AICR OPEN PROBLEM

Decomposition for Split Unpredictable Sources

Two sources, arbitrary query budget

Status. Statement: AI-written, not yet formalized. The case $q = 0$ is proved and tight; $q = 1$ is proved under an extra hypothesis; $q \geq 2$ is open.

Category. *Idealized Models & Non-Uniformity.*

Conjecture (informal). *Two parties each read an entire random function and each output one input point, together with a little leakage about the function. They never communicate, and neither party's point can be guessed by anyone who sees all of the leakage and the whole function. The conjecture says that, given the leakage, the random function is then close to a mixture of functions that have been fixed in advance on a short list of points, where the list is chosen without any knowledge of the two parties' points. Closeness is measured against an observer who sees the leakage, the value of the function at the pair of points, and a bounded number of adaptive queries, but never the points themselves.*

1 The setting

Presampling, or the bit-fixing model, is the standard route to non-uniform security in idealized models. A random oracle together with S bits of oracle-dependent advice is close to a mixture of oracles fixed on P points and uniform elsewhere, at a cost of roughly ST/P against T-query adversaries [Unruh, DGK, CDGS]. The technique assumes a single preprocessing stage. It is not known how to run it when the advice is produced by *several* parties who read the oracle independently, which is the setting of multi-source randomness extraction in the random-oracle model [CFHS]. That paper obtains its bounds by compression instead, and raises the decomposition question; compression is what stalls for non-monolithic constructions such as Merkle-Damgård and sponge, which is the reason to want a decomposition route.

The obstruction is that the two leakages can be individually harmless and jointly informative, so decomposing the oracle with respect to each leakage separately does not work [CFHS, §1.4]. The statement below therefore asks for a single mixture, selected using the whole oracle and both leakages at once, but forbidden to use the two points; correspondingly the observer sees the value of the function at those points but not the points themselves. Both restrictions are necessary, and the statement is also false with one source in place of two (Remark 4).

Progress to date. The query-free case is proved, with a bound better than the one conjectured here and tight to within a constant factor. The one-query case is proved up to polylogarithmic factors under the extra hypothesis that $M\delta$ is at most polylogarithmic in N, M and $1/\delta$. Queries whose positions do not depend on the challenge value are almost free at every query budget. What is open is $q \geq 2$, and, separately, the regime of large P with growing q (Remark 6).

2 Notation and parameters

For $n \in \mathbb{N}$ write $[n] := \{1, \ldots, n\}$, and let $\{0,1\}^*$ denote the set of finite binary strings, with $|z|$ the length of z in bits. All logarithms are base two. For a finite set $\mathcal{X}$ we write $x \leftarrow \mathcal{X}$ for x drawn uniformly from $\mathcal{X}$, and for a distribution Y we write $w \leftarrow Y$ for w drawn according to Y; each such draw is independent of all others unless stated otherwise. For a predicate E, $\mathbb{1}[E]$ is 1 if E holds and 0 otherwise. An *absolute constant* is one that does not depend on any parameter below, and $O(\cdot)$ and $\Theta(\cdot)$ mean up to such constant factors; no parameter tends to a limit anywhere in this document.

An algorithm A written with a superscript, as A^f, has *oracle access* to the function f: it may submit a point of the domain and receive the value of f there, adaptively, and it learns nothing about f except through such queries. Algorithms are *computationally unbounded*: no bound is placed on their running time, only on their number of queries where one is stated.

Fix throughout:

- $N, M \in \mathbb{N}$, and let $\mathsf{Fun} := \{ f : [N] \times [N] \to [M] \}$. The oracle is $H \leftarrow \mathsf{Fun}$, a uniformly random such function. Its domain is a pair of coordinates, one produced by each source.

- $\sigma_1, \sigma_2 \in \mathbb{N}$, bounds on the number of bits of leakage that each source outputs, and $\sigma := \sigma_1 + \sigma_2$.
- $\delta \in (0,1]$, the unpredictability parameter of Definition 1.
- $q \in \mathbb{N} \cup \{0\}$, the observer's query budget, and $q^+ := q + 1$, the budget together with the one evaluation of the function at the challenge.
- $P \in \mathbb{N}$, the number of points the mixture may fix, and $\gamma \in (0,1)$, an additive slack.
- $\sigma' := \sigma + 2 \log N$, the leakage length together with the number of bits needed to name a point of the domain.

3 The conjecture

Definition 1 (sources, splitness, unpredictability). A *source* is a computationally unbounded randomized algorithm S with oracle access to H, taking no input and using private coins, whose number of queries is unbounded, so that it may read all of H, and which outputs a pair $(x, z) \in [N] \times \{0,1\}^*$. We write $(x, z) \leftarrow S^H$ for one such run.
Fix two sources S_1, S_2. They are *split* if they are run on *independent* private coins, so that, conditioned on H, the pairs they output are independent. Write $(x_i, z_i) \leftarrow S_i^H$ for the output of S_i, with $|z_i| \le \sigma_i$, and put $\mathbf{x} := (x_1, x_2) \in [N] \times [N]$ and $\mathbf{z} := (z_1, z_2) \in \{0,1\}^* \times \{0,1\}^*$.
A *predictor* is a computationally unbounded algorithm P with oracle access to H and unbounded query count, which takes $\mathbf{z}$ as input and outputs an element of $[N]$. The pair (S_1, S_2) is *δ-unpredictable* if

$$\Pr\left[\mathsf{P}^H(\mathbf{z}) = x_i\right] \le \delta \qquad \text{for } i = 1,2$$

and every predictor P, the probability being over $H \leftarrow \mathsf{Fun}$, the coins of S_1 and S_2, and the coins of P. Since a predictor may ignore its inputs and output a fixed element of $[N]$, necessarily $\delta \ge 1/N$.

Definition 2 (bit-fixing sources, and consistency). A distribution Y over Fun is *P-bit-fixing* if there are a set $I \subseteq [N]\times[N]$ with $|I| \le P$ and a map $a : I \to [M]$ such that a draw $g \leftarrow Y$ satisfies $g(u) = a(u)$ for every $u \in I$ and is uniform on $[M]$ and independent across the points $u \notin I$. We call I its *fixed set* and a its *fixed values*, and we say Y is *consistent with* a function $f \in \mathsf{Fun}$ if $a(u) = f(u)$ for every $u \in I$. A *P-mixture* is a finite indexed family $(\lambda_j, Y_j)_{j\in\mathcal{J}}$ with $\lambda_j \ge 0$, $\sum_j \lambda_j = 1$ and each Y_j a P-bit-fixing source; the Y_j are its *components*, and the mixture is *consistent with* f if every component is. A *draw* from a P-mixture means: draw an index J with $\Pr[J = j] = \lambda_j$, then draw $g \leftarrow Y_J$; we write $(J, g) \leftarrow (\lambda, Y)$. The mixture, and not merely the law of g, is the object fixed below: one distribution over Fun can be presented as a P-mixture in many ways, with different fixed sets, and what follows refers to the fixed set of the component actually drawn.

Definition 3 (observers, and the two experiments). An *observer* is a computationally unbounded algorithm D with oracle access, making at most q queries, which takes a pair in $[M]\times(\{0,1\}^*\times\{0,1\}^*)$ and outputs a bit. It is *challenge-oblivious*: its input is $\mathbf{z}$ together with a single value of $[M]$, never $\mathbf{x}$ itself. Given sources S_1, S_2, a family $\mathcal{Y} = \{Y_{f,\zeta}\}$ of P-mixtures indexed by $f \in \mathsf{Fun}$ and $\zeta \in \{0,1\}^*\times\{0,1\}^*$, and an observer D, define two experiments.

- Real: draw $H \leftarrow \mathsf{Fun}$; draw $(x_1, z_1) \leftarrow S_1^H$ and $(x_2, z_2) \leftarrow S_2^H$ on independent coins; set $y := H(\mathbf{x})$; output $D^H(y, \mathbf{z})$.
- Dec: draw H and $(\mathbf{x}, \mathbf{z})$ exactly as in Real, and do not resample them; then draw $(J, H^*) \leftarrow Y_{H,\mathbf{z}}$, set $y^* := H^*(\mathbf{x})$, and output $D^{H^*}(y^*, \mathbf{z})$. Write I_J for the fixed set of the component drawn.

Write $\mathsf{Adv}_{\mathcal{Y},D} := \big|\Pr[\mathsf{Real} = 1] - \Pr[\mathsf{Dec} = 1]\big|$, the probabilities being over every draw listed, including the coins of D.

Conjecture 1 (decomposition for two split unpredictable sources). *There are absolute constants c, C such that the following holds for all N and M, every split δ-unpredictable pair of sources (S_1, S_2) with leakage lengths bounded by σ_1, σ_2, every $P \in \mathbb{N}$ and every $\gamma \in (0,1)$. There is a family*

$$\mathcal{Y} \;=\; \{\, Y_{f,\zeta} \,\}, \qquad f \in \mathsf{Fun}, \quad \zeta \in \{0,1\}^* \times \{0,1\}^*,$$

depending only on S_1, S_2, P and γ, in which every $Y_{f,\zeta}$ is a P-mixture consistent with f, such that for every $q \in \mathbb{N} \cup \{0\}$ and every q-query challenge-oblivious observer D,

$$\mathsf{Adv}_{\mathcal{Y},D} \;\le\; \frac{c\big(\sigma' + \log \gamma^{-1}\big)\, q^{+}}{P} \;+\; C\sqrt{\sigma'\, q^{+}\, \delta} \;+\; \gamma.$$

Remark 1 (what the indexing does and does not allow). The family is indexed by the oracle as well as by the leakage. This is the one departure from the classical single-source lemma, where it is indexed by the advice alone, and it is what permits a single *joint* decomposition rather than one per source: the component may be chosen by inspecting all of H and both leakages. It may not be chosen using $\mathbf{x}$, which is why the index set is $\mathsf{Fun} \times (\{0,1\}^*)^2$ and not anything larger. Consistency with f is likewise essential rather than cosmetic, since it keeps the fixed values compatible with the leakage the observer already holds.

Remark 2 (quantifier order). The order is part of the statement. The family is chosen after P and γ but *before* q, so no part of it may be tuned to the observer's query budget. Relaxing this makes the statement easier in a way the intended applications cannot use.

Remark 3 (the two forms of unpredictability agree here). Definition 1 hands the predictor the whole vector $\mathbf{z}$, which is formally a stronger hypothesis than handing it z_i alone. For split sources the two coincide exactly. Conditioned on H, the pairs (x_1, z_1) and (x_2, z_2) are independent, because the coins are; hence z_2 is independent of (x_1, z_1) given H, so x_1 is independent of z_2 given (H, z_1), and $\Pr[x_1 = u \mid H, \mathbf{z}] = \Pr[x_1 = u \mid H, z_1]$ for every $u \in [N]$. The same is not true of a predictor without oracle access, for which the oracle-free notion is strictly weaker.

Remark 4 (why two sources). Conjecture 1 is false with a single source. Take M even, let S draw $x \leftarrow [N]$, query H at x, and output the point x together with the one-bit leakage $z := \mathbb{1}[H(x) \le M/2]$, and let the observer be the 0-query D that outputs $\mathbb{1}\big[\mathbb{1}[y \le M/2] = z\big]$ on input (y, z). Then S is δ-unpredictable with $\delta \le 2/N$, while Real outputs 1 with probability 1 and, for any fixed set of size $P \ll N$, Dec outputs 1 with probability close to $1/2$, so $\mathsf{Adv}_{\mathcal{Y},D}$ is close to $1/2$ for every admissible family $\mathcal{Y}$. Any proposed proof must therefore fail visibly when the two sources are replaced by one, and it cannot draw its use of their independence solely from a bound on the probability that the fixed set contains $\mathbf{x}$, since such bounds hold for one source as well.

4 A consequence

The conjecture is worth having because it yields a bound on extraction from split unpredictable sources: the value of the oracle at the pair of points looks uniform. That statement mentions no mixtures and no fixed sets, and it is the form the applications use.

Definition 4 (extraction advantage). For sources S_1, S_2 and a q-query challenge-oblivious observer D, define two further experiments, both drawing H and $(\mathbf{x}, \mathbf{z})$ as in Real:

- Real_0: set $y \leftarrow [M]$ and output $D^H(y, \mathbf{z})$;
- Dec_0: draw $(J, H^*) \leftarrow Y_{H,\mathbf{z}}$, for a family $\mathcal{Y}$ as in Definition 3, set $y \leftarrow [M]$, and output $D^{H^*}(y, \mathbf{z})$.

The *extraction advantage* of the pair at q queries is

$$\kappa(q) \;:=\; \sup_{D} \big|\Pr[\mathsf{Real} = 1] - \Pr[\mathsf{Real}_0 = 1]\big|,$$

the supremum over q-query challenge-oblivious observers; it involves Real and Real_0 only, so it does not depend on $\mathcal{Y}$. The oracle is the real H in both experiments; only the challenge value changes.

Lemma 1 (the fixed set misses the challenge). *Let $\mathcal{Y}$ be any family as in Conjecture 1, let J index the component drawn in* Dec *or in* Dec_0,

and let I_J be its fixed set. Then $\Pr\left[\mathbf{x} \in I_J\right] \le P\delta$.

Proof. Consider the predictor P that, on input $\mathbf{z}$ and with oracle access to H, computes the family member $Y_{H,\mathbf{z}}$, draws a component index J from its weights, and outputs a uniformly random element of $\{u_1 : (u_1,u_2) \in I_J\}$, or 1 if that set is empty. It can do this because $\mathcal{Y}$ depends only on (S_1,S_2,P,γ) and is indexed by the oracle and the leakage, both of which P holds, and because predictors are unbounded in time and in queries. Here and below, "computationally unbounded algorithm" should be read as "arbitrary, possibly randomized, function", since $\mathcal{Y}$ is not assumed constructive and nothing in this document needs predictors to be computable. Its draw of J has the same law as the one in the experiment, since there too the component is drawn from $Y_{H,\mathbf{z}}$ after $(H,\mathbf{x},\mathbf{z})$ and hence independently of $\mathbf{x}$ given $(H,\mathbf{z})$. The first coordinates of I_J number at most $|I_J| \le P$, so $\delta \ge \Pr\left[\mathsf{P}^H(\mathbf{z}) = x_1\right] \ge \Pr\left[x_1 \in \{u_1 : u \in I_J\}\right]/P \ge \Pr\left[\mathbf{x} \in I_J\right]/P$. □

Lemma 2 (the observer misses the challenge). *In* Dec_0*, let Q be the set of points D queries. Then* $\Pr\left[\mathbf{x} \in Q\right] \le q\delta$*, the case $q = 0$ holding because Q is then empty.*

Proof. Consider the predictor that, on input $\mathbf{z}$ and with oracle access to H, draws $H^* \leftarrow Y_{H,\mathbf{z}}$ internally, draws $y \leftarrow [M]$, runs $D^{H^*}(y,\mathbf{z})$ while recording its queries, and outputs the first coordinate of a uniformly chosen one, or 1 if there are none. Every object it draws has the law it has in Dec_0, conditioned on $(H,\mathbf{x},\mathbf{z})$: the component because it is drawn from $Y_{H,\mathbf{z}}$, and y because it is uniform there. As $|Q| \le q$, we get $\delta \ge \Pr\left[\mathbf{x} \in Q\right]/q$ for $q \ge 1$. □

Theorem 1 (extraction for split unpredictable sources). *Assume Conjecture 1, with constants c and C. Then for every N, M, every split δ-unpredictable pair of sources with leakage lengths bounded by σ_1,σ_2, and every $q \in \mathbb{N} \cup \{0\}$,*

$$\kappa(q) \;\le\; (4c + 2C + 4)\sqrt{\sigma' q^+ \delta} \;+\; q\delta.$$

The second term is dominated by the first whenever the bound is below one, so the content is $\kappa(q) = O(\sqrt{\sigma' q^+ \delta})$.

Proof. Fix a q-query challenge-oblivious D, put $\gamma := N^{-2}$ and $P := \lceil \sqrt{\sigma' q^+/\delta} \rceil$, and let $\mathcal{Y}$ be the family the conjecture supplies for these P and γ. Choosing P as a function of q does not disturb the quantifier order of Remark 2: the conjecture supplies a family for *every* P, and we may instantiate it at whichever P we please; what the order forbids is a single family that depends on q, and none here does. Write β for the right-hand side of the conjecture at these parameters. Split

$$\begin{aligned}\left|\Pr[\mathsf{Real} = 1] - \Pr[\mathsf{Real}_0 = 1]\right| \le{}& \left|\Pr[\mathsf{Real} = 1] - \Pr[\mathsf{Dec} = 1]\right| && \text{(i)}\\ &+ \left|\Pr[\mathsf{Dec} = 1] - \Pr[\mathsf{Dec}_0 = 1]\right| && \text{(ii)}\\ &+ \left|\Pr[\mathsf{Dec}_0 = 1] - \Pr[\mathsf{Real}_0 = 1]\right|. && \text{(iii)}\end{aligned}$$

Term (i) is $\mathsf{Adv}_{\mathcal{Y},D} \le \beta$ by the conjecture. For term (iii), let D' be the observer that ignores its challenge input, draws $y \leftarrow [M]$ itself, runs $D(y, \mathbf{z})$ and forwards its oracle queries; D' is q-query and challenge-oblivious, and running D' in Real and in Dec reproduces Real_0 and Dec_0 for D, so term (iii) is $\mathsf{Adv}_{\mathcal{Y},D'} \le \beta$ by the conjecture again.

For term (ii), couple the two experiments. Draw H, $(\mathbf{x}, \mathbf{z})$ and a component index J with fixed set I_J, let H° be a draw of that component, and let $U \leftarrow [M]$ be independent of everything. Define $H^* := H^\circ$ except that, if $\mathbf{x} \notin I_J$, its value at $\mathbf{x}$ is overwritten by U. Both H° and H^* are distributed as that component, since off I_J its values are uniform and independent, so replacing the single value at $\mathbf{x}$ by an independent uniform one changes nothing. Run Dec with oracle H^*, whose challenge value $H^*(\mathbf{x})$ then equals U on the event $\mathbf{x} \notin I_J$, and run Dec_0 with oracle H° and challenge value U. The two executions receive the same input and the same oracle answers unless $\mathbf{x} \in I_J$, of probability at most $P\delta$ by Lemma 1, or unless D queries $\mathbf{x}$, in which case H° answers with a value independent of U while H^* answers with U. The executions agree up to and including the step before the first query to $\mathbf{x}$, so that event has the same probability in both and may be bounded in Dec_0, where Lemma 2 gives $q\delta$. Hence term (ii) is at most $P\delta + q\delta$.

Adding, $\kappa(q) \le 2\beta + P\delta + q\delta$. Now $\log \gamma^{-1} = 2 \log N \le \sigma'$, so $\beta \le 2c\sigma' q^+/P + C\sqrt{\sigma' q^+ \delta} + \gamma$; the choice of P gives $2c\sigma' q^+/P \le$

$2c\sqrt{\sigma' q^+ \delta}$ and $P\delta \le \sqrt{\sigma' q^+ \delta} + \delta$; and $\gamma = N^{-2} \le \delta^2 \le \delta$ because $\delta \ge 1/N$. Collecting, and using $\delta \le \sqrt{\sigma' q^+ \delta}$ since $\delta \le 1 \le \sigma' q^+$,

$$\kappa(q) \;\le\; (4c+2C+1)\sqrt{\sigma' q^+ \delta} + 3\delta + q\delta \;\le\; (4c+2C+4)\sqrt{\sigma' q^+ \delta} + q\delta.$$

□

Remark 5 (what the consequence does and does not use). Theorem 1 uses three features of the family and no others: that each fixed set has size at most P, which is what makes Lemma 1 cost $P\delta$; that the family is indexed by the oracle and the leakage and by nothing else, which is what lets the predictor in that lemma rebuild it; and that the component is drawn after $(H, \mathbf{x}, \mathbf{z})$, which is what makes the drawn index independent of $\mathbf{x}$ given $(H, \mathbf{z})$. The consistency requirement of Definition 2 is *not* used anywhere in this section, so the theorem follows from the weaker conjecture obtained by deleting it. It is retained in Conjecture 1 because the intended applications need the fixed values to agree with the oracle the leakage was computed from, and because the known single-source decomposition delivers it.

Remark 6 (the converse, and where it stops). The implication also runs backwards, which is why the conjecture is not merely sufficient. There is a family, independent of q and D, for which $\mathsf{Adv}_{\mathcal{Y},D} \le \kappa(q) + Aq/P_0 + \gamma + P_0\delta + q\delta$, with $A := \sigma + 2 + \log(1/\gamma)$ and $P_0 := \min\bigl(P, \lceil\sqrt{A/\delta}\,\rceil\bigr)$. So $\kappa(q) = O(\sqrt{\sigma' q^+ \delta})$ implies Conjecture 1 for $P \le \sqrt{A/\delta}$, and for every P when $q = O(1)$, while Theorem 1 gives the reverse at $P = \lceil\sqrt{\sigma' q^+/\delta}\,\rceil$. The two therefore coincide for $q = O(1)$. For growing q they do not: the directions hold at values of P separated by $\Theta(\sqrt{q^+})$, and closing that is a second open problem, distinct from bounding $\kappa(q)$.

5 Bibliography

[CDGS] S. Coretti, Y. Dodis, S. Guo, J. Steinberger. *Random Oracles and Non-Uniformity.* EUROCRYPT 2018; ePrint 2017/937.

[CFHS] S. Coretti, P. Farshim, P. Harasser, K. Southern. *Multi-Source Randomness Extraction and Generation in the Random-Oracle Model.* ITC 2025, LIPIcs 343, art. 10; ePrint 2025/1258.

[DGK] Y. Dodis, S. Guo, J. Katz. *Fixing Cracks in the Concrete: Random Oracles with Auxiliary Input, Revisited.* EUROCRYPT 2017.

[Unruh] D. Unruh. *Random Oracles and Auxiliary Input.* CRYPTO 2007.