# 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 Bondesan**[1]

[1] **Department of Computing, Imperial College London, London SW7 2AZ, UK**

[2] **Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK**

**{marcel.mordarski25, benjamin.gras25, abdelrahman.shehata25, daniel.budina25, r.bondesan}@imperial.ac.uk**

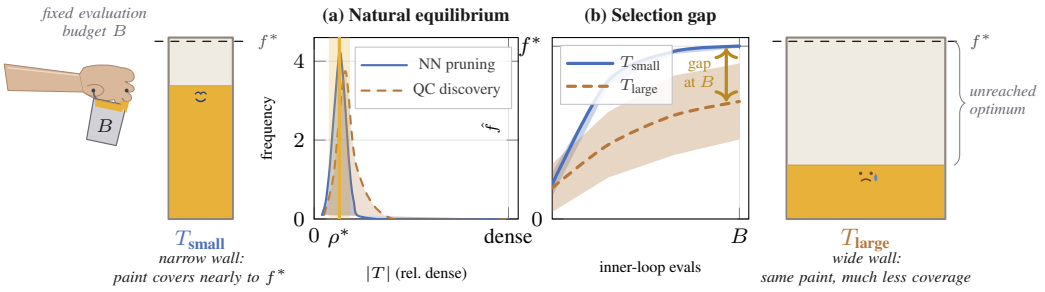
## Abstract

Joint optimisation over a discrete topology $T$ and continuous parameters $\theta$ under an evaluation budget $B$ arises in neural architecture search, pruning, and variational quantum algorithms. A related budget-induced effect, *dimensionality pressure*, is identified here: enlarging $T$ expands the search space of $\theta$, slows inner-loop optimisation at fixed $B$, and yields noisier fitness estimates that systematically bias outer-loop selection toward smaller structures, even without explicit sparsity regularisation. Dimensionality pressure is formalised in a qualitative Proposition composing standard Covariance Matrix Adaptation Evolution Strategy (CMA-ES) dimensionality results with rank-based noisy-comparison selection, predicting a stationary distribution over $|T|$ concentrated below the dense baseline, and as a budget-gated Corollary. To test the mechanism, EVOLUCMAES is introduced – a minimal two-loop framework whose domain-specific components are the topology-mutation operator and the fitness. On a teacher–student benchmark in which surplus weights cost search effort but add no attainable fitness, the equilibrium density stays near $0.09$ up to $B = 2^{11}$ and, from the budget at which independently measured estimator variance collapses, rises monotonically with $B$; on a pruning task, it is reproduced under four inner optimisers, including Adam. In pruning of UCI tabular models, the density ends training at $\rho \approx 0.06$ from an initial $0.05$ under a drift-neutral mutation operator; in eavesdropping on the BB84 quantum key distribution (QKD) protocol, the discovered Eve circuits use 4–6 gates from an arbitrary-depth search space and come within $2\%$ of the noiseless Pauli-channel cloning bound at the lowest bit-flip noise. In both applications, controls without selection stay as small over the same horizons. Adapting only the mutation operator suffices to reach a $95\%$ sparsity target. As a downstream consequence, each library cell is short enough to feed directly to an off-the-shelf RL agent: a single Rainbow-DQN with feasibility masking recovers $98.8\%$ (BB84) and $98.4\%$ (E91/DIQKD) of a dynamic-programming reference over the same library, at zero detection.

*Correspondence to: marcel.mordarski25@imperial.ac.uk

†Equal contribution.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).




*The Natural Equilibrium: where the paint just reaches the top.*

Figure 1: **Dimensionality pressure as paint coverage (schematic).** A fixed dose of paint – the evaluation budget $B$ – is poured equally onto two walls of the same height (the noiseless optimum $f^*$) but different widths (the topology size, i.e. parameter dimension). The narrow wall ($T_{\text{small}}$, left) is covered nearly to the top: the inner CMA-ES loop converges close to $f^*$ within $B$ evaluations. The wide wall ($T_{\text{large}}$, right) receives the same dose, but the paint spreads thin and reaches only a fraction of the way up – the inner loop is still well below $f^*$ at the budget cutoff. The unpainted brace on the right wall is the quantitative gap that drives the outer-loop selection bias against larger topologies; cumulatively, this bias produces equilibrium distributions such as that sketched in panel (a). Panel (b) shows the per-iteration trajectories that sum to this gap. Figure 3 shows measured equilibria.

## 1 Introduction

A wide range of machine-learning problems require joint optimisation over a discrete topology $T$ – which connections exist in a network, which gates appear in a quantum circuit, which operators feature in a symbolic expression – and the continuous parameters $\theta$ that fill it in. Neural architecture search and network pruning, symbolic regression, and the design of variational quantum algorithms [Cerezo et al., 2021] are all instances of this structure. In every case, evaluation is expensive: a single fitness query requires training a network or simulating a quantum circuit. Practical optimisers therefore operate under a strict total-evaluation budget $B$.

This raises the question of what the budget constraint does to the search. Topology growth is essentially free on its own: adding a connection or a gate to $T$ costs nothing per se. However, it enlarges $\dim(\theta)$, and the inner search over good parameters must now share the fixed budget $B$ across more dimensions. Under such a budget, does topology grow without bound, or does the budget itself induce a brake?

**Empirical answer.** Growth stays bounded. Across two unrelated domains – neural-network pruning on UCI tabular benchmarks and quantum-circuit discovery for eavesdropping on the BB84 quantum key distribution (QKD) protocol – a generic two-loop optimiser (evolutionary outer search over $T$, CMA-ES inner search over $\theta$ [Hansen and Ostermeier, 2001]) stays in a low-density/low-gate-count regime with no explicit size penalty. The phenomenon is named here *natural growth bounds* and is attributed to *dimensionality pressure*: larger topologies enlarge the parameter space, slow CMA-ES convergence at fixed budget, and yield noisier fitness estimates that systematically bias outer-loop selection toward smaller structures. Figure 1 pictures the budget as a fixed quantity of paint and the parameter dimension as the width of walls of equal height. A teacher–student benchmark, in which extra parameters add search dimensions without adding attainable fitness, tests the mechanism, and the equilibrium is reproduced under four inner optimisers, including a gradient method (§4.1); in the two applications, controls without selection stay as small (Appendix D), so the mechanism is isolated only on that benchmark. Whether it holds in a third domain (e.g. symbolic regression) remains open (§5).

**The mechanism proposition.** Let $f(T, \theta)$ be the fitness of a given topology $T$ filled in by parameters $\theta \in \mathbb{R}^{\dim(\theta)}$, and let $\hat{f}_B(T)$ denote the best fitness CMA-ES finds for $T$ under an inner-loop

budget of $B$ evaluations. Standard CMA-ES results suggest that $\mathrm{Var}[\hat{f}_B(T)]$ grows as $\dim(\theta)$ grows with $B$ held fixed [Hansen, 2016] (§4.1 weighs this assumption). The observation is that this dimension-dependent fitness noise, propagated through outer-loop selection, suffices to bound the equilibrium $|T|$ *without* an explicit penalty:

> **Proposition (informal; see Proposition 1).** *Under a generic two-loop topology-parameter search with a fixed total-evaluation budget $B$, outer-loop selection bias against high-variance fitness estimates produces a stationary distribution over topology size $|T|$ that is concentrated below the dense baseline $T_{dense}$ (the fully-connected architecture). The location of this equilibrium depends on the rate at which CMA-ES inner-loop variance grows with $\dim(\theta)$ at fixed $B$.*

A budget-gated Corollary 1 adds that the equilibrium is insensitive to $B$ until the estimator variance of denser topologies collapses. These are qualitative observations, and neither yields a closed-form expression for where the equilibrium sits.

**The framework.** To investigate the phenomenon, EvoluCMAES is built as a minimal two-loop framework that (i) treats the topology mutation operator as a domain-specific adapter and (ii) keeps all the other variables – population dynamics, selection, inner CMA-ES – fixed across applications. Across the domains studied here, only the mutation operator and the fitness change; the inner loop is CMA-ES except for the E91 library, which uses SPSA.

**Downstream uses.** When the application requires a specific target density (e.g. $95\%$ sparsity for deployment), the natural-bound regime is too loose, and adapting only the mutation operator, without changing the inner loop or the score function, suffices to hit the target on the UCI suite. The resulting library also transfers to sequential decision-making: under time-varying channel noise, a single Rainbow-DQN agent with feasibility masking matches the greedy-Oracle policy on BB84 and comes within $0.0023$ of it on E91/DIQKD.

**Scope of the empirical study.** The cross-domain evidence is qualitatively consistent and the proposed mechanism is stated in domain-independent terms, but the experimental scale is deliberately small. In practice, this means testing on UCI tabular networks rather than ImageNet-class models, restricting Eve to a one-qubit memory, and running relatively few seeds across variable fitness landscapes. The testbeds have known or controllable optima, so that the mechanism can be measured. The QKD experiments concern a restricted attacker and make no statement about the soundness of QKD security proofs; a companion study treats the cryptographic questions [Mordarski et al., 2026]. Claims that are mechanism-level (and expected, though not shown, to generalise) are distinguished throughout from claims that are run-level (and therefore preliminary).

**Contributions.**

- **An empirical phenomenon (§3.3; Proposition 1).** *Natural growth bounds:* under fixed evaluation budgets, joint topology–parameter optimisation stays at low complexity without any explicit size penalty, in both NN pruning and quantum-circuit discovery. On a teacher–student benchmark that separates search cost from capacity, the equilibrium responds to the budget where independently measured estimator variance collapses, as Corollary 1 predicts, and it is reproduced under four inner optimisers (§4.1).
- **A minimal apparatus (§3).** EvoluCMAES (Algorithm 1), a two-loop framework whose domain-specific components are the random-modification operator and the fitness.
- **Target-density control via mutation alone (§4.2).** Adapting the random-modification operator alone – without modifying the score function or inner loop – hits a $95\%$-sparsity target on the UCI suite.
- **Library transfer to sequential decision-making (§4.4).** Each library cell is a circuit of $4$–$9$ gates, small enough to feed directly to an off-the-shelf RL agent, which recovers $98.8\%$ (BB84) and $98.4\%$ (E91/DIQKD) of a dynamic-programming reference over the same library at zero detection.

## 2 Related Work

**Gradient-free architecture search.** Differentiable NAS methods such as DARTS [Liu et al., 2019] require gradients and rely on biased relaxations. Discrete-search methods such as Regularized Evolution [Real et al., 2019] and CMANAS [Sinha and Chen, 2024] have explored gradient-free topology search but target single domains. EvoluCMAES inherits the evolutionary outer/continuous inner pattern from this literature; the contribution here is not the apparatus but the cross-domain phenomenon the apparatus exposes.

**Neural-network pruning.** Classical magnitude pruning [Han et al., 2016] and the more recent saliency methods SNIP [Lee et al., 2019] and SynFlow [Tanaka et al., 2020] are known to degrade in extreme-sparsity regimes; the Lottery Ticket Hypothesis [Frankle and Carbin, 2019] and Iterative Magnitude Pruning [Frankle et al., 2020] provide alternative training-time strategies. The collapse of one-shot heuristics at very low density has been previously documented [Frankle et al., 2021]; the present work positions an evolutionary alternative that reaches the target density on the UCI suite without a score-level sparsity penalty (Table 2).

**Quantum-circuit learning.** Most variational quantum algorithms [Cerezo et al., 2021] use fixed ansätze. Recent work explores topology discovery (e.g. MoG-VQE [Chivilikhin et al., 2020]), and the present paper follows that thread. For QKD specifically, BB84 [Bennett and Brassard, 1984] and E91 [Ekert, 1991] are the canonical prepare-and-measure and entanglement-based protocols; the device-independent regime is treated by Acín et al. [2007], Pironio et al. [2009]. The Pauli-channel cloning primitive of Cerf [2000] provides the analytical reference; for optimal eavesdropping bounds on BB84 and its six-state extension see Fuchs et al. [1997], Bruß [1998]. Decker et al. [2025] train the angles of a fixed-template circuit attack on BB84 against a fidelity objective; the BB84 experiments here adopt their setting and also search the gate structure. Appendix H gives a short, ML-oriented primer on QKD and a diagram of the BB84 protocol.

**Reinforcement learning under constraints.** The library-transfer extension uses Rainbow-DQN [Hessel et al., 2018] with action masking [Huang and Ontañón, 2022]; PPO [Schulman et al., 2017] and DAgger [Ross et al., 2011] appear as ablation baselines in Appendix K.

## 3 Method: A Minimal Two-Loop Framework

### 3.1 Problem formulation

The objective is to minimise a noisy, gradient-free cost function $f(T, \theta)$ over a discrete topology $T \in \{0, 1\}^n$ (a binary mask, or a sequence of discrete choices) and continuous parameters $\theta \in \mathbb{R}^{\dim(\theta)}$ [Liu et al., 2019, Real et al., 2019], where $\dim(\theta)$ depends on $T$. The fitness estimate returned by a fixed-budget inner loop is stochastic through the randomised inner optimiser and, on the teacher–student benchmark, through injected evaluation noise; the QKD figures of merit are computed from simulated density matrices.

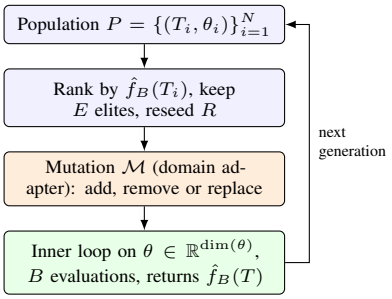
### 3.2 The two-loop framework

EvoluCMAES nests two deliberately minimal loops, ensuring that natural growth bounds emerge from their interaction instead of bespoke machinery.

The *outer loop* maintains a population $P$ of individuals $\{T_i, \theta_i\}$. At each generation, selection is performed by fitness, followed by a random-modification step on the topology. This loop explores the discrete space of architectures [Real et al., 2019]. The *inner loop* runs a fixed-budget ($B$ evaluations) optimisation of the continuous parameters $\theta$ using CMA-ES [Hansen, 2016] for each new or modified topology $T$. CMA-ES is chosen for its ability to handle non-convex, noisy landscapes without requiring gradients [Hansen et al., 2003]. Figure 2 summarises the procedure.

### 3.3 Natural growth bounds

The central empirical claim is that, under the dynamics above with a fixed evaluation budget $B$, no explicit sparsity penalty ($\alpha = 0$) and a mutation operator without net growth bias, topologies do not




**Algorithm 1** EvoluCMAES

**Require:** population size $N$, elites $E$, reseeds $R$, generations $G$, inner budget $B$, initial density $\rho_0$, mutation operator $\mathcal{M}$

$P \leftarrow N$ random topologies at density $\rho_0$
**for** $g = 1, \ldots, G$ **do**
  **for** each new or modified $T \in P$ **do**
    $(\theta_T, \hat{f}_B(T)) \leftarrow \textsc{Inner}(T, B)$ ▷ CMA-ES
  rank $P$ by $\hat{f}_B$ and keep the top $E$ unchanged
  reseed the bottom $R$ at density $\rho_0$
  replace the rest by $\mathcal{M}(T)$, parents $T$ from top ranks
**return** best $(T, \theta_T)$ and the density trace of $P$

Figure 2: The two-loop framework. Left, schematic of one generation; right, Algorithm 1. Only the mutation operator $\mathcal{M}$ and the fitness change between domains, with SPSA [Spall, 1992] as the inner optimiser for the E91 library; on BB84, SPSA and CMA-ES return the same fidelity and gate count at matched budgets (Appendix J). Implementation details of parent selection are given in Appendix D.

grow without bound and the population settles well below the dense baseline. The claim is tested on the teacher–student benchmark of §4.1, which matches the premise of the Proposition below, and it is observed in the two applications.

**Mechanism: dimensionality pressure.** As $\dim(\theta)$ (determined by the size of $T$) increases, the inner search space expands. Under a fixed evaluation budget $B$ and with stochastic fitness, CMA-ES converges more slowly and produces a noisier fitness estimate $\hat{f}_B(T)$ for high-dimensional structures [Hansen, 2016]. The selection pressure in the outer loop, which compares the (noisy) $\hat{f}_B(T)$ values across topologies, thus favours smaller topologies whose inner loops do converge under $B$, provided that the topologies compared share a noiseless optimum. The argument does not refer to the application, but it requires a mutation operator without net growth bias (§3.4) and loses force when the fitness saturates (Appendix E). This effect is what is termed *dimensionality pressure* throughout this paper.

**Proposition 1** (Dimensionality pressure on outer-loop selection)**.** *Fix an inner-loop budget $B$ and let $\hat{f}_B(T)$ denote the best fitness CMA-ES locates for topology $T$ on a noisy stochastic objective. Let $T_1, T_2$ be two topologies with parameter dimensions $d_1 = \dim(\theta_1) < d_2 = \dim(\theta_2)$ and identical noiseless optima $f^*$. Assume* (A1) *the per-iteration progress rate of $(\mu/\mu_w, \lambda)$-CMA-ES (with parent size $\mu$ and offspring size $\lambda$) on the underlying landscape decreases monotonically in $\dim(\theta)$ at fixed $B$ [Hansen, 2016], and* (A2) *the outer-loop selection rule is rank-based with a strict noisy-comparison preference: when $\mathbb{E}[\hat{f}_B(T_1)] = \mathbb{E}[\hat{f}_B(T_2)]$ but $\mathrm{Var}[\hat{f}_B(T_1)] < \mathrm{Var}[\hat{f}_B(T_2)]$, then $\Pr[T_1$ selected$] > \Pr[T_2$ selected$]$. Then:*

1. (Estimator-variance ordering.) *A1 implies that $\mathrm{Var}[\hat{f}_B(T_1)] < \mathrm{Var}[\hat{f}_B(T_2)]$.*
2. (Selection bias.) *Consequently, per A2, the rule assigns strictly higher selection probability to $T_1$ than to $T_2$, conditional on $\mathbb{E}[\hat{f}_B(T_1)] = \mathbb{E}[\hat{f}_B(T_2)]$.*
3. (Stationary distribution.) *Any selection-driven outer-loop dynamic satisfying (A2) admits a stationary distribution $\pi$ over $|T|$ concentrated strictly below the dense maximum $T_{dense}$ (the fully-connected architecture), i.e., $\mathbb{E}_\pi[|T|] < |T|_{dense}$.*

A proof sketch is given in Appendix A. (A1) is standard for textbook CMA-ES on convex-quadratic(-plus-noise) families; (A2) is an assumption on the selection rule, since with equal means and symmetric noise a pairwise comparison picks each candidate half the time and elite truncation can favour the noisier one. (A1) also orders the means (Appendix A). Item 3 alone does not show that selection acts; the nulls of Appendix D test this.

The Proposition predicts *that* the equilibrium $|T|$ is bounded below the dense maximum; it does not predict where. The dependence on $B$ is the natural follow-on:

**Corollary 1** (Budget-gated equilibrium, qualitative)**.** *Under the assumptions of Proposition 1, let $B^\star(\rho)$ denote the budget at which $\mathrm{Var}[\hat{f}_B]$ for topologies of density $\rho$ falls below the fitness separation that selection must resolve. For a pair of densities $\rho_1 < \rho_2$ and budgets $B < B^\star(\rho_2)$, selection cannot*

*order the pair, and the stationary distribution over $|T|$ is insensitive to $B$. For $B > B^\star(\rho_2)$, the stationary distribution shifts monotonically toward the denser topology as $B$ grows. The equilibrium lies at the largest density whose estimator variance has collapsed at the operating budget, and in the limit $B \to \infty$ the selection bias against denser topologies disappears.*

Corollary 1 converts the Proposition into a falsifiable empirical prediction: $B^\star(\rho)$ can be measured without an outer loop, from cold-start inner runs at fixed density, and the population density should begin to respond to $B$ at these budgets (§4.1). The fitness separation entering $B^\star$ is a property of the task, set on the teacher–student benchmark by the scale of the evaluation noise.

**Cross-domain observations.** The bound is qualitatively consistent across both domains:

- **Neural-network pruning** (UCI tabular, §4.2): with the score penalty $\alpha = 0$ and a drift-neutral mutation operator, the population ends training at $\rho \approx 0.06$ from its initial $0.05$ across Iris, Wine, and Cancer (Table 2), and at $\rho \approx 0.12$–$0.16$ under a density-agnostic operator, whose net growth bias selection amplifies (§3.4). Since accuracy saturates on these datasets, these values capture the density at the budget cut-off, prior to any stationary equilibrium (Appendix E).
- **Quantum-circuit discovery** (BB84 cloning under bit-flip noise, §4.3): the discovered Eve circuits use $4$–$6$ gates from a search space that admits arbitrary depth. The analytical Pauli-channel cloning machine (PCCM) requires $4$ gates, so the equilibrium sits one or two gates above the analytical optimum.

Both sizes are bounded but not zero. Over the same horizons, however, the selection-off nulls stay as small ($\rho = 0.049$–$0.051$ for the drift-neutral UCI operator, 5–6 gates for the QKD operators; Appendix D), so the mechanism is isolated only on the teacher–student benchmark (§4.1). When an application requires a specific density, the operator-level adaptation of §4.2 is used instead.

### 3.4 Target-density control via the mutation operator alone

For applications that require a specific density (e.g. $95\%$ sparsity for deployment), the natural-bound regime of §3.3 is too loose. Target control is obtained by adapting only the random-modification operator. The selection process and inner CMA-ES loop remain fixed across domains, leaving the operator as the only change.

In the pruning case, the adaptation is a density-driven mutation: if $\rho(T) > \rho_{\text{target}}$, the probability of removing a connection is increased; if $\rho(T) < \rho_{\text{target}}$, the probability of adding one is increased. Concretely, the density-agnostic swap/remove/add primitive is replaced with an add/remove primitive whose add-probability, $\text{clip}(0.5 - 4(\rho(T) - \rho_{\text{target}}), 0.05, 0.95)$, decreases with the current density. The ablation in Table 2 confirms that the operator alone, without any score-level sparsity penalty, suffices to hit $\rho_{\text{target}} = 0.05$.

Without a target, the operator needs to be size-neutral, since selection will amplify any drift in the proposals that improves fitness. The density-agnostic operator (swap/remove/add with probabilities $0.40/0.20/0.40$) favours growth, and on a downsampled MNIST task a drift-neutral operator (swap/remove/add $0.40/0.30/0.30$) lowers the end-of-run density from about $0.5$ to $0.19$–$0.23$, against $0.10$–$0.11$ in a selection-off null (Appendix D).

## 4 Experiments

### 4.1 Controlled test of the mechanism

A teacher–student benchmark realises the premise of Proposition 1 by construction. A randomly initialised $49 \to 16 \to 10$ teacher network is frozen with a random $10\%$ of its 970 parameters active. Students share the architecture, evolve masks in the outer loop, and are scored by the mean squared error to the teacher's outputs on random inputs, with fresh Gaussian noise at every evaluation, so the fitness never saturates. A mask covering the teacher's active weights reproduces it, and further weights add search dimensions but no attainable fitness. A clean error from one noiseless evaluation is used for reporting only. Budgets $B = 2^7, \ldots, 2^{14}$ (and $2^{15}$ for $\rho_0 = 0.10$) are run from initial

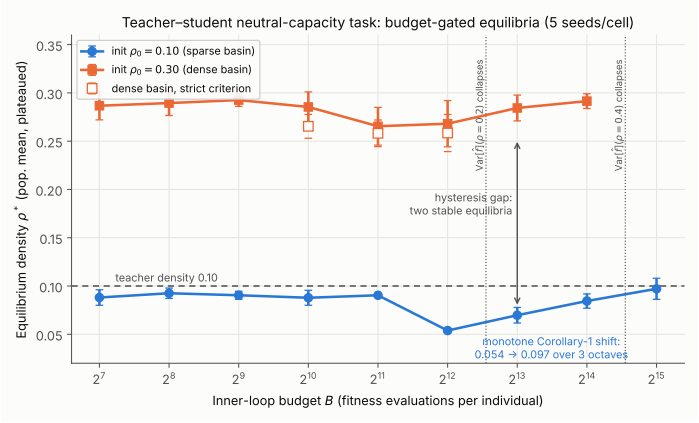



Figure 3: Teacher–student benchmark: population-mean equilibrium density $\rho^\star$ against the inner-loop budget $B$ for two initial densities (five seeds per cell; error bars $\pm 1$ standard deviation over seeds). Open squares, dense-basin re-runs under a stricter plateau criterion (seeds that did not optimise omitted); dashed line, teacher density; dotted lines, variance-collapse budgets for $\rho = 0.2$ and $0.4$ read from Figure 4(b) ($B \approx 6{,}000$ and $24{,}000$). For $B \leqslant 2^9$, most dense-initialised seeds do not optimise (Appendix B).

densities $\rho_0 = 0.10$ and $0.30$, with five seeds per cell, until the population-mean density stops changing (Appendix B).

Sparse-initialised populations plateau at $\rho^\star \approx 0.09$ for $B = 2^7$ to $2^{11}$, fall to $0.054$ at $B = 2^{12}$, and then rise monotonically to $0.097$ at $B = 2^{15}$ (Figure 3), while the best individual reaches a clean error near $0.003$ at every budget, so the plateau is an equilibrium of a search that succeeds. Cold-start inner runs at fixed density without an outer loop (50 per cell, 2,700 in total) locate the collapse of estimator variance at $B \approx 2^{12}$–$2^{14}$ for $\rho = 0.2$ and $B \approx 2^{15}$ for $\rho = 0.4$, where the population trace turns and the dense equilibrium begins to persist. The response is thus flat below the measured threshold and monotone above it, as Corollary 1 states. Populations initialised at $\rho_0 = 0.30$ settle at a second equilibrium, $\rho^\star \approx 0.26$–$0.29$; the dynamics are hysteretic, but both equilibria lie far below the dense value of 1. The selection-off nulls at these initial densities, $0.102$ and $0.300$ (Appendix D), lie $0.005$–$0.05$ above both basins.

On MNIST $7 \times 7$ pruning at $B = 2048$, with three seeds and a meter that charges one unit per fitness evaluation whether or not a gradient is returned, the equilibrium density is $0.215$ (CMA-ES), $0.186$ (random search), $0.191$ (finite-difference SGD) and $0.208$ (Adam [Kingma and Ba, 2014]), against a dense value of 1 (Appendix C). Adam reaches the dense-model accuracy at this budget and still does at $B = 256$ ($\rho^\star = 0.222 \pm 0.010$, accuracy $0.842 \pm 0.006$), so these experiments show the equilibrium under a gradient method without probing it under budget pressure.

### 4.2 Neural-network pruning at extreme sparsity

EvoluCMAES is evaluated on three UCI datasets (Iris [Fisher, 1936], Wine [Aeberhard and Forina, 1992], Breast Cancer [Wolberg et al., 1993]) at a target density of $0.05$ (i.e. $95\%$ sparsity), each split once into $70\%$ for training and $30\%$ held out. The biased-mutation strategy is compared against five baselines: magnitude pruning [Han et al., 2016], iterative magnitude pruning [Frankle et al., 2020], SNIP [Lee et al., 2019], SynFlow [Tanaka et al., 2020], and a random-mask control. Results are means of five seeds (`NNpruning_experiment_2.ipynb`). Because the outer loop of EvoluCMAES selects on the held-out split used for Table 1 (which the baselines never touch), its reported accuracies are biased upwards.

Despite achieving the highest single-dataset score on Iris ($0.938$), magnitude pruning collapses on Wine ($0.411$), reflecting a known failure mode of one-shot heuristics under extreme sparsity [Frankle et al., 2021]. SNIP and SynFlow tie at mean $0.946$ and degrade on Iris ($0.92$ and $0.90$ respectively); IMP is more stable but still suffers a Wine drop ($0.885$). EvoluCMAES reaches the target density on every dataset, but its selection on the held-out split prevents a ranking against the baselines. It also

| Method | Iris | Wine | Cancer | Mean | Density |
|---|---|---|---|---|---|
| Random | $0.449 \pm 0.231$ | $0.659 \pm 0.206$ | $0.758 \pm 0.162$ | 0.622 | 0.053 |
| Magnitude [Han et al., 2016] | $0.938 \pm 0.022$ | $0.411 \pm 0.044$ | $0.958 \pm 0.009$ | 0.769 | 0.050 |
| IMP [Frankle et al., 2020] | $0.911 \pm 0.020$ | $0.885 \pm 0.095$ | $0.951 \pm 0.009$ | 0.916 | 0.050 |
| SNIP [Lee et al., 2019] | $0.920 \pm 0.011$ | $0.959 \pm 0.032$ | $0.959 \pm 0.022$ | 0.946 | 0.050 |
| SynFlow [Tanaka et al., 2020] | $0.902 \pm 0.011$ | $0.970 \pm 0.009$ | $0.966 \pm 0.008$ | 0.946 | 0.050 |
| EvoluCMAES (biased)* | $0.933 \pm 0.000$ | $0.996 \pm 0.007$ | $0.986 \pm 0.003$ | 0.972 | 0.051 |

Table 1: Accuracy on the held-out $30\%$ split at $95\%$ sparsity (mean $\pm$ std over five seeds, target density $\rho = 0.05$). EvoluCMAES (Adam inner loop, biased mutation, score penalty $500\,(\rho - 0.05)^2$) versus pruning baselines on three UCI datasets; the Adam inner loop runs 80 epochs on the training split. *EvoluCMAES selects individuals by their accuracy on this split, which the baselines do not use, so its row is biased upwards and no entry is marked as best. Source: `nnpruning2_results.pkl`.

costs more per solution than a single-pass saliency method, since every candidate mask is trained in an inner loop: a run takes $19.5$–$30.8$ s on eight worker processes, against $0.1$–$0.6$ s per baseline on one. On four further tabular datasets, with selection on a validation split, its mean test accuracy is $0.792$, against $0.789$–$0.814$ for the four pruning baselines (Appendix G). These results serve to validate the apparatus across domains, without claiming to benchmark the pruning methods themselves (§5).

**Mutation-operator ablation.** The role of mutation versus score-level penalty is isolated in Table 2: with $\alpha = 0$ (no score-level penalty), a CMA-ES inner loop and *biased* mutation, the framework reaches $\rho = 0.052$–$0.057$ against $\rho_{\text{target}} = 0.05$, so the score-level penalty is not needed for density control. With $\alpha = 0$ and the drift-neutral operator, the framework ends at $\rho = 0.055$–$0.067$ from $0.05$, against $0.049$–$0.051$ for its selection-off null; the density-agnostic operator ends at $\rho \approx 0.12$–$0.16$ against a null of $0.063$–$0.065$, so most of its rise comes from its growth bias, which selection amplifies (Appendices D and E). All runs keep $\geqslant 95\%$ accuracy on the held-out split, which selection also uses.

| Mutation | Dataset | Test accuracy | Final density $\rho$ |
|---|---|---|---|
| **Biased** ($\rho_{\text{target}} = 0.05$) | Iris | $0.978 \pm 0.014$ | $0.057 \pm 0.002$ |
| | Wine | $1.000 \pm 0.000$ | $0.055 \pm 0.004$ |
| | Cancer | $0.979 \pm 0.006$ | $0.052 \pm 0.002$ |
| **Unbiased** (drift-neutral) | Iris | $0.987 \pm 0.011$ | $0.066 \pm 0.008$ |
| | Wine | $0.996 \pm 0.007$ | $0.067 \pm 0.007$ |
| | Cancer | $0.982 \pm 0.004$ | $0.055 \pm 0.004$ |

Table 2: Mutation-operator ablation under $\alpha = 0$ (no score penalty); mean $\pm$ std over five seeds, CMA-ES inner loop (150 evaluations), initial density $0.05$. *Biased* mutation alone hits $\rho_{\text{target}} = 0.05$; *unbiased* mutation (drift-neutral, swap/remove/add $0.40/0.30/0.30$) does not target a value. Under the density-agnostic operator ($0.40/0.20/0.40$), the unbiased runs end at $\rho = 0.153 \pm 0.032$, $0.120 \pm 0.010$ and $0.156 \pm 0.010$, with accuracy $0.996$, $1.000$ and $0.991$ (Appendix E). Accuracy is measured on the held-out split that selection also uses. Sources: `NNpruning_experiment_3.ipynb` (biased), `mutation_operator_ablation_uci.py` (drift-neutral), `NNpruning_experiment_4.ipynb` (density-agnostic).

### 4.3 Quantum-circuit discovery for QKD eavesdropping

The same framework, with the random-modification operator now acting on a discrete gate alphabet $\{R_X, R_Y, R_Z, CNOT\}$ [Chivilikhin et al., 2020] with add/remove/replace primitives, is run on the problem of discovering an Eve circuit for BB84 eavesdropping. The fitness follows Decker et al. [2025]: Eve's fidelity $F_{AE}$ with a quadratic penalty that pins Bob's fidelity $F_{AB}$ to a prescribed value (E91: Appendix K.4), with no gate-count penalty. Because Eve cannot alter the physical channel noise and holds only a one-qubit memory, each value is at most a constructive lower bound for a restricted attacker, leaving the soundness of security proofs (which attribute all disturbance to Eve [Scarani et al., 2009]) intact. Appendix H gives a one-page QKD primer.

**BB84 cloning.** Following Decker et al. [2025], a bit flip of rate $p$ acts on the travelling qubit before Eve's interaction, so it also degrades her copy; Eve's two-qubit unitary acts on the travelling qubit and one ancilla, which she measures in the announced basis (an individual attack); and $F_{AB}$, averaged over both bases, equals $1 - p/2$ without Eve. The framework finds circuits with 4–6 gates. At $p = 0.02$, they reach 98.0–99.0% of the noiseless Pauli-channel cloning value $F_{AE} = \frac{1}{2} + \sqrt{F_{AB}(1 - F_{AB})}$ [Fuchs et al., 1997, Cerf, 2000] at their own $F_{AB}$, and at $p = 0.25$, 70–87% of it, since the bit flip degrades Eve's copy. The abort threshold $F_{AB} \geqslant 0.87$, a basis-averaged error rate of 13% that lies above the 11% at which the one-way key rate of BB84 vanishes [Shor and Preskill, 2000], is a modelling choice for this testbed. The selection-off null stays at 5–6 gates (Appendix D); this size is already deployment-relevant, so no target-control step is needed.

**Cross-protocol generalisation to E91/DIQKD.** The same library framework ports to E91 with three adjustments: (i) the channel becomes *bilateral depolarising* $\Lambda_p^A \otimes \Lambda_p^B$ on the entangled pair, acting after Eve's three-qubit preparation; (ii) the public observable shifts to the CHSH violation $S_{\text{obs}}$ with detection threshold $S_{\text{sec}} + 0.02 \approx 2.444$; and (iii) Eve's information is measured by the Holevo quantity $\chi(A : E)$ for collective attacks. Both local channels contribute: $\Lambda_A$ damps $\chi$ via data-processing on Alice's measured qubit, while $\Lambda_B$ leaves $\chi$ untouched but tightens Eve's CHSH budget (Appendix K). Gate-set mutation over $\{R_X, R_Y, R_Z, H, CNOT\}$ with an SPSA inner loop yields a 26-cell library (5–9 gates each), where the library DP value $\chi \approx 0.350$ agrees with a numerical Pareto search over the full three-qubit pre-channel state (Appendix K.7). Since the attacker's dimension remains fixed and only $S_{\text{obs}}$ is constrained, these figures hold only within a restricted, fixed-dimension regime, prior to any device-independent guarantee.

### 4.4 Library transfer to sequential decision-making

The natural-growth-bound regime has a downstream consequence beyond static optimisation. Each library cell is a discovered circuit of 4–6 (BB84) or 5–9 (E91) gates rather than an arbitrary-depth one, so an RL agent navigating the library (24 cells for BB84, 26 for E91) faces a discrete action set of pre-built circuits small enough to enumerate without compression. This property is exploited here to test whether an off-the-shelf RL agent, given the EvoluCMAES library as its only action set, can approach a dynamic-programming (DP) reference on a constrained Markov decision process where the channel noise varies over time.

Concretely, the channel parameter $p$ follows an Ornstein–Uhlenbeck (OU) drift that Eve observes without error, an idealisation that favours her; she picks one library cell per round to maximise her information measure, and detection is declared if the mean of the monitored statistic over a 50-round episode falls below its threshold. A single Rainbow-DQN agent [Hessel et al., 2018] with feasibility masking [Huang and Ontañón, 2022] and matched hyperparameters across protocols recovers 98.8% (BB84) and 98.4% (E91) of the DP reward at zero detection (Table 3). With its coarse grid and finite terminal penalty, the E91 DP policy provides only a baseline reference, prior to any certified upper bound. On BB84, a tolerated-detection variant exceeds the greedy Oracle on the same 50 episodes by 0.0047 (paired 95% interval $\pm 0.0011$; sign test $p \approx 4 \times 10^{-10}$) at a 4% detection rate. However, because the DP reference reaches 0.753 at zero detection, this advantage can already be achieved without risking detection. Across 200 E91 episodes with shared channel realisations, the agent trails the Oracle by 0.0023 and exceeds MaskablePPO by 0.0017, both significant under a Wilcoxon signed-rank test (Table 8). Full ablations and additional agents appear in Appendix I.

## 5 Discussion and Limitations

**Takeaway.** The argument of this work is that sparsity in topology–parameter search has been treated, in both NAS/pruning and BB84/E91, as a hyperparameter to be controlled via penalty terms, target densities, and training-time schedules. The empirical phenomenon documented here is that, in the fixed-evaluation-budget regimes studied with a noisy stochastic objective, sparsity is a property the optimiser already maintains *for free*, in two domains whose fitness landscapes are unrelated, and Proposition 1 gives a qualitative account of how selection contributes, which the teacher–student benchmark supports. The implication is that "how sparse should it be?" is the wrong question for fixed-budget search; *"what is the natural equilibrium under your evaluation budget, and how do you adjust it without losing the bound?"* is the right one. The 95%-sparsity result of §4.2 is one answer to the latter, and the library transfer of §4.4 is a downstream consequence.

| Protocol | Method | $F_{AE}/\chi$ | det % | % of DP |
|---|---|---|---|---|
| BB84 | Identity | 0.5000 | 0.0 | 66% |
| | Static safe (noise-indexed lookup)$^\dagger$ | 0.7197 | 0.0 | 96% |
| | Oracle (greedy) | 0.7434 | 0.0 | 99% |
| | Rainbow-DQN, strict ($\kappa = 2.0$) | 0.7434 | 0.0 | 99% |
| | Rainbow-DQN, tolerated ($\kappa = 0.5$)$^\dagger$ | **0.7477** | 4.0 | 99% |
| | DP reference | 0.7527 | 0.0 | 100% |
| E91 | Identity | 0.000 | 0.0 | 0% |
| | Oracle (greedy) | 0.3507 | 0.0 | 99% |
| | Rainbow-DQN | 0.3484 | 0.0 | 98% |
| | DP reference (coarse grid) | 0.3540 | 7.0 | 100% |

Table 3: Rainbow-DQN on the EvoluCMAES library across both protocols (best checkpoint). Evaluation over 50 (BB84) and 200 (E91) episodes; per-episode standard deviations are about $0.03$ (BB84) and $0.065$ (E91). "% of DP" divides by the DP reference over the same library (BB84: fine grid, 200 episodes; E91: coarse grid). The analytical envelopes at the threshold, $F_{AE} \leqslant 0.836$ for individual attacks on BB84 and $\chi \leqslant 0.607$ for collective attacks on E91, also bound attacks outside the library. $\kappa$ weights the detection rate in the checkpoint-selection score; bold marks an agent value above greedy Oracle on the same episodes. $^\dagger$Second set of 50 episodes, on which greedy Oracle scores $0.7429$ (Appendix I).

**Limitations.** The natural-growth-bounds claim is empirical at the cross-domain scale. The Proposition of §3.3 establishes that the equilibrium $|T|$ is bounded below the dense baseline under generic CMA-ES dimensionality assumptions, but does not pin down where the equilibrium sits. Corollary 1 makes the budget dependence falsifiable, and the teacher–student benchmark supports it in its budget-gated form (§4.1); the threshold is measured and has no closed form in terms of $B$ and the noise variance, and the dynamics are hysteretic, so the low-density equilibrium is not unique. Changing the evaluation-noise scale and repeating the inner-optimiser comparison at larger scale would test the Corollary and (A1) further.

The pruning experiments are at small scale (UCI tabular). Because the pruning baselines were originally validated on CIFAR/ImageNet-class networks, failure modes like magnitude pruning collapsing on Wine ($0.411$) may simply be tiny-dataset artefacts, without reflecting fundamental limits of the methods. EvoluCMAES has not been validated at CIFAR/ImageNet scale, where full covariance adaptation in CMA-ES, at $\mathcal{O}(d^2)$ memory and cost per update, would need replacing by variants that scale linearly in $d$, such as separable CMA-ES [Ros and Hansen, 2008], VD-CMA [Akimoto et al., 2014] or LM-CMA [Loshchilov, 2017], and the inner-optimiser ablation of §4.1 suggests that the mechanism does not require a full covariance model. UCI and MNIST accuracies are biased upwards by selection on the reporting split (§4.2); the densities are not.

A second caveat applies to the applications. Test accuracy on the UCI datasets saturates at $1.0$ within very few epochs, after which selection no longer favours smaller topologies; density then drifts upward under the density-agnostic operator (Appendix E) and stays near its initial $0.05$ under the drift-neutral one. In both applications the selection-off nulls stay as small (Appendix D), so the mechanism and the location claim are tested on the non-saturating teacher–student benchmark.

Quantum-circuit discovery is demonstrated with a one-qubit Eve memory; larger ancillas, non-Pauli noise and non-QKD VQAs are left to future work. The QKD values are constructive lower bounds for a restricted attacker under the trusted-noise convention, with asymptotic statistics and a noise level known to Eve; security proofs attribute all observed disturbance to the eavesdropper and extend to adaptive strategies [Metger and Renner, 2023]. The BB84 experiments monitor the basis-averaged fidelity and the E91 library constrains the CHSH value alone [Mordarski et al., 2026]. The BB84 Rainbow-DQN result is the best checkpoint of one training run.

## Acknowledgments and Disclosure of Funding

The authors declare no competing interests. In the past three years, MM received a UKRI DLA, USRA Richard Feynman, and OIST Research Internship Program scholarships; AS received University of Leeds International Excellence scholarship; DB received UCL Cost of Living Bursary.

## References


Antonio Acín, Nicolas Brunner, Nicolas Gisin, Serge Massar, Stefano Pironio, and Valerio Scarani. Device-independent security of quantum cryptography against collective attacks. *Physical Review Letters*, 98(23):230501, 2007. doi: 10.1103/PhysRevLett.98.230501.

Stefan Aeberhard and M. Forina. Wine. UCI Machine Learning Repository, 1992.

Youhei Akimoto, Anne Auger, and Nikolaus Hansen. Comparison-based natural gradient optimization in high dimension. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pages 373–380. ACM, 2014. doi: 10.1145/2576768.2598258.

Charles H. Bennett and Gilles Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing*, pages 175–179, Bangalore, India, December 1984.

Peter Brown, Hamza Fawzi, and Omar Fawzi. Device-independent lower bounds on the conditional von Neumann entropy. *Quantum*, 8:1445, 2024. doi: 10.22331/q-2024-08-27-1445. arXiv preprint posted 2021.

Dagmar Bruß. Optimal eavesdropping in quantum cryptography with six states. *Physical Review Letters*, 81(14):3018–3021, 1998. doi: 10.1103/PhysRevLett.81.3018.

M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, 2021. doi: 10.1038/s42254-021-00348-9.

Nicolas J. Cerf. Pauli cloning of a quantum bit. *Physical Review Letters*, 84(19):4497–4500, 2000. doi: 10.1103/PhysRevLett.84.4497.

Małgorzata Charytanowicz, Jerzy Niewczas, Piotr Kulczycki, Piotr Kowalski, and Szymon Łukasik. Seeds. UCI Machine Learning Repository, 2010.

D. Chivilikhin, A. Samarin, V. Ulyantsev, I. Iorsh, A. R. Oganov, and O. Kyriienko. MoG-VQE: Multiobjective genetic variational quantum eigensolver. *arXiv preprint arXiv:2007.04424*, 2020.

Thomas Decker, Marcelin Gallezot, Sven Florian Kerstan, Alessio Paesano, Anke Ginter, and Wadim Wormsbecher. Quantum key distribution as a quantum machine learning task. *npj Quantum Information*, 11(1):140, 2025. doi: 10.1038/s41534-025-01088-9.

Artur K. Ekert. Quantum cryptography based on Bell's theorem. *Physical Review Letters*, 67(6): 661–663, 1991. doi: 10.1103/PhysRevLett.67.661.

R. A. Fisher. Iris. UCI Machine Learning Repository, 1936.

Jonathan Frankle and Michael Carbin. The lottery ticket hypothesis: Finding sparse, trainable neural networks. In *International Conference on Learning Representations (ICLR)*, 2019.

Jonathan Frankle, Gintare Karolina Dziugaite, Daniel M. Roy, and Michael Carbin. Linear mode connectivity and the lottery ticket hypothesis. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, volume 119 of *Proceedings of Machine Learning Research*, pages 3259–3269, 2020.

Jonathan Frankle, Gintare Karolina Dziugaite, Daniel M. Roy, and Michael Carbin. Pruning neural networks at initialization: Why are we missing the mark? In *International Conference on Learning Representations (ICLR)*, 2021.

Christopher A. Fuchs, Nicolas Gisin, Robert B. Griffiths, Chi-Sheng Niu, and Asher Peres. Optimal eavesdropping in quantum cryptography. I. information bound and optimal strategy. *Physical Review A*, 56(2):1163–1172, 1997. doi: 10.1103/PhysRevA.56.1163.

B. German. Glass identification. UCI Machine Learning Repository, 1987.

Song Han, Huizi Mao, and William J. Dally. Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding. In *International Conference on Learning Representations (ICLR)*, 2016.

Nikolaus Hansen. The CMA evolution strategy: A tutorial. *arXiv preprint arXiv:1604.00772*, 2016.

Nikolaus Hansen and Andreas Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001. doi: 10.1162/106365601750190398.

Nikolaus Hansen, Sibylle D. Müller, and Petros Koumoutsakos. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary Computation*, 11(1):1–18, 2003. doi: 10.1162/106365603321828970.

Carl W Helstrom. Quantum detection and estimation theory. *Journal of Statistical Physics*, 1(2): 231–252, 1969.

Matteo Hessel, Joseph Modayil, Hado van Hasselt, Tom Schaul, Georg Ostrovski, Will Dabney, Daniel Horgan, Bilal Piot, Mohammad Azar, and David Silver. Rainbow: Combining improvements in deep reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018. doi: 10.1609/aaai.v32i1.11796.

Shengyi Huang and Santiago Ontañón. A closer look at invalid action masking in policy gradient algorithms. In *Proceedings of the Thirty-Fifth International Florida Artificial Intelligence Research Society Conference (FLAIRS-35)*, volume 35, 2022. doi: 10.32473/flairs.v35i.130584.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. doi: 10.1109/5.726791.

Namhoon Lee, Thalaiyasingam Ajanthan, and Philip H.S. Torr. SNIP: Single-shot network pruning based on connection sensitivity. In *International Conference on Learning Representations (ICLR)*, 2019.

Hanxiao Liu, Karen Simonyan, and Yiming Yang. DARTS: Differentiable architecture search. In *International Conference on Learning Representations (ICLR)*, 2019.

Ilya Loshchilov. LM-CMA: An alternative to L-BFGS for large-scale black box optimization. *Evolutionary Computation*, 25(1):143–171, 2017.

Tony Metger and Renato Renner. Security of quantum key distribution from generalised entropy accumulation. *Nature Communications*, 14:5272, 2023. doi: 10.1038/s41467-023-40920-8.

Marcel Mordarski, Benjamin Gras, Abdelrahman Shehata, Daniel Budina, and Roberto Bondesan. Learnt attacks on quantum key distribution under channel noise and device drift. *arXiv preprint arXiv:2610.01792*, 2026.

Miguel Navascués, Stefano Pironio, and Antonio Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(7):073013, 2008. doi: 10.1088/1367-2630/10/7/073013.

Stefano Pironio, Antonio Acín, Nicolas Brunner, Nicolas Gisin, Serge Massar, and Valerio Scarani. Device-independent quantum key distribution secure against collective attacks. *New Journal of Physics*, 11(4):045021, 2009. doi: 10.1088/1367-2630/11/4/045021.

Esteban Real, Alok Aggarwal, Yanping Huang, and Quoc V. Le. Regularized evolution for image classifier architecture search. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 4780–4789, 2019. doi: 10.1609/aaai.v33i01.33014780.

Raymond Ros and Nikolaus Hansen. A simple modification in CMA-ES achieving linear time and space complexity. In *Parallel Problem Solving from Nature – PPSN X*, volume 5199 of *Lecture Notes in Computer Science*, pages 296–305. Springer, 2008.

Stéphane Ross, Geoffrey J. Gordon, and J. Andrew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 15 of *Proceedings of Machine Learning Research*, pages 627–635. PMLR, 2011.

Valerio Scarani, Helle Bechmann-Pasquinucci, Nicolas J. Cerf, Miloslav Dušek, Norbert Lütkenhaus, and Momtchil Peev. The security of practical quantum key distribution. *Reviews of Modern Physics*, 81:1301–1350, 2009. doi: 10.1103/revmodphys.81.1301.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Terry Sejnowski and R. Gorman. Connectionist bench (Sonar, mines vs. rocks). UCI Machine Learning Repository, 1988.

Peter W. Shor and John Preskill. Simple proof of security of the BB84 quantum key distribution protocol. *Physical Review Letters*, 85(2):441–444, 2000. doi: 10.1103/PhysRevLett.85.441.

V. Sigillito, S. Wing, L. Hutton, and K. Baker. Ionosphere. UCI Machine Learning Repository, 1989.

Nilotpal Sinha and Kuan-Wen Chen. Neural architecture search using covariance matrix adaptation evolution strategy. *Evolutionary Computation*, 32(2):177–204, 06 2024. ISSN 1063-6560. doi: 10.1162/evco_a_00331. URL `https://doi.org/10.1162/evco_a_00331`.

James C Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control*, 37(3):332–341, 1992.

Hidenori Tanaka, Daniel Kunin, Daniel L.K. Yamins, and Surya Ganguli. Pruning neural networks without any data by iteratively conserving synaptic flow. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 6377–6389, 2020.

Joaquin Vanschoren, Jan N. van Rijn, Bernd Bischl, and Luis Torgo. OpenML: Networked science in machine learning. *ACM SIGKDD Explorations Newsletter*, 15(2):49–60, 2014. doi: 10.1145/2641190.2641198.

Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.

William Wolberg, Olvi Mangasarian, Nick Street, and W. Street. Breast cancer Wisconsin (Diagnostic). UCI Machine Learning Repository, 1993.

# A Proof sketch for Proposition 1

This appendix provides a heuristic proof sketch of Proposition 1 and Corollary 1. Serving as a consistency check for the proposed mechanism, the argument falls short of a formal proof and instead composes two standard intuitions: dimension-dependent convergence in CMA-ES and selection under noisy fitness estimates. Throughout, $f^*$ denotes the noiseless optimum shared by $T_1$ and $T_2$.

**Step 1: estimator-variance ordering (heuristic).** Under assumption *(A1)*, the per-iteration progress rate $\varphi_d$ of $(\mu/\mu_w, \lambda)$-CMA-ES decreases with dimension $d$. On convex-quadratic landscapes, bounds of the form $\varphi_d = \mathcal{O}(1/d)$ are known [Hansen et al., 2003], with analogous scaling under additive Gaussian noise [Hansen, 2016].

After a fixed budget $B$ of evaluations, higher-dimensional problems therefore remain further from the optimum in expectation; under stochastic fitness, this yields noisier estimates $\hat{f}_B(T)$ at larger $d$. Decomposing the estimator variance into (i) a residual-gap term governed by convergence speed and (ii) an evaluation-noise term captures this directly: for budgets below the convergence horizon of the higher-dimensional topology, the residual-gap term dominates and grows with $d$. This motivates the ordering

$$\mathrm{Var}[\hat{f}_B(T_2)] \gtrsim \mathrm{Var}[\hat{f}_B(T_1)] \quad \text{for } d_2 > d_1 \text{ at fixed } B, \tag{1}$$

which serves here as a modelling assumption justified by known CMA-ES scaling behaviour, as opposed to a general theorem. The strict inequality need not hold uniformly across all regimes (e.g. near convergence or under different noise structures).

**Step 2: noisy-comparison selection bias (informal).** Under assumption *(A2)*, the outer-loop selection rule compares candidates via their noisy fitness estimates $\hat{f}_B(T)$, as in standard tournament or rank-based schemes where selection probability depends on pairwise comparisons.

When two candidates share a comparable expected fitness, lower estimator variance produces more reliable rankings, yet symmetric noise models (e.g. Gaussian perturbations) do not automatically favour the lower-variance candidate: a single pairwise comparison between equal means selects either candidate with probability $1/2$, and elite truncation can actually favour the noisier candidate because its estimate more often crosses a high threshold. Assumption *(A2)* therefore enters directly as a condition on the selection rule. By contrast, Step 1 alone induces a bias in the mean even without *(A2)*: within $B$ evaluations, the higher-dimensional topology ends further from its optimum, lowering its expected estimate whenever this shortfall exceeds any gain in its true optimum and causing rank-based selection to favour the lower-dimensional topology.

Combining assumption *(A2)* with Step 1 yields a qualitative selection bias towards lower-dimensional topologies under limited evaluation budgets.

**Step 3: induced bias on topology size (qualitative).** The outer loop defines a stochastic process over topology size $|T|$, driven by mutation proposals and noisy selection. The sketch assumes a drift-neutral mutation operator with equal addition and removal probabilities so that proposals do not favour any size, whereas the density-agnostic operator used in the UCI and MNIST runs (swap/remove/add at $0.40/0.20/0.40$) carries a net growth bias, which Appendix D measures against a selection-off null. Against neutral proposals, the selection bias from Step 2 acts in the opposite direction: under fixed budgets, larger topologies incur higher-variance fitness estimates and are selected less often when competing against smaller candidates of similar expected fitness.

Heuristically, this induces a negative drift in $|T|$ above a scale set by the evaluation budget and noise level, concentrating the stationary distribution over $|T|$ away from the maximal (dense) topology. While the drift conditions required for a formal Foster–Lyapunov argument are left unverified here, the combined mutation–selection dynamics remain consistent with a bounded equilibrium. Because a mean below the dense maximum holds for almost any stationary distribution, including a drift-neutral walk without selection. This step alone does not demonstrate that selection is active; Appendix D isolates the selection effect by comparing the evolved populations against selection-off nulls.

**Corollary 1 (qualitative).** Increasing the evaluation budget $B$ sharpens the inner-loop estimates $\hat{f}_B(T)$. Once $B$ exceeds a density-dependent threshold $B^\star(\rho)$ (which grows with $\rho$), the estimator variance for topologies of density $\rho$ falls below the fitness separation that selection must resolve.

For a pair of densities $\rho_1 < \rho_2$, budgets $B < B^\star(\rho_2)$ leave the denser topology dominated by variance, preventing selection from ordering the pair by expected fitness and leaving the dynamics insensitive to moderate changes in $B$. Once $B > B^\star(\rho_2)$, the denser topology is estimated reliably; raising $B$ further brings still denser topologies below their thresholds and shifts the equilibrium towards denser structures. In the limit $B \to \infty$, all topologies are evaluated near their optima and the variance-induced selection bias vanishes.

This yields a falsifiable empirical prediction: population density should respond to $B$ at the thresholds $B^\star(\rho)$ (which can be measured independently without an outer loop) as tested on the teacher–student benchmark in §4.1.

**Scope of the sketch.** The argument above establishes the plausibility of the dimensionality-pressure mechanism under standard assumptions, without constituting a formal proof. Specifically, it leaves open (i) deriving a general variance–dimension law for CMA-ES under arbitrary noise, (ii) characterising selection bias across all rank-based schemes, (iii) rigorously proving the existence, uniqueness, or location of the stationary distribution (the teacher–student benchmark exhibits two stable equilibria selected by initialisation), and (iv) obtaining a closed form for the equilibrium density as a function of $B$ and the noise variance. These formal questions are deferred to future work, leaving the primary validation of the mechanism in this paper empirical.

## B Teacher–student benchmark: protocol and diagnostics

The teacher uses the $49 \to 16 \to 10$ architecture from the MNIST experiments (Appendix F) with 970 parameters, where a random $10\%$ mask leaves 95 weights active and fixes the remainder at zero. The teacher retains its random initial weights (PyTorch default initialisation) and remains frozen along with its mask. Students share this architecture and evolve their masks under the outer loop of Algorithm 1, while the inner CMA-ES optimises the active weights. Student fitness is the negative mean squared error (MSE) between the student and teacher outputs on random inputs, with i.i.d. Gaussian noise ($\sigma = 0.10$) freshly drawn and added to every teacher output at each evaluation. Inputs are drawn i.i.d. from $\mathcal{N}(0, \mathbb{I}_{49})$, partitioned into 2,048 fixed inputs for the inner-loop fitness and a separate set of 1,024 inputs for both the selection score and the clean error. The reported clean error is a single noiseless evaluation of the same MSE on the 1,024-input split and never enters selection. In the outer loop ($N = 30$, $E = 4$, $R = 6$), six additional individuals are replaced by mutated copies of the top three elites, and the remaining 14 are mutated in place using the drift-neutral operator of Appendix D ($\alpha = 0$), paired with an inner CMA-ES with initial step size $0.3$. After a minimum of 120 outer epochs, a run terminates when the least-squares slope of the population-mean density over the last 40 epochs falls below $2 \times 10^{-4}$ per epoch in absolute value, up to a cap of 300 epochs. All 85 standard runs stopped on this slope criterion (72 at epoch 120, and the latest at epoch 173), with the equilibrium density $\rho^\star$ defined as the population-mean density averaged over the final 40 epochs. Repeating the dense-basin runs at $B = 2^{10}$–$2^{12}$ under a stricter threshold of $5 \times 10^{-5}$ (capped at 600 epochs) yielded termination at epochs 120–178.

Figure 4(a) plots the best-individual clean error at equilibrium. Under sparse initialisation, the error lies near $0.003$ across all budgets, confirming that the density plateau below $B = 2^{12}$ in Figure 3 reflects a successful search equilibrium. Under dense initialisation at $B \leqslant 2^9$, 11 of the 15 seeds fail to optimise and remain near their initial error at $0.13$–$0.22$ (medians $0.157$–$0.166$), whereas the other four reach $0.003$–$0.005$. At intermediate budgets, individual seeds similarly either learn or stay near that initial level. Under the stricter stopping criterion, the dense-basin equilibrium persists at $\rho^\star = 0.258$–$0.265$ for the seeds that optimise (the open squares in Figure 3, which omit seeds with final clean error $\geqslant 0.05$) and at $0.266$–$0.277$ across all five seeds. Of the initial densities used elsewhere in this paper ($\rho_0 = 0.05$–$0.10$), only $\rho_0 = 0.10$ was evaluated on this benchmark, falling within the sparse basin.

Figure 4(b) shows the clean error after a single cold-start inner run at fixed density without outer-loop evolution, covering $\rho \in \{0.05, 0.10, 0.20, 0.40, 0.70, 1.00\}$ across budgets $2^7$ to $2^{15}$ with 50 cold starts per cell (forming the 2,700-run grid of §4.1). Spanning nearly five orders of magnitude in $\mathrm{Var}[\hat{f}_B]$, these curves provide the raw form of the variance-collapse thresholds plotted as dotted lines in Figure 3: runs at $\rho \leqslant 0.10$ converge at every tested budget, runs at $\rho = 0.2$ require $B \approx 2^{12}$–$2^{14}$, runs at $\rho = 0.4$ begin to converge only at $B = 2^{15}$, and runs at $\rho \geqslant 0.7$ do not converge by $B = 2^{15}$. Unlike the MNIST convergence horizons in Figure 10(B), which compare densities with different

capacity ceilings, these curves directly test the premise of assumption *(A1)* by evaluating topologies that share a common optimum.

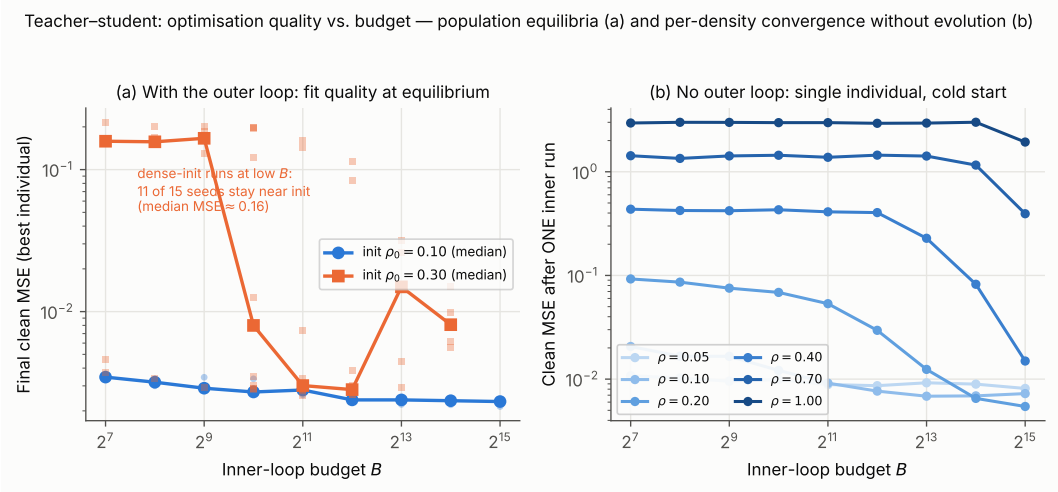



Figure 4: Optimisation quality against budget on the teacher–student benchmark. (a) Best-individual clean error at equilibrium for each initial density. Points show the clean error of the top-ranked individual of each seed at the final epoch (five per cell, and ten for $\rho_0 = 0.30$ at $B = 2^{10}$–$2^{12}$ where the standard and strict runs are pooled), while lines track the medians. (b) Mean clean error after a single cold-start inner run at fixed density without an outer loop (50 cold starts per cell).

## C Inner-optimiser ablation on MNIST pruning

Four inner optimisers were run on the MNIST $7 \times 7$ pruning task of Appendix F at $B = 2048$ across three seeds, keeping the outer loop ($N = 30$, $E = 4$, $R = 6$, $\rho_0 = 0.10$, 60 outer epochs), the density-agnostic mutation operator (swap/remove/add at $0.40/0.20/0.40$), and the selection rule unchanged. Here, $\rho^\star$ denotes the population-mean density over the final five epochs. Under a matched evaluation meter, each fitness evaluation counts as a single unit for every method regardless of whether it also returns a gradient, which advantages Adam [Kingma and Ba, 2014] (configured with a learning rate of $0.01$). As shown in Figure 5, the equilibrium density lies between $0.186$ and $0.215$ across all four optimisers, whereas the best accuracy on the held-out split (also used for outer-loop selection) ranges from $0.544$ (random search) to $0.847$ (Adam, matching the dense-model accuracy of this architecture). Adam also reaches the dense-model accuracy at $B = 256$ ($\rho^\star = 0.222 \pm 0.010$, accuracy $0.842 \pm 0.006$, three seeds), confirming that the evaluation budget does not bind for Adam at either setting.

## D Mutation operators and selection-off nulls

In the UCI and MNIST pruning runs, each mutation flips $2\%$ of the mask entries one at a time, where each flip swaps an active and an inactive entry, removes an active entry, or adds one. The density-agnostic operator draws these actions with probabilities $0.40$, $0.20$, and $0.40$ (increasing the expected number of active entries by $0.2$ per flip), whereas the drift-neutral operator draws them with probabilities $0.40$, $0.30$, and $0.30$ for zero expected change. By contrast, the density-driven operator of §3.4 sets the addition probability to $\mathrm{clip}(0.5 - 4(\rho - \rho_{\text{target}}), 0.05, 0.95)$. At each epoch, the pruning outer loop retains the $E$ best individuals unchanged and replaces the $R$ worst with random masks at the initial density. A further $R$ individuals (drawn from the remaining ranks with probability increasing in their fitness deficit) are replaced by mutated copies of elite parents, while the rest are mutated in place ($N = 100$, $E = 10$, $R = 20$ with an inner budget of 150 evaluations for Table 2; $N = 30$, $E = 5$, $R = 10$ for Table 1). For the E91 library, a gate-level operator replaces, removes, or inserts a single gate, or perturbs one angle, with probabilities $0.35$, $0.20$, $0.25$, and $0.20$, drawing parents with equal probability from the elites or the ranked population.

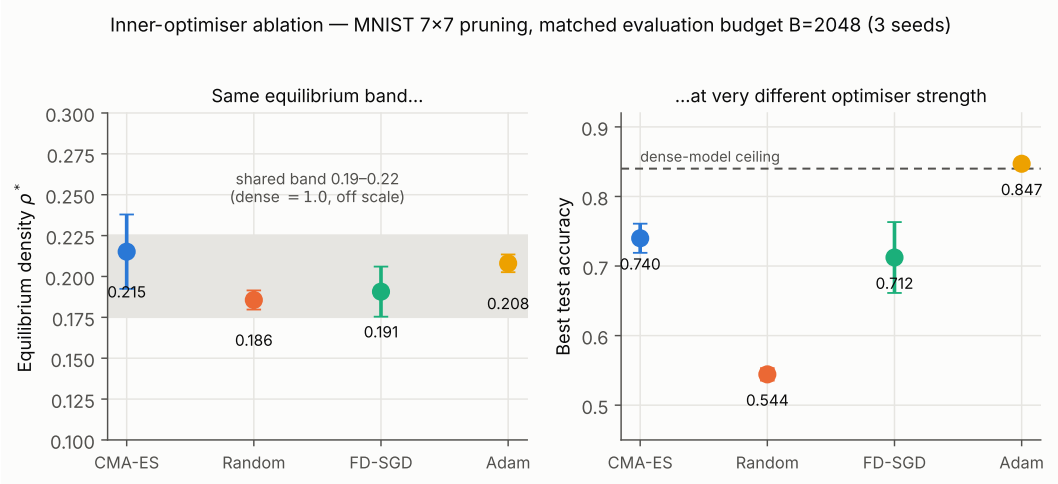



Figure 5: Inner-optimiser ablation on MNIST $7 \times 7$ pruning at a matched evaluation budget $B = 2048$ (three seeds; error bars indicate $\pm 1$ standard deviation). Left: equilibrium density $\rho^\star$ (dense value 1, off scale). Right: best accuracy on the held-out split used for selection, with the dense-model accuracy shown as a dashed line.

The selection-off null retains the elitism, reseeding, and mutation steps of the outer loop, but replaces the fitness ranking with a uniformly random permutation at every epoch and skips the inner loop (which can no longer affect survival). Consequently, any density change reflects the proposals and these structural steps alone. Figure 6 plots 400-epoch MNIST runs with three seeds per curve (two for the density-agnostic operator at $B = 8192$). Under the density-agnostic operator, the population density climbs to approximately $0.56$ at $B = 512$ and $0.49$ at $B = 8192$ (reversing the ordering predicted by Corollary 1), while the selection-off null stays at $0.112 \pm 0.001$. Selection therefore amplifies the growth bias of the proposals on MNIST, where extra weights improve accuracy. Under the drift-neutral operator, the density rises slowly to $0.194 \pm 0.033$, $0.232 \pm 0.030$, and $0.201 \pm 0.028$ by epoch 400 at $B = 512$, $2048$, and $8192$, remaining well below the dense maximum against $0.102 \pm 0.002$ for its selection-off null.

Because the null skips the inner loop, it depends on the operator, $N$, $E$, $R$, and $\rho_0$, independent of the task: the teacher–student null sits at $0.102 \pm 0.001$ for $\rho_0 = 0.10$ (epochs 81–120) and at $0.300$ for $\rho_0 = 0.30$. Applying the same construction across three seeds yields the UCI and QKD nulls in Table 4 (where the BB84 null copies the operator and settings of the BB84 library search). The UCI nulls stay at their initial density of $0.05$, except under the density-agnostic operator, whose growth bias raises the density to $0.063$–$0.065$ within 40 epochs. By comparison, the selected UCI populations end at $0.055$–$0.067$ under the drift-neutral operator and at $0.12$–$0.16$ under the density-agnostic operator (Table 2). The QKD nulls likewise remain at their initial circuit size, averaging $5.57 \pm 0.05$ gates for BB84 and $6.05 \pm 0.05$ gates for E91 over the final ten epochs. Because these values lie inside the ranges of the discovered circuits (4–6 gates for BB84; 5–9 gates, median 6, for E91), the operators alone suffice to keep the circuits short over these horizons.

**Population size and elitism.** Table 5 varies the population size $N \in \{15, 30, 60\}$ and the number of elites $E \in \{2, 4, 8\}$ on the teacher–student benchmark at $B = 2^{12}$ ($\rho_0 = 0.10$, drift-neutral operator, $R = 0.2N$ reseeds, three seeds per cell, using the plateau rule of Appendix B). Parents for the $R$ mutated copies are drawn from the top $R/2$ ranks, extending beyond the elites when $E < R/2$. At $N = 30$ and 60, $\rho^\star$ remains at $0.050$–$0.054$ regardless of $E$, whereas at $N = 15$ it is both higher and more variable ($0.061$–$0.081$). Every cell settles below its own selection-off null ($0.101$–$0.104$), and all 27 runs met the plateau rule between epochs 120 and 170. The $(N = 30, E = 4)$ cell reuses seeds 0–2 from the main sweep.

| Search | Operator | Start | End of run |
|---|---|---|---|
| UCI, loop of Table 2 ($N = 100$, $E = 10$, $R = 20$, 40 epochs) | density-agnostic $0.40/0.20/0.40$ | $\rho = 0.05$ | 0.063–0.065 |
| | drift-neutral $0.40/0.30/0.30$ | $\rho = 0.05$ | 0.049–0.051 |
| | density-driven (target 0.05) | $\rho = 0.05$ | 0.050–0.051 |
| UCI, loop of Table 1 ($N = 30$, $E = 5$, $R = 10$, 60 epochs) | density-driven (target 0.05) | $\rho = 0.05$ | 0.050–0.051 |
| BB84 ($N = 100$, $E = 10$, $R = 20$, 40 epochs) | replace/remove/add $0.40/0.20/0.40$ | 5 gates | $5.75 \pm 0.14$ gates |
| E91 ($N = 30$, 8 elites, 12 reseeds, 25 epochs) | replace/remove/insert/perturb $0.35/0.20/0.25/0.20$ | 6 gates | $6.03 \pm 0.10$ gates |

Table 4: Selection-off nulls for the UCI and QKD searches (three seeds each). The UCI ranges span the Iris, Wine, and Cancer network shapes, while the QKD values report gate counts at the final epoch (mean $\pm$ std across seeds).

| | $E = 2$ | $E = 4$ | $E = 8$ |
|---|---|---|---|
| $N = 15$, $R = 3$ | $0.073 \pm 0.013$ | $0.061 \pm 0.007$ | $0.081 \pm 0.010$ |
| $N = 30$, $R = 6$ | $0.051 \pm 0.002$ | $0.054 \pm 0.001$ | $0.050 \pm 0.002$ |
| $N = 60$, $R = 12$ | $0.053 \pm 0.002$ | $0.053 \pm 0.003$ | $0.053 \pm 0.006$ |

Table 5: Equilibrium density $\rho^\star$ on the teacher–student benchmark at $B = 2^{12}$ across population size $N$ and elites $E$ (mean $\pm$ std over three seeds). The selection-off nulls across the nine cells lie at 0.101–0.104.

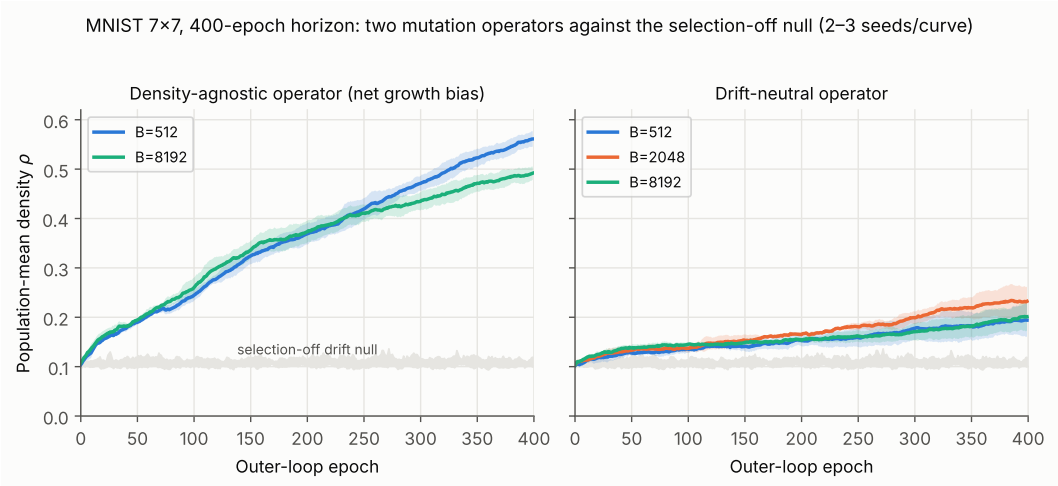



Figure 6: Population-mean density over 400 outer-loop epochs on MNIST $7 \times 7$ (three seeds per curve, and two for the density-agnostic operator at $B = 8192$; coloured bands show the per-epoch mean $\pm 1$ standard deviation over seeds). Left: density-agnostic operator (swap/remove/add at $0.40/0.20/0.40$). Right: drift-neutral operator ($0.40/0.30/0.30$). In both panels, the grey band shows the min–max envelope of the six selection-off runs (three per operator).

## E Accuracy saturation on UCI: the equilibrium drift

The natural-equilibrium claim holds most clearly in regimes where the inner-loop fitness surface continues to discriminate between topologies. On the UCI tabular benchmarks of §4.2, test accuracy saturates at $1.0$ within very few epochs on Iris and Wine, and within tens of epochs on Cancer. Once any topology in the population reaches this saturation regime, the inner-loop fitness gap between it and a strictly larger competitor approaches zero. At that point, the dimensionality-pressure mechanism of Proposition 1 loses traction: with no fitness advantage left to exploit, the rank-based selection bias against larger topologies remains inactive.

Figure 7 illustrates this effect by plotting the average population density across training epochs under the density-agnostic operator ($\alpha = 0$, with no score-level penalty and no target density). Across all three UCI datasets, the density climbs almost monotonically over the training horizon without settling into a flat asymptote, reinforced by the net growth bias of this operator (swap/remove/add at $0.40/0.20/0.40$; Appendix D). Under the drift-neutral operator (Figure 8), the population-mean density rises only slightly from $0.050$ to $0.064 \pm 0.006$ (Iris), $0.064 \pm 0.005$ (Wine), and $0.056 \pm 0.003$ (Cancer) by the final epoch, compared with $0.049$–$0.051$ for its selection-off null. Most of the upward drift under the density-agnostic operator therefore stems from its proposal growth bias, which selection amplifies. Consequently, the end-of-training values in Table 2 and its caption capture the density at the budget cut-off, prior to any stationary equilibrium. Figure 9 plots these density-agnostic end-of-training densities against accuracy alongside those of the biased operator.

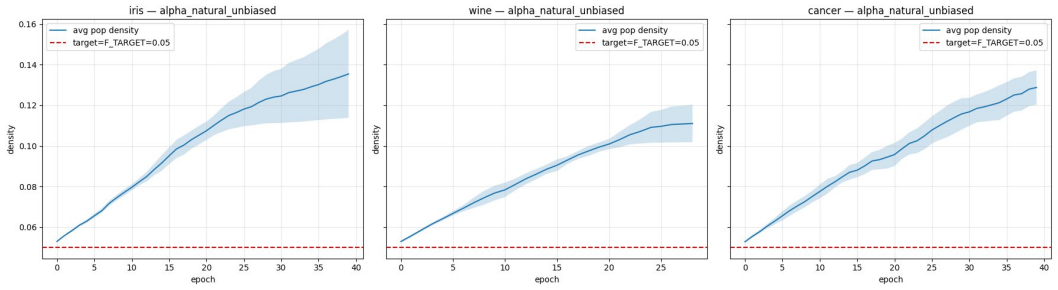



Figure 7: Average population density across training epochs under the density-agnostic operator ($\alpha = 0$, no target density) on the three UCI benchmarks. Solid lines show the per-seed average across five seeds, with shaded bands indicating $\pm 1$ standard deviation. The red dashed line marks the $\rho_{\text{target}} = 0.05$ reference used in the biased-mutation experiments of Table 1 (these runs do not target it). Across all three datasets, the density continues to climb without settling: the inner-loop fitness saturates near $1.0$ before a stationary distribution can form, and the selection mechanism of Proposition 1 loses pressure once the topologies in the population become functionally equivalent.

**Implications.** At any *finite* training horizon, the UCI density remains well below the dense baseline ($\rho < 0.16$ across all observed epochs versus $\rho_{\text{dense}} = 1$), as Proposition 1 requires, though the selection-off nulls show that this bound alone does not demonstrate that selection acts (Appendix D). Moreover, accuracy saturation invalidates interpreting the late-epoch density as a fixed point of the dynamics. In the QKD experiments of §4.3, the analytical optima lie strictly in the interior ($F_{AE} \approx 0.75$, well below the trivial $F_{AE} = 1$ ceiling), yet their selection-off nulls remain at the size of the discovered circuits over the run horizons (Appendix D), preventing those runs from isolating the mechanism on their own. The budget-gated test of §4.1 is therefore run on the teacher–student benchmark, where fitness does not saturate.

## F Preliminary budget sweep on MNIST

Testing Corollary 1 requires a budget sweep on a task where the dense model stops short of saturating at $1.0$ accuracy. As a first step, this appendix extends the EvoluCMAES methodology beyond the UCI tabular benchmarks to a larger dataset and a new architecture.

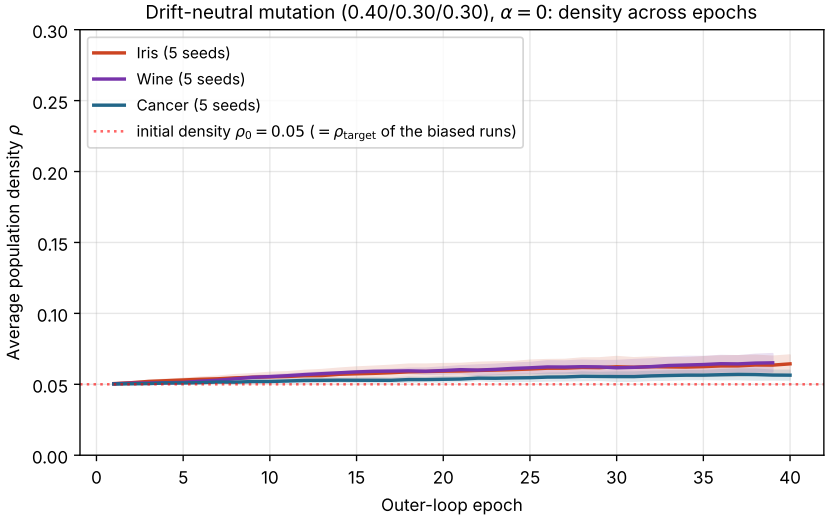



Figure 8: Average population density across training epochs under the drift-neutral operator (swap/remove/add at $0.40/0.30/0.30$, $\alpha = 0$, no target density) on the three UCI benchmarks, showing the mean over five seeds with a $\pm 1$ standard deviation band. Because runs stop early at different epochs, later epochs average over the seeds still running. The red dotted line marks the initial density of $0.05$.

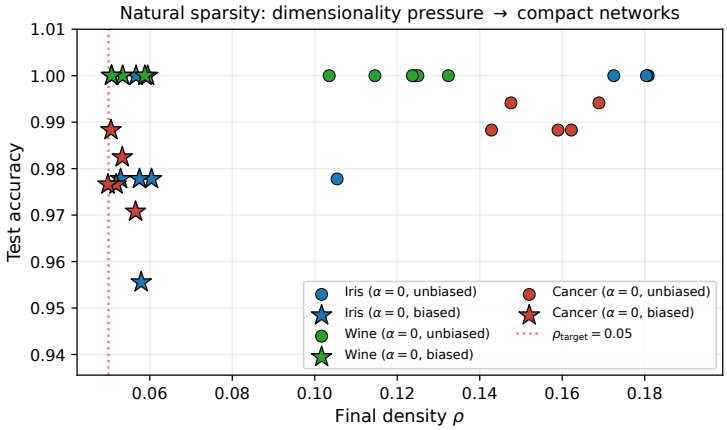



Figure 9: Density vs. test accuracy for $\alpha = 0$ runs. Stars show biased mutation hitting $\rho_{\text{target}} = 0.05$ (red dotted line) without any score-level penalty, while circles show density-agnostic mutation ending training at $\rho \approx 0.10$–$0.18$. Across all three UCI datasets, both regimes maintain $\geqslant 0.94$ accuracy on the held-out split (which is also used for selection).

**Setup.** A deliberately under-sized $49 \to 16 \to 10$ MLP ($\sim 970$ parameters) is trained on MNIST [LeCun et al., 1998] downsampled to $7 \times 7$ (49-dimensional input), capping the dense-model accuracy near $0.83$, well below ceiling saturation. The outer loop matches NNpruning_experiment_4 ($\alpha = 0$, unbiased `random_modif_quantum` with swap/remove/add at $0.40/0.20/0.40$, POP $= 30$, ELITISM $= 4$, RESEED $= 6$, and 60 outer-loop epochs), while the inner CMA-ES runs with `maxfevals` $= B$ swept over $B \in \{2^7, 2^8, \ldots, 2^{16}\}$ across 3 seeds. As in the UCI runs, outer-loop selection uses the same held-out split on which test accuracy is evaluated, biasing the reported accuracies upwards. Repeating the sweep under the drift-neutral operator of Appendix D follows the same protocol, with the values at $B = 2^9$, $2^{11}$, and $2^{13}$ read at epoch 60 of the 400-epoch runs in Appendix D.

**Result.** Table 6 reports the equilibrium density and best-individual test accuracy across 9 octaves of $B$ (from $2^7 = 128$ to $2^{16} = 65{,}536$). Under the density-agnostic operator, the equilibrium density rises from $\rho = 0.182$ at $B = 2^7$ to a peak of $\rho = 0.224$ at $B = 2^{15}$, while best-individual test accuracy climbs from $0.58$ to $0.84$. Under the drift-neutral operator, the density is lower at every budget (while remaining above its selection-off null of $0.102$ throughout), rising from $0.124$–$0.128$ for $B \leqslant 2^9$ to a peak of $0.162$ at $B = 2^{14}$ ($0.156$ at $2^{15}$, within one standard deviation) and $0.147$ at $2^{16}$. Although the upward shift in $\rho(B)$ matches the direction predicted by Corollary 1, denser masks also fit MNIST better. Model capacity is therefore rewarded alongside any budget effect, confounding the two factors until they are separated on the teacher–student benchmark of §4.1.

| | Density-agnostic operator | | Drift-neutral operator | |
|---|---|---|---|---|
| $B$ | $\rho_{\text{eq}}$ (mean $\pm$ std) | best test acc | $\rho_{\text{eq}}$ (mean $\pm$ std) | best test acc |
| $2^7 = 128$ | $0.182 \pm 0.016$ | 0.58 | $0.126 \pm 0.011$ | 0.57 |
| $2^8 = 256$ | $0.182 \pm 0.012$ | 0.66 | $0.124 \pm 0.013$ | 0.60 |
| $2^9 = 512$ | $0.197 \pm 0.018$ | 0.69 | $0.128 \pm 0.005$ | 0.61 |
| $2^{10} = 1024$ | $0.204 \pm 0.002$ | 0.71 | $0.134 \pm 0.012$ | 0.65 |
| $2^{11} = 2048$ | $0.208 \pm 0.002$ | 0.74 | $0.136 \pm 0.007$ | 0.70 |
| $2^{12} = 4096$ | $0.213 \pm 0.005$ | 0.79 | $0.143 \pm 0.009$ | 0.74 |
| $2^{13} = 8192$ | $0.201 \pm 0.008$ | 0.80 | $0.139 \pm 0.007$ | 0.78 |
| $2^{14} = 16384$ | $0.219 \pm 0.006$ | 0.84 | $0.162 \pm 0.009$ | 0.81 |
| $2^{15} = 32768$ | $0.224 \pm 0.016$ | 0.84 | $0.156 \pm 0.005$ | 0.80 |
| $2^{16} = 65536$ | $0.209 \pm 0.010$ | 0.84 | $0.147 \pm 0.007$ | 0.80 |

Table 6: MNIST $7 \times 7$ with the 970-parameter MLP (3 seeds per cell) under the density-agnostic and drift-neutral operators (at $B = 2^9$, $2^{11}$, and $2^{13}$, drift-neutral values are read at epoch 60 of the 400-epoch runs). Here, $\rho_{\text{eq}}$ is the population-mean density averaged over epochs 56–60, and best test acc is measured on the held-out split also used for selection.

**Analysis.** Table 6 highlights two main patterns. First, the equilibrium density rises with $B$ across most of the swept range, matching the qualitative direction predicted by Corollary 1 despite the confound between capacity and search cost on this benchmark. Second, across the entire sweep, the population remains simultaneously *sparse* (under $23\%$ active parameters at every $B$, an order of magnitude below the dense baseline) and *accurate*, with best-individual accuracy on the selection split reaching the architecture's capacity ceiling of $\sim 0.84$ for $B \geqslant 2^{14}$. Extending the framework across domains therefore preserves the two core features of the UCI result that motivate the natural-bound regime (low equilibrium density and competitive accuracy at that density), while also exhibiting the budget-dependent shift predicted by Corollary 1.

**Hypothesis.** In absolute terms, the shift in $\rho_{\text{eq}}$ in Table 6 is modest (peaking at $+0.04$ in $\rho$ over a $\sim 500\times$ range in $B$), and the population density $\rho$ continues to drift upwards across outer-loop epochs at the cut-off used here. Panel (B) of Figure 10 measures the per-$\rho$ convergence horizon directly: $\rho = 0.05$ saturates near $0.21$ accuracy at $B \approx 240$, $\rho = 0.20$ requires $B \approx 22{,}000$ to reach its ceiling near $0.65$, and larger-$\rho$ topologies require still greater budgets. Accordingly, the regime where Corollary 1 expects the variance-gap selection bias of Proposition 1 to act most strongly sits near the upper end of the swept range ($B \approx 2^{14}$–$2^{15}$), where smaller-$\rho$ topologies have reached their ceilings while larger-$\rho$ ones are still climbing. Because these densities have different capacity

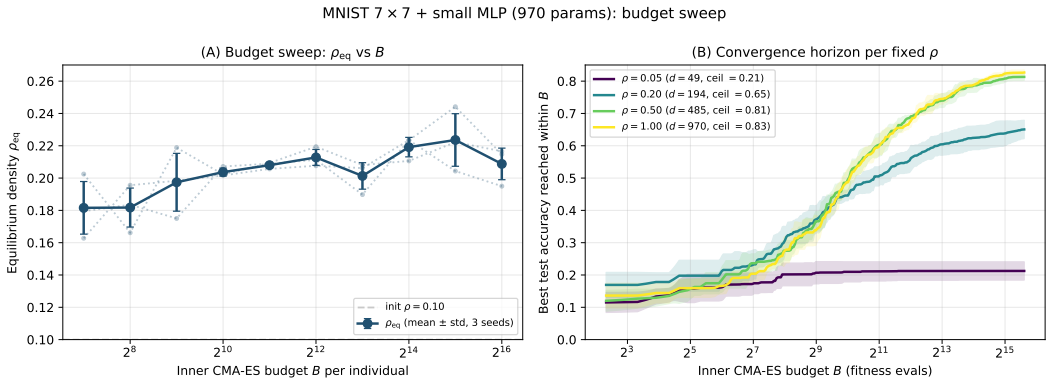



Figure 10: (A) Equilibrium density $\rho$ as a function of inner-loop budget $B$ on MNIST $7 \times 7$ (3 seeds, mean $\pm$ std), showing a monotone trend in $B$. (B) Best held-out accuracy reached within $B$ inner-loop fitness evaluations at four fixed mask densities $\rho \in \{0.05, 0.20, 0.50, 1.00\}$, illustrating the per-$\rho$ convergence horizons. Each density saturates at its own capacity-limited ceiling, and the budget required to reach that ceiling grows steeply with $\rho$.

ceilings, Panel (B) describes per-density horizons without testing assumption *(A1)*, which concerns topologies sharing an optimum (a condition met instead by the teacher–student grid of Appendix B). Running the outer loop for 400 epochs with the density-agnostic operator of this sweep does not sharpen the predicted trend: the density climbs towards $0.5$ and the budget ordering inverts, reflecting a drift driven by the growth bias of the operator (Appendix D).

**Limitation.** Because the population density $\rho$ continues to drift upwards across the outer-loop epochs and has not yet plateaued at the cut-off used here, the values in Table 6 represent conservative lower bounds on the stationary distribution. A definitive sweep on this benchmark would either run the outer loop until the per-individual *inner* loop has plateaued at every $B$ (giving the variance-gap selection bias of Proposition 1 a fair chance to act on near-optimal $\hat{f}_B(T)$ values beyond the transient regime), or use budgets large enough for the inner CMA-ES to saturate within a small number of outer-loop epochs.

## G Additional small tabular benchmarks

Four additional small tabular datasets from OpenML [Vanschoren et al., 2014] (IDs 1499, 41, 59, and 40) were evaluated at $\rho = 0.05$ across five seeds using the five baselines and the $[d, 64, 32, k]$ network shape of Table 1: Seeds [Charytanowicz et al., 2010] (210 samples), Glass [German, 1987] (214), Ionosphere [Sigillito et al., 1989] (351), and Sonar [Sejnowski and Gorman, 1988] (208). Each dataset undergoes a single stratified split into $70\%$ training and $30\%$ test data. Under the validation-split protocol, EvoluCMAES trains its inner loop on $80\%$ of the training partition, selects on the remaining $20\%$, and reports accuracy on the untouched test split. Because the baselines require no selection split and train on the entire $70\%$ training partition, this protocol is slightly conservative for EvoluCMAES. Under the protocol of Table 1 (included for comparison only), both selection and reported accuracy use the test split. Across every method and dataset, the final density settles at $0.050$–$0.054$.

Table 7 reports the resulting test accuracies. With selection restricted to the validation split, EvoluCMAES ranks first on Seeds, finishes within $0.006$ of the top baseline on Glass, and trails the best baseline on Ionosphere and Sonar. Its mean accuracy across the four datasets ($0.792$) lies within the range of the four pruning baselines ($0.789$–$0.814$). On its own selection split, EvoluCMAES reaches $1.000$, $0.707$, $0.988$, and $0.952$, while the row following the Table 1 protocol (mean $0.895$) illustrates how strongly selecting on the reporting split inflates accuracy. Using eight worker processes, an EvoluCMAES run takes $16.5$–$36.4$ s under the validation-split protocol and $24.9$–$43.1$ s under the Table 1 protocol, compared with $0.1$–$0.5$ s for each baseline running in a single process. Running the same script under the Table 1 protocol reproduces 17 of the 18 cells in that table to three decimal places (IMP on Cancer differs in a single seed, yielding a mean of $0.952$ versus $0.951$).

| Method | Seeds | Glass | Ionosphere | Sonar | Mean |
|---|---|---|---|---|---|
| Random | $0.543 \pm 0.161$ | $0.428 \pm 0.051$ | $0.674 \pm 0.064$ | $0.692 \pm 0.079$ | 0.584 |
| Magnitude | $0.857 \pm 0.120$ | $0.612 \pm 0.083$ | $0.919 \pm 0.017$ | $0.825 \pm 0.050$ | 0.803 |
| IMP | $0.860 \pm 0.032$ | $0.662 \pm 0.048$ | $0.902 \pm 0.048$ | $0.733 \pm 0.045$ | 0.789 |
| SNIP | $0.838 \pm 0.037$ | $0.668 \pm 0.021$ | $0.919 \pm 0.013$ | $0.775 \pm 0.035$ | 0.800 |
| SynFlow | $0.867 \pm 0.008$ | $0.662 \pm 0.067$ | $0.913 \pm 0.026$ | $0.816 \pm 0.022$ | 0.814 |
| EvoluCMAES (validation) | $0.883 \pm 0.038$ | $0.662 \pm 0.046$ | $0.896 \pm 0.015$ | $0.727 \pm 0.025$ | 0.792 |
| EvoluCMAES (Table 1 protocol)* | $0.975 \pm 0.008$ | $0.732 \pm 0.023$ | $0.970 \pm 0.004$ | $0.902 \pm 0.025$ | 0.895 |

Table 7: Test accuracy at $95\%$ sparsity on four additional OpenML datasets (mean $\pm$ std over five seeds; network $[d, 64, 32, k]$). Neither the baselines nor the validation row use the test split for selection. *Performs both selection and accuracy evaluation on the test split (as in Table 1), biasing this row upwards; included for comparison only.

# H Background: Quantum Key Distribution

This appendix introduces the minimum vocabulary needed to follow the QKD material in §4.3 and Appendix K. For a comprehensive textbook treatment, readers should consult Scarani et al. [2009] and the references therein.

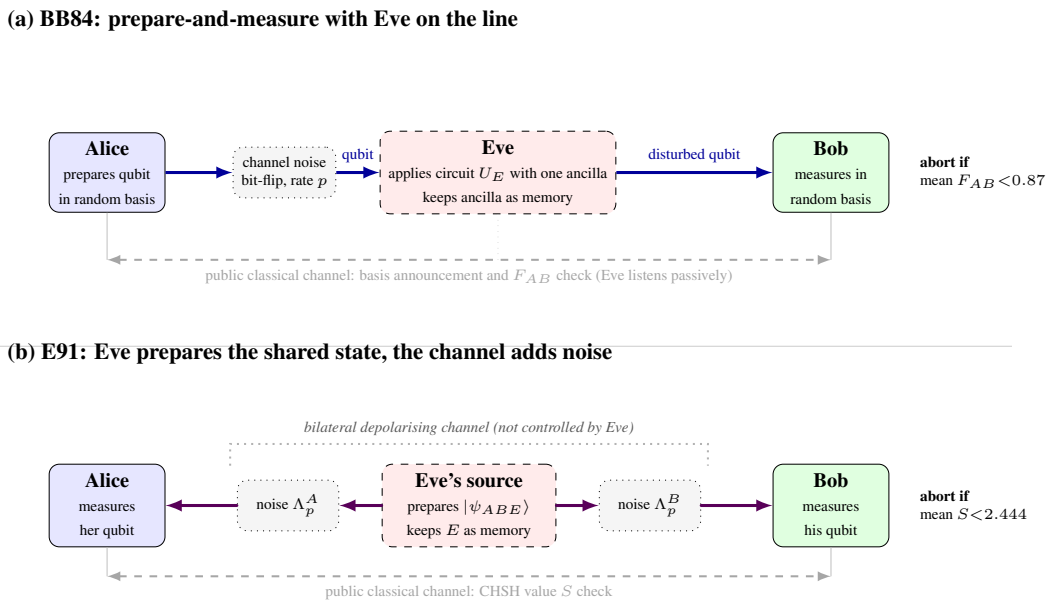



Figure 11: **The two QKD protocols used in this work. (a) BB84 (prepare-and-measure).** Alice sends single qubits to Bob. In the experiments, fixed bit-flip noise of rate $p$ (which Eve cannot alter) acts on each qubit before reaching Eve. She then passes the qubit and a single ancilla through a circuit $U_E$ (learnt by EvoluCMAES), forwards the qubit to Bob, and retains the ancilla in memory. Once the bases are announced publicly, Eve measures her ancilla in the announced basis. Alice and Bob abort if the episode mean of their matched-basis agreement $F_{AB}$ falls below $0.87$, while Eve's reward is $F_{AE}$ (the probability that her measurement recovers Alice's bit). **(b) E91 (entanglement-based).** Eve controls the source, preparing a three-qubit state whose $A$ and $B$ qubits travel to Alice and Bob while she retains $E$. A bilateral depolarising channel $\Lambda_p^A \otimes \Lambda_p^B$ acts on the two travelling qubits as physical noise beyond Eve's control, where $\Lambda_p^A$ reduces Eve's information and $\Lambda_p^B$ lowers the CHSH value (Appendix K). Alice and Bob abort if the episode mean of the CHSH value $S$ (bounded at $2\sqrt{2}$ in quantum theory) falls below $2.444$. Eve's reward is the Holevo information $\chi(A{:}E)$.

**What problem QKD solves.** Quantum key distribution (QKD) is a family of protocols that enables two parties (by convention, Alice and Bob) to agree on a shared random secret string, while ensuring that any eavesdropper (Eve) who tries to read the messages in transit leaves a statistical signature that Alice and Bob can detect and abort on. The security argument relies on the no-cloning property of quantum states: because copying an unknown quantum state without disturbing it is impossible, any measurement Eve performs perturbs the transmitted state and inflates a publicly checkable error rate. Whereas formal security proofs attribute all observed disturbance to Eve [Scarani et al., 2009], the experiments in this paper fix a physical noise process beyond Eve's control and measure what a restricted Eve can gain on top of it.

**BB84 (prepare-and-measure).** In the canonical prepare-and-measure protocol BB84 [Bennett and Brassard, 1984], Alice prepares each qubit (a two-level quantum system) in one of four states drawn from two "bases" (analogous to two orthogonal coordinate systems in which the qubit can be measured). Alice picks a random basis for each qubit, and Bob picks a random basis for each measurement. Their basis choices match in roughly half of the rounds, in which case Bob's outcome equals Alice's bit deterministically in the noiseless regime, while the remaining rounds are discarded after a public basis announcement. The security check acts on the matched-basis rounds: Alice and Bob publicly compare a small fraction of their bits to compute the *Alice–Bob fidelity* $F_{AB}$ (the fraction of agreeing bits). While $F_{AB} = 1$ on a noiseless channel, channel noise reduces this value, and if its episode mean falls below a fixed threshold (here, $0.87$), Alice and Bob declare the link compromised (whether by noise or by Eve) and abort. Because this threshold corresponds to a basis-averaged error rate of $13\%$, above the $11\%$ limit where the one-way key rate of BB84 vanishes [Shor and Preskill, 2000], it serves as a modelling choice for this testbed. Figure 11(a) summarises the flow.

**What Eve does.** Positioned between Alice and Bob, Eve intercepts each transmitted qubit and passes it, together with one or more ancilla qubits under her control, through a quantum circuit $U_E$ that entangles the two systems. She then forwards the Bob-bound qubit to Bob and stores her ancillas as a quantum memory. Once the bases are announced publicly, Eve measures her memory in the announced basis; her *Alice–Eve fidelity* $F_{AE}$ is the fraction of Alice's bits she guesses correctly. An optimal (Helstrom) measurement on her memory would yield at least this value [Helstrom, 1969], and because each ancilla is measured separately, $F_{AE}$ characterises individual attacks. Because Eve cannot raise $F_{AE}$ without lowering $F_{AB}$, the public detection threshold caps what she can extract.

On a noiseless channel, the optimal individual attack obeys the closed-form relation $F_{AE} = \frac{1}{2} + \sqrt{F_{AB}(1 - F_{AB})}$ [Fuchs et al., 1997], which is attained within the *Pauli-channel cloning machines* (PCCM) of Cerf [2000] and recovered on that channel by the same evolutionary search [Mordarski et al., 2026]. Following Decker et al. [2025], the experiments in this paper apply the bit-flip noise to the travelling qubit before Eve's interaction, which degrades Eve's copy as well. Under such pre-interaction noise, a cloner imbalanced between the two bases outperforms a symmetric one [Decker et al., 2025]; the circuits discovered by EvoluCMAES come within $2\%$ of the noiseless closed form at low noise and fall further below it as the noise grows (§4.3).

**E91/DIQKD (entanglement-based) and why both protocols are used here.** In E91 [Ekert, 1991], the entanglement-based counterpart to BB84, an external (potentially untrusted) source emits an entangled qubit pair and sends one half to Alice and the other to Bob for independent measurement, replacing single-qubit preparation by Alice. The public statistic is the *CHSH violation* $S$, a Bell-inequality test that reaches its maximal quantum value $S = 2\sqrt{2} \approx 2.83$ for a perfect entangled pair and satisfies $S \leqslant 2$ for any state explainable by classical correlations. Detection is declared if the episode mean of $S$ falls below a fixed operational threshold (here, $\approx 2.444$). In the experiments, Eve controls the source and retains a third system as memory, while the channel noise acts on both travelling qubits. Figure 11(b) sketches this entanglement-based flow.

The device-independent (DIQKD) regime [Acín et al., 2007, Pironio et al., 2009] additionally drops the assumption that Alice and Bob trust their own devices: only the observed CHSH value is taken at face value, and the security analysis must hold for any quantum apparatus that produces those statistics. Because the experiments here fix both the dimension of Eve's system and the measurement settings, they describe a dimension-restricted attacker within this setup and stop short of device-independent guarantees. Eve's information is quantified by the Holevo information $\chi(A : E)$ (where higher $\chi$ favours Eve), which bounds the classical information about Alice's outcome that Eve can

extract via a joint measurement across many stored copies of her system, placing the E91 results in the collective-attack regime.

Including both protocols exposes the EvoluCMAES library to two distinct operating regimes: BB84 combines a permissive detection threshold with a single-qubit channel (bit-flip), whereas E91 pairs a tighter threshold relative to the honest baseline with a two-qubit channel acting simultaneously on both halves of the entangled pair. Appendix K gives the technical details of the E91 channel structure.

**Glossary.** Quick reference for the QKD terminology used throughout the paper:

| Term | Meaning in this paper |
|---|---|
| Alice/Bob/Eve | Sender/receiver/eavesdropper. |
| Qubit | Two-level quantum system; the basic unit of quantum information. |
| BB84 | Prepare-and-measure QKD protocol [Bennett and Brassard, 1984]. |
| E91/DIQKD | Entanglement-based QKD protocol [Ekert, 1991]; DIQKD [Acín et al., 2007] adds the untrusted-device assumption. |
| Channel | Physical noise process that Eve cannot alter: bit-flip noise on the travelling qubit before Eve's interaction in BB84, and bilateral depolarising noise on the entangled pair after Eve's preparation in E91. |
| $F_{AB}$ | Alice–Bob fidelity: fraction of matched-basis rounds where Bob's bit equals Alice's (publicly observed and used for detection in BB84). |
| $F_{AE}$ | Alice–Eve fidelity: fraction of Alice's bits Eve guesses correctly after measuring her ancilla in the announced basis under individual attacks (Eve's per-round reward in BB84). |
| CHSH violation $S$ | Bell-inequality statistic, taking $S = 2\sqrt{2}$ for a perfect entangled pair and $S \leqslant 2$ for classical correlations (publicly observed and used for detection in E91). |
| Holevo $\chi(A:E)$ | Upper bound on the classical information about Alice's outcome that Eve can extract from her stored systems via joint measurement under collective attacks (Eve's per-round reward in E91). |
| PCCM | Pauli-channel cloning machines [Cerf, 2000], which attain the optimal individual-attack trade-off for BB84 on a noiseless channel [Fuchs et al., 1997]. |
| $U_E$ | Eve's circuit: a unitary acting on the transmitted qubit and one ancilla (BB84), or a three-qubit state preparation (E91). |
| Detection threshold | The value of $F_{AB}$ (or $S$) below which Alice and Bob abort the protocol. |
| Library cell | A single entry in the EvoluCMAES library, pairing a (target, noise-anchor) point with a discovered circuit. |

# I Library-as-Action-Basis: RL Extension on QKD

**Context.** This appendix frames the eavesdropping task as a constrained Markov decision process (MDP) in which Eve (the eavesdropper) acts as the RL agent. At each round, she observes a noisy channel parameter $p$ and selects a circuit from the EvoluCMAES library to clone the qubit transmitted between the legitimate parties, Alice and Bob. Her per-round reward is an information measure ($F_{AE}$ for BB84; the Holevo information $\chi(A:E)$ for E91), while her constraint is a public statistic of Alice and Bob's outcomes ($F_{AB}$ for BB84; the CHSH violation $S$ for E91). Alice and Bob monitor the mean of this statistic over a 50-round episode and declare the attack detected if it falls below a known threshold. Because the channel parameter $p$ drifts over time according to an Ornstein-Uhlenbeck (OU) process, the optimal action depends on the current noise level, making static optimisation (a

single fixed circuit) suboptimal. This appendix tests whether an off-the-shelf RL agent, given the EvoluCMAES library as its action set, can match the greedy-Oracle policy on this MDP. Whereas the BB84 case (which uses the standard bit-flip channel) is self-contained here, the full derivation of the E91/DIQKD setup (channel structure, detection threshold, and library construction) appears in Appendix K.

**The MDP.** Under OU channel-noise drift, the EvoluCMAES library of §4.3 serves as a discrete action set, converting the static cloning problem into a constrained MDP. At each round, Eve picks one library cell based on the current noise level (observed without error, an idealisation that favours the attacker), its recent history, the recent constraint observable, her mean reward so far, and the time remaining. A feasibility mask restricts the available actions: on BB84, it admits cells whose single-round $F_{AB}$ at the current noise level reaches the threshold, whereas on E91, it admits cells that keep the projected ten-round mean of $S$ above the threshold. Under this same admissibility rule, the greedy Oracle plays the highest-reward cell at each step, thereby avoiding detection while ignoring future rounds. By comparison, the BB84 static-safe baseline acts as a noise-indexed lookup that adapts to the drift by playing the most conservative library level at the noise anchor nearest the current $p$. Finally, the dynamic programming (DP) reference is computed via backward induction over the round, a grid in $p$, and a grid in the accumulated statistic (using 100 noise levels and 2,000 bins of accumulated $F_{AB}$ on BB84, versus 16 noise levels, 250 bins of accumulated $S$, and a finite terminal penalty on E91, which explains why its E91 rollouts incur detection).

To keep the comparison consistent across protocols, the headline evaluation below focuses on a single RL agent (Rainbow-DQN with feasibility masking) run on *both* BB84 and E91 with matched hyperparameters. Additional agents and ablations (vanilla PPO, RecurrentPPO, MaskablePPO, behaviour cloning, and MPC) are reported in the per-experiment Appendix K for completeness.

### I.1 BB84: detection-budget trade-off

In the BB84 protocol, a bit-flip channel of rate $p_{\text{bf}}$ acts on the travelling qubit before Eve's interaction. The public statistic is $F_{AB}$ (which equals $1 - p_{\text{bf}}/2$ without Eve), and the detection threshold requires an episode mean of $F_{AB} \geqslant 0.87$. The training reward is the per-round $F_{AE}$ combined with shaping penalties on the running mean of $F_{AB}$ and a terminal detection penalty. Checkpoints are ranked by an evaluation score using one of two weights $\kappa$ on the detection rate:

- **Strict:** Score $= F_{AE} - 2.0 \times$ (detection rate), heavily penalising detection.
- **Tolerated:** Score $= F_{AE} - 0.5 \times$ (detection rate), moderately penalising detection.

**Strict variant: matches greedy Oracle.** Figure 12 shows the training curve under the strict penalty. On the 50-episode test, the best checkpoint (step 185k) reaches $F_{AE} = 0.7434$ at a $0\%$ detection rate, matching the greedy Oracle ($F_{AE} = 0.7434$, det $= 0\%$) in every episode to within $7 \times 10^{-4}$, while the next two best checkpoints (steps 190k and 440k) lie within $0.0002$ of this value. This agreement indicates that the EvoluCMAES library combined with feasibility masking provides a sufficient action set, allowing the agent to recover the greedy policy directly.

**Tolerated variant: exceeds Oracle at $4\%$ detection.** Under the looser checkpoint score, the best-checkpoint policy (step 750k) reaches $F_{AE} = 0.7477$ at a $4\%$ detection rate (2 of 50 test episodes), compared with $0.7429$ for the greedy Oracle on the same 50 episodes. This yields a paired gain of $0.0047$ (paired $95\%$ interval $\pm 0.0011$), with the tolerated agent scoring higher in 46 of the 50 episodes (two-sided sign test, $p \approx 4 \times 10^{-10}$). The gain closes roughly half the gap from the greedy Oracle (which leaves this gap because it ignores future rounds of the episode) to the DP reference at $F_{AE} = 0.753$. Because the DP reference reaches $0.753$ with zero detection, incurring detection events is unnecessary to obtain this extra information.

Figure 13 compares each BB84 policy against the greedy Oracle evaluated on the same episodes: the strict variant matches the Oracle, whereas vanilla PPO and RecurrentPPO plateau $0.013$–$0.016$ below it.

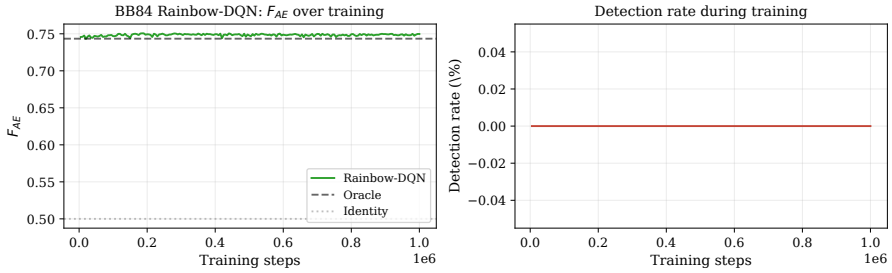



Figure 12: BB84 Rainbow-DQN with the strict checkpoint score ($\kappa = 2.0$). Left: $F_{AE}$ across 20-episode evaluations during 1M training steps (per-episode standard deviation $\approx 0.03$, giving a standard error of $\approx 0.007$ per point), peaking at $0.751$ near step 185k and remaining between $0.744$ and $0.751$ thereafter. The selected checkpoint scores $0.7434$ on the 50-episode test, matching the greedy Oracle. Right: detection rate, which remains at $0\%$. Reference lines indicate the greedy Oracle (dashed) and Identity (dotted). The tolerated variant (omitted here) reaches $F_{AE} = 0.7477$ at a $4\%$ detection rate at step 750k.

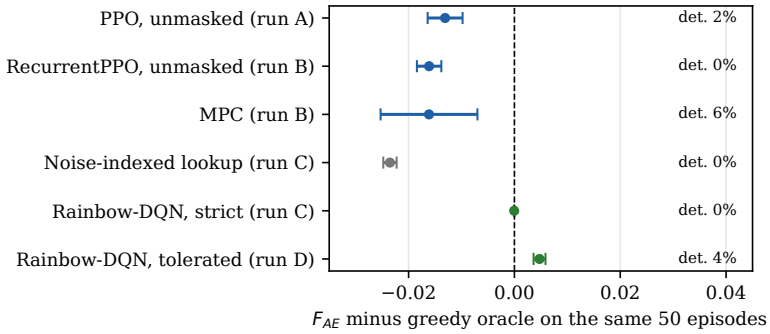



Figure 13: BB84/bit-flip channel: difference in mean $F_{AE}$ between each policy and the greedy Oracle evaluated on the same 50 episodes, shown with paired $95\%$ intervals and detection rates. Runs A to D are separate evaluations, each with its own Oracle score (A: $0.7357$; B: $0.7423$; C: $0.7434$; D: $0.7429$). Vanilla PPO and RecurrentPPO remain $0.013$–$0.016$ below the Oracle, the noise-indexed static-safe lookup trails by $0.024$, and the strict Rainbow-DQN agent matches it.

### I.2 E91/DIQKD: same agent, narrower OU range

Running the same Rainbow-DQN agent and feasibility masking on the E91 library of §K.4 requires no algorithmic changes: the protocol differences (tracking CHSH instead of $F_{AB}$ over a narrower OU drift range) pass directly through the environment by updating only the threshold constants and the information measure. Across 200 episodes, the agent reaches $\chi = 0.3484$ at det $= 0\%$, compared with $0.3507$ for the greedy Oracle on the same episodes (Table 13). Table 8 compares the policies episode by episode using an evaluation reconstruction that reproduces every aggregate value in Table 13. Because all policies share the per-episode channel seeds, this pairing removes the channel-to-channel variance that dominates unpaired comparisons at this effect size. On E91, the upper edge of the drift range lies $0.01$ below the noise level at which the honest link stops producing key, and whether this geometry removes the benefit of tolerating detection seen on BB84 remains untested here. Figures 14 and 15 present the corresponding training curves and methods bar chart (analogous to Figures 12 and 13).

## J Inner-loop comparison on BB84: CMA-ES, SPSA, and random search

This appendix reports an initial inner-optimiser ablation on the BB84 cloning task. The EvoluCMAES outer evolutionary loop is run with three different inner-loop optimisers over the rotation-angle parameters $\theta$: CMA-ES (the default used throughout the paper), SPSA [a gradient-approximation

| Comparison | Mean difference | 95% interval | Wilcoxon $p$ |
|---|---|---|---|
| Rainbow-DQN − MaskablePPO | $+0.0017$ | $[+0.0012, +0.0022]$ | $1.0 \times 10^{-8}$ |
| Rainbow-DQN − Oracle (greedy) | $-0.0023$ | $[-0.0028, -0.0018]$ | $2.5 \times 10^{-15}$ |
| Rainbow-DQN − DP (7% detection) | $-0.0056$ | $[-0.0065, -0.0047]$ | $1.3 \times 10^{-21}$ |

Table 8: Paired per-episode comparisons of $\chi(A{:}E)$ on E91 over $n = 200$ episodes with shared channel realisations, using two-sided Wilcoxon signed-rank tests [Wilcoxon, 1945]. Because the DP reference incurs $7\%$ detection whereas the agent incurs none, the two are evaluated under different effective constraints.

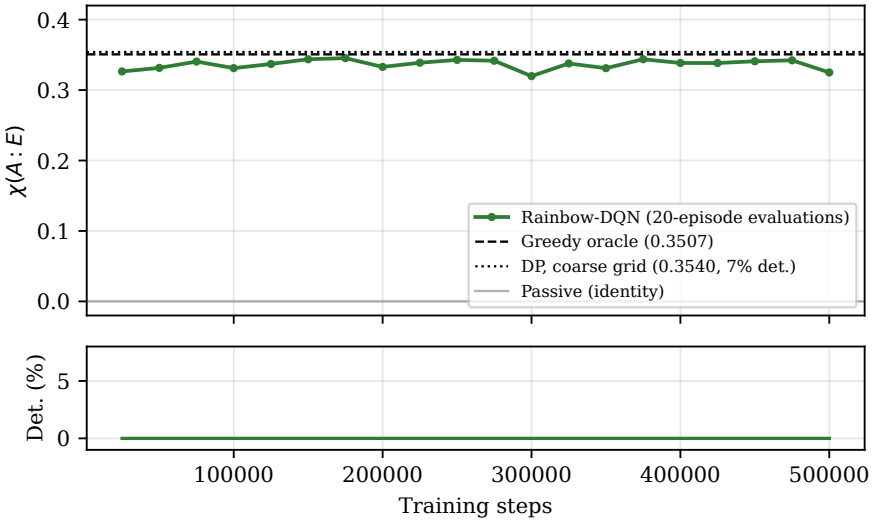



Figure 14: E91/bilateral depolarising channel: Rainbow-DQN training curve. Top: Holevo $\chi(A : E)$ across 20-episode evaluations (per-episode standard deviation $\approx 0.065$, giving a standard error of $\approx 0.015$ per point). Bottom: detection rate. Reference lines show the 200-episode test values from Table 13: greedy Oracle (dashed), coarse-grid DP (dotted), and Identity (solid grey).

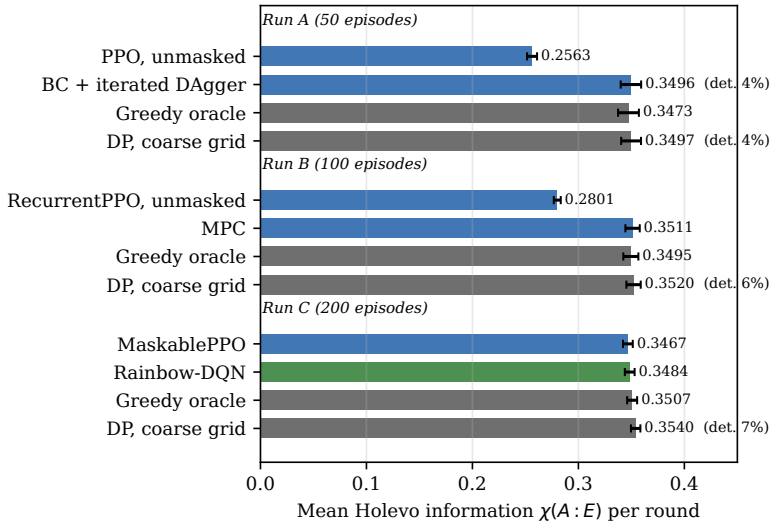



Figure 15: E91/DIQKD: mean $\chi$ per round with $\pm 1$ standard error, grouped by evaluation run (runs A, B, and C correspond to Tables 11, 12, and 13). Within each run, all policies share the channel realisations, and non-zero detection rates are annotated directly. Rainbow-DQN with feasibility masking comes within $0.0023$ of the greedy Oracle in run C, whereas vanilla PPO and RecurrentPPO collapse to $\chi \approx 0.26$–$0.28$.

baseline widely used in variational quantum algorithms; Cerezo et al., 2021], and uniform random search given twice the evaluation budget to compensate for its lack of adaptation. All three runs share identical outer-loop population dynamics, gate-set mutations, and selection rules, varying only the inner $\theta$-optimiser.

| Method | $F_{AE}$ | $F_{AB}$ | Loss | Gates | Evals | Best $F_{AE}$ | Best $F_{AB}$ | Best Gates |
|---|---|---|---|---|---|---|---|---|
| SPSA | $0.760 \pm 0.003$ | $0.794 \pm 0.002$ | $-0.756 \pm 0.002$ | $6.4 \pm 0.8$ | 2000 | 0.763 | 0.794 | 6 |
| CMA-ES | $0.764 \pm 0.001$ | $0.794 \pm 0.000$ | $-0.760 \pm 0.001$ | $5.2 \pm 0.7$ | 2000 | 0.763 | 0.794 | 4 |
| Random | $0.756 \pm 0.006$ | $0.796 \pm 0.002$ | $-0.754 \pm 0.004$ | $5.6 \pm 1.0$ | 4000 | 0.764 | 0.793 | 6 |

Table 9: Inner-loop optimiser comparison on the BB84 cloning task (mean $\pm$ std over five seeds). Random search receives $2\times$ the evaluation budget of CMA-ES and SPSA, and the "Best" columns report the per-seed best across the run. All three methods recover nearly identical final $F_{AE}$ values (within $0.01$ of one another, comparable to the cross-seed standard deviation) and the same equilibrium gate count ($4$–$6$), consistent with the view that Proposition 1 is driven by the evaluation budget rather than the specific optimiser on this task.

**Reading.** Table 9 highlights three main points. First, the evaluation budget drives the outcome at this scale: random search at $4{,}000$ evaluations matches CMA-ES at $2{,}000$ evaluations on the headline metric, with fully overlapping gate-count distributions. Second, the equilibrium gate count of $4$–$6$ reproduces across all three inner loops, matching the cross-optimiser consistency predicted by the dimensionality-pressure mechanism (where the equilibrium location is governed by the inner-loop *convergence-vs-budget trade-off* regardless of the specific algorithm). Third, because the loss values track $F_{AE}$ with the expected sign (as the BB84 outer-loop fitness is essentially $-F_{AE}$ on the matched-basis rounds), these runs provide a single-task null result against the hypothesis that CMA-ES is specifically required to reach the natural equilibrium. Extending this experiment across domains (in particular, to a non-saturating pruning benchmark per Appendix E) is left to follow-on work.

# K E91/DIQKD Extension: Derivation and Full Results

This appendix complements Appendix I by providing the technical derivation of the E91/DIQKD setup (channel structure, detection threshold, and library construction) alongside the full per-method ablation tables (vanilla PPO, RecurrentPPO, MaskablePPO, behaviour cloning, and MPC) omitted from the headline comparison in Appendix I. Readers satisfied with the BB84 case in Appendix I may skip this section, whereas QKD-theory readers will find the bilateral-depolarising channel analysis, the detection-threshold derivation, and the comparison against the noiseless Acín envelope.

## K.1 BB84 $\leftrightarrow$ E91 mapping

| Component | BB84 (exp_5) | E91/DIQKD (this work) |
|---|---|---|
| Protocol family | Prepare-and-measure (4 BB84 states) | Entanglement-based, CHSH game |
| Channel | Bit-flip on the travelling qubit before Eve, rate $p_{bf}$ | Bilateral depolarising $\Lambda_p^A \otimes \Lambda_p^B$ after Eve's preparation |
| Honest reference | $F_{AB} = 1 - p_{bf}/2$ | $S_{\text{obs}} = (1-p)^2 \cdot 2\sqrt{2}$ (exact) |
| Detection threshold (episode mean) | $F_{AB} \geqslant 0.87$ | $S_{\text{obs}} \geqslant S_{\text{sec}} + 0.02 \approx 2.444$ |
| Eve's information | $F_{AE}$ (Alice–Eve fidelity) | $\chi(A:E)$ (Holevo information) |
| OU drift ($\kappa = 0.1$) | $\mu = 0.10, \sigma = 0.03, p \in [0.01, 0.30]$ | $\mu = 0.03, \sigma = 0.015, p \in [0, 0.0642]$ |
| Cliff | $p = 2(1 - F_{\text{thresh}}) = 0.26$ | $p_{\text{cliff}} \approx 0.0742$ |

Table 10: Component-by-component mapping from BB84 (§4.3) to E91/DIQKD.

## K.2 Bilateral depolarising channel: necessity and Pauli transfer

The bilateral depolarising channel $\Lambda_p^A \otimes \Lambda_p^B \otimes \mathbb{I}_E$ acts on Eve's tripartite preparation $|\psi_{ABE}\rangle$. In the Pauli basis, $\Lambda_p$ damps every non-identity Pauli operator by $(1-p)$ while leaving the identity

untouched. Tracing out $E$ yields Alice and Bob's marginal CHSH value,

$$S_{\text{obs}}(U,p) \;=\; (1-p)^2 \cdot S_{\text{pre}}(U), \tag{2}$$

because the CHSH expectation $E(a,b) = \mathrm{Tr}[\rho_{AB}\, O(a) \otimes O(b)]$ is bilinear in the non-identity $A$-Pauli and $B$-Pauli operators (as $O$ is traceless), causing each CHSH term to pick up two damping factors.

**Why both channels.** Eve's information $\chi(A:E)$ depends only on $\rho_E$ and $\rho_{E|X}$. Tracing out $B$ retains only the identity-on-$B$ Pauli coefficients, which $\Lambda_B$ does not damp:

$$\rho_E^{\text{obs}} = \rho_E^{\text{pre}} \quad \text{independent of } \Lambda_B. \tag{3}$$

A numerical check confirms that $\chi$ remains identical to four decimal places across $p \in \{0, 0.05, 0.1, 0.2, 0.5\}$ when only $\Lambda_B$ is applied. Thus, while $\Lambda_B$ leaves $\chi$ unchanged directly, it tightens Eve's CHSH budget through the $(1-p)^2$ factor above. By contrast, $\Lambda_A$ reduces $\chi$ via the data-processing inequality (DPI) on Alice's measurement outcome, giving $\chi_{\text{obs}} \leqslant \chi_{\text{pre}}$ with strict inequality for $p > 0$. The two channels therefore constrain Eve through *separate* mechanisms: without $\Lambda_B$, the CHSH value would degrade by a single factor of $(1-p)$, halving the noise-induced loss of $S$ at small $p$ and roughly doubling the noise level at which the honest link fails.

### K.3 Detection threshold and cliff

With $Q = \frac{1}{2}(1 - S/2\sqrt{2})$ denoting the error rate of the honest depolarised pair, the DIQKD asymptotic key rate $r(S) = 1 - h(Q) - h\Big(\frac{1}{2} + \frac{1}{2}\sqrt{(S/2)^2 - 1}\Big)$ vanishes numerically at $S_{\text{sec}} \approx 2.4240$. Adding a margin of $0.02$ as a finite-statistics buffer gives the operational detection threshold:

$$\text{DET_THRESH} = S_{\text{sec}} + 0.02 \approx 2.444. \tag{4}$$

The honest Bell pair just meets the security threshold at the noise cliff

$$p_{\text{cliff}} = 1 - \sqrt{S_{\text{sec}}/2\sqrt{2}} \approx 0.0742. \tag{5}$$

In the experiments, the OU drift ($\mu = 0.03, \sigma = 0.015$) is clipped to $p \leqslant 0.0642$, sitting $0.01$ below $p_{\text{cliff}}$, while the honest CHSH value reaches the operational abort threshold at $p \approx 0.070$.

### K.4 EvoluCMAES library for E91

The library spans 7 absolute $S$ targets across $4$ noise anchors:

- Levels: `safer/safer2/safe/medium1/medium/medium2/agg` $= (2.55, 2.50, 2.43, 2.30, 2.20, 2.00, 1.80)$
- Anchors: $p \in \{0.000, 0.0214, 0.0428, 0.0642\}$

Skipping cells where $S_{\text{target}} > s_{\text{honest}}(p) - 0.01$ leaves 26 feasible cells. Each cell is built by the EvoluCMAES outer loop using gate-set mutation over $\{R_X, R_Y, R_Z, H, CNOT\}$ on three qubits (initialised from random six-gate circuits) and an SPSA inner loop over the rotation angles (population 30, 25 generations, 40 SPSA iterations, keeping the better of two seeds per cell). Fitness is defined as $\chi - \alpha\,(S - S_{\text{target}})^2$ with $\alpha = 100$, evaluated after the channel at the anchor noise with Alice and Bob using the standard CHSH settings. The resulting cells contain 5–9 gates (median 6). Following the convention of the envelope in Acín et al. [2007], the library constrains only the CHSH value, which is weaker than constraining the full set of outcome probabilities.

### K.5 Per-experiment headline tables (full method ablation)

Tables 11–13 report the complete ablation across all RL methods evaluated on E91, including the slower baselines (vanilla PPO and RecurrentPPO) omitted from the figures in Appendix I for readability. Unpaired standard errors are approximately $0.009$, $0.007$, and $0.005$ for the 50-, 100-, and 200-episode evaluations. Because all policies within each table share the same channel realisations, paired comparisons resolve smaller differences (Table 8).

| Method | Mean $\chi$ | Std $\chi$ | Mean $S$ | Det % | Distinct actions |
|---|---|---|---|---|---|
| PPO (vanilla, 500k) | 0.2563 | 0.033 | 2.522 | 0.0 | 3 (collapsed) |
| BC + iter-DAgger of Oracle | 0.3496 | 0.068 | 2.446 | 4.0 | 13 |
| DP (coarse grid) | 0.3497 | 0.066 | 2.449 | 4.0 | 19 |
| Oracle (greedy) | 0.3473 | 0.069 | 2.446 | 0.0 | 17 |
| Identity | 0.0000 | 0.000 | 2.659 | 0.0 | 1 |

Table 11: exp_5_E91: library + PPO + BC (50-episode evaluation). DP is computed via backward induction on a coarse grid of the accumulated CHSH statistic with a finite terminal penalty. The reported value is its rollout mean, which varies across the appendix tables within sampling error.

| Method | Mean $\chi$ | Std $\chi$ | Mean $S$ | Det % | Distinct actions |
|---|---|---|---|---|---|
| RecurrentPPO (LSTM) | 0.2801 | 0.033 | 2.508 | 0.0 | 4 |
| MPC (1-step + Oracle rollout) | 0.3511 | 0.068 | 2.448 | 0.0 | 15 |
| DP (coarse grid) | 0.3520 | 0.068 | 2.448 | 6.0 | 24 |
| Oracle | 0.3495 | 0.071 | 2.446 | 0.0 | 17 |
| Identity | 0.0000 | 0.000 | 2.660 | 0.0 | 1 |

Table 12: exp_6_E91: RecurrentPPO + MPC (100-episode evaluation, where MPC inference is the computational bottleneck). The DP policy is computed identically to Table 11, with the rollout mean ($0.3520$ versus $0.3497$) differing only due to the rollout sample.

| Method | Mean $\chi$ | Std $\chi$ | Mean $S$ | Det % | Distinct actions |
|---|---|---|---|---|---|
| MaskablePPO (feasibility mask) | 0.3467 | 0.065 | 2.448 | 0.0 | 14 |
| Rainbow-DQN (Double + Dueling + n-step + mask) | 0.3484 | 0.065 | 2.447 | 0.0 | 14 |
| DP (coarse grid) | 0.3540 | 0.063 | 2.448 | 7.0 | 24 |
| Oracle | 0.3507 | 0.067 | 2.446 | 0.0 | 17 |
| Identity | 0.0000 | 0.000 | 2.661 | 0.0 | 1 |

Table 13: exp_7_E91: MaskablePPO + Rainbow-DQN (200-episode evaluation). The DP policy is computed identically to Tables 11 and 12, with the rollout mean ($0.3540$ versus $0.3497$ and $0.3520$) reflecting rollout sampling across $N \in \{50, 100, 200\}$. The E91 rows of Table 3 and the paired tests of Table 8 use this run.

| Method | 4 levels (15 cells) | 7 levels (26 cells) | $\Delta$ |
|---|---|---|---|
| DP (coarse grid) | 0.3477 | 0.3497 | $+0.002$ |
| BC (Oracle) | 0.3459 | 0.3496 | $+0.004$ |
| Oracle | 0.3415 | 0.3473 | $+0.006$ |
| Vanilla PPO | 0.2520 | 0.2563 | $+0.004$ |

Table 14: Doubling library granularity from 4 levels to 7 levels across 50-episode evaluations (standard error $\approx 0.01$). All differences lie within sampling error, while the DP rows incur $14\%$ (4 levels) and $4\%$ (7 levels) detection.

### K.6 Library granularity ablation

Within sampling error, the DP, BC, and Oracle values remain insensitive to library granularity. Vanilla PPO collapses at both granularities ($\chi \approx 0.25$ at 4 levels and $\chi \approx 0.26$ at 7 levels), showing that the qualitative collapse persists across action-set sizes.

### K.7 Acín bound and the noisy ceiling

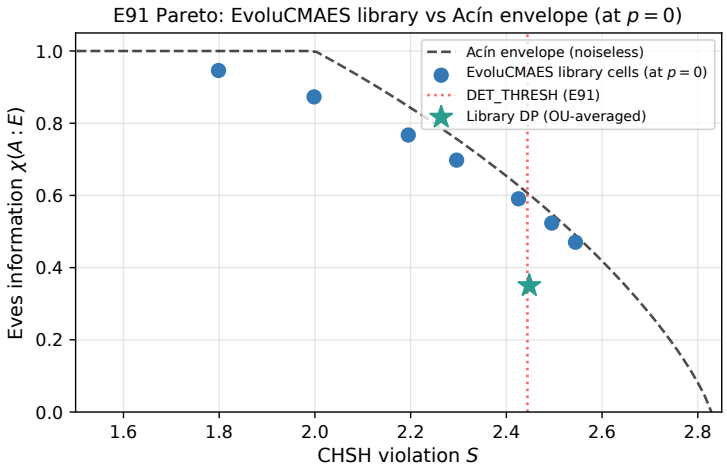



Figure 16: E91 Pareto front: EvoluCMAES library cells at $p = 0$ versus the noiseless Acín envelope. Library cells lie close to the analytical envelope for $S \geqslant 2.2$ (with small SPSA slack), whereas the OU-averaged library DP value (green star) sits below the envelope at $S = \text{DET_THRESH}$ because the bilateral depolarising channel restricts Eve at $p > 0$ and the library attacks use a fixed three-qubit preparation.

The library DP value (0.350) sits well below the noiseless Acín–Brunner–Gisin–Massar–Pironio–Scarani analytical envelope [Acín et al., 2007, Pironio et al., 2009], which bounds $\chi$ for collective attacks of arbitrary dimension:

$$\chi \leqslant h\left(\tfrac{1}{2} + \tfrac{1}{2}\sqrt{(S/2)^2 - 1}\right) \approx 0.607 \quad \text{at } S = \text{DET_THRESH}.$$

This gap stems from two factors: first, the analytical envelope is evaluated without channel noise, whereas the bilateral depolarising channel restricts the states Eve can present to Alice and Bob at $p > 0$; second, the library attacks use a fixed three-qubit preparation, whereas the envelope admits arbitrary dimensions. Consequently, the gap reflects these physical and dimensional constraints, without indicating any looseness in the envelope.

A direct *numerical* Pareto search using library-warm-started SLSQP over the full three-qubit pre-channel state (16 real parameters) and the OU stationary distribution yields an episode-mean $\chi \approx 0.36$ for this dimension-restricted attacker, closely matching the library DP value. While the agreement across three independent methods (library DP $= 0.350$, BC of DP $= 0.359$, and SLSQP Pareto + LP $= 0.36$) is suggestive, it stops short of a rigorous upper bound. Establishing a certified bound would require an NPA-SDP relaxation for noisy DIQKD [Navascués et al., 2008, Brown et al., 2024], which is not run here (a noiseless NPA relaxation for Eve's guessing probability was run solely as an implementation check against the analytical envelope).

Figure 16 overlays the EvoluCMAES library cells at $p = 0$ (no channel) on the Acín envelope. For $S \geqslant 2.2$, the library cells lie close to the noiseless envelope, whereas the OU-averaged DP value (green star) sits lower due to the bilateral channel restriction at $p > 0$ and the fixed preparation dimension. Overall, the engineered library captures the value attained by this numerical search for a three-qubit preparation constrained solely by the CHSH value.

## L Compute

The UCI pruning and QKD experiments ran on a single workstation with a 16-core x86-64 CPU (no GPU), while the preliminary MNIST sweep ran on a laptop (Intel Core i7-10875H, 16 threads, 16 GB). The teacher–student sweep, the variance grid, the drift-neutral 400-epoch runs, and one density-agnostic 400-epoch run were dispatched as separate single-node jobs across four lab workstations (Intel Core i7-13700, 24 threads, 16 GB, no GPU, using 14 or 20 worker processes per job as recorded in the job logs). The inner-optimiser ablation, the selection-off nulls, the remaining four density-agnostic 400-epoch runs, and all runs added for the camera-ready version executed on the laptop. Inner-loop arithmetic is dominated by small forward passes (UCI and MNIST networks with $\leqslant 10^4$ parameters) and 2–3-qubit density-matrix simulations (QKD circuits), neither of which benefits from GPU acceleration at these scales. No individual run was distributed across nodes.

**Per-experiment wall-clock.** The following wall-clock times were recorded on the reference hardware from the run logs that produced the reported results:

- **UCI NN pruning** (Tables 1 and 2; `NNpruning_experiment_2/3/4`): 5 seeds $\times$ 6 methods $\times$ 3 datasets, requiring $\sim$ 6 CPU-hours in total.
- **Appendix E density-agnostic density traces**: extracted as a subset of the `NNpruning_experiment_4` runs above at no additional compute cost.
- **Appendix F preliminary MNIST budget sweep**: 10 budgets $\times$ 3 seeds at $N_{\text{epoch}} = 60$, plus a 4-density convergence-horizon measurement, totalling 12.9 h of wall-clock time across runs ($\sim$ 103 core-hours).
- **Teacher–student sweep, 2,700-run variance grid, MNIST inner-optimiser ablation, and 400-epoch operator study** (§4.1, Appendices B–D): 40.8 h of cumulative wall-clock time ($\sim$ 605 core-hours), excluding the variance grid, which adds 0.7 h ($\sim$ 10 core-hours).
- **Runs added for the camera-ready version** (drift-neutral rows of Table 2 and Appendix E; the nulls and population grid of Appendix D; the drift-neutral sweep of Appendix F; Appendix G and the timings in §4.2; Adam at $B = 256$): run on the laptop (mostly with 8 worker processes), requiring $\sim$ 45 h of wall-clock time ($\sim$ 354 core-hours), of which the drift-neutral MNIST sweep took 22.4 h and the population grid took 19.7 h.
- **BB84 EvoluCMAES library** (§4.3): 24 library cells at $\sim$ 1 CPU-hour per cell, totalling $\sim$ 20 CPU-hours.
- **E91/DIQKD EvoluCMAES library** (§4.3, Appendix K.4): 26 library cells under the EvoluSPSA inner loop (population 30, 25 generations, 40 SPSA iterations, two seeds per cell), taking $\sim$ 2–4 CPU-hours per cell and $\sim$ 60–100 CPU-hours in total.
- **Appendix J BB84 inner-loop comparison**: 3 optimisers (CMA-ES, SPSA, and random search) at a matched per-individual evaluation budget across 5 seeds each (the notebook producing Table 9 did not log wall-clock times).
- **BB84 RL training** (Table 3 BB84 rows; Appendix I): $\sim$ 8–12 CPU-hours per Rainbow-DQN run at 1M training steps ($\sim$ 24 CPU-hours combined for the headline strict and tolerated variants), plus $\sim$ 8–12 CPU-hours for the PPO and RecurrentPPO+MPC baselines.
- **E91 RL training** (Table 3 E91 rows; Tables 11–13): $\sim$ 4–10 CPU-hours per method, totalling $\sim$ 30–50 CPU-hours across the full ablation set (vanilla PPO, BC+iter-DAgger, RecurrentPPO, MPC+Oracle rollout, MaskablePPO, and Rainbow-DQN).
- **Numerical Pareto and NPA-SDP bound checks** (Appendix K.7): SLSQP-based numerical Pareto search over the three-qubit pre-channel state (16 real parameters) and a noiseless `ncpol2sdpa`-based NPA implementation check, taking $\sim$ 4–8 CPU-hours in total.

**Total.** Across the categories above, all experiments outside §4.1 and the camera-ready additions reproduce in roughly 150–250 CPU-hours on the reference hardware. Peak memory reached 1–2 GB during the heavier RL training runs and stayed within 500 MB for all other tasks. The released code is designed to run on commodity laptops without high-memory or accelerator-specific hardware.

**Failed and preliminary runs.** The figures above cover only the runs that produced the *reported* values. Earlier exploratory runs during development (testing alternative seed counts, $B$ ranges, and mutation operators) required an estimated additional 50–100 CPU-hours on the same hardware. Those preliminary runs are omitted from the released artefact, which retains only the final-configuration pickles in `results/`.

## M Broader Impact and Ethical Considerations

Because the two domains studied in this paper carry distinct societal profiles, each is addressed separately below.

**Optimisation methodology.** The natural-growth-bounds phenomenon and the EvoluCMAES framework are domain-neutral contributions to gradient-free joint topology–parameter search. Their primary positive impact lies in neural architecture search and pruning: understanding why fixed-budget search yields sparse solutions without explicit regularisation can help reduce the energy footprint of architecture search pipelines, which currently rely on heavy regularisation schedules and large compute budgets to find sparse models. The optimisation framework itself presents no apparent direct misuse vector.

**Dual-use considerations of the QKD application.** By construction, the eavesdropping circuits discovered in §4.3 and Appendix K represent attack strategies against quantum key distribution protocols. Three factors bound the realistic misuse risk of these findings and support open publication.

First, the experiments operate strictly at the analytical-reference scale: the Eve circuits use a one-qubit memory (two-qubit circuits for BB84 and three-qubit preparations for E91) over idealised bit-flip and bilateral depolarising channels, approaching established analytical bounds from the public literature (the Pauli-channel cloning machine of Cerf [2000], the optimal individual-attack bounds of Fuchs et al. [1997], Bruß [1998], and the Acín–Brunner–Gisin–Masanes–Pironio–Scarani DIQKD envelope [Acín et al., 2007, Pironio et al., 2009]). Rather than exceeding these long-standing bounds, the results show that a generic gradient-free optimiser *rediscovers* circuits in their neighbourhood from a small gate alphabet, leaving the information-theoretic ceiling for a deployed QKD attacker unchanged. Moreover, the reported values represent constructive lower bounds for a restricted attacker under a trusted-noise convention, leaving formal security proofs (which attribute all observed disturbance to the eavesdropper and extend to adaptive strategies [Metger and Renner, 2023]) fully intact.

Second, a wide gap separates simulated two-qubit Eve circuits under idealised noise from a practical attack on a deployed QKD link. Real-world vulnerabilities are dominated by physical-layer factors unmodelled here, including detector blinding and side-channel attacks on hardware receivers, finite-key statistics, decoy-state intensity calibration, source imperfections, and authenticated classical post-processing. Historically, practical attacks on deployed QKD systems have exploited these physical-layer hardware channels (*e.g.* bright-light detector control) as opposed to improved cloning-circuit constructions in the asymptotic regime, placing the contributions of this paper outside the deployment-relevant attack surface.

Third, the defensive value of this analysis outweighs its offensive utility. Adaptive eavesdropping under time-varying channel noise forms the exact threat model that guides QKD operators in setting detection thresholds and finite-key buffers (such as the $S_{\text{sec}} + 0.02$ buffer used in Appendix K). Quantifying what an adaptive attacker can extract under realistic channel drift (and how closely that attacker approaches the analytical ceiling) provides the concrete baselines needed to set operational parameters defensively. This follows the standard cryptographic practice of publishing attack analyses to tighten deployment guidance, just as optimal-eavesdropping bounds have long informed QKD security [Fuchs et al., 1997, Bruß, 1998, Scarani et al., 2009].

**Disclosure choice.** In line with standard academic cryptography practice, the methodology and results are published openly. Restricted release would be warranted if the findings breached an analytical bound, exploited an unpatched implementation, or could be deployed against a fielded system without substantial further engineering. None of those conditions apply here: the circuits remain at or below existing analytical bounds, target no specific hardware implementation, and would require extensive physical-layer engineering to threaten a deployed link. Conversely, open publication

directly benefits defenders by allowing QKD operators to stress-test their threshold choices and enabling the wider community to reproduce, critique, and extend the analysis.

**Alignment with the post-quantum-cryptography transition.** QKD forms part of a broader portfolio of quantum-resistant key-distribution mechanisms whose comparative security margins remain under active evaluation. Concrete analyses of noise-aware adversaries contribute directly to that evaluation and complement parallel efforts to deploy lattice-based post-quantum public-key schemes.

**Environmental and access considerations.** Because the experiments rely on small setups (UCI tabular models, two- and three-qubit simulators, and CPU-only training), they avoid the energy and compute-access barriers associated with large-scale model training. Being gradient-free and runnable on commodity hardware, EvoluCMAES lowers the compute threshold for follow-on research.

**Limitations of this analysis.** While not every downstream application of the optimisation framework or eavesdropping methodology can be anticipated, researchers extending the QKD-attack direction should weigh release decisions against the deployment context of any target protocol and treat any future result that breaches an analytical ceiling as a candidate for coordinated disclosure.

**Release safeguards.** The public repository contains only the optimisation framework and the simulation harness for the idealised channels studied here, omitting hardware drivers for deployed QKD receivers, side-channel exploitation primitives, and finite-key post-processing attack code. Operating at the analytical reference scale standard in the public quantum cryptography literature [Cerf, 2000, Fuchs et al., 1997, Bruß, 1998, Acín et al., 2007], the released Eve circuits and channel models reproduce existing analytical bounds without exceeding them. Should future work built on this framework breach an analytical ceiling or expose an implementation-specific vulnerability in a deployed system, the authors commit to engaging with the QKD-defender community through standard coordinated-disclosure channels.

**Code and reproducibility.** The full EVOLUCMAES framework, the BB84 and E91/DIQKD experimental harnesses, the Rainbow-DQN training and evaluation scripts, and the analysis notebooks generating every table and figure in this paper are released under the MIT licence – click here. The repository includes a top-level `README` mapping each notebook to its corresponding table or figure, the random seeds for all reported runs, the conda environment specification, and the expected wall-clock times on the reference hardware.

## N Use of Large Language Models

Large language models (LLMs) were used in this study to assist with code implementation, figure generation, and manuscript editing.

## NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and Introduction (§1) state the four contributions delivered: (i) the empirical natural growth bounds phenomenon, tested on a teacher–student benchmark where the equilibrium responds to the budget at independently measured variance thresholds and reproduced under four inner optimisers, and observed at $\rho \approx 0.13$ for UCI pruning and 4–6 gates for BB84 cloning; (ii) the EvoluCMAES framework with the mutation operator and the fitness as its domain-specific components; (iii) target-density control at $95\%$ sparsity on the UCI suite by the mutation operator alone; and (iv) the library-as-action-basis transfer where Rainbow-DQN recovers $98.8\%$ (BB84) and $98.4\%$ (E91/DIQKD) of a dynamic-programming reference. The Introduction explicitly demarcates mechanism-level claims from run-level claims, flags the small experimental scale (UCI tabular, one-qubit Eve memory), and states that the QKD results concern a restricted attacker and make no statement about security proofs, so the headline claims do not overreach the evidence.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: §5 ("Discussion and Limitations") is dedicated to this. It states (a) that Proposition 1 bounds the equilibrium $|T|$ below dense without pinning down where, that the budget-gated threshold is measured and has no closed form, and that the equilibrium is not unique; (b) that the pruning experiments are at small UCI scale, on which baselines (SNIP, SynFlow, IMP) were not originally validated, that scaling is limited by the parameter dimension, and that UCI and MNIST accuracies are measured on the split used for selection; (c) that quantum-circuit discovery is demonstrated only with a one-qubit Eve memory and that the QKD values are constructive lower bounds under a trusted-noise convention; and (d) that the Rainbow-DQN result has not been replicated for MaskablePPO on BB84. The first paragraph of the Limitations also names the tests that would probe the Corollary and assumption A1 further.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [No]

   Justification: The paper's only formal results are Proposition 1 and Corollary 1 (§3.3). Their assumptions A1 (per-iteration progress rate of $(\mu/\mu_w, \lambda)$-CMA-ES decreasing in $\dim(\theta)$ at fixed $B$, with citations to Hansen et al. [2003], Hansen [2016]) and A2 (rank-based selection with strict noisy-comparison preference) are stated in the proposition body. Appendix A provides a heuristic proof sketch, explicitly labelled as such, and its "Scope of the sketch" paragraph lists the steps that are not formally established, so a complete proof is not provided; both statements are presented as qualitative and are tested empirically (§4.1).

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: §4.1 and Appendices B–D describe the teacher–student benchmark, the budget and density grids, the inner optimisers and the mutation operators. §4.2 specifies the datasets (Iris, Wine, Breast Cancer from UCI), the split, the target density ($\rho = 0.05$), the inner loops, the five comparison baselines, and the seed count (five) for the pruning experiments, and references the underlying notebooks (`NNpruning_experiment_2.ipynb`, `_3.ipynb`, `_4.ipynb`) and scripts; Appendix G gives the four further tabular datasets and both selection

protocols. §4.3 describes the BB84 setup (bit-flip noise before Eve's interaction, gate alphabet $\{R_X, R_Y, R_Z, CNOT\}$ with add/remove/replace primitives, fitness following Decker et al. [2025]). Appendix K provides the E91 channel structure, the detection threshold derivation ($S_{\text{sec}} \approx 2.4240$, DET_THRESH $\approx 2.444$), the OU drift parameters ($\mu = 0.03$, $\sigma = 0.015$ for E91 and $\mu = 0.10$, $\sigma = 0.03$ for BB84), and the library construction (7 absolute $S$ targets $\times$ 4 noise anchors, 26 feasible cells, $\alpha = 100$, SPSA inner loop). Appendix I specifies the RL setup (Rainbow-DQN with feasibility masking, matched hyperparameters across protocols, 50 BB84 and 200 E91 evaluation episodes), and Appendix K provides per-method ablation tables. The code release (see Q5) provides the random seeds and configurations used for the reported runs.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The full EvoluCMAES framework, the BB84 and E91/DIQKD experimental harnesses, the Rainbow-DQN training and evaluation scripts, and the analysis notebooks that produced every table and figure are released under the MIT licence (Appendix M). The repository contains a top-level README mapping each notebook to the table or figure it generates, the random seeds used for the reported runs, an environment specification, per-experiment wall-clock estimates on the reference hardware, and a script that regenerates the reported values from the stored results. The benchmark datasets (Iris, Wine, Breast Cancer Wisconsin, Seeds, Glass, Ionosphere, Sonar, MNIST) are public; the QKD experiments use synthetic channel simulations with all parameters specified in Appendix K.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: For pruning, §4.2 and Appendix D state the split, the inner optimisers (Adam for Table 1, CMA-ES for Table 2), the outer-loop settings, the seed count (five), the target density, and the baselines compared. For quantum-circuit discovery, §4.3 and Appendix K state the gate alphabets, the mutation primitives, the fitness functions with their $\alpha$ value, the three-qubit preparation, the channel structure (bit-flip for BB84, bilateral depolarising for E91), the OU drift parameters, the detection thresholds, and the library-cell anchors. For the RL extension, Appendix I states the agent (Rainbow-DQN with feasibility masking), the strict and tolerated checkpoint-score weights ($\kappa = 2.0$ and $\kappa = 0.5$), the training horizon (1M steps on BB84), and the evaluation protocol (50 BB84 and 200 E91 episodes, best-checkpoint policy).

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Tables 1 and 2 report mean $\pm$ standard deviation over five seeds for every entry, the teacher–student and inner-optimiser figures report seed-level dispersion (five and three seeds), and the captions state how the dispersion is computed. The RL comparisons use paired tests over shared evaluation episodes: a sign test and a paired interval for the BB84 tolerated-versus-greedy-Oracle gap (50 episodes, Appendix I) and Wilcoxon signed-rank tests with paired intervals on E91 (200 episodes, Table 8). Where statistical significance is not established at the level the reader might want – e.g. the cross-domain match of equilibria, which is qualitative across two domains rather than tested with a hypothesis test – the paper says so explicitly in §5.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: Appendix L reports the hardware used for each experiment family (CPU type, memory, and wall-clock time) and the total compute footprint for the runs reported in the paper, including preliminary and failed runs not reported. The code release (Q5) reproduces the same wall-clock estimates per experiment as expected reference values for re-runners.

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The work uses only public, anonymised benchmarks (Iris, Wine, Breast Cancer Wisconsin, Seeds, Glass, Ionosphere and Sonar from UCI, and MNIST) and synthetic quantum-channel simulations; no human subjects are involved, no scraped data is used. The QKD-eavesdropping content is academic offensive-security research on idealised channels, of the same kind routinely published in the quantum-information-theory literature, and is grounded in well-known analytical references (Pauli-channel cloning machine, optimal BB84 attacks, the Acín DIQKD bound). Dual-use considerations are addressed explicitly in Appendix M.

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: Appendix M ("Broader Impact and Ethical Considerations") is dedicated to this. It discusses positive impacts on optimisation methodology (a clearer mechanistic understanding of why fixed-budget joint search produces sparse solutions, with energy-footprint implications for NAS pipelines) and addresses the dual-use nature of the QKD eavesdropping content directly. The dual-use discussion bounds realistic misuse risk on three axes – the analytical-reference scale of the experiments (one-qubit Eve memory, idealised bit-flip and bilateral-depolarising channels, targets approaching pre-existing public bounds rather than exceeding them, and values that are lower bounds for a restricted attacker under a trusted-noise convention), the gap between simulated and deployed-link attack surfaces (detector blinding, finite-key statistics, source-side imperfections, and authenticated post-processing all unmodelled here), and the defender-side value of clean adaptive-attack analysis for setting detection thresholds and finite-key buffers. The disclosure choice is framed in the coordinated-vulnerability-disclosure tradition standard to academic cryptography, and the work is positioned as complementary to the broader transition to post-quantum cryptography.

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [Yes]

    Justification: The "Release safeguards" paragraph of Appendix M states what is and what is not included in the released code: the optimisation framework and the simulation harness for the idealised channels studied are released, while hardware drivers for deployed QKD receivers, side-channel exploitation primitives, and finite-key post-processing attack code are not. The released circuits operate at the analytical-reference scale standard in the public quantum-cryptography literature [Cerf, 2000, Fuchs et al., 1997, Bruß, 1998, Acín et al., 2007] and reproduce rather than exceed pre-existing analytical bounds. The authors commit to engaging with the QKD-defender community via standard coordinated-disclosure channels should follow-on work using this framework yield results that breach an analytical ceiling or expose an implementation-specific weakness in a deployed system.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All third-party assets are cited at point of use: the UCI datasets (CC BY 4.0) are cited to their creators [Fisher, 1936, Aeberhard and Forina, 1992, Wolberg et al., 1993, Charytanowicz et al., 2010, German, 1987, Sigillito et al., 1989, Sejnowski and Gorman, 1988] and loaded through scikit-learn, the last four from OpenML [Vanschoren et al., 2014], and MNIST is cited to LeCun et al. [1998]; the algorithmic baselines have explicit citations (magnitude pruning [Han et al., 2016], IMP [Frankle et al., 2020], SNIP [Lee et al., 2019], SynFlow [Tanaka et al., 2020]); CMA-ES [Hansen, 2016, Hansen et al., 2003], SPSA [Spall, 1992], Adam [Kingma and Ba, 2014] and Rainbow-DQN [Hessel et al., 2018] with feasibility masking [Huang and Ontañón, 2022] are cited; PPO [Schulman et al., 2017] and DAgger [Ross et al., 2011] are cited; the QKD references (BB84 [Bennett and Brassard, 1984], E91 [Ekert, 1991], DIQKD bounds [Acín et al., 2007, Pironio et al., 2009], Pauli-channel cloning [Cerf, 2000], optimal-BB84-eavesdropping bounds [Fuchs et al., 1997, Bruß, 1998], and the learnt-attack formulation of Decker et al. [2025]) are cited at point of use. The terms of use of every asset (public benchmarks, open-source algorithm references, peer-reviewed analytical bounds) are respected.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The released code is accompanied by a top-level README that maps each notebook and script to the table or figure it generates in the paper, the random seeds used for the reported runs, an environment specification, and per-experiment wall-clock estimates on the reference hardware reported in Appendix L. The released framework (EvoluCMAES) is also described algorithmically in §3 (Algorithm 1), and the QKD-experimental instantiations are described in Appendices K and I, so the code is documented at both the per-script level (in the repository) and the per-experiment level (in the paper).

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The work involves no crowdsourcing and no research with human subjects. The experiments use only public benchmarks (UCI Iris, Wine, Breast Cancer Wisconsin, Seeds, Glass, Ionosphere and Sonar, and MNIST) and synthetic quantum-channel simulations.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The work involves no human subjects, so IRB or equivalent review is not applicable.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [Yes]

    Justification: LLMs were only used for code assistance, writing, and editorial formatting (Appendix N). They were not part of the research methodology, experimental design, or scientific core of the paper.