

Closing the Logarithmic Locality Gap for Almost k -Wise Independent Hashing

Input/output size n , error 2^{-n} , at word size 1 or $\Omega(\log^2 n)$

Status. Statement: AI-written from Yevgeniy Dodis, Shachar Lovett and Daniel Wichs, *Locally Computable High Independence Hashing*, IACR ePrint 2026/622. Not yet checked by a human. Settled up to a $\Theta(\log n)$ factor at word size 1 and at word size $\Omega(\log^2 n)$ and above; the source gives no matching construction at all for intermediate word sizes, so the gap is recorded only where both a bound and a construction exist.

Category. *Idealized Models & Non-Uniformity.*

Conjecture (informal). *Consider hash functions whose evaluations on any k inputs look statistically close to k independent uniformly random outputs — “almost k -wise independent” hashing — keyed by a string stored as a sequence of w -bit words, with the property that evaluating the function on one input reads only a small number t of those words (its locality). For the natural cryptographic regime where the input and output are both n bits, the independence parameter k is polynomial in n , and the target statistical error is as small as 2^{-n} , the source proves an unconditional lower bound on the locality t that holds for every word size w , and gives explicit constructions matching it up to a factor of $\log n$ — but only at word size $w = 1$ and at word size w at least about $(\log n)^2$. For word sizes strictly in between, no matching construction is given at all. The source explicitly records closing this logarithmic gap, in the regimes where it has a construction to compare against, as an open problem it does not resolve.*

1 The setting

Dodis, Lovett and Wichs study *local* hash functions: families $\mathcal{F} = \{f_s : \{0,1\}^n \rightarrow \{0,1\}^w\}_{s \in \Sigma^\ell}$, keyed by a seed s drawn from a word alphabet $\Sigma = \{0,1\}^w$, whose evaluation $f_s(x)$ reads only a small number of the ℓ words making up s . This continues a line initiated by Siegel [Sie89, Sieo4] on locally computable independent hashing,

later advanced by Larsen et al. [LPP+24]. A family is ε -almost k -wise independent if no adaptive distinguisher making at most k oracle queries to f_s , for random s , can tell it apart from a truly random function with advantage more than ε ; this relaxes *perfect* k -wise independence, where the k outputs must be exactly uniform and independent.

The source's central cryptographic regime fixes the output length equal to the input length n , takes an independence parameter $k = \text{poly}(n)$, and drives the statistical error down to $\varepsilon = 2^{-n}$. In this regime the source's Theorem 5.1 proves an unconditional locality lower bound $\Omega(n/(\sqrt{w} \log n))$ that holds for every word size w (given k large enough relative to t, w, n). Matching explicit constructions, achieving locality $O(n/\sqrt{w})$, are given only at $w = 1$ (Theorem 5.7, the bit-local case, where $t = O(n) = O(n/\sqrt{1})$) and at $w = \Omega(\log^2 k + \log^2 n)$ up to $w \leq n^2$ (Theorem 5.10, the word-local case); the source's own results summary states the latter requires a "sufficiently large word size." For the intermediate range $2 \leq w = o(\log^2 n)$, no construction achieving $O(n/\sqrt{w})$, or any explicit construction matching the lower bound, is given. The two bounds that *are* compared, at $w = 1$ and at $w = \Omega(\log^2 n)$, agree only up to a multiplicative $\log n$ factor, and the source's own summary states: "Another open problem is to close the logarithmic gaps between the lower bounds and explicit constructions of almost-independent hashing when the input and output size is n and the error is $\varepsilon = 2^{-n}$."

Separately, and not part of this conjecture, the source also records that it would be desirable to unify the bit-local and word-local constructions into a single construction spanning all $1 \leq w \leq n^2$, including the intermediate range where no construction is currently known at all; that is a different, easier-stated question about elegance of exposition rather than about the value the locality can take, and is not the log-gap question below.

2 Notation and parameters

- n is the common input and output length, in bits.
- $k = k(n)$ is the independence parameter, with $k = \text{poly}(n)$.
- w is the word size, $1 \leq w \leq n^2$; $\Sigma = \{0,1\}^w$ is the corresponding word alphabet.

- ℓ is the key length, in words, so a seed is $s \in \Sigma^\ell$.
- t is the locality: the number of words of s that evaluating $f_s(x)$ reads.
- ε is the statistical distinguishing advantage, fixed to $\varepsilon = 2^{-n}$.

3 The conjecture

Definition 1 (ε -almost k -wise independence). A family $\mathcal{F} = \{f_s : \{0,1\}^n \rightarrow \{0,1\}^n\}_{s \in \Sigma^\ell}$ is ε -almost k -wise independent if, for every (possibly adaptive) distinguisher D making at most k oracle queries,

$$\left| \Pr_{s \leftarrow \Sigma^\ell} [D^{f_s(\cdot)} = 1] - \Pr_R [D^{R(\cdot)} = 1] \right| \leq \varepsilon,$$

where $R : \{0,1\}^n \rightarrow \{0,1\}^n$ is a uniformly random function.

Definition 2 (t -word-locality). A family $\mathcal{F} = \{f_s : \{0,1\}^n \rightarrow \{0,1\}^n\}_{s \in \Sigma^\ell}$ is t -word-local if, for every s and every $x \in \{0,1\}^n$, evaluating $f_s(x)$ reads at most t of the ℓ words of s . The family is *explicit* if $f_s(x)$ can be computed in time polynomial in $n, t, w, \log \ell$ given oracle access to s .

Conjecture 1 (Closing the locality log-gap). Fix the regime where the output length equals the input length n , the error is $\varepsilon = 2^{-n}$, and the independence parameter is $k = \text{poly}(n)$, large enough that Theorem 5.1's standing hypothesis on k relative to t, w, n holds. For a word size $w \in \{1\} \cup [\Omega(\log^2 n), n^2]$ — the range where the source gives both a lower bound and a matching explicit construction — the best locality lower bound and the best locality achieved by an explicit construction currently known for an ε -almost k -wise independent t -word-local family $\mathcal{F} = \{f_s : \{0,1\}^n \rightarrow \{0,1\}^n\}_{s \in \Sigma^\ell}$ (Definitions 1 and 2) are

$$t_{\text{lb}}(n, w) = \Omega\left(\frac{n}{\sqrt{w} \log n}\right), \quad t_{\text{ub}}(n, w) = O\left(\frac{n}{\sqrt{w}}\right),$$

which differ by a multiplicative factor of $\Theta(\log n)$.

Determine the true order of the minimal locality $t^(n, w)$ achievable by an explicit ε -almost k -wise independent t -word-local family in this regime, by resolving one of the following for some $w \in \{1\} \cup [\Omega(\log^2 n), n^2]$:*

- (a) *exhibit an explicit ε -almost k -wise independent t -word-local family with $t = o(n/\sqrt{w})$, improving the construction; or*
- (b) *prove that every ε -almost k -wise independent t -word-local family, explicit or not, must have $t = \Omega(n/\sqrt{w})$, improving the lower bound;*

thereby closing the $\Theta(\log n)$ gap between t_{lb} and t_{ub} for at least one such w .

Remark 1 (Where this statement does and does not apply). For $2 \leq w = o(\log^2 n)$ the source establishes no $O(n/\sqrt{w})$ -locality construction at all, so Conjecture 1 is silent there: there is no known upper bound in that range for a log-factor gap to be measured against. That intermediate range is instead the subject of the source’s separate remark about unifying its bit-local and word-local constructions into one that spans every word size.

4 Bibliography

- [DLW26] Yevgeniy Dodis, Shachar Lovett and Daniel Wachs. *Locally computable high independence hashing*. IACR ePrint 2026/622. The source. The lower bound is Theorem 5.1; the constructions are Theorem 5.7 (bit-local, $w = 1$) and Theorem 5.10 (word-local, $w = \Omega(\log^2 k + \log^2 n)$); the open problem is Section 6 (“Summary and Open Problems”), page 20.
- [Sie89] Alan Siegel. *On universal classes of fast high performance hash function, their time-space tradeoff, and their applications (extended abstract)*. 30th Annual Symposium on Foundations of Computer Science (FOCS), 1989.
- [Sie04] Alan Siegel. *On universal classes of extremely random constant-time hash functions*. SIAM J. Comput., 33(3):505–543, 2004.
- [LPP+24] Kasper Green Larsen, Rasmus Pagh, Giuseppe Persiano, Toniann Pitassi, Kevin Yeo and Or Zamir. *Optimal non-adaptive cell probe dictionaries and hashing*. 51st International Colloquium on Automata, Languages, and Programming (ICALP 2024), LIPIcs vol. 297.