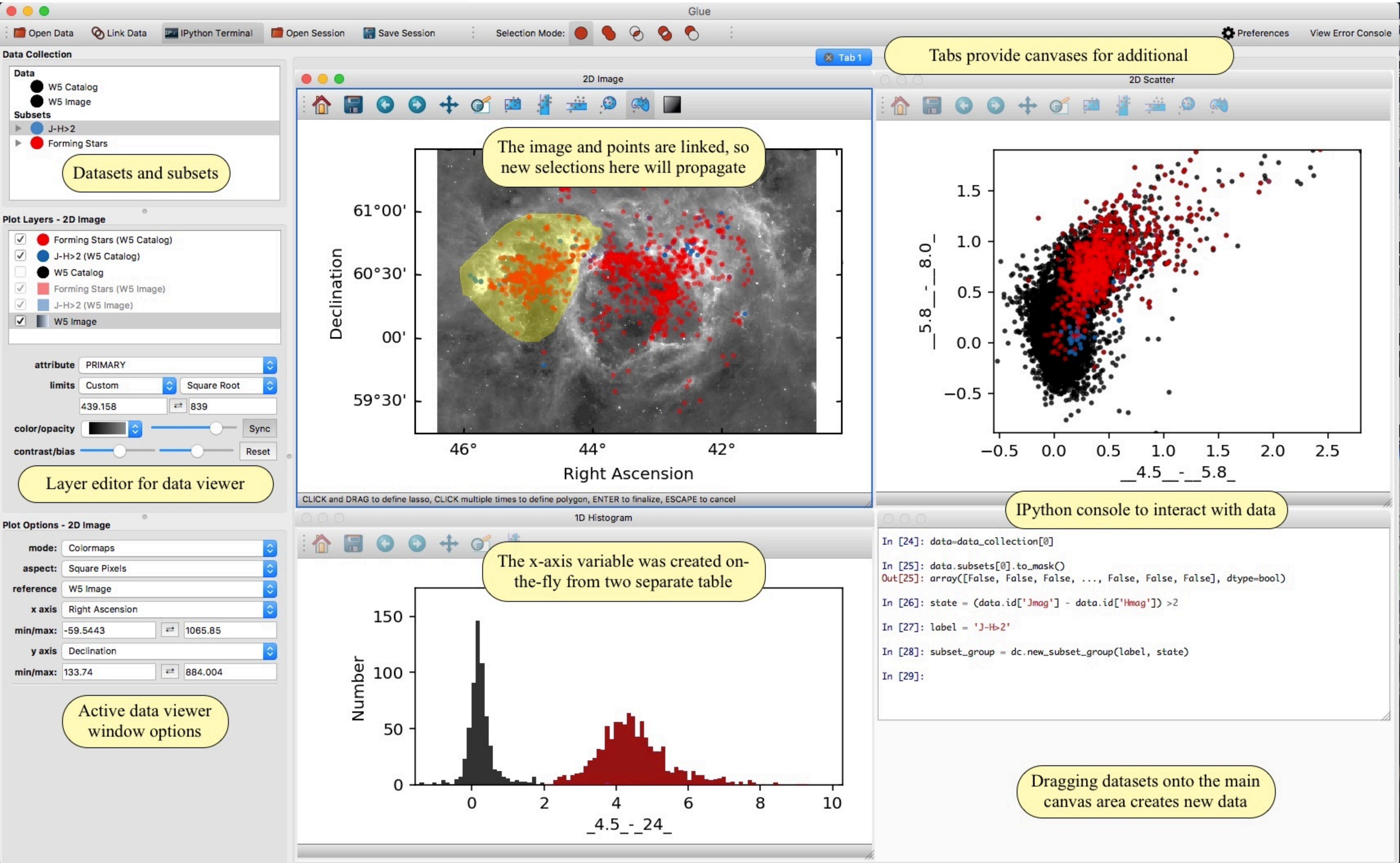


本图来源 *[A. Goodman(2012)]* 的*Fig. 3*，本图由*Michelle Borkin*创作。

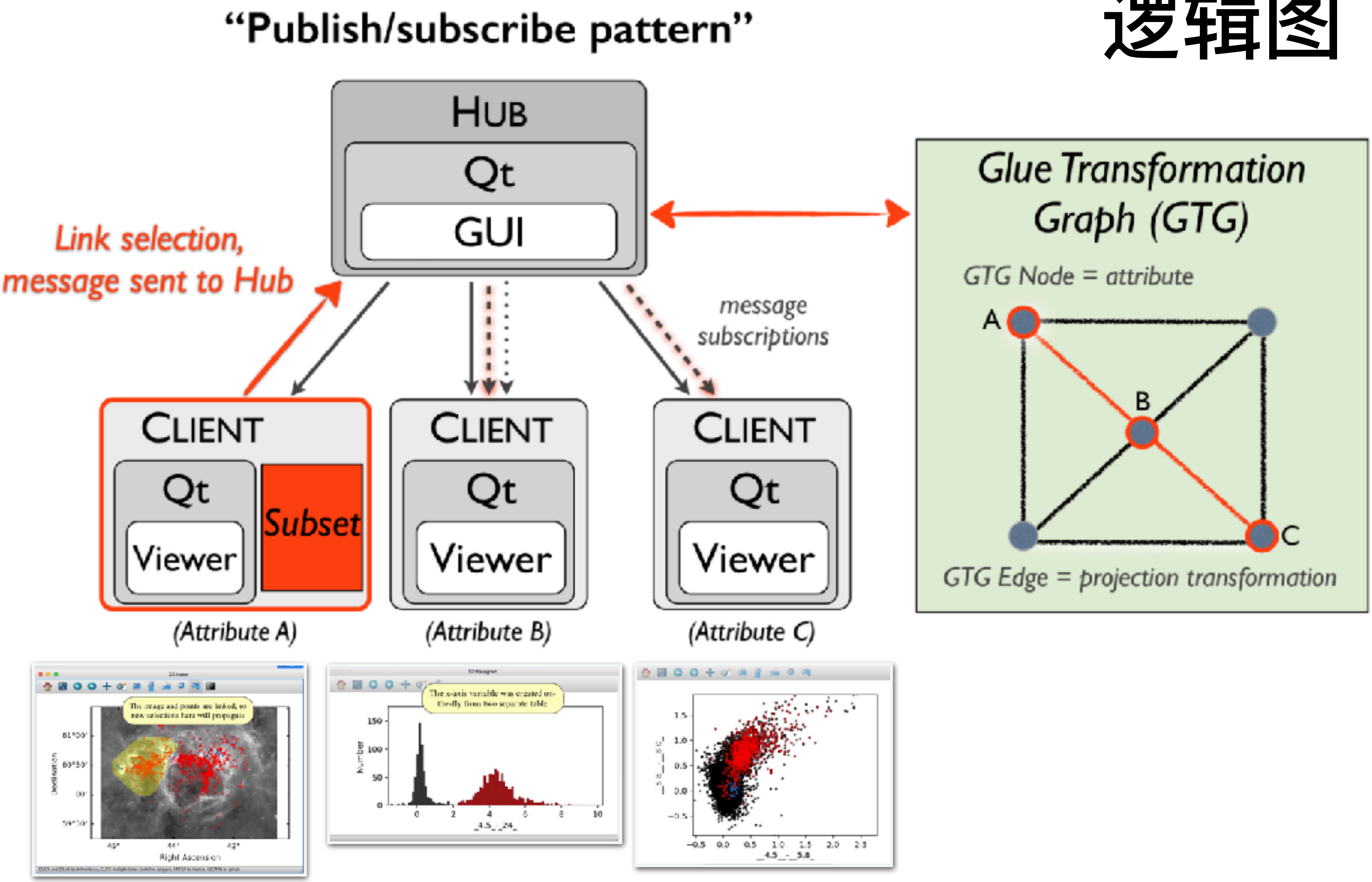
glue

multidimensional data exploration

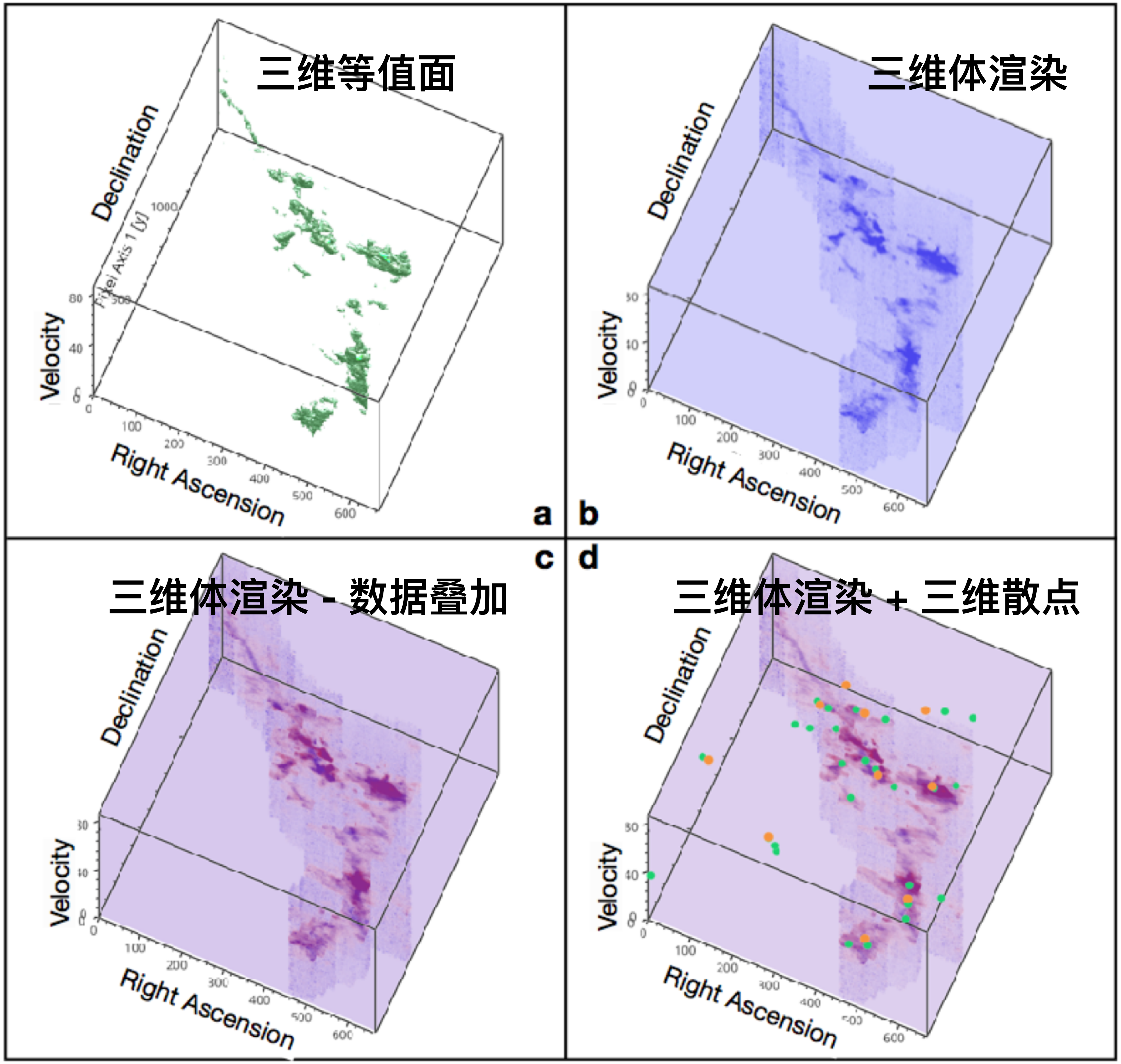
基于视图互联的多维度天文数据分析软件



逻辑图

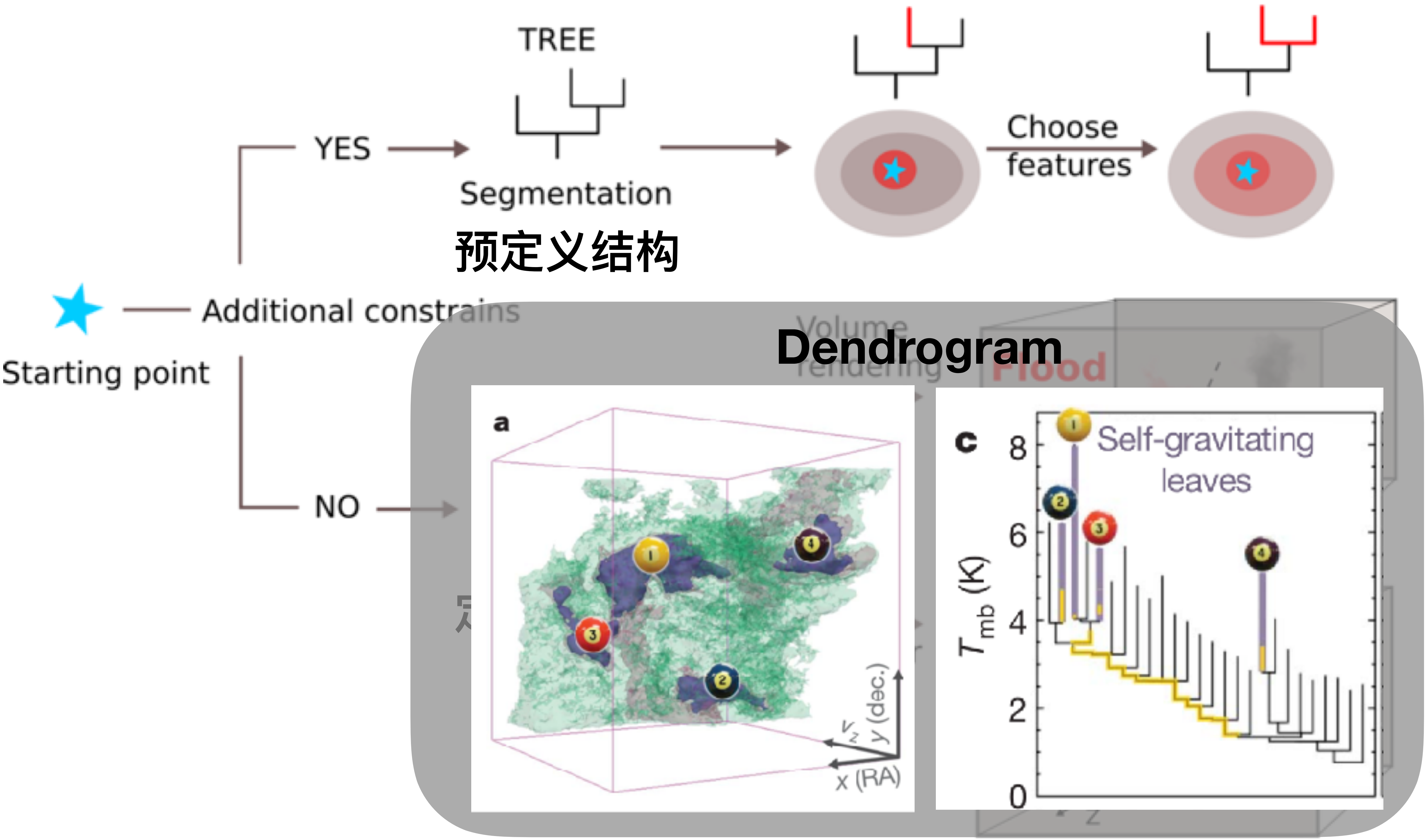


Borkin, Qian, et al. (In prep.)



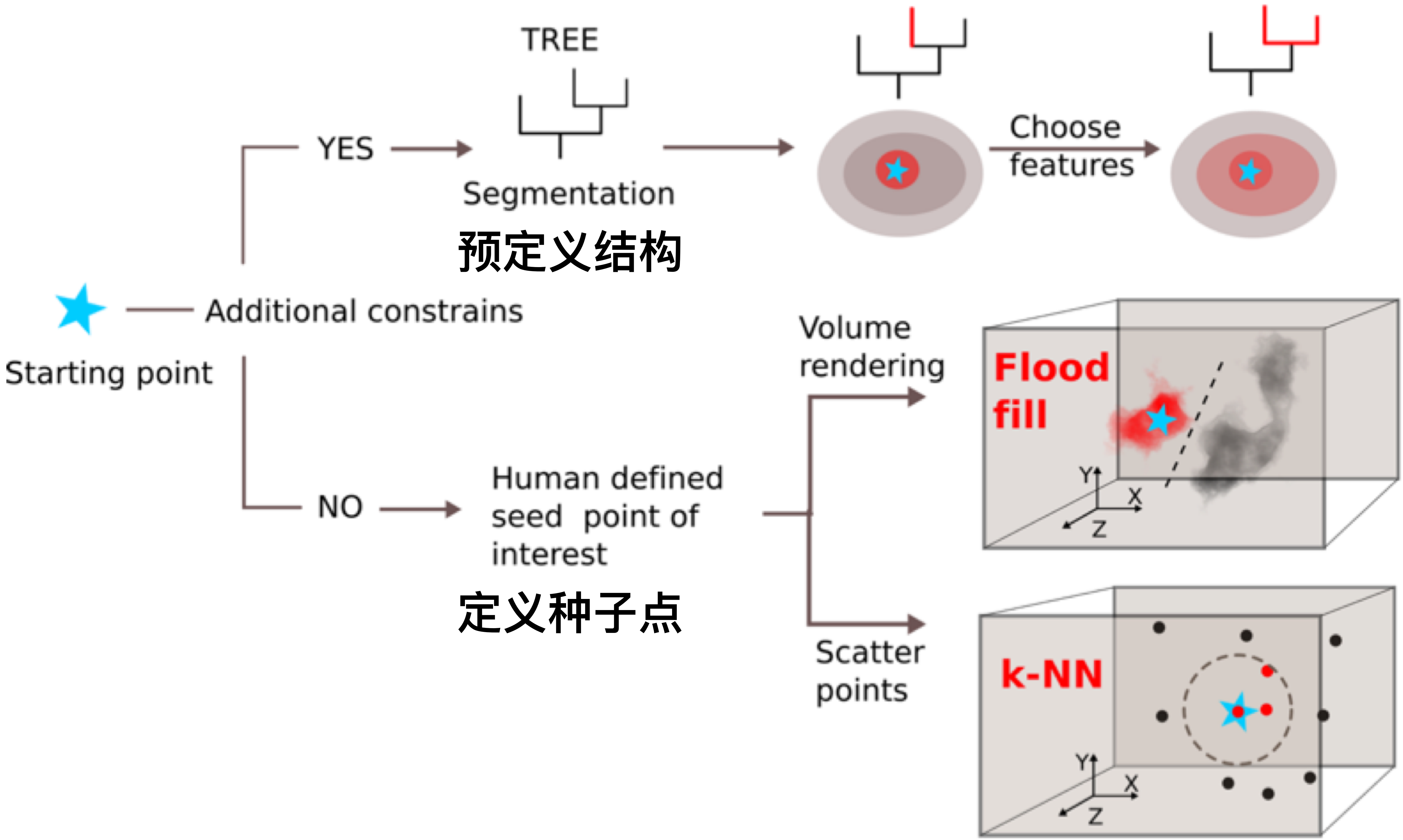
- OpenGL
- 三维交互
- 多数据叠加

三维选取算法



TREE
YES
Segmentation
预定义结构
Choose features
Additional constrains
Starting point
NO
Dendrogram
a
y (dec.)
x (RA)
c
Self-gravitating leaves
8
6
4
2
0
T_{mb} (K)

三维选取算法


TREE
YES
Segmentation
预定义结构
Choose
features
Starting point
Additional constrains
Volume
rendering
Flood
fill
Human defined
seed point of
interest
NO
定义种子点
Scatter
points
k-NN
Y
X
Z

Floodfill算法核心

输入设备

指数映射

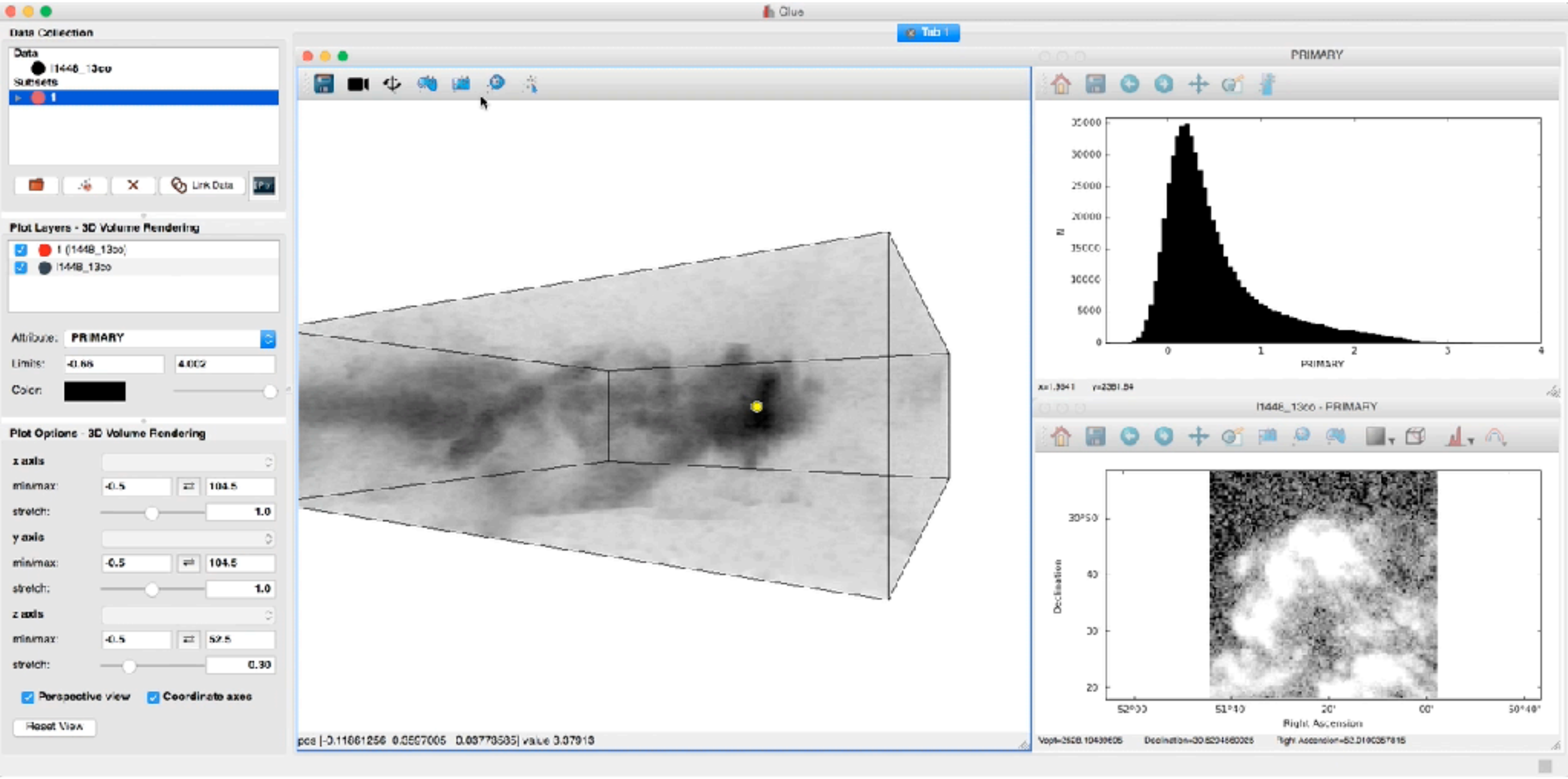
```
function Floodfill(x, y, z, threshold)
    if note[x][y][z].value > threshold:
        mark note[x][y][z] as visited
    if note[x][y][z].value <= threshold:
        return
        Floodfill(x-1, y, z, threshold)
        Floodfill(x, y-1, z, threshold)
        Floodfill(x, y+1, z, threshold)
        Floodfill(x+1, y, z, threshold)
        Floodfill(x, y, z-1, threshold)
        Floodfill(x, y, z+1, threshold)
```

$$t = m/s$$

$$t = 1 + 10^{(4t-2)}$$

m 是鼠标移动距离，s 是屏幕对角线，t 表示阈值；

视图互联1D/2D/3D



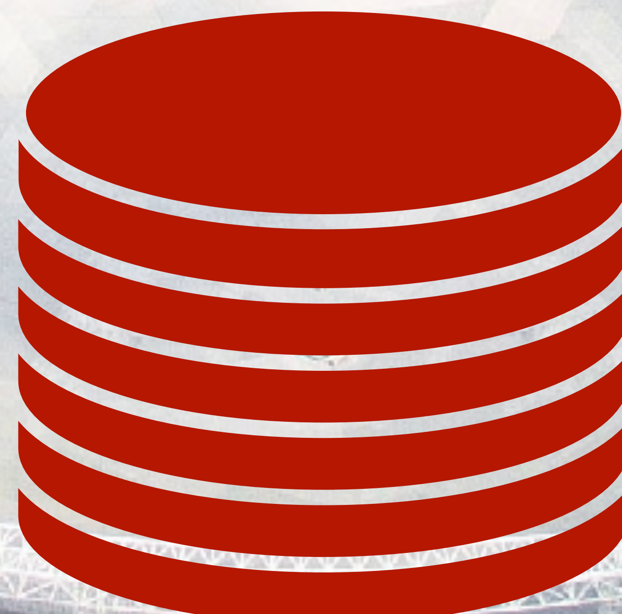
Glue
Data Collection
Data
Subsets
Link Data
Plot Layers - 3D Volume Rendering
Attribute: PRIMARY
Limits:
Color:
Plot Options - 3D Volume Rendering
x axis
min/max:
stretch:
y axis
z axis
Perspective view
Coordinate axes
Reset View
PRIMARY

glue
multidimensional data exploration

FAST可视化分析模块

FAST大数据背景下的需求分析

- 数据展示
- 数据选取
- 数据及视图互联



FAST数据

- 数据分析
- 数据检索
- 互操作性

可视化分析软件对比

商业软件	Tableau DataDesk Spotfire Plot.ly
其它领域	Igor Filtergraph
编程语言	IDL
现有天文软件	ds9 TOPCAT Mathematica ArcGIS CARTA

可视化分析软件对比

互联功能	数据互联，视图互联
二维展示与交互	标准化二维plot，自定义二维plot，二维基本选取，二维算法辅助选取
三维展示与交互	三维散点图，三维等值面图，三维体渲染，三维基本选取，三维算法辅助选取
远程连接	远程分析工具连接，远程数据连接
功能	GPU加速，Web端显示，软件互联
自定义化	脚本编程，自定义模块

View this table online at
tinyurl.com/viz-features

互联功能

二维展示与交互

三维展示与交互

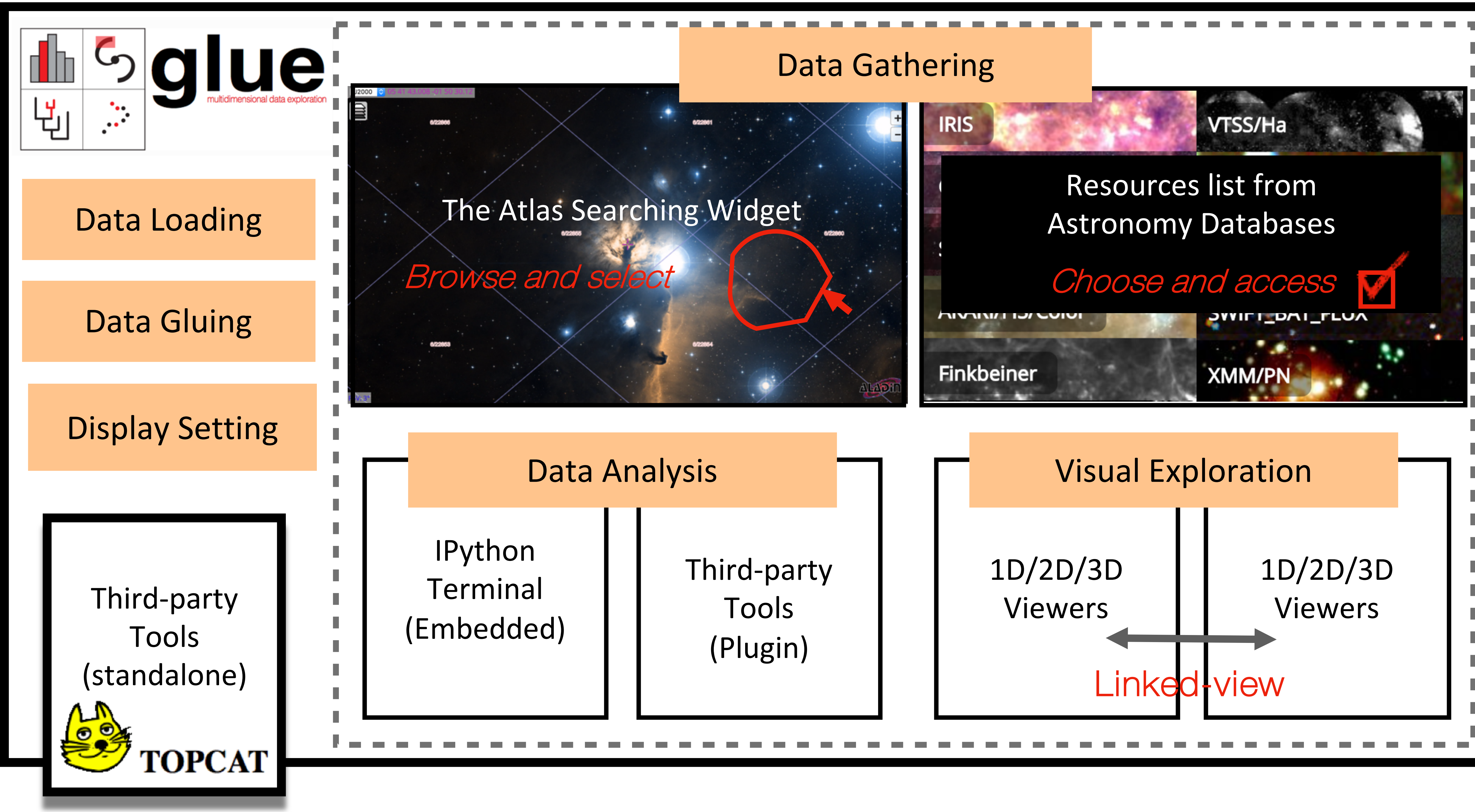
远程连接

功能

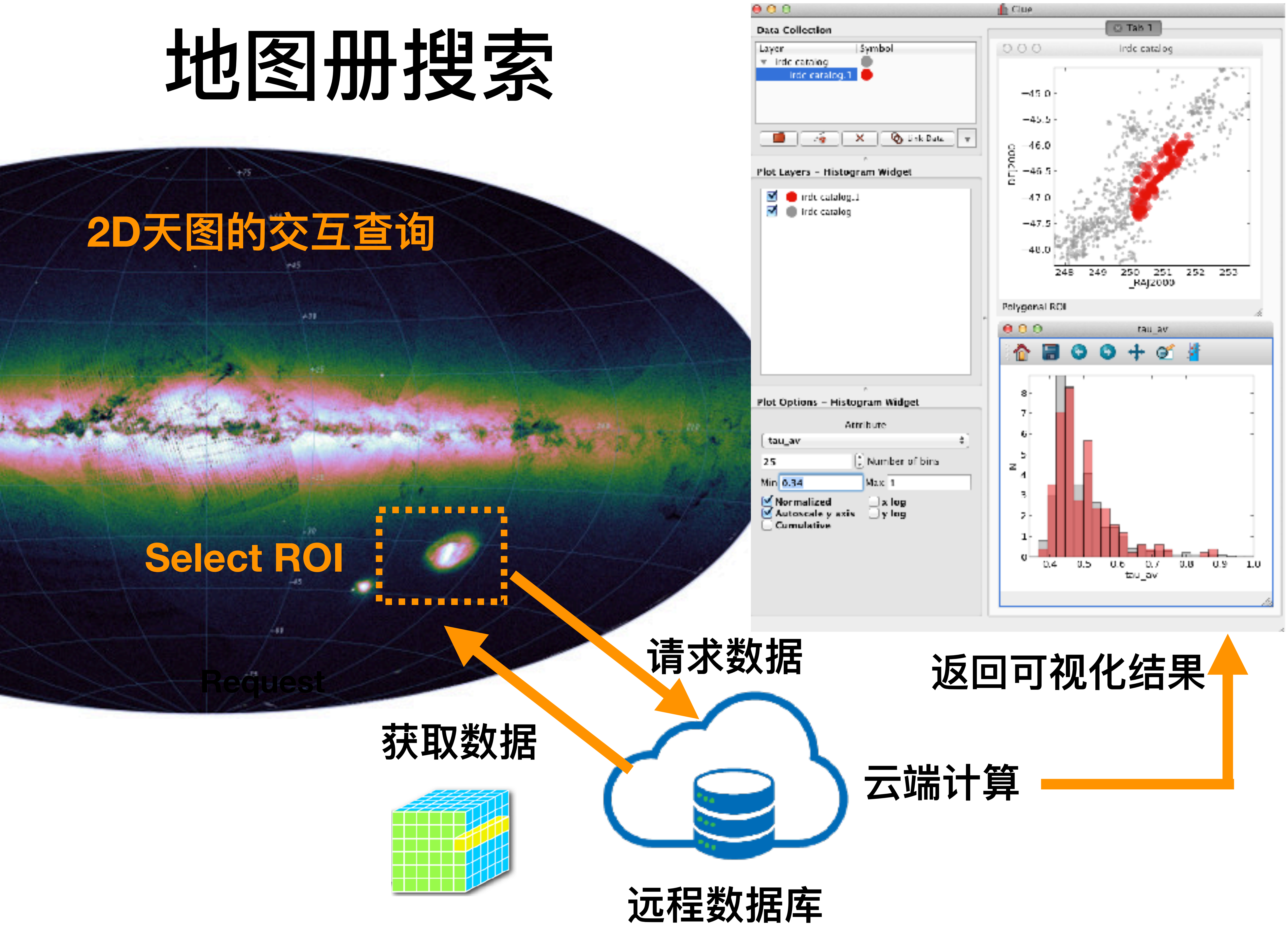
自定义化

	Glue (now)	Glue (planned)	Tableau	DataDesk	Spotfire	Plot.ly	Igor	Filtergraph	IDL	ds9	TOPCAT	Mathematica	ArcGIS	CARTA
linked views	✓	✓	✓	✓	✓	✗	~	✗	✗	✗	✓	~	✓	✗
linking data files	✓	✓	~	✗	✗	✗	✗	✗	✗	✗	~		~	✗
standard 2D plots (e.g. x-y, histograms)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✗
custom 2D plots (e.g. trees)	✓	✓	~	✗	✓	~	✓	✗	✓	✗	✗			✗
image display (formats)	✓ (astro, med, GIS)	✓	✗	✗	✗	✗		✗	✓	✓	✓ (tabular data)	✓ (astro, ...?)	✓ (GIS)	✓
2D selection	✓	✓	✓	✓	✓	✗		✓	✓	✗	✓			✗
2D selection (smart)	~	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
3D point cloud plots	✓	✓	✗	✓	✓ (Desktop)	✓			✓	✗	✓	✓	✗	✗
3D surface renderings	✓	✓	✗	✗	✗	✓	✓		✓	~	✗	✓	✗	✗
3D volumetric views	✓	✓	✗	✗	✗	✗			✓	✓	✓			✗
3D selection (basic)	✓	✓	✗	✓	✓				✗		✗			✗
3D selection (smart)	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
(access to) statistics & fitting	~	✓	✗	✓	✓	~	✓	~	✓	✗		✓	~	~
remote data access	✗	✓	✓	✗	✓				✗	✓	✓			✓
GPU acceleration	~	✓		✗	Not found				~		Not found			✗
web dashboard	✗	✓	✓	✗	✓	✓	✗	✓	✗	✗	✗	~	✓	✓
Hub integration (e.g. SAMP)	✗	✓	✗	✗	✗	✗	✗	✗	✗	✓	✓	✗	✗	✗
custom widgets	✓	✓	✗	✗	✗	✗	✓	✗	✓	✗	✗	✓	✗	✗
web output	✓	✓	✓	✗	✓	✓	✗	✓	✗	✗	✗	✓	✓	✓
scripting options	Jupyter	Jupyter			R						No			Python
primary development language(s)	python	python, javascript	C++		S+	d3, javascript					Java			C++ & Java!

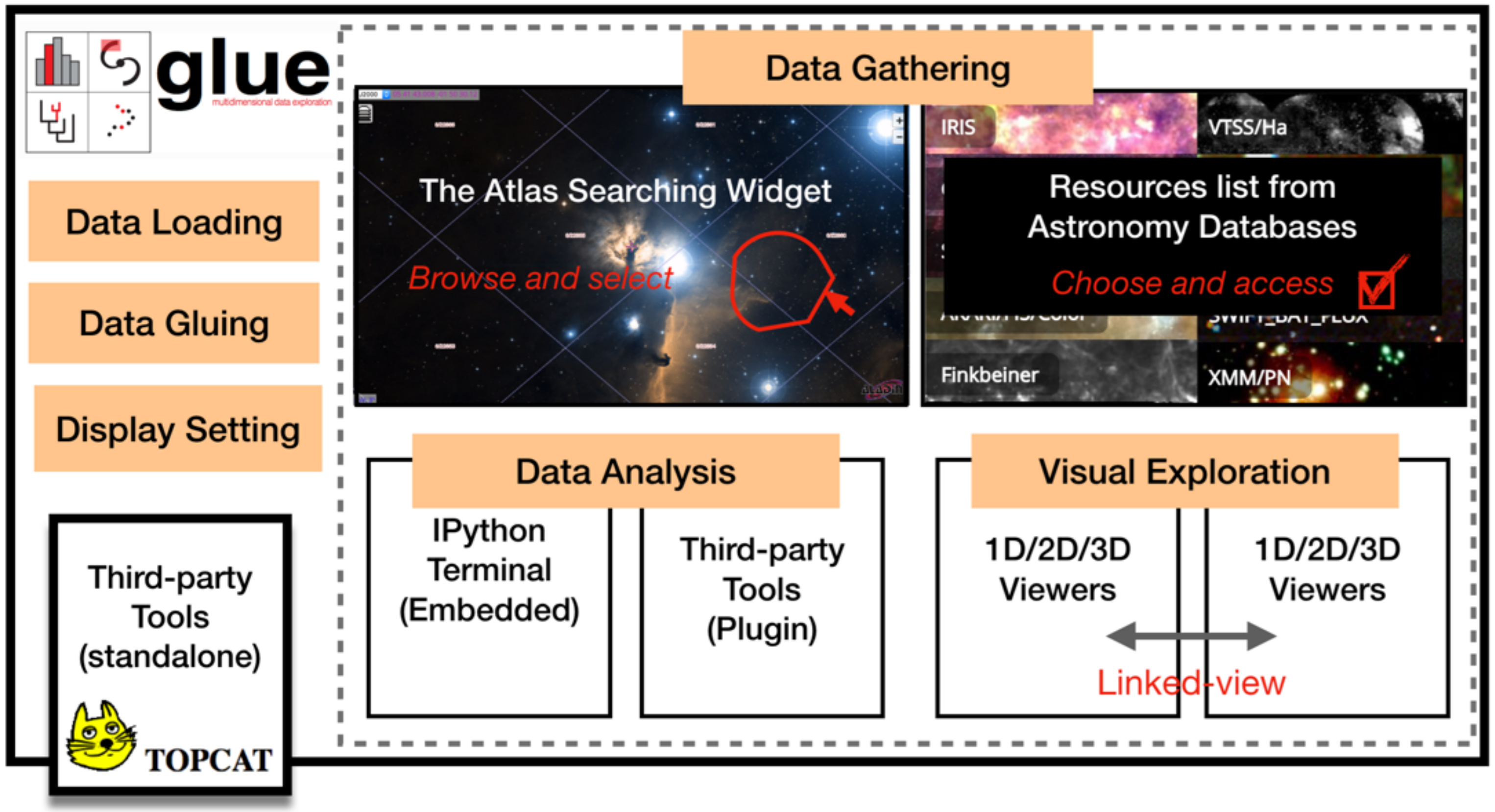
FAST可视化分析系统的概念框架图

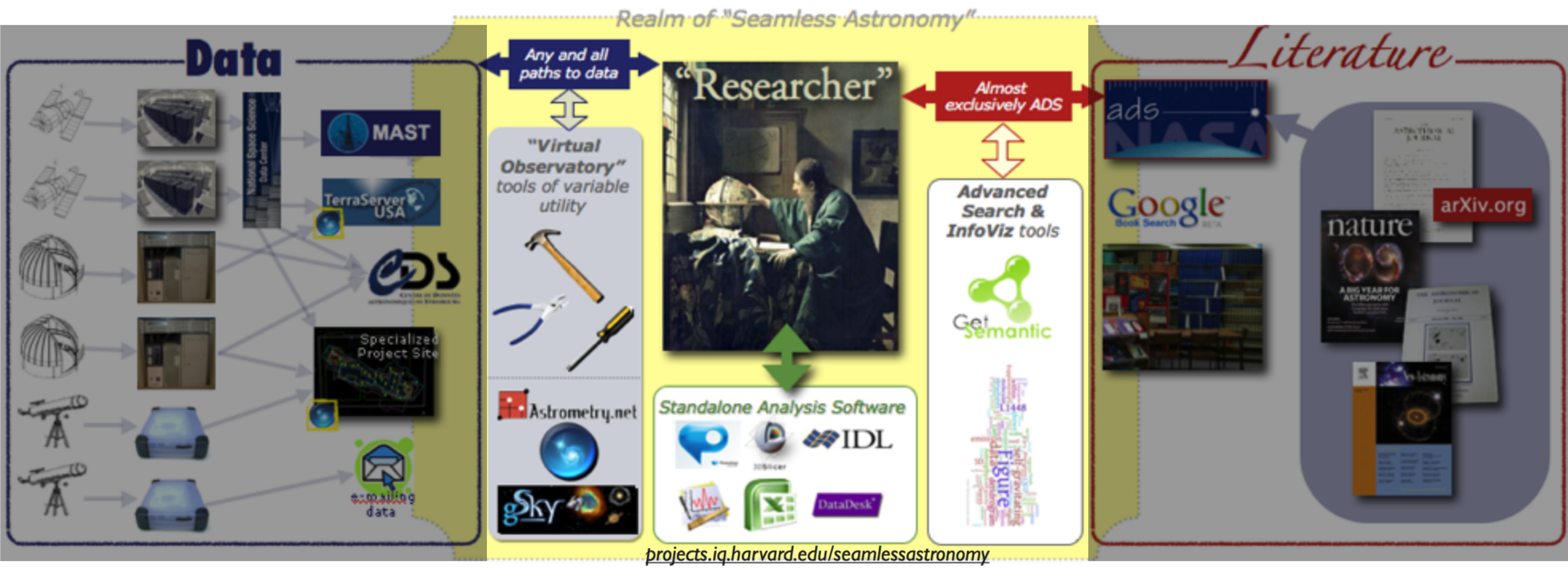


地图册搜索

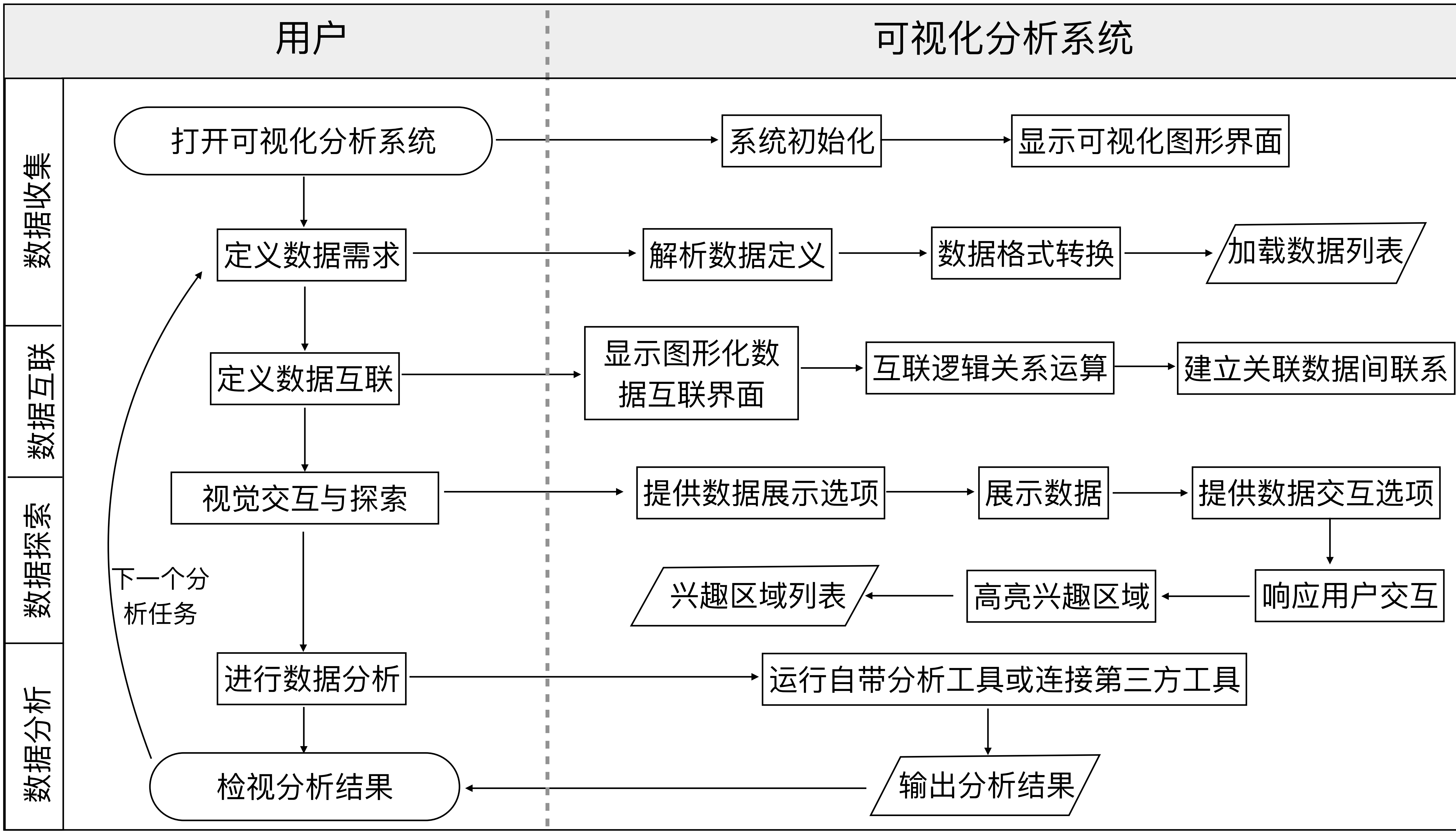

2D天图的交互查询
Select ROI
Request
请求数据
返回可视化结果
获取数据
云端计算
远程数据库
Glue
Data Collection
Layer
Symbol
irdc catalog
irdc catalog.1
Plot Layers - Histogram Widget
Plot Options - Histogram Widget
Attribute
tau_av
25
Number of bins
Min 0.34
Max 1
Normalized
Autoscale y axis
Cumulative
x log
y log
Tab 1
irdc catalog
_RAJ2000
Polygonal ROI
tau_av

"无缝"
天文研究


glue
Data Gathering
Data Loading
Data Gluing
Display Setting
The Atlas Searching Widget
Browse and select
IRIS
VTSS/Ha
Resources list from Astronomy Databases
Choose and access
Finkbeiner
XMM/PN
Third-party Tools (standalone)
TOPCAT
Data Analysis
IPython Terminal (Embedded)
Third-party Tools (Plugin)
Visual Exploration
1D/2D/3D Viewers
1D/2D/3D Viewers
Linked-view


Realm of "Seamless Astronomy"
Data
MAST
TerraServer USA
Specialized Project Site
Any and all paths to data
"Virtual Observatory" tools of variable utility
Astrometry.net
"Researcher"
Standalone Analysis Software
IDL
DataDesk
Almost exclusively ADS
Advanced Search & InfoViz tools
Literature
ads
Google
nature
arXiv.org
projects.iq.harvard.edu/seamlessastronomy

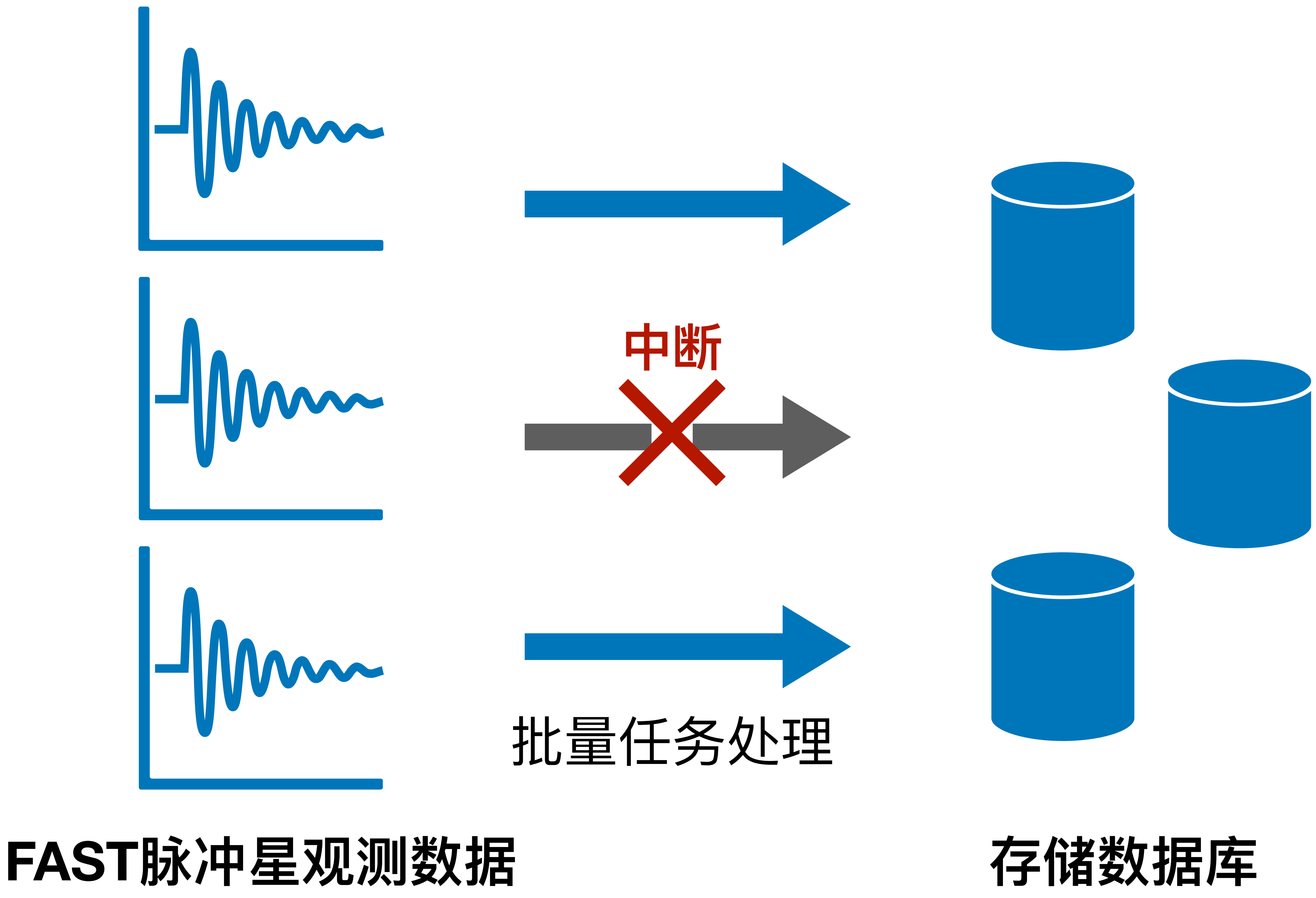
FAST可视化分析模块的工作流程图



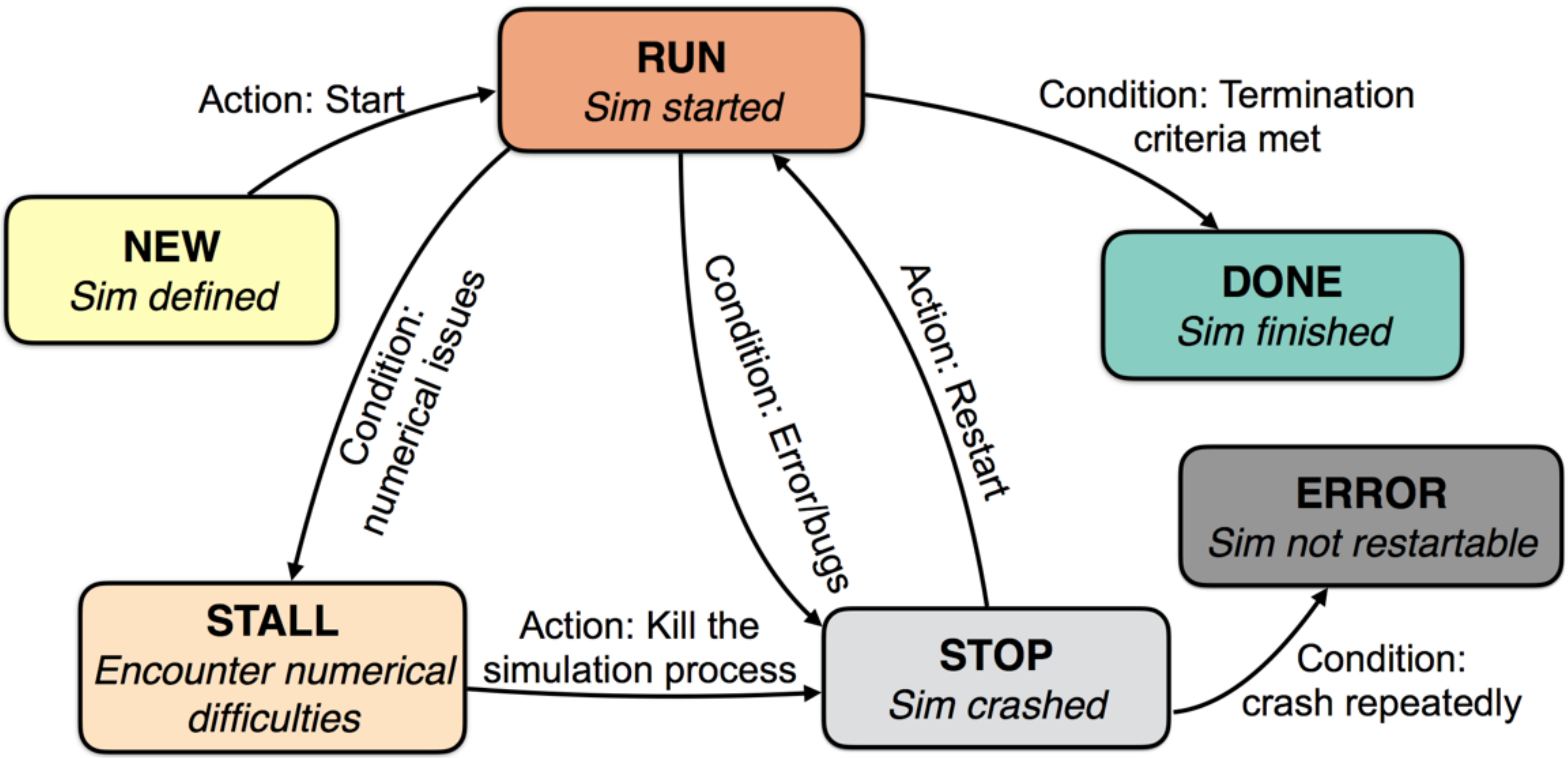
论文提纲

1. 研究背景与意义
2. 可视化分析模块的设计与实现
3. 批量任务管理工具的开发
4. FAST数据产品的三维展示
5. 总结与展望

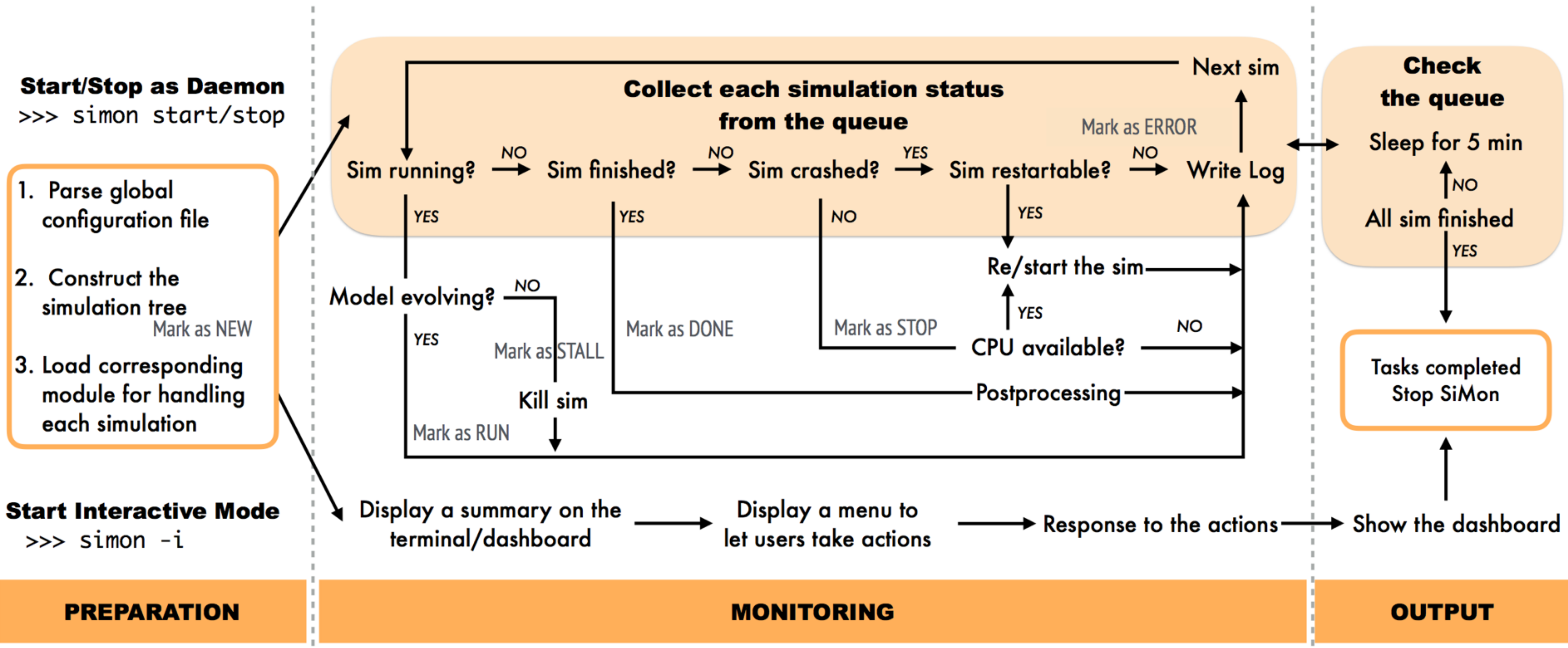
背景介绍



设计：任务抽象

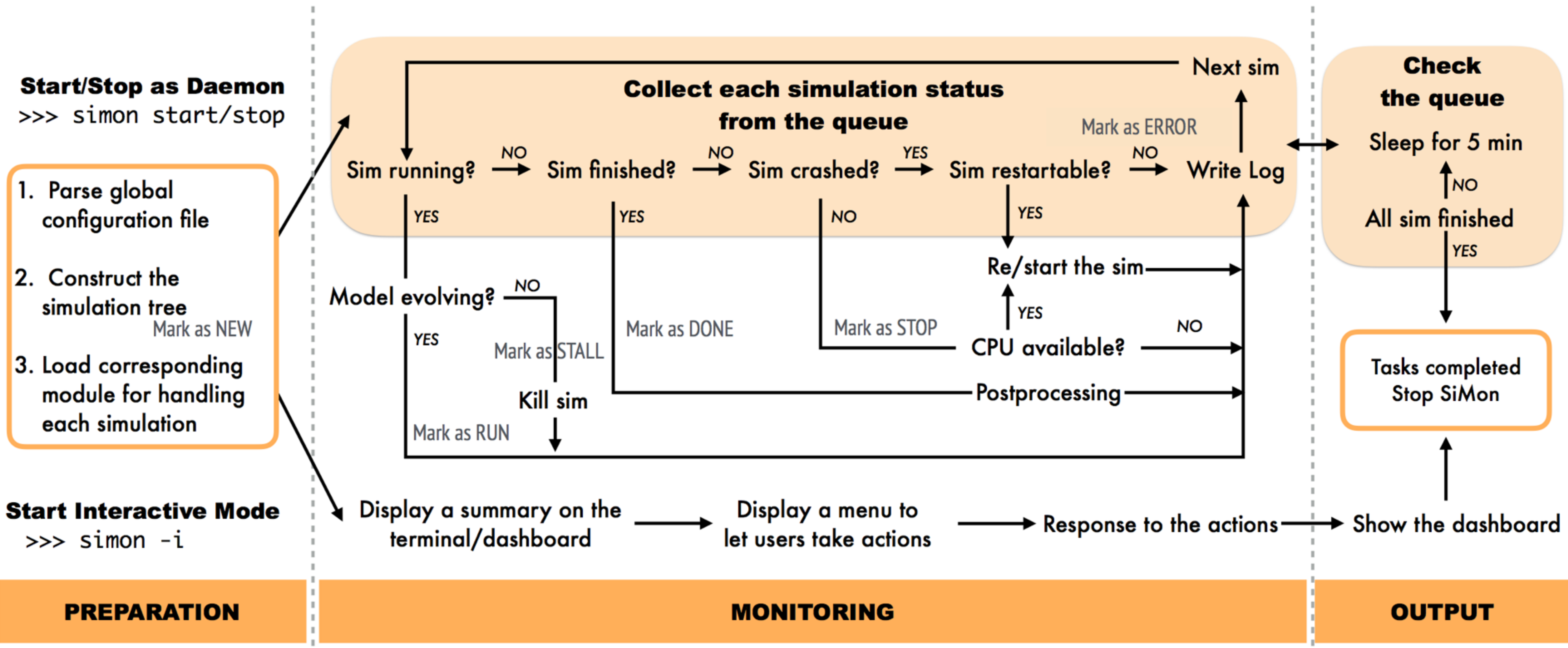

RUN
Sim started
Action: Start
NEW
Sim defined
Condition: Termination criteria met
DONE
Sim finished
Condition: numerical issues
Condition: Error/bugs
Action: Restart
ERROR
Sim not restartable
STALL
Encounter numerical difficulties
Action: Kill the simulation process
STOP
Sim crashed
Condition: crash repeatedly

设计：逻辑决策树



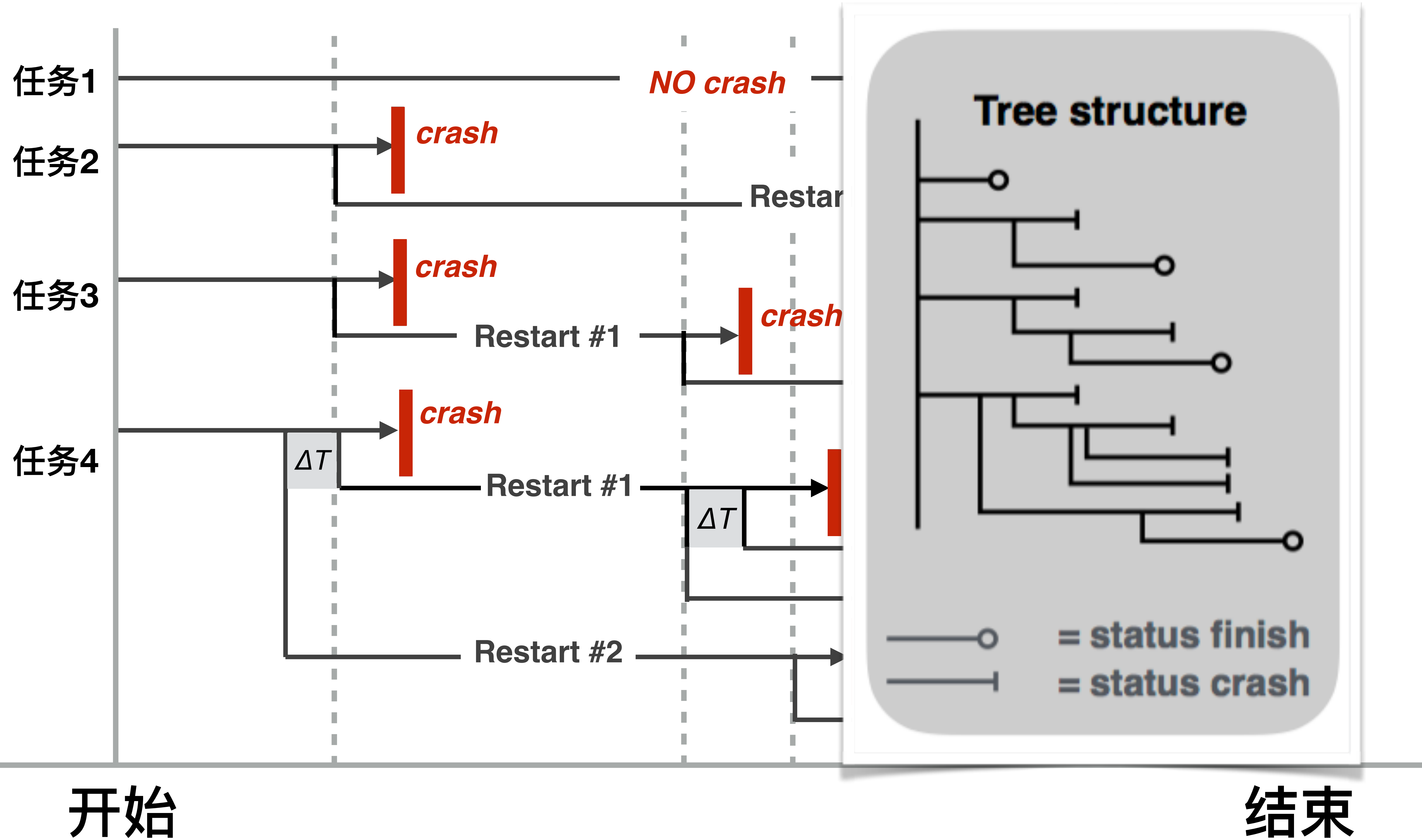
准备阶段　　监控与调度　　输出

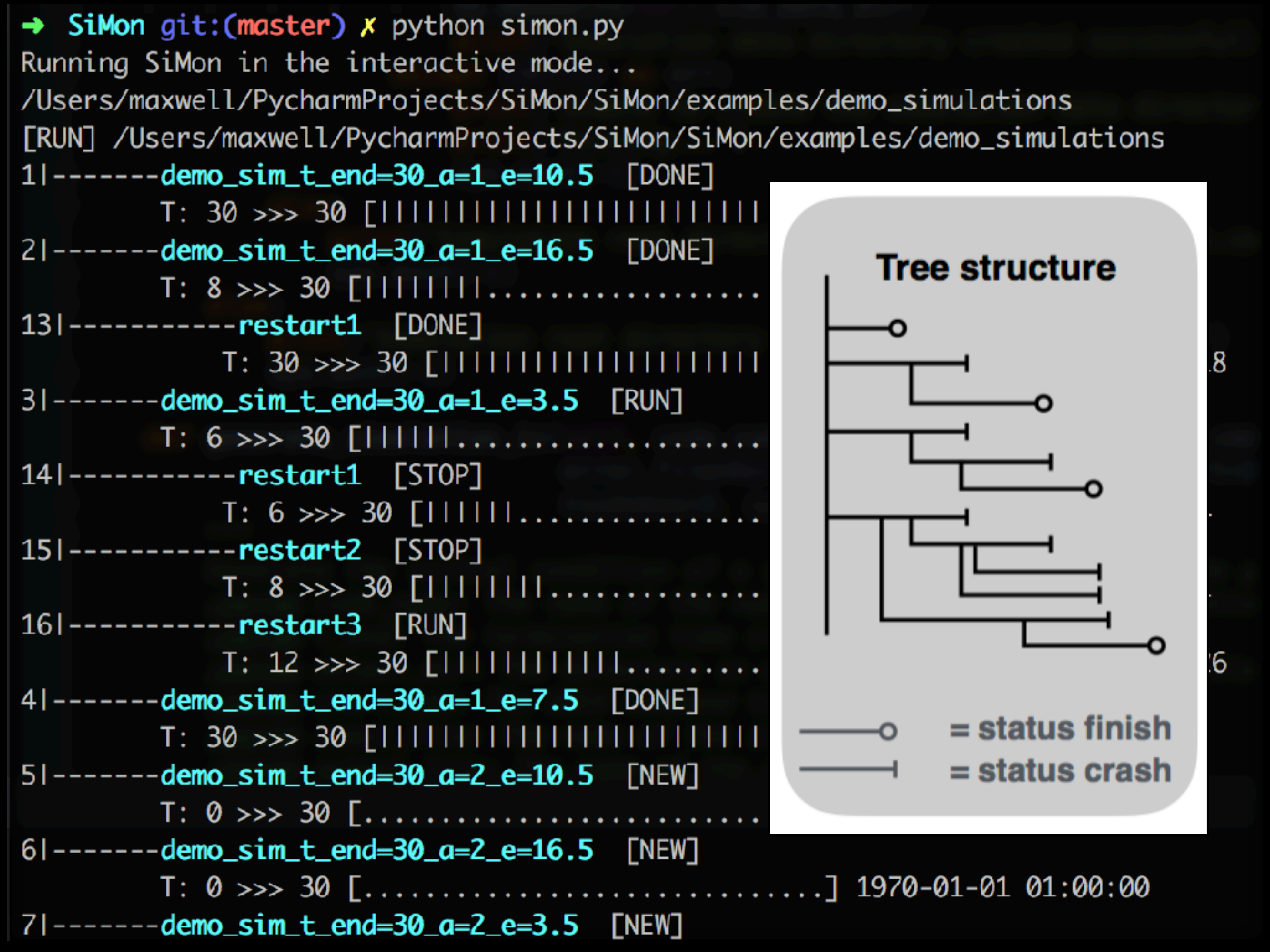
设计：逻辑决策树



准备阶段　　监控与调度　　输出

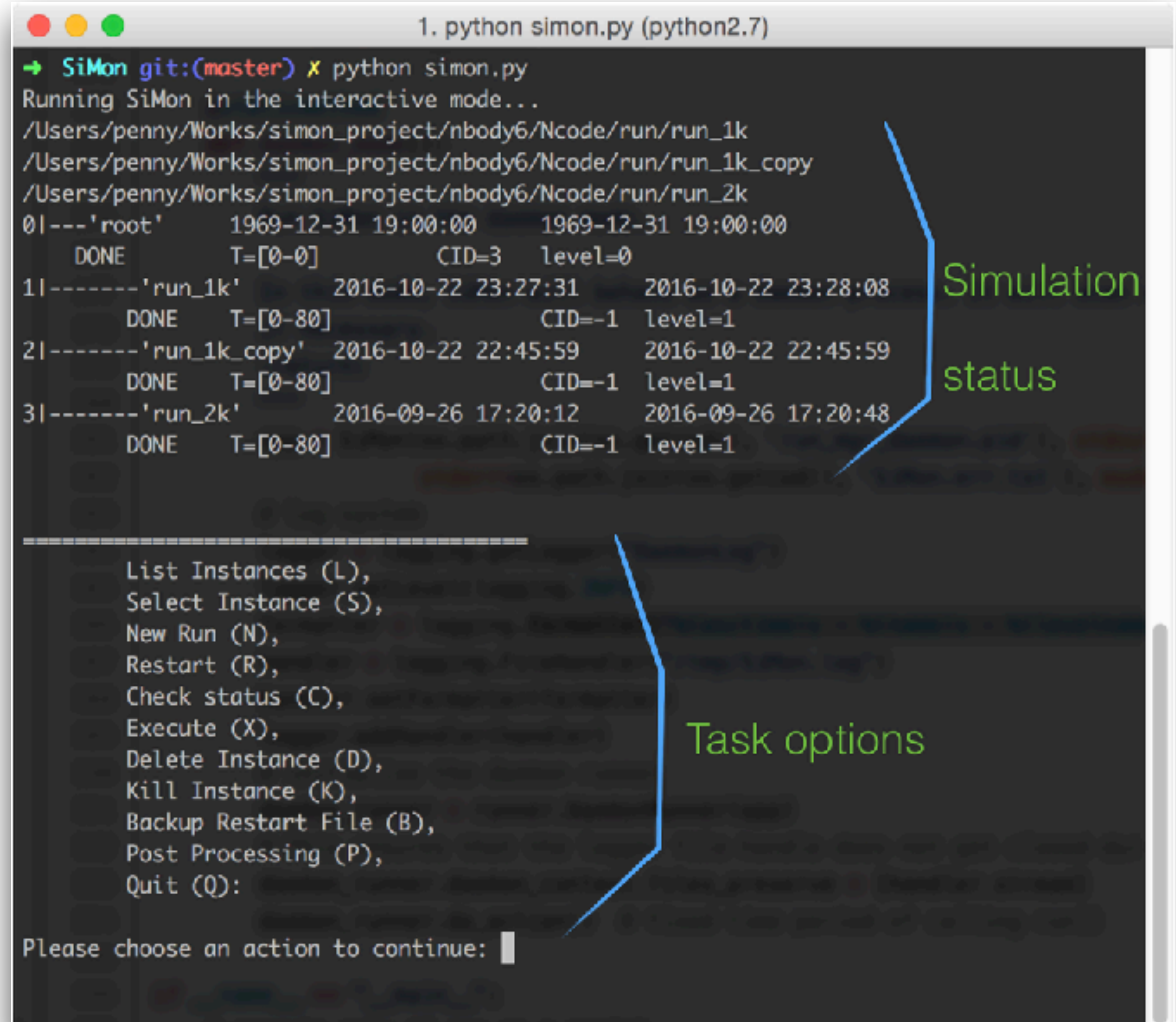
设计：任务集合的层次拓扑图



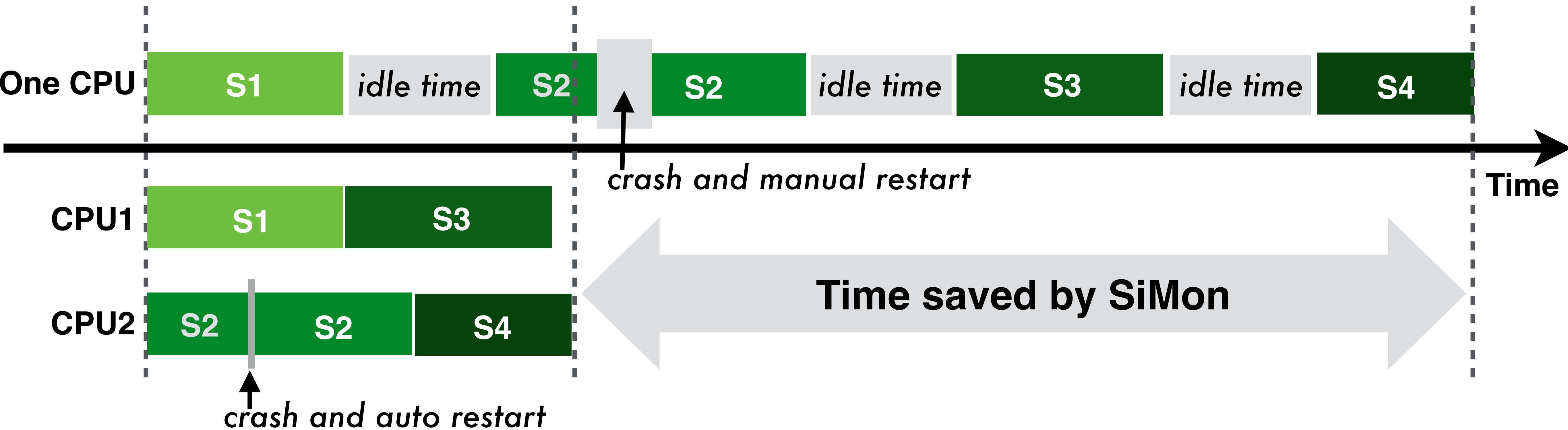

➜ SiMon git:(master) ✗ python simon.py
Running SiMon in the interactive mode...
/Users/maxwell/PycharmProjects/SiMon/SiMon/examples/demo_simulations
[RUN] /Users/maxwell/PycharmProjects/SiMon/SiMon/examples/demo_simulations
1|-------demo_sim_t_end=30_a=1_e=10.5 [DONE]
T: 30 >>> 30 [||||||||||||||||||||||||
2|-------demo_sim_t_end=30_a=1_e=16.5 [DONE]
T: 8 >>> 30 [||||||||.....................
13|-----------restart1 [DONE]
T: 30 >>> 30 [||||||||||||||||||||||
3|-------demo_sim_t_end=30_a=1_e=3.5 [RUN]
T: 6 >>> 30 [||||||......................
14|-----------restart1 [STOP]
T: 6 >>> 30 [||||||.................
15|-----------restart2 [STOP]
T: 8 >>> 30 [||||||||...............
16|-----------restart3 [RUN]
T: 12 >>> 30 [||||||||||||.........
4|-------demo_sim_t_end=30_a=1_e=7.5 [DONE]
T: 30 >>> 30 [||||||||||||||||||||||||
5|-------demo_sim_t_end=30_a=2_e=10.5 [NEW]
T: 0 >>> 30 [..........................
6|-------demo_sim_t_end=30_a=2_e=16.5 [NEW]
T: 0 >>> 30 [..............................] 1970-01-01 01:00:00
7|-------demo_sim_t_end=30_a=2_e=3.5 [NEW]
Tree structure
= status finish
= status crash

实现：运行模式

- 后台模式
- 交互模式



实现：自动调度

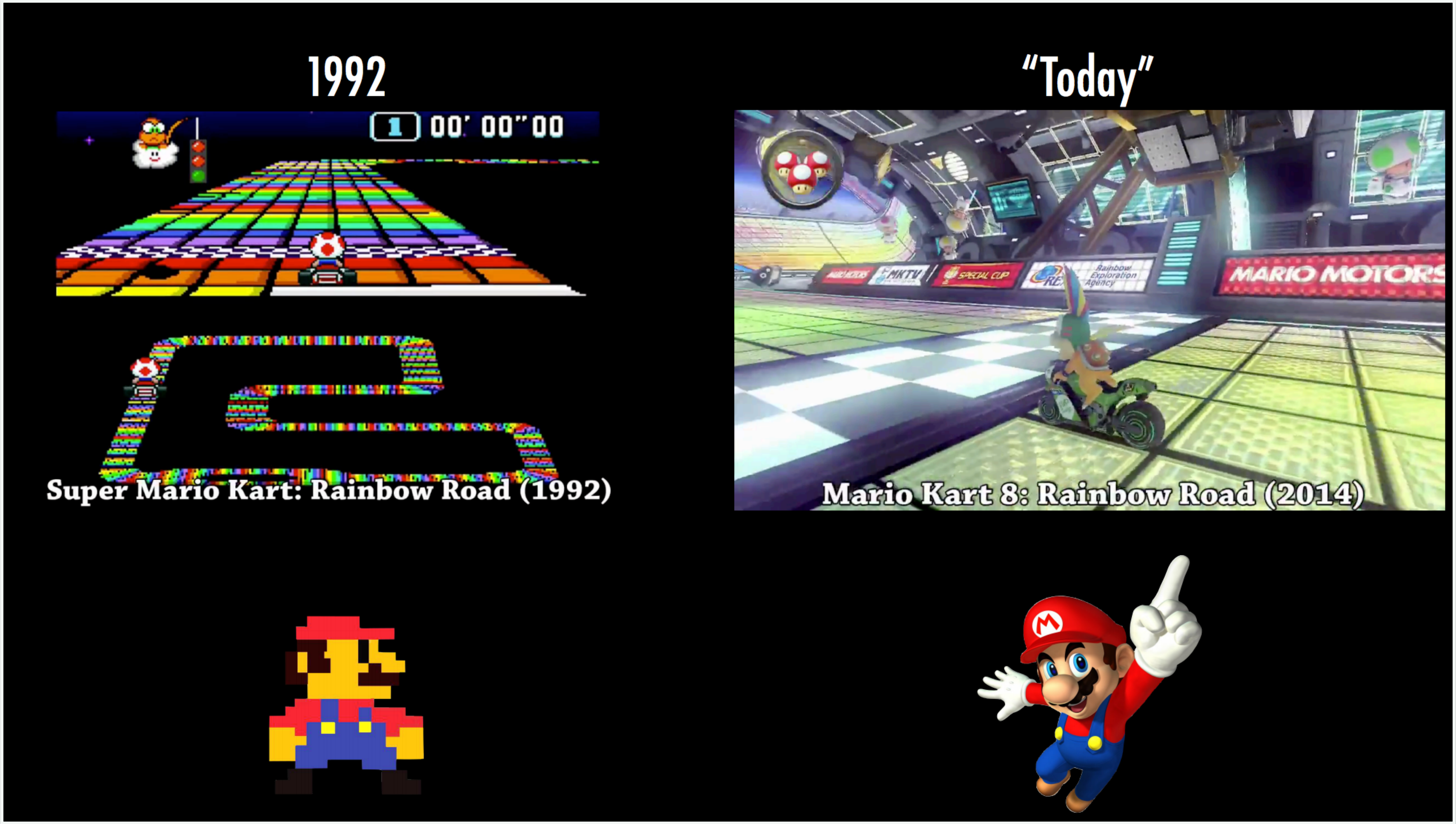


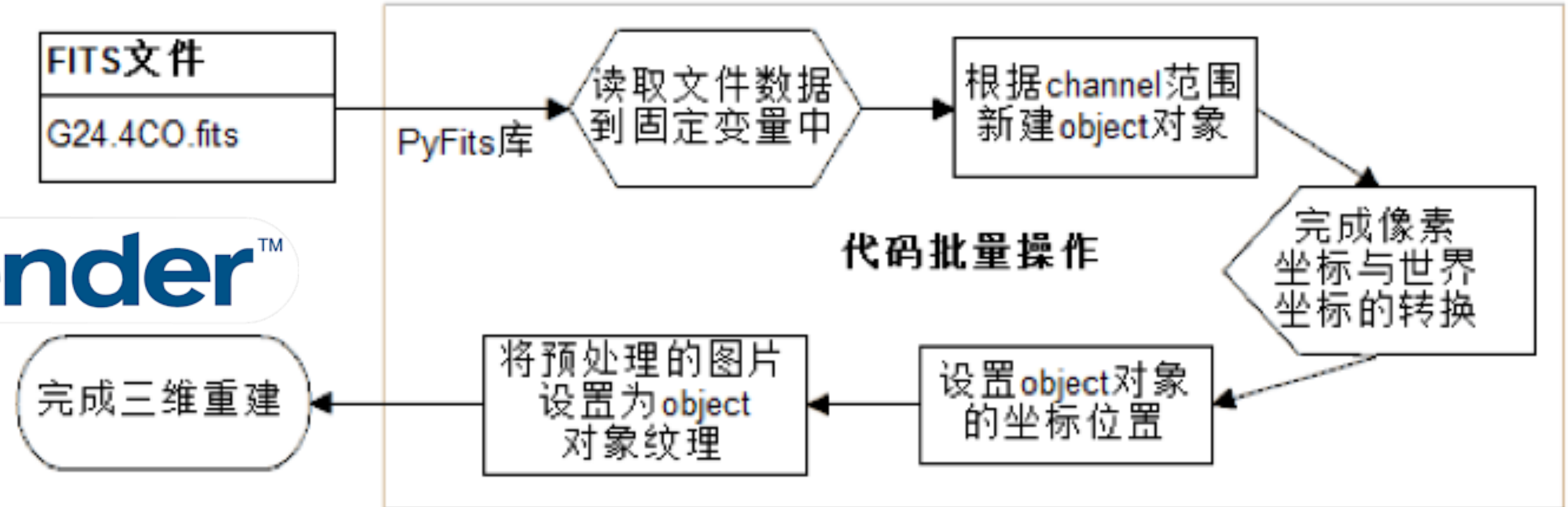
- 数值模拟任务管理
- FAST脉冲星数据处理

论文提纲

1. 研究背景与意义
2. 可视化分析模块的设计与实现
3. 批量任务管理工具的开发
4. FAST数据产品的三维展示
5. 总结与展望

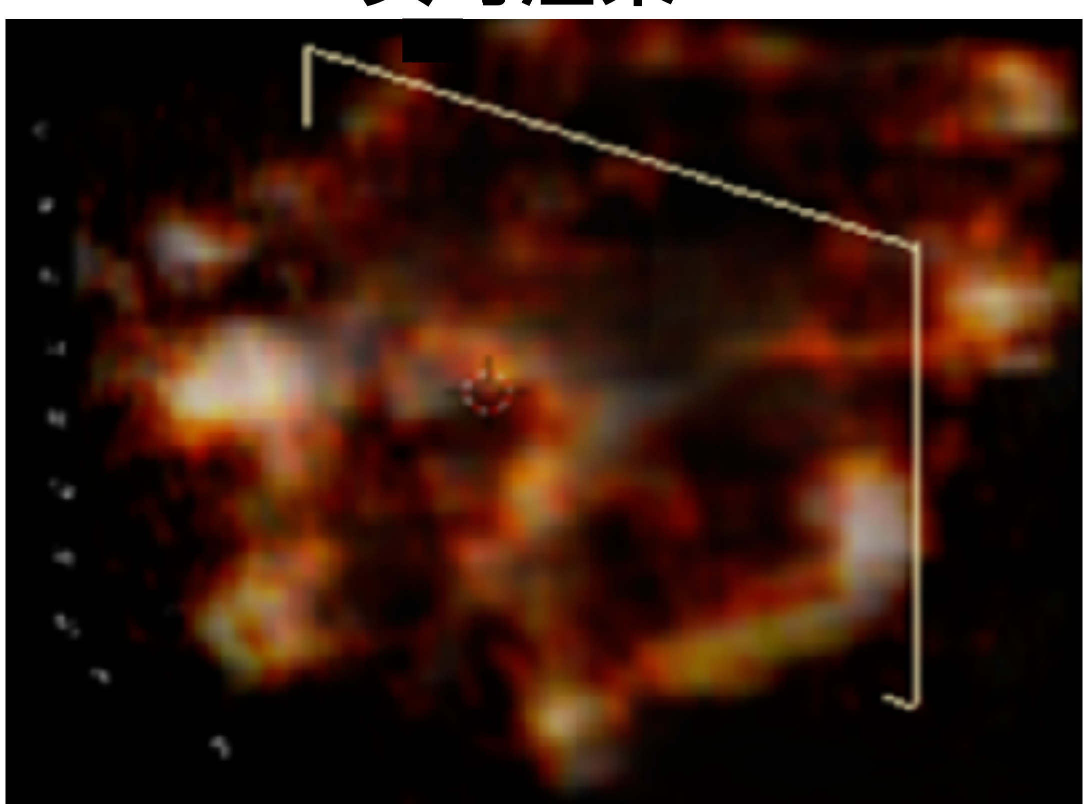
研究目的


1992
“Today”
00’ 00” 00
MARIO MOTORS
Super Mario Kart: Rainbow Road (1992)
Mario Kart 8: Rainbow Road (2014)

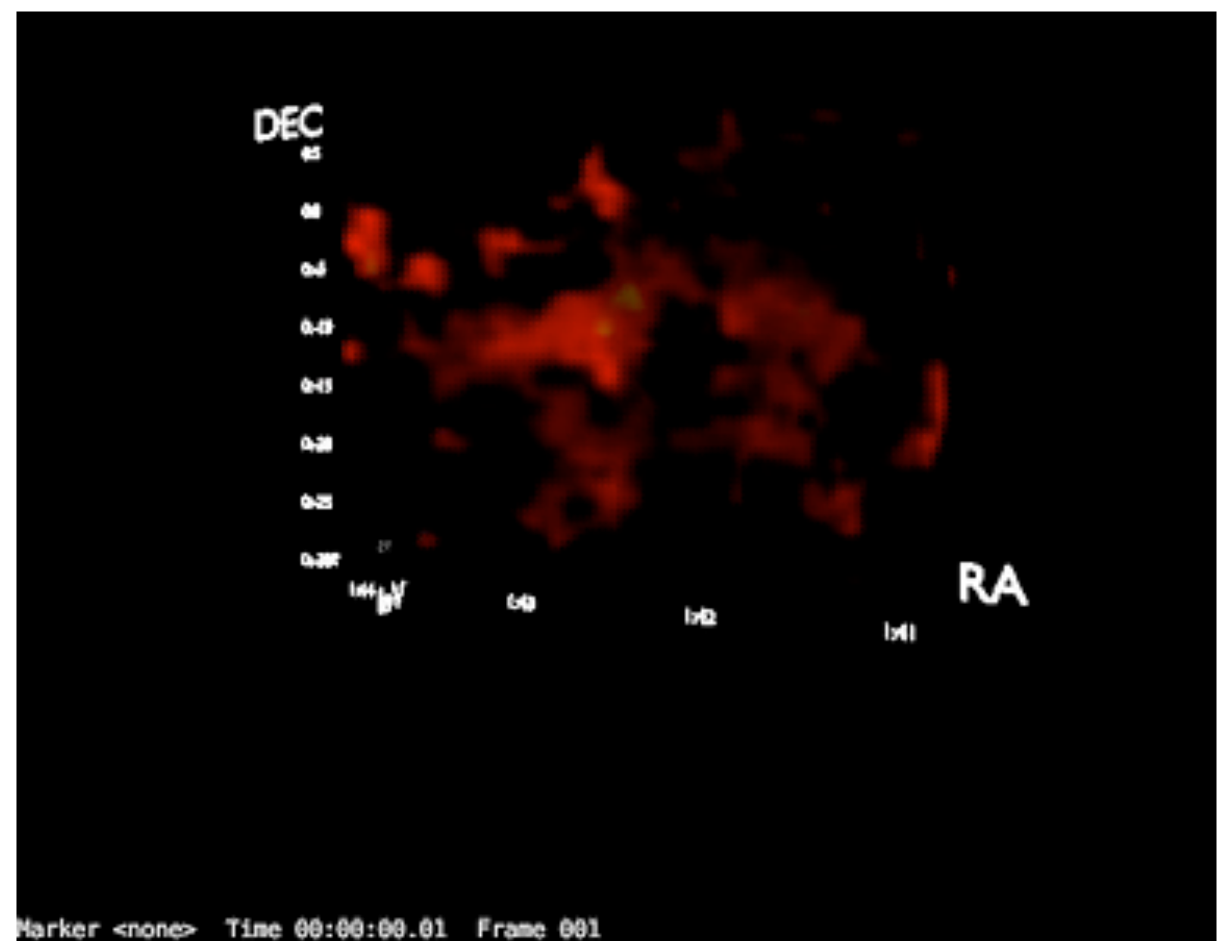


企业级图像软件

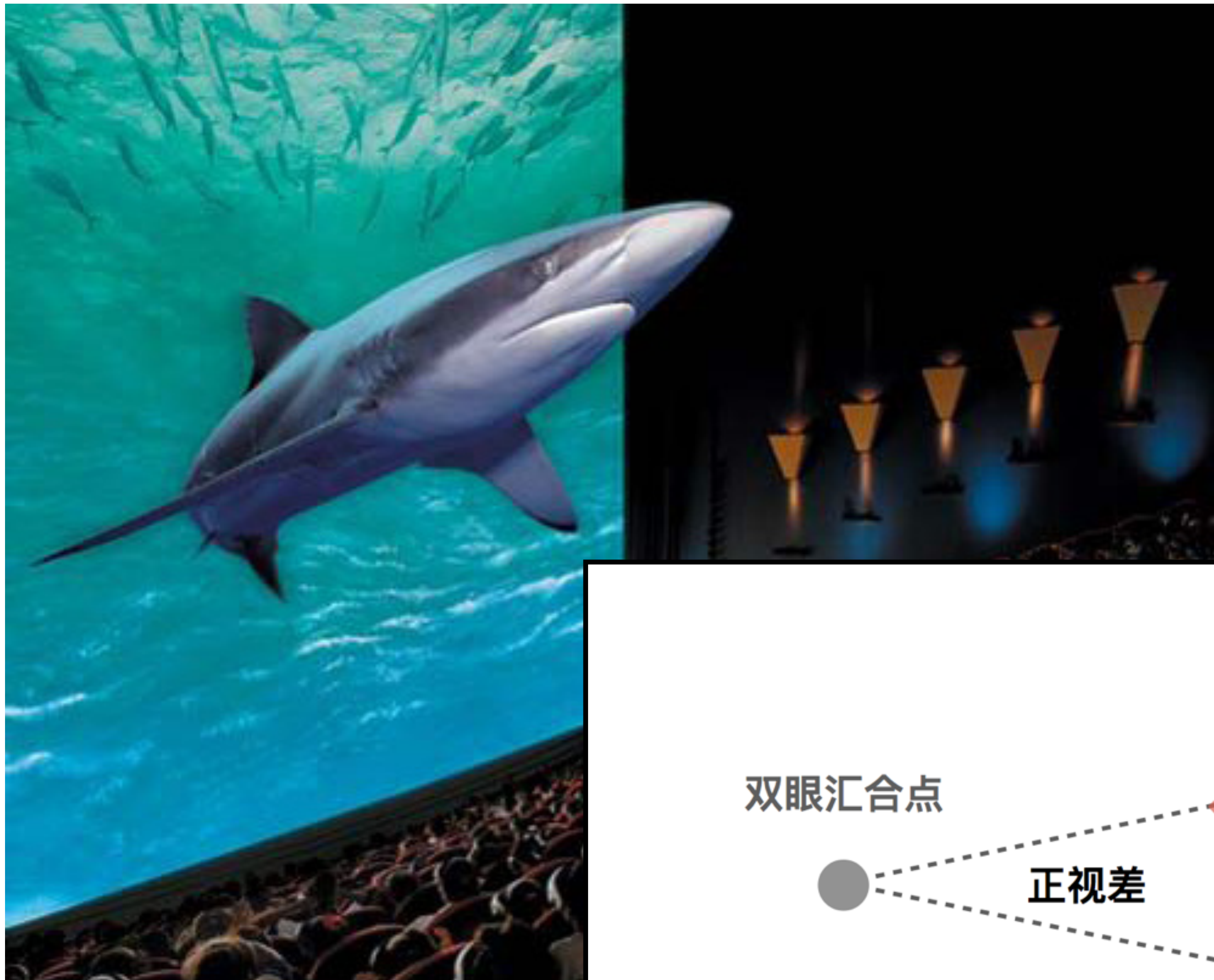
实时渲染

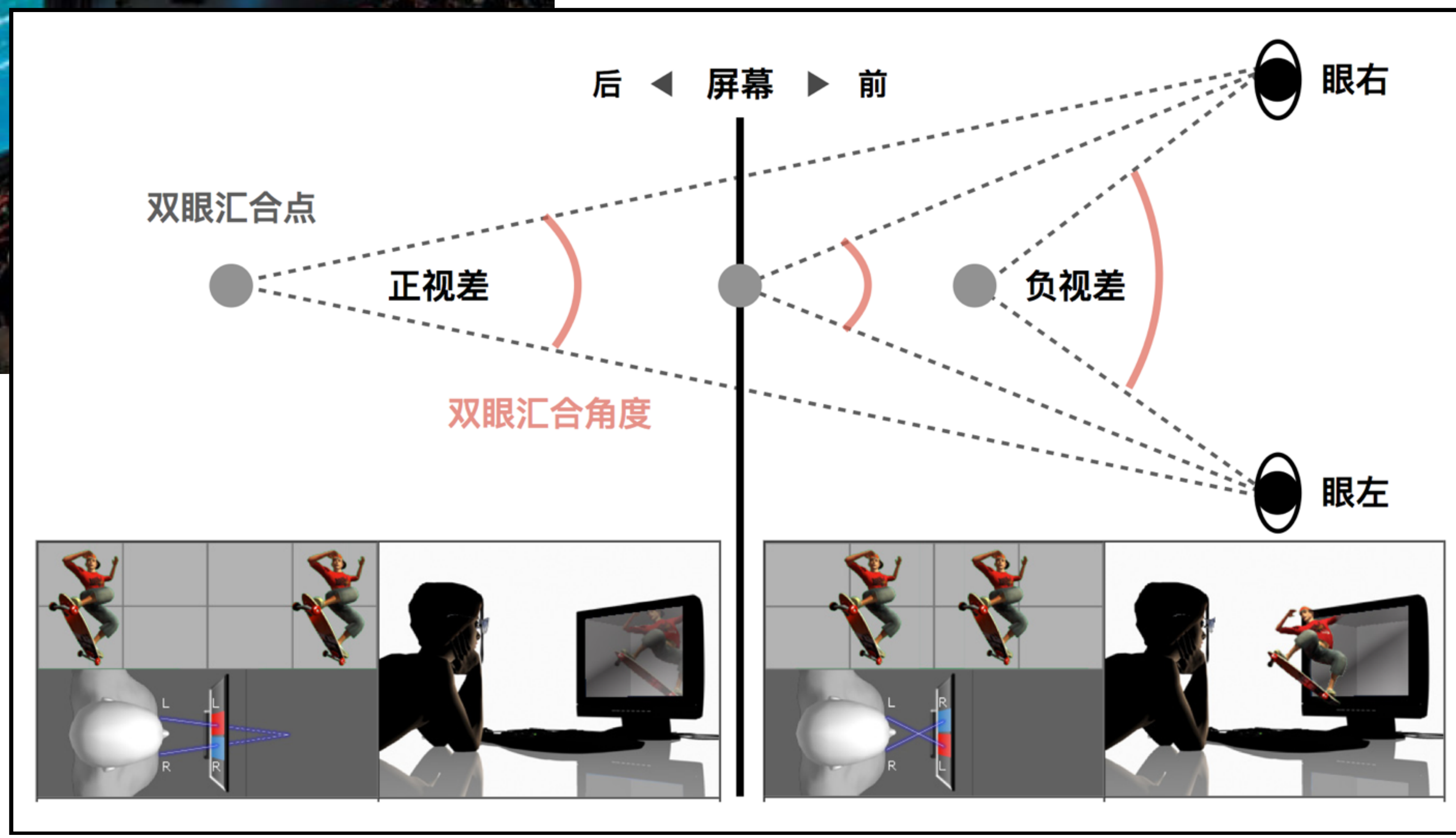


动画输出



立体显影



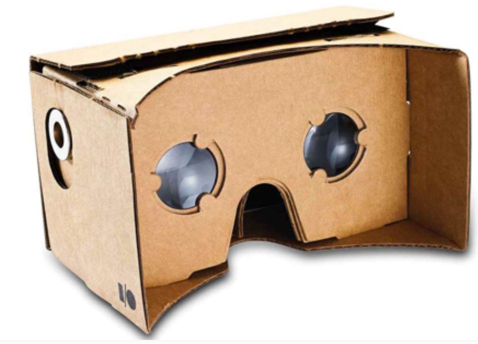


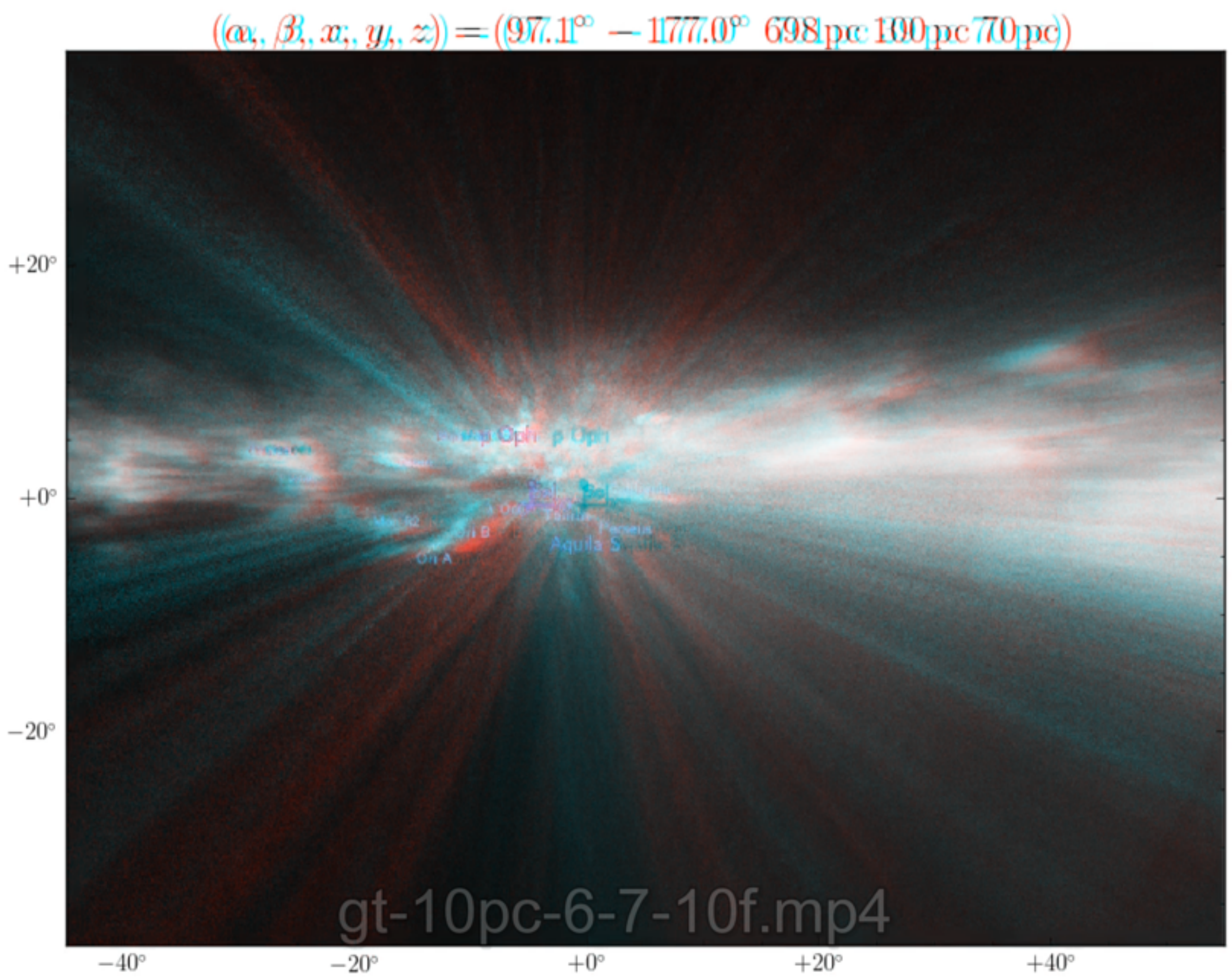
source: reallusion.com

立体显影实现

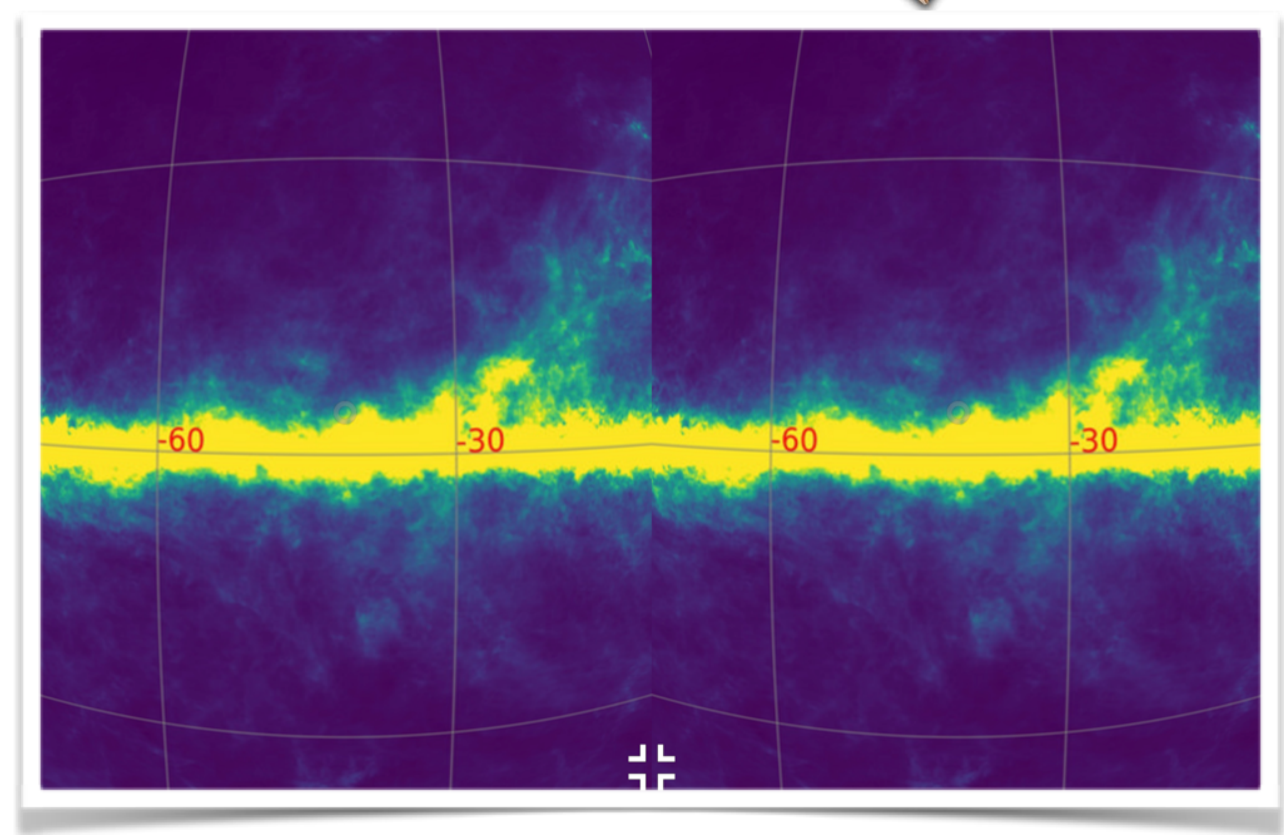
三维立体图片
Anaglyph 影像

VR眼镜
360度全景图



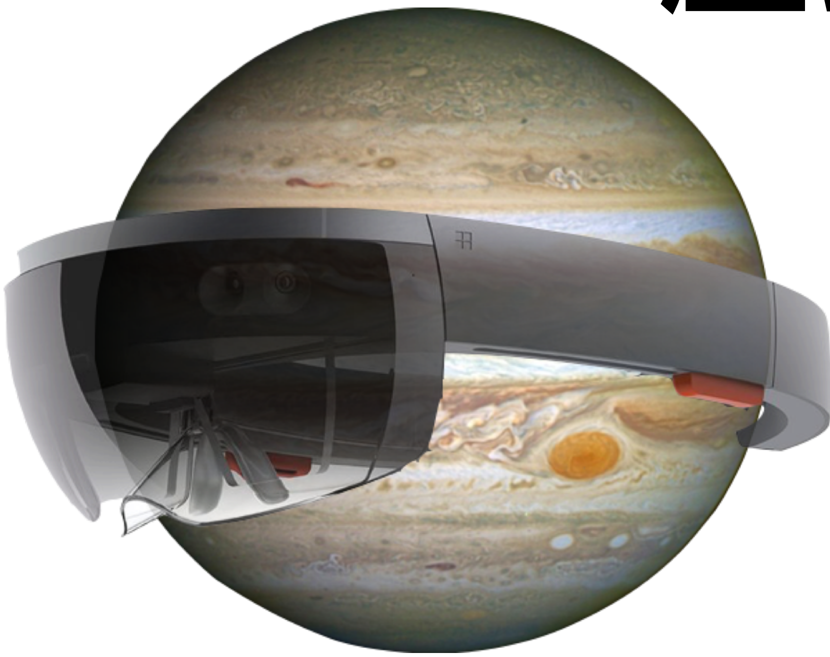


Pan-STARRS1(PS1) 巡天和*The Two Micron All-Sky Survey(2MASS)* 巡天的星际尘埃红化的三维天图数据

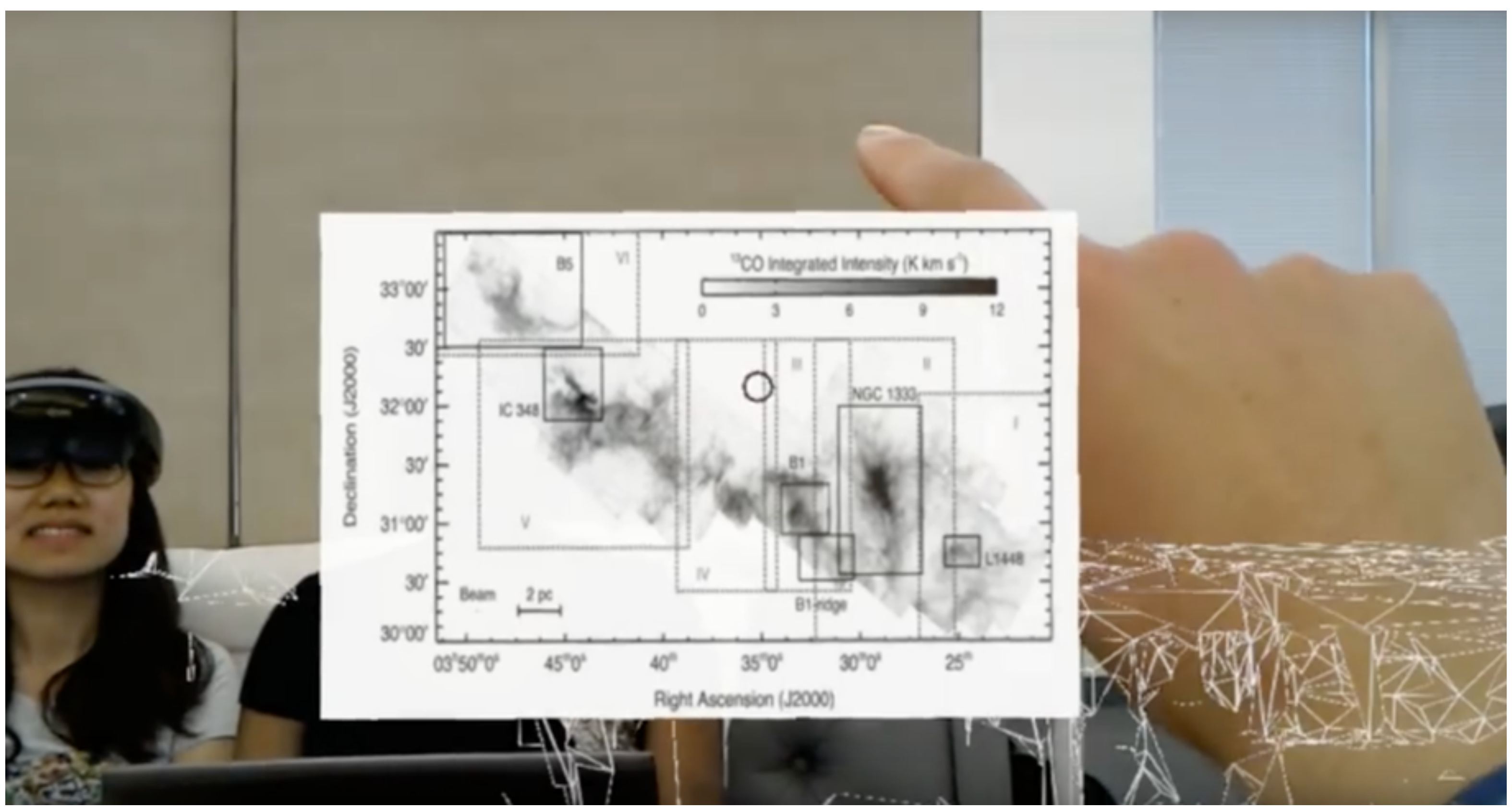


尘埃三维天图数据的*SDF* 图像

虚拟现实+交互



- 加载天文数据到Unity 环境下
- 三维体渲染 - Ray Casting 算法



混合现实显示

声音+手势交互

论文提纲

1. 研究背景与意义
2. 可视化分析模块的设计与实现
3. 批量任务管理工具的开发
4. FAST数据产品的三维展示
5. 总结与展望

论文名称：FAST 数据处理系统可视化分析与自动化管理技术研究

- **FAST数据产品和数据处理工作流**
- **高维度复杂数据的可视化分析**
- **FAST可视化分析模块**
- **批量任务管理**
- **FAST三维可视化应用**

展望：人工智能，云平台

致谢

朱明导师，美国导师Alyssa Goodman和Michelle Borkin

美国合作者：Tom Robitaille, Catherine Zucker, Hope Chen

国台老师们：艾华老师，马怀宇老师，杜洪荣老师，张海燕老师

国台小伙伴们：艾美，岳楠楠

更多…

谢谢！