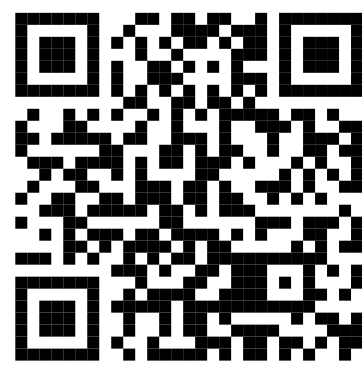




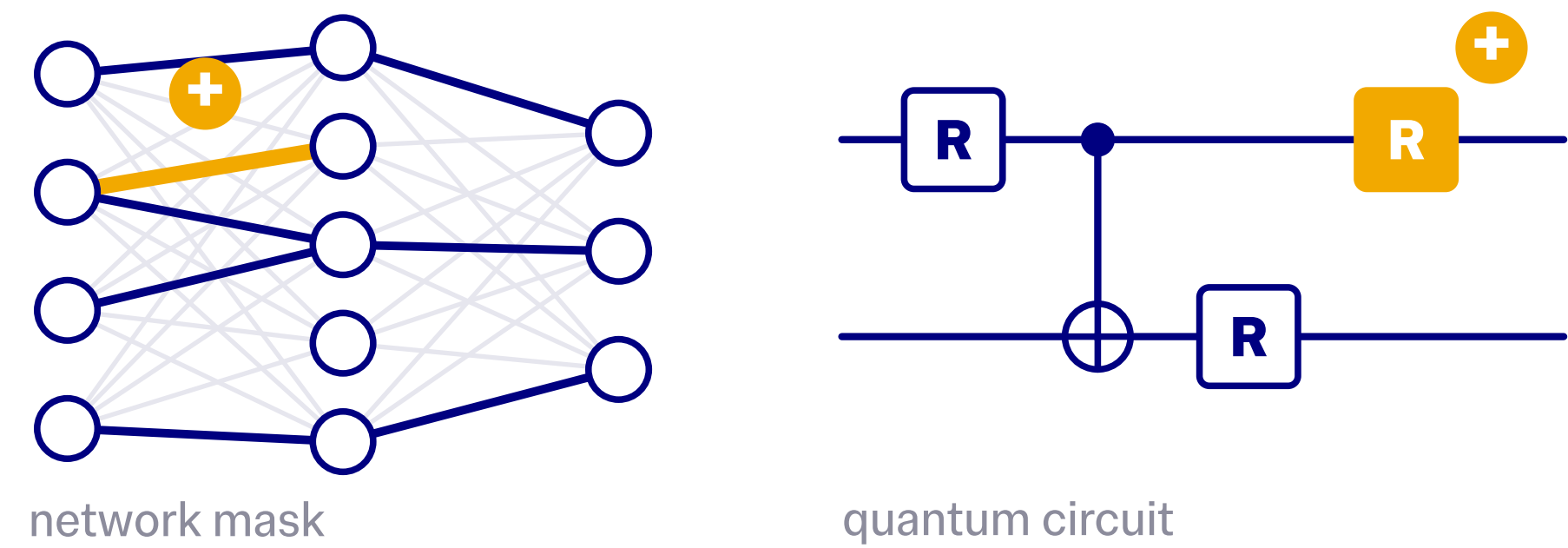
Sparsity for Free: A Budget-Induced Equilibrium in Joint Topology–Parameter Search

Marcel Mordarski^{1*}, Benjamin Gras^{2†}, Abdelrahman Shehata^{2†},
Daniel Budina^{2†}, Roberto Bondesan¹

¹Dept. of Computing, ²Dept. of EEE, Imperial College London ^{*}equal contribution marcel.mordarski25@imperial.ac.uk

1 The problem

Adding a weight or a gate is free.
What stops the topology growing?



Architecture search [1, 2], pruning [3] and quantum circuit design [4, 5] choose a topology T (which weights or gates exist) and tune its parameters θ . Nothing in the score charges for size, but each new element adds parameters, all tuned within the **same budget of B evaluations**.

So far, sparsity is imposed as a knob:

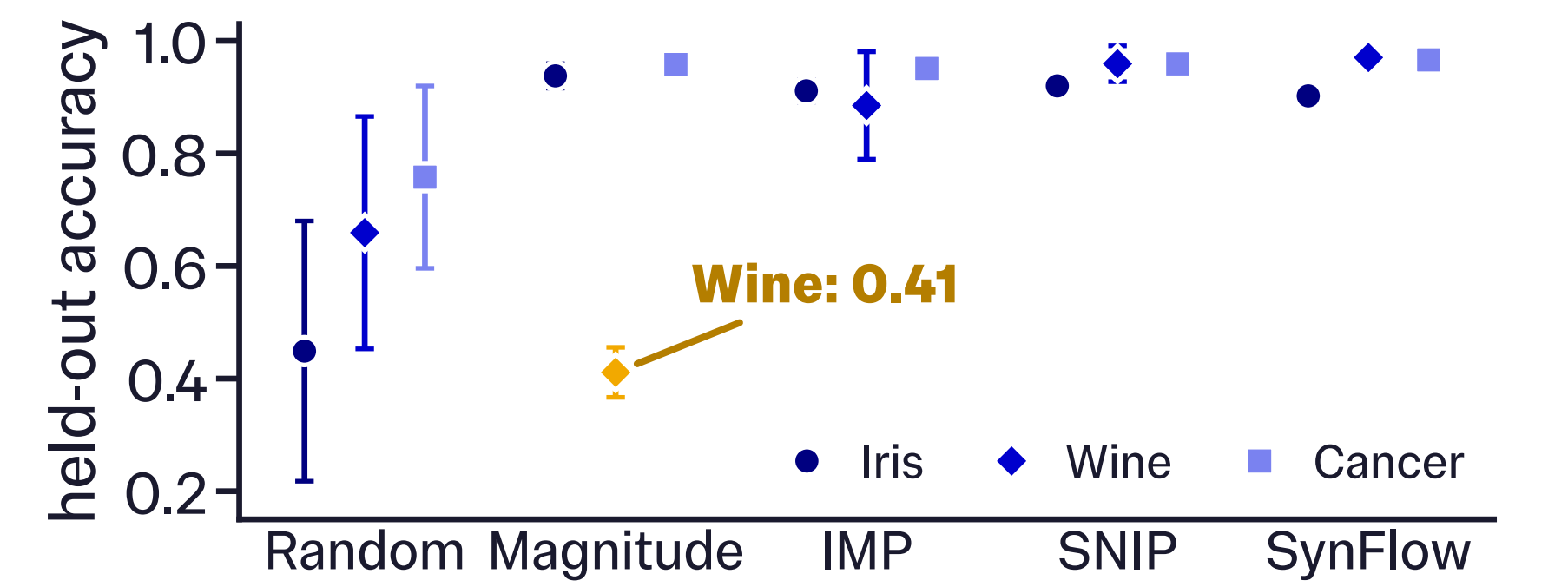
$\lambda|T|$ size penalty
or extra objective [5]

target density
[3, 6, 7]

pruning schedule
[8, 9]

fixed ansatz
[4]

Knobs need tuning per task, and standard pruning heuristics can fail at extreme sparsity [10]:



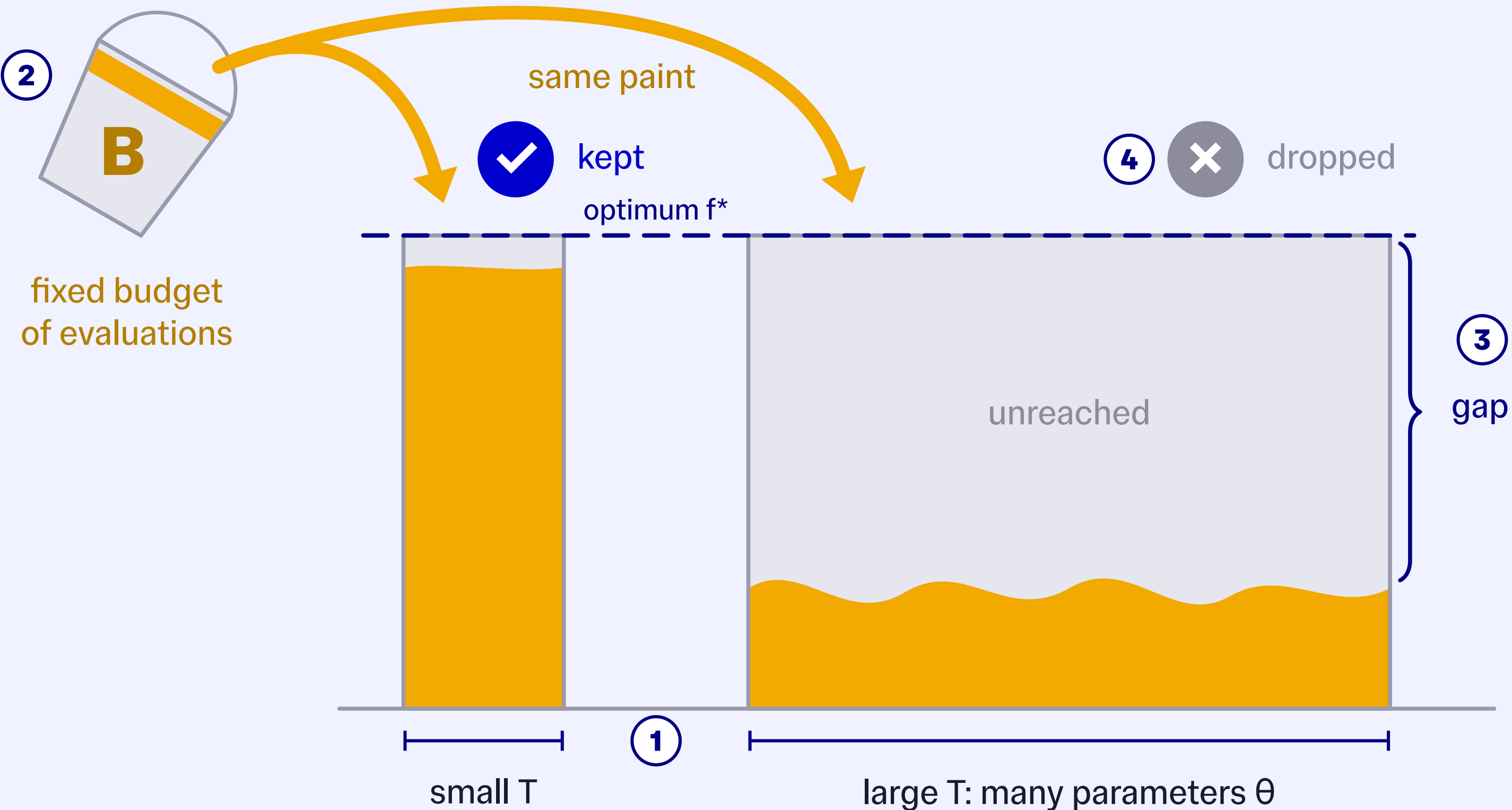
Held-out accuracy at 95% sparsity, mean $\pm$ sd over five seeds, on three small UCI datasets [11–13]: Magnitude [3], IMP [9], SNIP [6], SynFlow [7]. Magnitude pruning collapses on Wine, below a random mask, possibly a small-data artefact.

Missing: what does the budget B itself do?

2 Dimensionality pressure

The budget itself favours small topologies.

With more parameters but the same B evaluations, a larger topology ends up less optimised: its score is lower and noisier. Selection, which compares scores, therefore keeps smaller topologies, **with no size penalty**.

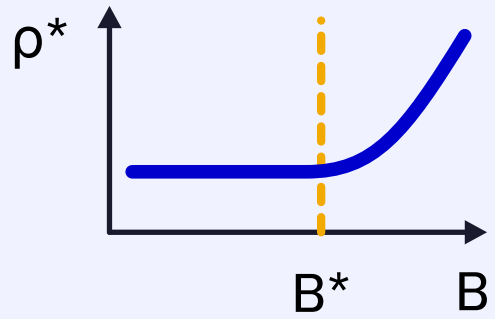


- 1 wall width = number of parameters
- 2 paint = budget B , same for every wall
- 3 coat height = score reached within B
- 4 selection keeps the higher coat

Theory

Proposition 1. If CMA-ES progress slows with dimension (A1) [14], a larger topology with the same optimum scores lower on average within B (the gap) and noisier. If selection also prefers the less noisy of two equal scores (A2, an assumption), the stationary topology size lies below dense.

Corollary 1. The equilibrium density ρ^* (share of elements kept) is flat in B until B scores denser topologies reliably (B^*), then rises: **tested in 3**.



EvoluCMAES: the test bed



Outer loop: evolution over T [1]; each generation keeps the best (elites), reseeds the worst at the initial density ρ_0 and mutates the rest. Inner loop: CMA-ES tunes θ within B [15] (SPSA [16] for E91). Between domains, only the mutation operator and the fitness change; **no size penalty**.

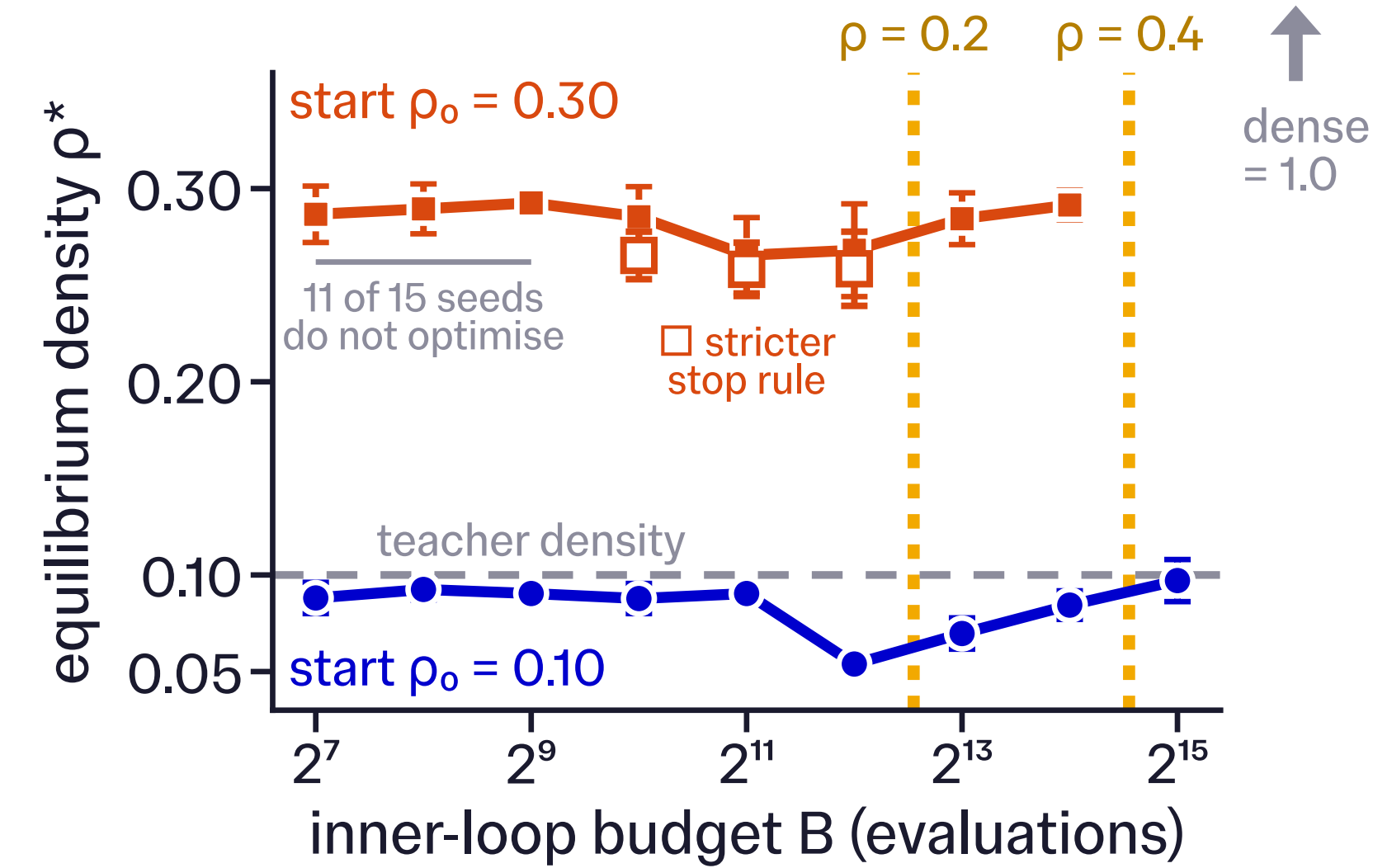
3 Multidisciplinary applications

First a controlled test of the prediction; then the same search on neural networks and quantum cryptography, and a use for its small circuits in reinforcement learning.

Controlled test · teacher–student

Does density ρ^* respond to B as predicted?

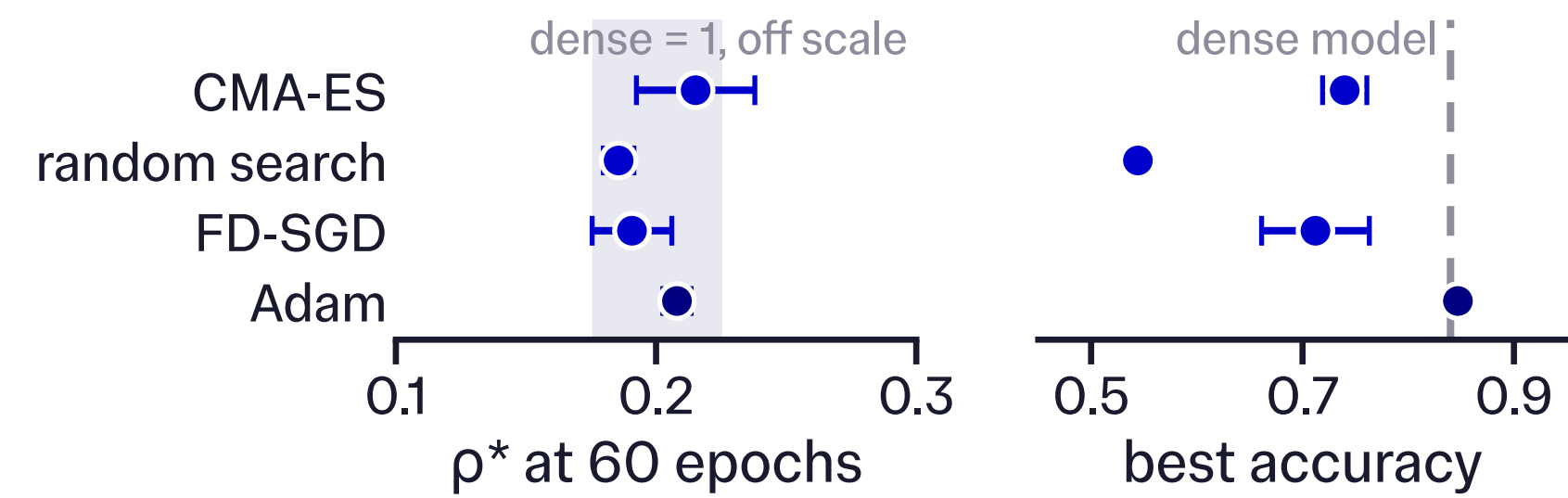
Setup: Student networks evolve weight masks to mimic a frozen teacher that uses only **10%** of its weights ($\rho = 0.10$). Extra weights add search dimensions, not score.



Result: Starting sparse ($\rho_0 = 0.10$, blue), ρ^* stays flat at low budgets, dips at 2^{12} , and **rises monotonically from $2^{12} \rightarrow 2^{15}$** once B scores denser masks reliably (dotted). Starting denser ($\rho_0 = 0.30$, red) settles in a second basin (hysteresis). Both sit below random-selection controls (0.102, 0.300).

Neural networks · MNIST

Is this specific to CMA-ES?

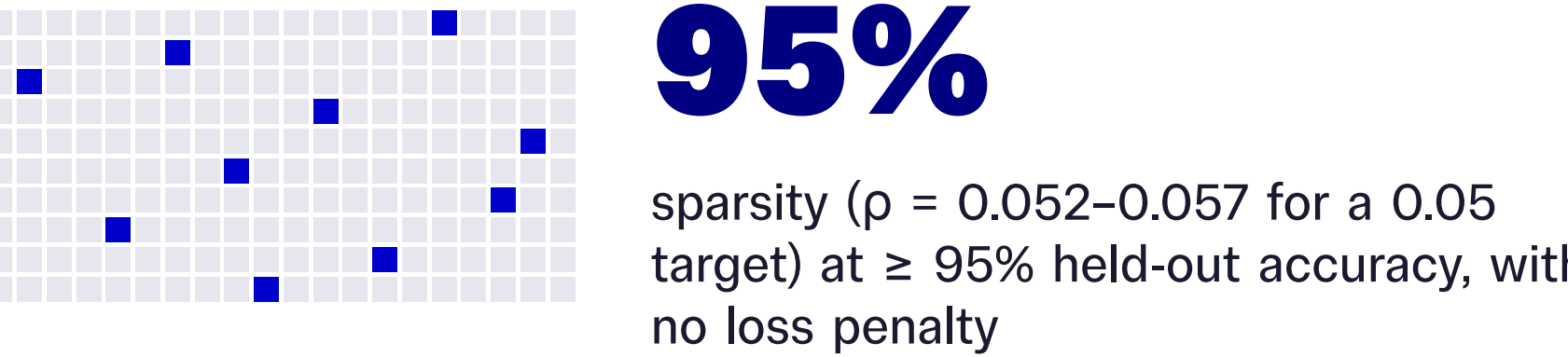


Result: Across four inner optimisers (downsampled MNIST [17] 7×7, $B = 2048$), 60-epoch densities all land near $\rho = 0.19$ – 0.22 despite wide accuracy gaps.

Caveat: Adam [18] already hits the dense accuracy ceiling here, so B does not bind; under this add-heavy mutation operator, density drifts to 0.49–0.56 over 400 epochs.

Neural networks · UCI pruning

Need a target sparsity? Tilt the mutation operator to add weights less often whenever $\rho > \rho_{\text{target}}$:

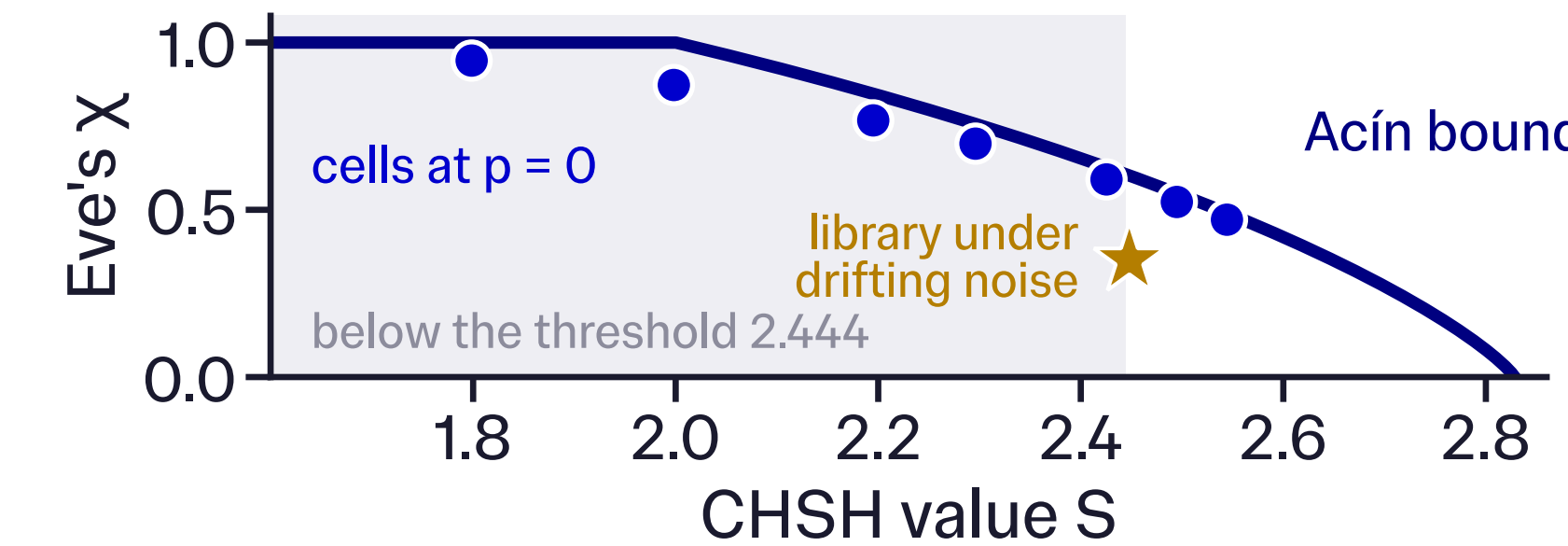


Quantum cryptography · BB84 & E91

Do quantum circuits stay small too?

An eavesdropper (**Eve**) evolves circuits to maximise stolen key info (F_{AE}, χ) at a fixed disturbance (F_{AB}, S), with **no gate penalty**.

BB84 [19, 20]: Evolves compact **4–6 gate** circuits reaching **98–99%** of the theoretical cloning bound [21, 22] at low noise ($p = 0.02$; random-selection control: 5–6 gates).



E91 [23]: 26 circuits (**5–9 gates**, median 6). At zero noise ($p = 0$, dots) cells hug the theoretical Acín bound [24] for $S \geq 2.2$; under drifting noise, the library's DP value (★) sits below it.

Reinforcement learning · QKD

What are small circuits good for?

Why compact circuits matter: Over 50-round episodes of drifting noise, our 24–26 circuits form a discrete action set. Filtering out unsafe actions (**feasibility masking** [25]), both a **greedy oracle** and **Rainbow-DQN** [26] nearly match dynamic programming (DP) with **zero detections**:

	BB84	E91
greedy oracle	98.8%	99.1%
Rainbow-DQN	98.8%	98.4%

E91 reference: coarse grid with a finite terminal penalty, detected in 7% of episodes.

References

[1] Real et al., AAAI 2019. [2] Liu et al., ICLR 2019. [3] Han et al., ICLR 2016. [4] Cerezo et al., Nat. Rev. Phys. 3, 2021. [5] Chivilikhin et al., arXiv:2007.04424, 2020. [6] Lee et al., ICLR 2019. [7] Tanaka et al., NeurIPS 2020. [8] Frankle & Carbin, ICLR 2019. [9] Frankle et al., ICML 2020. [10] Frankle et al., ICLR 2021. [11] Fisher, Iris, UCI repository, 1936. [12] Aeberhard & Forina, Wine, UCI repository, 1992. [13] Wolberg et al., Breast Cancer, UCI repository, 1993. [14] Hansen, arXiv:1604.00772, 2016. [15] Hansen & Ostermeier, Evol. Comput. 9, 2001. [16] Spall, IEEE Trans. Autom. Control 37, 1992. [17] LeCun et al., Proc. IEEE 86, 1998. [18] Kingma & Ba, arXiv:1412.6980, 2014. [19] Bennett & Brassard, Proc. IEEE ICCSP, 1984. [20] Decker et al., npj Quantum Inf. 11, 2025. [21] Fuchs et al., Phys. Rev. A 56, 1997. [22] Cerf, Phys. Rev. Lett. 84, 2000. [23] Ekert, Phys. Rev. Lett. 67, 1991. [24] Acín et al., Phys. Rev. Lett. 98, 2007. [25] Huang & Ontañón, FLAIRS-35, 2022. [26] Hessel et al., AAAI 2018. [27] Ros & Hansen, PPSN X, 2008. [28] Loshchilov, Evol. Comput. 25, 2017. [29] Scarani et al., Rev. Mod. Phys. 81, 2009. [30] Mordarski et al., arXiv:2610.01792, 2026. [31] Next: ImageNet scale with linear-memory CMA-ES [27, 28]; larger quantum memories; symbolic regression. [32] Scope: controls without selection stay as small in the applications, so the mechanism is isolated on the teacher–student test. QKD values are constructive lower bounds for a restricted attacker (trusted noise) and say nothing about security proofs [29]; companion study: [30].