AI for Code-based Cryptography

SAC 2025

Aug 14th

Mohamed Malhou[1,3], Ludovic Perret[2,3], Kristin Lauter[1]

FAIR[1] EPITA[2] Sorbonne Université[3]

Post-Quantum Cryptography Standards

NIST Standardization Process & Algorithm Selection

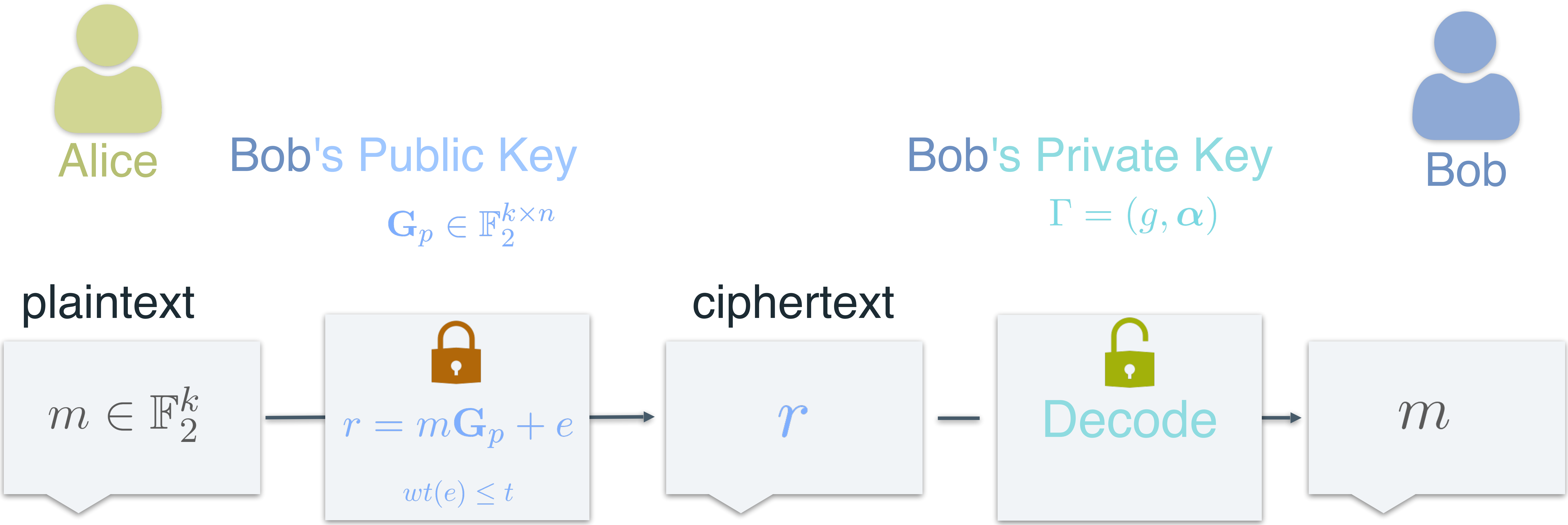
Category	Primary Algorithm	Alternate Algorithms
Public-Key Encryption/KEMs	**CRYSTALS-Kyber**	HQC
Digital Signatures	**CRYSTALS-Dilithium**	FALCON, SPHINCS+

Fourth Round KEM Finalists (2022-2025)

- BIKE
- Classic McEliece
- HQC *(selected as alternate in March 2025)*
- SIKE *(withdrawn due to security concerns)*

Source: National Institute of Standards and Technology (NIST)

Classic McEliece



Binary Irreducible Goppa Codes

- A family of **error-correcting codes** defined over a finite field $\mathbb{F}_q$ with **q=2**.

- **Parameterized by $\Gamma = (g, \alpha)$:**
  - ▶ A set of distinct elements $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{F}_{q^m}$ called the **support**.
  - ▶ A polynomial $g(x) \in \mathbb{F}_{q^m}[x]$ of degree t, **irreducible**.

Key Properties

- Code dimension **k ≥ n - mt**

✓ Corrects up to t errors in a codeword

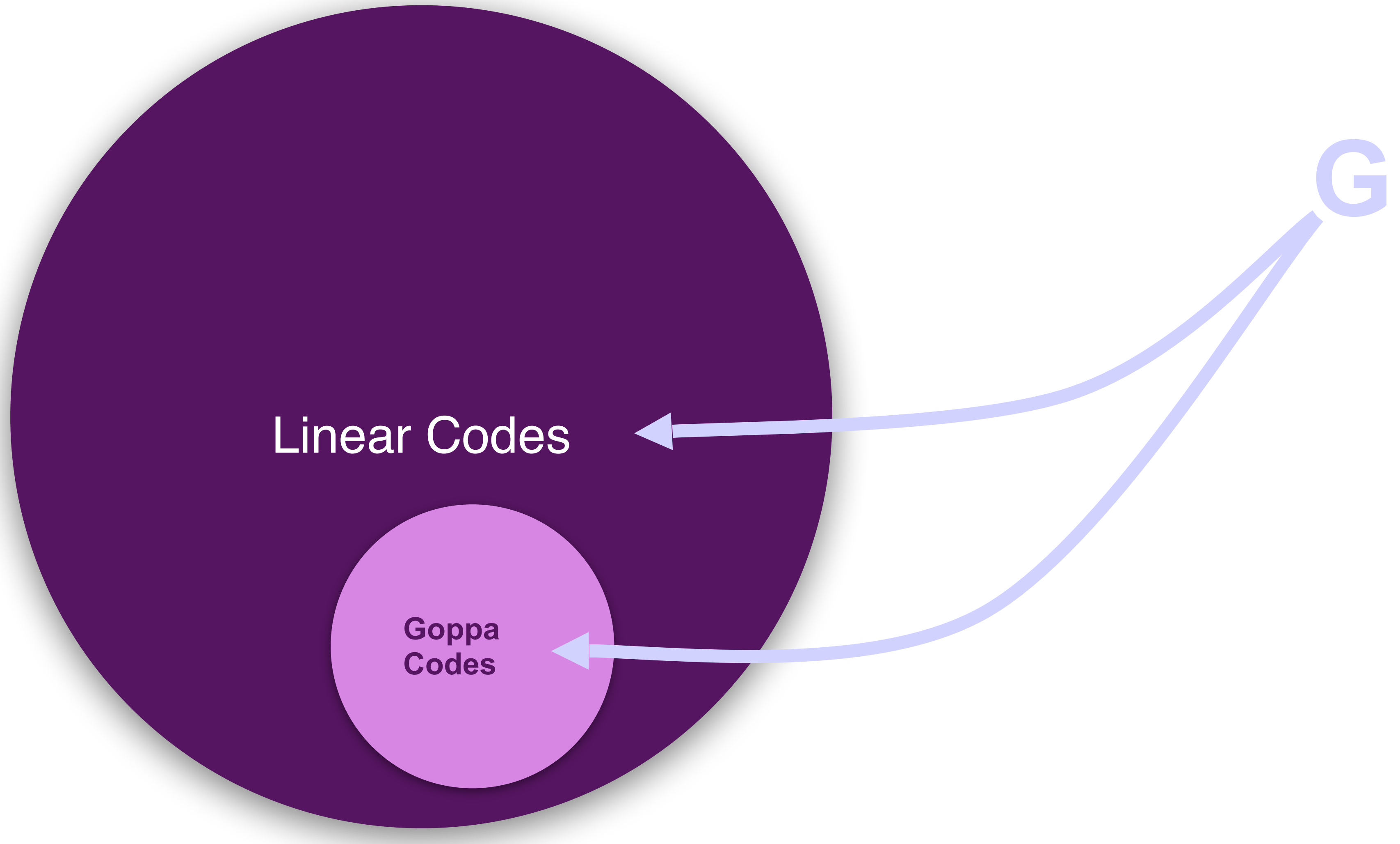
Defined as the $\mathbb{F}_q$-kernel of $V_t[\alpha, \beta]$

Where $\beta = (g(\alpha_1)^{-1}, g(\alpha_2)^{-1}, \ldots, g(\alpha_n)^{-1})$

$$V_t[\alpha, \beta] = \begin{pmatrix} \beta_1 & \beta_2 & \cdots & \beta_n \\ \beta_1\alpha_1 & \beta_2\alpha_2 & \cdots & \beta_n\alpha_n \\ \vdots & \vdots & \ddots & \vdots \\ \beta_1\alpha_1^{t-1} & \beta_2\alpha_2^{t-1} & \cdots & \beta_n\alpha_n^{t-1} \end{pmatrix}$$

Security Analysis of McEliece

- Hardness of Decoding
- Goppa Distinguishing Problem



Some Distinguishers in the Literature

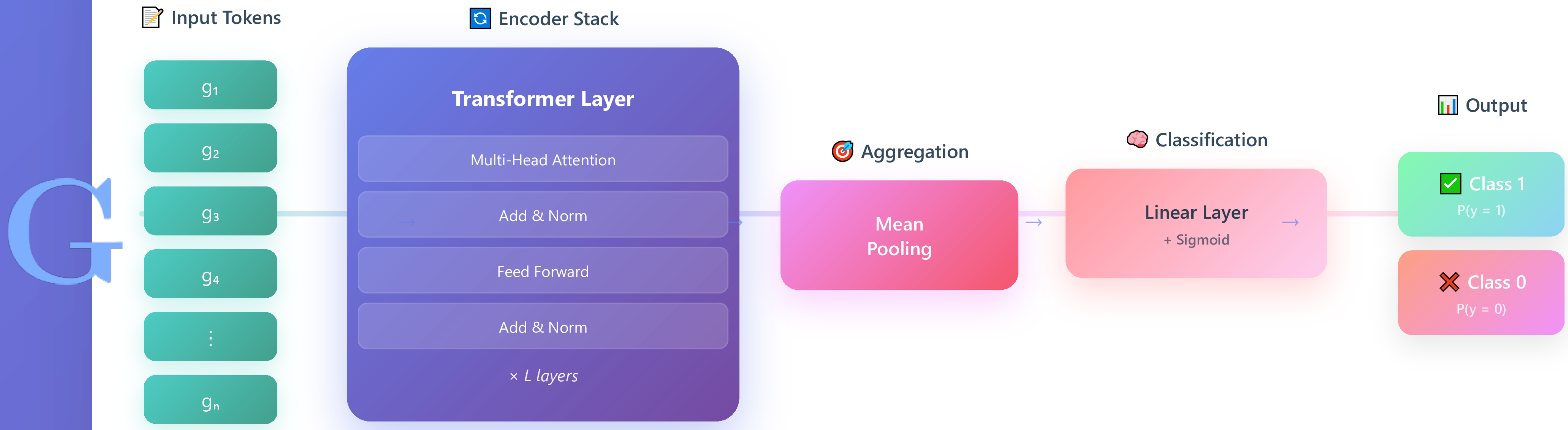
J.-C. Faugère, V. Gauthier-Umana, A. Otmani, L. Perret, J.-P. Tillich.
A Distinguisher for High Rate McEliece Cryptosystems.
IEEE-IT 2013

A. Couvreur, R. Mora, J.-P. Tillich.
A New Approach Based on Quadratic Forms to Attack the McEliece Cryptosystem.
Asiacrypt 2023.

H. Randriambololona.
The Syzygy Distinguisher.
EUROCRYPT 2025.

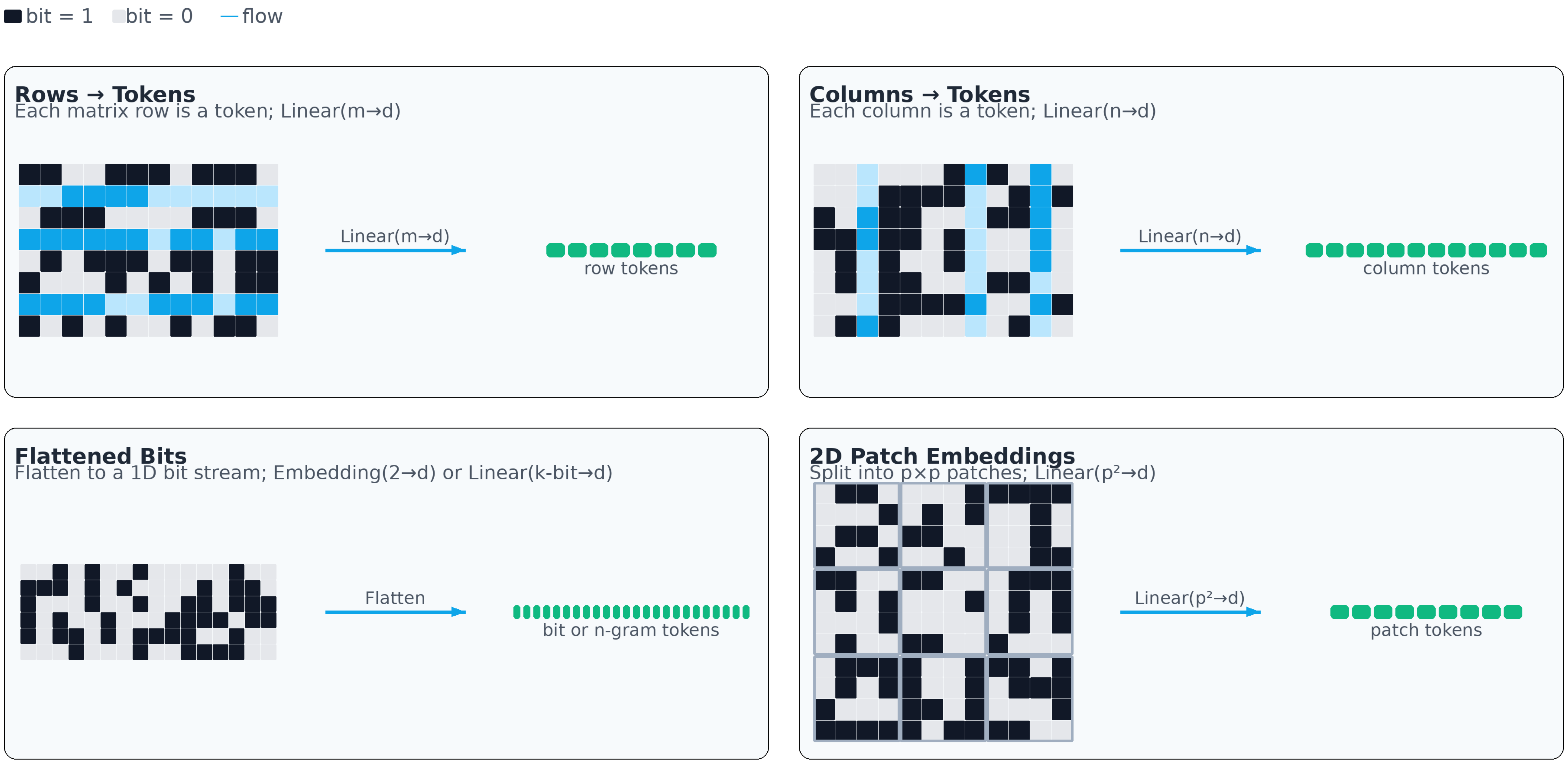
DEEP LEARNING

Transformer-Based Binary Classification



Architecture: Encoder-Only • Mean Pooling • Self-Attention • Residual Connections • Layer Normalization

Tokenizing & Encoding Binary Matrices for Transformers

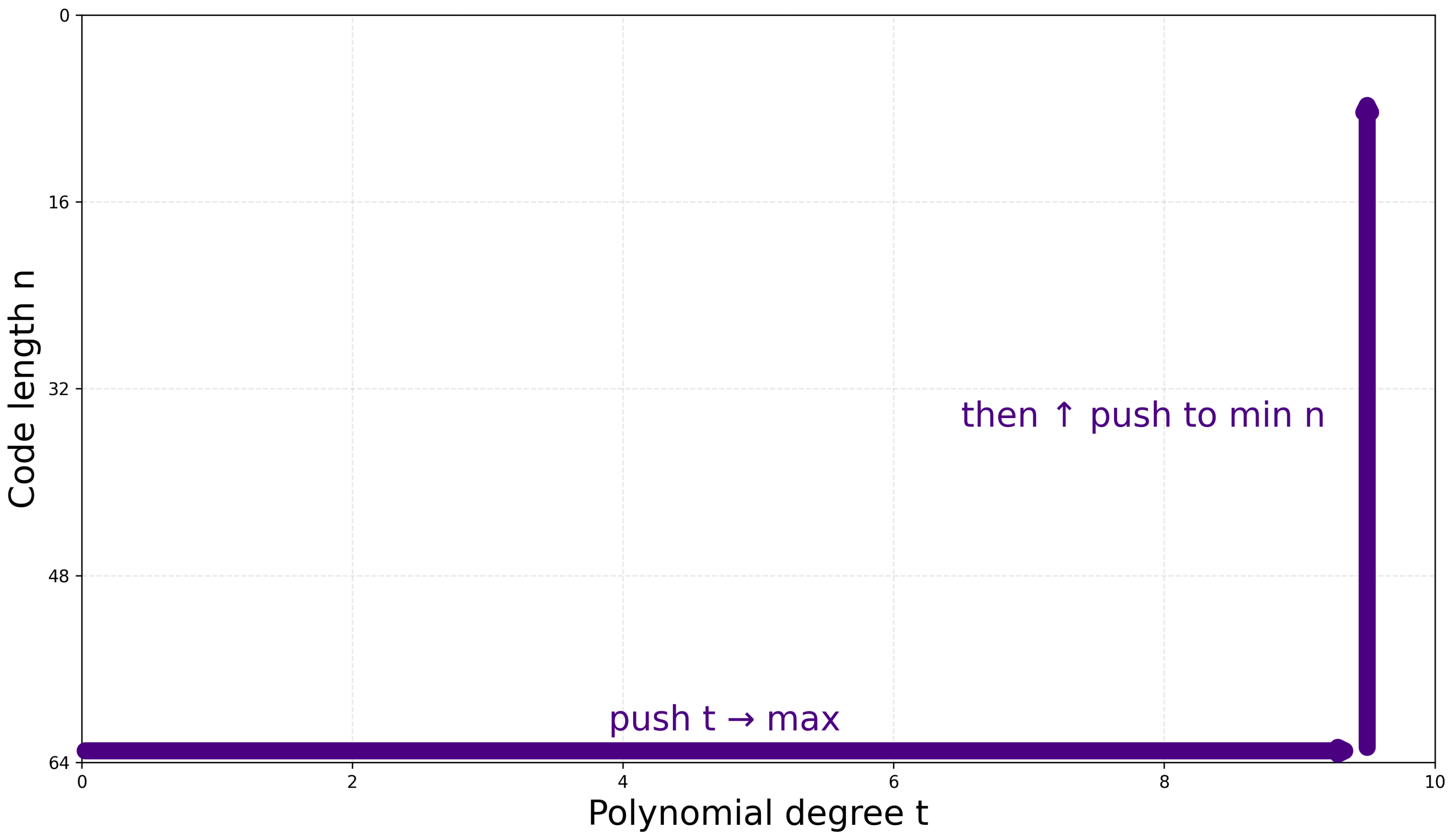


Evaluation of the distinguishers in the literature [Randriam 24 & CMT 23]

Consider a field extension degree m (e.g. m= 6)

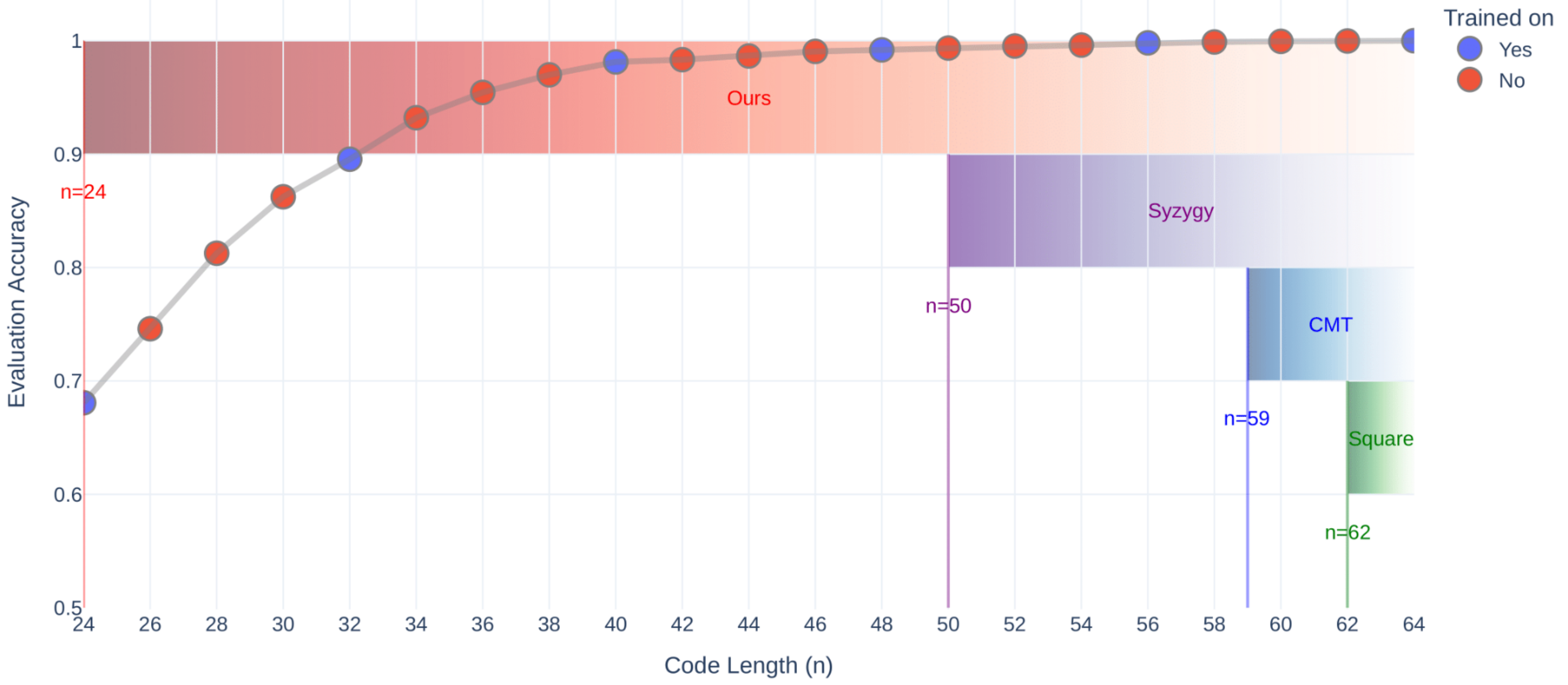
For maximum $n = q^m$ (e.g. n= 64), determine maximal distinguishable degree t.

Fix t, determine minimal distinguishable n.

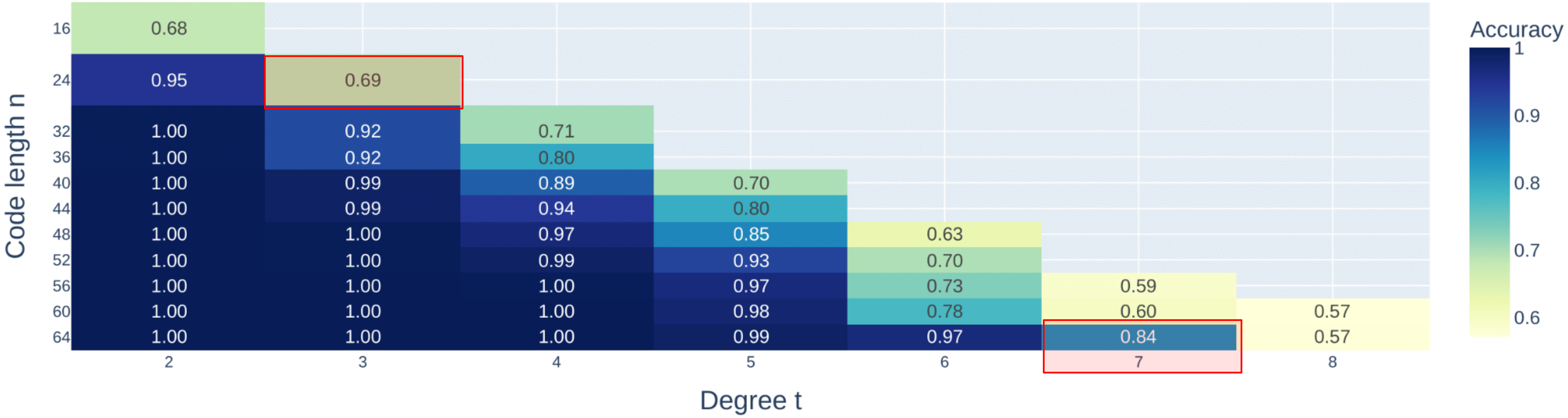


Results:

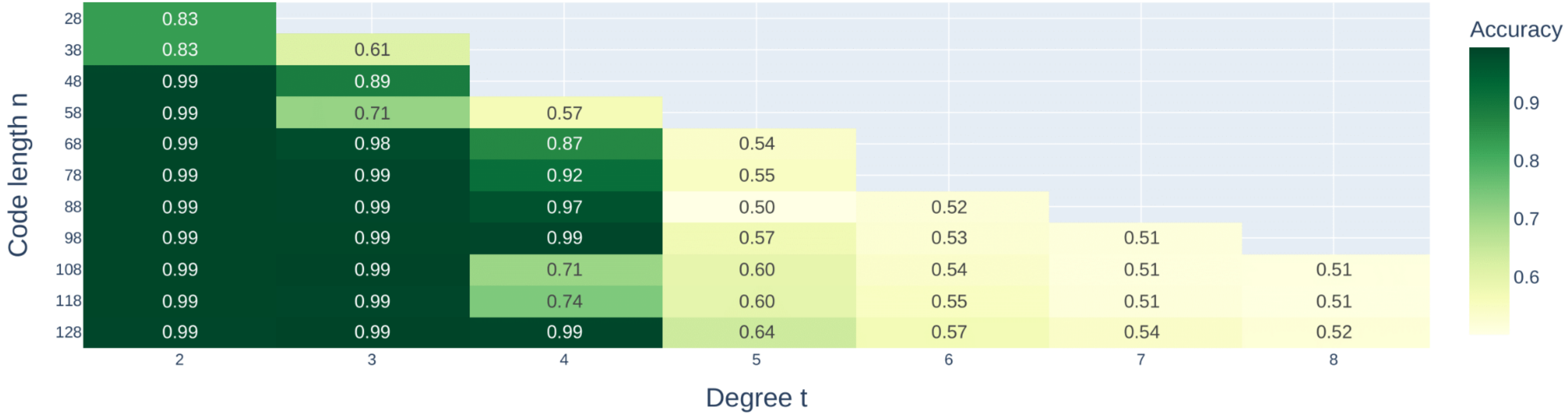
- Model accuracy as a function of code length. The model is trained on Binary Goppa Codes with m=6 and t=3.
- Scatter points indicate the evaluation accuracy of our model.
- The scatter color indicates the range where the model was trained (n = 24 + 8k for k=0,1,..)



- Heatmap of model accuracy for q=2, m=6 as a function of code length n and degree parameter t.

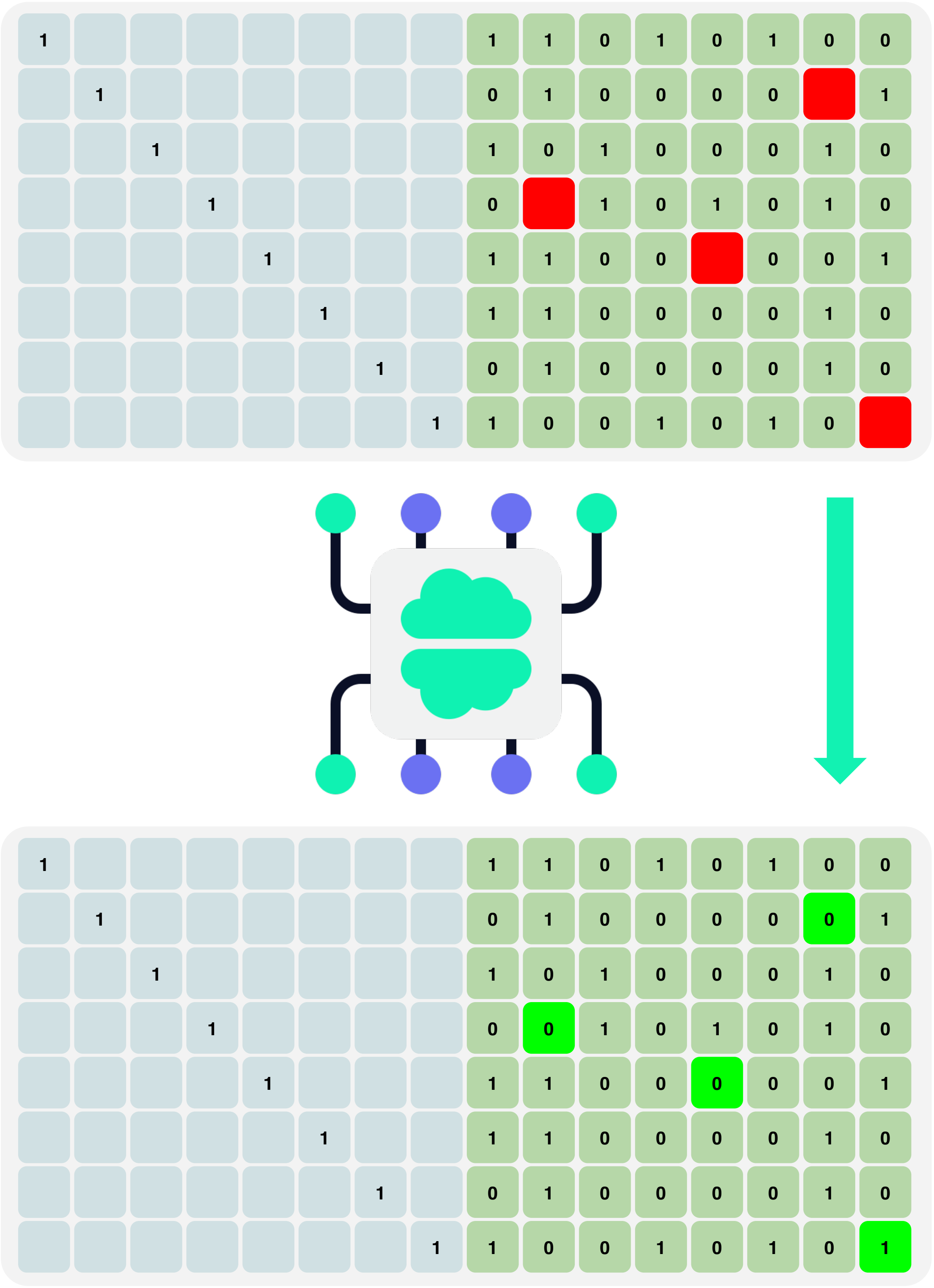


❑ Heatmap of model accuracy for q=2, m=7 as a function of code length n and degree parameter t.

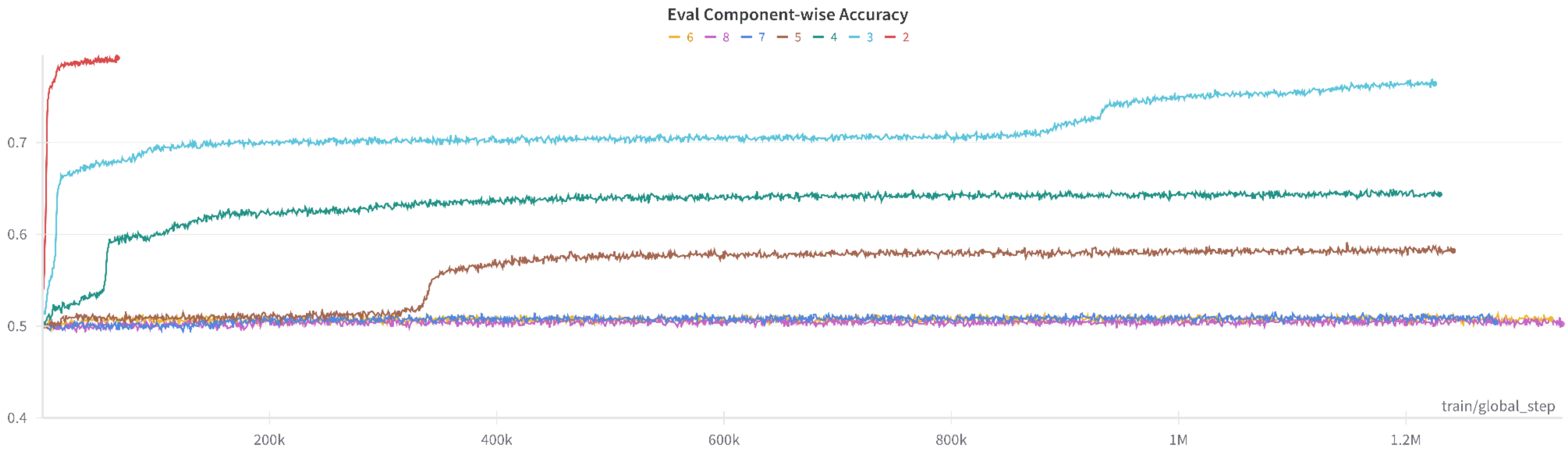


Can we do better than just distinguish ?

Training on a new task:
Goppa Code Completion



Test Accuracy on n=64, m=6



Summary and Conclusion

Paper: Mohamed Malhou, Ludovic Perret, Kristin Lauter
AI for Code-based Cryptography
Code available at github.com/facebookresearch/ai4code-cryptanalysis

★Improve SOTA Goppa distinguishers in toy examples.
- What are the limits of this approach and how to estimate the complexity?

★Results can be extended to (QC) MDPC Codes (BIKE).

Solution icons created by Canticons - Flaticon