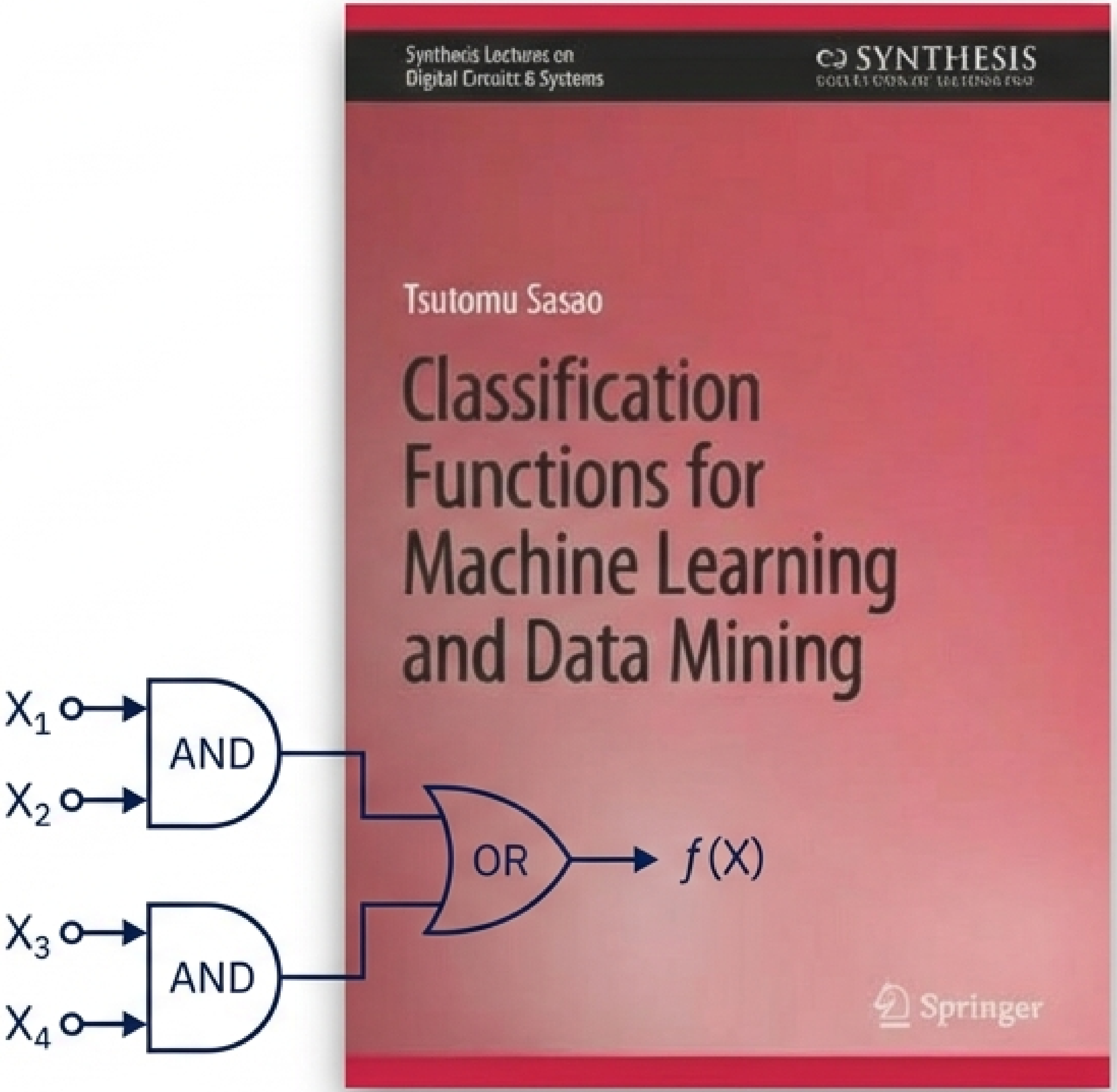
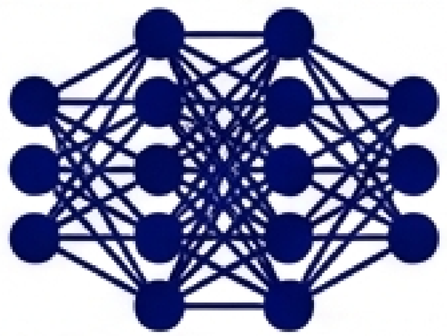
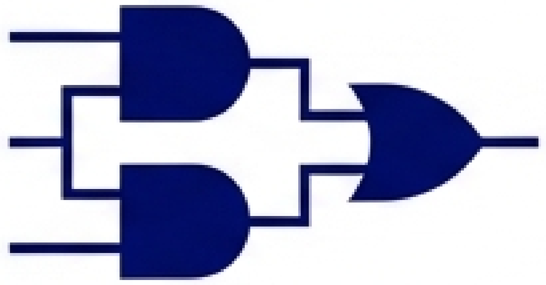
Classification Functions for Machine Learning

A Logic Synthesis Approach to Data Mining and Pattern Recognition

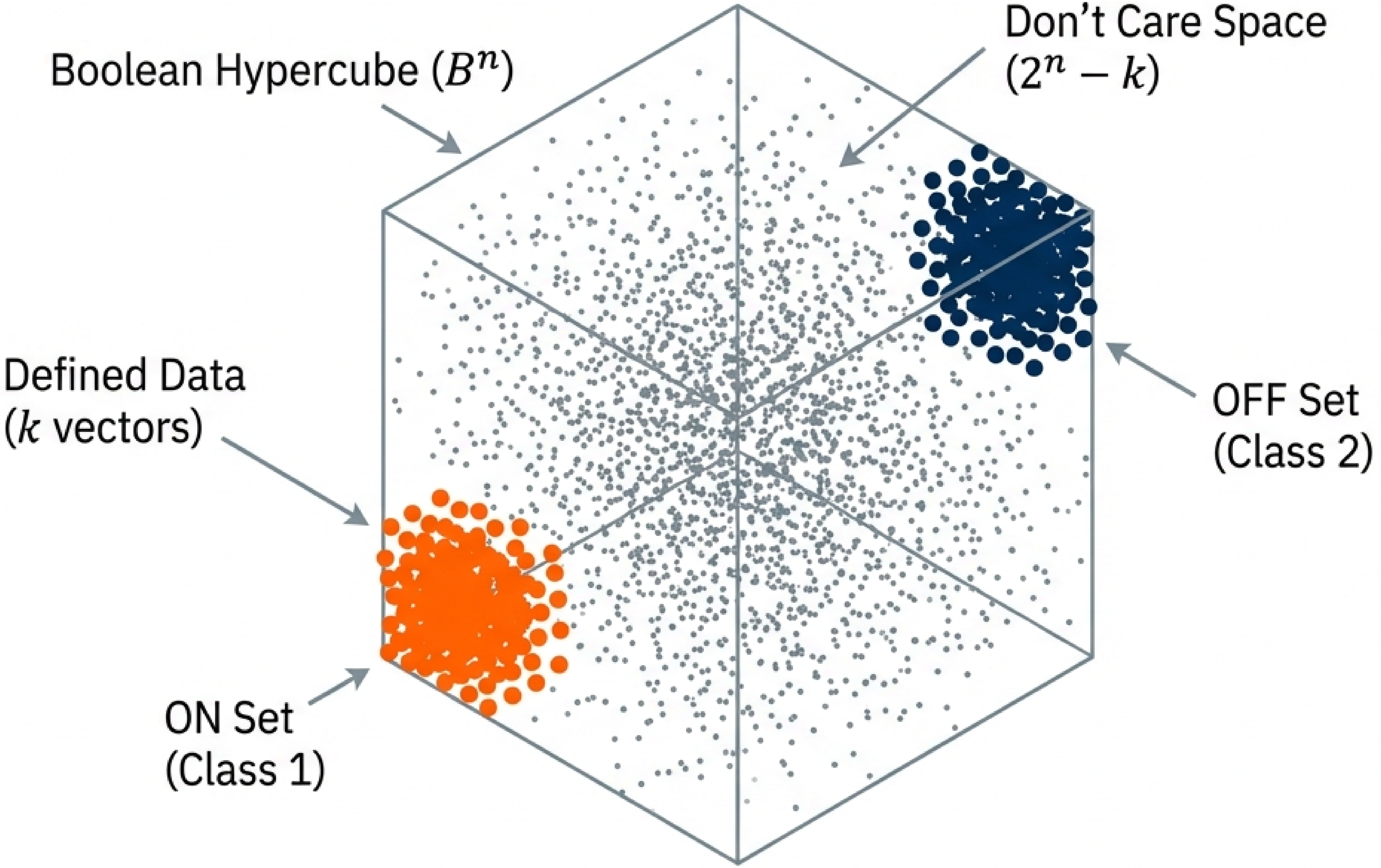


Based on the research of
Tsutomu Sasao | Meiji University, Japan

Trading Accuracy for Radical Hardware Simplicity

	Neural Networks (The Current Paradigm)	Logic Synthesis / SOP (The Proposed Solution)
Circuit Structure	 Multi-level threshold network	 AND-OR two-level structure
Learning Method	Backpropagation / SGD (Continuous)	Variable Reduction (Discrete)
Hardware Cost	High Power / Massive GPU Memory	**Low Power / Compact LUTs**
Interpretability	Black Box (Opaque)	**White Box** (Explicit Boolean Rules)
Generalization Strategy	High Accuracy, Weights-based	Lower Accuracy, Logic-based

Classification as a Partially Defined Function



$f : D \to M$

$D \subset B^n$
(Binary inputs)

$M = \{1, 2, \ldots, m\}$
(Classes)

Key constraint:
Training size $k \ll 2^n$

The Objective: Minimization of Variables

Support Set S

x_1	x_2	x_3	x_4	f
0	0	1	0	0
0	0	1	1	0
0	1	0	0	1
0	1	0	1	1
1	0	1	0	0
1	0	1	1	0
1	1	0	0	1
1	1	0	1	1

If $F_i|_S \cap F_j|_S = \emptyset$, the reduced variables are sufficient.

Exact Minimization via Difference Vectors

Identify Essential Variables

Is unit vector e_i in the Difference Set?

Generate Minimal Difference Vectors

Calculate $d = a \oplus b$ for all pairs.

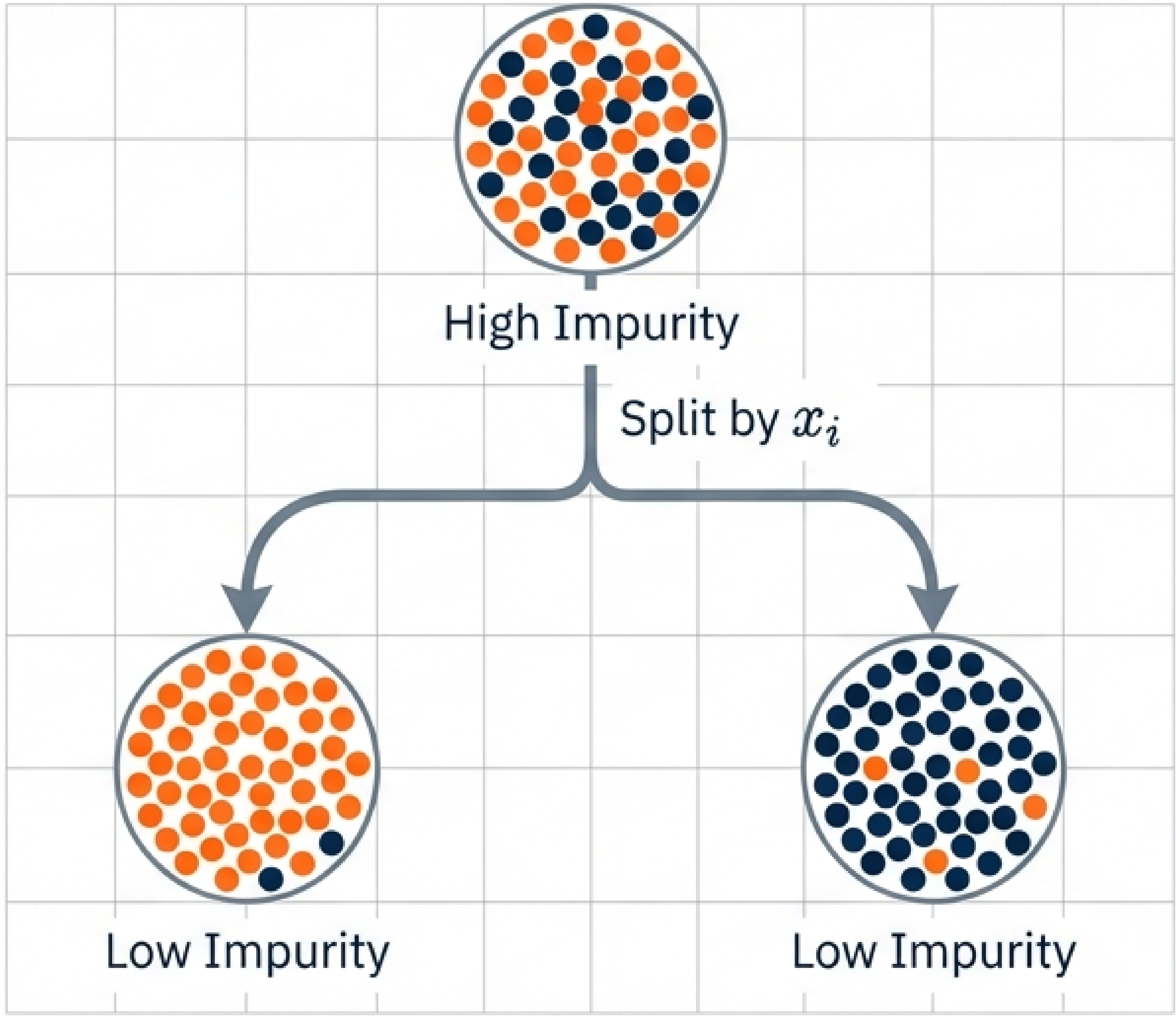
Solve Covering Problem

Find minimum support set S that covers all differences.

Difference Vector Definition:

- $\vec{d} = \vec{a} \oplus \vec{b}$
- Variable x_i is essential iff difference set contains e_i.

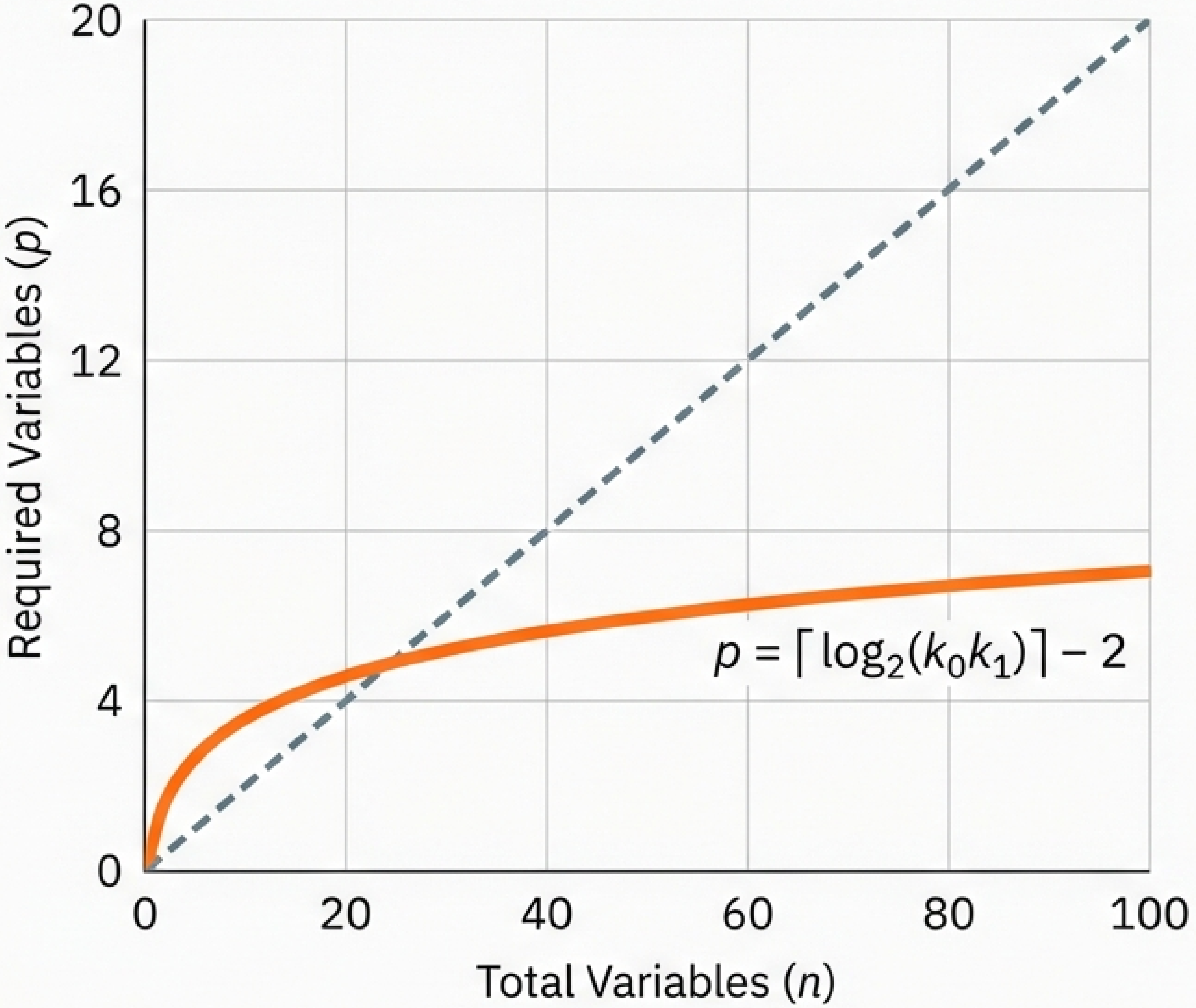
Heuristic Selection: The Impurity Measure



$$\mu(\vec{a}) = \sum \left[Size(j)^2 - \sum Hist(j, Value)^2 \right]$$

Greedy Strategy: At each step, select variable x_i that minimizes μ.

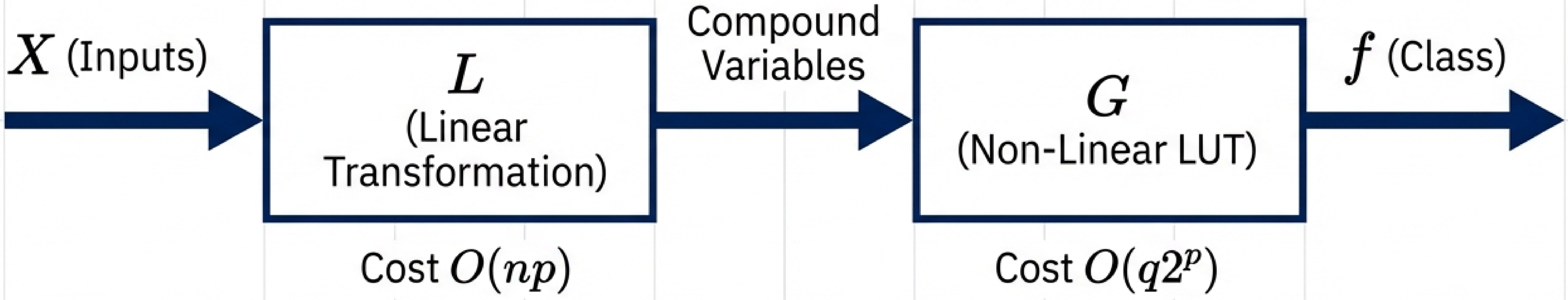
Statistical Bounds of Variable Reduction



The Parsimony Guarantee

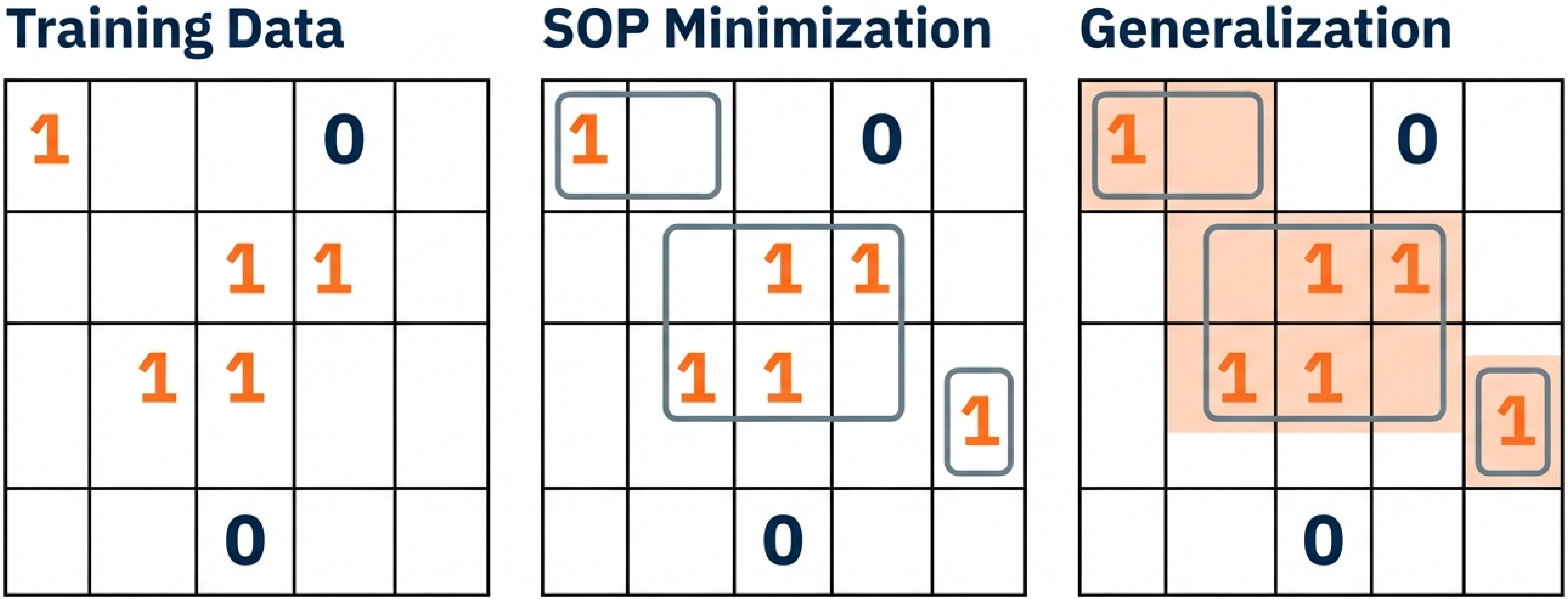
For random functions, as $n \to \infty$, the probability of needing **needing** $> p$ variables approaches 0.

Handling Non-Linearity: Linear Decomposition



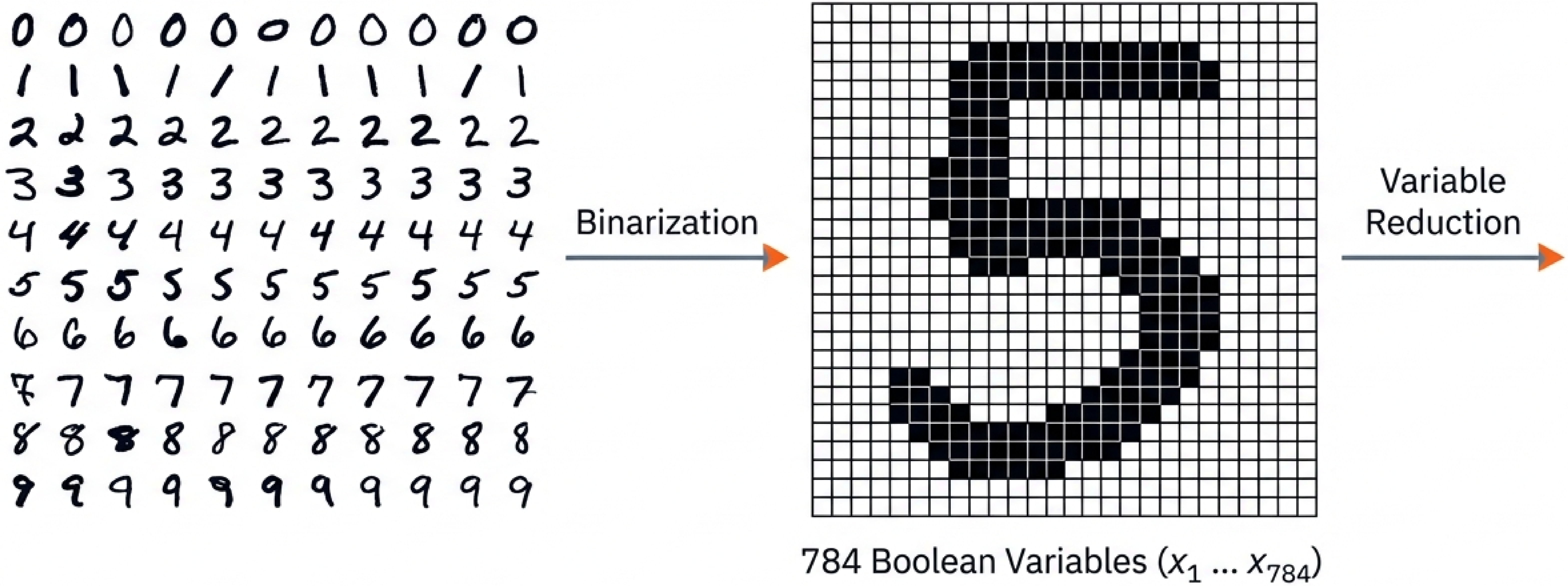
Solves the "Achilles Heel" (e.g., XOR functions) by shifting complexity to the linear block.

Generalization via Logic Minimization



Minimizing the logic naturally “fills in” the Don’t Care space with the simplest explanation.

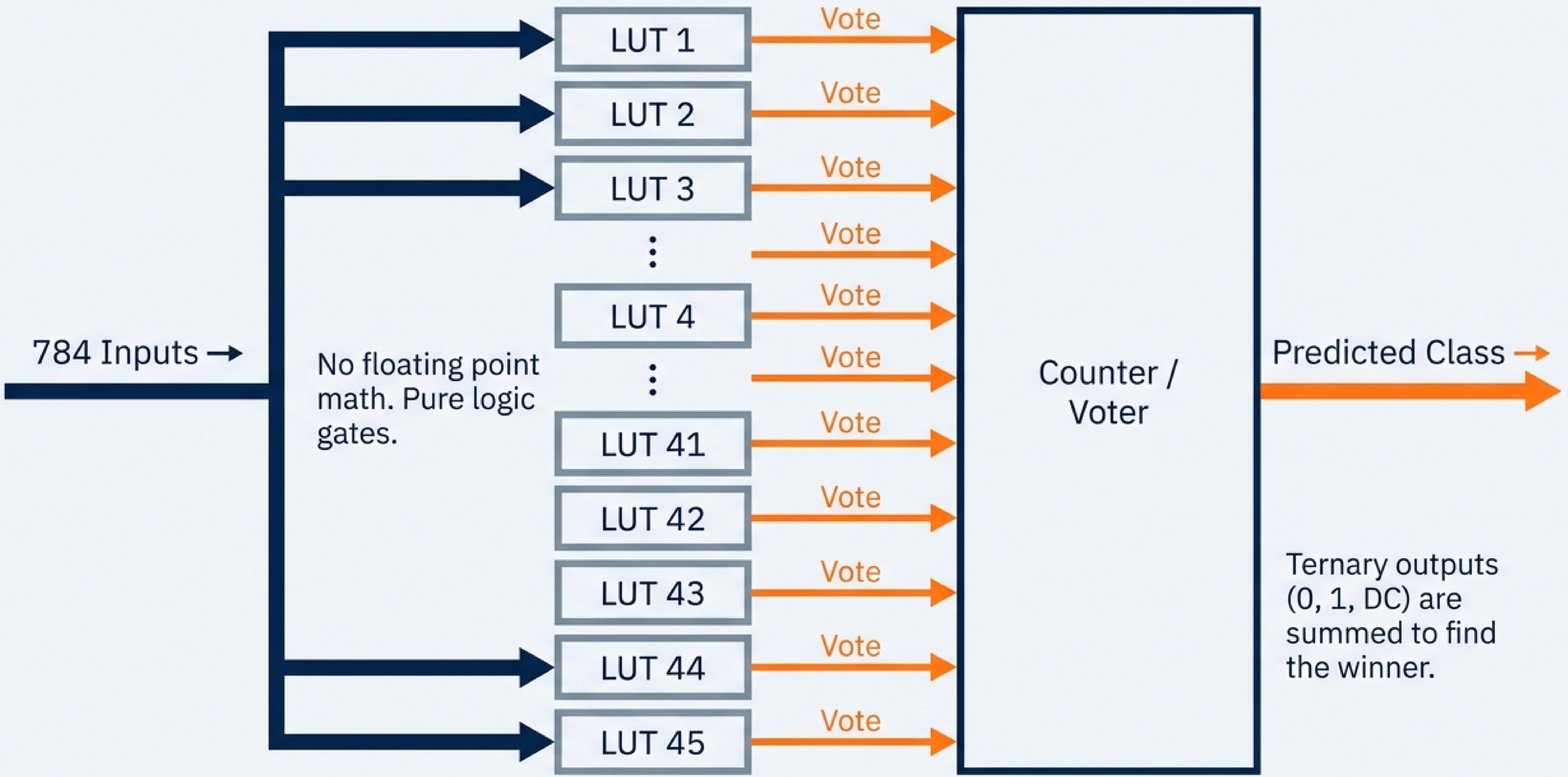
Case Study: MNIST Handwritten Digits



x_{42}, x_{105}, x_{300}, x_{300}, x_{512}, x_{512}, x_{618}, x_{678}, x_{220}, x_{236}, x_{321}, x_{328}, x_{335}, x_{333}, x_{335}, x_{401}, x_{435}, x_{450}, x_{426}, x_{378}, x_{390}, x_{438}, x_{448}, x_{452}, x_{475}, x_{500}, x_{512}, x_{521}, x_{613}, x_{667}, x_{678}...

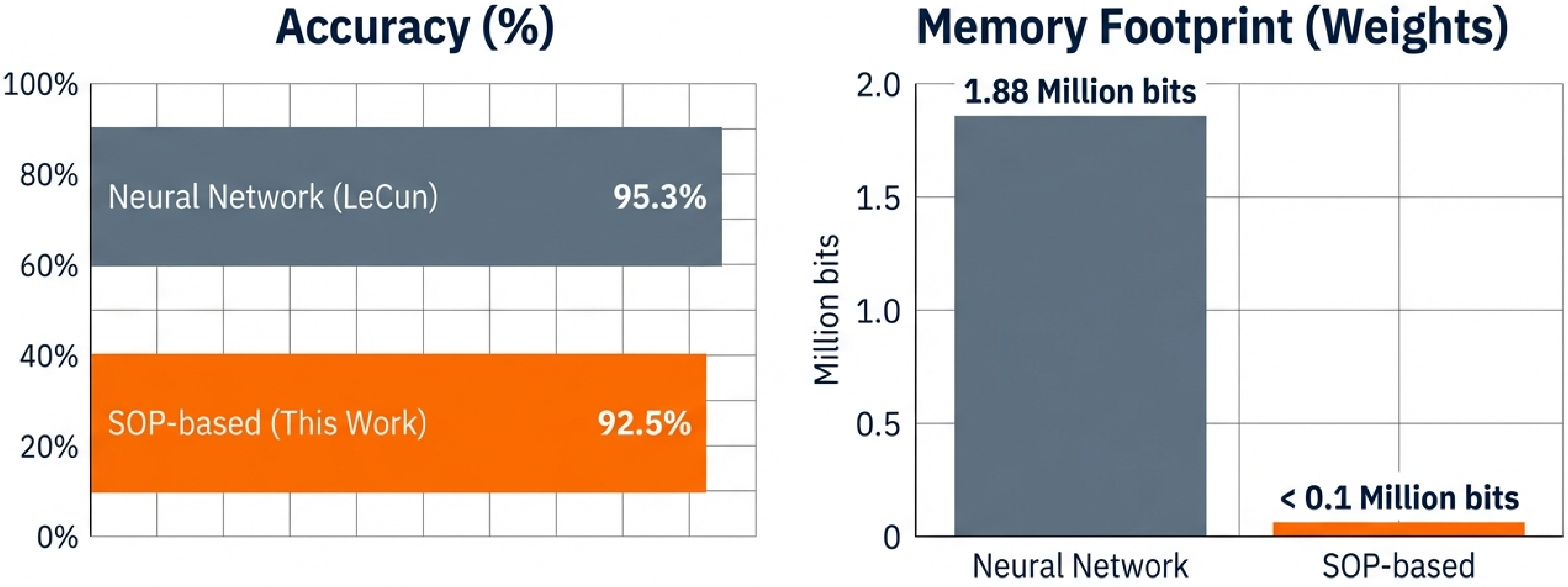
Training: 60,000 images. Testing: 10,000 images.

The 45-Unit Realization Architecture



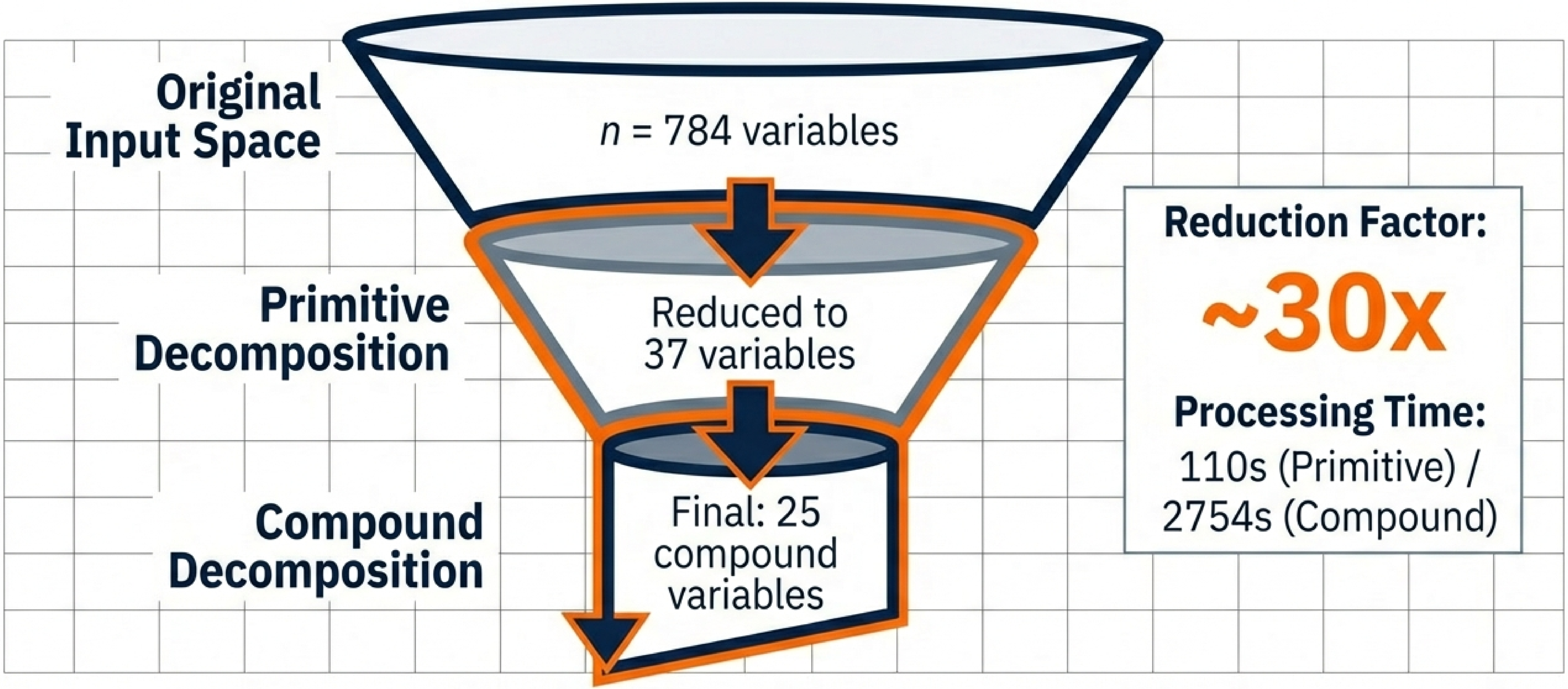
NotebookLM

Performance Comparison: The Trade-off



Interpretability: Neural Network = Low (Black Box) | SOP = High (White Box)

Decomposing High-Dimensional Data



Reduction Factor:

~30x

Processing Time:

110s (Primitive) / 2754s (Compound)

Future Directions & Open Problems

Continuous Variables

Developing better discretization for sensor/medical data.

Multi-Valued Inputs

Minimization for non-binary (>2 state) logic functions.

Closing the Gap

Improving ensemble methods to reach 95%+ accuracy.

Applications

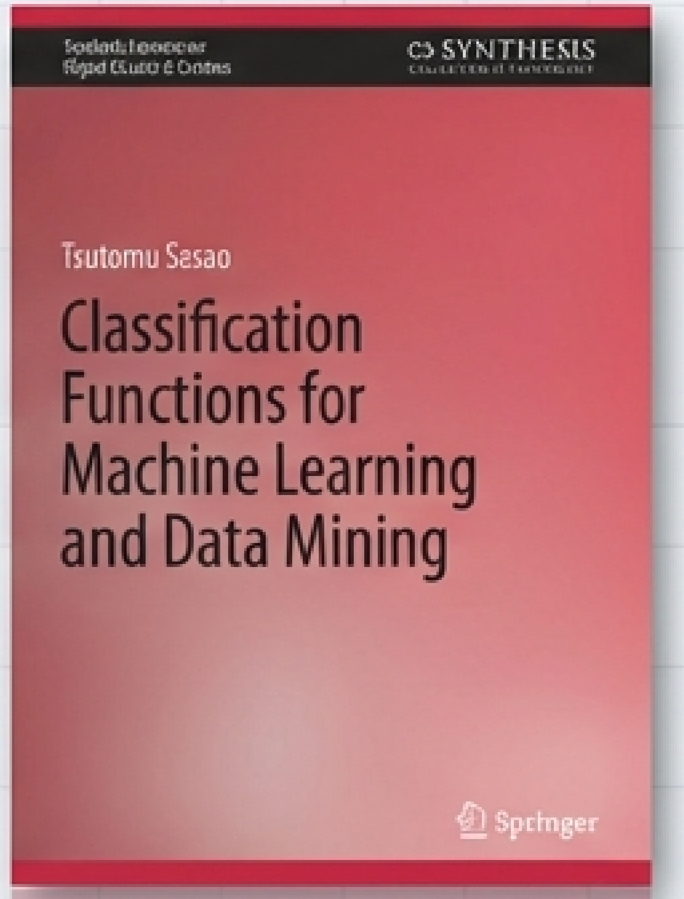
Packet classification, Intrusion Detection, High-Energy Physics.

Summary of Contributions

1. **Parsimony:** Classification possible with $p \approx \log_2 k$ variables.

2. **Transparency:** Logic synthesis yields interpretable ‘White Box’ rules.

3. **Efficiency:** Orders of magnitude reduction in memory for FPGA/ASIC.



Sasao, T. (2023). *Classification Functions for Machine Learning and Data Mining*. Springer.