AICR · OPEN PROBLEM

Collision Resistance of the Symmetric Binary Perceptron

An intermediate window of κ where neither the algorithms nor the hardness arguments reach

Status. Statement: AI-written from Marco Benedetti, Andrej Bogdanov, Enrico M. Malatesta, Marc Mézard, Gianmarco Perrupato, Alon Rosen, Nikolaj I. Schwartzbach and Riccardo Zecchina, *Collision Resistance of Single-Layer Neural Nets*, IACR ePrint 2026/1143. Not yet checked by a human. The source poses this as a displayed "Open Question" and calls it "a central open challenge". It gives an efficient online collision finder for small κ and an overlap-gap-based online lower bound for a *different*, randomized activation — introduced precisely because the SBP's own collision space resists a first-moment analysis.

Category. *Fine-Grained Cryptography.*

Conjecture (informal). *A single-layer neural net with random weights compresses its input, so collisions exist. Whether they can be found efficiently depends on the activation function. If the activation flattens out on a positive tail the answer is yes and the source gives the algorithm: drive every margin above the flat region and the outputs agree. The symmetric binary perceptron does not flatten — it is a window indicator — and for very large window width collisions become trivial for a different reason. Between those two regimes nothing is known: no algorithm reaches, and no hardness argument applies. The question is whether that middle window is where a collision-resistant hash function lives.*

1 The setting

Fix a random matrix A and an activation φ, and consider the map $x \mapsto \varphi(Ax)$ on binary inputs. Collision resistance here means the hardness of finding two distinct inputs with the same output. Two parameters govern the picture: the density $\alpha = m/n$, and the activation's threshold κ.

The source's two results bracket the question from opposite sides.

Positive. For activations that are constant on a positive tail $[\kappa, \infty)$, collision finding is easy at low densities: it suffices to drive all margins above κ, after which the activation is flat and the outputs coincide. The source gives an efficient online algorithm doing this, at fixed positive algorithmic rates, for $\alpha < \alpha_0$ and $\kappa \le c/\sqrt{\alpha}$. The asymmetric binary perceptron falls in this class.

Negative. For a *randomized oscillating* activation — the SBP's fixed window boundary replaced by independent random thresholds at each neuron — the source proves an overlap gap property for extensive collisions and derives an online lower bound. It is explicit about why it changed the model: "This extra randomness removes the rigid alignments that make the SBP collision space difficult to analyze, while preserving the obstruction to the positive-tail algorithm."

The SBP itself sits between. For small κ it falls inside the positive-tail algorithmic window. For very large κ the activation is nearly constant on typical Gaussian margins and collisions become trivial. The source states the gap plainly: "The difficult case is the intermediate window, where the positive-tail escape route no longer applies but no OGP-based hardness result is known for collision finding. Resolving whether SBP is collision resistant in this middle regime, for some $\alpha < 1$ and some κ, remains a central open challenge."

2 Notation and parameters

- n is the input dimension, m the number of neurons, $\alpha = m/n$ the collision ratio; the open regime is $\alpha < 1$, so the map is compressing.
- $\kappa > 0$ is the SBP threshold: the activation is the indicator of $|\cdot| \le \kappa$, a window rather than a tail.
- A collision is δ-*extensive* if the two inputs $x, y \in \{\pm 1\}^n$ satisfy $x^\top y \le (1-\delta)n$ — that is, they are far apart, not near-duplicates.
- The source's online algorithm applies when $\kappa \ll 1/\sqrt{\alpha}$; the trivial-collision regime is κ large enough that the activation is nearly constant on typical margins.

3 The conjecture

Conjecture 1 ([BBMMPRSZ26, §1.3, Open Question]). *For some $\alpha < 1$, there exists an intermediate window of κ with $1/\log n \ll \kappa \ll 1/\sqrt{\alpha}$ in which* SBP *is collision resistant against polynomial-time algorithms.*

Remark 1 (This is the source's own phrasing, and it is a question). The source displays it as a question — "For some $\alpha < 1$, does there exist an intermediate window of κ with $1/\log n \ll \kappa \ll 1/\sqrt{\alpha}$ in which SBP is collision resistant against polynomial-time algorithms?" — and predicts nothing. It is recorded here in the affirmative because a statement needs a direction, not because the source endorses one. The concrete parameters the source names as living in the hard middle are $\kappa \approx 0.675$ and $\alpha \approx 0.9$, "where statistical physics suggests nontrivial structure"; that is the natural first point to attack from either side.

Remark 2 (Why the two known techniques both stop short). This is the useful part of the statement, because both failures are specific. The positive-tail algorithm needs the activation to be constant above κ so that pushing margins past it is enough; the SBP's activation is a window, so pushing margins up eventually leaves the window rather than saturating it. And the OGP argument needs a first-moment computation on the collision space, which the source could not carry out for the SBP — hence the randomized oscillating activation, whose independent per-neuron thresholds destroy the "rigid alignments" that block the analysis. So a resolution needs either an algorithm that does not rely on saturation, or a hardness argument that survives the SBP's rigidity.

Remark 3 (Neighbouring statements in the same paper, both distinct from this one). Its Conjecture 1.4 (Hardness of Collision Finding) states that for fixed constants $K \in \mathbb{N}$, $\kappa > 0$ and $\delta > 0$ there is a threshold $\alpha^\star < 1$ such that for all $\alpha \geq \alpha^\star$, no polynomial-time algorithm finds a δ-extensive collision with non-negligible probability. That is about the randomized oscillating activation, where the source has the OGP and an online lower bound, and it asks to upgrade "online" to "polynomial-time"; the source leaves "confirming this conjecture, or finding algorithmic counterexamples, as an open

direction for future work." Separately, the source notes its formal lower bound is for online algorithms and that "lower bounds for broader algorithmic models, including low-degree, message-passing, and quantum algorithms, remain open." Neither is Conjecture 1, which is about the SBP itself.

Remark 4 (What is on the other side of the ledger). The SBP is susceptible to a second-preimage attack: sample $y = \text{sign}(Ax)$ for random x and solve the resulting teacher-student problem. That is a different attack goal from collision finding, and the source presents it as motivation for care rather than as an obstruction. There is also complementary evidence in the other direction: the source notes work giving SBP hardness from worst-case lattice assumptions, in a regime where the number of variables is polynomially larger than the number of constraints, with the proportional regime $m = \Theta(n)$ explicitly left open — so it "does not directly address collision finding or the intermediate-density SBP window considered here."

Remark 5 (The other algorithmic caveat). The source's practical algorithm is proved correct only at inverse logarithmic algorithmic rates and it leaves its performance at constant algorithmic rates open, using it as a component in a more complicated algorithm that provably works at fixed positive rates for sufficiently small δ and α. It also notes that whether the sharper discrepancy-minimization approach adapts to collision finding "remains open." Progress there would move the algorithmic boundary and could shrink the window in Conjecture 1 from below.

4 Bibliography

[BBMMPRSZ26] Marco Benedetti, Andrej Bogdanov, Enrico M. Malatesta, Marc Mézard, Gianmarco Perrupato, Alon Rosen, Nikolaj I. Schwartzbach and Riccardo Zecchina. *Collision resistance of single-layer neural nets.* IACR ePrint 2026/1143. The source. The framing of the SBP as the limiting case is on page 3, the displayed Open Question on page 5, Conjecture 1.4 on page 7, and Theorem 4.1 (the positive-tail collision finder) in Section 4.

[APZ19] Benjamin Aubin, Will Perkins and Lenka Zdeborová. *Storage capacity in symmetric binary perceptrons.* Journal of Physics A, 2019. Introduces the symmetric binary perceptron. [UNVERIFIED] — the source uses

numeric citations that were not resolved against its reference list during this harvest, so the bibliographic detail here is inferred from the surrounding literature rather than read off the source.

[Gam21] David Gamarnik. *The overlap gap property: a topological barrier to optimizing over random structures.* PNAS, 2021. The survey of the overlap gap property whose framework the source's hardness results sit in. [UNVERIFIED] — bibliographic detail inferred as above.