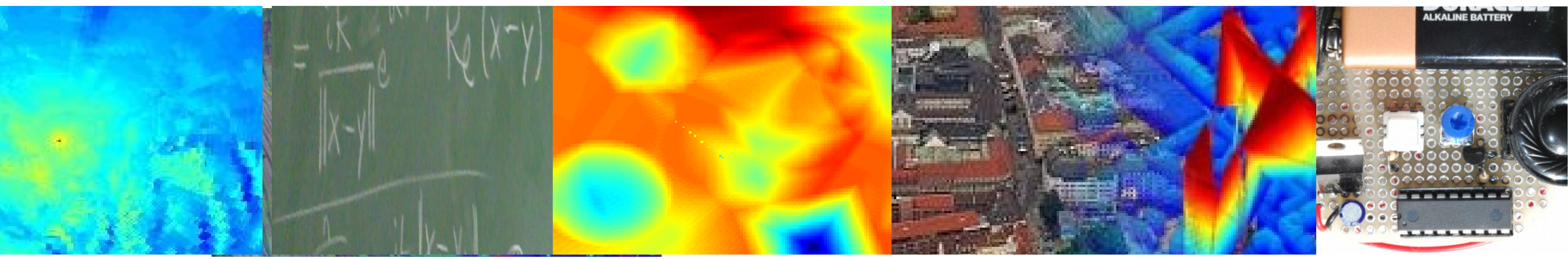
What is a sheaf?



Michael Robinson

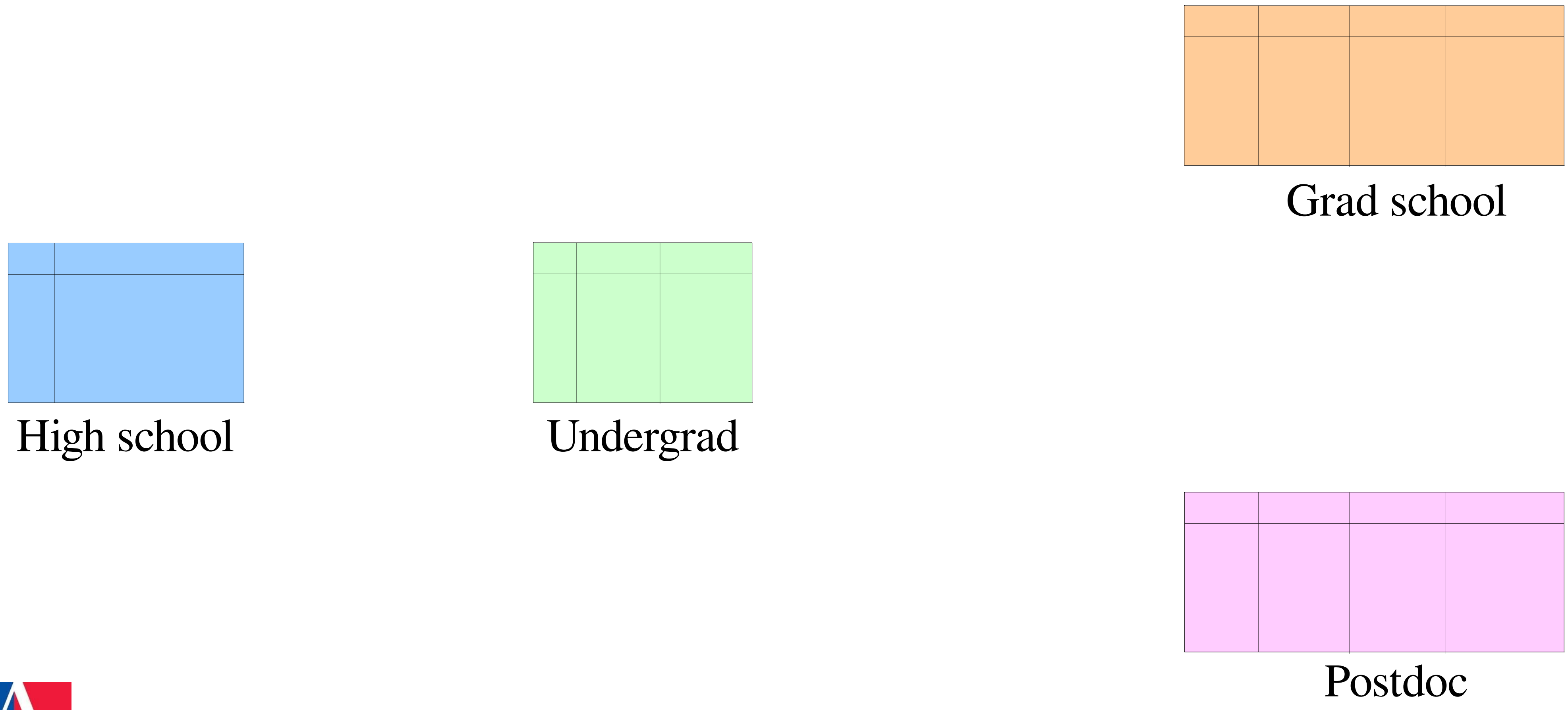
SIMPLEX Program

Session objectives

- What is heterogeneous data fusion?
- What is a sheaf?
- What happens if I don't have a sheaf?

Axiom 1: There is a set of data sources

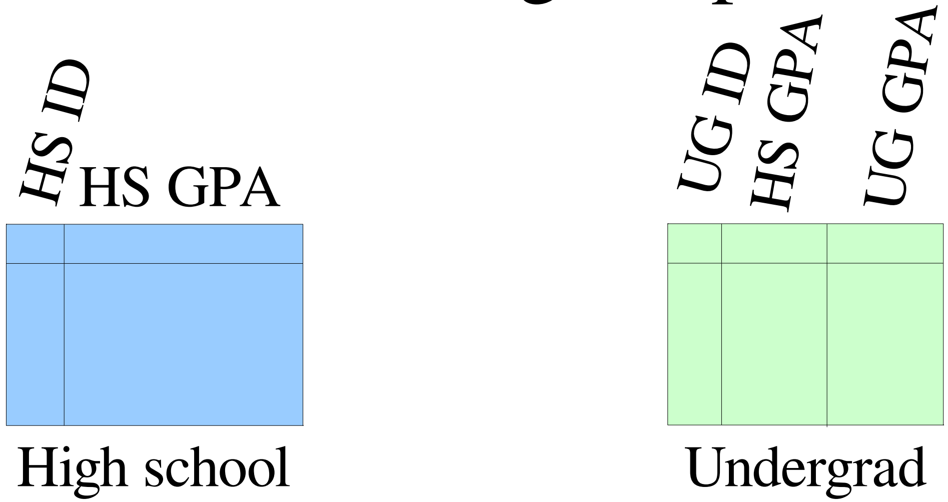
- A list of sensors
- A collection of tables

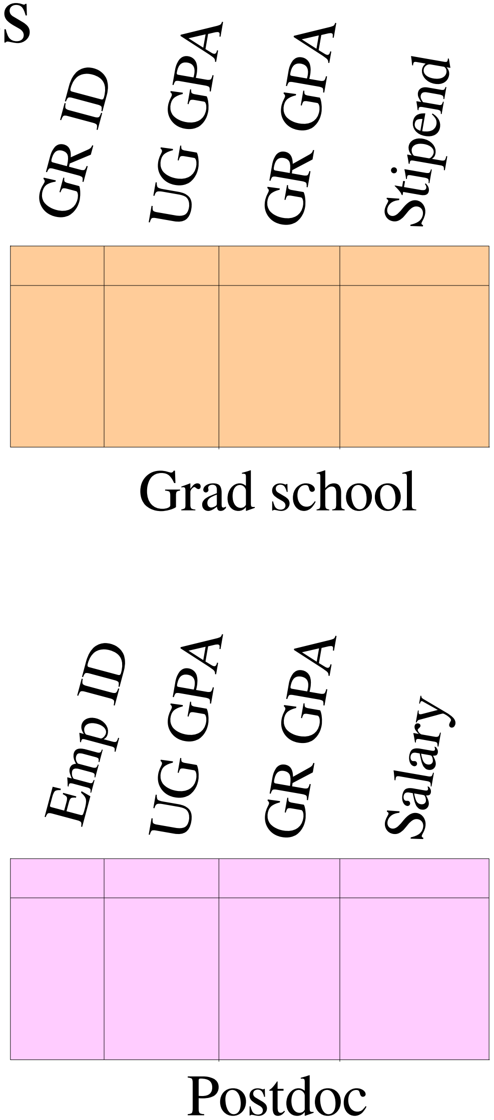


Axiom 2: There is a list of possible attributes

Different sensors will read out in attributes

- Possible columns or keys
- Possible signal spaces

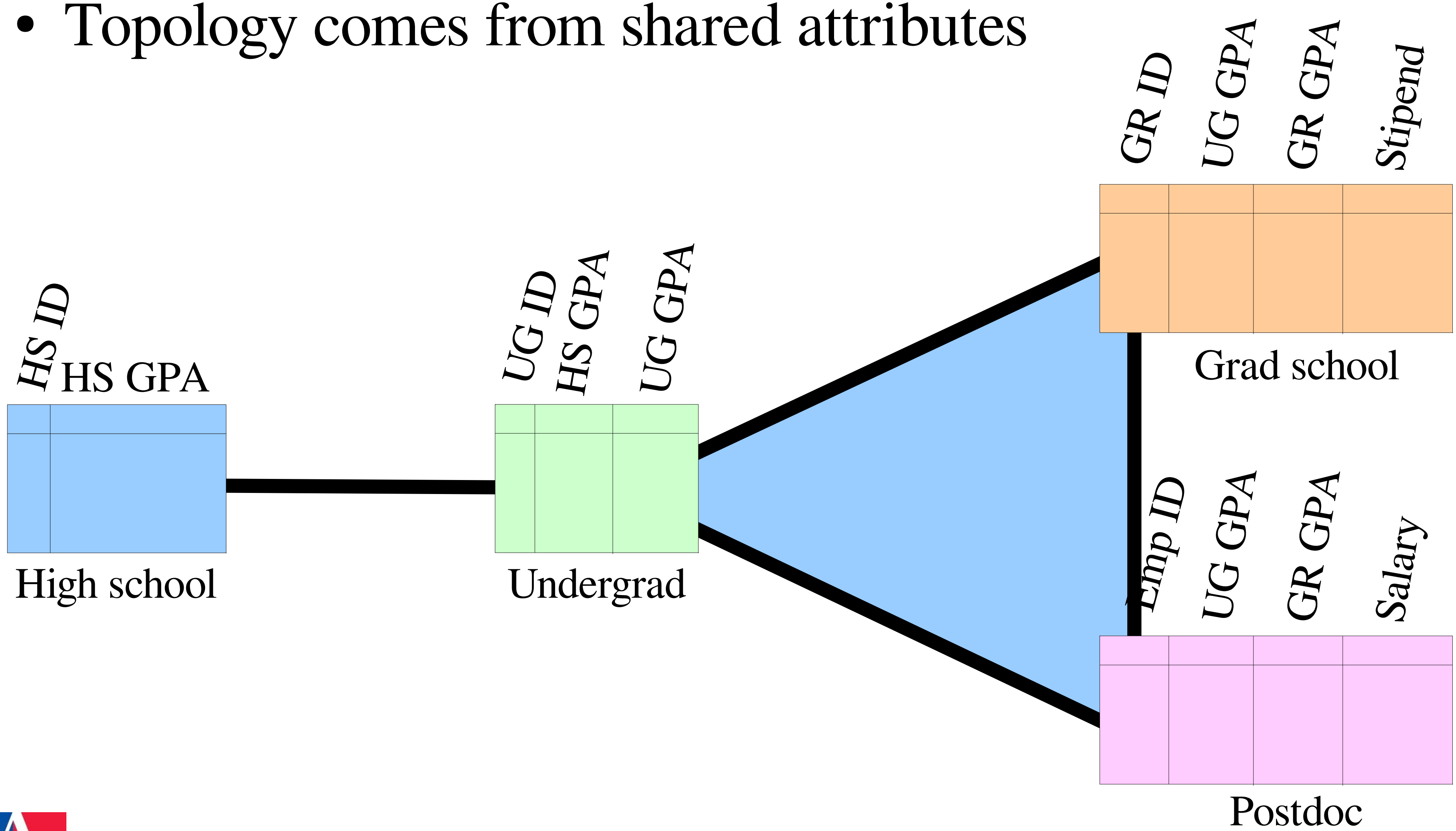




Michael Robinson

Axioms 3 and 4: The data sources are topologized

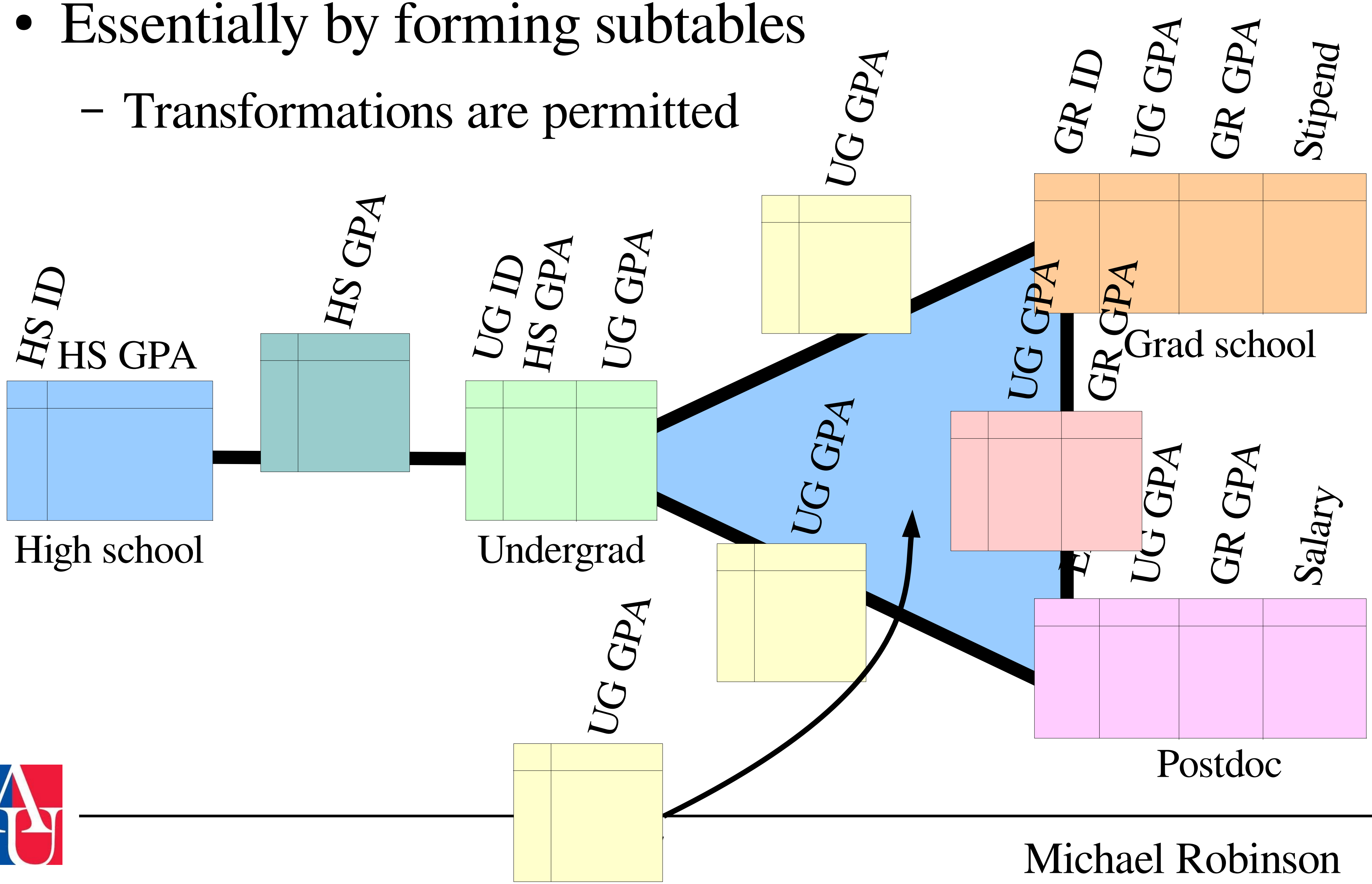
- Topology comes from shared attributes



Michael Robinson

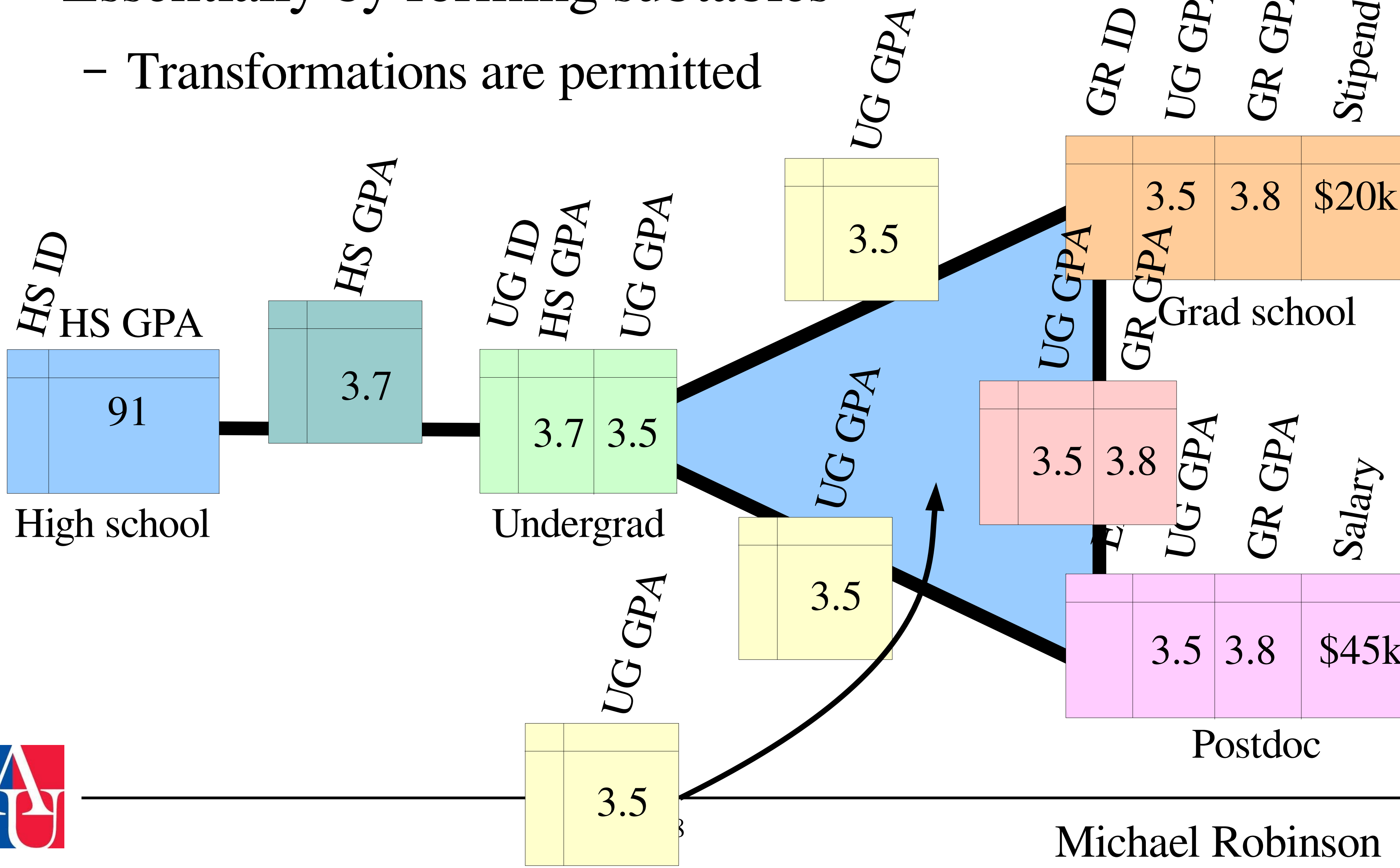
Axiom 5: Data sources can be compared via *restriction*

- Essentially by forming subtables
  - Transformations are permitted



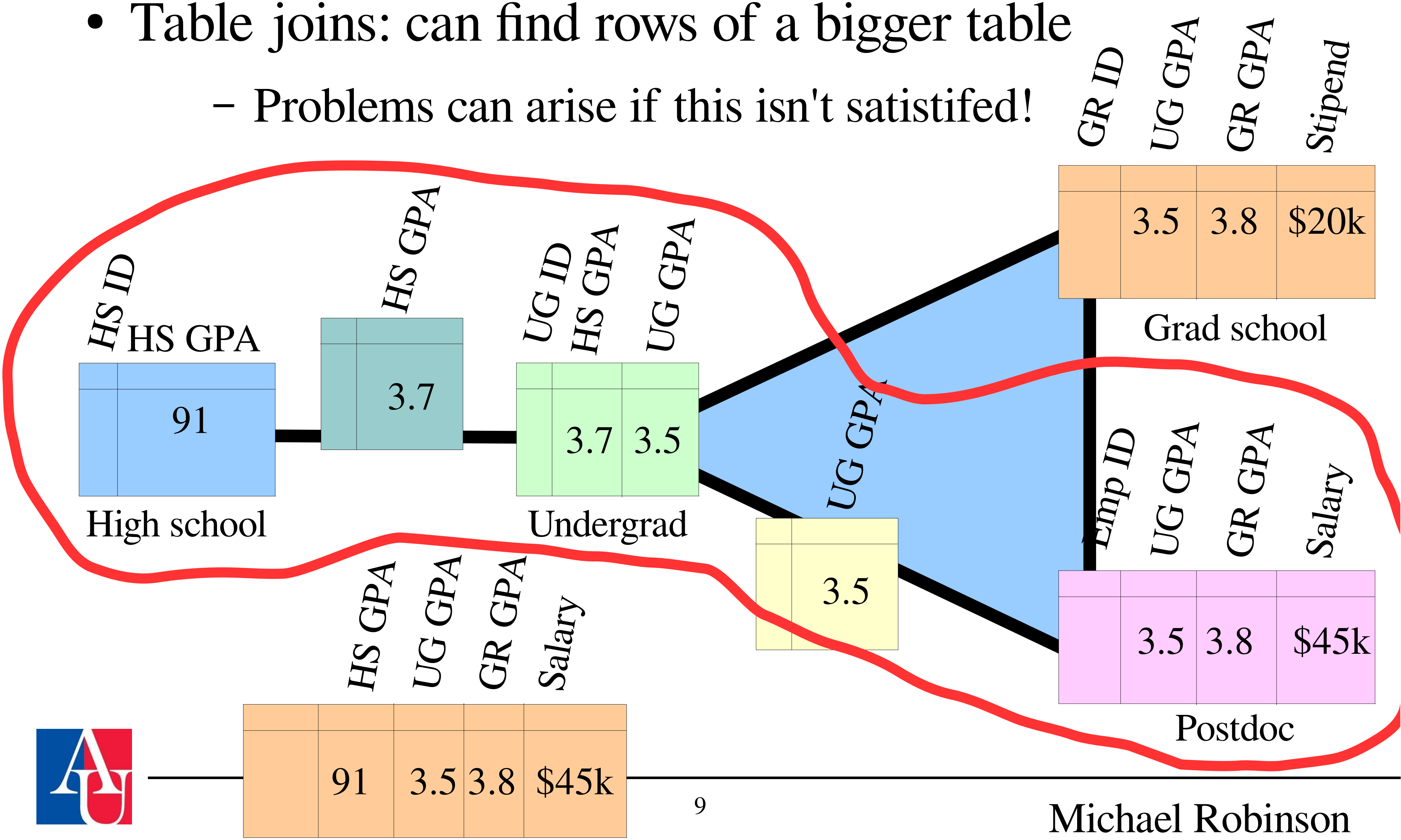
Axiom 5: Data sources can be compared via *restriction*

- Essentially by forming subtables
  - Transformations are permitted



Axioms 6 and 7: Data sources can be combined in a unique way

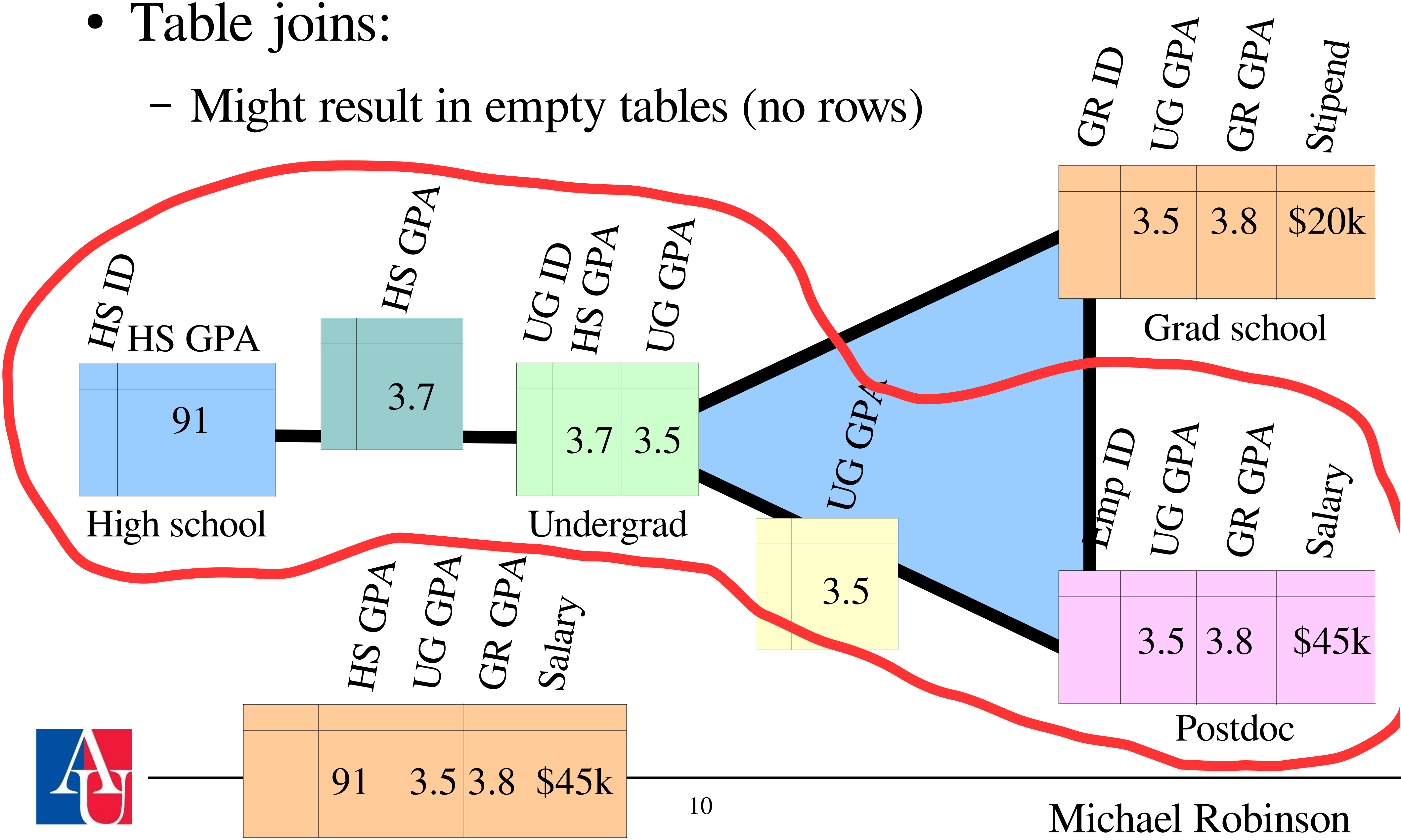
- Table joins: can find rows of a bigger table
  - Problems can arise if this isn't satistifed!



Michael Robinson

Axioms 6 and 7: Data sources can be combined in a unique way

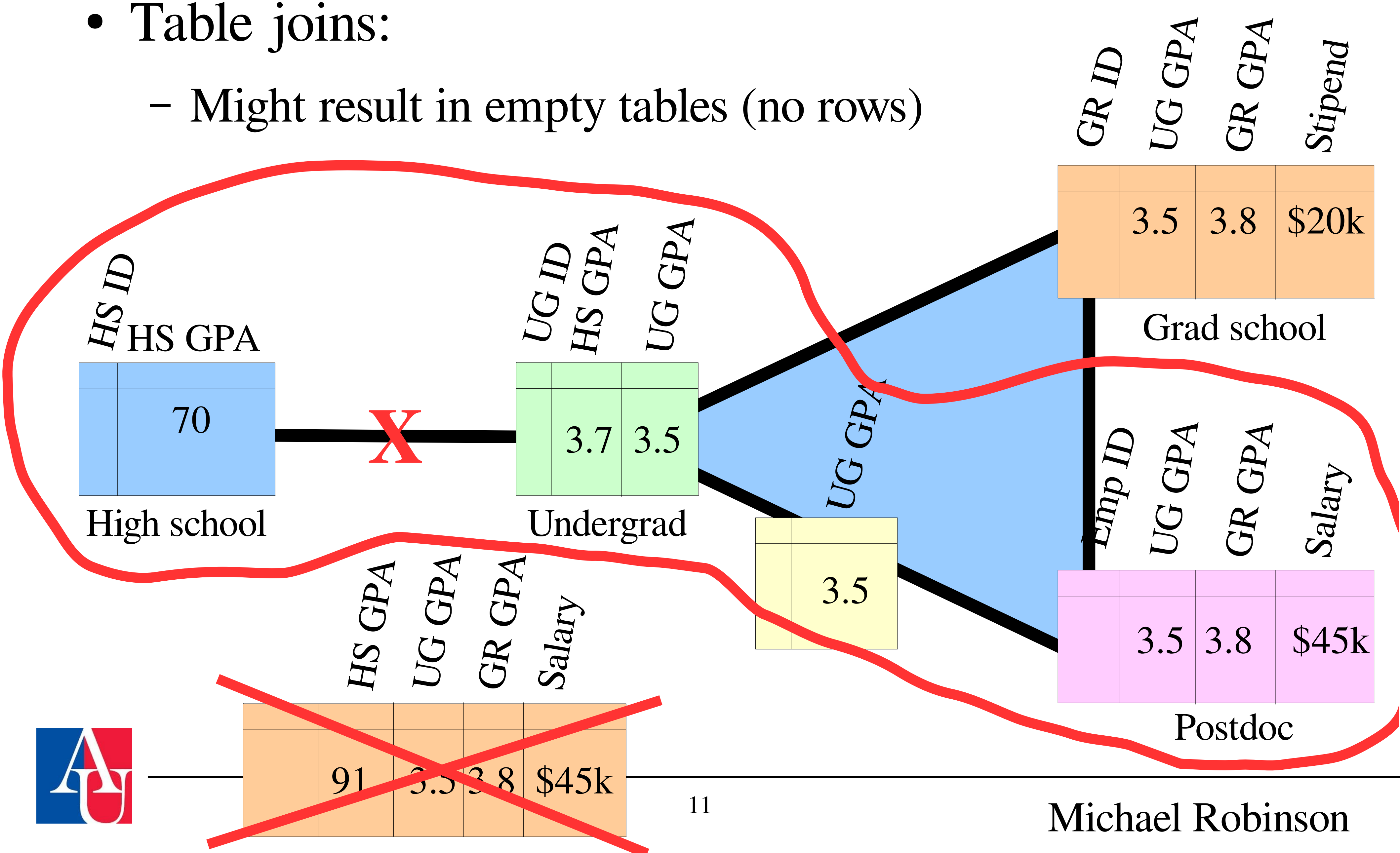
- Table joins:
  - Might result in empty tables (no rows)



Michael Robinson

Axioms 6 and 7: Data sources can be combined in a unique way

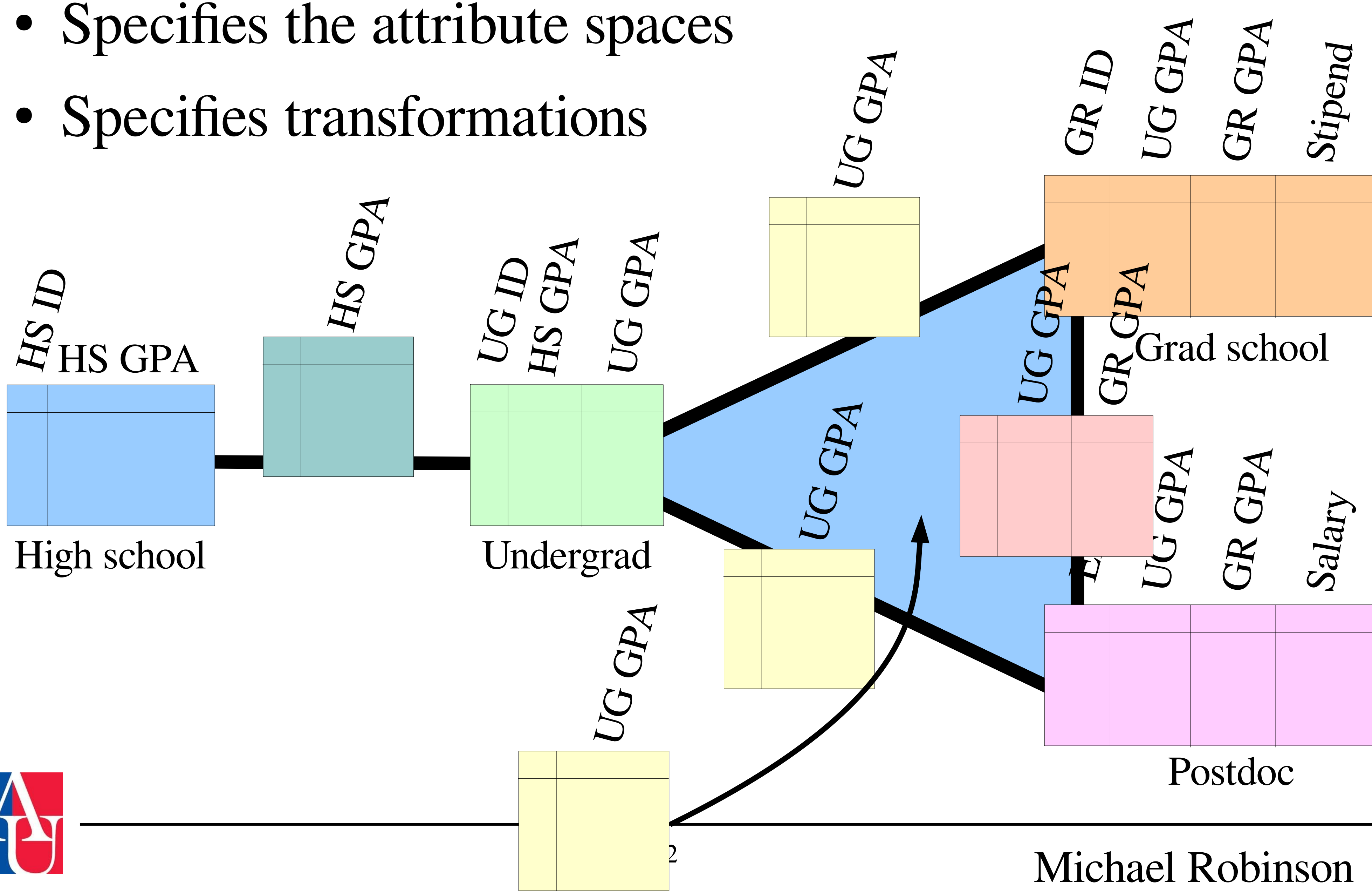
- Table joins:
  - Might result in empty tables (no rows)



Michael Robinson

Mathematical representation

- Specifies the attribute spaces
- Specifies transformations



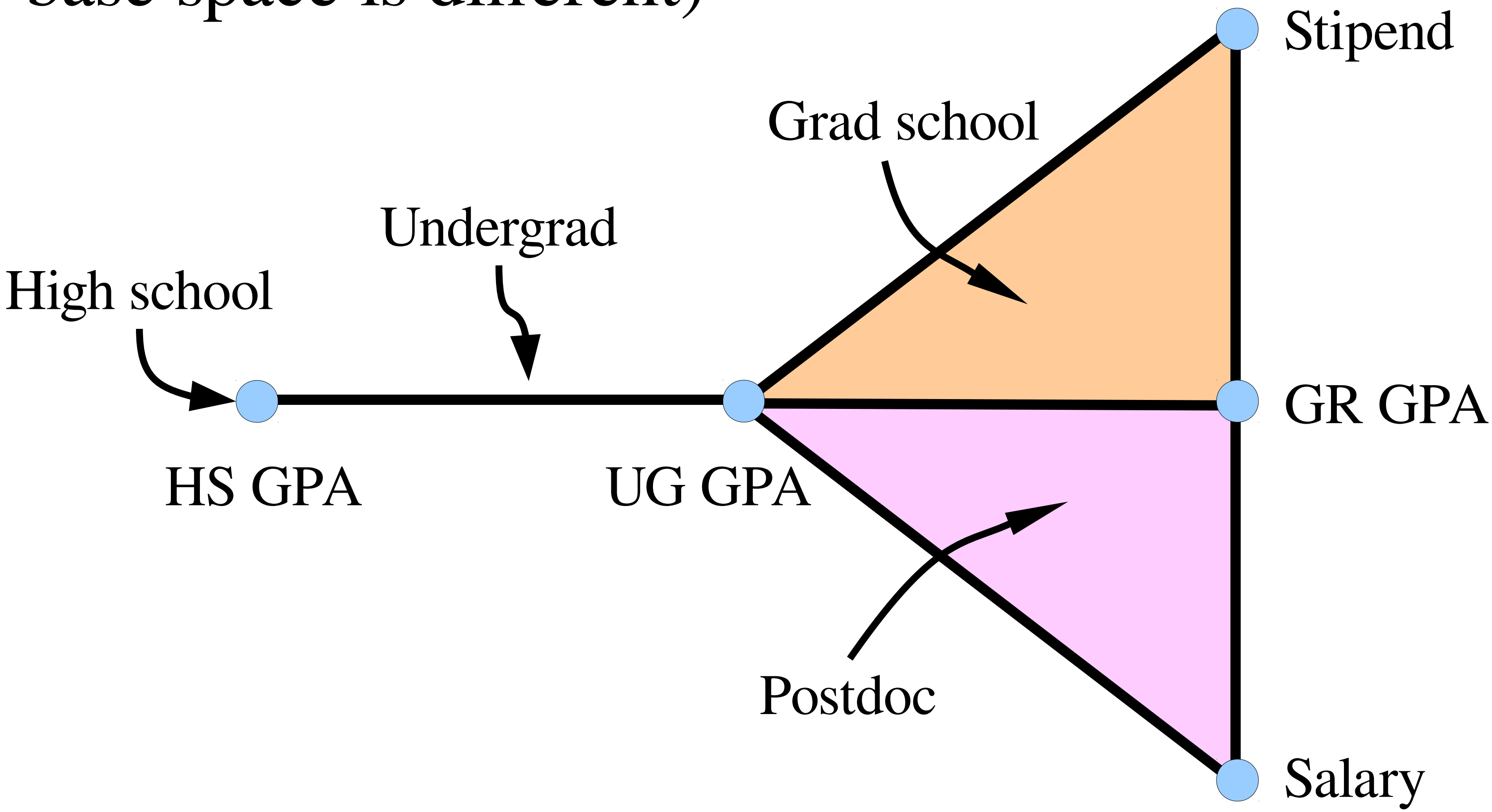
Key/table duality

Two obvious viewpoints:

- Table-centric → Tables on vertices (leads to *sheaf*)
  - "Bottom up"
  - *Restrictions* go from low dimensional simplices to higher dimensional simplices
- Key-centric → Keys on vertices (leads to *cosheaf*)
  - "Top down"
  - *Extensions* go from high dimensional simplices to lower dimensional simplices

Building a cosheaf model

- Keys are vertices, and tables are simplices (so the base space is different)



Idea credit: Jose Perea (Michigan) (Note: ID fields ignored for simplicity)

What happens when the axioms fail?

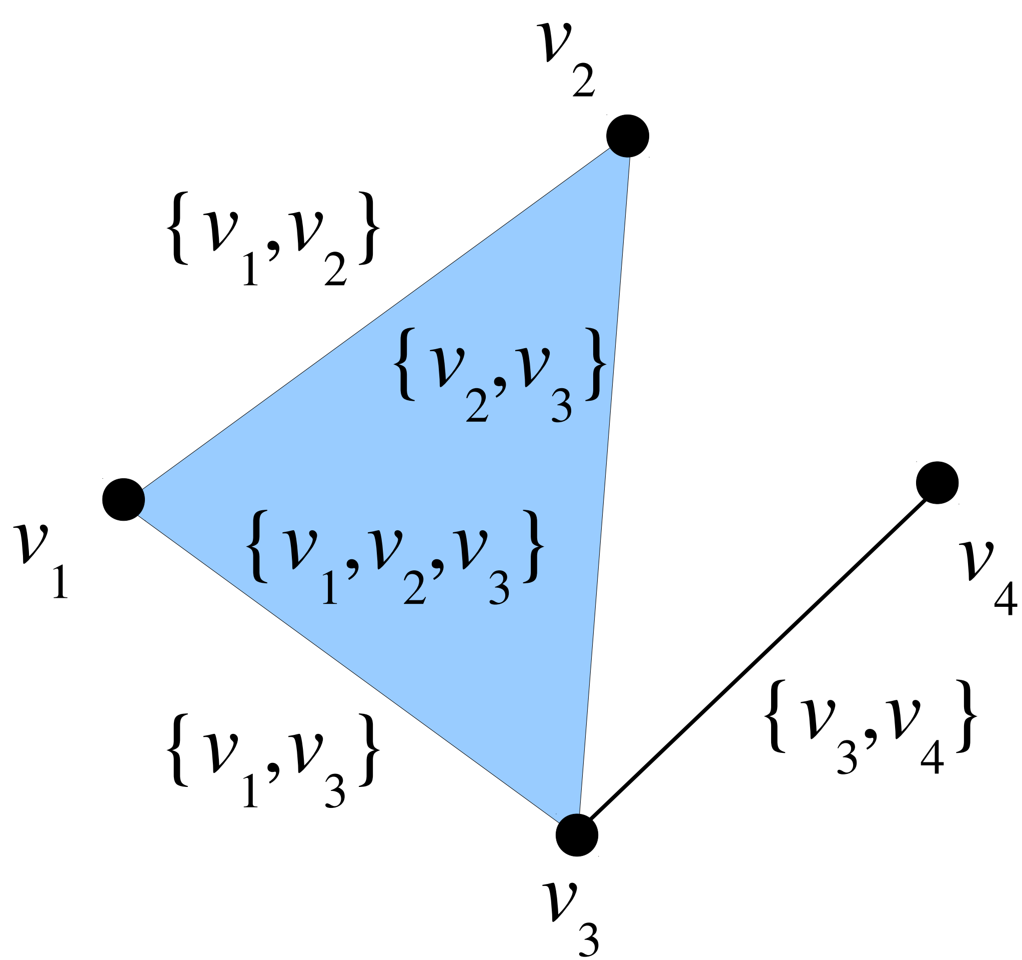
- Axiom 1 or 2: Data isn't in a set…
  - You won't be able to do much computationally
- Axiom 3 or 4: Sources not topologized
  - No basis for combining data sources…
- Axiom 5: No transformations to identify commonalities between observations
  - Although information about a given entity might be available through different sources, they can't be joined
- Axiom 6 or 7: Cannot uniquely fuse
  - Even if two observations are comparable, it is impossible to infer anything else about nearby observations
  - This often happens in databases – two rows with overlapping keys and matching values doesn't mean they're the same!

Now, abstractly...

A *sheaf* of ______________ on a _______________
(data type) (topological space)

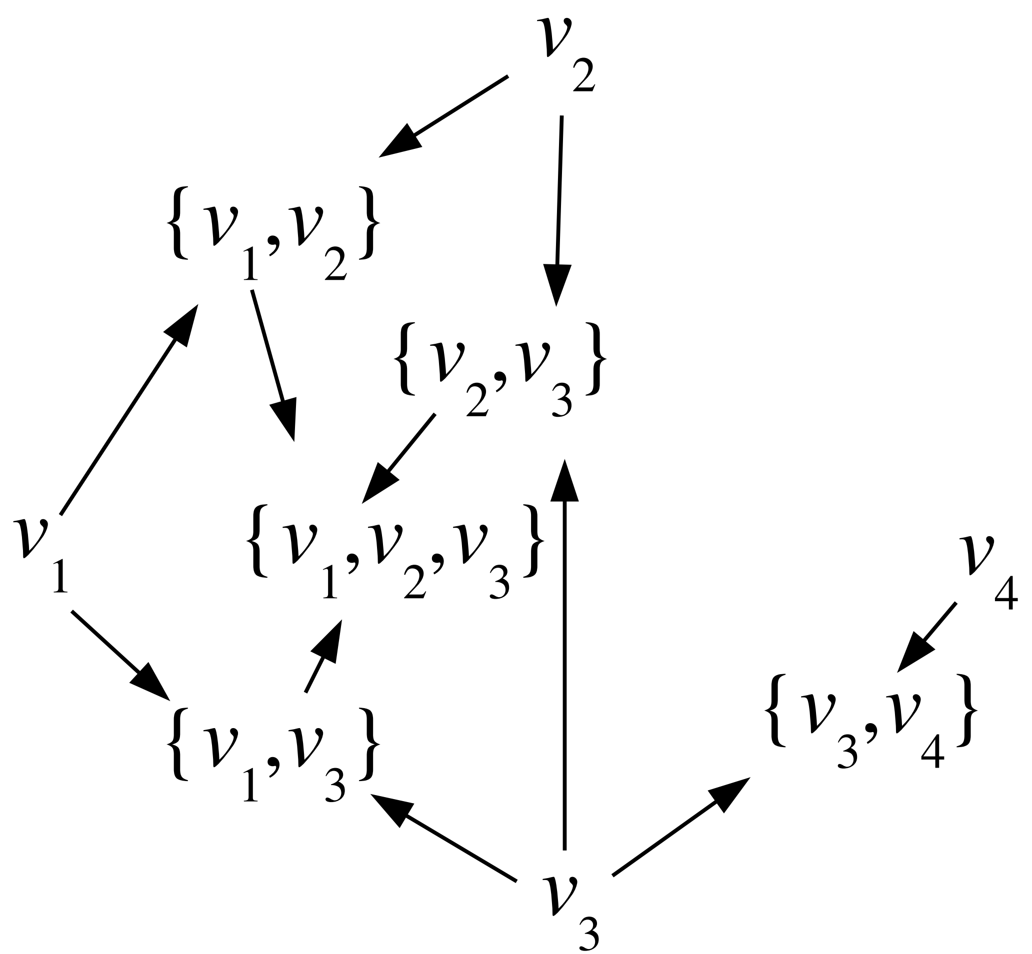
Simplicial complexes

- ... higher dimensional *simplices* (tuples of vertices)



Simplicial complexes

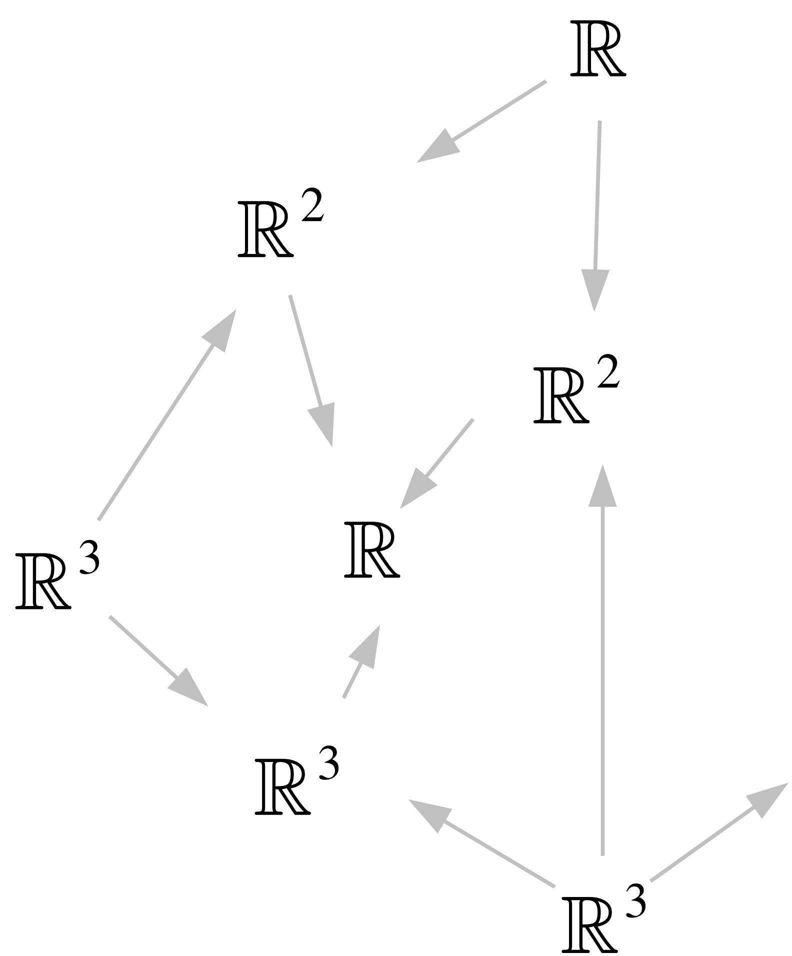
- The *attachment diagram* shows how simplices fit together



Michael Robinson

A *sheaf* is ...

- A set assigned to each simplex and ...



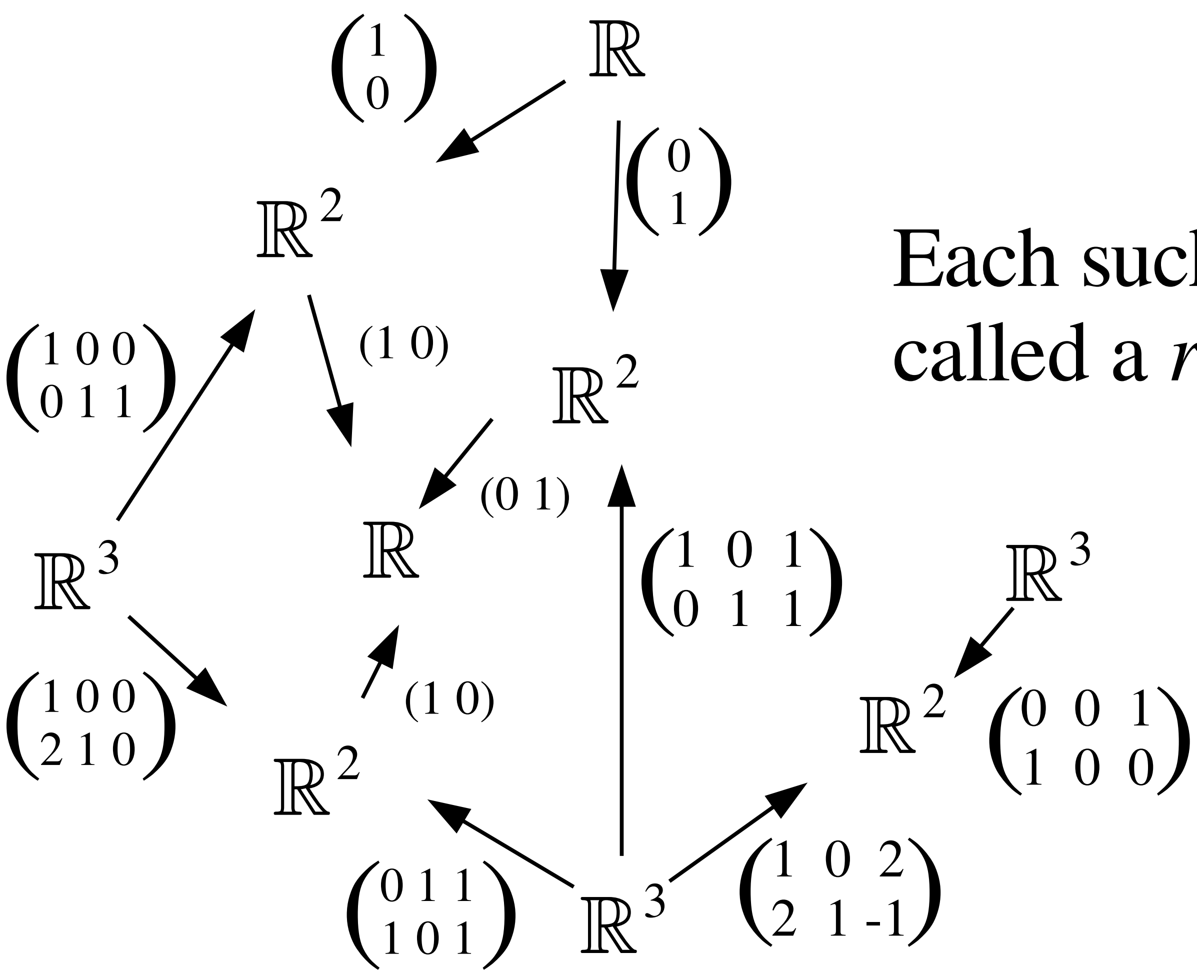
Each such set is called the *stalk* over its simplex

This is a sheaf **of** vector spaces **on** a simplicial complex

Michael Robinson

A *sheaf* is ...

- … a function assigned to each simplex inclusion

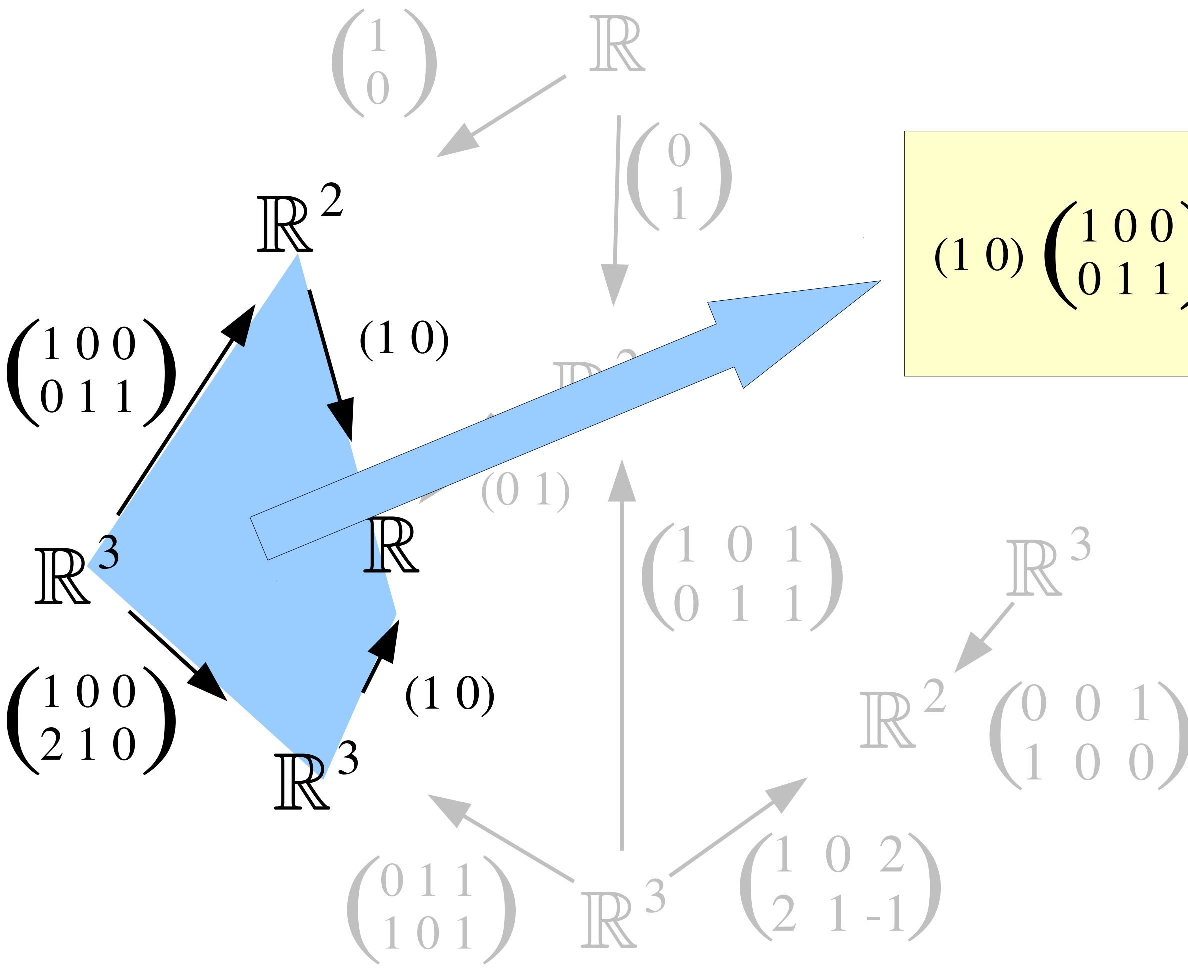


Each such function is called a *restriction*

Michael Robinson

A *sheaf* is ...

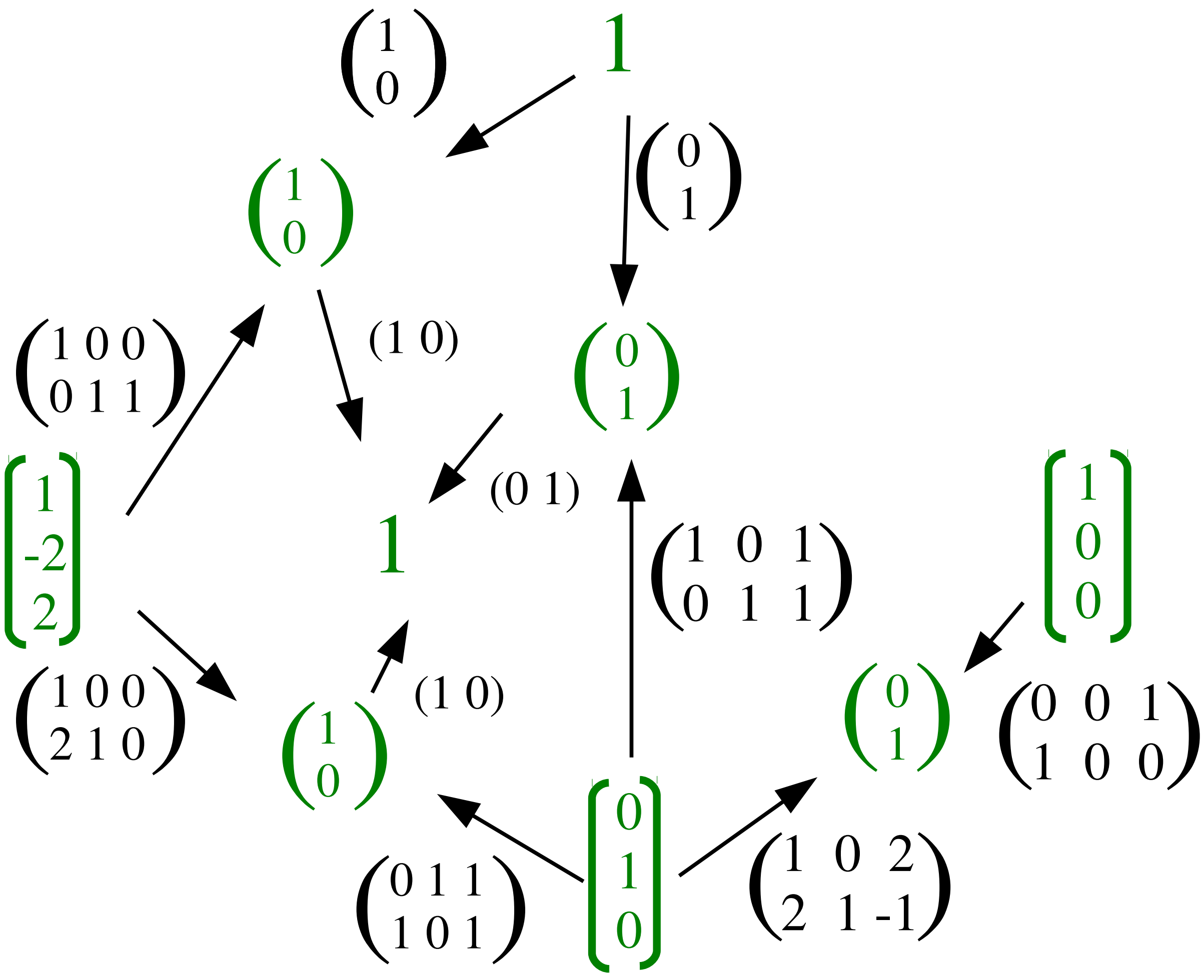
- … so the diagram commutes.



$$(1\ 0)\begin{pmatrix}1&0&0\\0&1&1\end{pmatrix} = (1\ 0)\begin{pmatrix}1&0&0\\2&1&0\end{pmatrix}$$

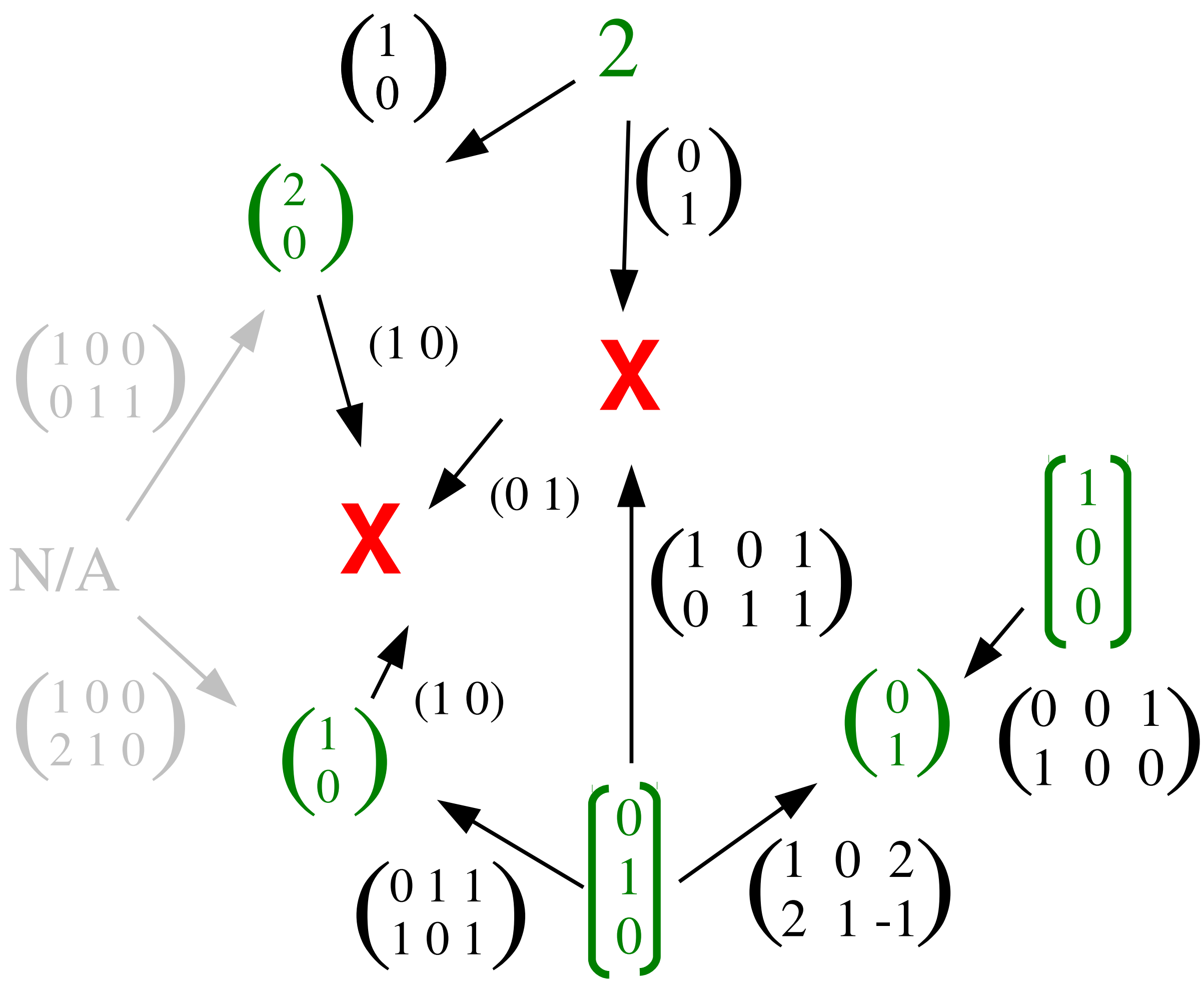
A *global section* is ...

- An assignment of values from each of the stalks that is consistent with the restrictions



Some sections are only local

- They might not be defined on all simplices or disagree with restrictions



Flabbiness

- If all local sections defined on vertices extend to global sections, the sheaf is called *flabby* (or *flasque*)
  - These sheaves don't have interesting invariants
  - They are good for decomposing other sheaves

$$\mathbb{R}^3 \rightarrow 0 \leftarrow \mathbb{R}^3$$

- Example: Vertex- or edge-weighted graphs with no further constraints
- Flabby sheaves mean there is a lack of constraints imposed by the model

Queue model example

A queue as a sheaf

- Contents of the shift register at each timestep
- $N = 3$ shown

$$\rightarrow \mathbb{R}^2 \xleftarrow{\begin{pmatrix}0&1&0\\0&0&1\end{pmatrix}} \mathbb{R}^3 \xrightarrow{\begin{pmatrix}1&0&0\\0&1&0\end{pmatrix}} \mathbb{R}^2 \xleftarrow{\begin{pmatrix}0&1&0\\0&0&1\end{pmatrix}} \mathbb{R}^3 \xrightarrow{\begin{pmatrix}1&0&0\\0&1&0\end{pmatrix}} \mathbb{R}^2 \xleftarrow{\begin{pmatrix}0&1&0\\0&0&1\end{pmatrix}} \mathbb{R}^3 \xrightarrow{\begin{pmatrix}1&0&0\\0&1&0\end{pmatrix}} \mathbb{R}^2 \xleftarrow{\begin{pmatrix}0&1&0\\0&0&1\end{pmatrix}} \mathbb{R}$$

Michael Robinson

A single timestep

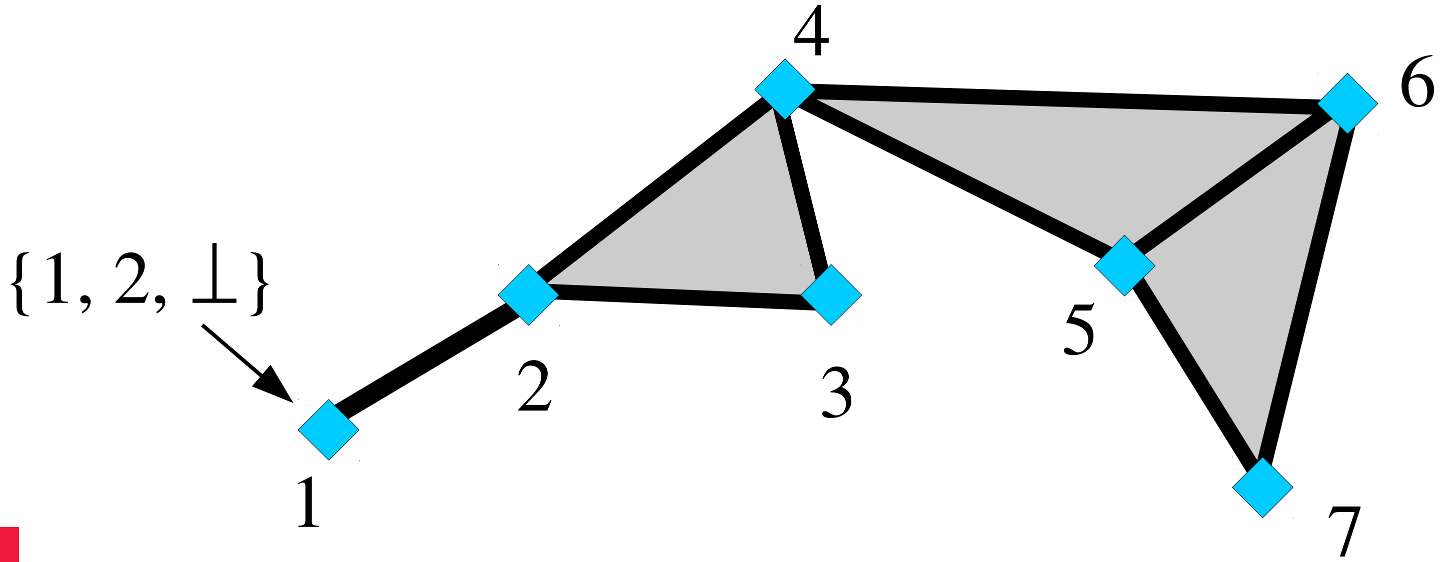
- Contents of the shift register at each timestep
- $N = 3$ shown

$$\to (9,2) \xleftarrow[\begin{pmatrix}0&1&0\\0&0&1\end{pmatrix}]{} (1,9,2) \xrightarrow{\begin{pmatrix}1&0&0\\0&1&0\end{pmatrix}} (1,9) \xleftarrow[\begin{pmatrix}0&1&0\\0&0&1\end{pmatrix}]{} (1,1,9) \xrightarrow{\begin{pmatrix}1&0&0\\0&1&0\end{pmatrix}} (1,1) \xleftarrow[\begin{pmatrix}0&1&0\\0&0&1\end{pmatrix}]{} (5,1,1) \xrightarrow{\begin{pmatrix}1&0&0\\0&1&0\end{pmatrix}} (5,1) \xleftarrow[\begin{pmatrix}0&1&0\\0&0&1\end{pmatrix}]{} (2,5$$

Wireless network example

Wireless activation sheaf

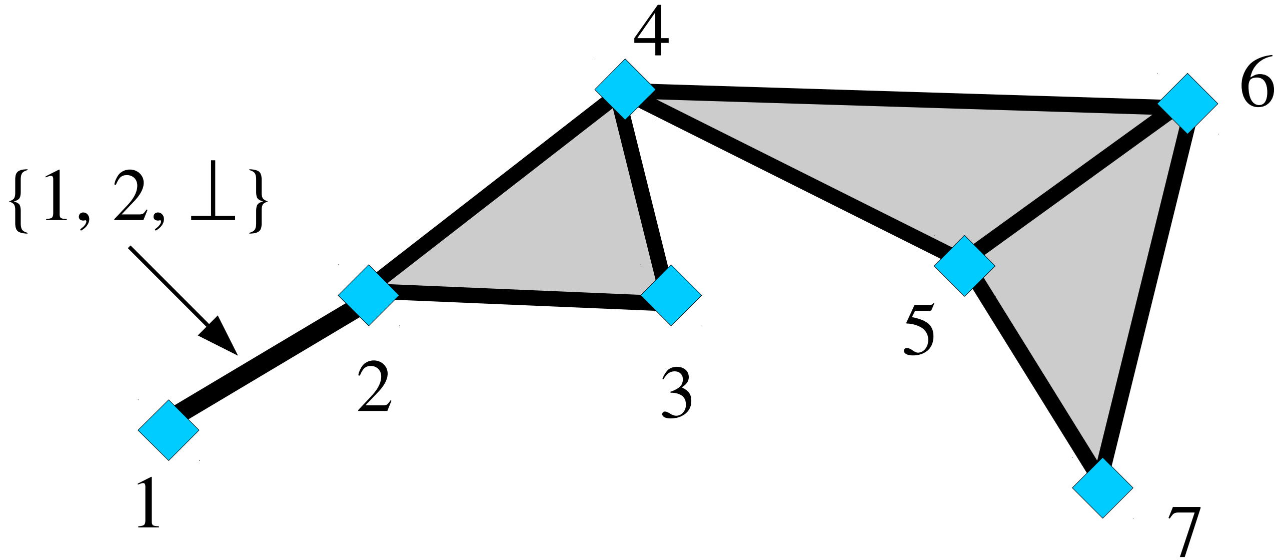
- Each node in the wireless complex has a unique ID
- The stalk over a face consists of the set of nodes adjacent to that face, and the special symbol ⊥



Michael Robinson

Wireless activation sheaf

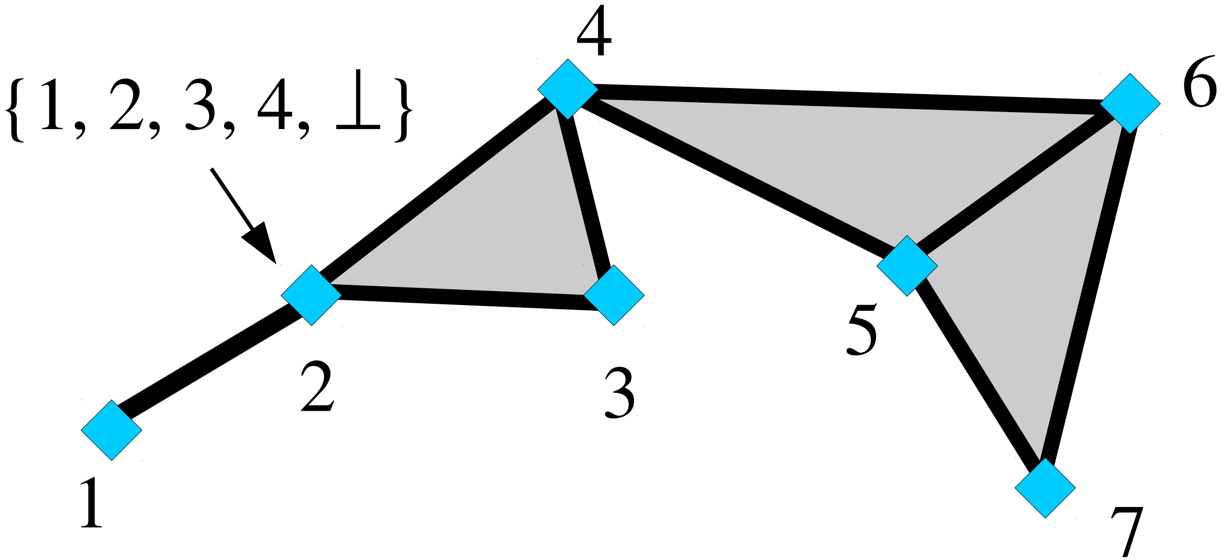
- Each node in the wireless complex has a unique ID
- The stalk over a face consists of the set of nodes adjacent to that face, and the special symbol $\perp$



Michael Robinson

Wireless activation sheaf

- Each node in the wireless complex has a unique ID
- The stalk over a face consists of the set of nodes adjacent to that face, and the special symbol ⊥

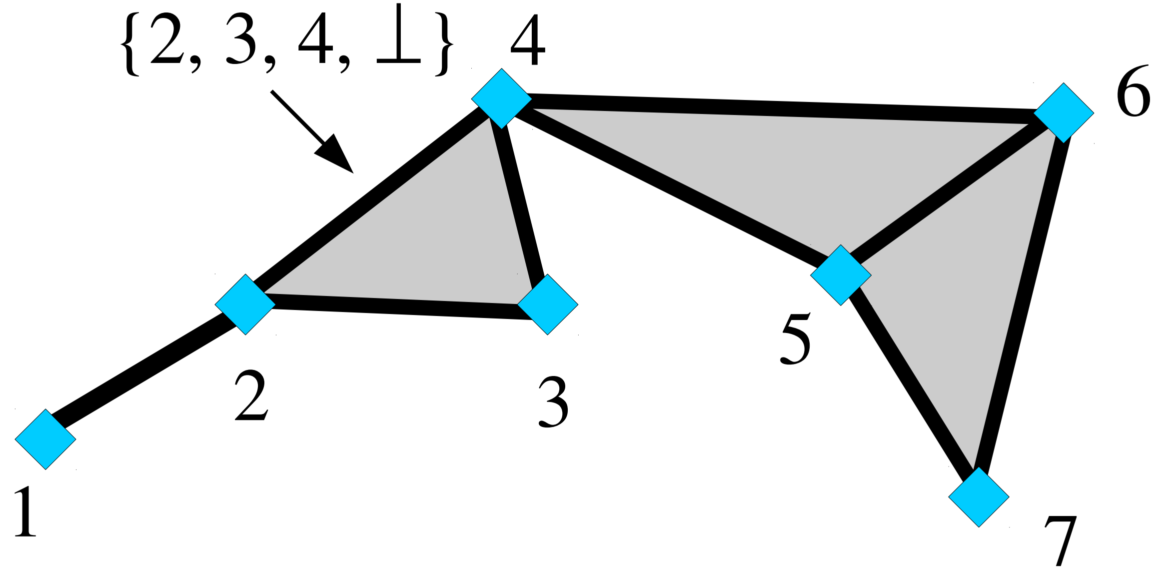


Michael Robinson

Wireless activation sheaf

- Each node in the wireless complex has a unique ID
- The stalk over a face consists of the set of nodes adjacent to that face, and the special symbol ⊥

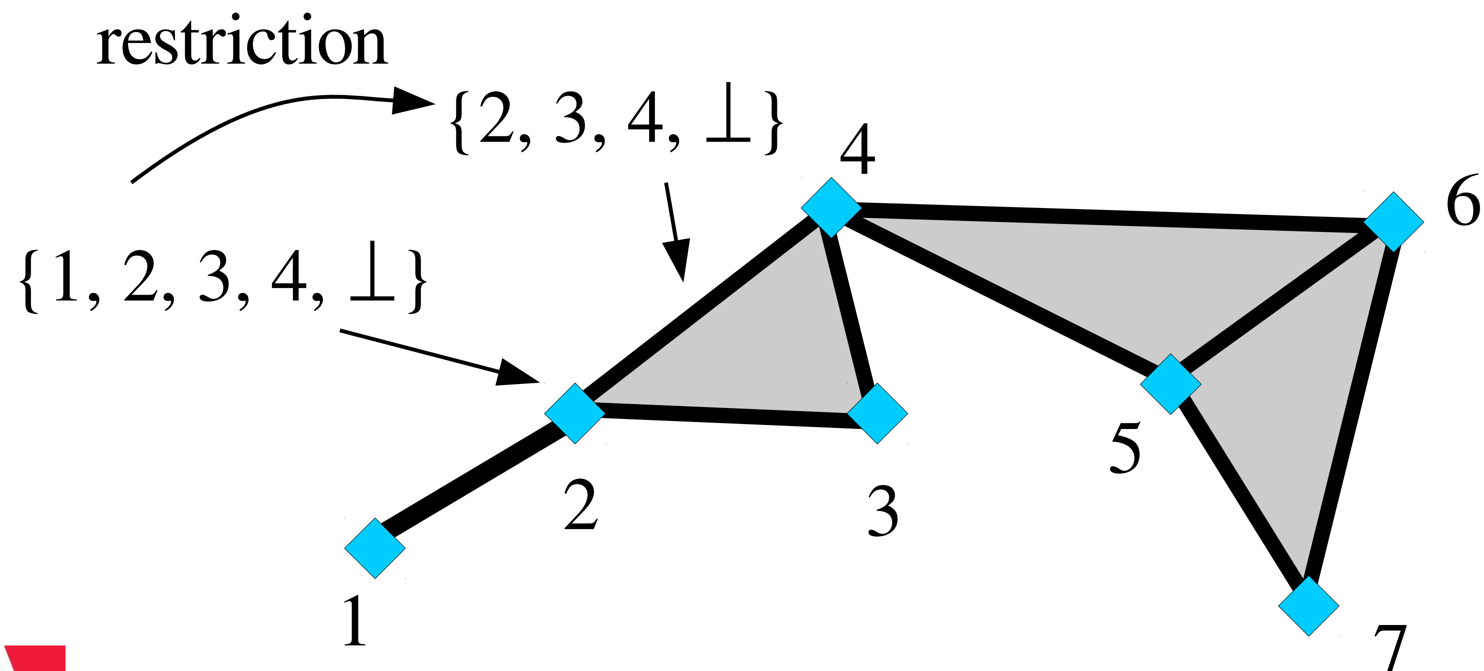
Note: "adjacent" means "has a higher-dimensional face in common"



Michael Robinson

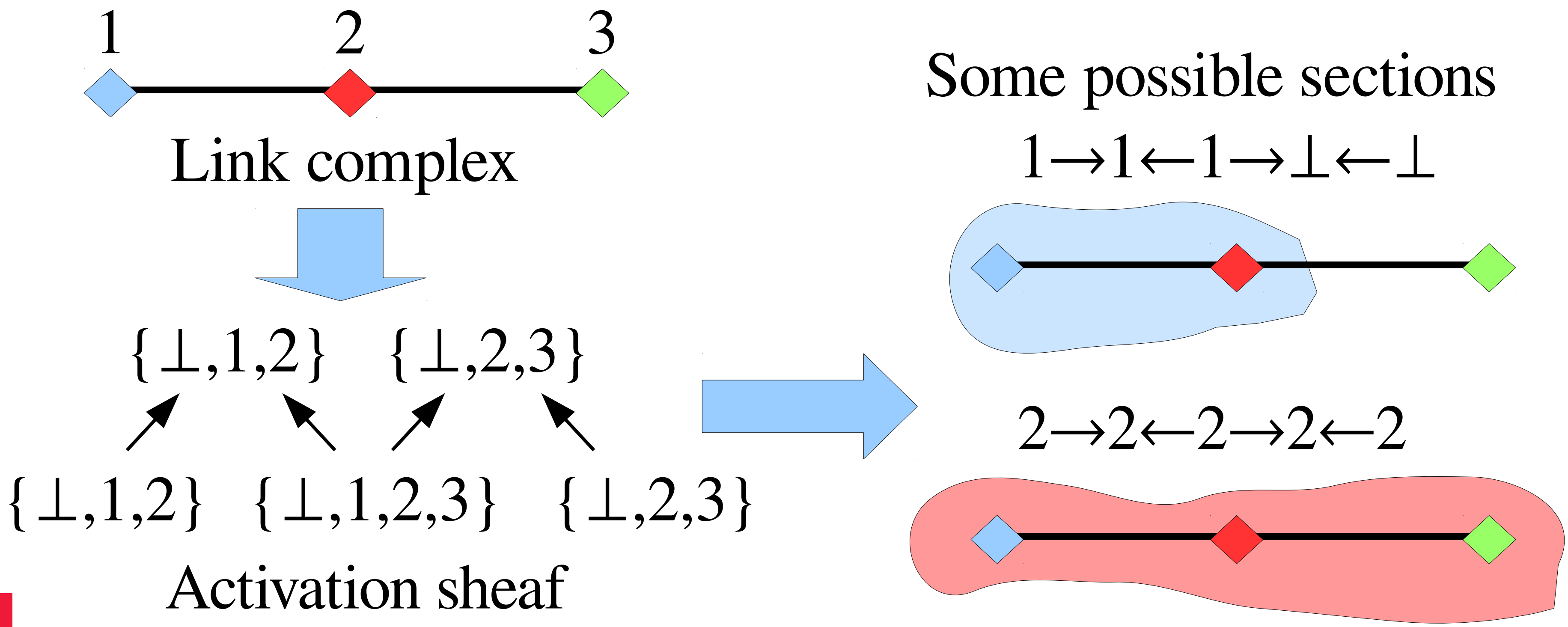
Wireless activation sheaf

- Restrictions map node IDs via identity whenever possible, and otherwise send to ⊥



Wireless activation sheaf

- Sheaves model node activation and traffic-passing protocols, by encoding **local consistency** constraints
- Node activation patterns when interference is possible. (⊥ means no activity)

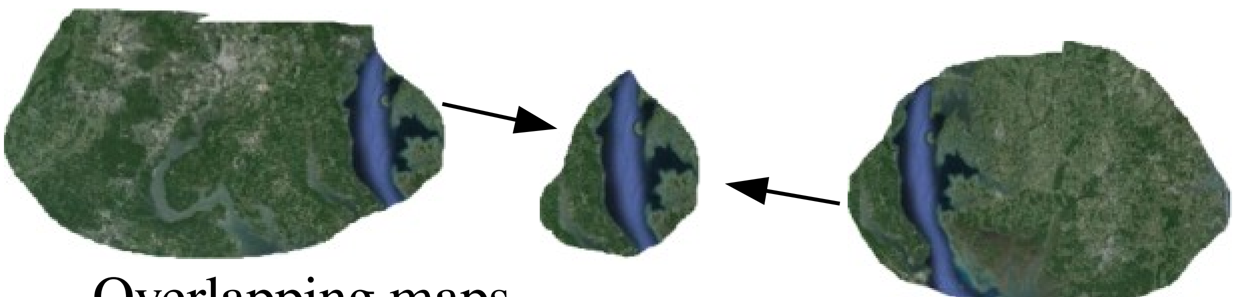


Michael Robinson

Shared situational awareness

Forming mosaics

- Multiple, overlapping images can be assembled into a mosaic by stitching together similar regions
  - Many algorithms exist
  - Most are robust to perspective (or other) changes

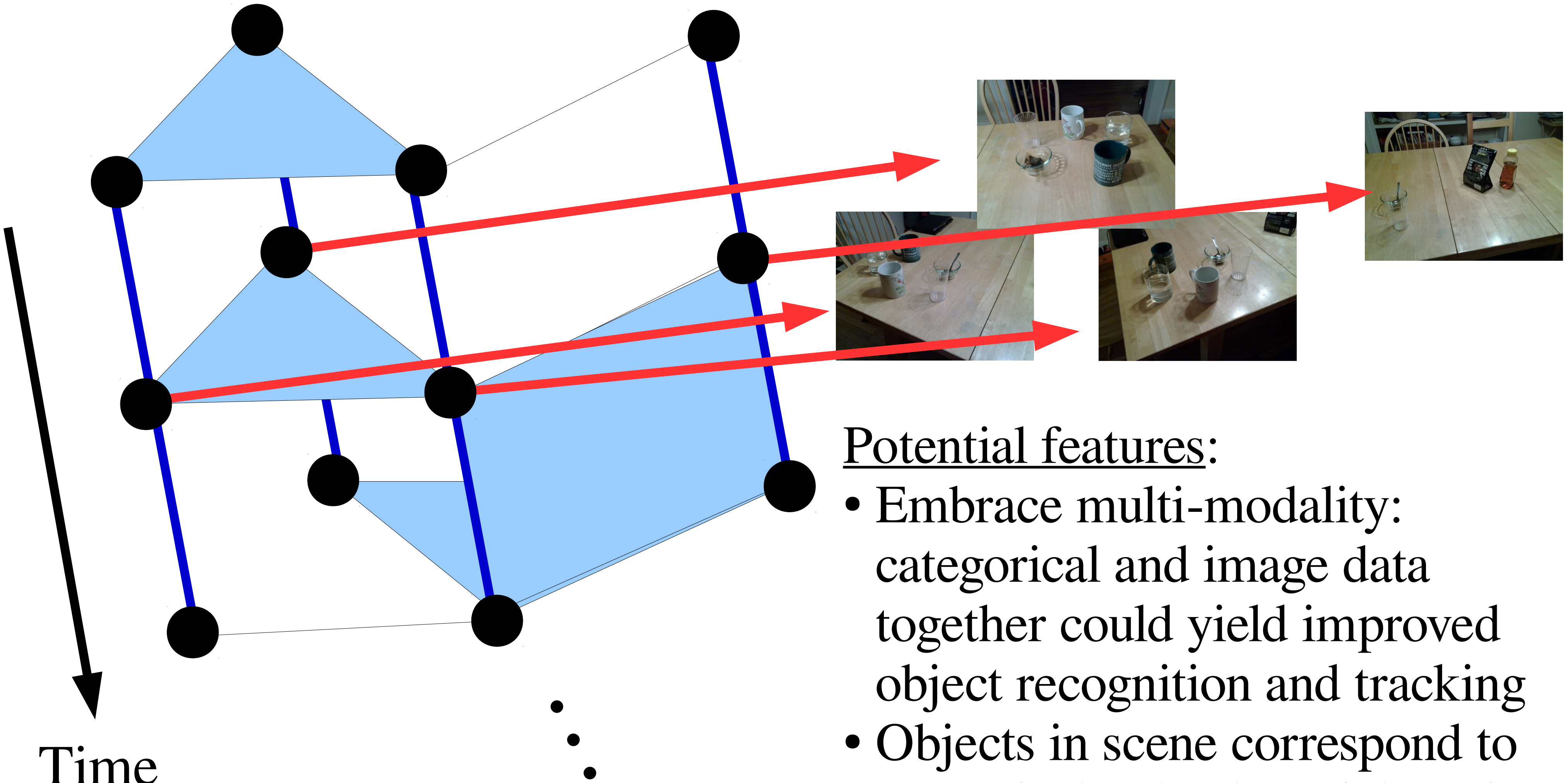


Overlapping maps

(Image courtesy of NASA/JPL)

Michael Robinson

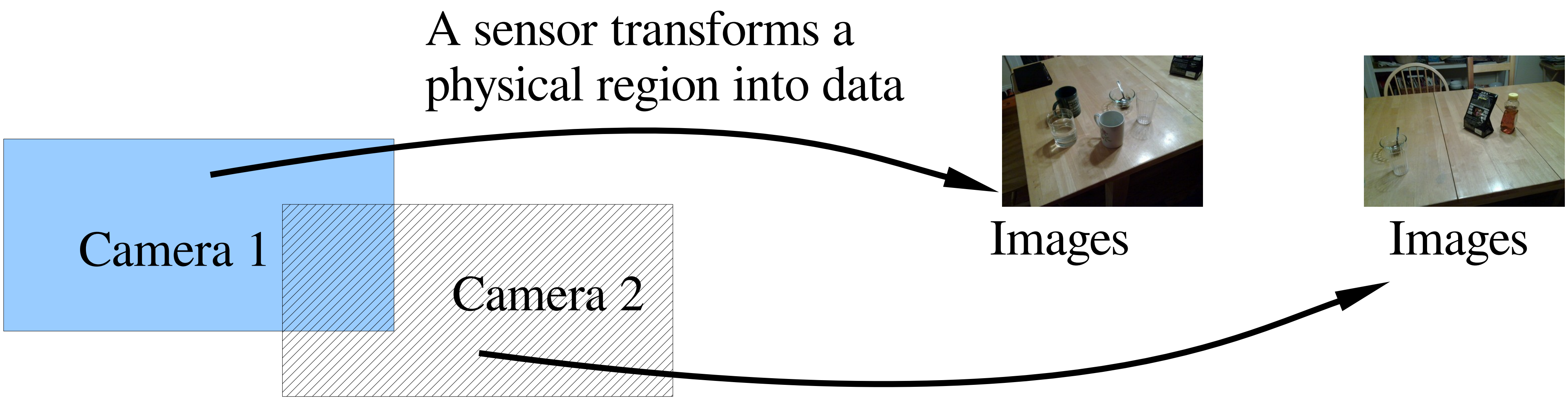
Shared visual situational awareness
(In collaboration with UCLA)



Potential features:

- Embrace multi-modality: categorical and image data together could yield improved object recognition and tracking
- Objects in scene correspond to categorical-valued partial sections

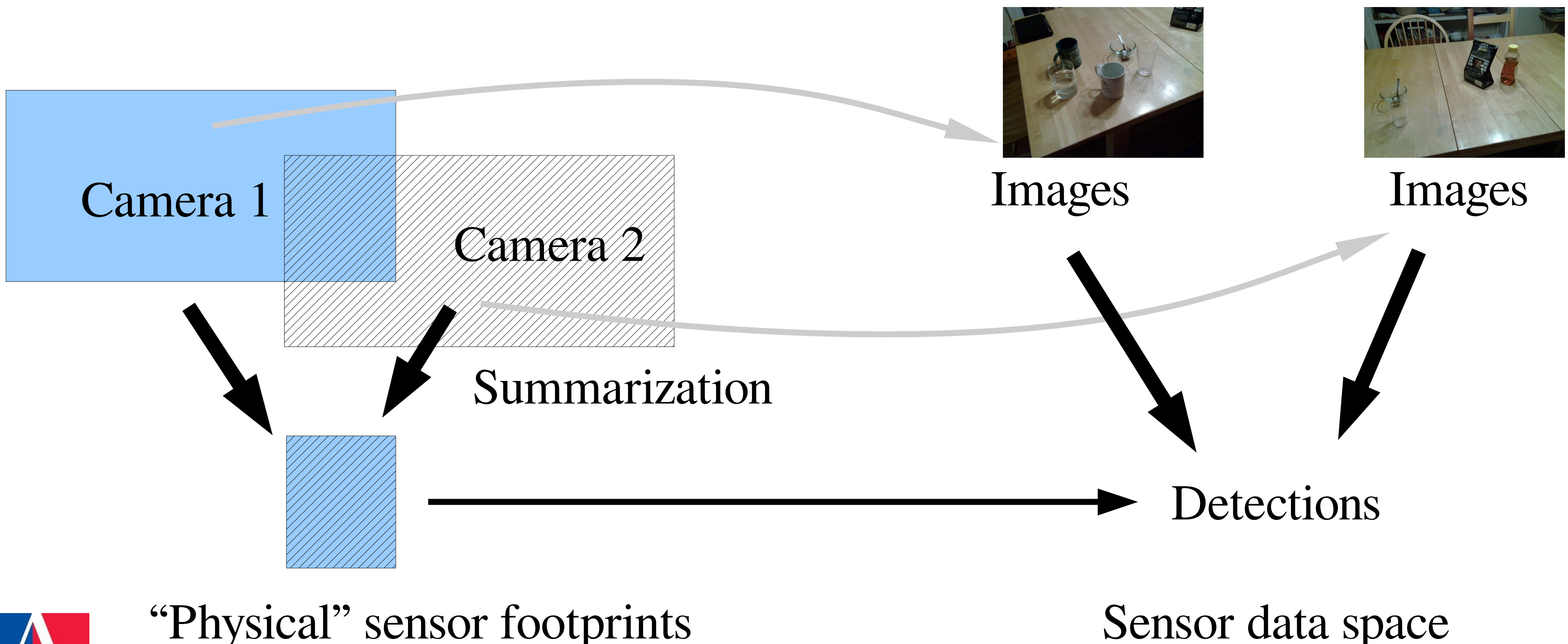
Heterogeneous fusion among homogeneous sensors



"Physical" sensor footprints

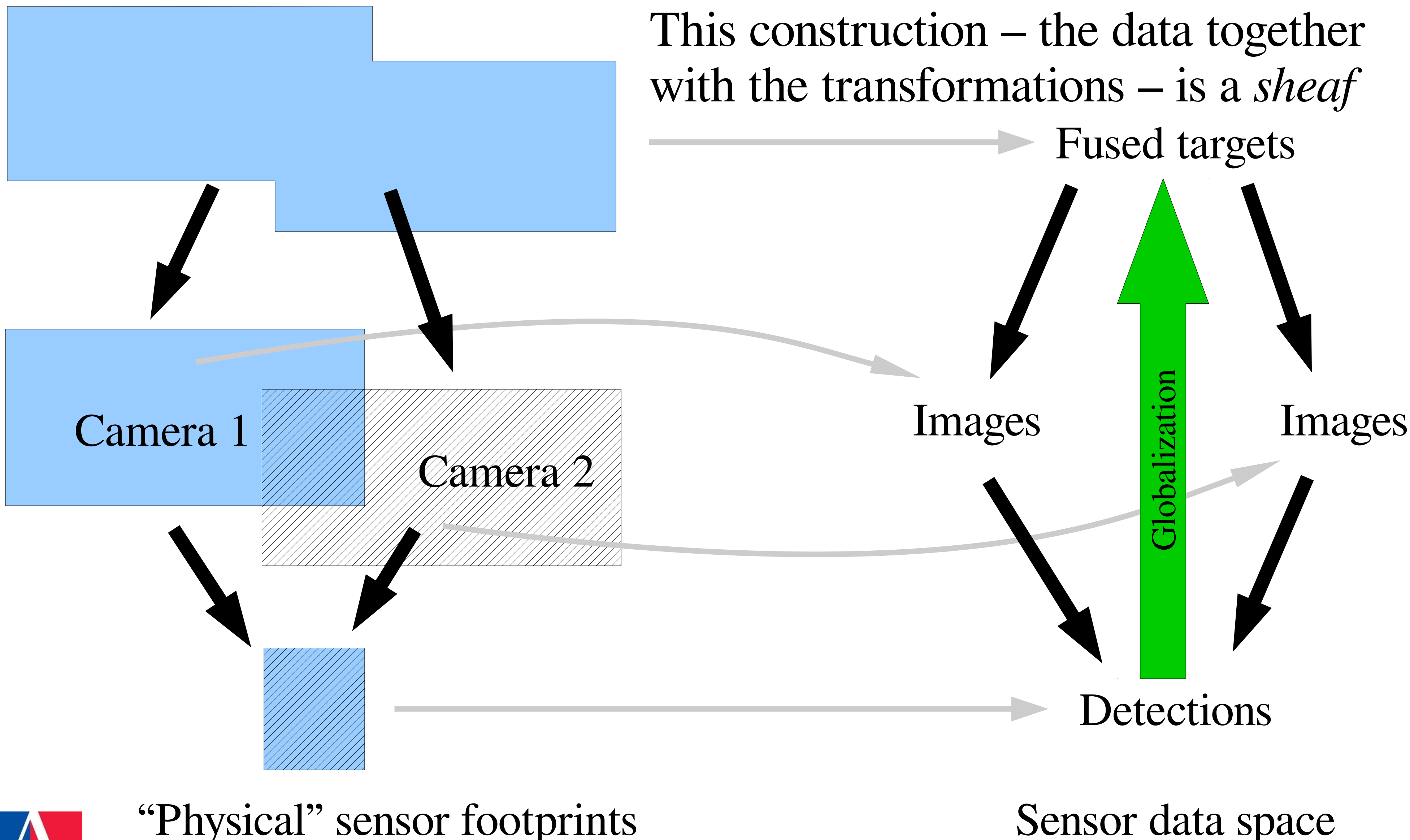
Sensor data space

Heterogeneous fusion among homogeneous sensors



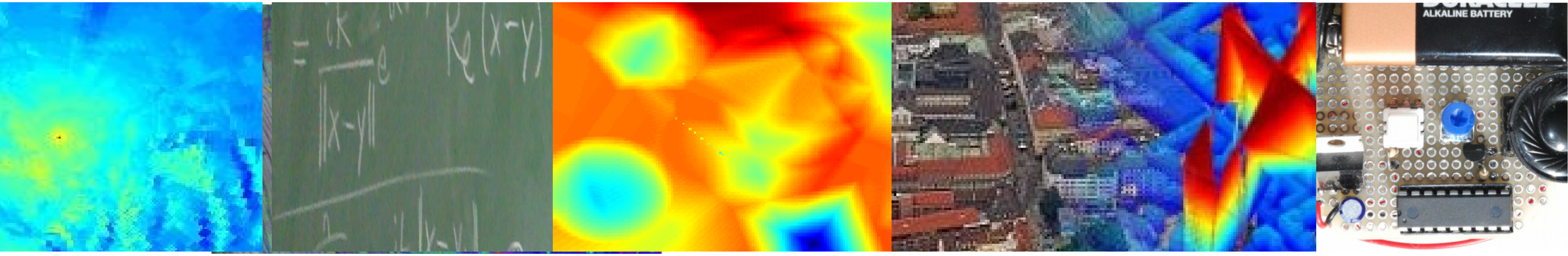
Heterogeneous fusion among homogeneous sensors

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



Michael Robinson

Data Structures as Sheaves



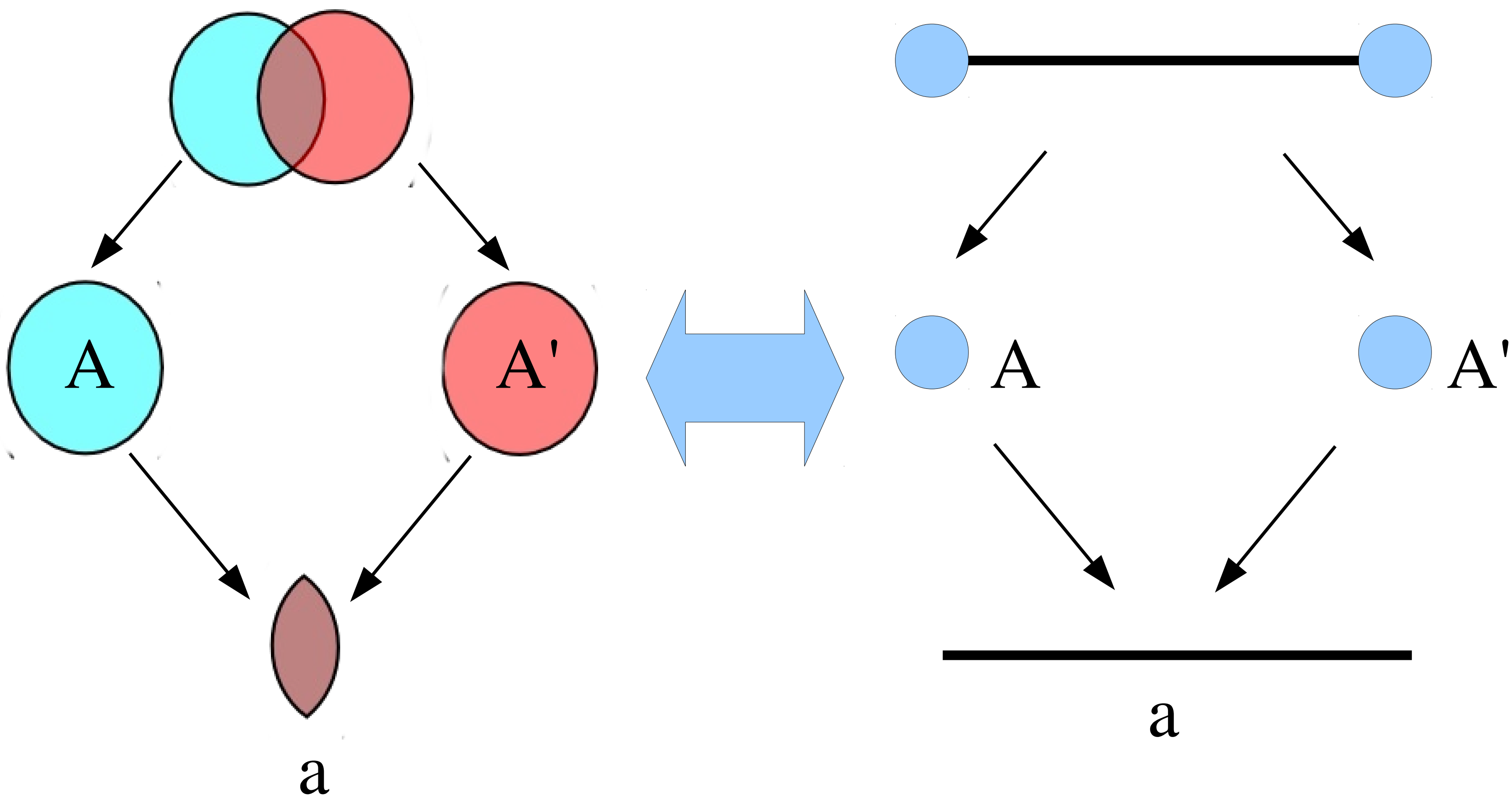
Michael Robinson

DARPA SIMPLEX Program

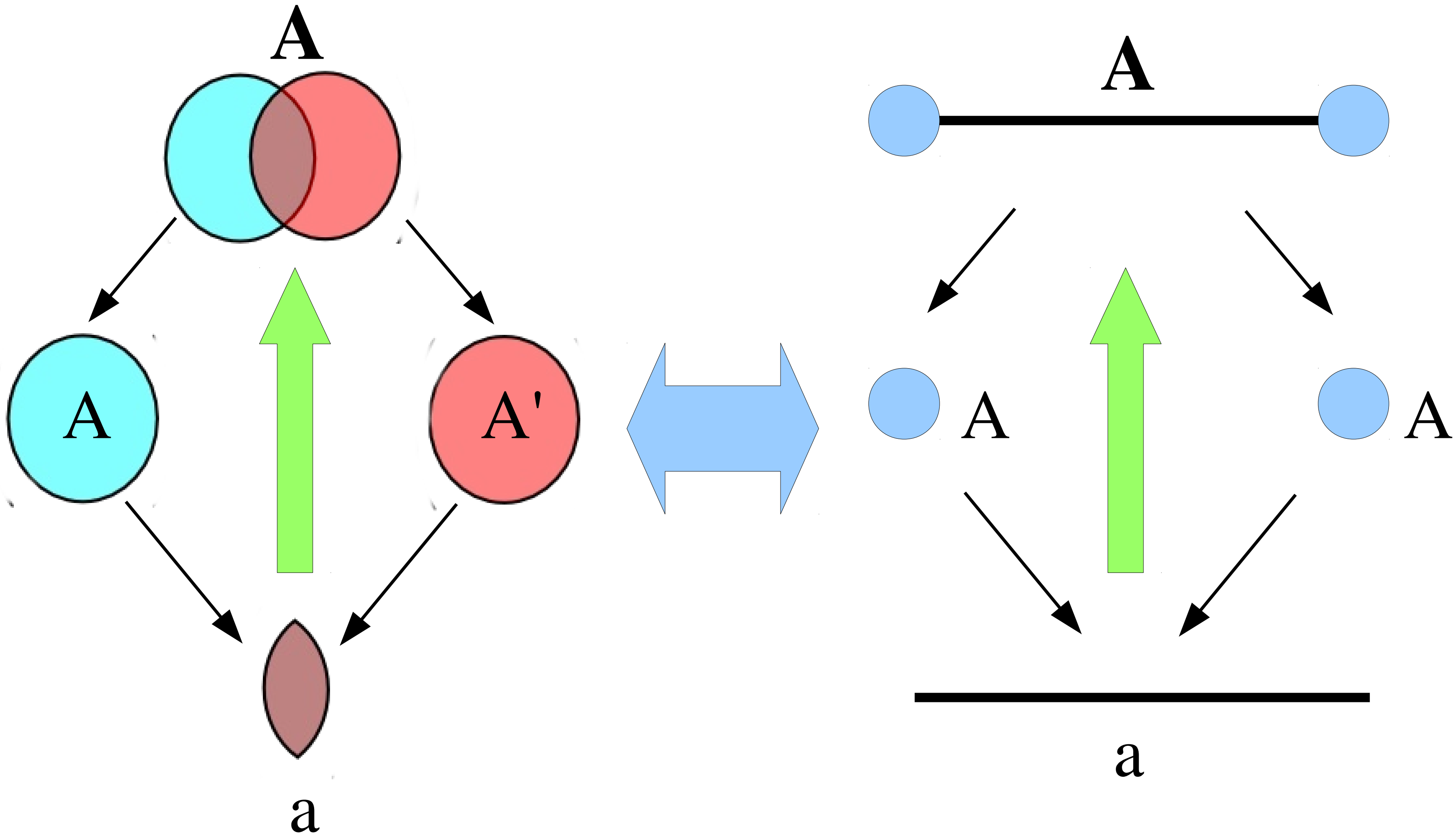
Session objectives

- How do sheaves extend well-known data structures?
- How do I translate between sheaf-based data structures?
- What can I do once I have a sheaf?

What is a sheaf?



What is a sheaf?

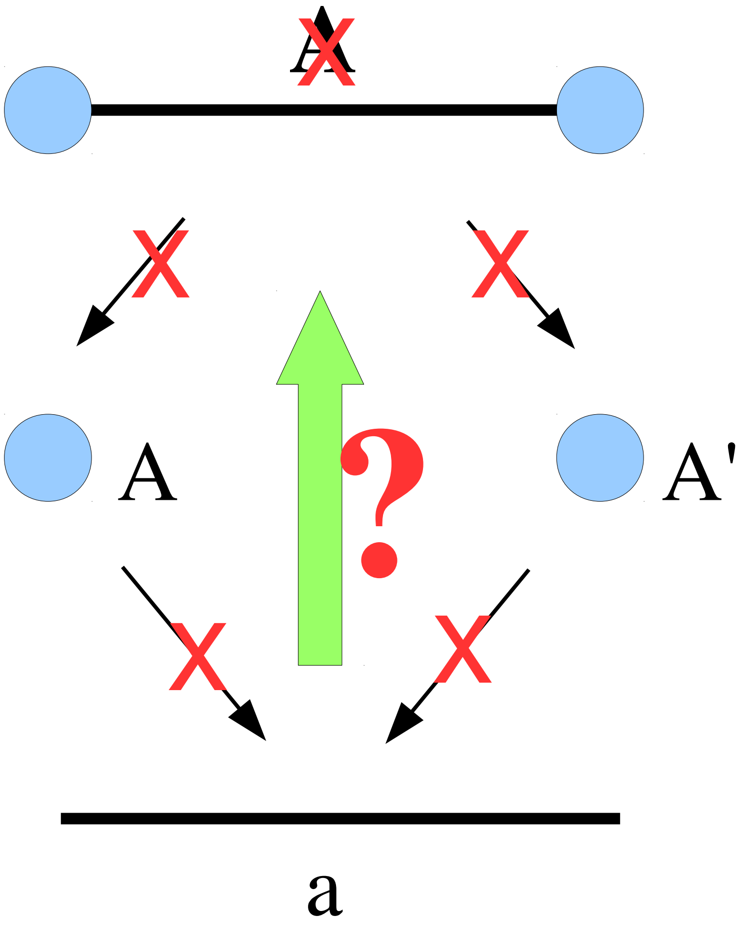


What isn't a sheaf?

If labels on the graph are not systematically related to one another...

Consistency checks become *ad hoc*

Cross-modality inference is no longer possible



Michael Robinson

Vertex- or (hyper)edge-weighted (hyper)graphs

Vertex weighted → sheaves

- Vertex has nontrivial stalk
- All restrictions are zero maps
- The resulting sheaf is flabby

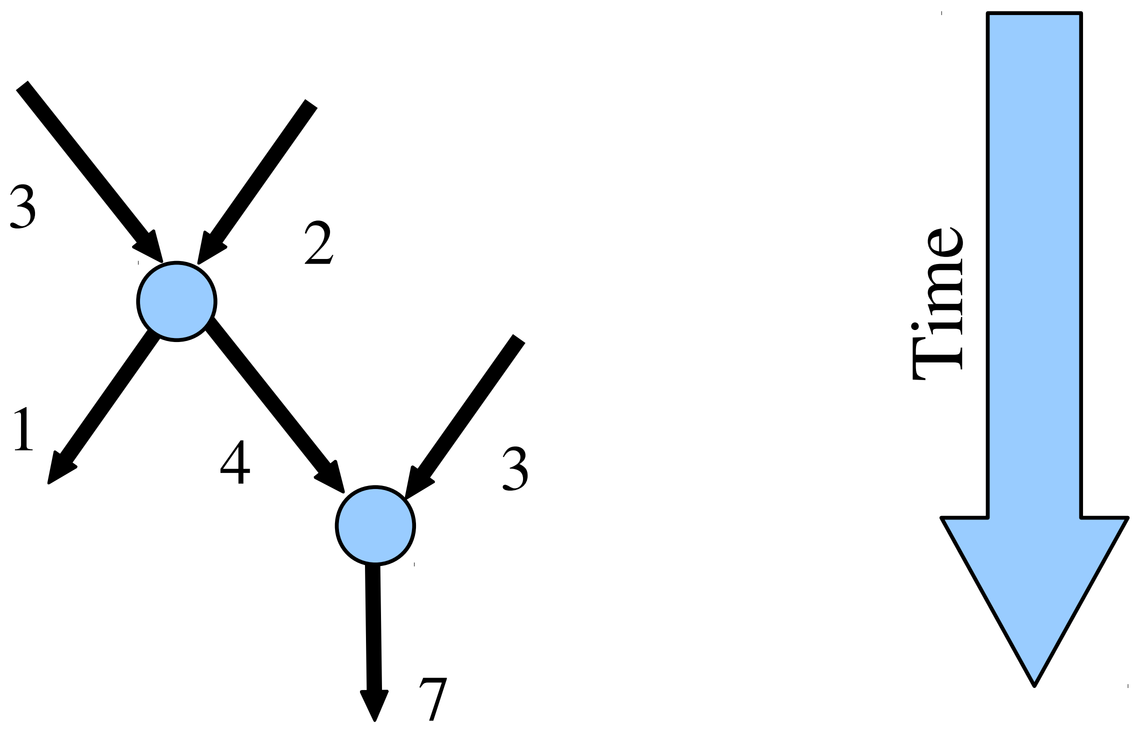
Hyperedge-weighted → cosheaves

- Toplex has nontrivial stalk
- All extensions are zero maps
- The resulting cosheaf is coflabby*

* This is such a fun term – but I don't see it much in use

Flow sheaves

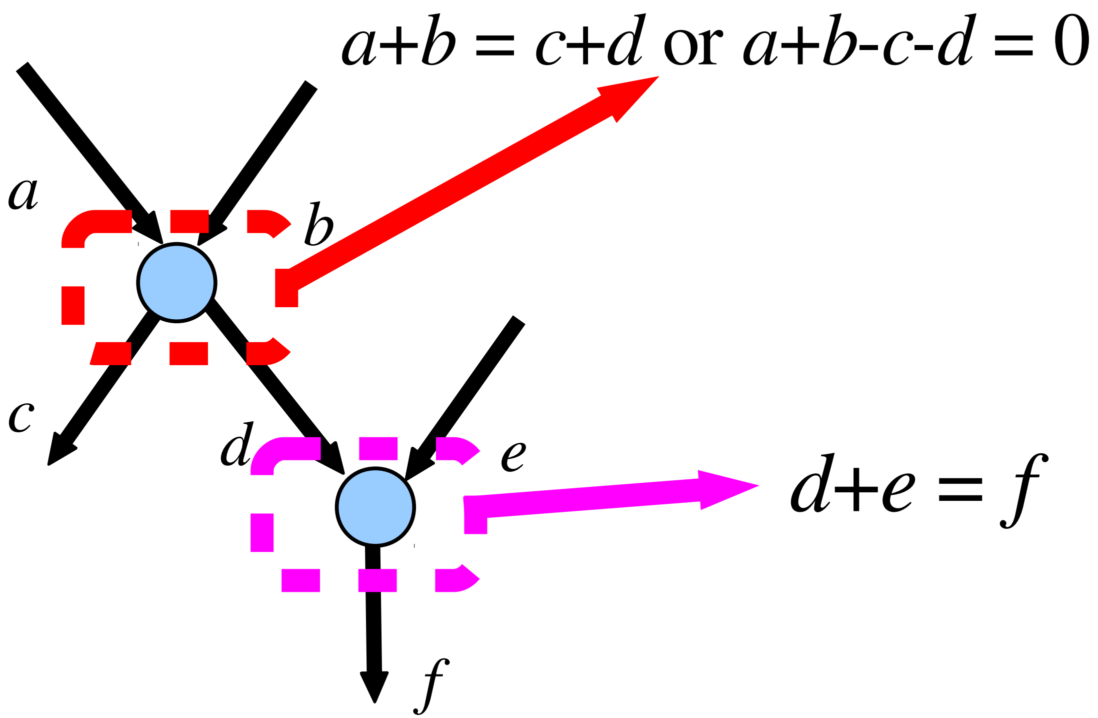
- Start with a collection of paths along which material flows
- Label each track segment with amount of material on that segment



Michael Robinson

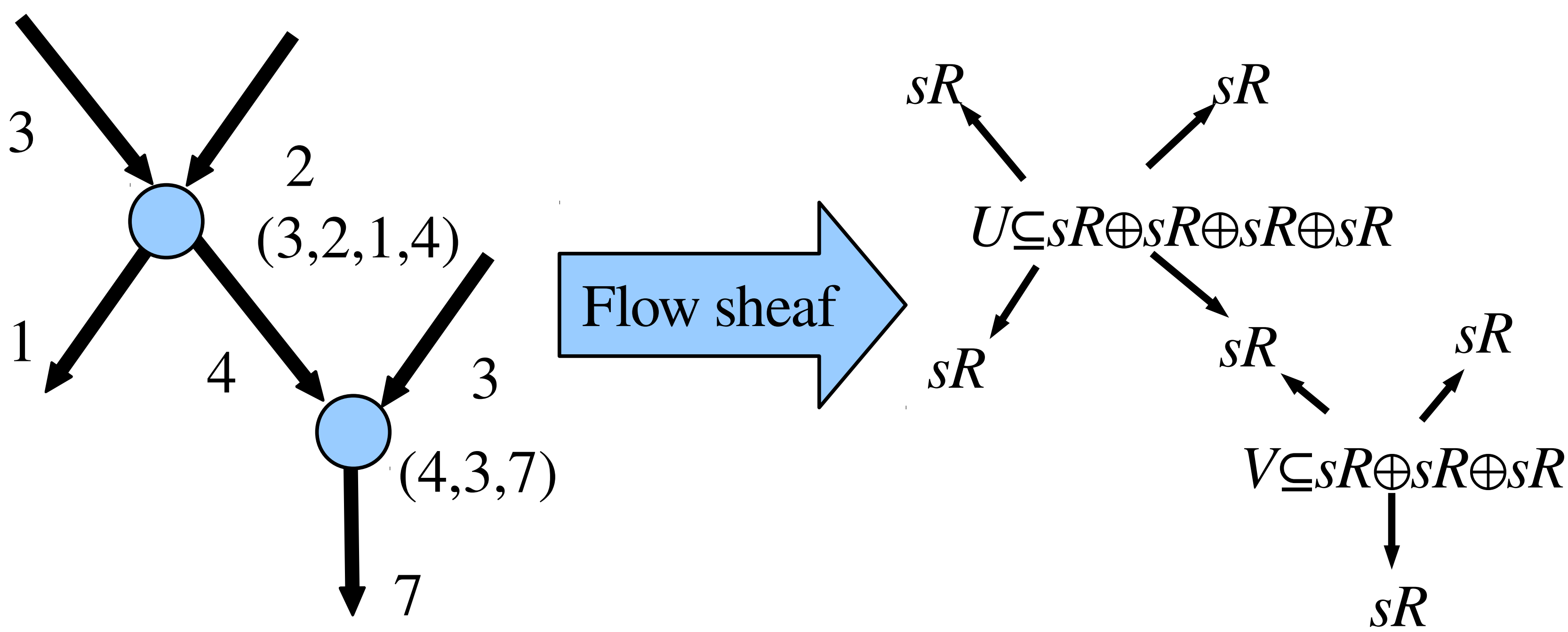
Flow sheaves

- Conservation law enforced at each vertex
- Depending on precisely how we count material (in $\mathbb{N}$ or $\mathbb{R}$, for instance), we might write the conservation law as



Flow sheaf

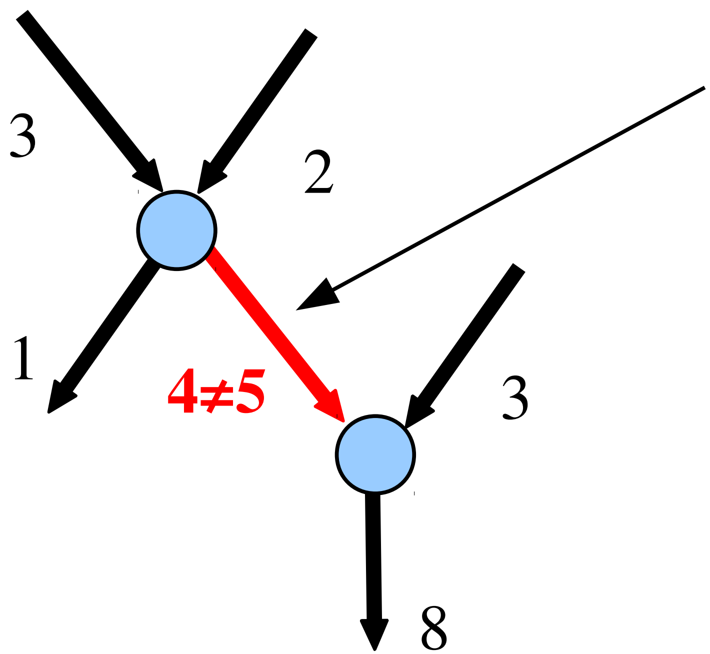
- Each degree *n* vertex is assigned a free *sR*-module* of rank (n – 1) for material measured in a semiring *sR*
- Restriction maps are projections



* or just a vector space if it seems easier

Local consistency

A flow sheaves encodes a notion of consistency between adjacent faces



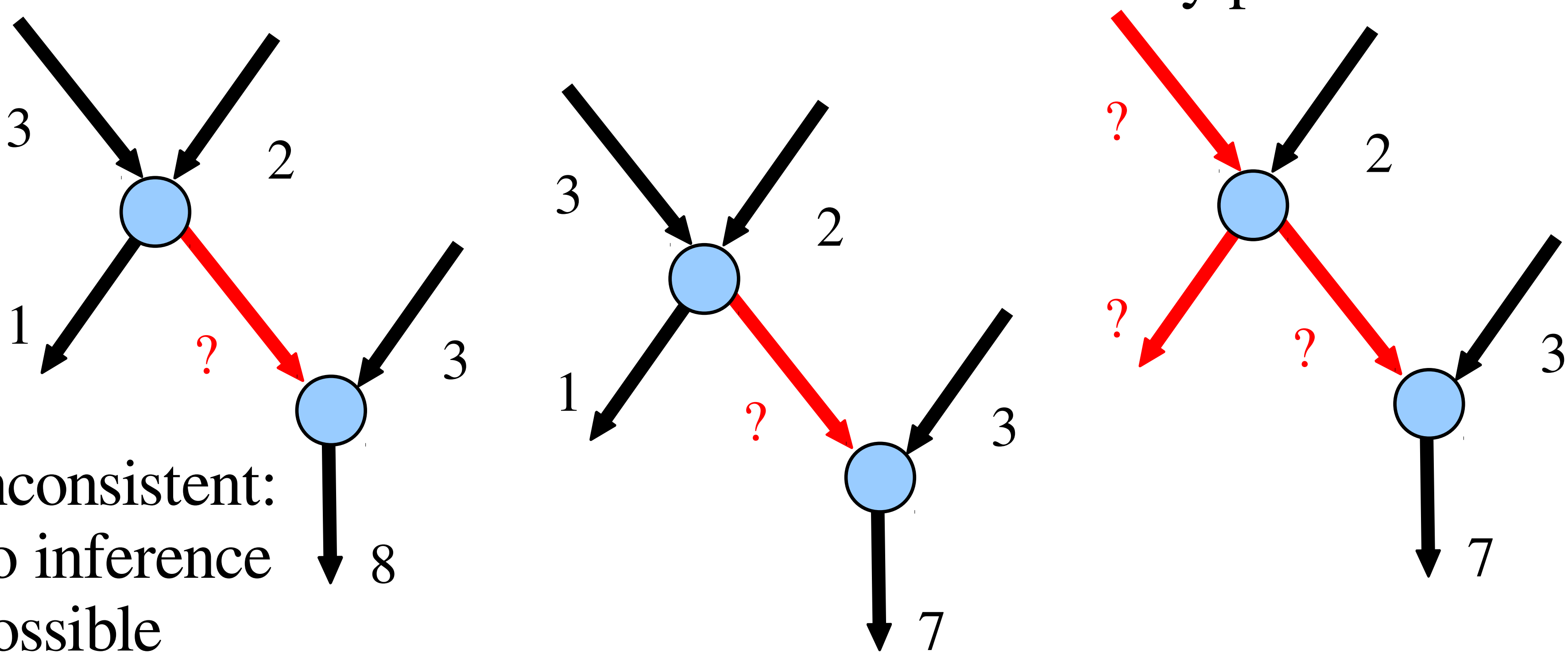
Cannot label this edge in a way consistent with the data

It's an indication that a flow was incorrectly measured somewhere in this vicinity

Michael Robinson

Inferential ambiguity

Depending on how we make measurements, we might not get "the full story" of the flow



Bayesian networks

Probability spaces

Start with a set of random variables $X_0, X_1, \ldots X_n$

Consider the set $P(X_0, X_1, \ldots, X_n)$ of all joint probability distributions over these random variables

- These are the nonnegative measures (generalized functions)

$$f = f(X_0, X_1, \ldots, X_n)$$

  with unit integral

- This is not a vector space – adding probability distributions doesn't yield another distribution

Michael Robinson

Marginalization cosheaf

There is a natural map

$$P(X_0,X_1,\ldots, X_n) \rightarrow P(X_0,X_1,\ldots, X_{n\text{-}1})$$

via *marginalization*, namely

$$f(X_0,X_1,\ldots, X_{n\text{-}1}) = \int f(X_0,X_1,\ldots, X_n)\, dX_n$$

- There similar maps for marginalizing out the other random variables, too
- This yields a cosheaf on the complete n-simplex!

Michael Robinson

Bayes' rule

Conditional probabilities produce maps going the other way… For instance,

$$P(X_0,X_1,\ldots, X_{n-1}) \rightarrow P(X_0,X_1,\ldots, X_n)$$

is parameterized by functions C

$$F(X_0,X_1,\ldots, X_n) = C(X_0,X_1,\ldots, X_n)\, f(X_0,X_1,\ldots, X_{n-1})$$

usually, one writes the arguments to C like

$$C = C(X_n \mid X_0,X_1,\ldots, X_{n-1})$$

So… conditional probabilities yield a sheaf on part of the n-simplex

Michael Robinson

Small Bayes net example

- Consider two binary random variables X and Y with a given conditional $C(Y \mid X)$

$$P(X) \xleftarrow{\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}} P(X,Y) \xrightarrow{\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}} P(Y)$$

Marginalization cosheaf

$$P(X) \longrightarrow P(X,Y) \qquad P(Y)$$

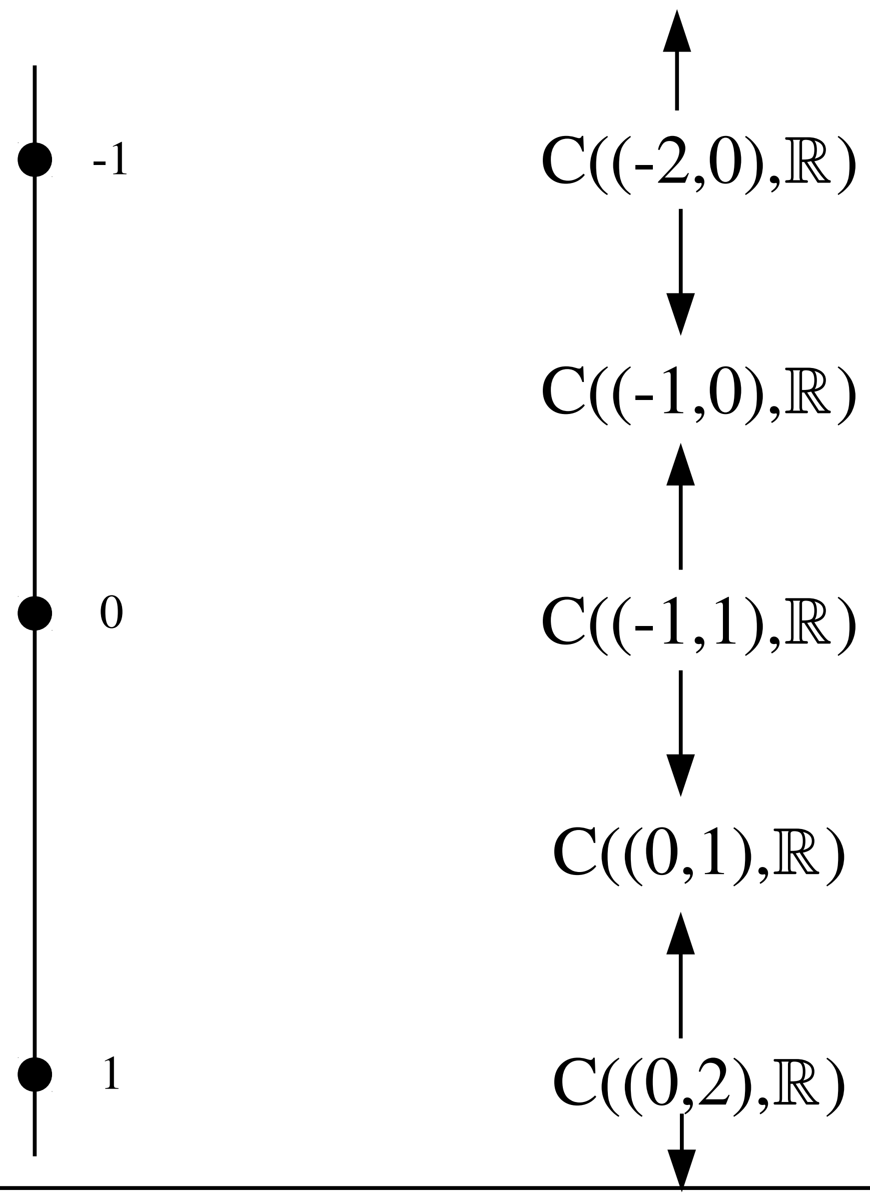
$$\begin{pmatrix} p(0|0) & 0 \\ p(1|0) & 0 \\ 0 & p(0|1) \\ 0 & p(1|1) \end{pmatrix} = C(Y \mid X)$$

Conditional sheaf

Linear translation-invariant filters

How does a sheaf model a signal?

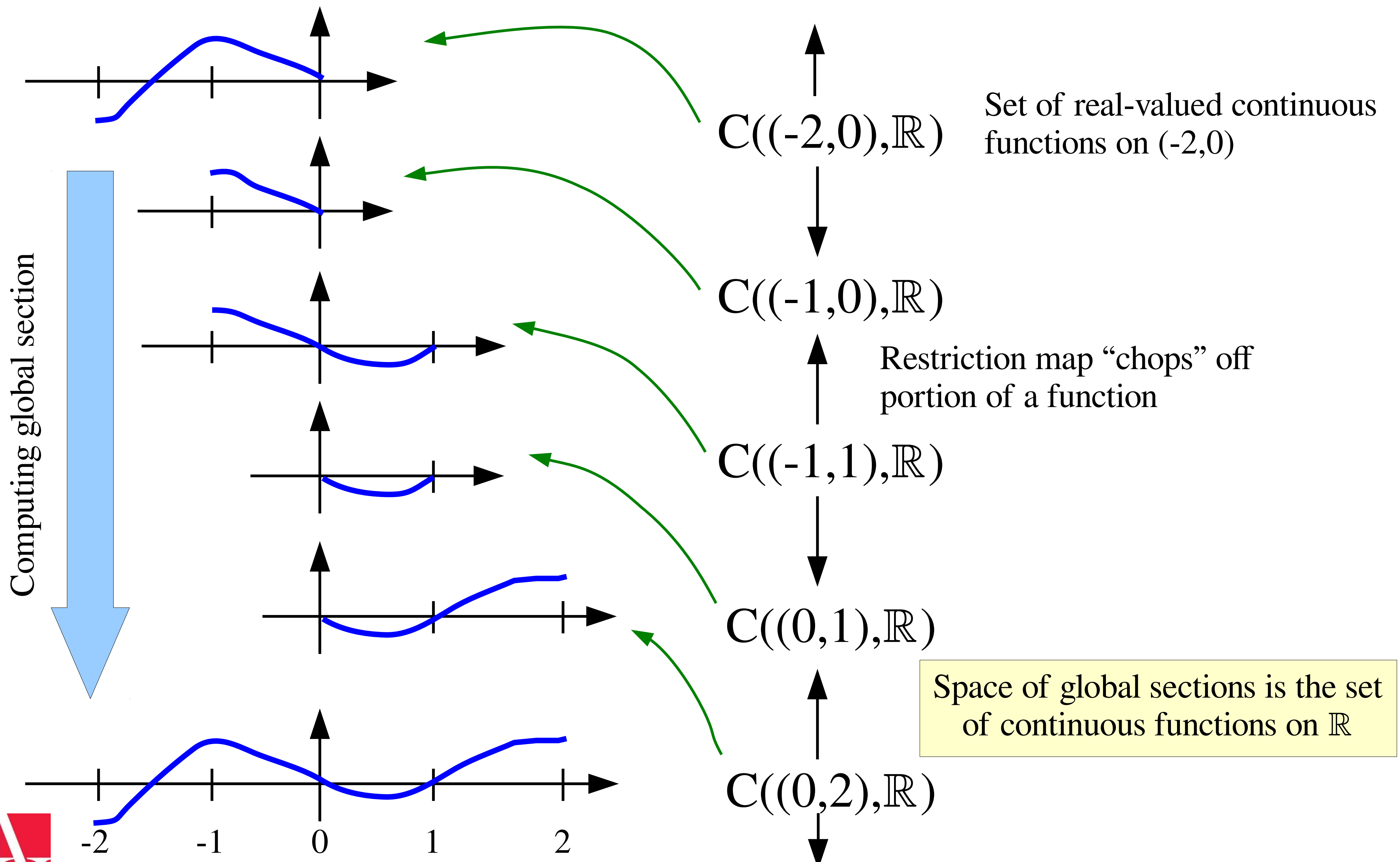
Simplicial complex for ℝ



The sheaf of continuous functions

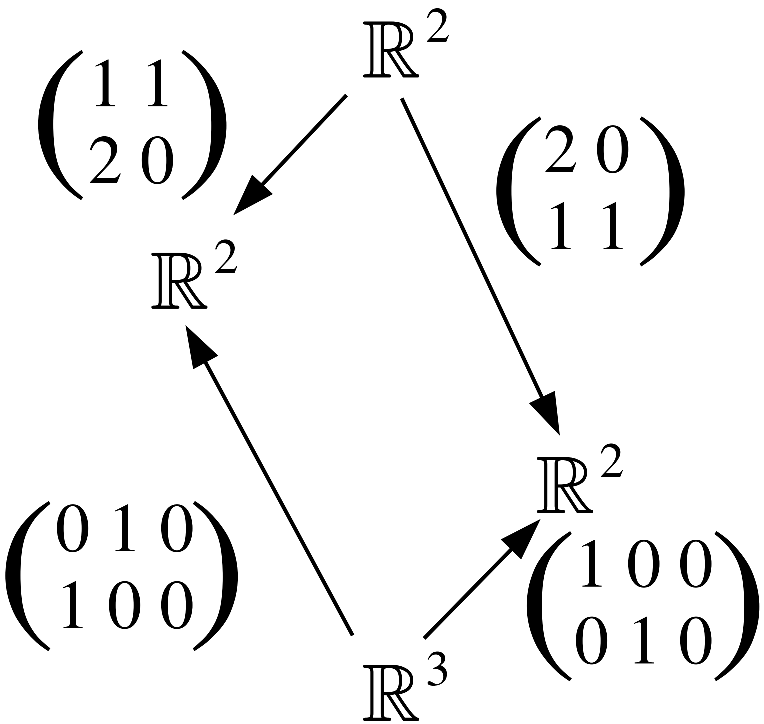
Michael Robinson

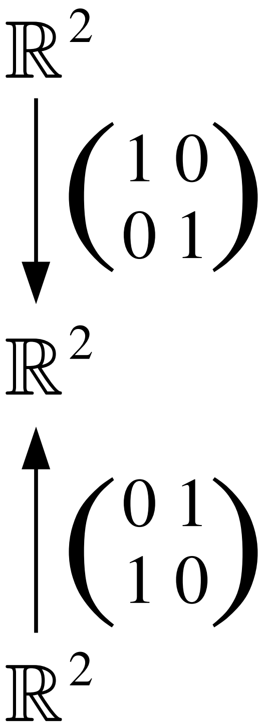
How does a sheaf model a signal?



A *sheaf morphism* ...

- … takes data in the stalks of two sheaves …

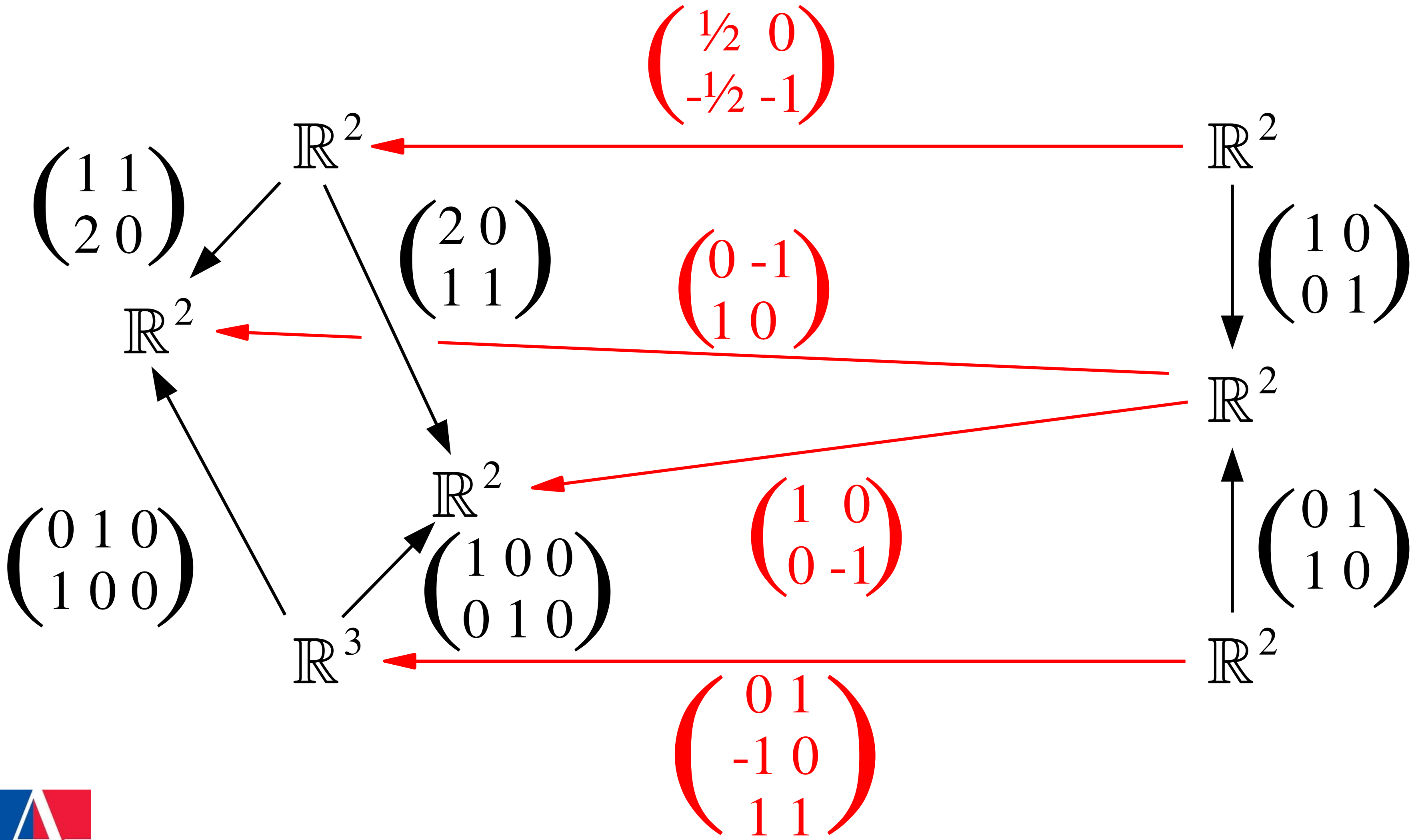




Michael Robinson

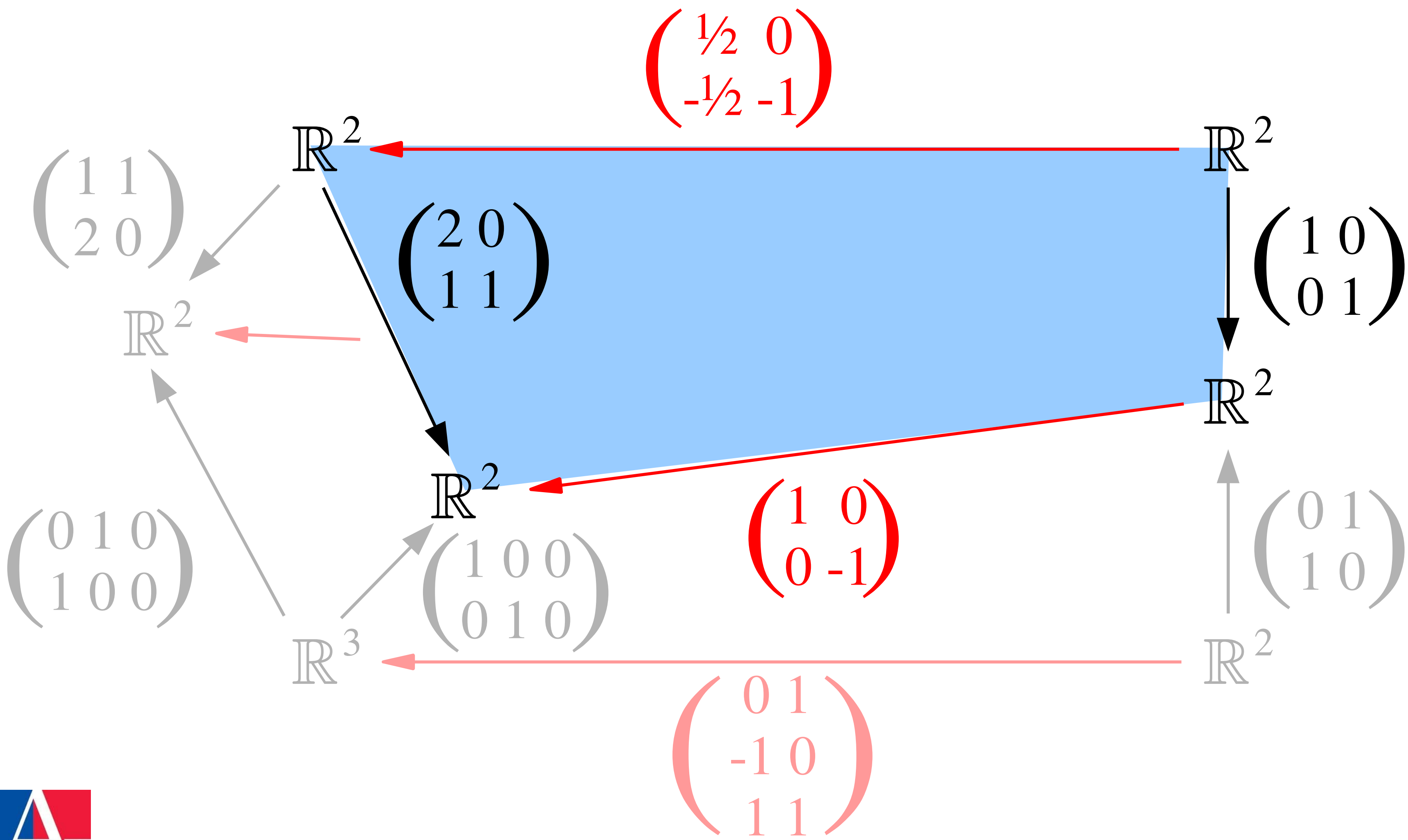
A *sheaf morphism* …

- … and relates them through linear maps …



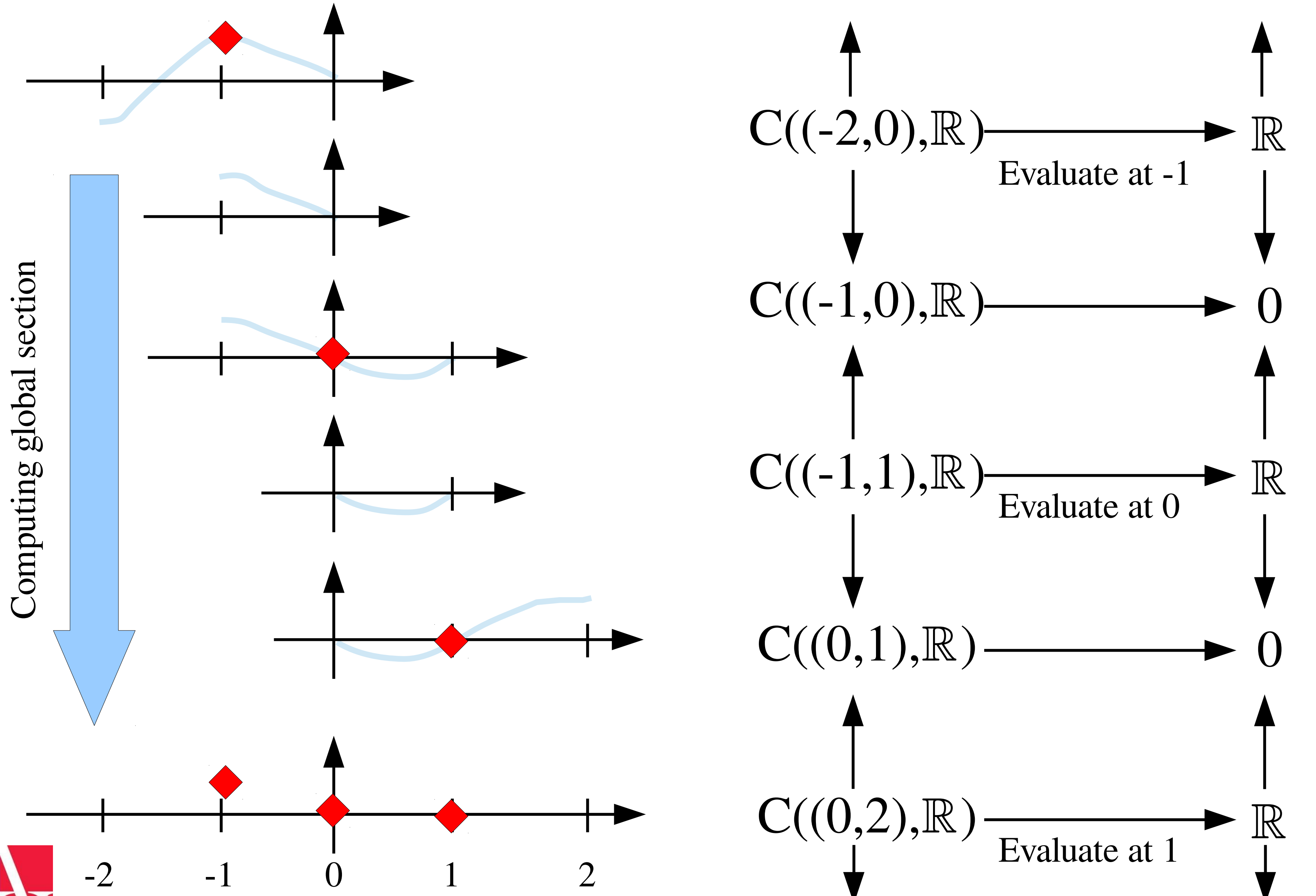
A *sheaf morphism* ...

- … so the diagram commutes!



Michael Robinson

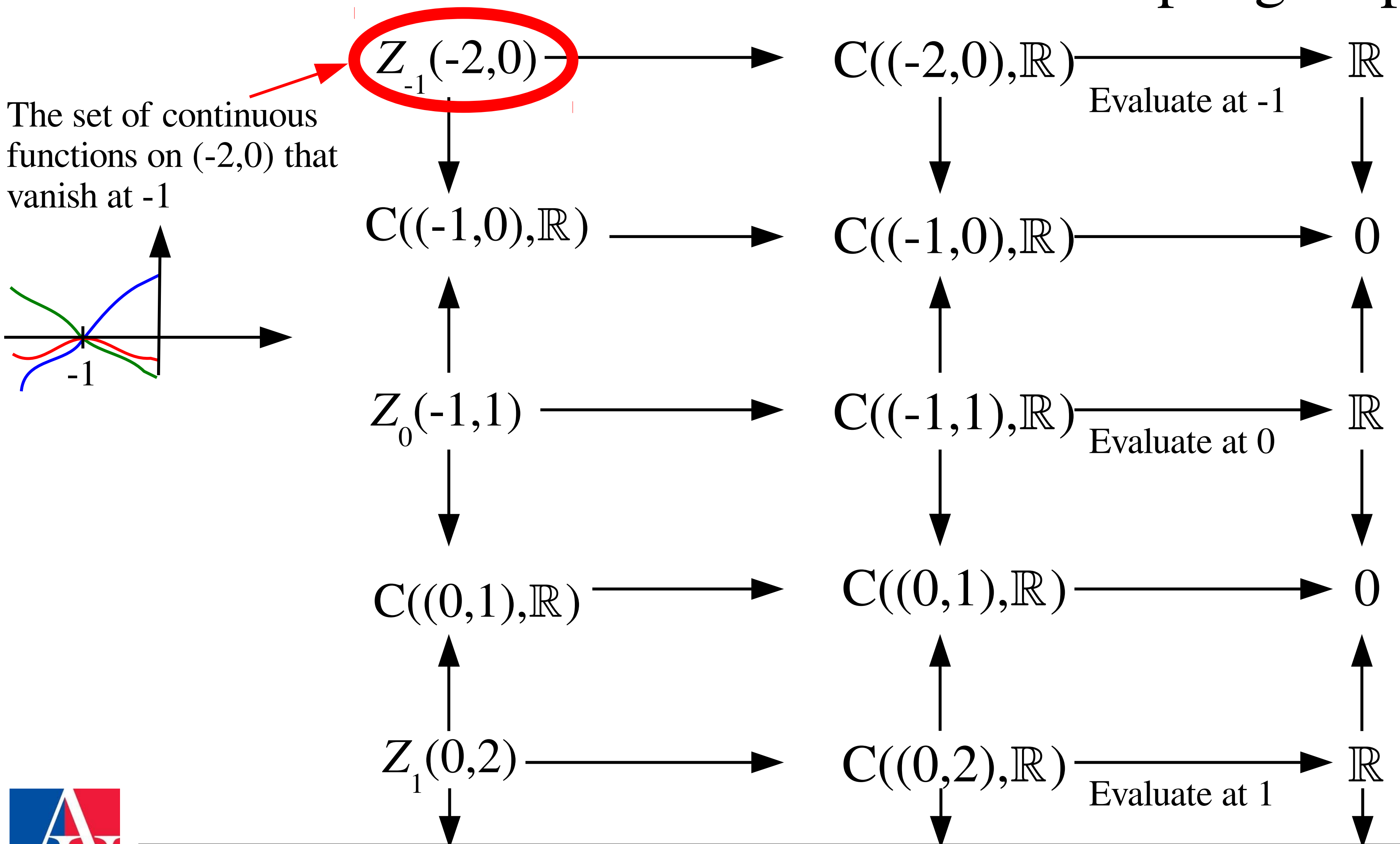
A *sampling morphism* is ...



Michael Robinson

An *ambiguity sheaf* is ...

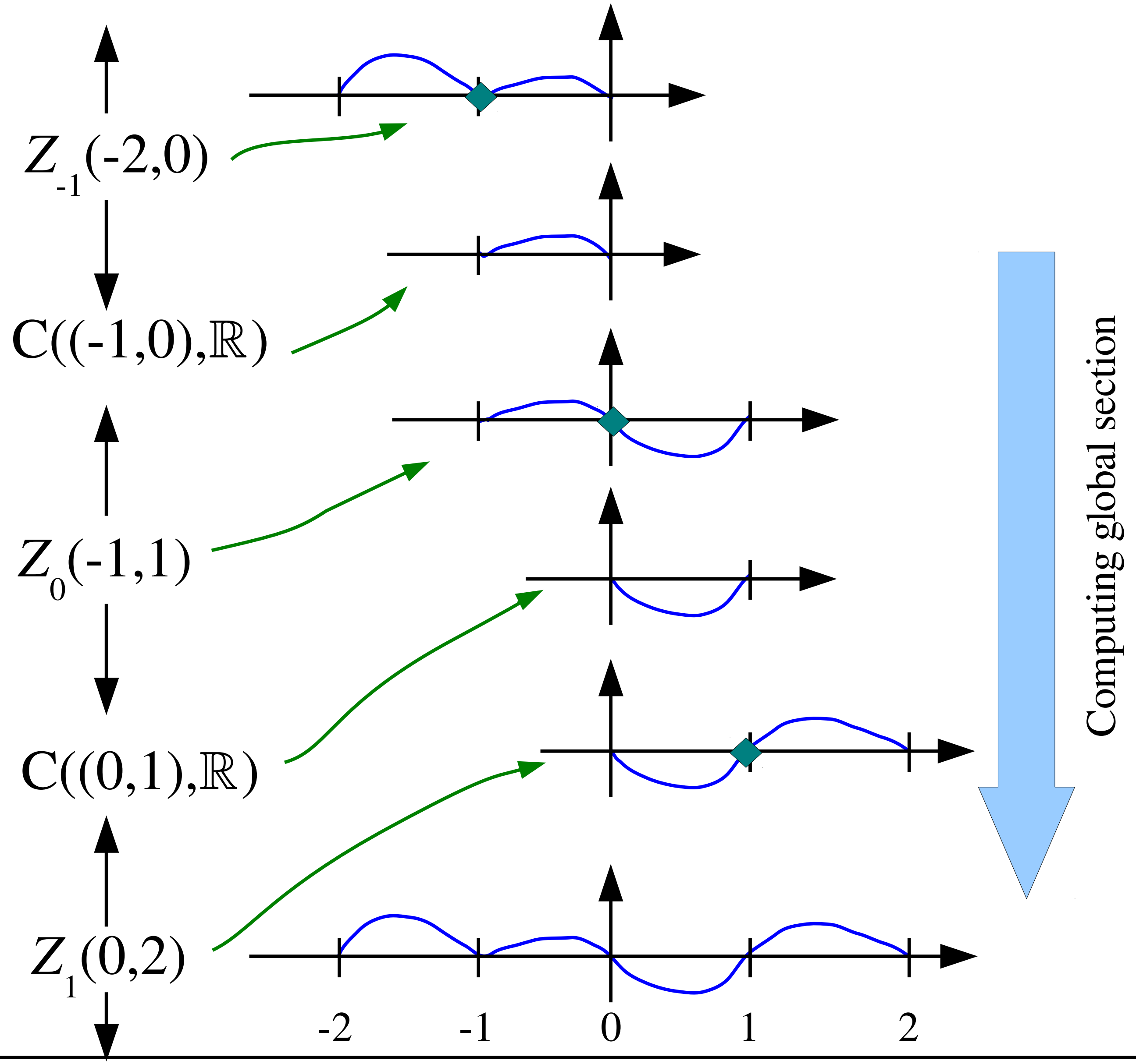
- The collection of *kernels* of these sampling maps



Michael Robinson

An *ambiguity sheaf* is ...

Sections of the ambiguity sheaf are functions that appear to be all zero under the sampling



The general sampling theorem

- Given a sheaf *S* and a sampling of it, construct the ambiguity sheaf *A*

Theorem:

- Perfect reconstruction of global sections of *S* is possible if and only if the only global section of *A* is the zero function

- "No ambiguities means it's possible to reconstruct"

Michael Robinson

Nyquist-Shannon sampling

- Encode signals as a sheaf of bandlimited functions *BF* over $\mathbb{R}$, with bandwidth *B*
- It's easier to work in the frequency domain:

$$BF = \{ f \in C(\mathbb{R},\mathbb{C}) \mid \operatorname{supp} f \subseteq [-B, B] \}$$

- Samples are taken at integers, obtained by inverse Fourier transform
- For instance at *n*, we sample using the function M_n

$$M_n(f) = \int_{-B}^{B} f e^{-2n\pi i x} dx$$

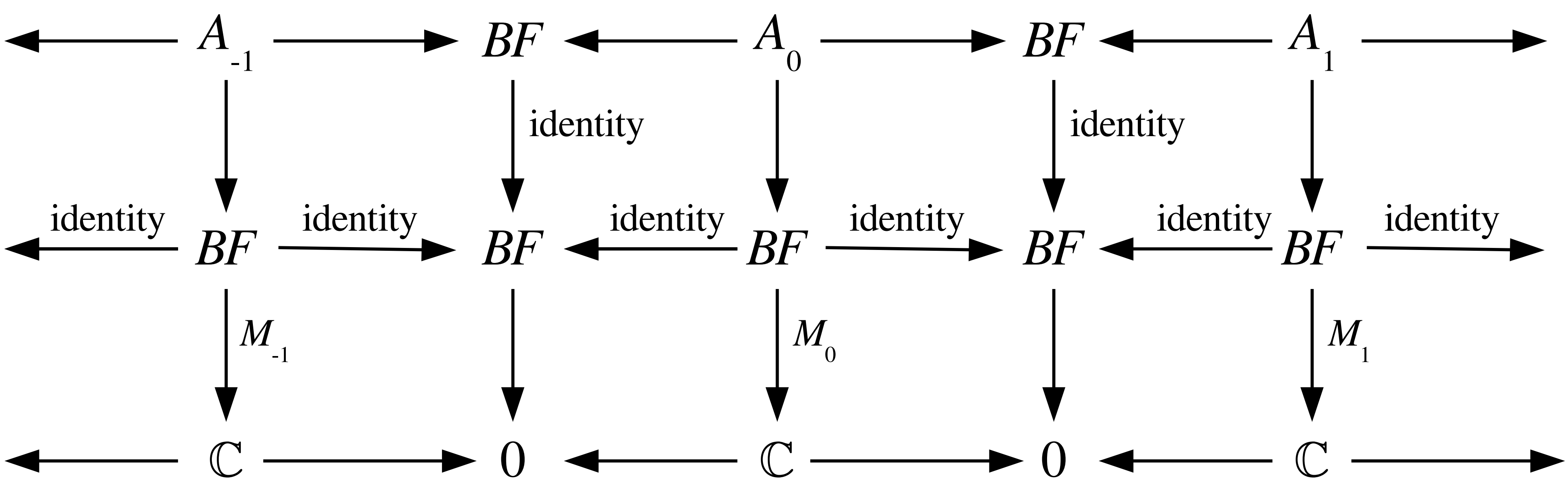
Michael Robinson

Nyquist-Shannon sampling

- The ambiguity sheaf identifies bandlimited functions with zeros at specific locations

$$A_n = \{ f \in BF \mid M_n(f) = 0 \}$$

$$= \{\text{Bandlimited functions that are zero at } n\}$$



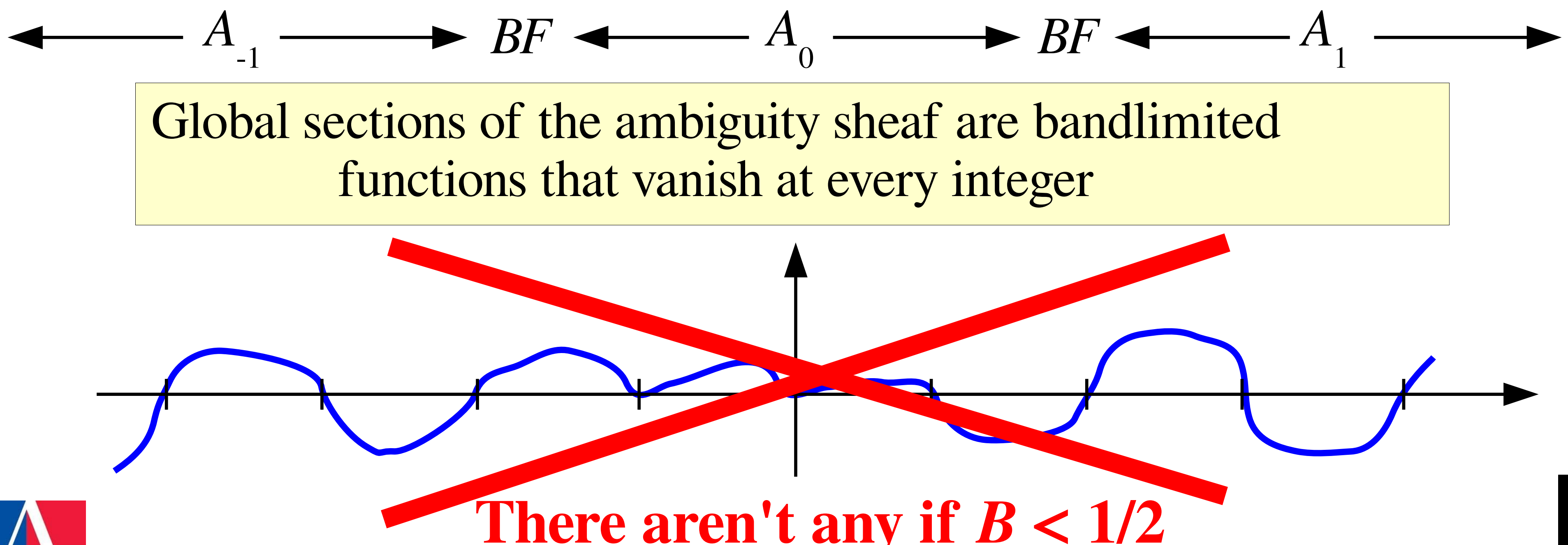
Michael Robinson

Nyquist-Shannon sampling

- The ambiguity sheaf identifies bandlimited functions with zeros at specific locations

$$A_n = \{ f \in BF \mid M_n(f) = 0 \}$$

$$= \{\text{Bandlimited functions that are zero at } n\}$$



Michael Robinson

Filters as sheaf morphisms

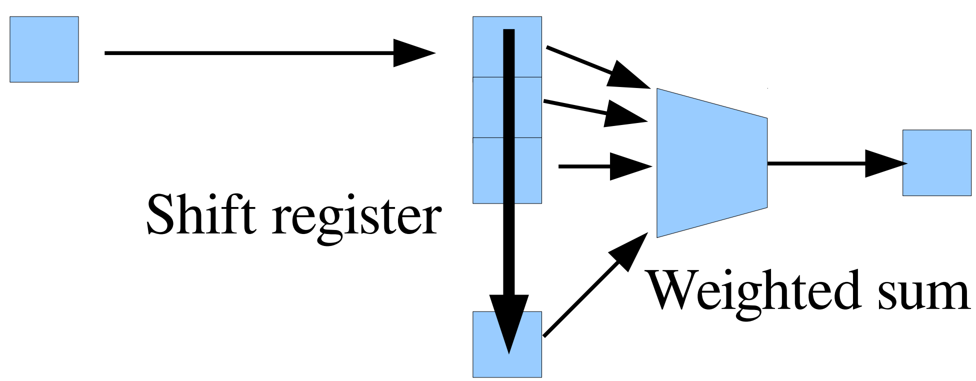
- <u>Theorem</u>: Every discrete-time LTI filter can be encoded as a sequence of two sheaf morphisms

$$S_1 \longleftarrow S_2 \longrightarrow S_3$$

Sheaf formalism

Input — Internal state — Output

Hardware



Michael Robinson

Input sheaf

- Sections of this sheaf are timeseries, instead of continuous functions

$$\rightarrow 0 \leftarrow \mathbb{R} \rightarrow 0 \leftarrow \mathbb{R} \rightarrow 0 \leftarrow \mathbb{R} \rightarrow 0 \leftarrow \mathbb{R}$$

Output sheaf

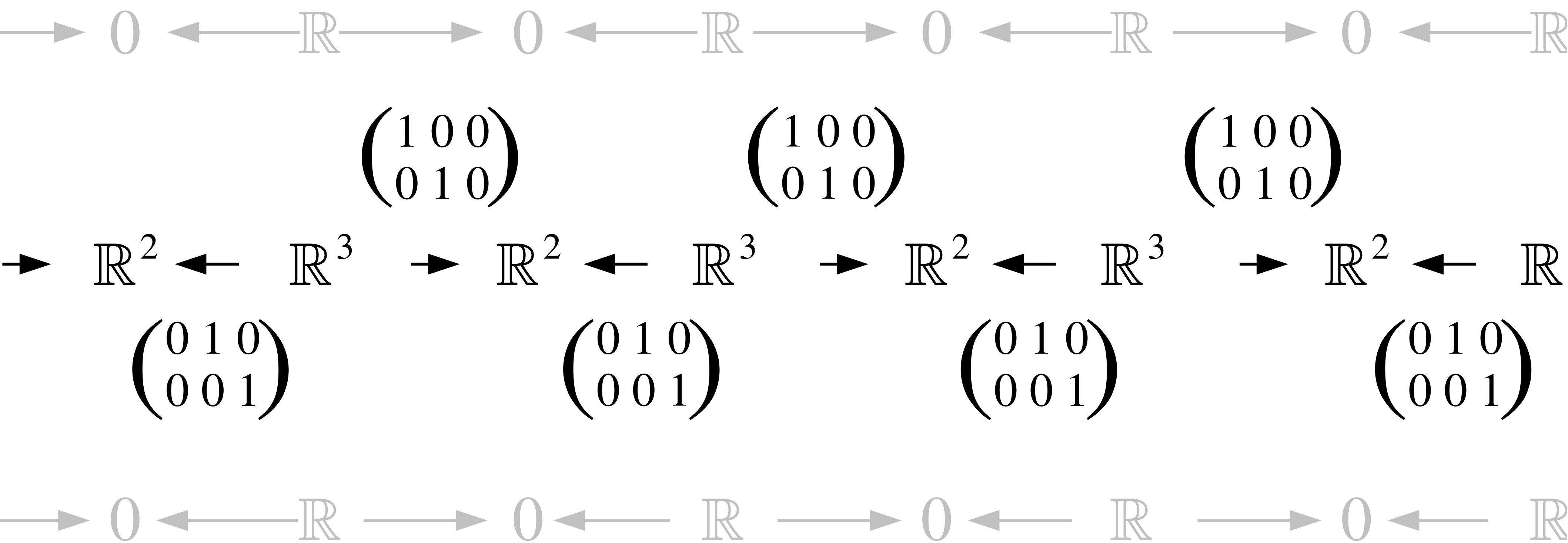
- The output sheaf is the same

$$\longrightarrow 0 \longleftarrow \mathbb{R} \longrightarrow 0 \longleftarrow \mathbb{R} \longrightarrow 0 \longleftarrow \mathbb{R} \longrightarrow 0 \longleftarrow \mathbb{R}$$

$$\longrightarrow 0 \longleftarrow \mathbb{R} \longrightarrow 0 \longleftarrow \mathbb{R} \longrightarrow 0 \longleftarrow \mathbb{R} \longrightarrow 0 \longleftarrow \mathbb{R}$$

The internal state

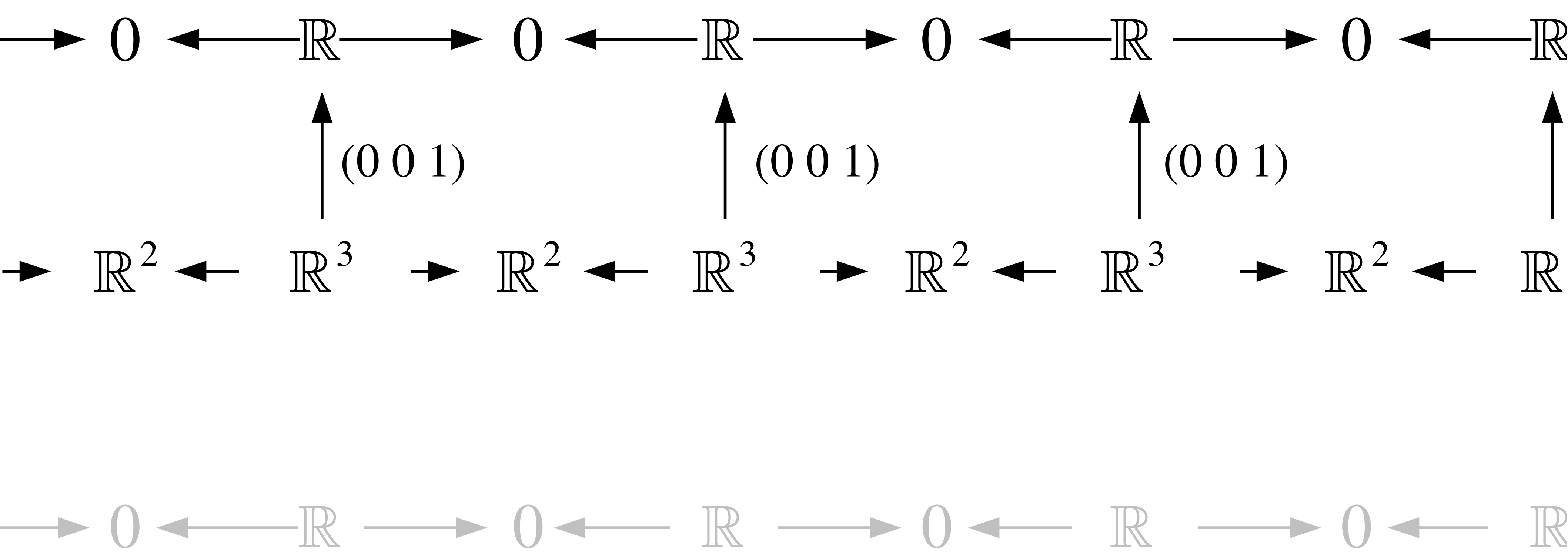
- Contents of the shift register at each timestep
- $N = 3$ shown



Michael Robinson

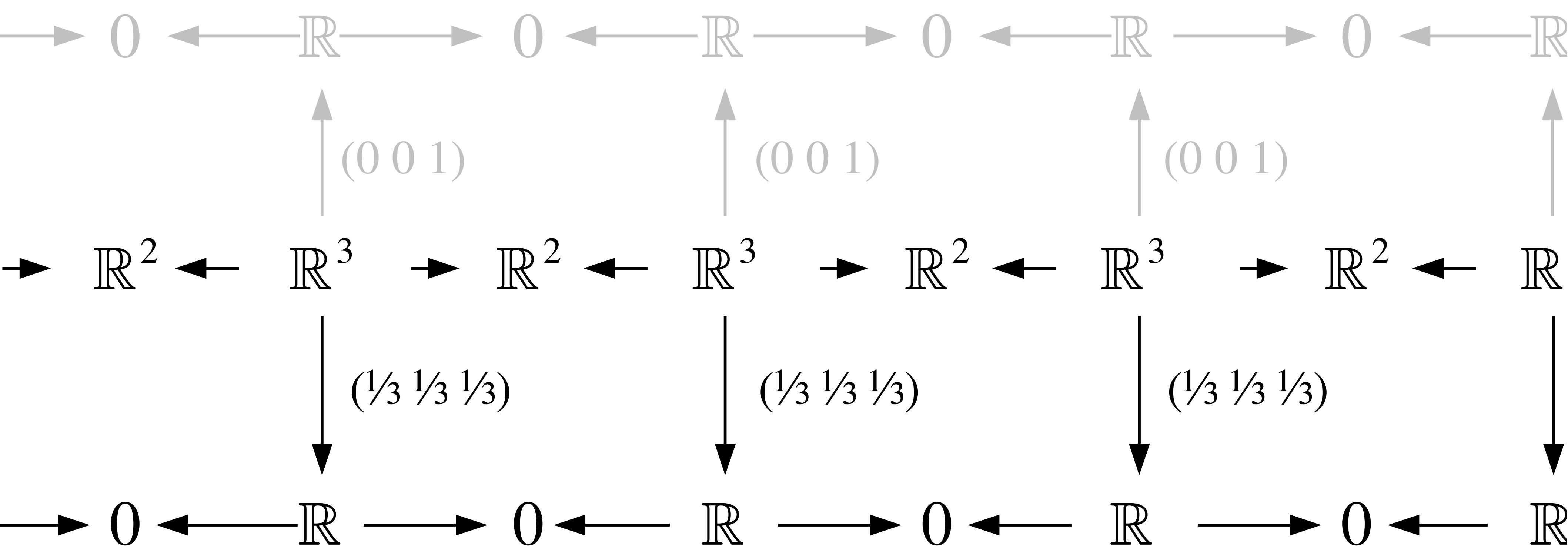
The internal state

- Loads a new value with each timestep



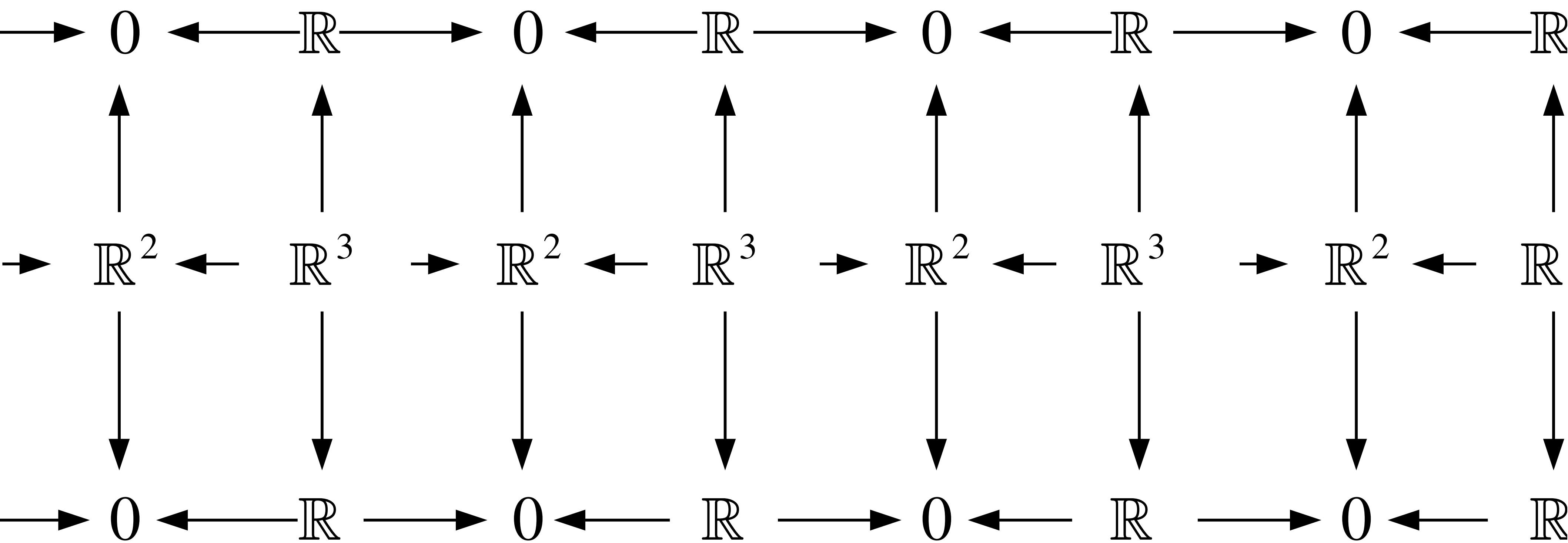
The internal state

- Produces average of the shift register at each timestep



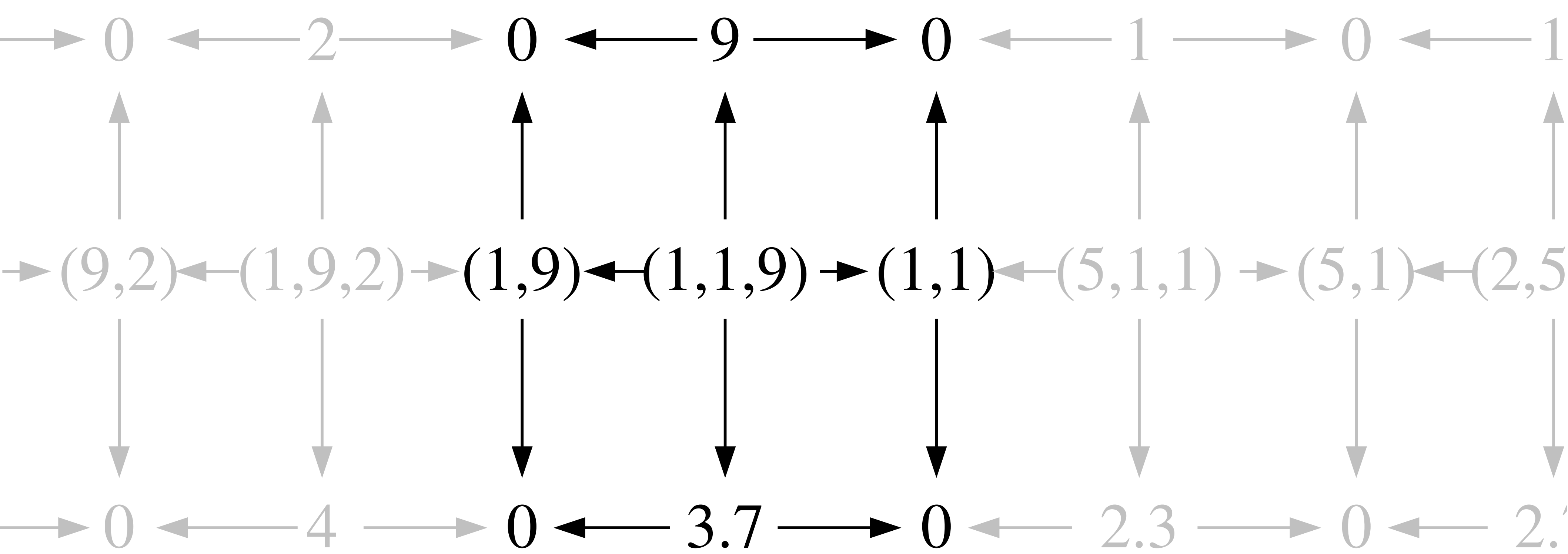
Finishing both morphisms

- Put in a few zero maps!



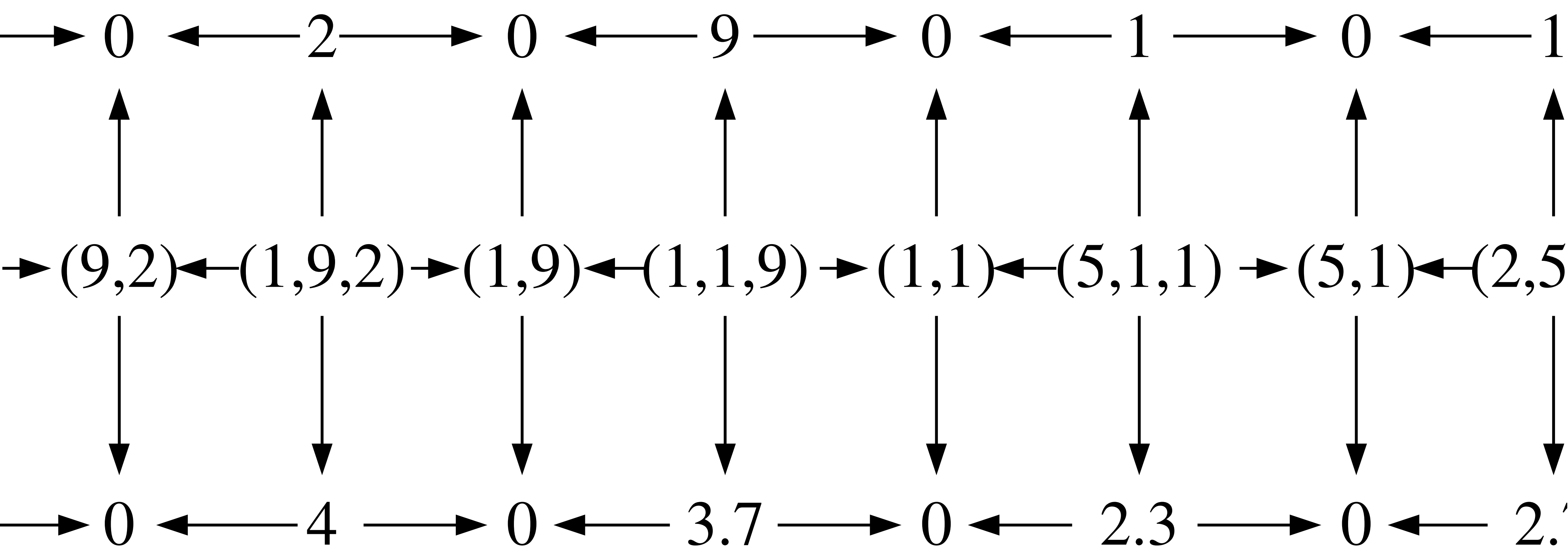
A single timestep

- Sections of the sheaves linked together give possible input/output combinations of the filter



Of course, this extends...

- Sections of the sheaves linked together give possible input/output combinations of the filter



Michael Robinson

How this formalism helps...

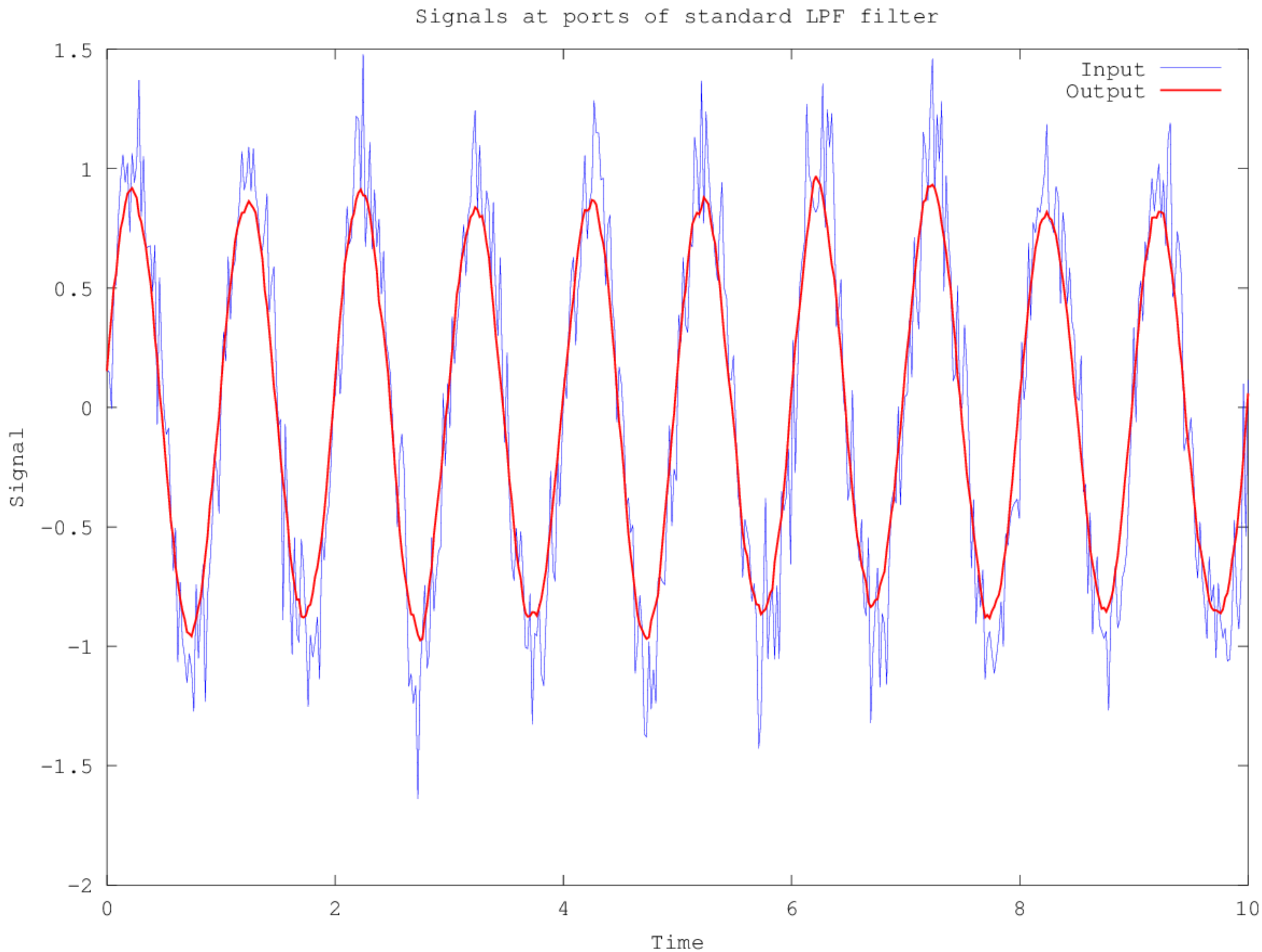
- Of course, it corresponds nicely to the hardware

… BUT...

- It's easy to splice in **nonlinear** operations
- It works on nontrivial base spaces: A **systematic study** of filtering is **now possible** on cell complexes

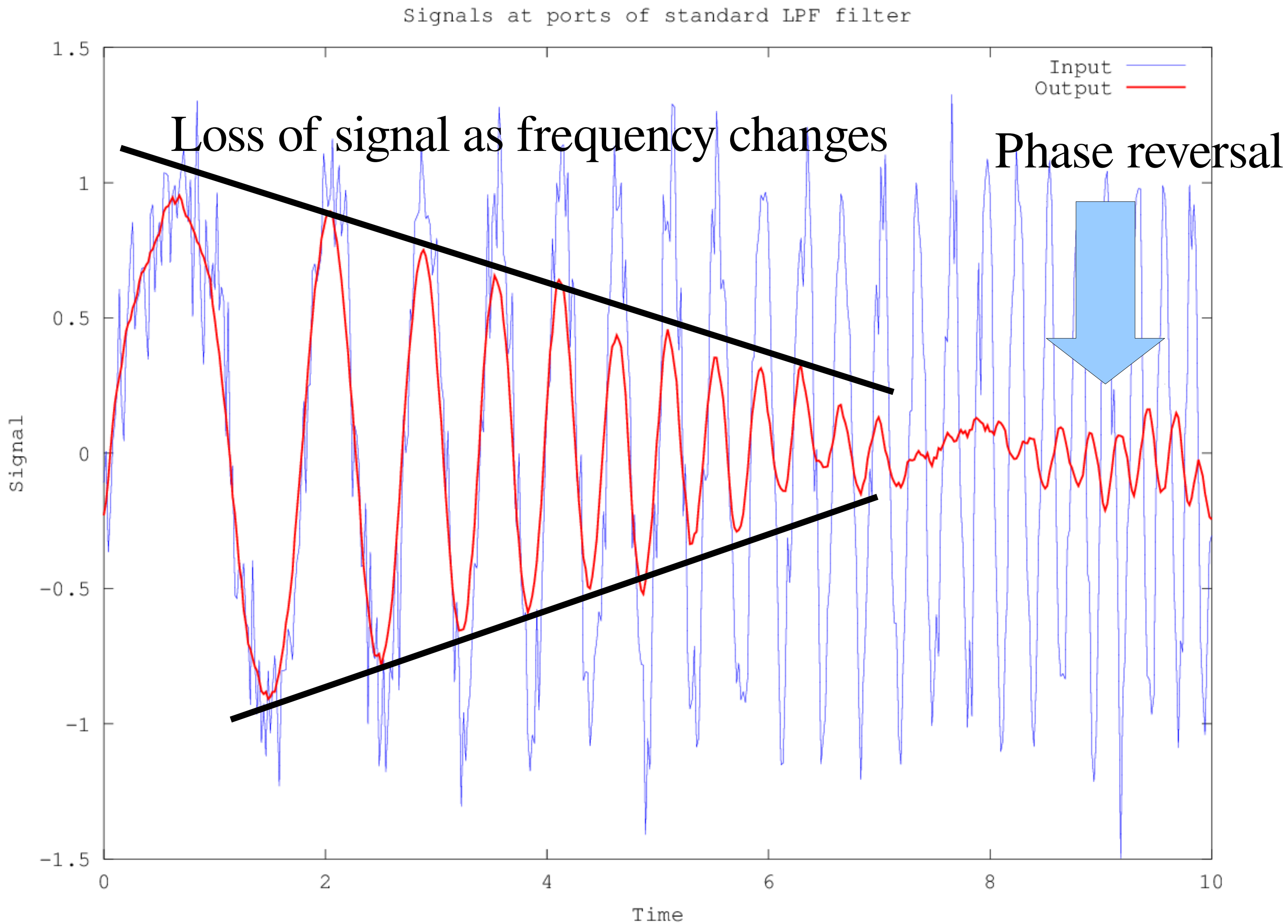
Sheaves make it easy to invent *topological filters* that have controlled performance characteristics

Filtering sinosoids from noise



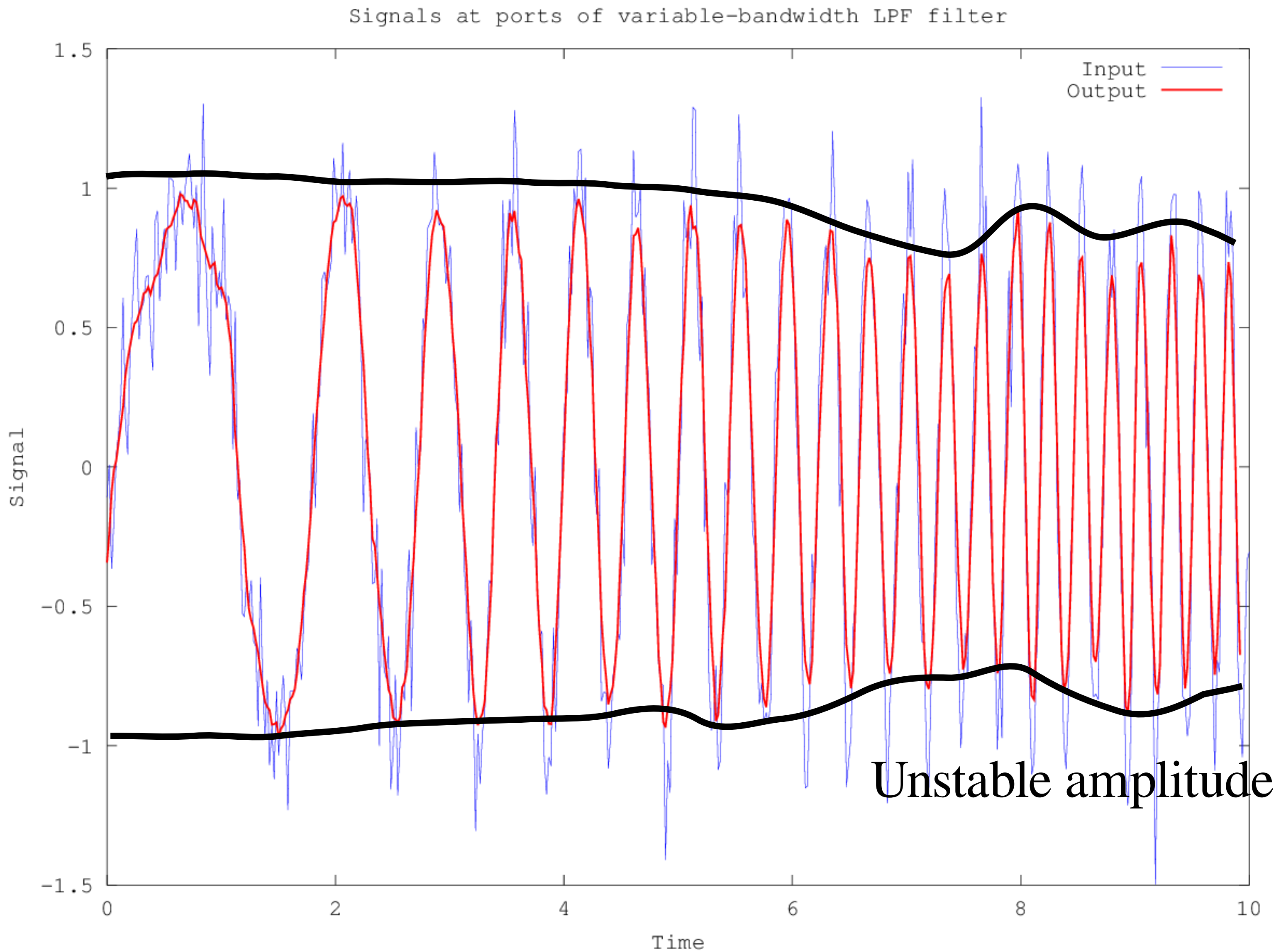
Michael Robinson

Filtering out chirpy signals



Michael Robinson

Filtering out chirpy signals

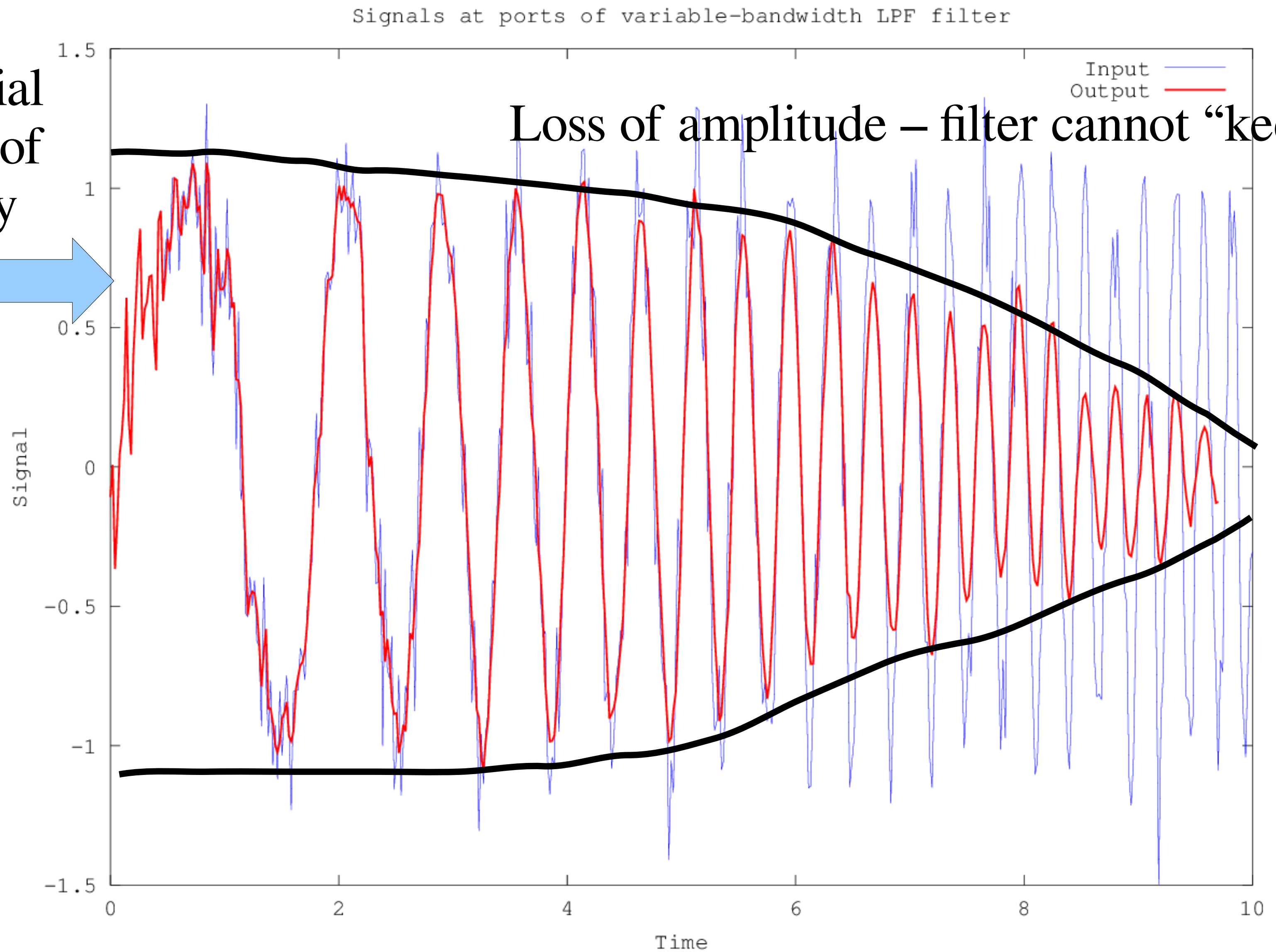


Michael Robinson

Filtering out chirpy signals

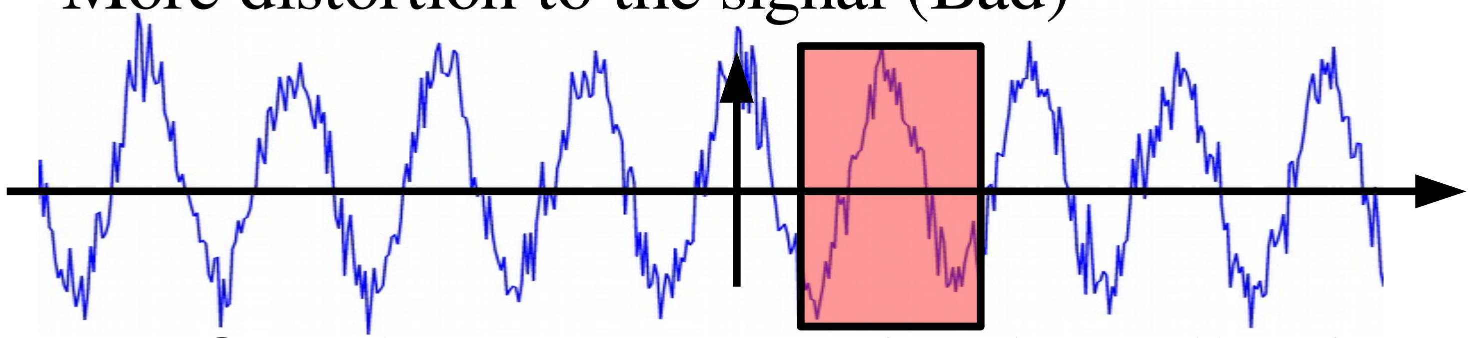
Poor initial estimate of frequency

Loss of amplitude – filter cannot "keep up"!

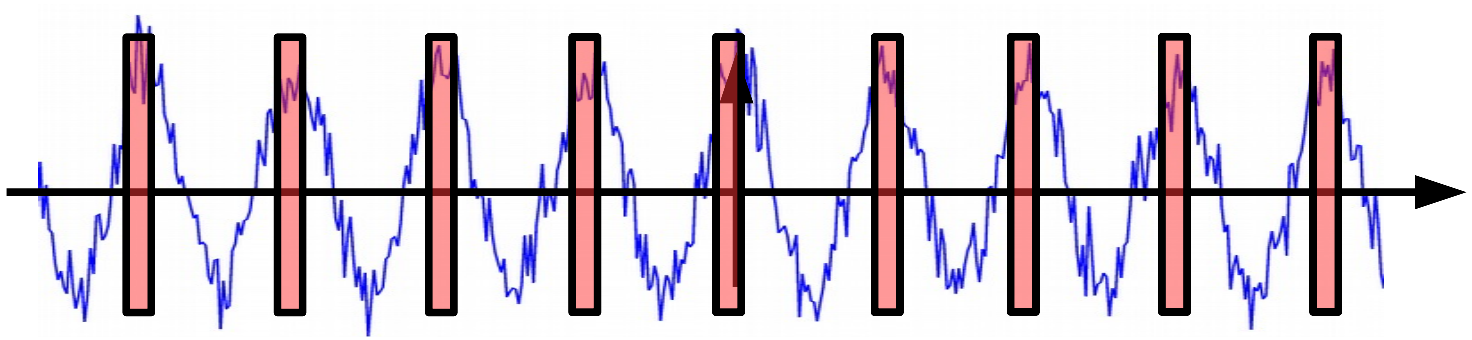


Circumventing bandwidth limits

- More averaging in a connected window leads to:
  - More noise cancellation (Good)
  - More distortion to the signal (Bad)

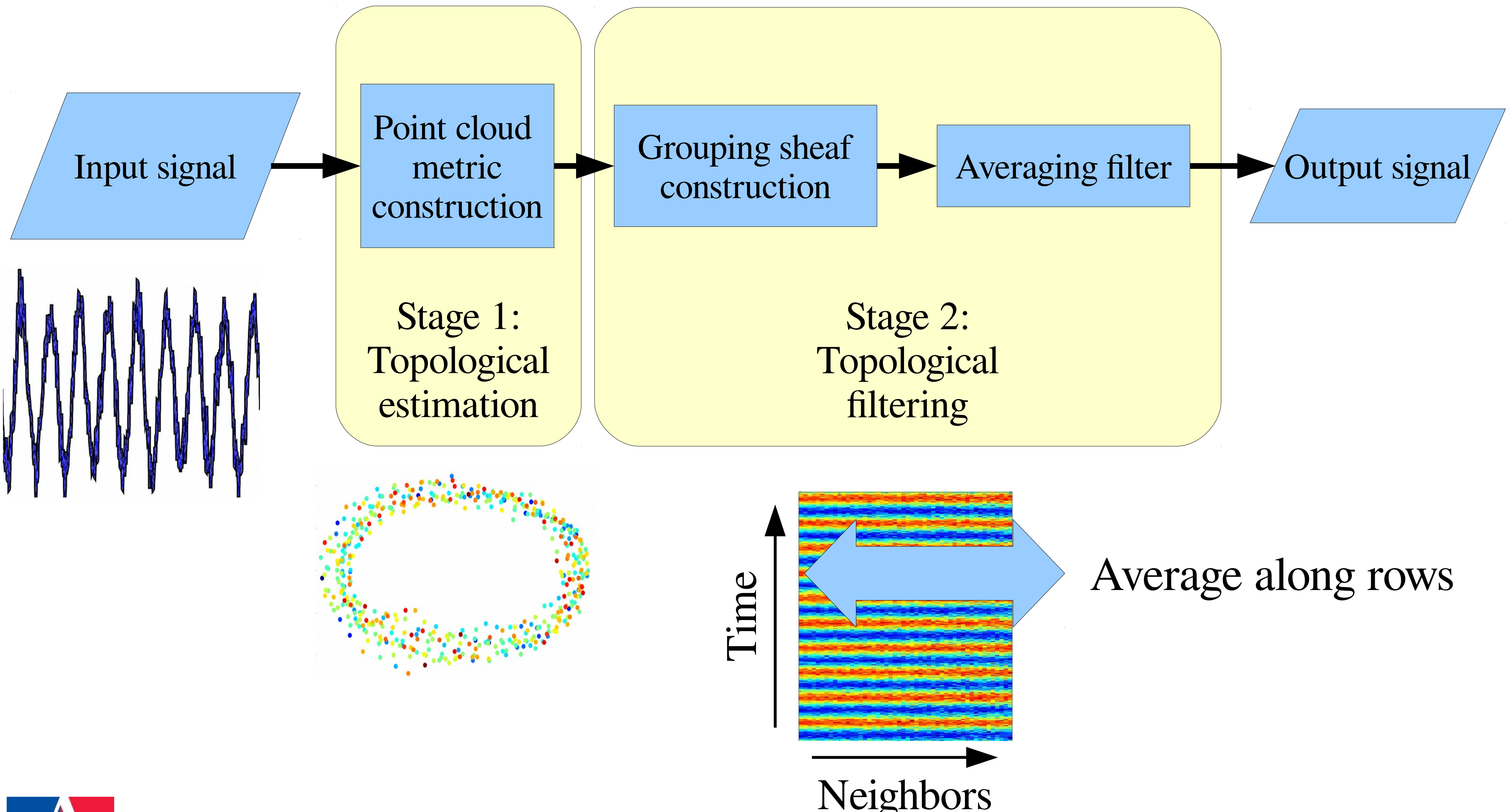


- Can **safely** do **more** averaging by collecting samples at “similar places” across the **entire** signal



Michael Robinson

Filter block diagram



Stage 1: Topological estimation

Image values

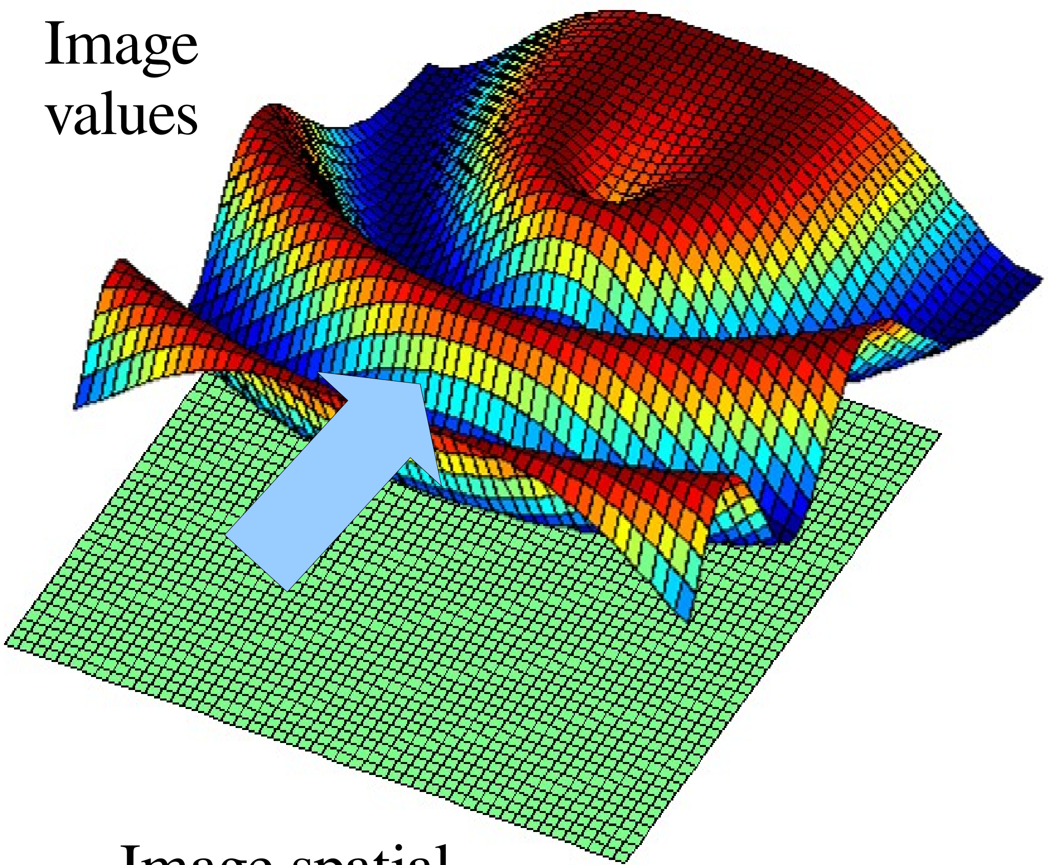
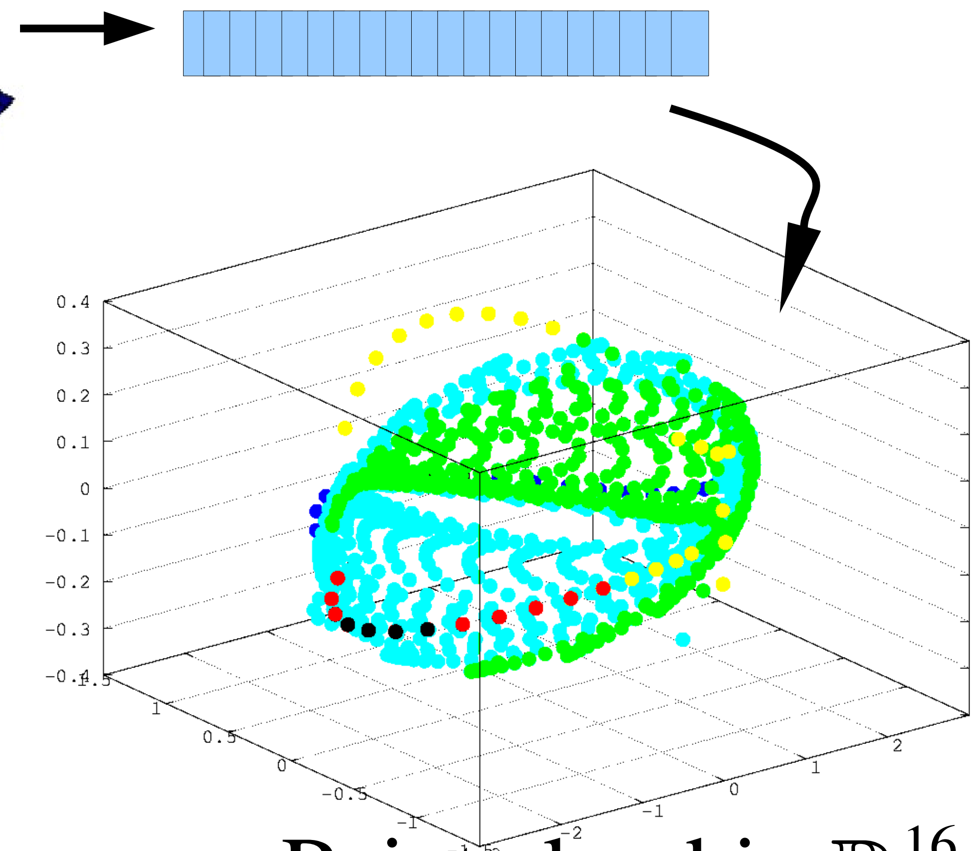


Image spatial domain

Pixel neighborhood



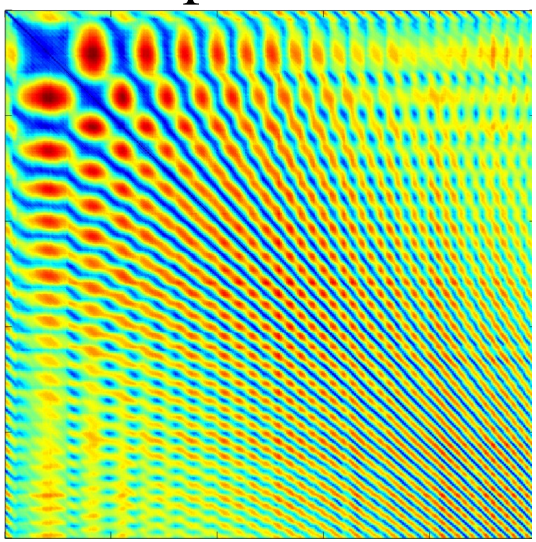
Point cloud in $\mathbb{R}^{16}$

Stage 2: Grouping sheaf

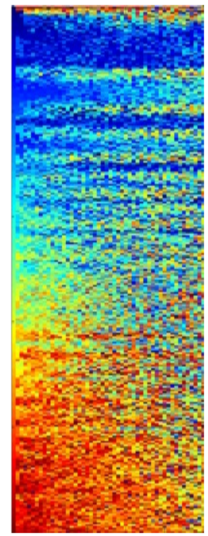
Distance matrix of point cloud

Sample number

Sample number



Sample number



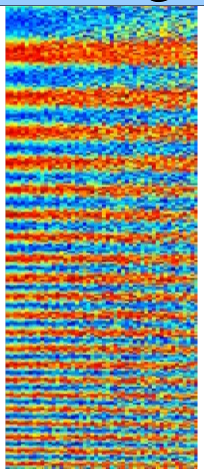
neighbors

Locations of nearest neighbors

Values of the signal at the neighbors

Average

Sample number



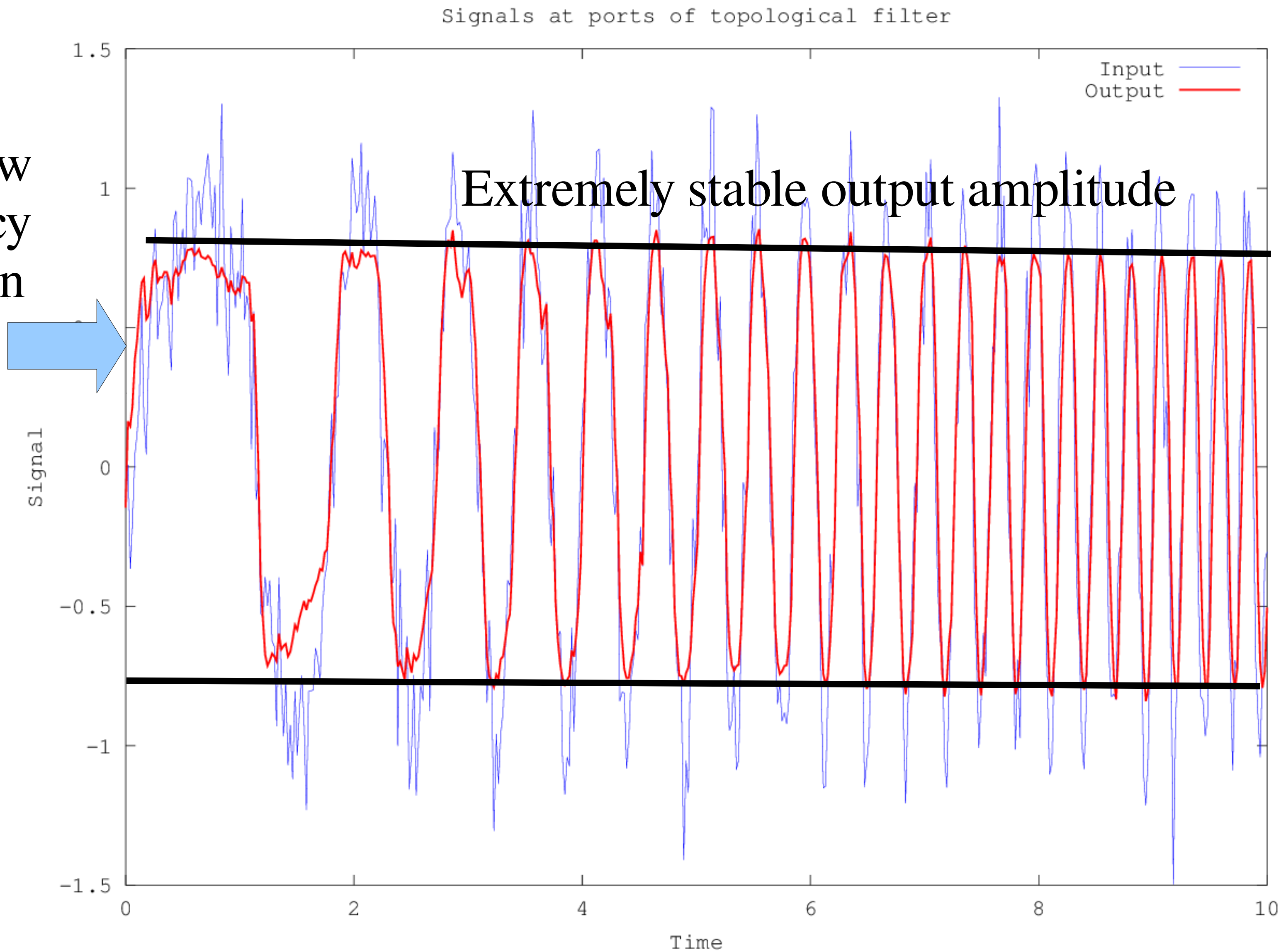
neighbors

Process execution

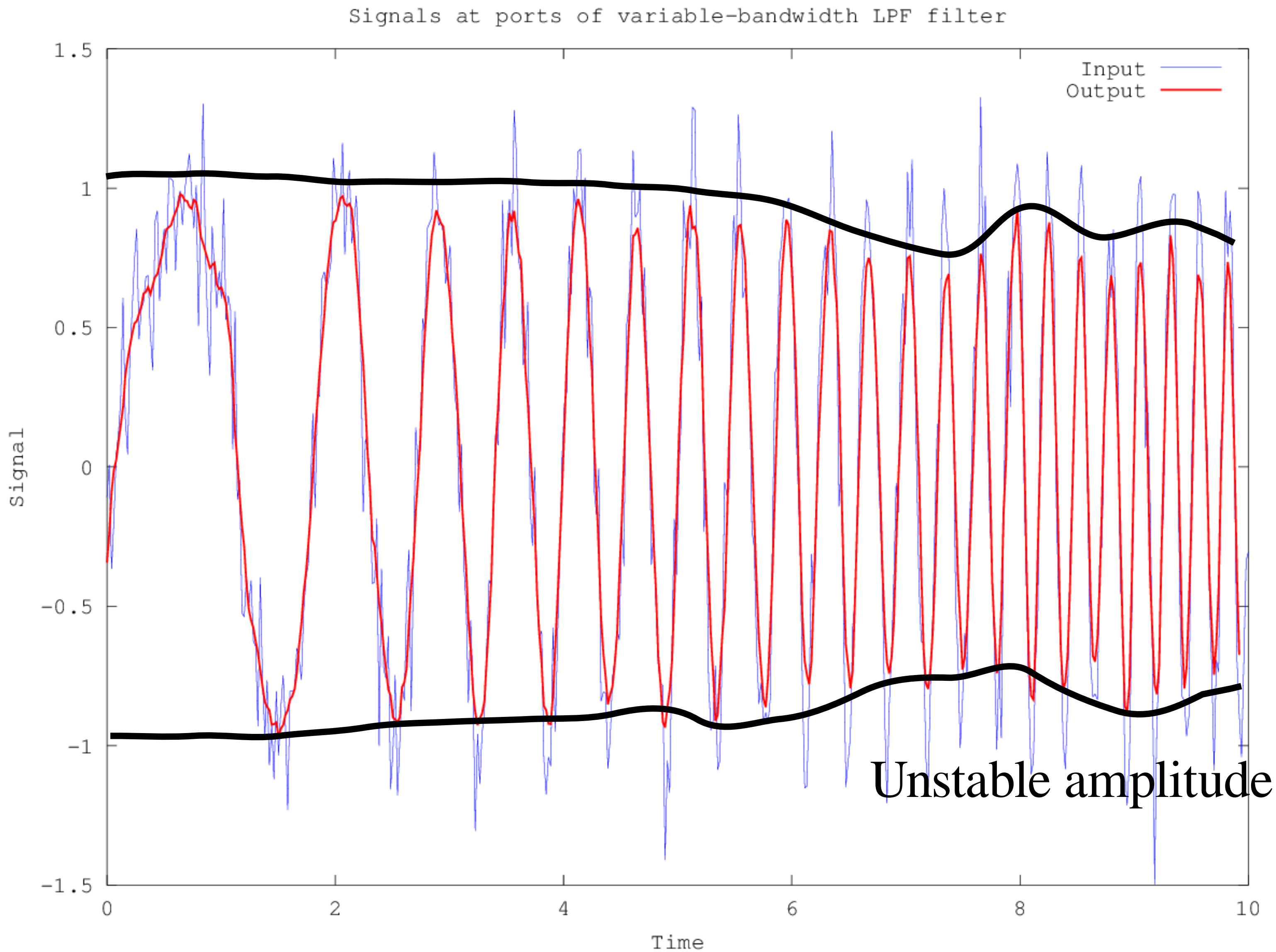
Michael Robinson

Topological filter results

Some low frequency distortion

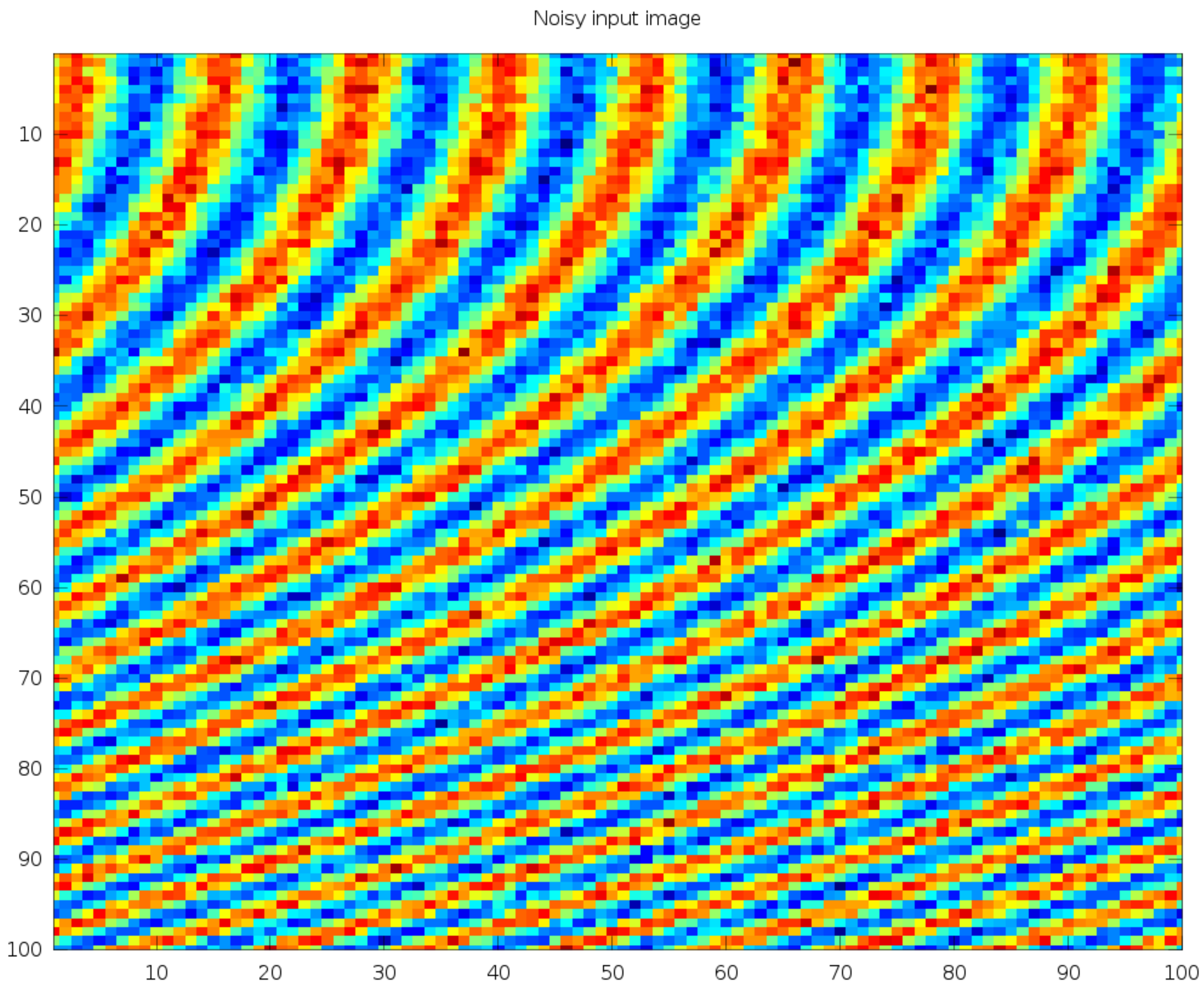


Compare: standard adaptive filter



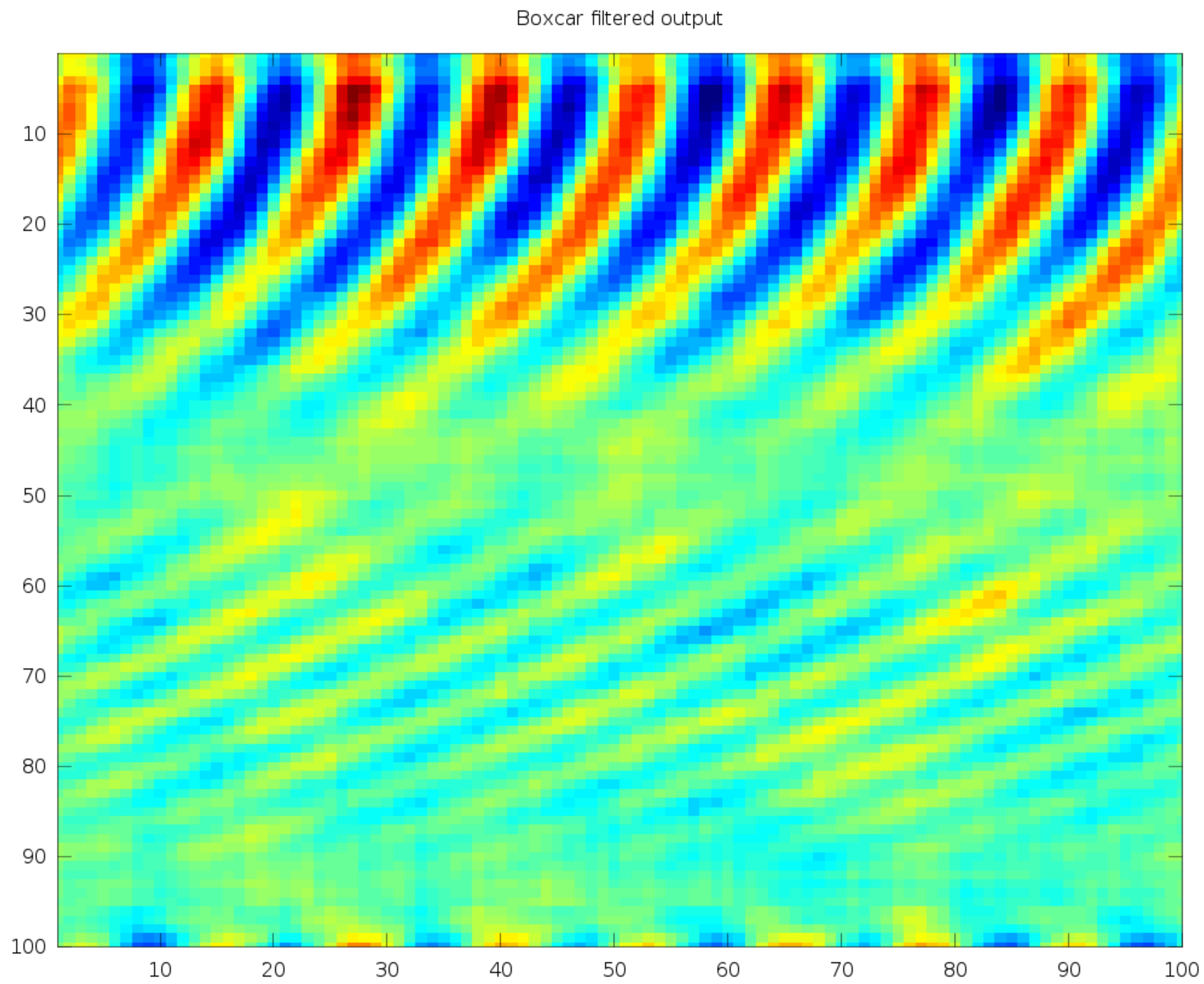
Michael Robinson

Input image

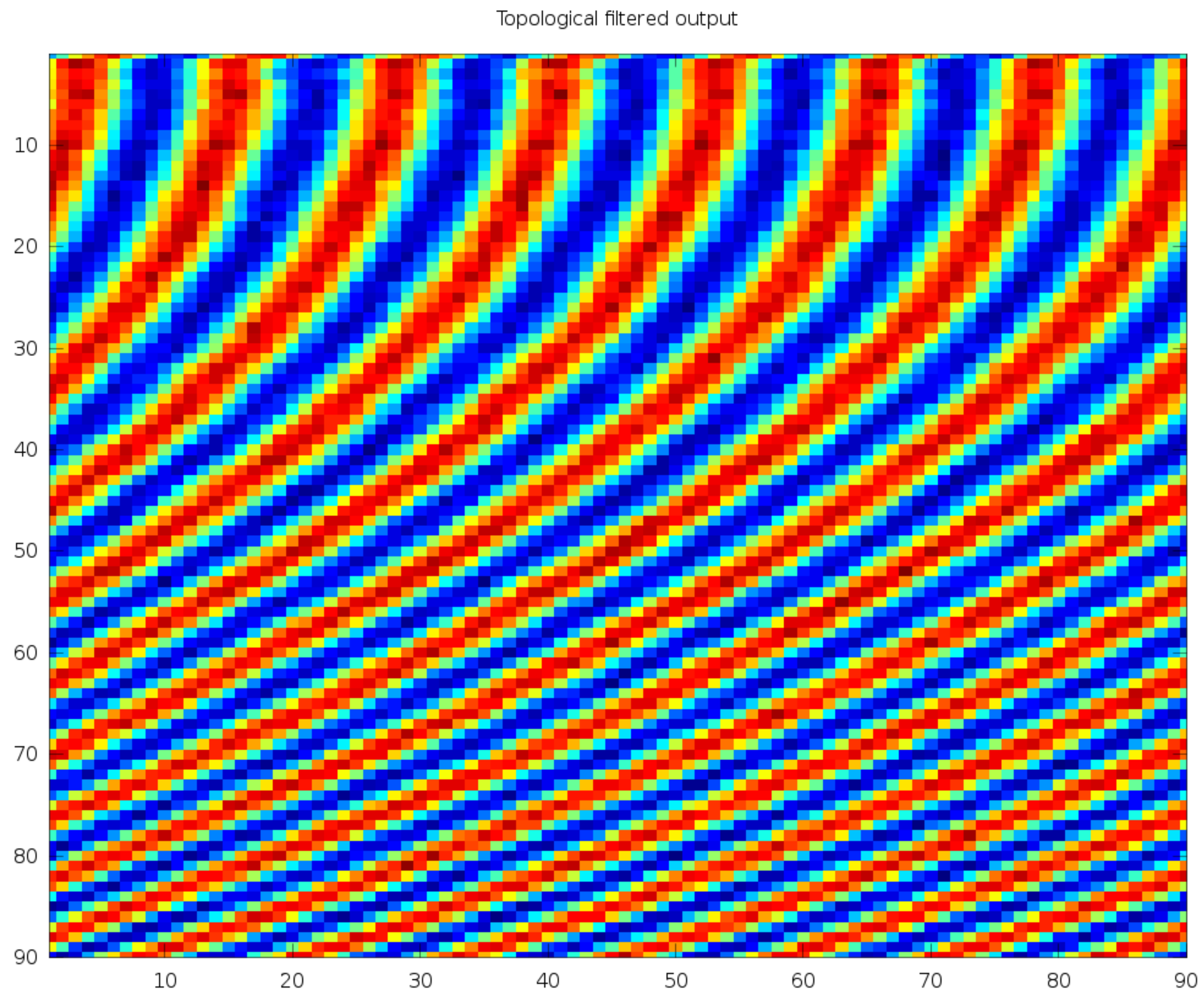


Michael Robinson

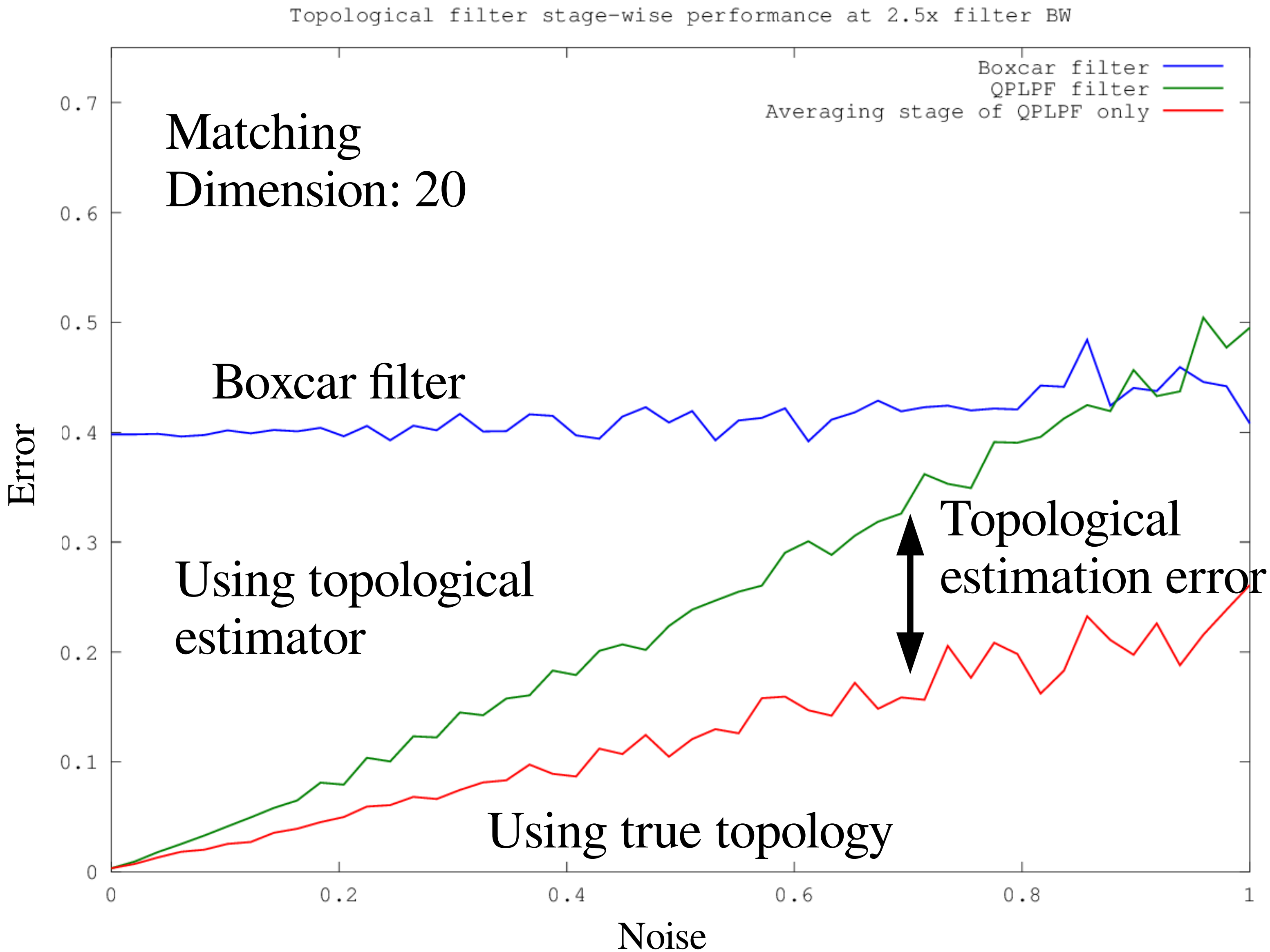
Fixed frequency image filter



Topological filter output

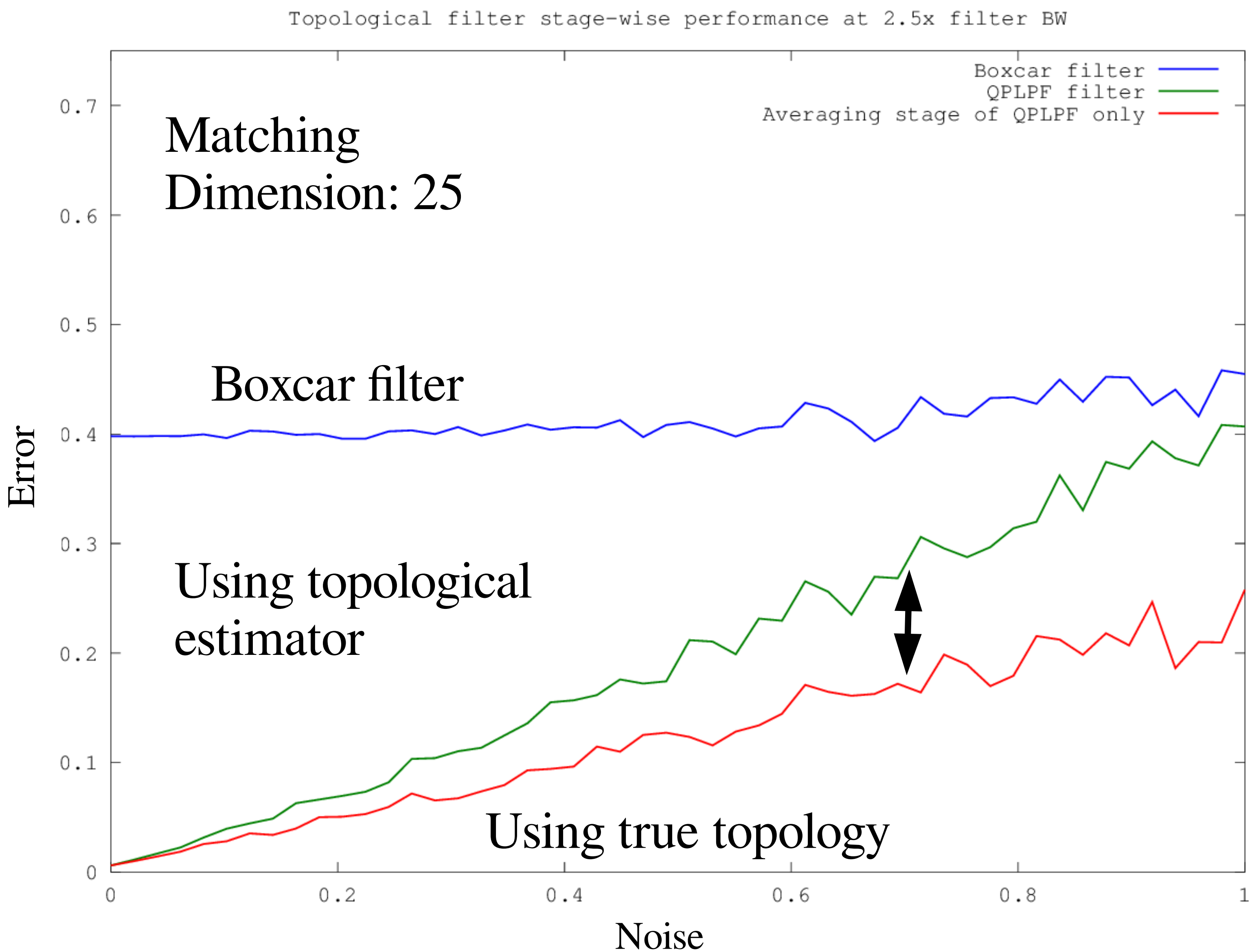


Error contributions

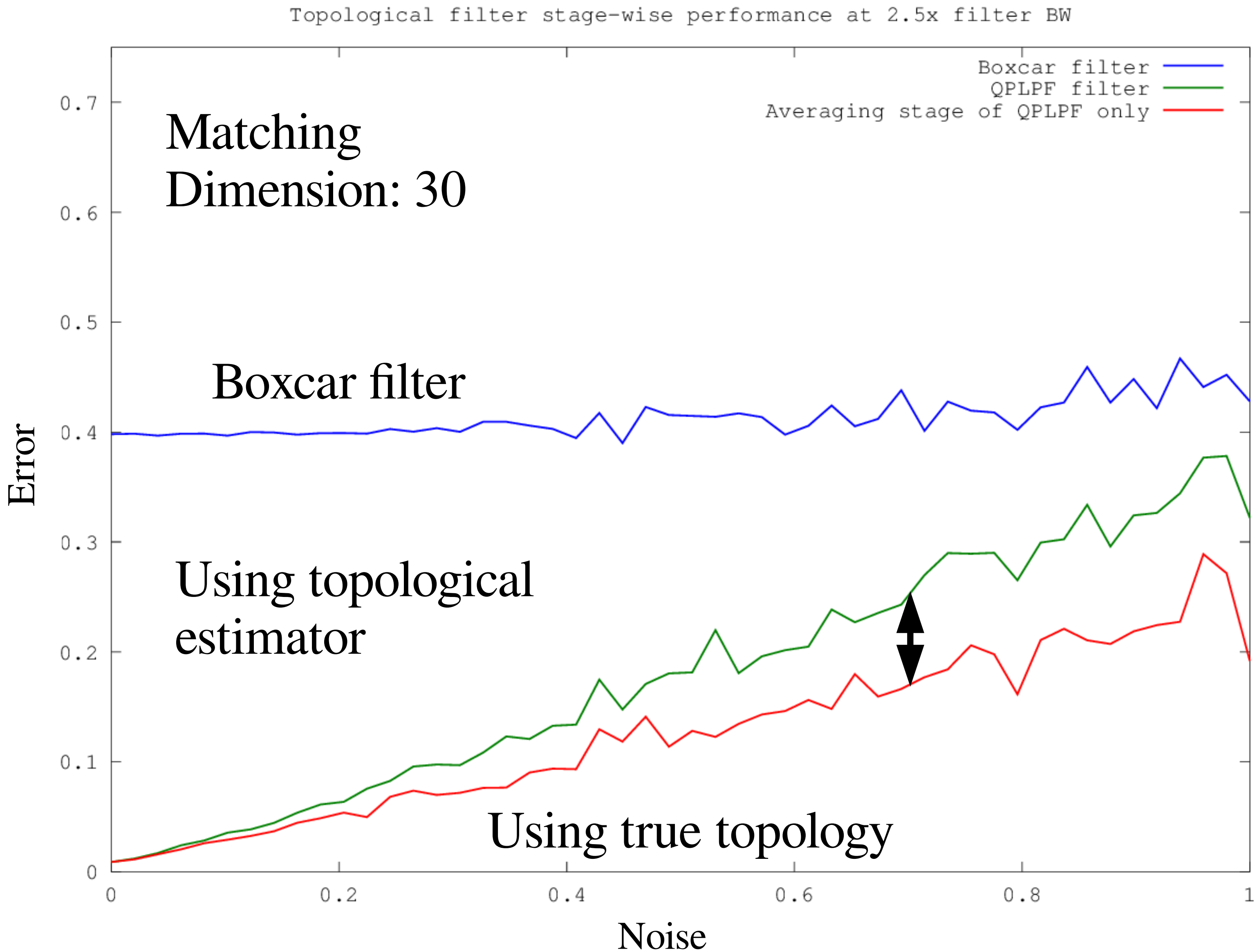


Michael Robinson

Error contributions



Error contributions



Michael Robinson

Further reading...

- Sanjeevi Krishnan, "Flow-cut dualities for sheaves on graphs," http://arxiv.org/abs/1409.6712
- Robert Ghrist and Sanjeevi Krishnan, "A Topological Max-Flow Min-Cut Theorem," *Proceedings of Global Signals. Inf.*, (2013).
- Michael Robinson, "Understanding networks and their behaviors using sheaf theory," *IEEE Global Conference on Signal and Information Processing (GlobalSIP)* 2013, Austin, Texas.
- Michael Robinson, "The Nyquist theorem for cellular sheaves," *Sampling Theory and Applications 2013*, Bremen, Germany.