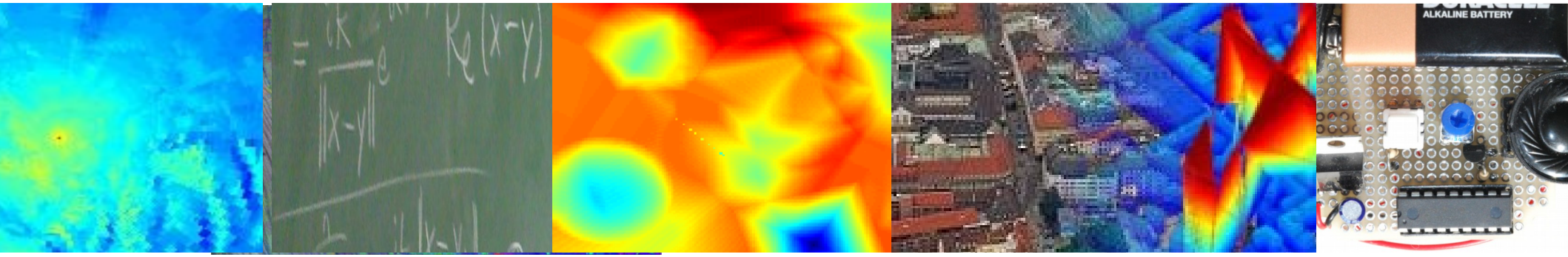
Sheaf Theory: The Mathematics *of* Data Fusion



Michael Robinson

DARPA SIMPLEX Program

Acknowledgements

Reza Ghanadan (DARPA/DSO)

Collaborators:

- Cliff Joslyn & Emilie Hogan
  - Semantic technologies
  - Discrete mathematics
- Chris Capraro
  - Signal processing system design
  - High performance computing
- Other SIMPLEX teams: Georgetown University, Baylor Medical College, Vanderbilt University, MIT, ...

Logistics

- What's where?
- Discussion areas
- Wireless access
- Streaming, archiving
  - Lecture “ground rules”
- How the interactive sessions work

Tutorial schedule

	Tuesday, August 25	Wednesday, August 26
8:00-9:00am	Registration	
9:00-9:15am	Opening remarks	
9:15-10:00am	Sheaf theory: the Mathematics of Data Fusion	Categorification and Chain Complexes
10:15-11:00am	Topological Features	Constructing and Categorifying
11:15-12:00pm	What is Topology?	Computing Topological Features
12:00-1:30pm	Lunch	
1:30-2:15pm	What is a Sheaf?	Sheaf Cohomology and its Interpretation
2:30-3:15pm	Building Complexes and Sheaves	Computing Homology and Cohomology
3:30-4:15pm	Data Structures as Sheaves	How do we Deal with Noisy Data?

Tutorial objectives

- What are sheaves?
  - The “local to global” viewpoint
- Encoding existing data into sheaves
  - “Sheafification”
- Data analytic capabilities enabled by sheaves
  - “Sections,” “cohomology”
- Practice analyzing sheaves in software
  - “Persistence,” “local homology”
- Interpret this analysis into the context of the data

The SIMPLEX project

<u>Problem</u>: Heterogeneous data fusion has been stymied by the lack of effective quantitative – qualitative fusion techniques

<u>Key points</u>:

- *Sheaf-based* heterogeneous fusion is the sole canonical mathematical framework for multi-modal, multi-INT integration systems
- Demonstrate the framework by showing cross-modality inference using *topological invariants*

Heterogeneous integration systems

A mathematical framework for heterogeneous integration should

- Be general enough to accurately represent the data in all its richness
- Be capable of summarization of the data into important, actionable features

<u>Theorem</u>: These two properties lead directly to a *sheaf-based* data representation

- Any other representation recapitulates sheaf theory

Benefits of sheafification

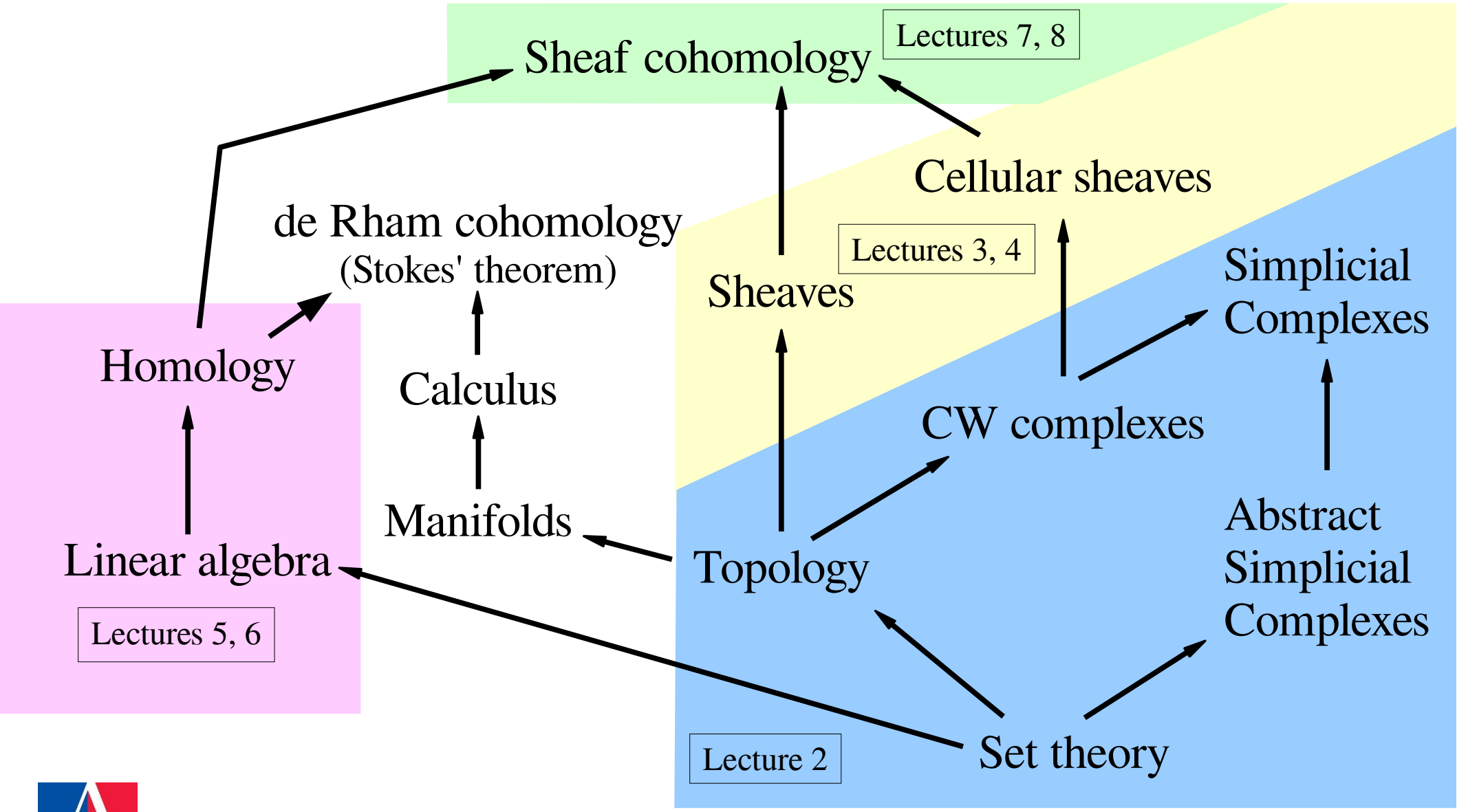
Solid theoretical grounding yields powerful inference capabilities:

- **Globalization** of sensor data
  - Sheaves are the correct theory of data fusion
  - Track generation and disambiguation
- Topological sensor **retasking**
  - Detection of misinformation and deception
  - Detection of “loopy” belief propagation
  - Removal of inferential vulnerabilities and self-inconsistencies

Steps of heterogeneous fusion

1. Form the *base topological space*: multiway interactions between data sources
2. *Sheafify*: build the model of relations between data sources
3. *Categorify*: place the data streams in vector spaces to aid in computation and analysis
4. Compute *cohomology*: globalize the data to find robust invariants

Mathematical dependency tree



Session objectives

- Provide an overview of topological data fusion
- When does a problem involve “topological” features?
- When do sheaves add a new capability beyond topology?

What are topological features?

- Clusters
- Holes
- Voids
- Boundaries
- Dimension
- Local “model” spaces

Sheafification

Getting your data into a *sheaf*

(and what's that anyway?)

Michael Robinson

What is a sensor?

- Anything that provides a measurement
- It could be something that records an **image**...

What is a sensor?

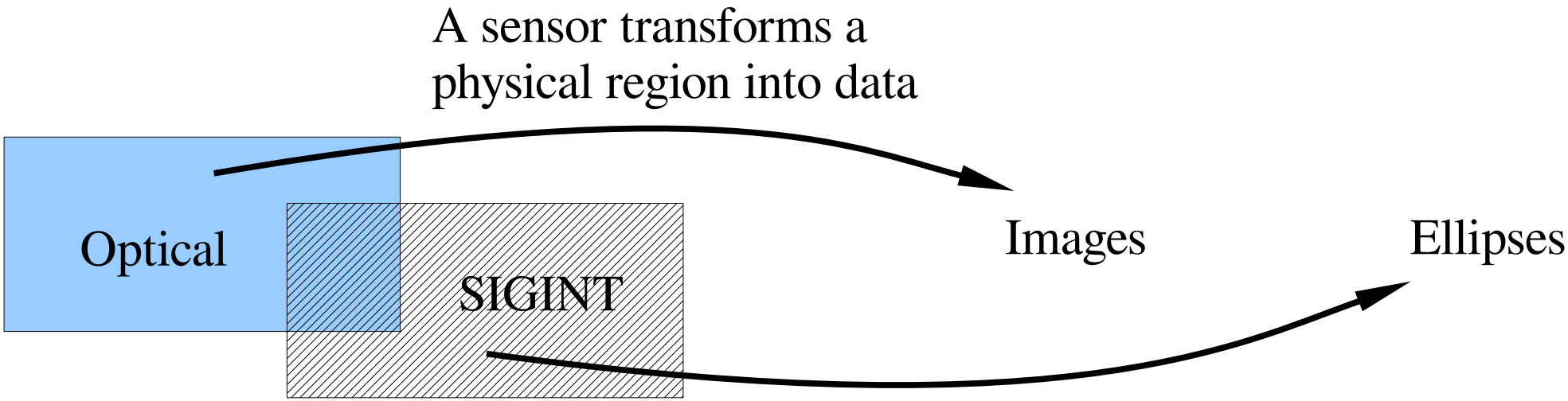
- Anything that provides a measurement
- … or **a pixel** ...

What is a sensor?

- Anything that provides a measurement
- … or even some **text**

```
"21 train cars, two of
which are yellow"
```

Sheaf: a heterogeneous sensor system

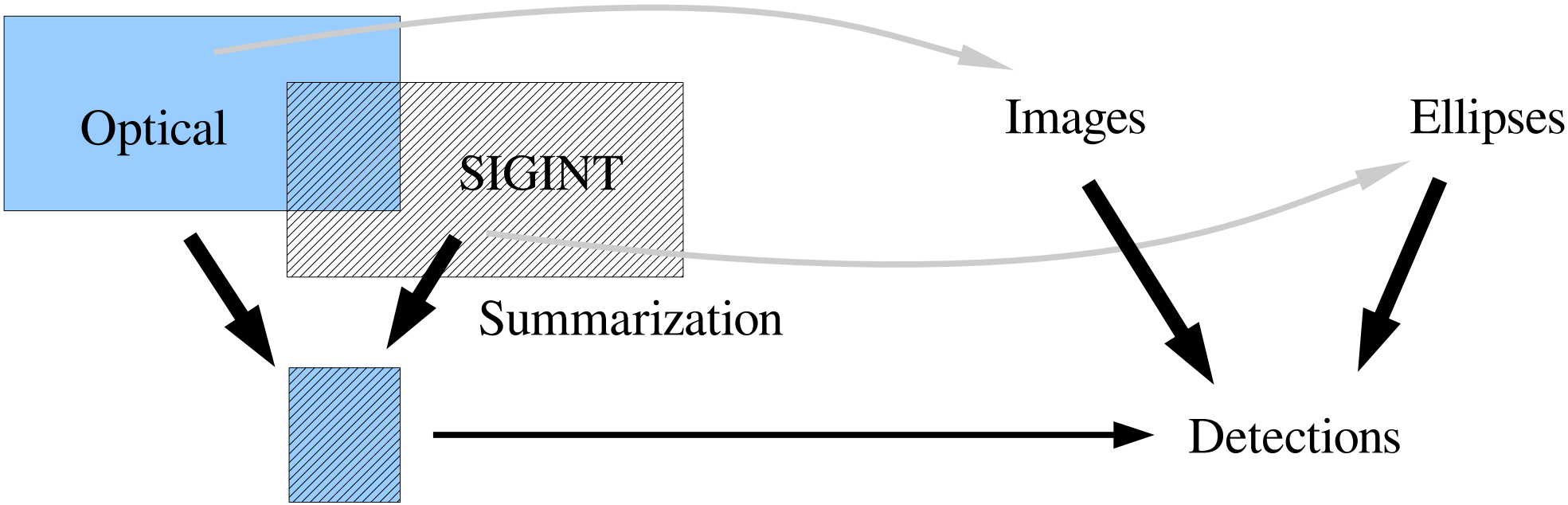


“Physical” sensor footprints

Sensor data space

Michael Robinson

Sheaf: a heterogeneous sensor system



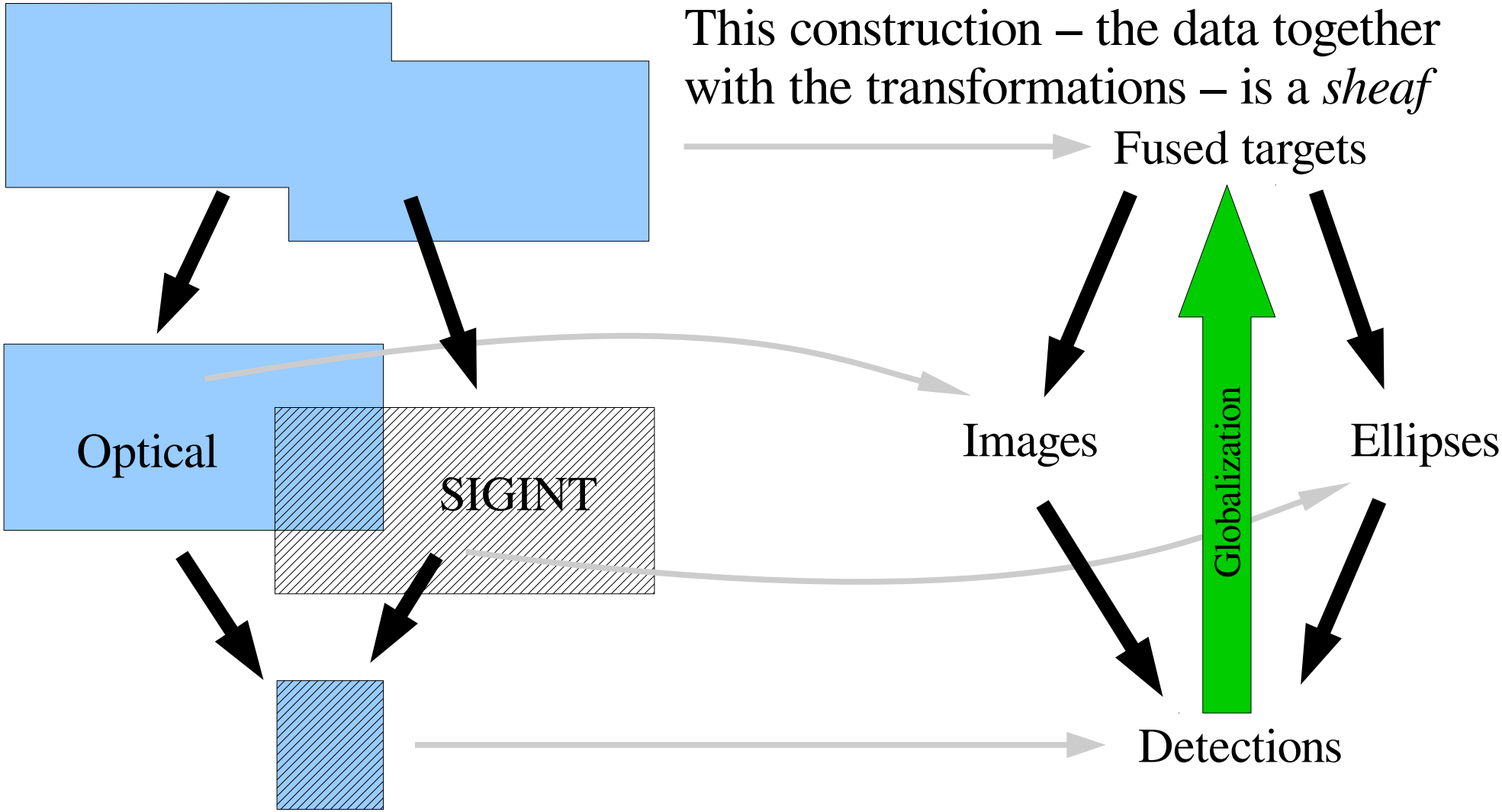
"Physical" sensor footprints

Sensor data space

Michael Robinson

Sheaf: a heterogeneous sensor system

This construction – the data together with the transformations – is a *sheaf*



"Physical" sensor footprints

Sensor data space

Michael Robinson

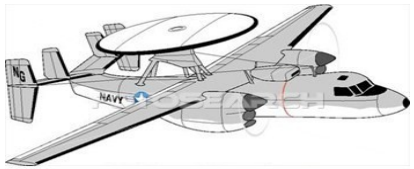
Globalizing data in a sheaf

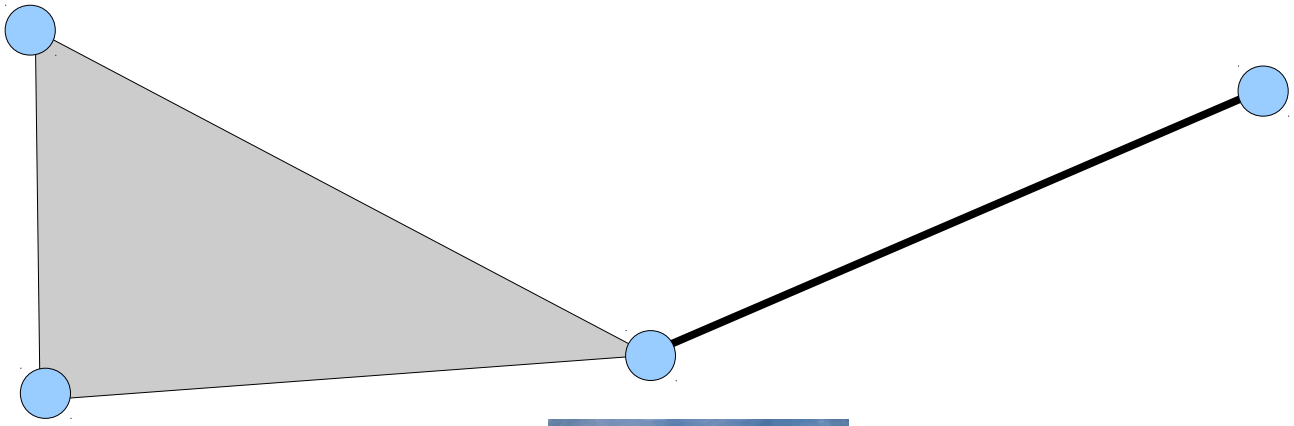
Multi-INT tracking

- Combine measurements of a scene from **multiple sensors** returning **arbitrary data types** to make inferences about the scene

Topological fusion

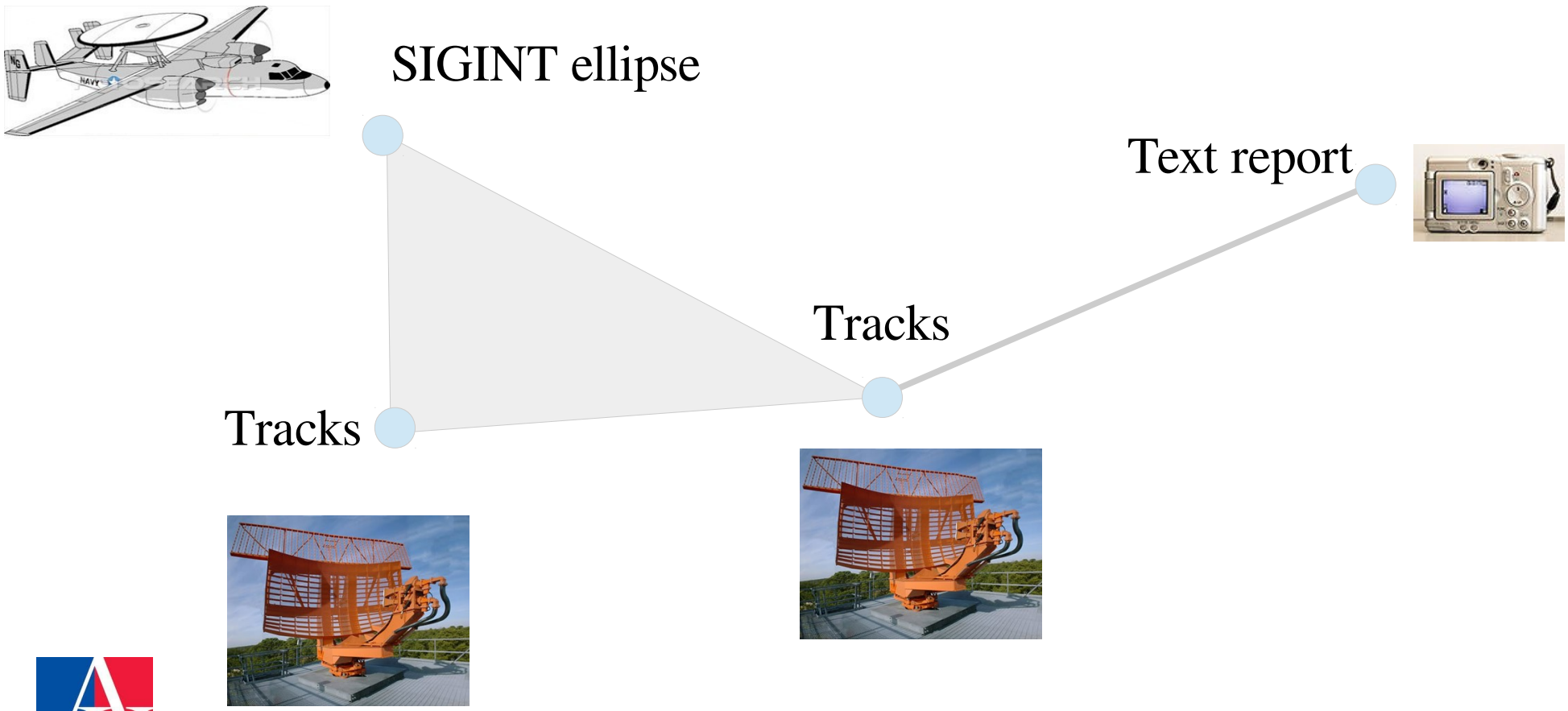
- **Topologize** the space of sensors, abstractly





Topological fusion

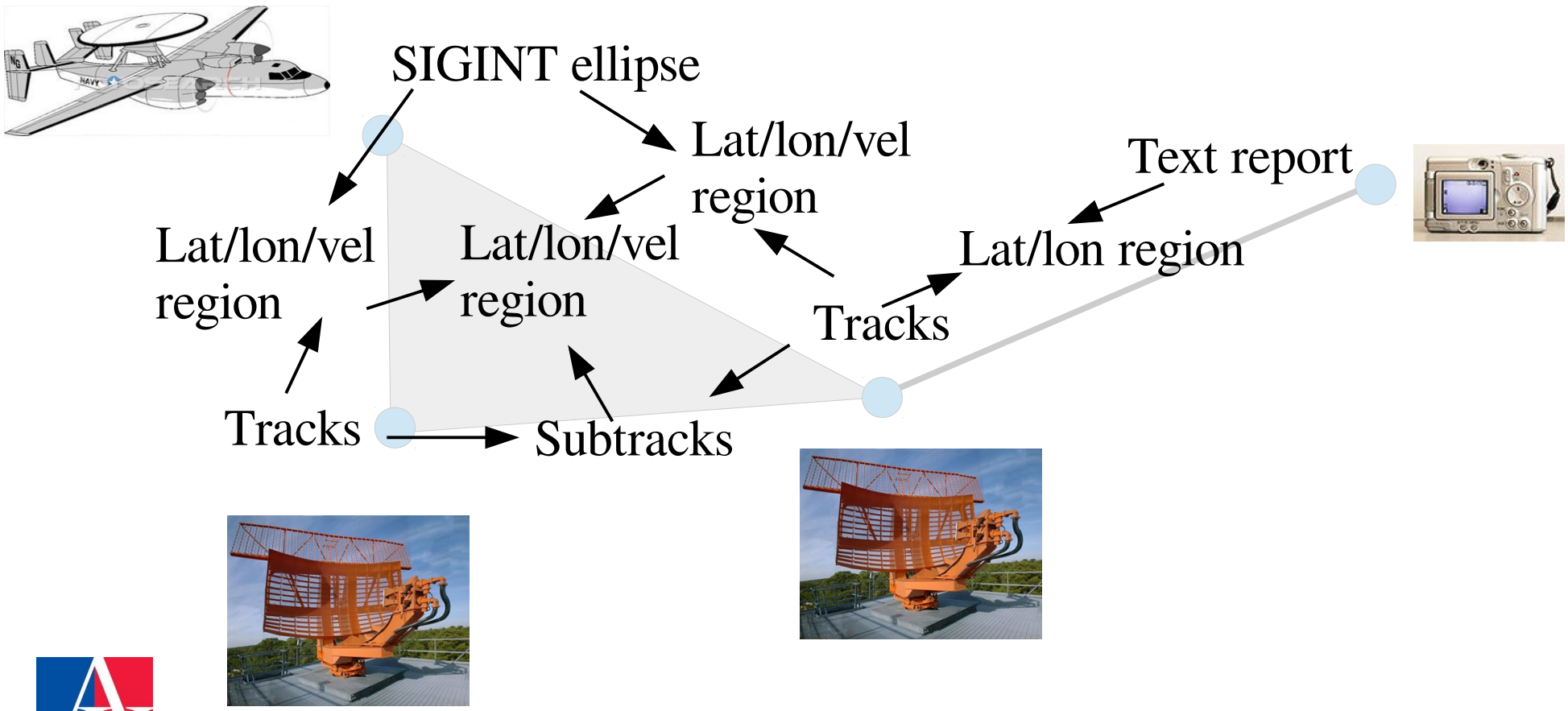
- **Encode** the data collected by each sensor



Michael Robinson

Topological fusion

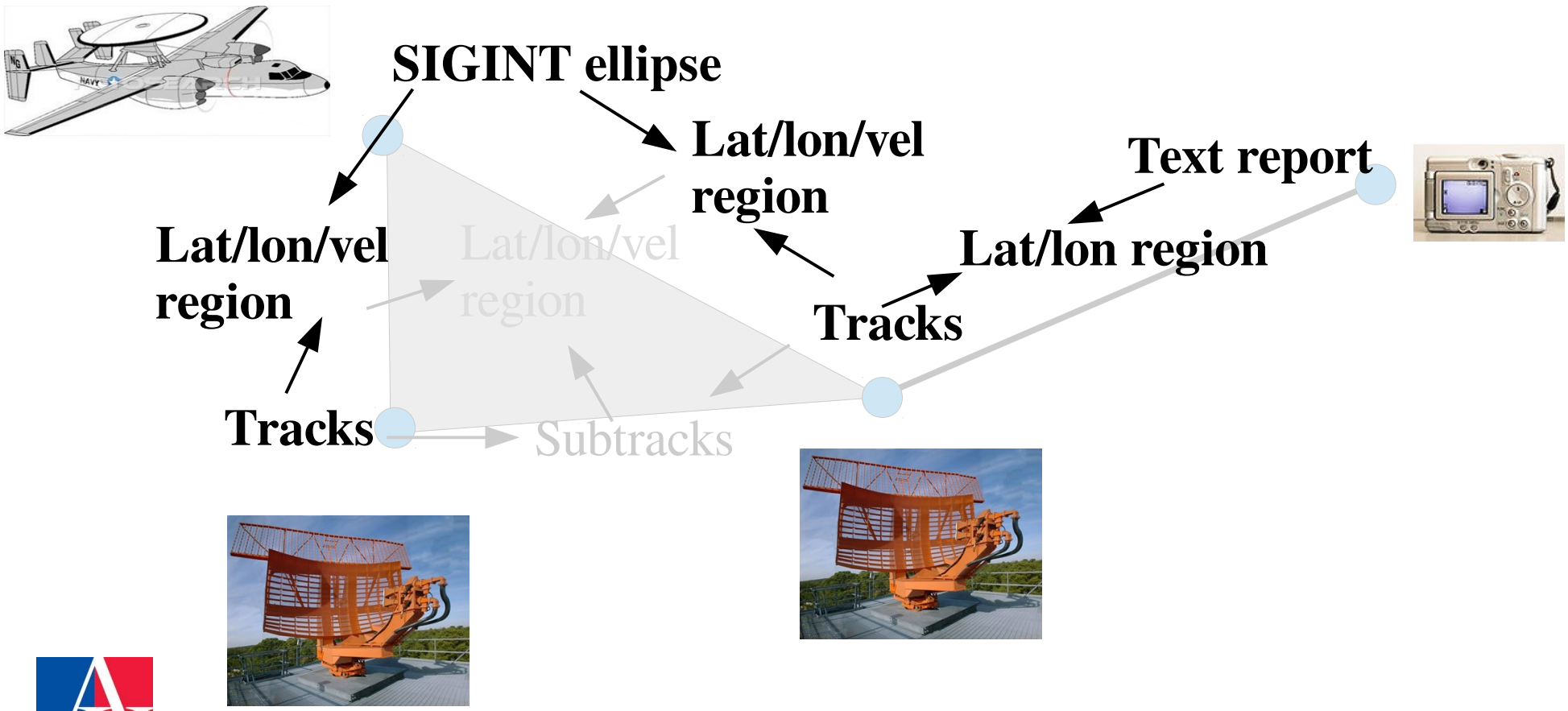
- Express **consistency requirements** for adjacent sensors



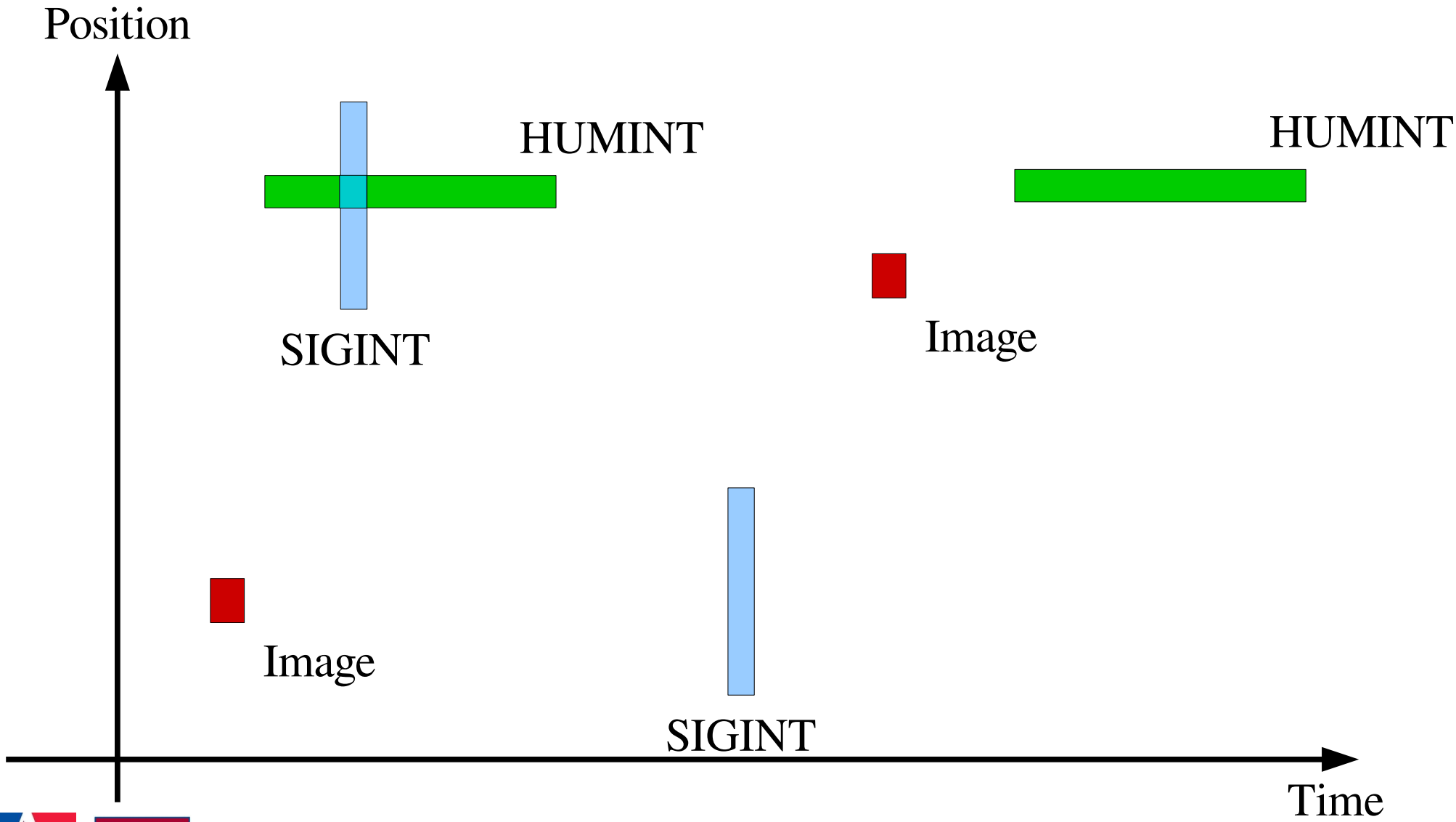
Michael Robinson

Globalization of data

- Finding **maximally consistent information** through this network amounts to **tracking**

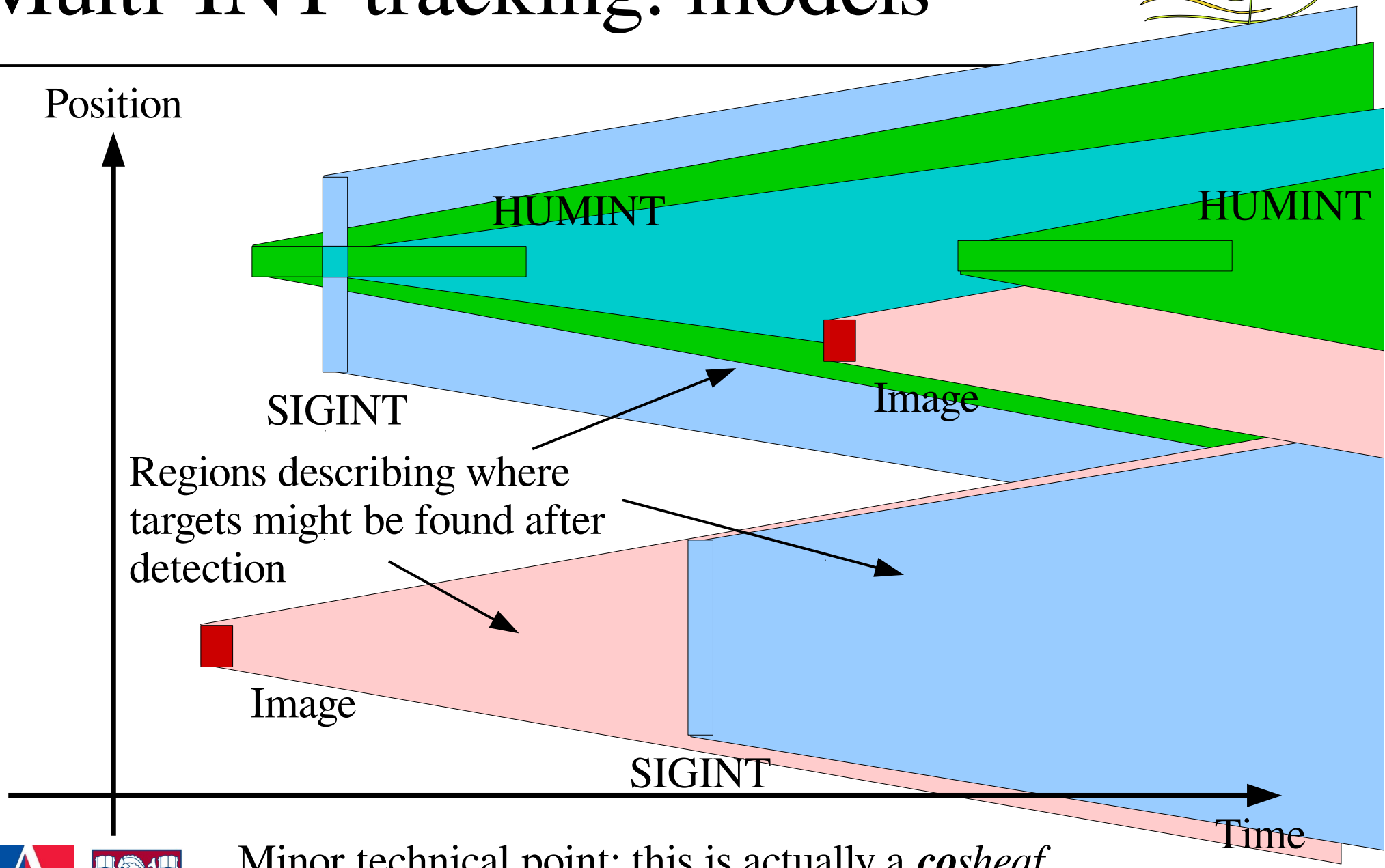


Multi-INT tracking: input data



Michael Robinson

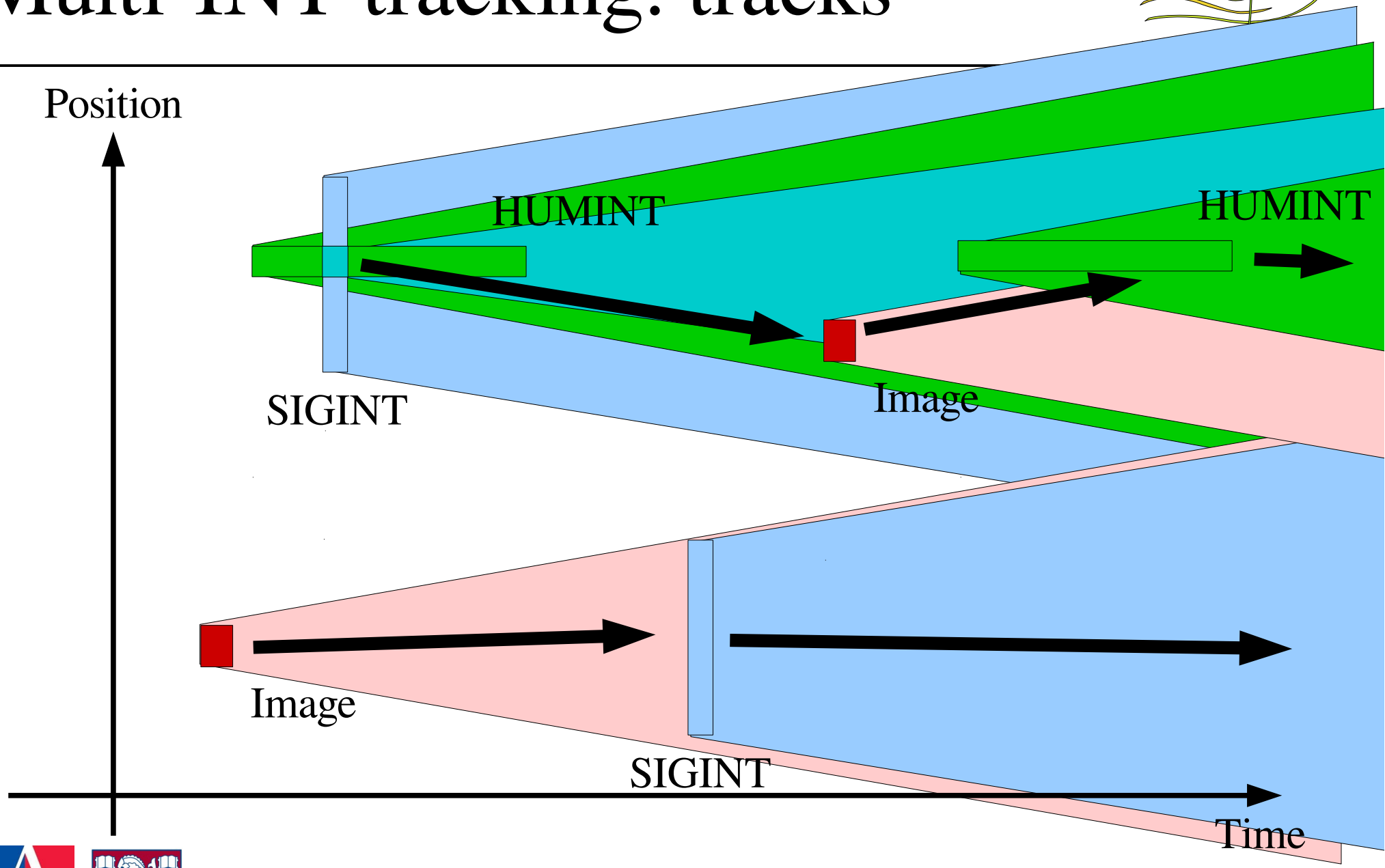
Multi-INT tracking: models



Minor technical point: this is actually a ***cosheaf***

Michael Robinson

Multi-INT tracking: tracks

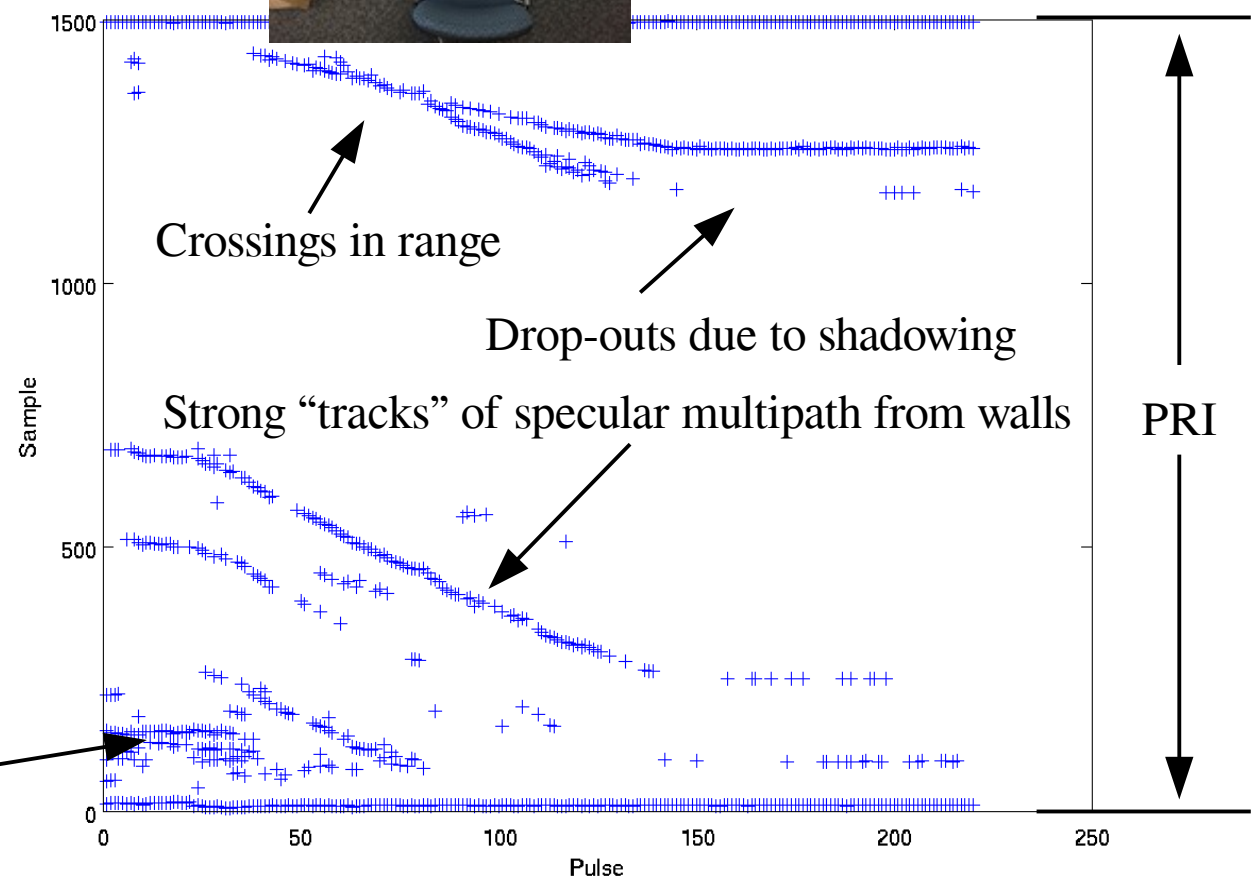


Michael Robinson

Single-INT demonstration

- Processing: pulse-wise CFAR followed by centroid
- Moving receiver, stationary transmitter
- 44.1 kHz sample rate
- 1500 range samples per pulse

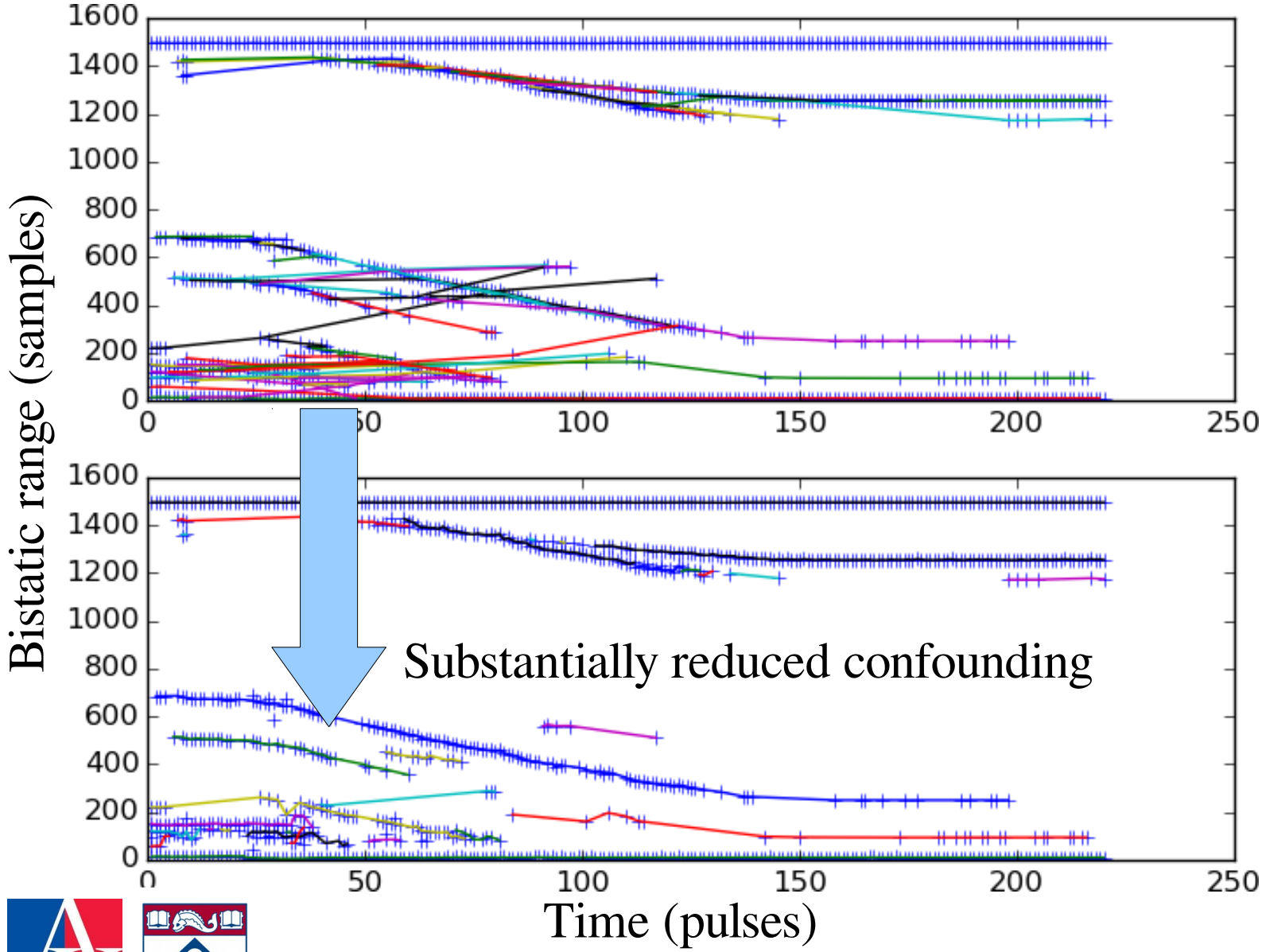
Collection system:
Netbook
Wired microphone
10W Laptop speakers



Fine structure of near-in clutter

Michael Robinson

Single-INT tracker results



Kalman filter

Substantially reduced confounding

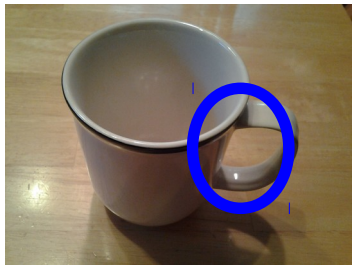
Cosheaf tracker

Michael Robinson

Detecting defects in the data

Misleading fused data: Loops

- Sheaves are sensitive to *topological features*
  - These are features that are robust to **large** perturbations
  - Clusters, loops, voids, etc.



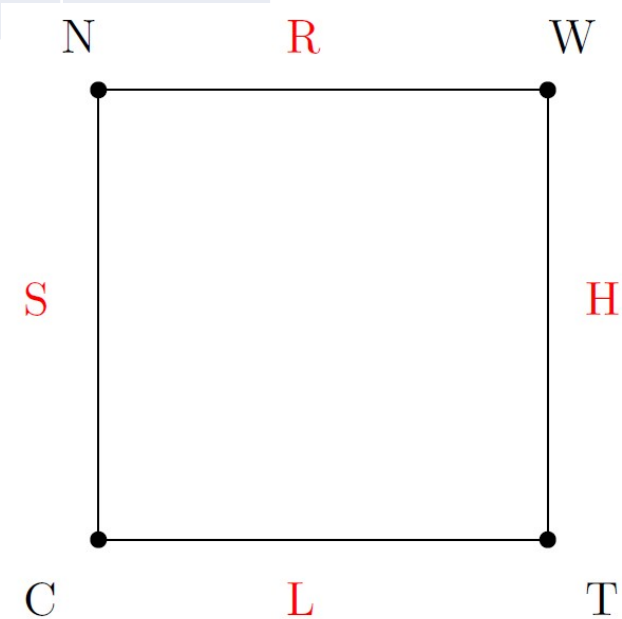
Nontrivial $H^1(X; \mathbb{Z})$

- The presence of loops in the sensor structure indicates
  - Coverage holes
  - Possibility of spurious information going unnoticed

“Weather Loop” a simple model

Sensors/ Questions	Rain? (R)	Humidity % (H)	Clouds? (L)	Sun? (S)
News (N)	X			X
Weather Website (W)	X	X		
Rooftop Camera (C)			X	X
Twitter (T)		X	X	

Make simplicial complex

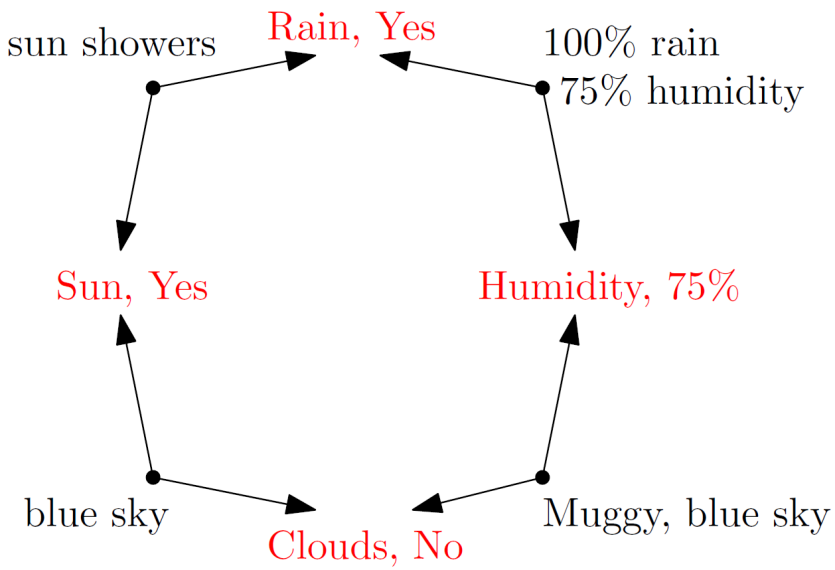


<u>Question</u>: Can misleading globalized information be detected?

Michael Robinson

Misleading globalized data

News: "There are some sun showers in the area"
Weather website: 100% chance of rain; Humidity is at 75%
Rooftop camera: Only blue sky in view
Twitter: "@weatherguy Dude it's so muggy out, but at least I see some blue sky!"



- Each edge is consistent – so this is globalized
- No global consistency check (no higher dimensional face), so there can be conflicting information
  - For instance: rain without clouds

Under the right conditions, this can be detected algorithmically

Michael Robinson

Next up...

- Interactive session: Topological Features
- Next lecture: What is Topology?