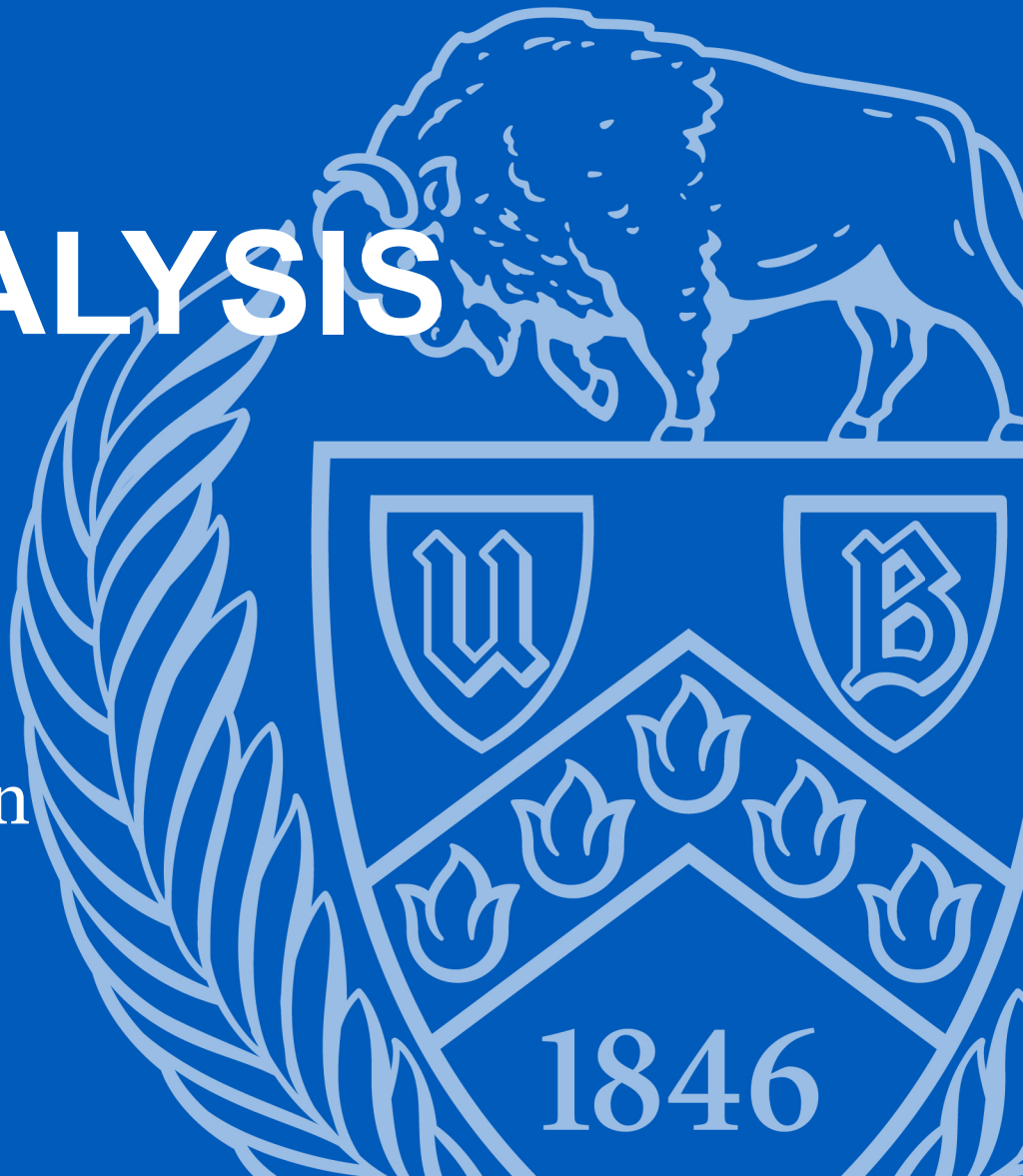
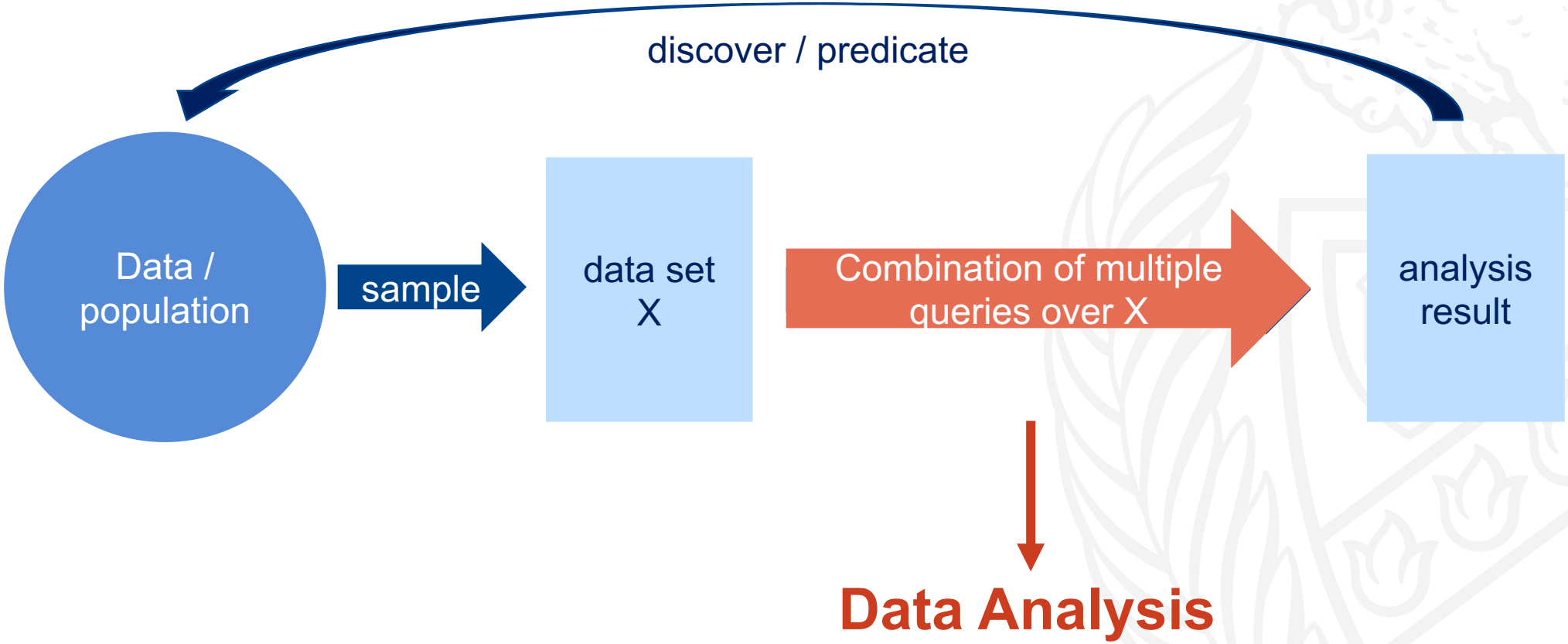
TYPE SYSTEM FOR ADAPTIVE DATA ANALYSIS

Presenter: Jiawen Liu
Joint Work : Marco Gaboardi, Weihao Qu,
Deepak Garg, Jonathan Ullman

University at Buffalo
School of Engineering and Applied Sciences

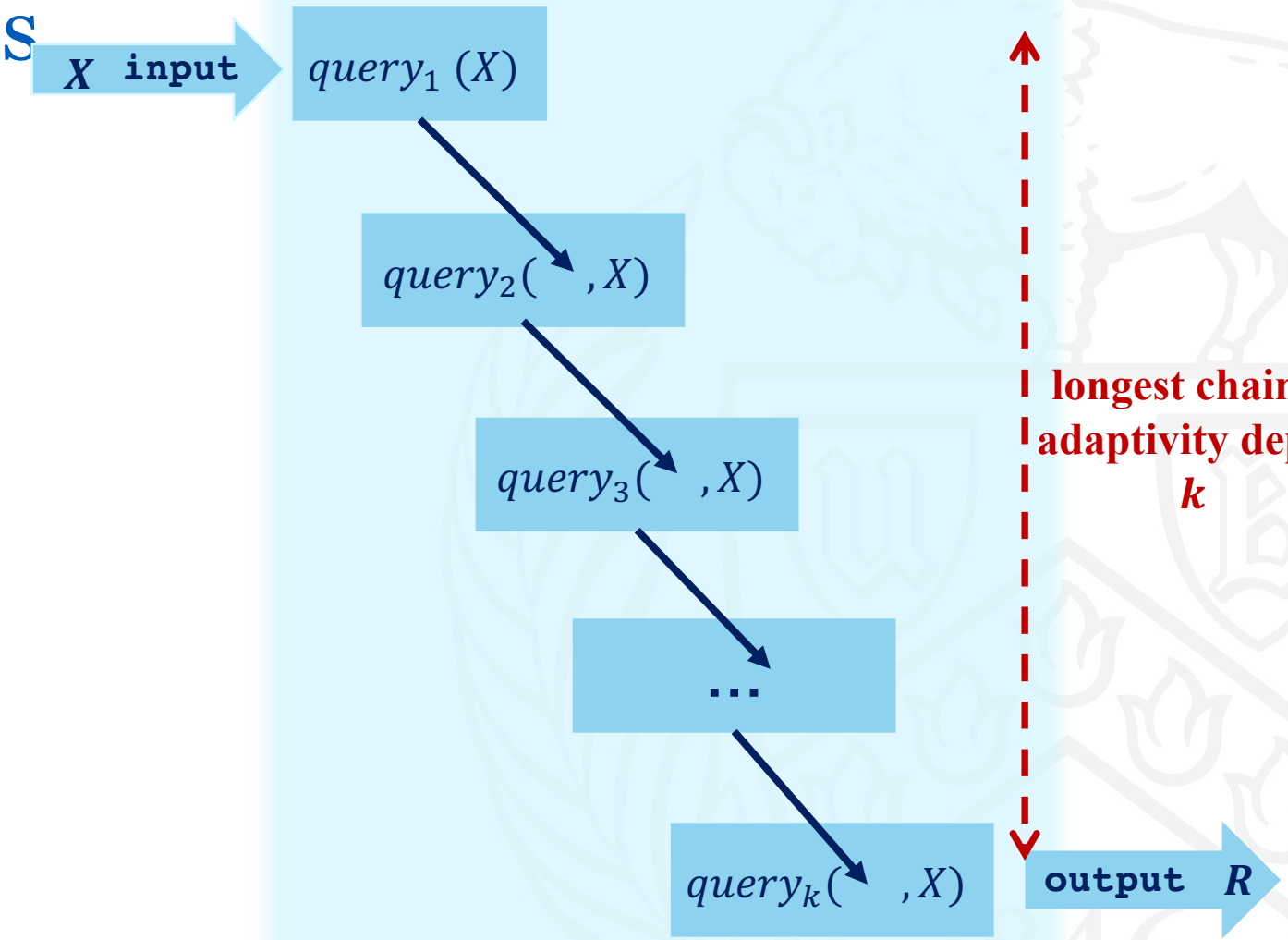


Data Analysis

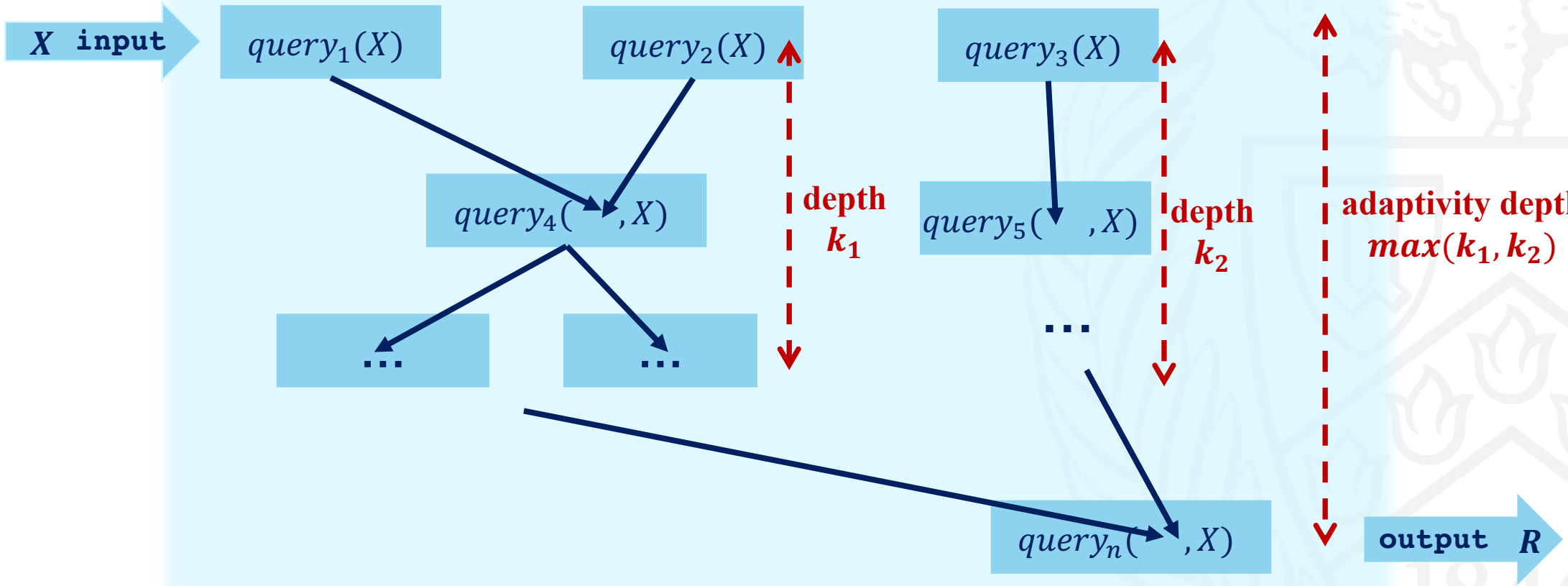


Adaptive Data Analysis

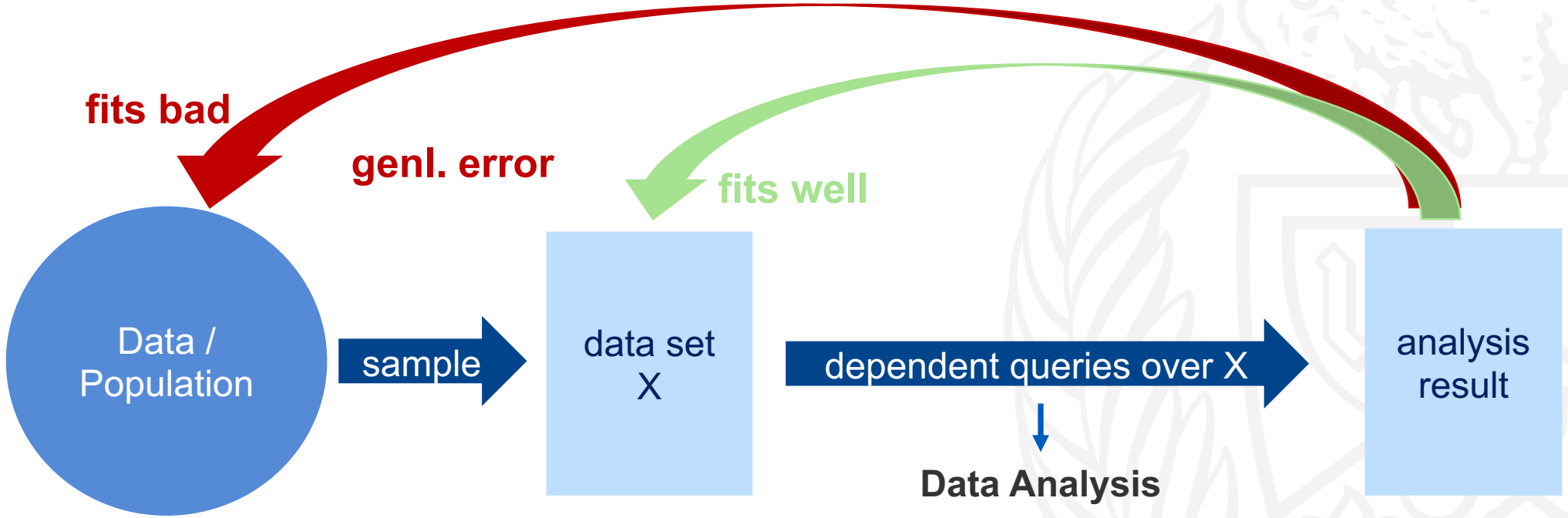
some queries rely on the results of other queries



Adaptive Data Analysis - example



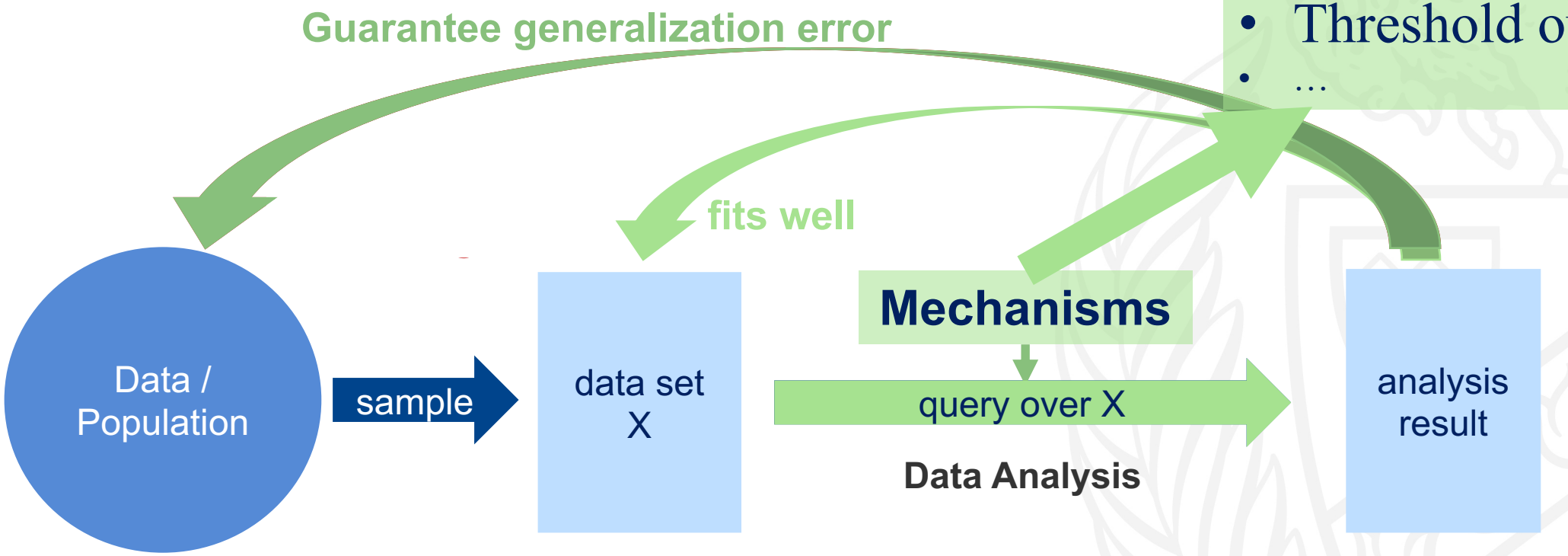
Motivation
– Generalization Error / Overfitting



[Adaptivity in analysis will propagate the overfitting]

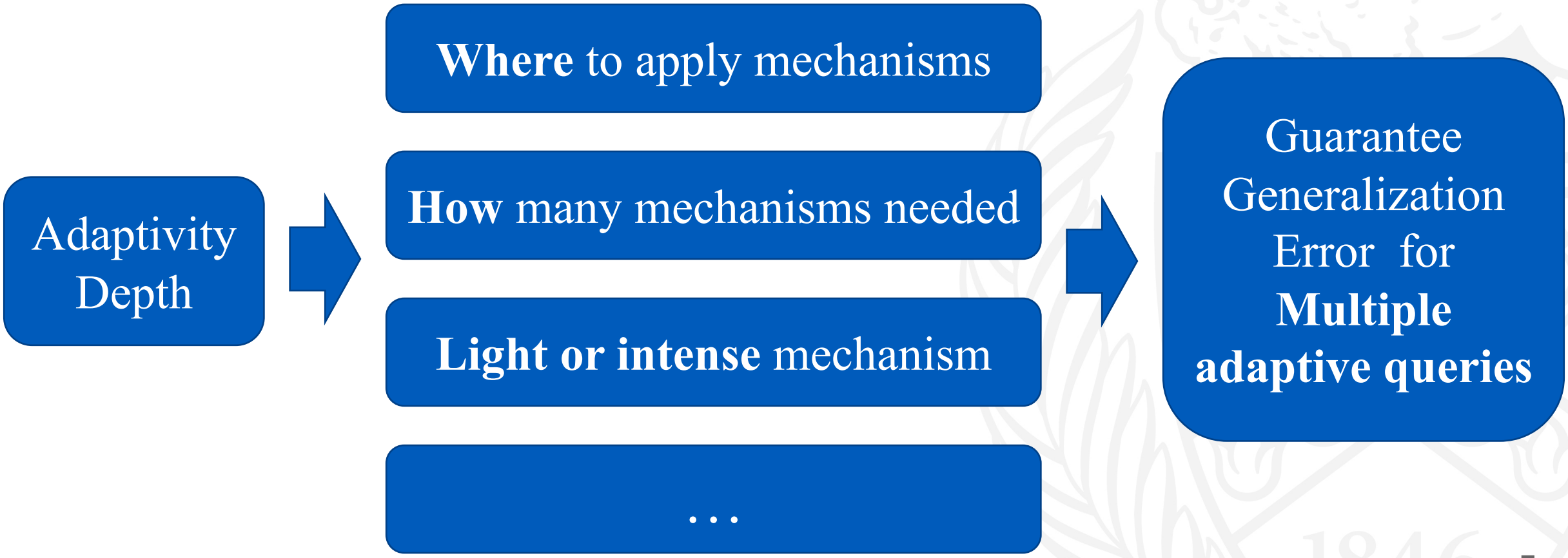
Existing Methods – 1 Query Guarantee

- Gaussian Mechanism
- Laplace Mechanism
- Threshold out
- …



[Guarantee will lose in multiple adaptive queries]

Motivation – Multiple Queries Guarantee

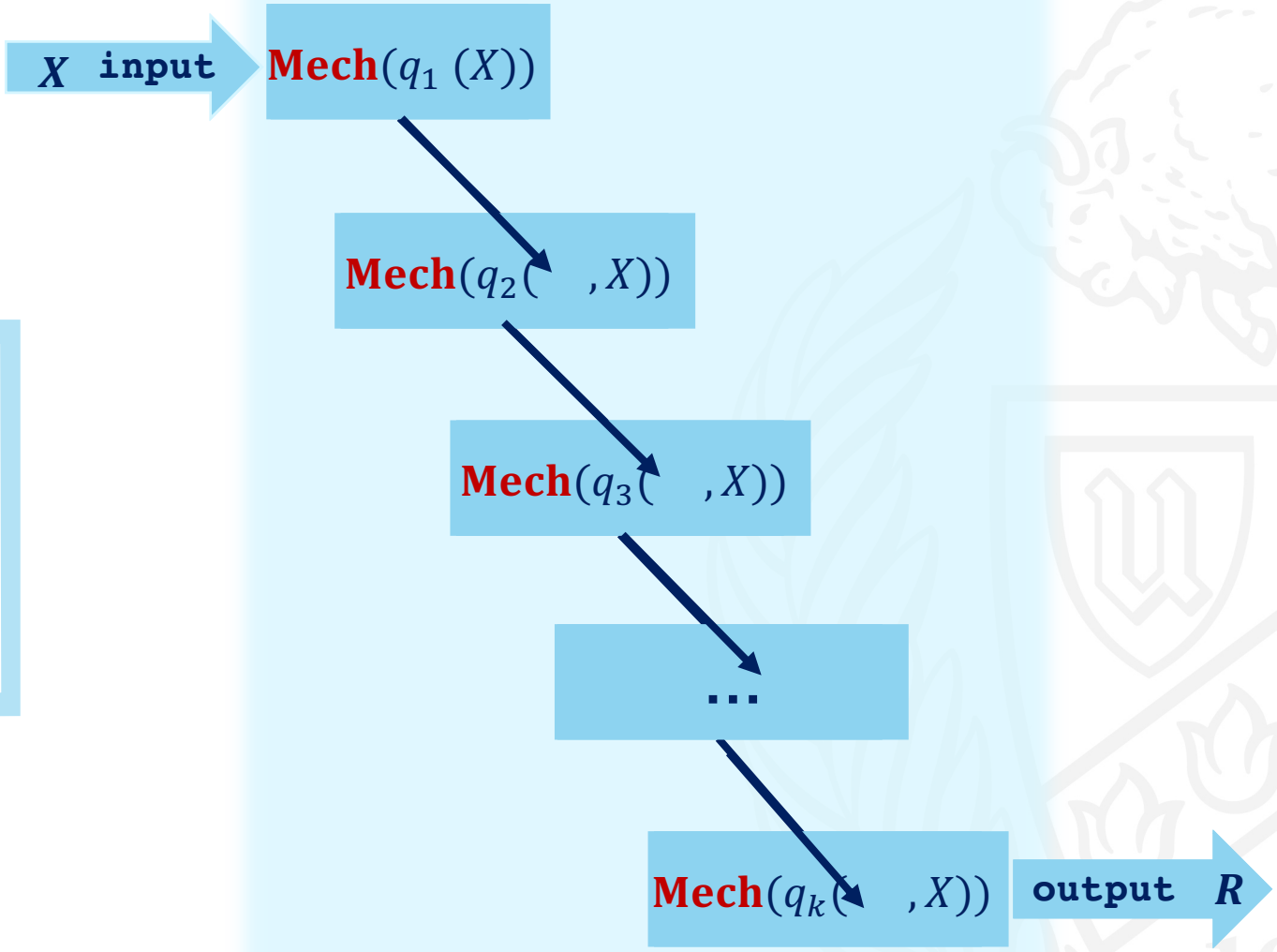


OUR WORK

analyze the *adaptivity depth*
for data analysis program.

Novelty

use mechanism to encapsulate queries and combine them.



Challenges in Language Design

adaptivity depends on the **Runtime Information**

represent the **probabilistic computing**

provide precise **Upper Bound** for adaptivity depth

Refinement Types

singleton type: **int[I]**

representing the run time information

index term: indicate the value of an integer

domain: $\mathbb{N}$

Expressions

$\delta(q)$

represent the mechanism δ applied over a query q

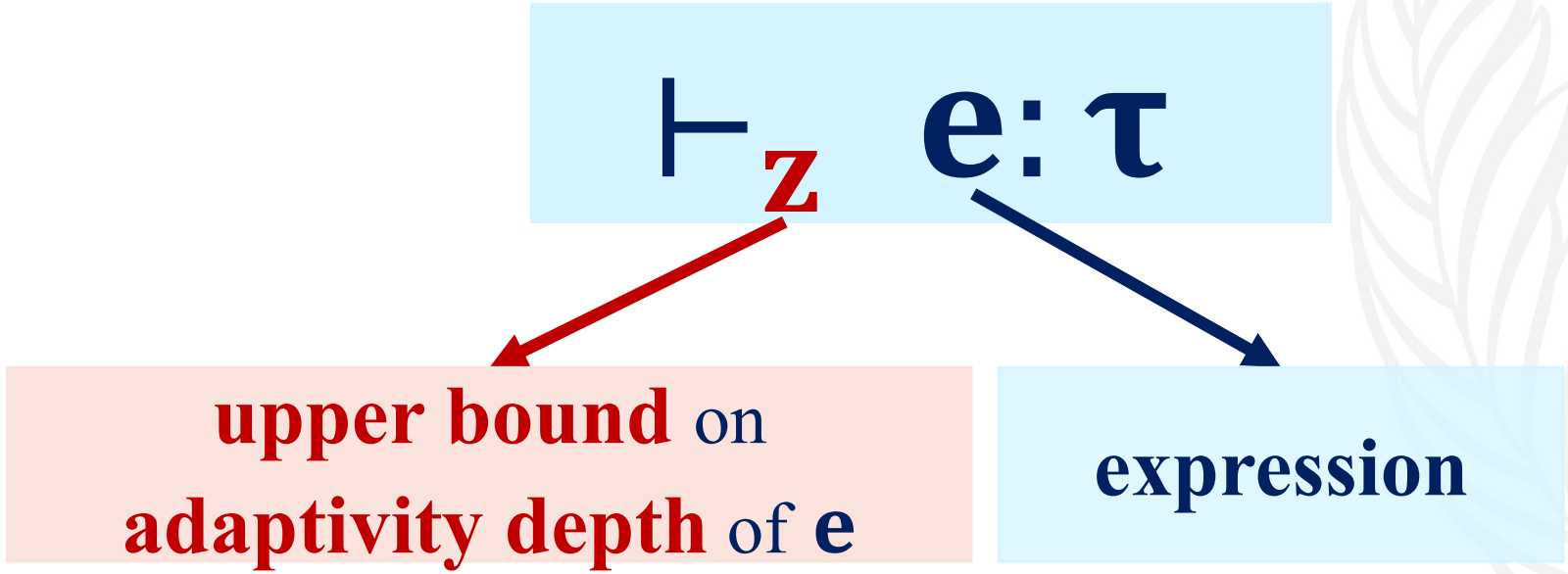
uniform v_1 v_2

parameterized const represents values sampled from distributions

the range $[v_1, v_2]$, where sample from

Typing Judgement

Annotated typing judgement:
approximates an **upper bound** on **adaptivity depth** of expressions



Typing Judgement

Important rule for calculating the adaptivity:

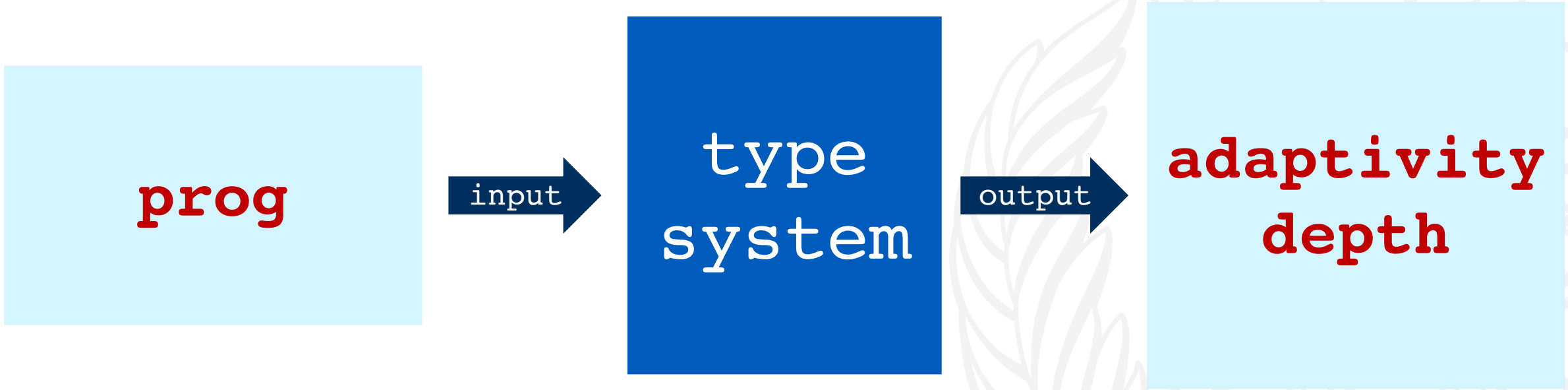
$$\frac{\vdash_{z} q : \tau}{\vdash_{z+1} \delta(q) : \tau}$$

Soundness

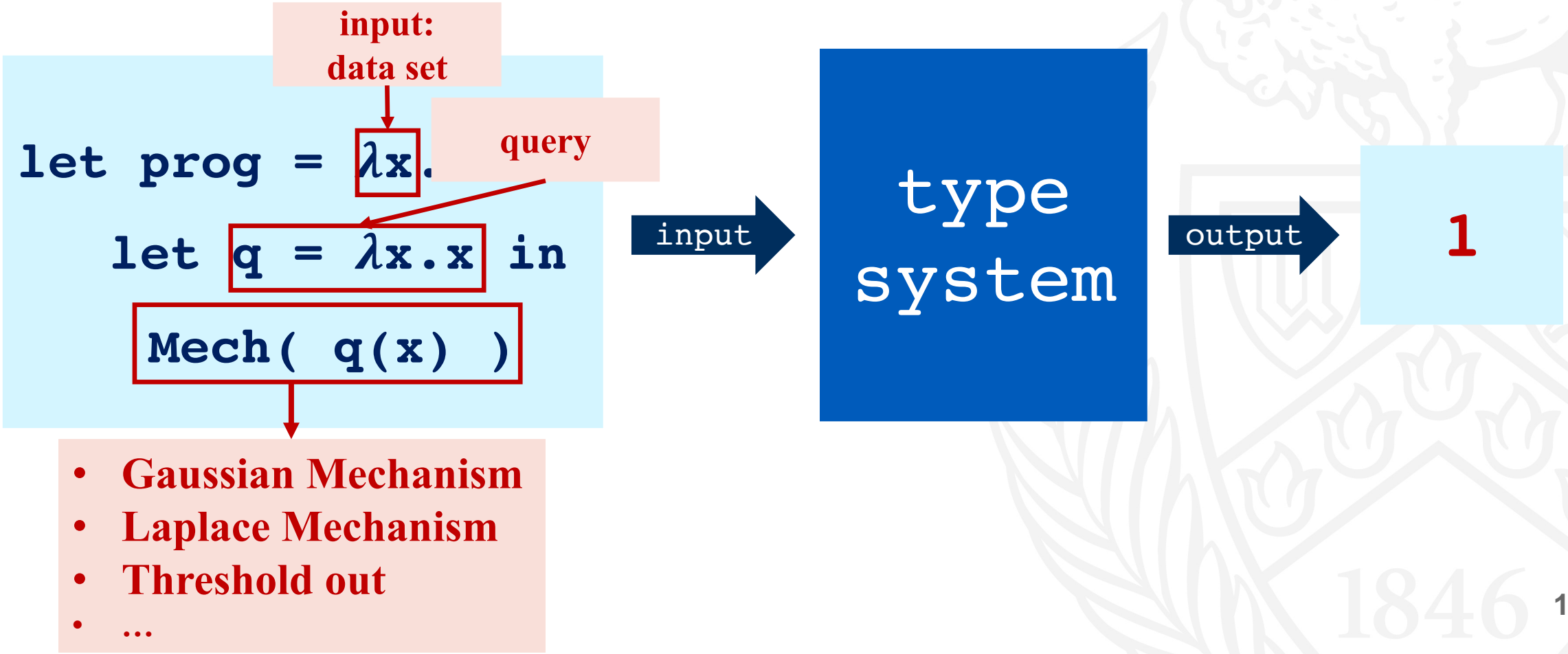
step-indexed logical relations → $[\![\tau]\!]_v$ $[\![\tau]\!]_e^Z$

fundamental theorem → if $\vdash_z e : \tau$ then $e \in [\![\tau]\!]_e^Z$

System Overview



System Overview – Simple Example



THANKS