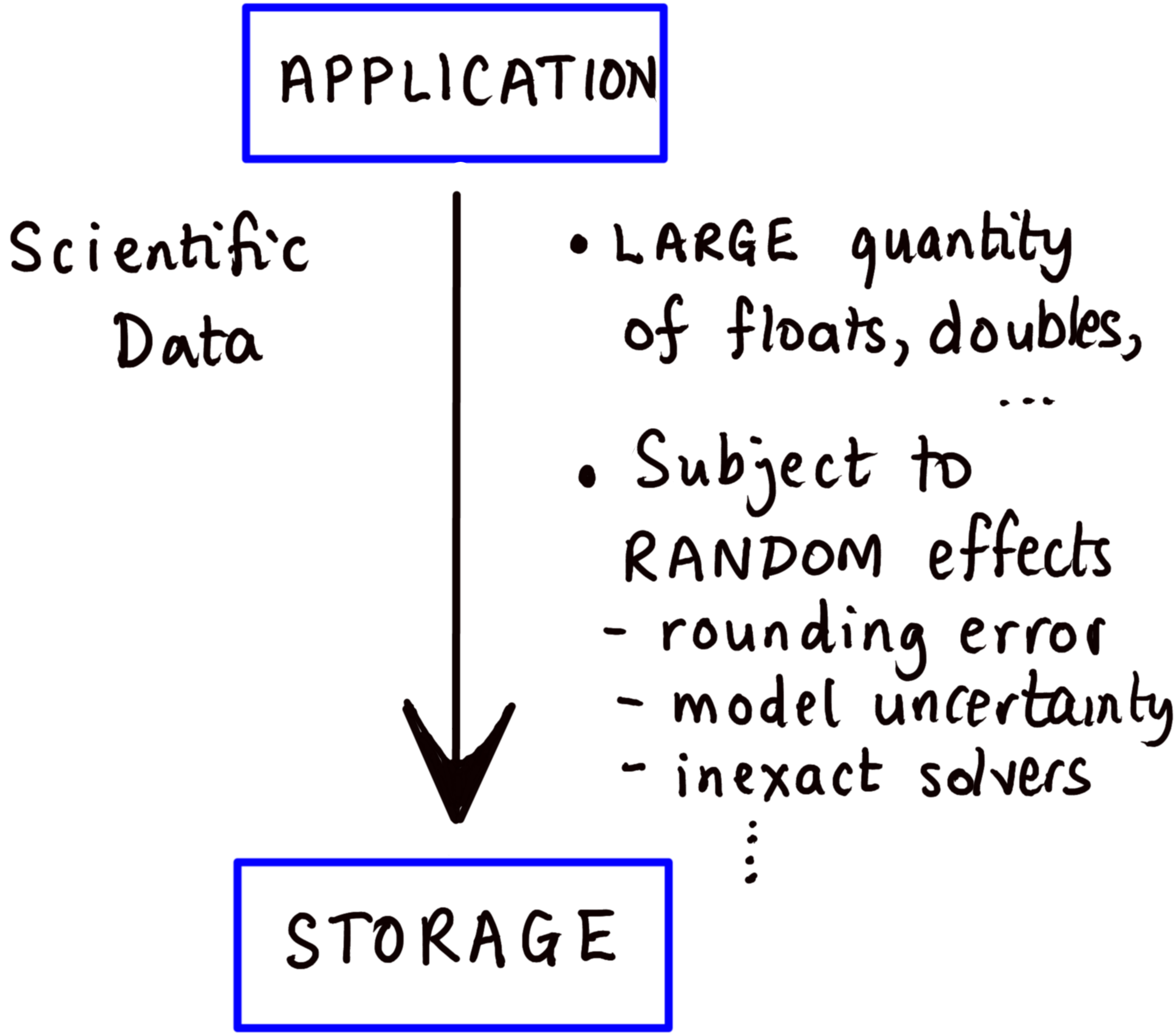
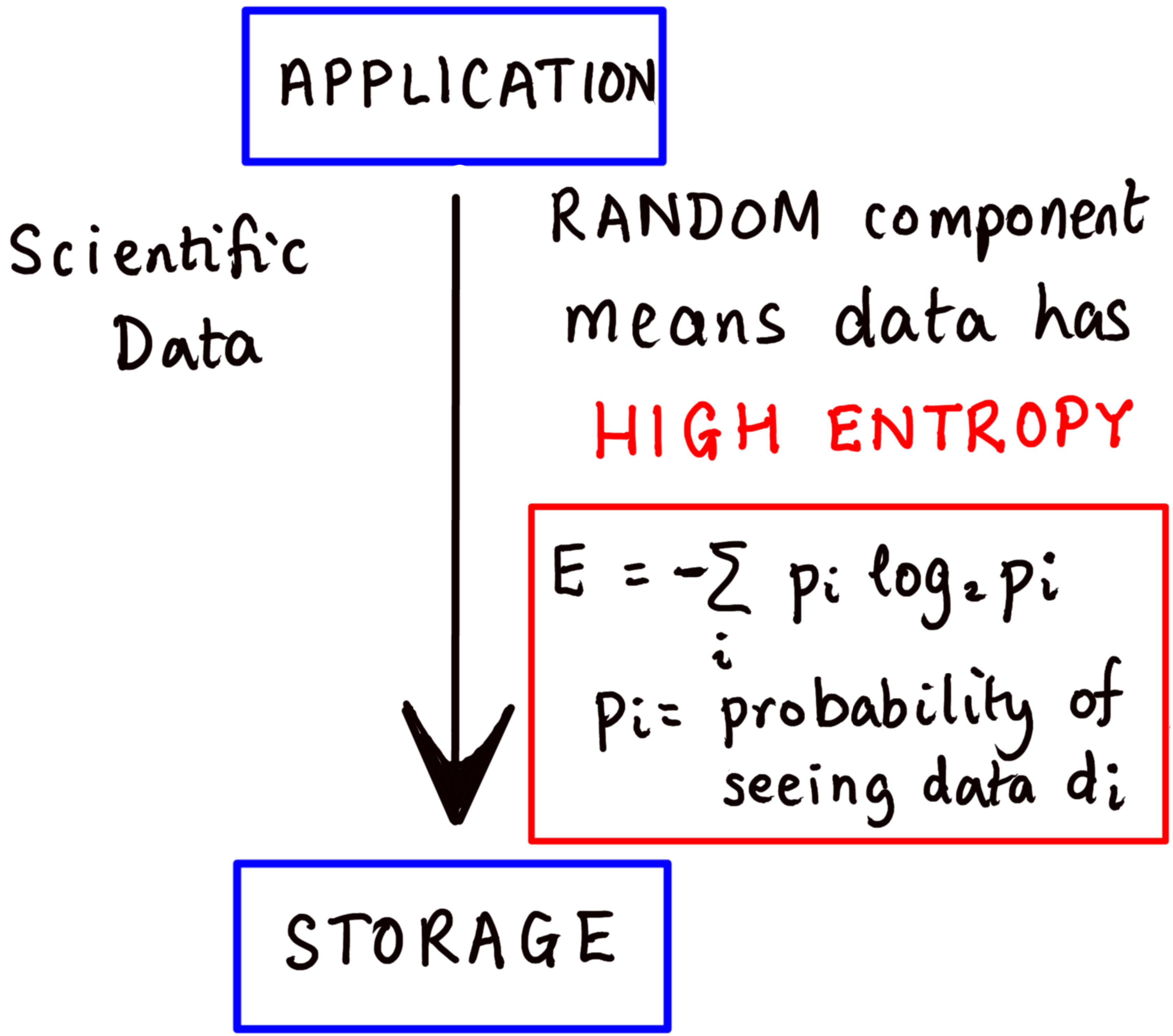
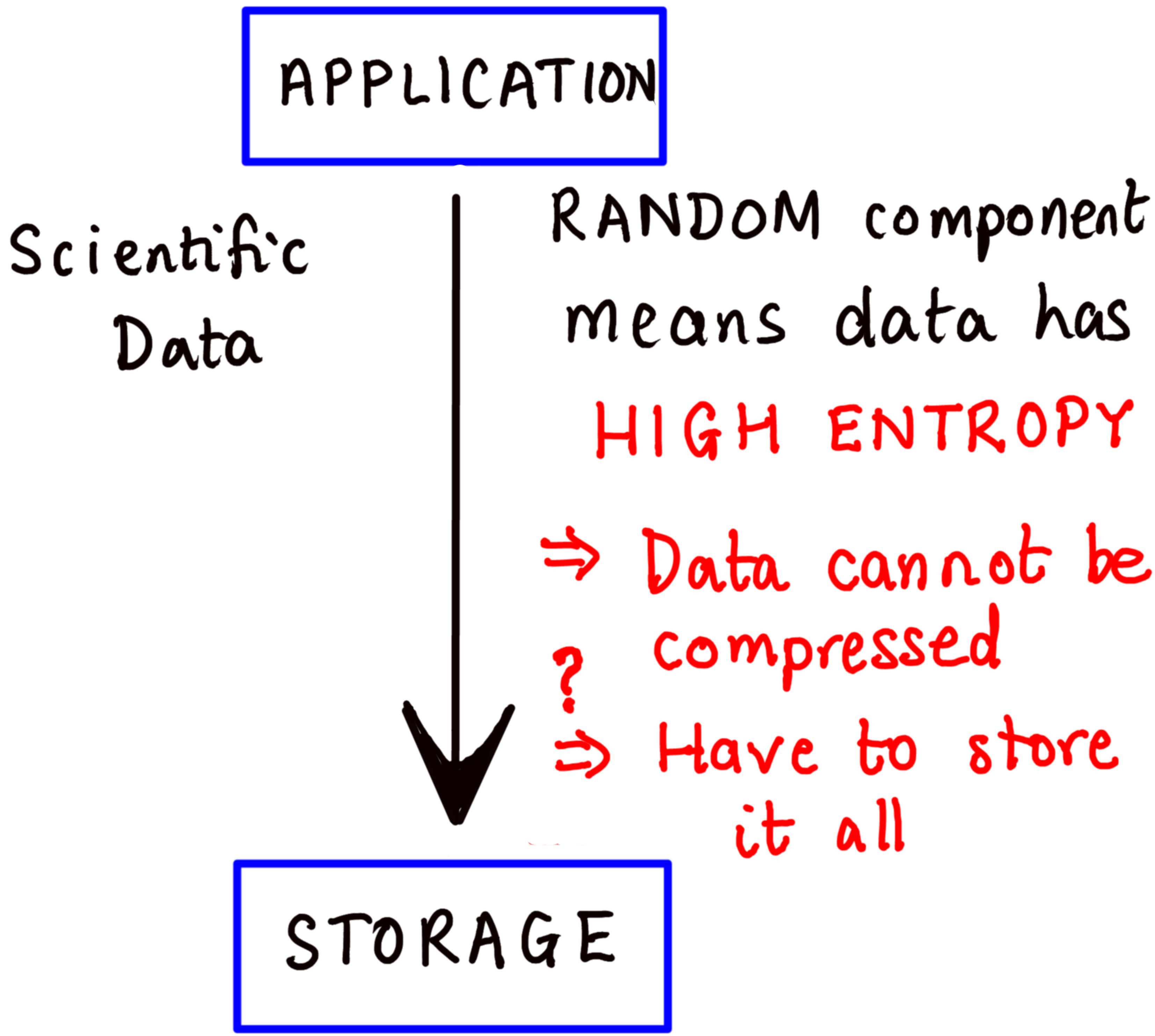
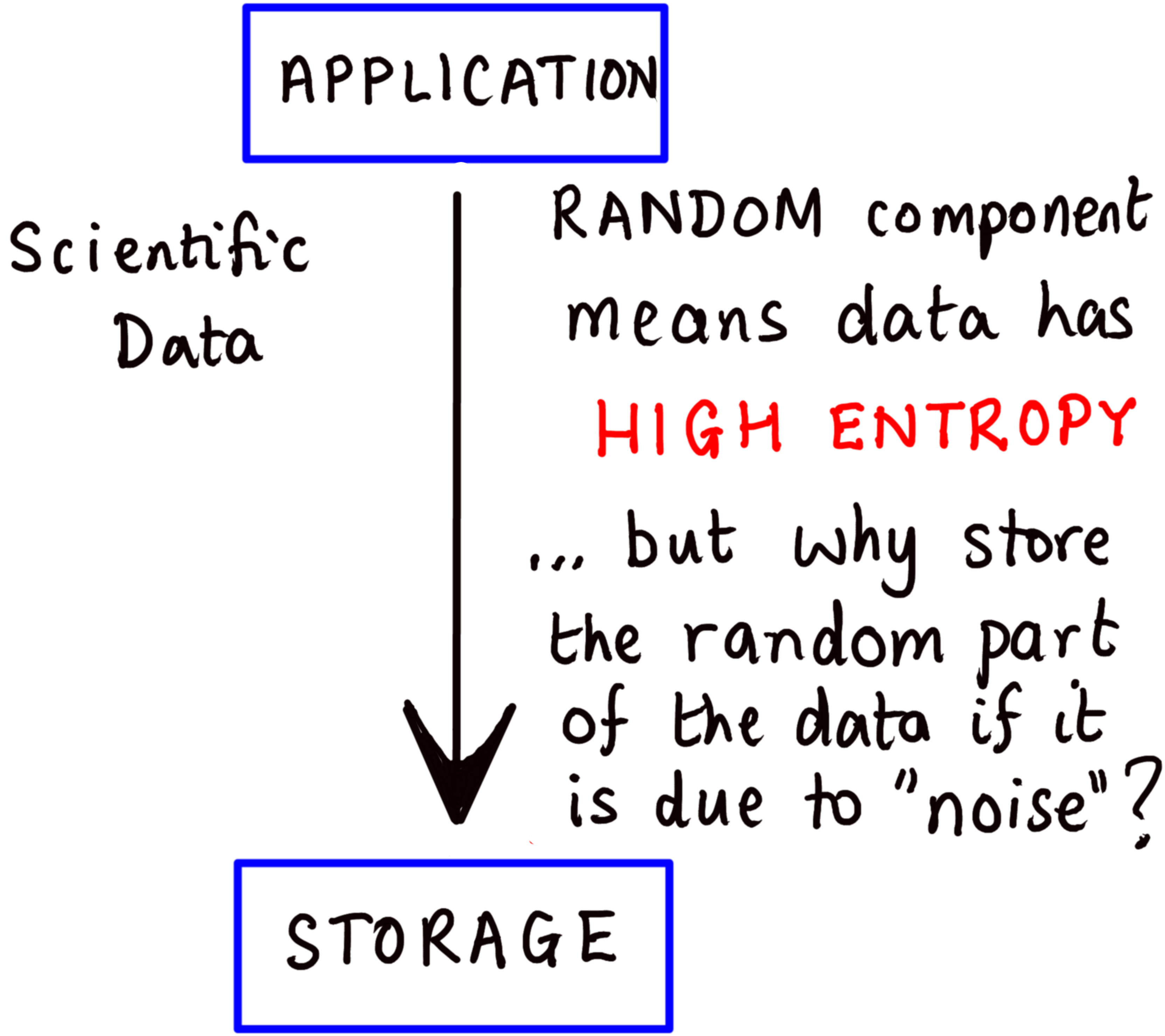
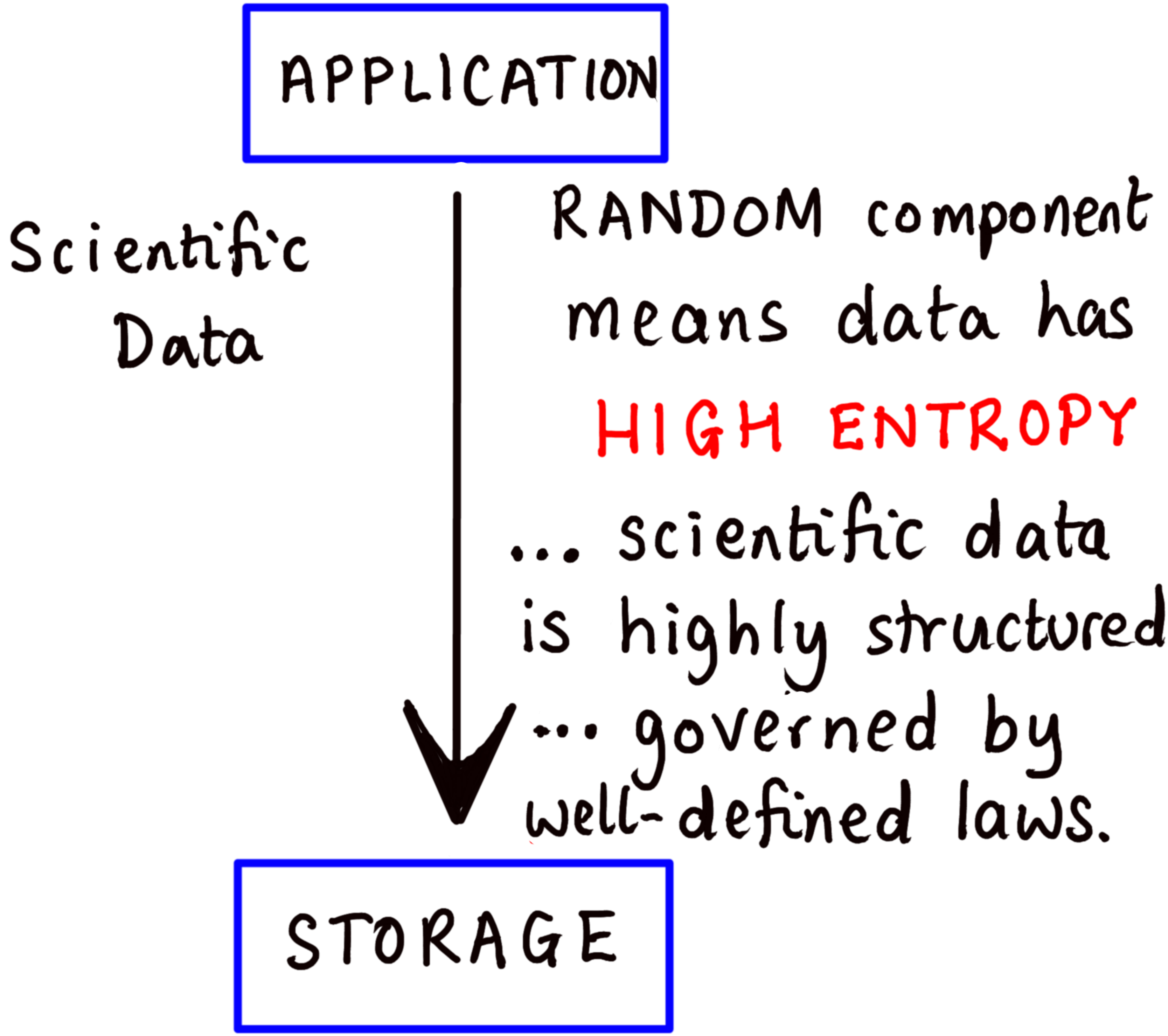
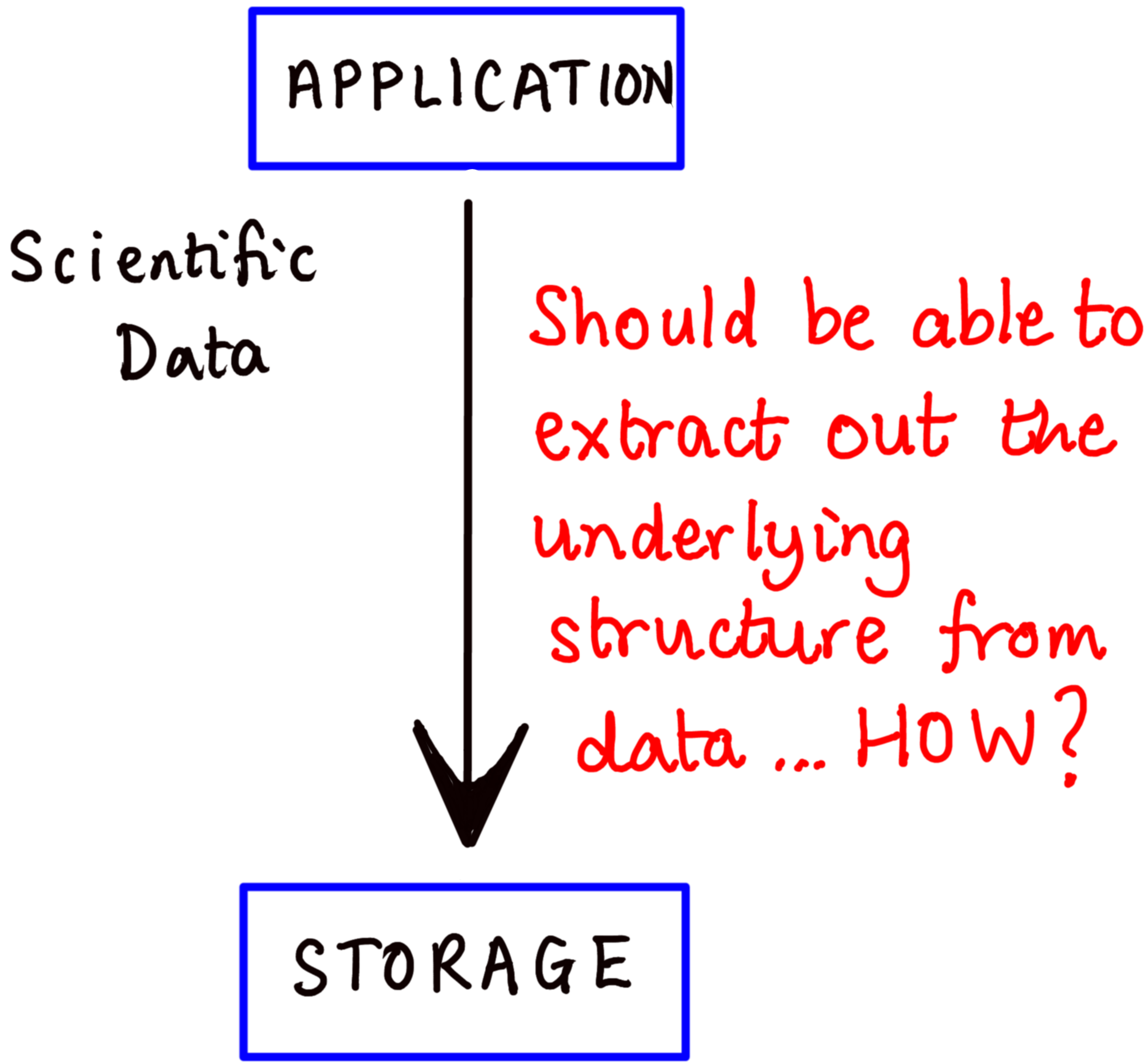

APPLICATION
Scientific Data
• LARGE quantity of floats, doubles, ...
• Subject to RANDOM effects
- rounding error
- model uncertainty
- inexact solvers
⋮
STORAGE

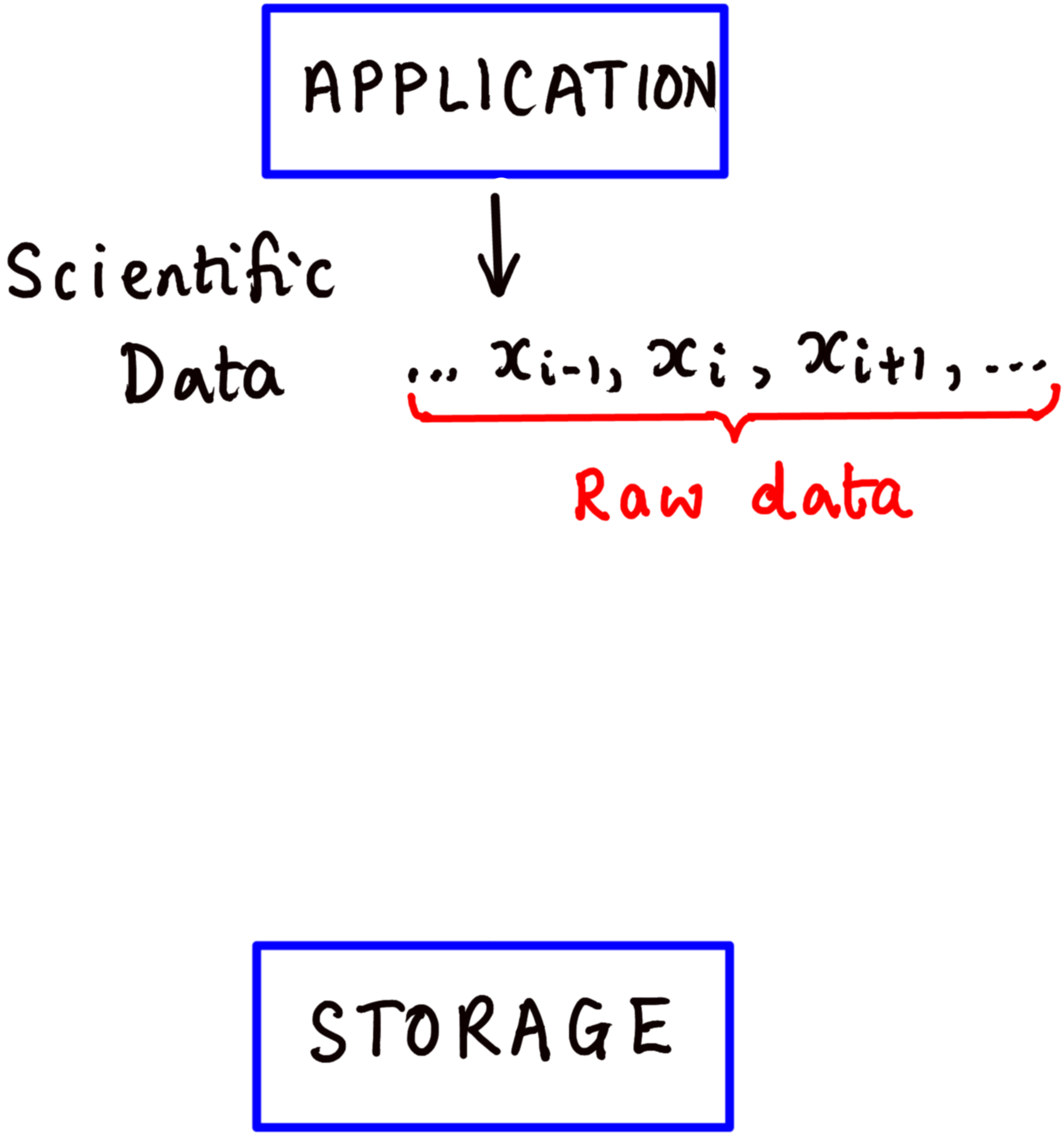

APPLICATION
Scientific Data
RANDOM component means data has
HIGH ENTROPY
$E = -\sum_i p_i \log_2 p_i$
p_i = probability of seeing data d_i
STORAGE


APPLICATION
Scientific Data
RANDOM component means data has
HIGH ENTROPY
⇒ Data cannot be compressed
?
⇒ Have to store it all
STORAGE


APPLICATION
Scientific Data
RANDOM component means data has HIGH ENTROPY
... but why store the random part of the data if it is due to "noise"?
STORAGE


APPLICATION
Scientific Data
RANDOM component means data has HIGH ENTROPY
... scientific data is highly structured
... governed by well-defined laws.
STORAGE


APPLICATION
Scientific Data
Should be able to extract out the underlying structure from data ... HOW?
STORAGE



APPLICATION
Scientific Data
..., x_{i-1}, x_i, x_{i+1}, ...
Raw data
STORAGE

$\boxed{\text{APPLICATION}}$

$\downarrow$

Scientific Data $\underbrace{\ldots,\ x_{i-1},\ x_i,\ x_{i+1},\ \ldots}_{\text{Raw data}}$

Knowing x_i means we can often guess/predict x_{i+1} quite well...

$\boxed{\text{STORAGE}}$

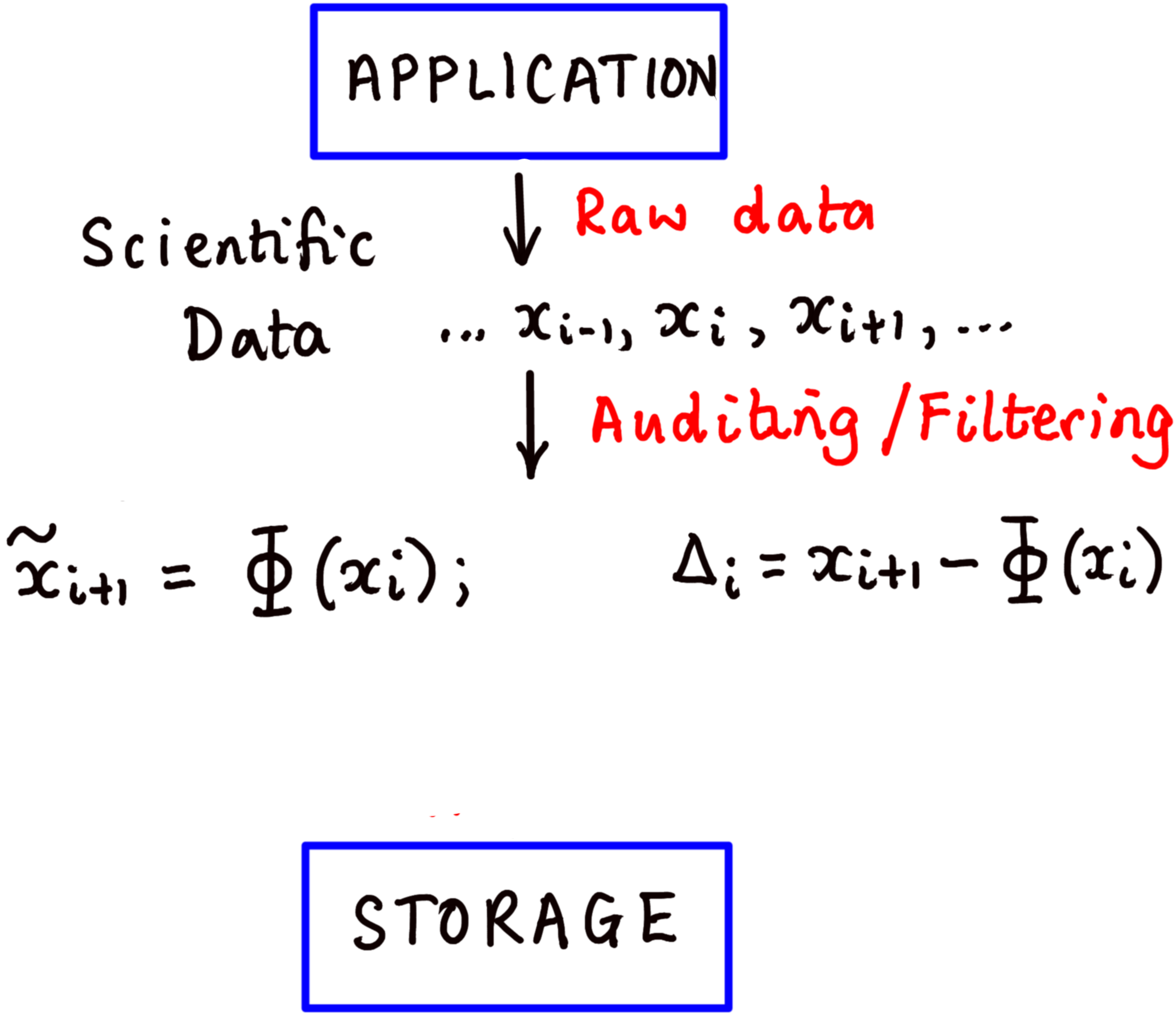
APPLICATION

$\downarrow$

Scientific Data $\quad \underbrace{\ldots, x_{i-1}, x_i, x_{i+1}, \ldots}_{\text{Raw data}}$

$$\text{i.e.} \quad x_{i+1} = \underbrace{\Phi(x_i)}_{\text{Guess / Predictor / Auditor}} + \overbrace{\text{"noise"}}^{\text{small}}$$

STORAGE



APPLICATION
Raw data
Scientific Data
$\ldots, x_{i-1}, x_i, x_{i+1}, \ldots$
Auditing / Filtering
$\tilde{x}_{i+1} = \Phi(x_i);$
$\Delta_i = x_{i+1} - \Phi(x_i)$
STORAGE

APPLICATION

$\downarrow$ Raw data

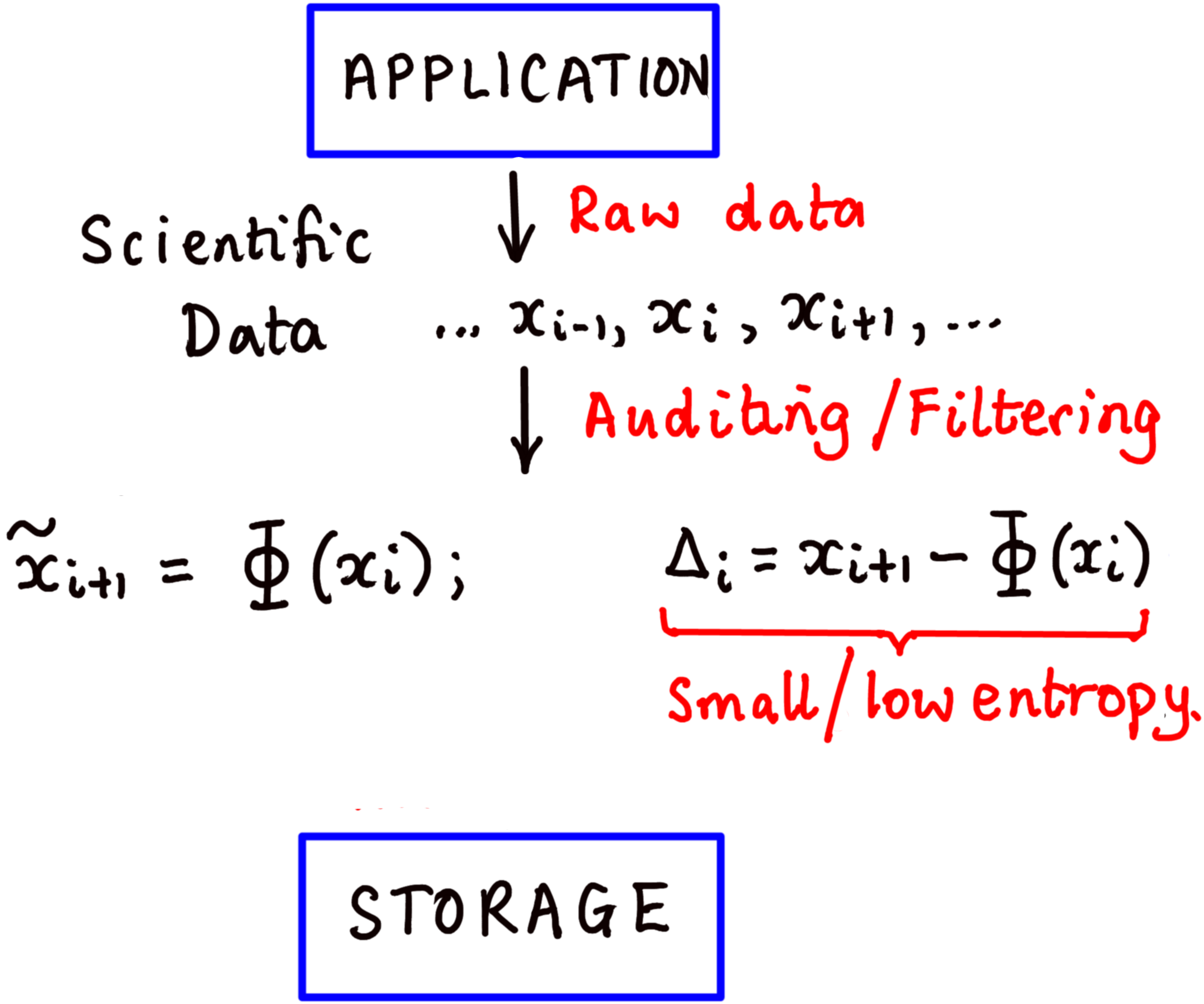
Scientific Data

$\ldots, x_{i-1}, x_i, x_{i+1}, \ldots$

$\downarrow$ Auditing / Filtering

$$\underbrace{\tilde{x}_{i+1} = \Phi(x_i);}_{\text{Store the RULE } \Phi \text{ instead of DATA}} \qquad \Delta_i = x_{i+1} - \Phi(x_i)$$

STORAGE


APPLICATION
Raw data
Scientific Data
$\ldots, x_{i-1}, x_i, x_{i+1}, \ldots$
Auditing / Filtering
$\tilde{x}_{i+1} = \Phi(x_i);$
$\underbrace{\Delta_i = x_{i+1} - \Phi(x_i)}_{\text{Small / low entropy.}}$
STORAGE

APPLICATION

↓ Raw data

Scientific Data $\ldots, x_{i-1}, x_i, x_{i+1}, \ldots$

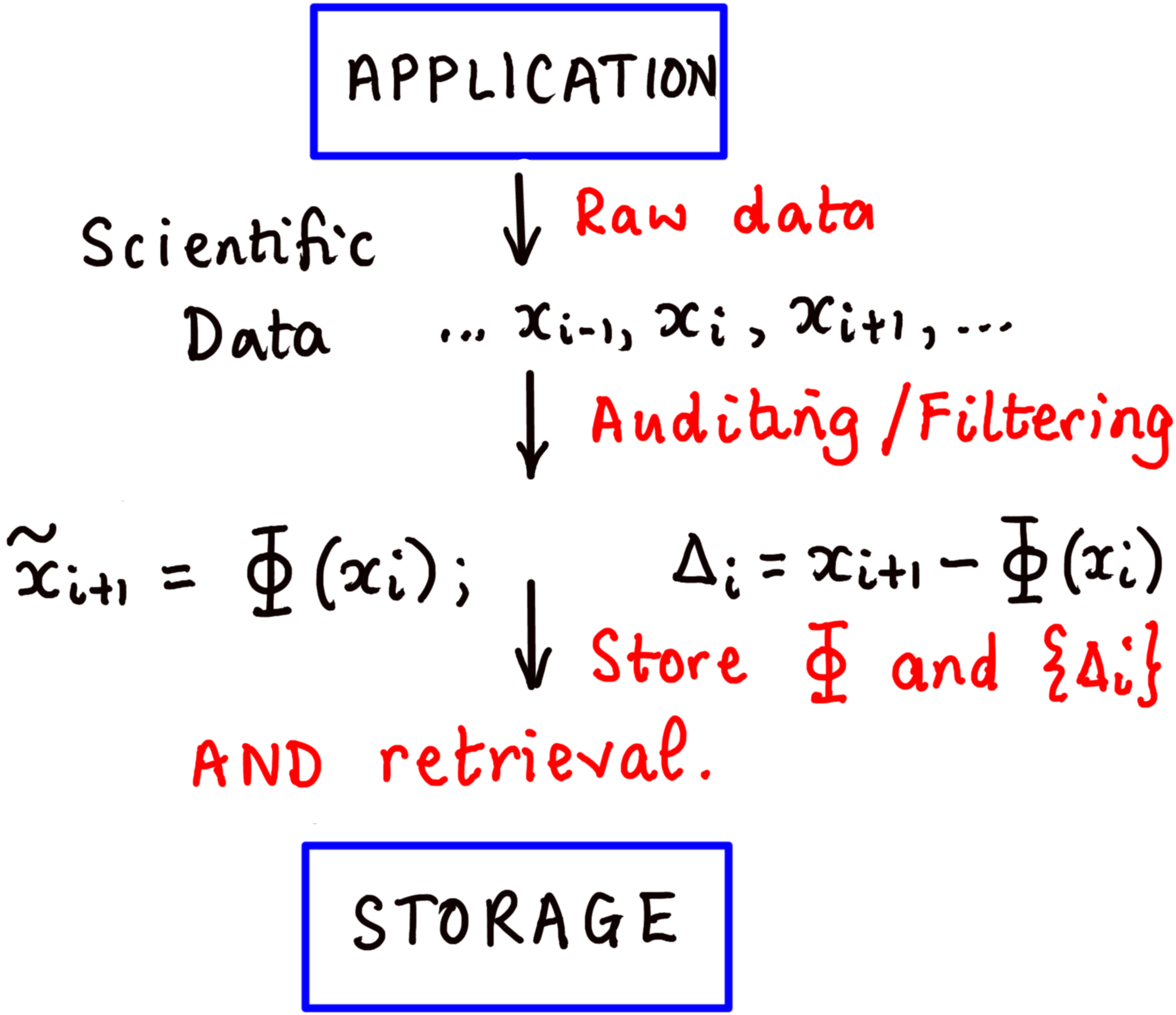
↓ Auditing/Filtering

$\tilde{x}_{i+1} = \Phi(x_i)$; $\Delta_i = x_{i+1} - \Phi(x_i)$

↓ Store Φ and $\{\Delta_i\}$

$\Rightarrow$ Need more sophisticated storage

STORAGE


APPLICATION
Raw data
Scientific Data
$\ldots, x_{i-1}, x_i, x_{i+1}, \ldots$
Auditing / Filtering
$\tilde{x}_{i+1} = \Phi(x_i)$;
$\Delta_i = x_{i+1} - \Phi(x_i)$
Store Φ and $\{\Delta_i\}$
AND retrieval.
STORAGE

How to choose an Auditor?

- Lazy Auditor

$$\Phi(x_i) = x_i$$

Guess data did not change (much).

Used in MANY compression algorithms

e.g. fpc Burtscher et al.

HOW TO CHOOSE AN AUDITOR?

- Domain Specific Knowledge

True / exact physics

$$x_{i+1} = \Phi_F(x_i)$$

Expensive

Simplified / Reduced Model

$$\tilde{x}_{i+1} = \Phi_C(x_i)$$

Cheap

e.g. Isenburg & Lindstrom mesh compression

HOW TO CHOOSE AN AUDITOR?

- Wavelet / Multi-resolution compression

$$\left.\begin{aligned} x_i &= F(t_i) \\ F(t) &= \sum_{k,\ell} c_{k\ell}\,\psi(2^{\ell} t - k) \end{aligned}\right\} \begin{array}{c}\text{Exact} \\ \& \\ \text{Expensive}\end{array}$$

$\downarrow$ TRUNCATE $\downarrow$

$$x_{i+1} \approx \sum_{\substack{|k| \le K \\ \ell \le L}} c_{k\ell}\,\psi(2^{\ell} t_{i+1} - k)$$