Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Certifying the quantum local differential privacy ofan untrusted mechanism: a no-go theorem and acomposable trusted-probe protocol

**Clifford N. Jones**

cnj3v@mtmail.mtsu.edu

Middle Tennessee State University

**Khem Poudel**

Middle Tennessee State University

**Research Article**

**Keywords:** quantum local differential privacy, quantum information processing, privacy certification, untrusted quantum channels, composable security, federated aggregation

**Posted Date:** November 11th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-10297234/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# Certifying the quantum local differential privacy of an untrusted mechanism: a no-go theorem and a composable trusted-probe protocol

Clifford N. Jones[1*] and Khem Poudel[1]

[1*]Middle Tennessee State University, Murfreesboro, Tennessee, USA.

*Corresponding author(s). E-mail(s): cnj3v@mtmail.mtsu.edu;

**Abstract**

Quantum local differential privacy (QLDP) bounds how much any measurement can distinguish the outputs of a mechanism on neighboring inputs, but existing theory quantifies privacy under the assumption that the mechanism is trusted and characterized. We ask whether the QLDP parameter of an untrusted quantum mechanism can be certified from observed statistics, and we answer in two parts. First, we prove a no-go theorem: under full device independence, where the measurement is also untrusted, no protocol can certify any finite bound on the privacy parameter, because the privacy action of a channel can be absorbed into the untrusted measurement without changing the observed statistics. Second, we identify the minimal trusted-probe model that escapes this obstruction: the verifier prepares informationally complete test states, applies a characterized informationally complete measurement to the output, and reports the worst-case privacy parameter over all mechanisms consistent with the data. The resulting certificate is a sound upper bound on the true leakage, admits a closed form on the optimal mechanism family used for private inference, and remains valid under finite statistics against adaptive, memoryful adversarial strategies through a martingale concentration bound. The certified mechanism retains the known quantum utility advantage for non-binary data up to a quantified noise threshold, and the certificate is composably secure in the abstract-cryptography framework, yielding a certified-private-aggregation resource for federated settings. The results connect quantum differential privacy with quantum certification by treating privacy leakage itself as the operational quantity to be certified.

**Keywords:** quantum local differential privacy, quantum information processing, privacy certification, untrusted quantum channels, composable security, federated aggregation

# 1 Introduction

A central promise of privacy-preserving computation is that a party applying a randomizing mechanism to sensitive data actually applied it—and at the strength claimed. Whether that promise can be *verified*, without trusting the hardware that implements the mechanism, is a question about certification: can an operational privacy guarantee be established for an untrusted device from what it can be made to output? Quantum information has answered the analogous question for states, dimensions, and structural channel properties through device-independent certification and self-testing, but never for a privacy parameter. This paper asks whether an operational privacy guarantee of an untrusted quantum mechanism can itself be certified, shows that full device independence makes this impossible, and identifies the verifier-controlled preparation-and-measurement model in which such certification becomes sound and composable.

The question has a concrete home in federated and distributed learning, where many participants contribute to a shared computation without surrendering their raw data: a central server aggregates locally processed contributions, and privacy is supposed to be enforced by a randomizing mechanism applied before any data is released. Local differential privacy formalizes this promise by bounding how much any single contribution can shift the distribution of what is published. Yet, the promise rests on an assumption that is rarely examined, that the party applying the mechanism actually applied it. A server charged with privatizing client contributions may, through negligence, misconfiguration, or malice, apply a weaker mechanism, or none, and the clients have no way to tell from the published output alone.

This gap is fundamental rather than incidental. Classically, any differential-privacy guarantee that can be *efficiently verified* from a mechanism's behavior can also be efficiently *broken*: the same access that lets a verifier confirm the privacy level lets an adversary distinguish the very inputs the mechanism was meant to protect [1]. The leading classical route to verifiable differential privacy therefore retreats to a weaker setting, certifying privacy only through computational zero-knowledge proofs supplied by a semi-trusted prover, under cryptographic assumptions [2, 3]. In the local model relevant to federated aggregation, where the untrusted party is the one we wish to hold accountable, no classical protocol certifies the privacy level of an uncharacterized mechanism information-theoretically and without a trusted prover.

Quantum information offers a way past this barrier. Device-independent certification and self-testing establish that properties of untrusted quantum devices—entanglement, dimension, the fidelity of a prepared state—can be verified from observed nonlocal statistics alone, with no assumption on the internal workings of the apparatus [4]. Independently, quantum local differential privacy (QLDP) has been developed as the quantum counterpart of local differential privacy, with optimal mechanisms and composition theory now in hand [5, 6], and quantum mechanisms are known to deliver a strict utility advantage over their classical counterparts for non-binary data [7, 8]. These two lines have stayed separate: certification has addressed states and structural channel properties, never a privacy parameter, while the QLDP literature has quantified privacy under the assumption that the mechanism is trusted and characterized. This work builds a bridge between them. The machinery of

device-independent and composable certification—the worst-case optimization over consistent devices, ideal resources, simulators, and composable soundness against an adversary—was developed to certify states, dimensions, and structural channel properties; we show it can be turned on an operational *privacy* functional, letting a client verify, without trusting the server's hardware, that an untrusted mechanism has privacy parameter no larger than the reported bound. That a privacy parameter, defined by a universal quantification over all measurements, can be certified at all by these tools and made composable is the conceptual content of the paper.

The point is broader than the motivating learning application. It suggests that quantum certification can be used not only to certify physical properties of a device, such as a state fidelity or dimension, but also to certify operational information-theoretic guarantees carried by that device. Privacy leakage is the first such functional that we treat. The same perspective raises a wider program: which other operational quantities—distinguishability, leakage, robustness, or resource monotones—admit sound and composable certificates when the mechanism that realizes them is untrusted?

In this work, we answer the privacy-certification question affirmatively: we recast the certification of an untrusted mechanism's privacy as a composable cryptographic primitive, joining the certification toolkit of quantum cryptography to the quantum differential-privacy functional, with federated aggregation as the motivating application. Our contributions are fourfold. First, we establish an obstruction that fixes the right trust model: under full device independence, where the measurement is also untrusted, no nontrivial bound on the QLDP parameter can be certified, because any privatization can be absorbed into the untrusted measurement and so hidden from the verifier (Section 4). This shows that the verifier must control a reliable, characterized, informationally complete measurement; without it, certification is vacuous. The untrusted party throughout is the mechanism itself. Second, in this trust model we construct the certificate as the supremum of the privacy parameter over all mechanisms consistent with the observed statistics of trusted, informationally complete probes; the resulting bound never reports a mechanism as more private than it is, is tight on the optimal mechanisms, and is sound under finite statistics against a general, correlated, and memoryful adversarial mechanism through a martingale concentration bound (Section 5). The certified value is the operational privacy guarantee read from the consistent set, and—unlike channel reconstruction for its own sake—it is sound against *any* mechanism the server may employ. Third, we show that certification does not destroy the quantum utility advantage: on the certified mechanism, and for non-binary data, the quantum-to-classical utility ratio remains above unity up to a substantial noise threshold (Section 6). Fourth, we prove that the certificate is composably secure in the abstract-cryptography framework, constructing an ideal certified-private-aggregation resource within an explicit, additively composing error; this is what allows the certificate to serve as a reusable subroutine, and we illustrate it in the federated-aggregation setting that motivates the work, where the composable guarantee accumulates across rounds (Sections 7 and 8). Federated aggregation serves as the paper's motivating use case and as the natural setting in which the composable certificate becomes operational.

The certification target is what is new: where prior methods certify a state, a dimension, or a structural device property, we certify an operational *privacy functional* of an untrusted mechanism—impossible under full device independence, but recovered in a minimal trusted-measurement model, made sound under finite statistics, and lifted to a composable cryptographic resource. The contribution is thus a new point in the certification landscape, a new certifiable operational quantity with its own no-go theorem, minimal-trust remedy, and composable semantics, rather than a new mechanism or a sharper bound within an existing one.

The result is scoped as follows. The verifier trusts its own preparation and measurement; this is not a convenience but a necessity, since leaving the measurement untrusted makes certification provably vacuous (Section 4), and in the regime where the trusted measurement is informationally complete, the reconstruction of the mechanism coincides with quantum process estimation—the content of our result is not that reconstruction but the operational privacy guarantee and its soundness against an adversarial mechanism. The certified advantage is claimed only for non-binary data and only as $Q/C > 1$ robust to realistic noise; the stronger optimal ratio is more fragile, and for binary data no advantage exists at all, in keeping with the known no-go [7, 8]. The classical comparison is scoped just as carefully: the paper does not claim that differential privacy can never be verified classically. The claim is that known classical verifier models do not provide the same combination proved here—certification of an uncharacterized mechanism's privacy parameter, prover-free operation, and information-theoretic soundness—without relaxing at least one of those requirements.

The remainder of the paper is organized as follows. Section 2 situates our contribution among closely related lines of work it builds on. Section 3 fixes the privacy functional, the channel and measurement model, and the trust and threat models. Section 4 proves the device-independent obstruction that fixes the trust model. Section 5 develops the privacy-certifying test and its finite-statistics soundness, the technical core of the article. Section 6 shows that the quantum advantage survives certification and Section 7 establishes composable security, which Section 8 assembles into the federated-learning application and the joint guarantee. Section 9 discusses assumptions, limitations, and outlook. The proofs and details of all numerical results are collected in the appendices.

## 2 Prior work

Our result is at the confluence of several lines of work. We review each in turn, isolating the precise respect in which our contribution differs; Table 1 summarizes the comparison.

### 2.1 Quantum local differential privacy and its utility advantage.

Differential privacy itself originates in the classical literature, in the sensitivity-calibrated noise of Dwork *et al.* [9] and the foundations surveyed by Dwork and Roth [10]; its *local* model, in which each contributor randomizes locally before release,

traces back to randomized response [11] and was formalized through local learning [12] and the minimax analysis of Duchi, Jordan and Wainwright [13]. The quantum analog of local differential privacy was placed on firm ground by Hirche, Rouzé, and Stilck França [5], who cast the guarantee in terms of quantum divergences, and by Guan [6], who identified optimal mechanisms and proved an additive composition theorem that holds even when inputs are entangled. On the utility side, Nam *et al.* [7] established that a quantum mechanism can strictly outperform any classical one for locally private statistical inference, and Yoshida [8] determined the optimal high-privacy mechanisms and the exact quantum-to-classical utility ratio, while also delimiting the advantage: for binary data no quantum advantage exists, a no-go we respect throughout. Quantum differential privacy has since developed along several lines: the gentle-measurement connection of Aaronson and Rothblum [14], the computational treatment of Zhou and Ying [15], the local-model analysis of Angrisani and Kashefi [16, 17], and applications to private learning [18]. This body of work quantifies and optimizes privacy under the standing assumption that the mechanism is trusted and fully characterized. We adopt its privacy functional and its advantage as given and ask the orthogonal question it does not address—whether the privacy level of an *untrusted* mechanism can be certified from observed statistics.

## 2.2 Device-independent and measurement-device-independent certification.

Self-testing and device-independent certification show that properties of untrusted devices can be established from nonlocal statistics: the fidelity of a prepared state, the dimension of a system, or the presence of entanglement, without assumption on the apparatus [19, 20]. The idea originates with Mayers and Yao [21] and underlies results ranging from robust analytic self-tests [22] to the classical command of quantum systems [23] and the device-independent certification of measurements [24]. The measurement-device-independent variant, introduced for key distribution to close detector side channels [25] and extended to the witnessing of entanglement [26], trusts only the preparation of known probe states and certifies channel properties such as whether an untrusted channel breaks entanglement. Most directly related to us, Bhavsar, Wooltorton, and Bae [4] give a composable framework for device-independent state certification, with security cast as closeness to an ideal resource and quantified through an extractability that plays the role of a fidelity. We adopt this composable finite-statistics methodology. The difference is the object certified: every protocol in this line certifies a state, a dimension, or a structural channel property, whereas we certify a *differential-privacy parameter*—a quantity defined by a universal quantification over all measurements—which, as Section 4 shows, behaves differently under the device-independent assumption: it cannot be certified when the measurement is untrusted, and so requires a trusted measurement in addition to trusted preparation.

## 2.3 Finite-statistics soundness.

The guarantee against a general, non-i.i.d. adversary is, for protocols that *extract a secret key or random string*, the role of the entropy accumulation theorem [27] and

its generalized form [28], applied to practical device-independent cryptography [29] and quantum key distribution [30, 31]: these lower-bound the smooth conditional min-entropy of the extracted string against an adversary's quantum side information. Our certificate does not extract a key. The quantity it controls is a bounded classical statistic—the empirical agreement of the trusted probes—and the soundness requirement is one-sided concentration of that statistic against an untrusted device that may carry memory and act adaptively across rounds. This is exactly the regime of the Azuma–Hoeffding inequality for bounded-difference martingales, the tool used to certify an estimated statistic against a memoryful device in device-independent randomness generation [32] and in composable device-independent certification of a property from observed statistics [4]. We therefore use martingale concentration, not entropy accumulation, to turn the per-round statistics into a sound bound on the certified privacy parameter; the certified value converges to the truth as the number of rounds grows. Accumulating the privacy divergence *directly* across rounds would require a relative-entropy accumulation theorem that is not known to hold, and we do not assume one.

### 2.4 Self-tested secure computation.

Closest in spirit on the cryptographic side is the verifiable quantum secure summation of Hayashi and Koshiba [33], which uses self-testing to verify, without trusting the measurement devices, that a multiparty summation was performed correctly and without leaking the inputs beyond the sum. The security object there is the *secrecy and correctness* of a computation. Ours is different: we certify a quantitative *privacy parameter* of the mechanism—a bound on the indistinguishability of neighboring inputs under every measurement—rather than the correctness of a computed result. The two are complementary guarantees on different things. Our composability is stated in the abstract-cryptography framework, the quantum instance of the universal-composability paradigm [34–36].

### 2.5 Classical verifiable differential privacy.

The classical literature on verifying privacy splits into two families, and our guarantee is distinguished from both by a precise combination of features: it is *prover-free*, *information-theoretically sound* (no computational assumption, common reference string, or trusted setup) and certifies a sound *upper bound* on the privacy parameter of an *uncharacterized* mechanism. The first family verifies a *prescribed* mechanism with the aid of a prover and cryptographic assumptions, running from the cryptographic randomized-response techniques of Ambainis *et al.* [37] and the verifiable differential privacy of Narayan *et al.* [38], through verifiable polling and confidential proofs of private training [39, 40], to recent input-authenticated schemes in the local and shuffle models [41] and the publicly verifiable counting of Biswas and Cormode [2]—who in fact prove that computational assumptions are *necessary* for their notion of verifiable differential privacy. Closest in aim, the certification of private probabilistic mechanisms of Bell, Goldwasser, Kim, and Watson [3] asks exactly our motivating question but answers it classically, with a prover and cryptographic assumptions. The second family

**Table 1** Where this work sits among certification and verifiable-privacy lines. "Privacy leakage of an untrusted mechanism" is the column that isolates the present contribution: prior certification targets a state, device, or calibrated property, while classical verifiable DP either certifies a prescribed mechanism under computational assumptions or audits a property without a sound upper certificate.

| Line of work | Certifies | Trust model | Finite-stat. | Composable |
|---|---|---|---|---|
| DI state certification [4] | State / extractability | Fully device-independent | Yes | Yes |
| Certification & benchmarking [43] | States / gates / devices | Varies | Varies | Not the focus |
| Classical black-box auditing [1, 42] | DP property (oracle access) | Classical black-box | Sometimes | Usually no |
| Verifiable DP with proofs [2, 3] | DP + correctness | Classical, computational | Yes | Computationally |
| **This work** | **QLDP leakage, untrusted mech.** | Trusted prep. + IC meas. | Yes | Yes |

is prover-free and assumption-light but delivers statistical *lower* bounds or violation detection rather than sound upper certificates: the property-testing impossibility of Gilbert and McMillan [1]—efficiently verifiable privacy is efficiently breakable—and the black-box auditing of Shi and Ma [42], who establish a fundamental impossibility for controlling both error types without further assumptions. These impossibility results are precisely the barrier our quantum protocol circumvents: the informationally complete probe structure, together with complete positivity and trace preservation, furnishes an information-theoretic, prover-free *upper* bound on an uncharacterized mechanism's privacy parameter—a combination that known classical verifier models attain only by introducing a prover, computational assumptions, or a weaker guarantee.

To our knowledge, existing classical verifier models do not simultaneously provide prover-free operation, information-theoretic soundness, and a sound upper bound on the privacy parameter of an uncharacterized mechanism; recent classical work attains verifiability only by relaxing one of these requirements. The remainder of the paper constructs a quantum certificate that meets these three requirements in the stated trusted-probe model, establishes its soundness and its surviving utility advantage, and makes it composable.

# 3 Setup and definitions

Constructing the certificate promised in Section 2 requires first fixing the language in which a privacy claim is made and checked. Three actors appear throughout. *Clients* hold data they wish to keep private; a *server* runs the aggregation mechanism that is supposed to privatize those data before release; and a *verifier*—which the clients trust, and which may be a client itself—seeks to confirm that the server's mechanism

is at least as private as claimed. The server is untrusted; the question is what the verifier can establish about it. We fix in turn the privacy functional that quantifies the claim, the model by which the verifier probes the mechanism, and the trust and threat assumptions that bound the whole exchange; Figure 1 shows the resulting protocol in a glance.

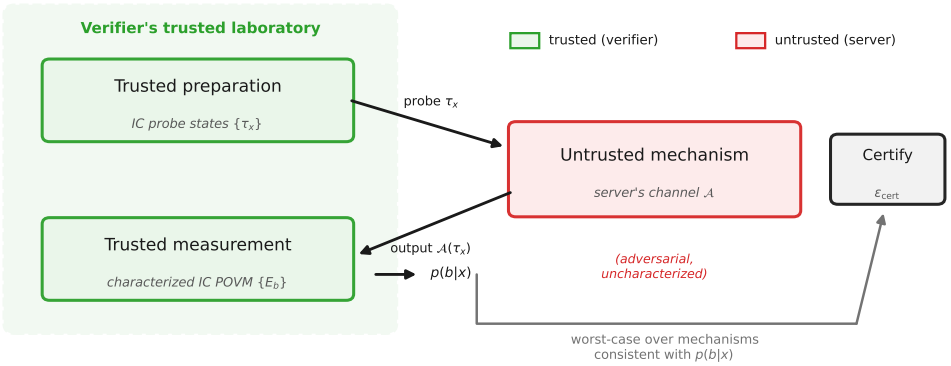



**Fig. 1** The certification protocol. The verifier's laboratory (green) holds two trusted, characterized capabilities—preparation of informationally complete probe states $\{\tau_x\}$ and an informationally complete measurement $\{E_b\}$. Probes are sent through the server's untrusted mechanism $\mathcal{A}$ (red), the output is measured, and the resulting statistics $p(b\,|\,x)$ feed the certificate $\varepsilon_{\mathrm{cert}}$, defined as the worst case of the privacy parameter over all mechanisms consistent with the data. The mechanism is the only untrusted element; it is treated as adversarial and uncharacterized.

## 3.1 Quantum local differential privacy

Let $\mathcal{H}$ be a finite-dimensional Hilbert space, and let $\mathcal{A}\colon \mathcal{S}(\mathcal{H}) \to \mathcal{S}(\mathcal{H}')$ be the quantum channel—a completely positive, trace-preserving map—that models the server's mechanism acting on a client's input state. Privacy is the requirement that the mechanism's output reveal little about which of two neighboring inputs was supplied. We make this quantitative through the QLDP functional in the hockey-stick sense [5, 6], taking the neighboring relation in its strongest local form: every pair of input states is neighboring, so that privacy must hold uniformly across all inputs rather than only for a restricted set.

**Definition 1** ($\varepsilon$-QLDP) For $\varepsilon \geq 0$, the channel $\mathcal{A}$ is *$\varepsilon$-QLDP* if, for every pair of input states $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ and every POVM element $0 \leq M \leq \mathbb{I}$,

$$\mathrm{Tr}\big[M\,\mathcal{A}(\rho)\big] \;\leq\; e^{\varepsilon}\,\mathrm{Tr}\big[M\,\mathcal{A}(\sigma)\big]. \tag{1}$$

Quantification over *all* POVM elements $M$ is the operative feature of (1): it demands that no measurement, however chosen, distinguishes the output of neighboring inputs by more than a factor $e^{\varepsilon}$. We write $\varepsilon(\mathcal{A})$ for the smallest $\varepsilon$ that satisfies (1), with

$\varepsilon(\mathcal{A}) = \infty$ when none exists. A *smaller* $\varepsilon$ is *stronger* privacy; "certifying privacy" accordingly means establishing an *upper* bound on $\varepsilon(\mathcal{A})$. The privacy budget is composed additively: across $R$ rounds of aggregation, the effective mechanism is $\big(\sum_{r=1}^{R} \varepsilon^{(r)}\big)$-QLDP, and this additivity is insensitive to whether the input per-round is independent, classically correlated, or entangled [6]. We use this basic additive composition throughout and make no sharper composition claim.

It is useful to record how $\varepsilon(\mathcal{A})$ can be evaluated, since this is the quantity that the certificate of Section 5 limits. For a qubit mechanism with affine Bloch action $\boldsymbol{r} \mapsto T\boldsymbol{r} + \boldsymbol{t}$, each output state has eigenvalues $\frac{1}{2}\big(1 \pm \|T\boldsymbol{r} + \boldsymbol{t}\|\big)$ and the universal quantification in (1) collapses to an extremal eigenvalue ratio,

$$\varepsilon(\mathcal{A}) \;=\; 2\,\operatorname{artanh}\Big(\max_{\|\boldsymbol{n}\|=1} \|T\boldsymbol{n} + \boldsymbol{t}\|\Big), \tag{2}$$

a closed form that serves as the ground truth against which we calibrate the certificate. In general dimension the same functional is the supremum over neighboring inputs of $\ln \lambda_{\max}\big[\mathcal{A}(\sigma)^{-1/2}\mathcal{A}(\rho)\,\mathcal{A}(\sigma)^{-1/2}\big]$, which is the measured hockey-stick divergence of the outputs and is computable by semidefinite programming [44]; this is what allows the certificate to extend beyond the qubit case.

## 3.2 Probing model

The verifier learns about the untrusted mechanism only through what it can prepare and what it observes. It prepares trusted, informationally complete (IC) probe states $\{\tau_x\}_x$ from a known ensemble—a symmetric informationally complete (SIC) set [45, 46] or other 2-design—and submits each to the mechanism $\mathcal{A}$. A trusted, characterized, informationally complete measurement $\{E_b\}$ is then applied to the output, yielding outcome $b$ with probability

$$p(b\,|\,x) \;=\; \operatorname{Tr}\big[\,E_b\,\mathcal{A}(\tau_x)\,\big], \qquad E_b \geq 0, \quad \textstyle\sum_b E_b = \mathbb{I}. \tag{3}$$

The trusted, informationally complete probes are what make this informative: because the $\{\tau_x\}$ span the operator space, the observed statistics $p(b\,|\,x)$ are a linear functional of the Choi operator of $\mathcal{A}$. This linearity is the structural fact on which Section 5 rests—it renders the search for the worst-case privacy parameter consistent with the data a convex program.

## 3.3 Trust and threat model

The object under test is the server's *mechanism*, and it alone is untrusted: it may deviate arbitrarily from any prescribed channel and is computationally unbounded. To probe it, the verifier relies on two trusted laboratory capabilities: the preparation of the informationally complete probe ensemble $\{\tau_x\}$ and a characterized informationally complete measurement $\{E_b\}$ applied to the output of the mechanism. Both are standard, well-calibrated quantum-optical primitives under the verifier's own control; neither is supplied by the server. We are explicit that the measurement is trusted, because this is exactly what makes certification possible: as Section 4 shows, if the

**Table 2** Trust ladder for certifying the QLDP functional (1). "Trusted" means characterized and relied upon by the verifier; "untrusted" means treated as adversarial and computationally unbounded. The mechanism under test is untrusted at every rung; what distinguishes the rungs is how much of the verifier's own apparatus is trusted. Certification is possible only when the measurement is trusted: the untrusted-measurement rung is the device-independent setting in which Theorem 1 makes every certificate vacuous.

| | State prep. | Mechanism | Measurement |
|---|---|---|---|
| Full DI | untrusted | untrusted | untrusted |
| Our setting | trusted | untrusted | trusted |

measurement is instead left untrusted, the privacy functional (1) cannot be certified at all—any privatization can be absorbed into the measurement, and every certificate becomes vacuous (Theorem 1). Trusting the measurement is therefore not a convenience but the minimal addition that breaks this obstruction; what the protocol never trusts is the mechanism whose privacy it is asked to certify.

Table 2 places this on a ladder ordered by how much of the apparatus is trusted. The fully device-independent (DI) rung, which does not trust preparation or measurement, is included for contrast: certification of the universal functional (1) is provably vacuous there (Section 4). Our setting trusts the verifier's own preparation and measurement while leaving the mechanism adversarial; this is the minimal trust in the apparatus under which the mechanism's privacy parameter can be certified, and it is the assumption made throughout. We emphasize the asymmetry that is the point of the result: the party whose honesty is in question—the server running the mechanism—is exactly the party the verifier does *not* trust, and the certificate is sound against any strategy it may adopt.

The verifier is honest and aims only to certify or reject the server's privacy claim; the accept-or-reject verdict is what protects the clients, and the verifier never attempts to recover client data. We assume that the verifier has a private source of trusted randomness and that classical communication is authenticated. The trusted preparation and measurement steps are taken to be reliable: hardware side channels and tampering with the verifier's own apparatus lie outside our model, and we make no experimental claim. What our guarantee delivers, and what no amount of trust in the verifier's apparatus trivializes, is soundness against an adversarial *mechanism*: the certificate upper-bounds the privacy parameter of *every* channel consistent with the observed statistics, so a server cannot make a non-private mechanism appear private.

## 4 The device-independent vacuousness obstruction

Before constructing the certificate, we establish why a trusted measurement is indispensable and why the fully device-independent setting, where the measurement too is untrusted, cannot support certification at all. The reason is not a limitation of any particular protocol, but a structural feature of the privacy functional itself: under full device independence, where the measurement apparatus is untrusted, no protocol can

certify any finite upper limit on $\varepsilon(\mathcal{A})$. The certificate is *vacuous*—it can never rule out a mechanism that leaks everything.

The obstruction is the interplay between universal quantification over measurements in Definition 1 and the freedom an untrusted measurement enjoys. In the device-independent setting, the verifier observes only the outcome statistics $p(b \mid x) = \mathrm{Tr}[M_b\, \mathcal{A}(\tau_x)]$ and controls neither the mechanism $\mathcal{A}$ nor the POVM $\{M_b\}$. Two pairs $(\mathcal{A}, \{M_b\})$ and $(\mathcal{A}', \{M'_b\})$ that reproduce the same statistics are therefore indistinguishable to the verifier, and a malicious server is free to present whichever pair it likes. The question is whether some fully-leaking mechanism can always hide behind a benign-looking one. It can.

**Theorem 1** (Device-independent vacuousness) *In the device-independent probing model, where the verifier observes only the outcome statistics $p(b \mid x)$ and controls neither the mechanism nor the measurement, the following is true. For self-dual unital mechanisms (the QLDP-optimal family), for* every *transcript $\{p(b \mid x)\}$ obtainable from a private mechanism $\mathcal{A}$ with $\varepsilon(\mathcal{A})$ arbitrarily small under some measurement $\{M'_b\}$, there exists a measurement $\{M_b\}$ under which the fully leaking identity channel* id*, with $\varepsilon(\mathrm{id}) = \infty$, reproduces the same transcript exactly:*

$$\mathrm{Tr}\big[M'_b\, \mathcal{A}(\tau_x)\big] \;=\; \mathrm{Tr}\big[M_b\, \mathrm{id}(\tau_x)\big] \qquad \textit{for all } b, x. \tag{4}$$

*A fully leaking implementation is therefore consistent with every such transcript, so the worst-case privacy parameter consistent with any device-independent statistics is $\sup \varepsilon = \infty$, and any sound worst-case device-independent certificate is vacuous: no finite certified bound exists.*

The proof is a one-line consequence of self-duality and is given in Section A; we describe the mechanism here because it makes the obstruction concrete. Take $\mathcal{A}$ as a depolarizing mechanism, the QLDP-optimal unital channel, which is *self-dual*: $\mathcal{A}^\dagger = \mathcal{A}$. Then for any honest POVM $\{M'_b\}$ the operators $M_b \coloneqq \mathcal{A}^\dagger(M'_b)$ form a valid POVM—they are positive and sum to the identity because $\mathcal{A}$ is unital and trace-preserving—and by the adjoint relation $\mathrm{Tr}[M_b\, \mathrm{id}(\tau_x)] = \mathrm{Tr}[\mathcal{A}^\dagger(M'_b)\, \tau_x] = \mathrm{Tr}[M'_b\, \mathcal{A}(\tau_x)]$. The depolarizing mechanism measured with $\{M'_b\}$ and the identity channel measured with $\{M_b\}$ thus produce identical statistics for every input. A verifier restricted to those statistics cannot tell a strongly private mechanism from one that performs no privatization at all: the privatization has been absorbed into the untrusted measurement (Figure 2).

This is the measurement-level form of a familiar obstruction. The QLDP value (1) is a property of the output *states* $\mathcal{A}(\tau_x)$, but device-independent statistics fix only the contraction of those states against an unknown POVM, and that contraction is invariant under moving a channel's action onto the measurement via the adjoint. What breaks the symmetry is trusting the measurement and the input. Trusting the informationally complete inputs alone is not enough: the statistics (3) then fix only the effective operators $\mathcal{A}^\dagger(N_b)$ that compose the channel's adjoint with the unknown POVM, and the channel can still be traded against the measurement through the adjoint (Section B). Once the measurement is also trusted to be the characterized informationally complete POVM $\{E_b\}$, statistics (3) determine $\mathcal{A}^\dagger$ on a complete operator basis, hence the Choi operator of $\mathcal{A}$ and the output states themselves, leaving no room to absorb the channel into the measurement. This is exactly the trust model

of Section 3.3, and Theorem 1 is the precise sense in which a trusted measurement is the *minimal* addition under which certification is possible: relaxing it returns the verifier to the vacuous device-independent regime. The remainder of the paper works in this model and shows that there the certificate is not only non-vacuous but tight.

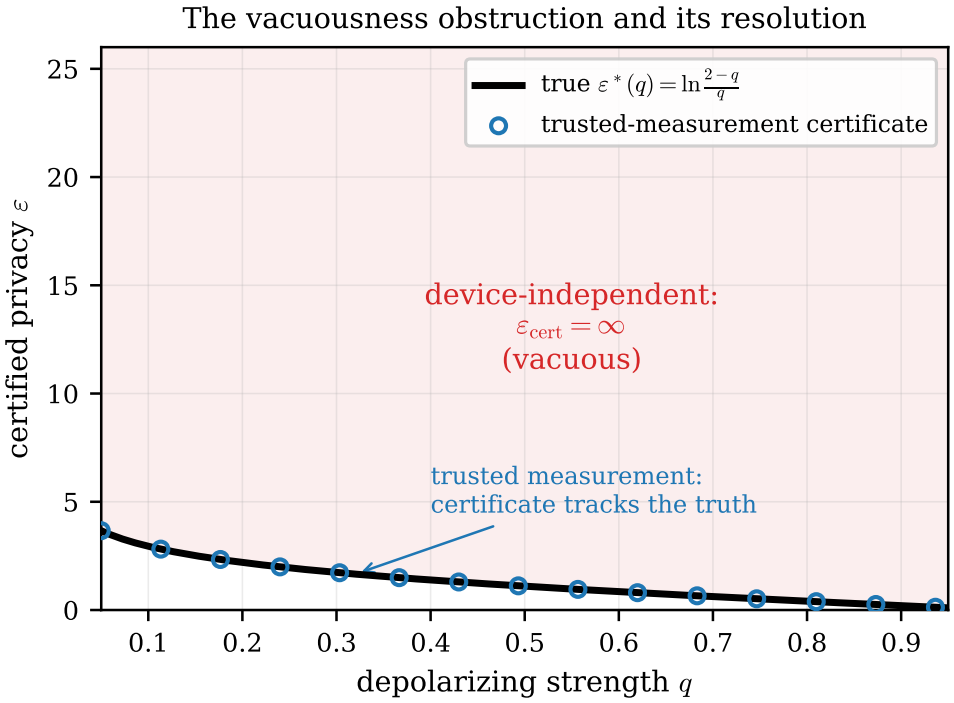



**Fig. 2** The vacuousness obstruction and its resolution. In the device-independent regime, where the measurement is untrusted, the certified parameter is unbounded (shaded region): an untrusted measurement can absorb any privatization, so no finite bound is sound. Once the measurement is trusted and informationally complete, the certificate (open circles) tracks the true QLDP value $\varepsilon^*(q)$ of a depolarizing mechanism. The finite-statistics convergence of the certificate is shown separately in Figure 4.

## 5 Certifying the privacy parameter

We now reach the central result: in our trust model, the privacy parameter of the untrusted mechanism can be certified from the statistics of trusted probe states. The certificate is the largest privacy parameter consistent with what the verifier observes—a worst case over every mechanism that could have produced the data—so that accepting it guarantees that the true mechanism leaks no more than the reported value. We first define this worst-case certificate and separate its mathematical soundness from its computational evaluation: the feasible set is convex, but the outer problem is a convex maximization rather than a single SDP. We then identify the regimes in

which the certificate does real work beyond reconstruction, give a closed-form expression on the target family, and confirm that the certified mechanism is the one carrying the quantum advantage.

One objection is worth addressing at the outset, because it shapes what the result is about. With a reliable and informationally complete measurement and unlimited data, certifying the privacy parameter might appear to be nothing more than reconstructing the channel by process estimation and evaluating a functional of it. That reading is correct—but only in the *degenerate limit* in which the consistent set collapses to a single mechanism. The object of interest is the generic case, where the measurement is coarse-grained or the statistics are finite, the data leave a nontrivial set of consistent mechanisms, no single channel is identified, and the only sound report is the worst case over that set. There, the certificate is provably not reconstruction-followed-by-evaluation (Section 5.2), and it is on that case that the conceptual content of the result rests. We therefore present the worst-case construction first in full generality, then exhibit the nondegenerate regime where it does work that no reconstruction estimate can, and only afterward specialize to the informationally complete target family on which a clean closed form and an exact value are available.

## 5.1 The certified parameter as a worst case

Fix the trusted, informationally complete probe ensemble $\{\tau_x\}$ and the trusted, characterized, informationally complete measurement $\{E_b\}$. The untrusted device implements only the mechanism $\mathcal{A}$. Writing $J$ for the Choi operator of the mechanism $\mathcal{A}$, the statistics of Equation (3) read

$$p(b\,|\,x) \;=\; \mathrm{Tr}\big[(\tau_x^{\mathsf{T}} \otimes E_b)\, J\big], \tag{5}$$

a *linear* functional of $J$ because the probes are trusted. The verifier does not know $\mathcal{A}$; it only knows that the true Choi operator lies in the set of completely positive, trace-preserving maps that reproduce the observed statistics. The certified privacy parameter is the worst case over that set,

$$\begin{aligned}\varepsilon_{\mathrm{cert}} = \sup_J \Big\{ \varepsilon(J) : J \succeq 0, \quad \mathrm{Tr}_{\mathrm{out}}\, J = \mathbb{I}, \\ \mathrm{Tr}\big[(\tau_x^{\mathsf{T}} \otimes E_b) J\big] = p(b\,|\,x) \quad \forall b, x \Big\}.\end{aligned} \tag{6}$$

where $\varepsilon(J)$ is the QLDP value Equation (1) of the mechanism with Choi operator $J$. The constraints in Equation (6) are linear in $J$, so the feasible set is a convex spectrahedron. However, the objective is *convex* in $J$: the hockey-stick divergence is jointly convex and the output $\mathcal{A}(\rho)$ is linear in $J$, so $\varepsilon(J)$ inherits convexity. The outer optimization is therefore a convex maximization over the consistent set, so the certificate is not a single black-box SDP in general; sound evaluation relies on conservative upper-bounding strategies or structured parameterizations, as used below. The crucial feature is the direction of the optimization: the reported value is a supremum over all mechanisms consistent with the data, so any computational relaxation used to evaluate it must be taken from above and remain conservative. For a fixed

candidate $J$, the inner privacy value $\varepsilon(J)$ is SDP-certifiable, with dual-feasible points giving machine-checkable upper bounds (Section B) [44]. In the structured regimes studied below—the informationally complete limit, the coarse depolarizing example, and the EITFF target family—the worst-case privacy certificate is evaluated either by the single identified $J$ or by a low-dimensional parametrization. This terminology is deliberate: "worst-case privacy certificate" denotes the sound leakage upper bound; The SDP language refers only to the fixed-$J$ divergence evaluation or to the dual certificates used in its evaluation.

**Theorem 2** (General soundness of the worst-case certificate) *Let the verifier feed the trusted, informationally complete probes $\{\tau_x\}$ to an untrusted mechanism $\mathcal{A}$ and measure the output with the fixed, trusted, informationally complete measurement $\{E_b\}$, and let $\varepsilon_{\mathrm{cert}}$ be defined by Equation (6). Then for* every *mechanism consistent with the observed statistics,*

$$\varepsilon(\mathcal{A}) \;\leq\; \varepsilon_{\mathrm{cert}}, \tag{7}$$

*so acceptance certifies that the true mechanism has a privacy parameter no larger than the reported bound. The bound holds in any finite dimension and for any mechanism, with no assumption on the family to which $\mathcal{A}$ belongs; it is moreover non-vacuous—finite whenever the observed statistics are incompatible with a near-unitary mechanism.*

This is the general guarantee, and it is all that soundness requires: $\varepsilon_{\mathrm{cert}}$ is a worst case over the consistent set, so it can only overstate leakage, never understate it. Two further statements sharpen what the certificate delivers on the target family—first that it is not merely an upper bound but *exact* there, and second (Section 5.5) that the bound survives finite statistics. We separate them from Theorem 2 because their scope is narrower: Theorem 2 is general, whereas the exactness and closed-form finite-sample constants are properties of the target family.

**Proposition 1** (Exactness on the target family) *In the depolarizing family of mechanisms (EITFF) targeted in Section 6, the worst-case certificate is tight: $\varepsilon_{\mathrm{cert}} = \varepsilon(\mathcal{A})$. Equivalently, the consistent set is pinned tightly enough so that the worst case coincides with the true privacy parameter, and the closed analytic form of Section 5.3 evaluates it exactly. Off this family, the certificate remains a sound upper bound by Theorem 2 but need not be tight.*

The bound direction is the one that privacy requires: $\varepsilon_{\mathrm{cert}}$ is an *upper* bound on leakage, so a verifier who accepts knows that the mechanism leaks no more than the certified amount. Two features of these statements deserve an emphasis because they are where a naive certificate fails.

*The full statistics are necessary.* Certifying from a single scalar test—one Bell-type number $\mathrm{Tr}[WJ] = w$—is vacuous: the set of mechanisms consistent with one linear constraint includes near-unitary channels, so the worst-case privacy parameter is essentially unbounded. Only by constraining the *entire* informationally complete statistics vector $\{p(b\,|\,x)\}$ can the feasible set in Equation (6) shrink enough to pin the privacy parameter to a finite and, in the target case, exact value. This is the quantitative reason the certificate uses informationally complete probes rather than a single witness.

**Table 3** Claim map. "Sound" means the reported value is a valid upper bound on the true privacy parameter. The table separates what is general from what is family-specific or application-level.

| Claim | Assumptions | Guarantee | Limitation/scope |
|---|---|---|---|
| Worst-case certificate | Trusted probes and measurement | Sound leakage upper bound | Tractable in structured regimes |
| Fixed-$J$ privacy value | Candidate Choi operator fixed | SDP/dual upper certificate | Fixed-channel only |
| Analytic closed form | Depolarizing/EITFF family | Exact expression | Unsound off-family |
| Finite statistics | Memoryful adaptive device | One-sided martingale bound | Target-family statistic |
| Composable security | Abstract cryptography | Additive security error | Separate budget |
| Federated aggregation | $R$ independent invocations | Joint corollary | Application corollary |

*The certificate never underestimates.* As a genuine supremum over the consistent set, $\varepsilon_{\text{cert}} \geq \varepsilon(\mathcal{A})$ always; the certificate can only err on the safe side, reporting a mechanism as *less* private than it is, never more. On pathological near-unitary statistics, it correctly returns a large (vacuous) value rather than a false certificate—it declines to certify privacy that the data do not support. Table 3 maps the claim of the paper by scope, separating what is general from what is family-specific or application-level.

## 5.2 When the worst case is nondegenerate: incomplete information

The worst-case program Equation (6) is more transparent when one asks what it certifies as the measurement becomes less informative. With a reliable informationally complete measurement, the consistent set collapses to a single Choi operator, and the certificate is the evaluation $\varepsilon(J_{\text{data}})$ (Section B); in that limit the procedure coincides with quantum process estimation, and the worst-case maximization, although still correct, does no work beyond reconstruction. This degenerate extreme is the exception, not the rule, and it is the only regime in which the reading of certification as "just tomography" is correct. The certificate earns its name precisely when the data leave a nontrivial consistent set, and there are two natural and operationally important regimes in which they do so.

*Coarse-grained measurement.* Suppose the trusted measurement is not informationally complete—it records, say, only output populations in a fixed basis, or any POVM with fewer linearly independent effects than span the operator space. The statistics then constrain $J$ only through the measured expectation values and the

consistent set

$$\mathcal{C} = \Big\{ J \succeq 0 : \operatorname{Tr}_{\text{out}} J = \mathbb{I}, \\ \operatorname{Tr}\left[(\tau_x^{\mathsf{T}} \otimes M_k) J\right] = p(k \,|\, x) \quad \forall k, x \Big\} \tag{8}$$

is a spectrahedron of *positive dimension*—and it remains so in the limit of infinitely many measurement repetitions because the missing constraints are structural, not statistical. The certified parameter is the genuine worst case $\bar{\varepsilon} = \sup_{J \in \mathcal{C}} \varepsilon(J)$, an upper bound on the privacy parameter of every mechanism consistent with the coarse data, including the true one. Here reconstruction is not merely harder; it is impossible—no single $J$ is identified—so the certificate cannot be "tomography followed by evaluation." It is the optimization that produces the guarantee.

A minimal example makes the gap concrete. Take a mechanism that has been promised to lie in the qubit depolarizing family, probed with the trusted informationally complete ensemble but read out by a coarse measurement that fixes the depolarizing strength only to an interval $p \in [p_{\text{lo}}, p_{\text{hi}}]$. Since $\varepsilon = 2\,\operatorname{artanh}(1-p)$ decreases monotonically in $p$ (more noise is more private), the worst case is at the *lower* endpoint, $\bar{\varepsilon} = \varepsilon(p_{\text{lo}})$, while a naive point estimate at the interval midpoint reports $\varepsilon(\hat{p}) < \bar{\varepsilon}$ and would *overstate privacy*. For $[p_{\text{lo}}, p_{\text{hi}}] = [0.30, 0.50]$ the worst-case certificate is $\bar{\varepsilon} = \varepsilon(0.30) \approx 1.73$ against a point estimate $\varepsilon(0.40) \approx 1.39$—a gap of 0.35 that is exactly the conceptual work the worst-case optimization does and that a reconstruction estimate misses. This concrete gap is the operational difference between a point estimate and a sound worst-case certificate.

**Proposition 2** (Nontriviality under partial information) *Suppose that the trusted measurement is coarse-grained, so that the consistent set $\mathcal{C}$ of Equation (8) has positive dimension—a property that persists in the limit of infinitely many measurement repetitions. Then:*

*(i) The worst-case certificate $\bar{\varepsilon} = \sup_{J \in \mathcal{C}} \varepsilon(J)$ is a sound upper bound on the privacy parameter of every mechanism in $\mathcal{C}$, including the true one.*

*(ii) No reconstruction estimate is well defined, because no single Choi operator is identified by the data; and any point estimate $\varepsilon(\widehat{J})$ at an interior $\widehat{J} \in \mathcal{C}$ satisfies $\varepsilon(\widehat{J}) \leq \bar{\varepsilon}$, with strict inequality whenever $\varepsilon$ is non-constant on $\mathcal{C}$—so a point estimate can report a mechanism as more private than the worst case consistent with the data, an unsound certificate.*

*For the qubit depolarizing family read out so that only $p \in [0.30, 0.50]$ is fixed, $\bar{\varepsilon} = \varepsilon(0.30) \approx 1.73$ strictly exceeds the midpoint point estimate $\varepsilon(0.40) \approx 1.39$.*

The proposition formalizes the reply to the reconstruction objection: in any regime where the consistent set is nondegenerate, the certificate is not the evaluation of a reconstructed channel—reconstruction is undefined—and the only sound report is the worst-case optimization, which strictly dominates point estimation. The informationally complete, infinite-data case is the degenerate limit in which $\mathcal{C}$ collapses to a point and the two coincide.

*Finite-statistics confidence region.* Even with a trusted informationally complete measurement, finite sampling determines the statistics only within a confidence region, so the consistent set is again nondegenerate: it is the preimage of that region under Equation (3). The sound certificate is the worst-case $\varepsilon$ over the region, which contains the true mechanism with the region's confidence; this is the optimization underlying the finite-sample bound of Section 5.5, and it reduces to the point evaluation only as the sample size diverges. The coarse-grained and finite-statistics regimes are thus the same worst-case program over a nondegenerate set, differing only in whether the set has positive dimension for structural or for statistical reasons.

In both regimes, the objective is convex in $J$, so the worst case is a convex maximization rather than a semidefinite program. The certificate remains sound because any relaxation used to evaluate it is conservative: it either evaluates a low-dimensional parametrization exactly, as in the depolarizing example, or returns an explicit upper bound on the supremum over $\mathcal{C}$ (Section B). The informationally complete, infinite-data case is the common degenerate limit in which $\mathcal{C}$ shrinks to a point.

## 5.3 Closed form on the target family

For the mechanisms that carry the quantum advantage of Section 6—the depolarizing 2-design family (EITFF), which contains the QLDP-optimal unital mechanisms—the worst case Equation (6) admits a closed form. Let

$$A \;=\; \sum_{x} \operatorname{Tr}\big[E_x\, \mathcal{A}(\tau_x)\big] \tag{9}$$

be the informationally complete agreement statistic, the average probability that the trusted measurement returns the outcome matched to the input. Then

$$\varepsilon_{\mathrm{cert}} \;=\; 2\,\operatorname{artanh}\big(A-1\big), \tag{10}$$

which coincides with the worst-case privacy certificate in this family and exactly equals the true privacy parameter (Figure 3).

*Remark 1* (Scope of the closed form) The closed form Equation (10) is exact *only* in the depolarizing family. On anisotropic mechanisms it *undercounts*—it can report a mechanism as more private than it is—and is therefore unsound as a general certificate. The general object is the worst-case privacy certificate Equation (6), which never undercounts; Equation (10) is a transparent corollary that must be used only when the target family is the object of certification, as it is here. We state this scope explicitly because mislabeling the closed form as general would be unsound.

## 5.4 Certifying the genuine mechanism

The certificate is not an artifact of the qubit case. Using the $d$-dimensional 2-design measurement and the measured hockey-stick (eigenvalue-ratio) form of the privacy

functional from Section 3.1, the certified parameter matches the exact QLDP value $\varepsilon^*(d,q) = \ln\big((d-(d-1)q)/q\big)$ of the EITFF mechanism to numerical precision for $d = 2$ and $d = 3$, and on the binding input pair for $d = 4$ (Figure 4). The advantage established in Section 6 therefore attaches to a genuinely certified mechanism in each of these dimensions, not to a low-dimensional idealization. The certificate is also stable under noise: as a white-noise fraction is added to the mechanism, $\varepsilon_{\mathrm{cert}}$ decreases smoothly and monotonically, while remaining a valid upper bound on the effective mechanism observed throughout, with no threshold pathology.

## 5.5 Finite-statistics soundness

The construction thus far treats $\varepsilon_{\mathrm{cert}}$ as a function of the exact statistics. In practice, the verifier estimates the agreement from a finite number of rounds, and a malicious server may correlate its behavior across rounds—the server's mechanism may carry internal memory or an adaptive controller that updates its action from one round to the next and may choose its action in each round adaptively on the basis of all previous rounds. A sound certificate must convert the empirical agreement of $n$ such rounds into a high-confidence bound that holds against this fully general, memoryful, adaptive adversary.

*What must be controlled is a bounded classical statistic.* The certified parameter Equation (10) is a monotone function of the single scalar agreement $A$ of Equation (9), which the verifier estimates as the empirical mean $\widehat{A} = 1 + \frac{1}{n}\sum_{i=1}^{n} m_i$ of the per-round *match indicators* $m_i \in \{0,1\}$, where $m_i = 1$ when the trusted measurement returns the result matched to the probe input of the round. The protocol extracts no secret key and amplifies no privacy; the only quantity inferred from the data is this bounded classical average, and the soundness requirement is simply that the true agreement does not exceed its empirical estimate by more than the sampling fluctuation. This is a one-sided concentration statement for a bounded, adaptively generated classical sequence—precisely the setting of the Azuma–Hoeffding inequality for bounded-difference martingales. It is *not* a statement about an adversary's uncertainty regarding a secret, so no conditional min-entropy needs to be accumulated; entropy accumulation, which lower-bounds the smooth min-entropy of an extracted string against quantum side information, is the tool for key distribution and randomness expansion, not for certifying a function of a classical test statistic. We return to this point below, as this is where a natural but mismatched choice of machinery would overstate the cost.

*The martingale bound against a memoryful adversary.* Let $\mathcal{F}_{i-1}$ denote all information available before round $i$—past inputs, outcomes, and the public record—and write $\xi_i = m_i - \mathbb{E}[m_i \mid \mathcal{F}_{i-1}]$. Regardless of internal memory and adaptive strategy of the device, each $m_i$ is a bounded random variable in $\{0,1\}$ and $\xi_i$ is a martingale difference with $\mathbb{E}[\xi_i \mid \mathcal{F}_{i-1}] = 0$ and $|\xi_i| \le 1$; adaptive memory can shift the conditional mean $\mathbb{E}[m_i \mid \mathcal{F}_{i-1}]$ round to round, but cannot break this bounded-difference structure. The one-sided Azuma–Hoeffding inequality then gives, except with probability $\delta$,

$$\bar{A}_n \;\le\; \widehat{A} \;+\; t(\delta,n), \qquad t(\delta,n) = \sqrt{\frac{\ln(1/\delta)}{2n}}, \tag{11}$$

where $\bar{A}_n = 1 + \frac{1}{n}\sum_i \mathbb{E}[m_i \mid \mathcal{F}_{i-1}]$ is the time-averaged agreement achieved by the certified mechanism. No independence or memorylessness assumption is used; conditioning on $\mathcal{F}_{i-1}$ absorbs the adversary's memory directly, which is exactly why the martingale bound is the natural and sufficient tool here.

*Mapping the certified statistic through the certificate.* Since $\varepsilon_{\mathrm{cert}}(A) = 2\,\mathrm{artanh}(A-1)$ increases with $A$, a sound *upper* bound on the privacy parameter is obtained by evaluating the certificate at the agreement value least favorable to privacy, the upper confidence end Equation (11), since a higher agreement means a less private mechanism. Writing the finite-sample certificate accordingly,

$$\varepsilon_{\mathrm{cert}}^{(n)} \;=\; 2\,\mathrm{artanh}\Big(\widehat{A} + t(\delta, n) - 1\Big), \tag{12}$$

we have $\varepsilon(\mathcal{A}) \le \varepsilon_{\mathrm{cert}}^{(n)}$ except for probability $\delta$, against an arbitrary memoryful adaptive adversary.

**Corollary 1** (Finite-statistics soundness on the target family) *In the target family, where the certificate is the monotone map $\varepsilon_{\mathrm{cert}}(A) = 2\,\mathrm{artanh}(A-1)$ of the scalar agreement Equation (9), the finite-sample certificate Equation (12) satisfies $\varepsilon(\mathcal{A}) \le \varepsilon_{\mathrm{cert}}^{(n)}$ with probability at least $1-\delta$ against an adversary that is memoryful and adaptive over rounds, with half-width $t(\delta, n) = \sqrt{\ln(1/\delta)/2n} \to 0$. The constants quoted here are those of this closed-form map; for a general mechanism, the same martingale argument controls the scalar agreement, but the map from the agreement to $\varepsilon$ is the worst-case program of Theorem 2 rather than the analytic form, and the finite-sample bound is obtained by propagating the confidence band through that program.*

The sign is the one soundness requires: the finite-statistics correction *adds* to the estimated agreement before the monotone map, so it can only make the certified mechanism appear *less* private, never more. We verify the direction and coverage numerically against an adaptive memoryful device (Section C): the empirical coverage of $\varepsilon_{\mathrm{cert}}^{(n)} \ge \varepsilon(\mathcal{A})$ meets the nominal $1-\delta$, and as $n \to \infty$ the penalty vanishes as $1/\sqrt{n}$ and $\varepsilon_{\mathrm{cert}}^{(n)}$ converges to the true parameter from above. In the high-agreement regime where an honest device operates, the per-round variance $\mathbb{E}[m_i \mid \mathcal{F}_{i-1}](1-\mathbb{E}[m_i \mid \mathcal{F}_{i-1}])$ is small, and a variance-adapted Freedman bound Section C tightens Equation (11) further; we quote the transparent Hoeffding form here.

*Sample-size cost.* The finite-size coefficient is $c = \sqrt{\ln(1/\delta)/2} \approx 2.6$ at $\delta = 10^{-6}$, so certifying the privacy parameter to within a few percent of its true value takes of order $10^4$ probe rounds—a modest cost for a guarantee that holds against an adversarial mechanism, and far smaller than an entropy-accumulation analysis would suggest, because no entropic second-order or smoothing penalties enter the concentration of a classical statistic. This is the finite-statistic counterpart of Theorem 2: together they constitute an effective privacy-certifying guarantee: acceptance after $n$ rounds certifies, with confidence $1-\delta$ and against an arbitrary memoryful adversary, that the untrusted mechanism has a privacy parameter not larger than $\varepsilon_{\mathrm{cert}}^{(n)}$.

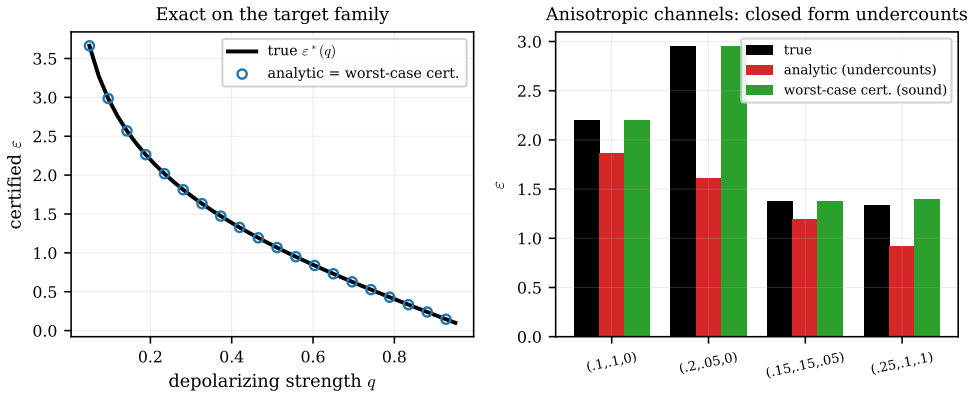



**Fig. 3** The privacy certificate and its scope. *Left:* on the depolarizing (EITFF) family the analytic closed form Equation (10) and the worst-case privacy certificate Equation (6) coincide and track the true value $\varepsilon^*(q)$. *Right:* on anisotropic mechanisms the closed form undercounts—reporting a mechanism as more private than it is, hence unsound as a general certificate—while the worst-case certificate remains a valid upper bound.

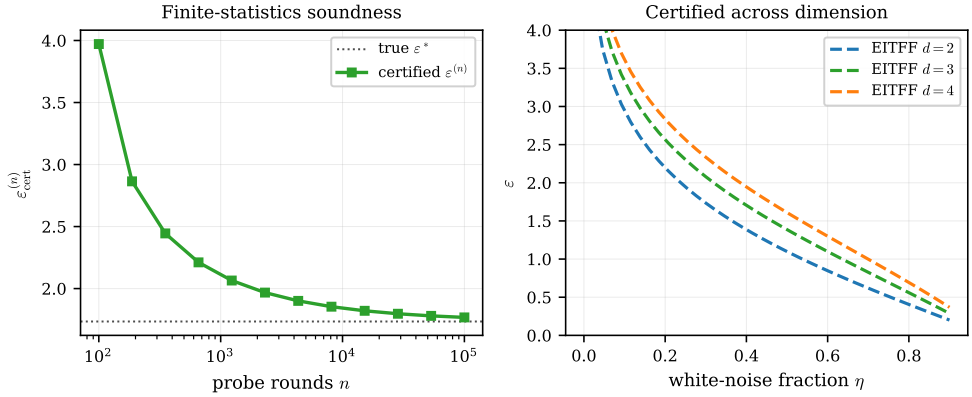



**Fig. 4** Robustness of the certificate. *Left:* under finite statistics the certified value $\varepsilon^{(n)}_{\mathrm{cert}}$ converges to the true parameter from above as the number of probe rounds grows, with the Azuma–Hoeffding half-width vanishing as $1/\sqrt{n}$. *Right:* the certificate extends across dimension, tracking the exact QLDP value $\varepsilon^*(d, q)$ for the EITFF mechanisms at $d = 2, 3, 4$ and degrading smoothly under added white noise.

# 6 Preservation of the quantum advantage

A certificate that destroyed the very advantage that motivated the use of quantum mechanisms would be of little value. We therefore ask whether the quantum utility advantage of locally private mechanisms survives on the *certified* ensemble, whether a mechanism can be simultaneously certified private and usefully more informative than any classical mechanism at the same privacy level. It can, within a quantifiable noise regime, and we are careful to state the advantage only where it provably holds.

## 6.1 The advantage and its scope

The utility of a locally private mechanism is measured by how distinguishable its outputs are across inputs—operationally, by an ensemble-distinguishability quantity such as the Holevo information [47], rather than a scalar error. The underlying task is quantum state discrimination, whose asymptotic error rates are governed by the quantum Chernoff [48] and quantum Stein [49, 50] bounds. For optimal mechanisms, the quantum-to-classical utility ratio is known in closed form. The equi-isoclinic tight-fusion-frame (EITFF) mechanisms of Yoshida [8], which are the QLDP-optimal mechanisms in the high-privacy regime, achieve

$$\frac{Q}{C} \;=\; \frac{n(n-1)}{2\,\lfloor n/2\rfloor\,\lceil n/2\rceil} \;=\; 2-\frac{1}{\lceil n/2\rceil}, \tag{13}$$

for data of alphabet size $n$, rising from unity toward the asymptotic value 2. This advantage is genuinely quantum and genuinely bounded: for binary data it is absent—no quantum advantage exists for binary private statistical inference, the no-go of Yoshida and Hayashi [7, 8]—and Equation (13) accordingly returns $Q/C = 1$ at $n = 2$. We claim an advantage only for $n \geq 3$, where Equation (13) exceeds unity. We are explicit about what this half of the paper takes on trust. The *optimality* of the EITFF mechanisms in the high-privacy regime is established in a single recent work [8], which we adopt as given and do not re-derive; our advantage claims inherit that optimality characterization. However, we do not rest the advantage on the cited formula alone: we reconstruct the mechanisms and recompute the quantum-to-classical ratio Equation (13) directly from the Holevo information of the encoded ensembles (Section D), so the *value* of the advantage and its survival under noise are verified here rather than imported. Should the optimality characterization be revised, the survival analysis of Section 6.2 continues to hold for whichever mechanisms are optimal, with the ratio recomputed on them.

## 6.2 Survival on the certified ensemble

The question this paper must answer is not whether the advantage exists for an ideal mechanism—that is known—but whether it survives once the mechanism is the one actually certified, carrying the noise that certification tolerates. We model the certification and device overhead conservatively, as an additional depolarizing layer of strength $p$ applied to the quantum ensemble while the classical baseline is held at its optimum; this is the least favorable account for the quantum side. Computing the Holevo-information ratio directly on the resulting certified ensemble—not by substituting into Equation (13), but by evaluating the utility of the noisy mechanism—yields the survival behavior summarized in Figure 6. This robustness analysis is our own contribution: the optimal-mechanism literature fixes the noiseless ratio but does not treat survival under a depolarizing certification overhead, and we derive the thresholds below directly (Section D).

Two statements result, and we lead with the robust one. First, a genuine quantum advantage, $Q/C > 1$, persists up to a substantial noise level: the ratio remains above unity for depolarizing strengths up to $p \approx 0.18$ for $n = 3, 4$, rising to $p \approx 0.24$ for larger

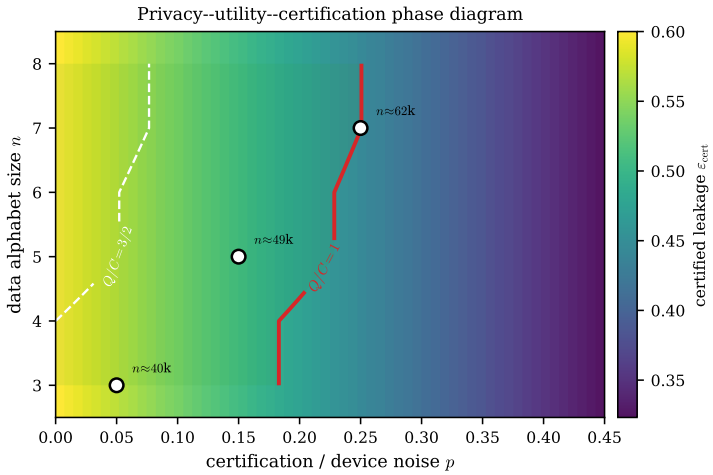



**Fig. 5** Privacy–utility–certification phase diagram, the paper's three quantitative guarantees in one view. The color field is the worst-case certified leakage $\varepsilon_{\mathrm{cert}}$, which decreases smoothly as the certification/device noise $p$ grows (more noise, more privacy). The red curve is the advantage-survival boundary $Q/C = 1$: to its left a certified mechanism retains a strict quantum utility advantage over every classical mechanism at the same privacy level, and the boundary moves to higher noise as the alphabet size $n$ grows (from $p \approx 0.18$ at $n = 3, 4$ to $p \approx 0.25$ at $n = 7, 8$); the dashed white curve is the stronger $Q/C = 3/2$ threshold. The markers give the order of probe rounds needed for the finite-statistics certificate to reach a few percent of the true value at $\delta = 10^{-6}$ (tens of thousands of rounds). A certified-private mechanism, a surviving quantum advantage, and a finite-sample-usable certificate thus coexist in the shaded region.

alphabets. A certified-private mechanism therefore retains a strict utility advantage over every classical mechanism at the same privacy level across a wide and realistic noise window. Second, the *strict* ratio $Q/C \geq 3/2$ of Equation (13) is more fragile: since it is the noiseless value of $n = 3, 4$, any added noise erodes it there, and a margin for the $3/2$ ratio under noise appears only at higher alphabet size ($n \geq 5$). We report both and base the paper's claim on the first: the certified quantum privacy–utility advantage $Q/C > 1$ survives certification overhead up to roughly 18–24% depolarizing noise, with the stronger $3/2$ advantage available in the low-noise regime and improving with $n$.

## 6.3 Coexistence with certification

The advantage-survival window overlaps the regime in which the certificate of Section 5 is non-vacuous, so the two guarantees hold simultaneously: in this regime the verifier both certifies a finite privacy parameter and retains a quantum utility advantage (Figure 5). This is the substantive content of the section—a certified mechanism need not be a useless one. We defer the formal statement that , *conditioned on acceptance*, the advantage holds together with the certified privacy bound except with the protocol's failure probability, to Section 8, where it is assembled from this survival result and the finite-statistics soundness of Section 5.5 into a single joint guarantee. The numerical construction of the EITFF mechanisms, the direct Holevo-information

computation reproducing Equation (13), and the survival thresholds are detailed in Section D.

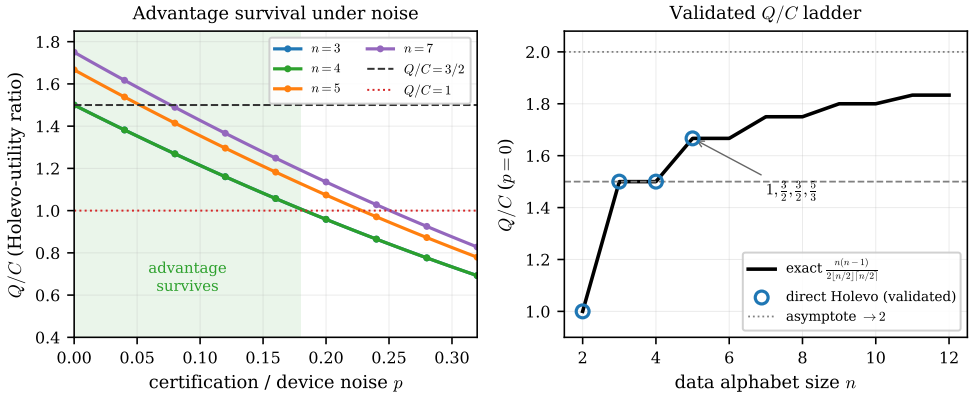



**Fig. 6** Survival of the quantum advantage on the certified ensemble. *Left:* the Holevo-information ratio $Q/C$ versus certification/device noise $p$ for alphabet sizes $n = 3, 4, 5, 7$, with the any-advantage threshold $Q/C = 1$ and the strict threshold $Q/C = 3/2$; the shaded region marks where a genuine advantage survives ($p \lesssim 0.18$–$0.24$). *Right:* the closed-form ratio Equation (13) (curve) with the directly computed Holevo-information values (points) at $n = 2, 3, 4, 5$ on $1, \frac{3}{2}, \frac{3}{2}, \frac{5}{3}$, approaching the asymptote 2.

# 7 Composable security

A certificate is most useful when it can be invoked as a subroutine inside a larger protocol without unraveling its guarantee. The standalone soundness of Section 5 is not enough for this: composition can amplify small errors, and a guarantee proved in isolation need not survive embedding. We therefore recast the certification protocol in the framework of abstract cryptography [51], where security means that the real protocol is indistinguishable from an ideal resource that is private by construction and where security errors are composed additively. This is the property that allows certified-private aggregation to serve as a building block for the federated-learning application of Section 8.

## 7.1 The ideal resource and the real protocol

In the abstract-cryptography framework one specifies an *ideal resource* that exhibits the desired guarantee by construction, and proves that the real protocol *constructs* it: that no distinguisher, interacting with all interfaces including the adversary's, can tell the real protocol from the ideal resource dressed by a simulator, beyond a distinguishing advantage $\varepsilon$. We write $\pi\, R \xrightarrow{\varepsilon} S$ for "the protocol $\pi$ constructs the resource $S$ within $\varepsilon$," meaning that there exists a simulator $\sigma$ on the adversarial interface with $d(\pi R,\ \sigma S) \le \varepsilon$, where $d$ is the distinguishing advantage [51, 52].

The ideal resource for our task, $S_{\mathrm{CPA}}$, captures certified-private aggregation directly. It is an abort-or-deliver resource: clients submit their input, and the resource raises an *abort* flag $\perp$, or outputs the pair (aggregate, certified parameter $\varepsilon_{\mathrm{cert}}$) while guaranteeing, by construction, an accepted branch that satisfies the reported $\varepsilon_{\mathrm{cert}}$-QLDP specification, so that neighboring input tuples are indistinguishable in the sense of Definition 1. The abort probability is matched to that of the real protocol, but is not itself revealed as an output. This structure follows the abort-or-deliver form of composable device-independent certification [4] and the leak-only-permitted-function form of composable privacy [53], with the permitted leakage quantified by the certified QLDP parameter. The feature that the ideal resource outputs a *certified privacy parameter* alongside the abort flag is, to our knowledge, new.

The ideal accepted branch represents the privacy specification certified by the protocol; the construction does not require the real accepted mechanism to be identical to a unique reconstructed Choi state, only indistinguishable from the ideal resource up to the stated trace-distance error.

The real protocol $\pi$ is the certification test of Section 5: the trusted, informationally complete probes are sent through the untrusted mechanism, the trusted measurement is performed, the certified parameter is computed by Equation (6) with the finite-statistics confidence bound of Equation (12), and the verifier delivers $(\,\text{aggregate}, \varepsilon_{\mathrm{cert}}^{(n)})$ on acceptance or $\perp$ on rejection. The construction is formulated in the abstract-cryptography framework with an explicit simulator and additive error accounting: we prove in Section E that $\pi$ constructs $S_{\mathrm{CPA}}$ within the error Equation (15). We exhibit the simulator explicitly and bound the distinguishing advantage by a trace distance, using the identity

$$d(\rho, \sigma) \;=\; \tfrac{1}{2}\|\rho - \sigma\|_1, \tag{14}$$

Helstrom's theorem that the optimal distinguishing advantage between two states equals their trace distance. The simulator is an explicit converter on the adversarial interface: it runs the device's own black-box strategy on the trusted probe rounds, recomputes the public agreement statistic and the accept/abort flag exactly as the protocol does—so the ideal resource aborts precisely when the real one does—and forwards the public certified pair. It draws nothing from the clients' private inputs, since the certified value is a function of the public statistics and the trusted probes alone, and so has exactly the information available to the real adversary. With it, the construction within $\varepsilon$ reduces to the trace-distance bound proved in Section E.

## 7.2 The composable error

The distinguishing advantage of the construction is the sum of three allocated error probabilities,

$$\varepsilon_{\mathrm{comp}} \;=\; \varepsilon_{\mathrm{compl}} \;+\; \delta_{\mathrm{sound}} \;+\; \varepsilon_{\mathrm{est}}. \tag{15}$$

Here $\varepsilon_{\mathrm{compl}}$ limits the probability that an honest run nevertheless aborts; $\delta_{\mathrm{sound}}$ is the finite-statistics soundness error of Section 5.5, the probability that an accepted certificate is wrong, equal to the martingale tail $e^{-2nt^2}$ of Equation (11); and $\varepsilon_{\mathrm{est}}$ is the residual error in estimating the certified parameter from the accepted statistics. Here $\varepsilon_{\mathrm{est}}$ denotes the residual numerical error of the acceptance-branch certificate

evaluation; in the closed-form target-family implementation used in the following, it vanishes $\varepsilon_{\text{est}} = 0$, since the certificate is the exact analytic map of the agreement statistic. Because the protocol extracts no key and performs no privacy amplification, no smoothing parameter is entered: the composable error is the sum of the abort, soundness, and estimation probabilities and nothing more. The load-bearing inequality, proved in Section E and not assumed, is that the real and ideal systems lie within $\delta_{\text{sound}} + \varepsilon_{\text{est}}$ in trace distance—a bound on the *joint* probability that the protocol accepts and the certified mechanism is nonetheless less private than claimed, not a conditional one, so that the abort branch contributes nothing because the two systems abort together; adding the honest-abort term $\varepsilon_{\text{compl}}$ and invoking Equation (14) gives Equation (15). We confirmed each term and the combined inequality numerically in the certified ensemble (Figure 7): the joint trace distance (after cancellation of the shared abort branch) stays within the allocated budget at every sample size and scales as predicted by the finite-statistics perturbation.

### 7.3 Two budgets that must not be conflated

The protocol carries two error parameters of fundamentally different type, and keeping them separate is essential to stating the guarantee correctly. The *privacy budget* $\varepsilon_{\text{cert}}$ is the QLDP parameter itself; it lives in the *specification* of the ideal resource and quantifies how much neighboring input can be distinguished. The *security error* $\varepsilon_{\text{comp}}$ is the distinguishing advantage of Equation (15); it measures how far the real protocol is from *meeting* that specification. The first is a privacy level and the second is a probability that the privacy level is not delivered; they have different units and different operational meanings.

Both compose additively, but along their own axes. Across $R$ rounds of aggregation, the privacy budget accumulates as $\varepsilon^{\text{tot}}_{\text{cert}} = \sum_r \varepsilon^{(r)}_{\text{cert}}$ by the QLDP composition theorem [6], while the security error accumulates as $\varepsilon^{\text{tot}}_{\text{comp}} = \sum_r \varepsilon^{(r)}_{\text{comp}}$ by the abstract-cryptography composition theorem [51]. We state the scope of the privacy accounting explicitly, as it bounds one of our broadest claims: the additive QLDP rule is the guarantee for rounds implemented *independently*, as a tensor product of per-round mechanisms; for federated rounds whose joint implementation correlates the per-round channels outside that structure, additive accounting should be read as applying to the independently-implemented regime, and a correlated-implementation analysis is left open. The two budgets are reported as a pair $\left(\varepsilon^{\text{tot}}_{\text{cert}}; \varepsilon^{\text{tot}}_{\text{comp}}\right)$ and never summarized, exactly as key-distribution protocols report a key length and a security error separately. Conflating them—adding a privacy level to a failure probability—would be a category error; we keep them distinct throughout, and the federated-learning guarantee of Section 8 tracks the pair.

## 8 Certified-private aggregation

We now assemble the pieces. The certificate of Section 5 bounds the privacy of an untrusted mechanism; the result of Section 6 shows that the quantum advantage survives in the certified mechanism; and the composable security of Section 7 allows the

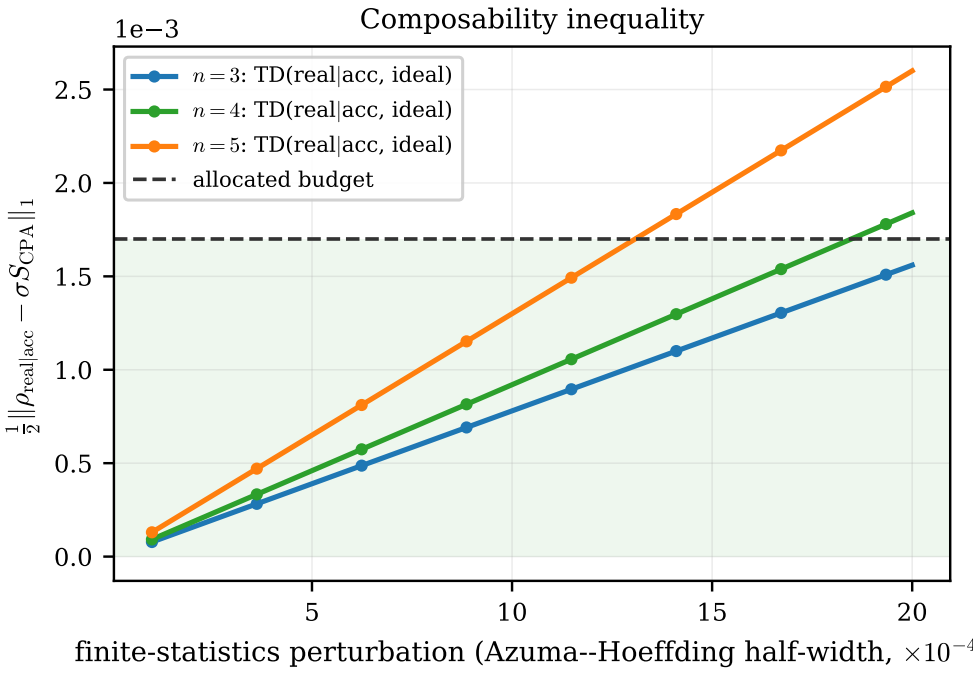



**Fig. 7** Composable-security verification. The composability inequality holds numerically on the certified ensemble: the joint trace distance after shared-abort cancellation—equivalently $p_{\text{acc}}$ times the accept-branch distance—between the real protocol and the ideal $\varepsilon_{\text{cert}}$-QLDP resource ($n = 3, 4, 5$) stays within the allocated composable budget and scales linearly with the finite-statistics perturbation, as the proof of Proposition 4 predicts. The privacy budget and the security error accumulate on independent axes (Section 7.3) and are not shown together, to avoid suggesting they are commensurable.

certificate to be used within a larger protocol. Combining them gives a single guarantee for federated aggregation and precisely locates the sense in which it has no classical counterpart. This section is an application: it assembles results already established into the aggregation setting that motivates the work, and introduces no technical claim beyond their composition. Federated aggregation is where the composable certificate is naturally used, not a source of additional assumptions or results.

## 8.1 Certified-private aggregation as a subroutine

In a round of federated aggregation, clients hold local data and a server is charged with combining locally privatized contributions into an aggregate—the setting of classical secure aggregation [54], here with the privatization performed by an untrusted quantum mechanism. Verifiable delegation of quantum computation to an untrusted party has a parallel history in blind and verifiable quantum computation [55], and the federated use of quantum data is itself an active direction [56]. The certified-private-aggregation resource $S_{\text{CPA}}$ of Section 7.1 is exactly the subroutine this round needs: invoking it, the clients obtain either an abort—in which case the round is discarded

and no data are exposed—or an aggregate together with a certified bound on how private the server's mechanism actually was, with no trust placed in the server's hardware. Because the resource is composable, a learning protocol may invoke it once per round in many rounds, and the guarantees accumulate in the controlled way established in Section 7.3: the privacy budget and the security error each add along their own axis, tracked as the pair $\left(\varepsilon^{\mathrm{tot}}_{\mathrm{cert}}; \varepsilon^{\mathrm{tot}}_{\mathrm{comp}}\right)$.

## 8.2 The joint guarantee

The three results combine into one statement, which is the operational summary of the paper.

**Theorem 3** (Certified-private aggregation with surviving advantage) *Run the certified-private-aggregation protocol for $R$ rounds, probing the untrusted server mechanism with trusted, informationally complete inputs. Then, conditioned on acceptance, except with a composable failure probability $\varepsilon^{\mathrm{tot}}_{\mathrm{comp}} = \sum_r \varepsilon^{(r)}_{\mathrm{comp}}$:*

*(i)* (Certification.) *The effective aggregation mechanism of the server is $\varepsilon^{\mathrm{tot}}_{\mathrm{cert}}$ -QLDP, with $\varepsilon^{\mathrm{tot}}_{\mathrm{cert}} = \sum_r \varepsilon^{(r)}_{\mathrm{cert}}$ by additive composition. The per-round value is a sound upper bound in general and is tight on the target family* (Theorem 2, Proposition 1, and Corollary 1).

*(ii)* (Advantage.) *For data of alphabet size $n \geq 3$, the certified mechanism retains a quantum utility advantage $Q/C > 1$ over every classical mechanism at the same privacy level, provided that the per-round certification and device noise stays below $p \approx 0.18$–$0.24$* (Section 6.2).

*(iii)* (Composability.) *The protocol constructs the ideal certified-privacy resource $S_{\mathrm{CPA}}$ within $\varepsilon^{\mathrm{tot}}_{\mathrm{comp}}$, so the guarantee holds when the protocol is used as a subroutine within any larger protocol* (Section 7.1).

*The privacy budget $\varepsilon^{\mathrm{tot}}_{\mathrm{cert}}$ and the security error $\varepsilon^{\mathrm{tot}}_{\mathrm{comp}}$ are reported as a pair and are never summed.*

The theorem is a composition of already established results—the certificate (Theorem 2), its finite-statistics soundness (Equation (12)), the advantage survival (Section 6.2) and the composable construction (Section 7.1)—and introduces no new claim. Its force is in the combined guarantee: a single accepted run delivers, with one quantified failure probability, both a certified privacy level for an untrusted mechanism and a surviving quantum advantage in a form that composes. The scope inherited from the parts is carried faithfully: the advantage clause is asserted only for $n \geq 3$, in keeping with the binary no-go, and the privacy clause is tight only on the optimal family while remaining a valid upper bound in general.

## 8.3 The classical-separation proposition

It is worth stating precisely what is and is not claimed to be impossible classically. We state the contrast as a scoped proposition, citing the two classical no-go results at exactly the strength they are proved.

**Proposition 3** (Classical limits of currently known verifier models) *No currently known classical verifier model certifies the local differential-privacy parameter of an untrusted mechanism while remaining simultaneously* prover-free *(using only black-box query access to the mechanism, with no semi-trusted prover),* information-theoretically sound *(sound against a computationally unbounded adversary, under no cryptographic assumption) and assumption-free. More specifically, black-box verification faces property-testing barriers, whereas interactive verifiable approaches achieve soundness only by introducing a prover and computational assumptions.*

*Justification* The two ways a classical verifier might proceed are ruled out. *(i) Prover-free, black-box verification.* If the verifier has only query access to the mechanism, then by the property-testing impossibility of Gilbert and McMillan [1] any differential-privacy guarantee that can be efficiently verified from the mechanism's input–output behavior can also be efficiently *broken*: there exist two databases that the verifier can efficiently distinguish, defeating the privacy it sought to confirm. Efficient prover-free verification of DP is therefore self-defeating, and the lower bounds of [1] show that the query complexity of verification is otherwise prohibitive. The same obstruction recurs in the recent black-box auditing of Shi and Ma [42], who proved that no prover-free black-box procedure can control both false-accept and false-reject errors without further structural assumptions, so an assumption-free one-sided *upper* certificate is not available by this route. *(ii) Interactive verification with a prover.* If instead a prover assists, then by the impossibility theorem of Biswas and Cormode [2, Thm. 5.2] a constant-round interactive protocol for a DP mechanism that satisfies their notion of verifiable DP *cannot* have unconditional (information-theoretic) soundness together with statistical zero-knowledge; achieving verifiability provably requires computational hardness assumptions. The same reliance is visible in the certification protocol of Bell, Goldwasser, Kim and Watson [3], which certifies private probabilistic mechanisms through an interactive protocol resting on a prover and cryptographic assumptions rather than information-theoretic soundness. Hence, neither route delivers prover-free *and* information-theoretically sound certification, which is the combination asserted. □

The guarantee of Theorem 3 occupies exactly the position Proposition 3 identifies as beyond currently known classical verifier models: it certifies the privacy parameter of an untrusted mechanism using trusted, informationally complete probe states and a characterized measurement, with no prover, and is sound against a computationally unbounded server. What makes this possible is not a computational assumption but the quantum structure of the probe: trusted informationally complete preparation lets the verifier read the privacy functional—a quantity defined by quantification over all measurements—off the mechanism's action on a tomographically complete input set, which no classical black-box query access to the mechanism can supply. The classical comparison is therefore precise: known classical verifier models cannot provide the same *combination*: certification of an uncharacterized mechanism's privacy parameter, prover-free operation, and information-theoretic soundness (Proposition 3) without relaxing at least one requirement. This is not the stronger—and false—claim that differential privacy can never be verified by classical means. The composable formulation of Section 7 makes the contrast rigorous by fixing exactly which resource the protocol constructs.

# 9 Discussion and outlook

We have shown that the certifying power of trusted quantum probing can be turned on a privacy guarantee: using trusted, informationally complete probes and a characterized measurement, a verifier can certify an upper bound on the quantum local differential privacy parameter of an untrusted mechanism, the certified mechanism retains a quantum utility advantage, and the certificate composes. We close by setting out the assumptions on which these results rest, the respects in which they are scoped, and the directions they open.

## 9.1 Assumptions and limitations

### 9.1.1 Trust in the verifier's apparatus.

The certificate trusts the verifier's own preparation and measurement, while leaving the server's mechanism adversarial. The measurement assumption is not an incidental convenience but a necessity: Theorem 1 shows that relaxing it to a fully device-independent setting, where the measurement is also untrusted, makes the certification of the privacy functional vacuous, because any privatization can be absorbed into the measurement. Trusting a characterized informationally complete measurement is the minimal addition that breaks this obstruction and renders the mechanism's privacy parameter identifiable from the statistics; it is the same trusted-apparatus assumption underlying quantum process estimation. The genuine and deliberate asymmetry of the model is that the untrusted party is the mechanism itself—the one whose privacy claim is in question—so that a miscalibrated or compromised verifier apparatus lies outside the guarantee, but an adversarial mechanism does not: the certificate is sound against every channel consistent with the data.

### 9.1.2 Scope of the closed-form witness.

The general certificate is the convex worst-case program Equation (6), which never underestimates the privacy parameter. The closed form Equation (10), by contrast, is exact only in the depolarizing family and *undercounts* outside it (Remark 1); it is a convenience valid in the optimal-mechanism family, which is the object of certification here, not a general tool. Any deployment certifying mechanisms outside this family must use the semidefinite program directly and propagate the finite-statistics confidence band through that worst-case optimization rather than through the closed-form map.

### 9.1.3 Fragility of the strict advantage.

The quantum advantage we claim robustly is $Q/C > 1$, which survives certification and device noise up to roughly 18–24%. The strongest ratio $Q/C \geq 3/2$ is the noiseless, low-privacy value and is more fragile: it erodes under any noise for the smallest alphabets and acquires a noise margin only at larger alphabet size. We have stated the advantage only for non-binary data, where it provably exists; for binary data there is no quantum advantage, and our claims respect that boundary [7, 8]. We also note that

the optimality of the mechanisms carrying the advantage rests on a single recent characterization [8]; we reproduce the resulting utility ratio independently (Section D), so our quantitative claims do not depend on that characterization being the last word, but the framing of these mechanisms as *optimal* does.

### 9.1.4 Composition and accounting.

The privacy budget is tracked under basic additive composition, which is insensitive to correlations between rounds [6]; we make no sharper advanced-composition claim, and a tighter accounting of the privacy budget over many rounds is left open. Under this additive rule, the budget grows linearly in the number of rounds, so the number of times the certificate can be composed at a fixed target privacy is bounded—a genuine limitation that an advanced-composition analysis would relax. On the security side, the composable error Equation (15) is proved in full in Section E: we give the ideal resource and the simulator explicitly, bound the joint distinguishing advantage by a trace distance through Helstrom's identity, and assemble the three error terms as separate completeness and soundness statements. What the precedent of composable device-independent certification [4] provides is the architecture; the bounds themselves are carried out for the certified-privacy resource. The construction of an ideal resource that outputs a certified privacy parameter is, to our knowledge, without direct precedent.

### 9.1.5 No experimental claim.

All results are theoretical. The numerical studies validate the certificate, the advantage, and the composability inequality on the optimal mechanisms, but we make no claim about a physical implementation; preparation side channels and hardware imperfections are outside the model.

## 9.2 Outlook

Several directions follow naturally. The most immediate is a proof-of-principle demonstration. The protocol asks only for trusted informationally complete preparation and a characterized informationally complete measurement on an untrusted channel; these are the same primitives used in quantum process tomography, so no capability beyond the present photonic platforms is required in principle. We stop short of claiming such an experiment is routine: a faithful demonstration must also control the preparation and measurement side channels that our model places outside its scope, and it is the honest characterization of those imperfections, rather than the probe and readout themselves, that would set the real difficulty. With that caveat, an experimental certification of a privacy parameter would be the natural counterpart to device-independent certifications of states and channels already demonstrated [4]. On the theory side, three questions stand out: whether the certificate can be extended soundly beyond the optimal-mechanism family with a tractable witness rather than the full semidefinite program; whether the finite-statistics analysis can be tightened, or the advantage-survival threshold improved, by propagating the actual soundness margin into the noise model rather than treating certification overhead as worst case; and whether

the strict $Q/C \geq 3/2$ advantage can be made robust at small alphabet size. More broadly, the certificate is one instance of a general possibility that this work is the first to realize—that the certification machinery of quantum cryptography can be turned on operational *privacy and information-theoretic functionals*, and not only on states, dimensions, and structural channel properties. Mapping out which such functionals admit composable, soundly certifiable resources of this kind is, we believe, a direction worth pursuing, and is the broader program this paper opens.


## Acknowledgements

No specific funding was received for this work.


## Statements and Declarations


**Funding.** The authors received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

**Competing interests.** The authors declare no competing interests.
**Ethics approval and consent to participate.** Not applicable.
**Consent for publication.** Not applicable.
**Data availability.** This is a theoretical work and reports no experimental data. The numerical studies supporting the figures and checks can be reproduced from the authors' code.
**Code availability.** The code used to generate the figures and numerical checks is available from the authors on reasonable request and will be deposited in a public repository with a permanent identifier upon publication.
**Author contributions.** Clifford N. Jones conceived the project, developed the certification framework, carried out the numerical validation and proofs, and drafted the manuscript. Khem Poudel supervised the project, provided conceptual guidance, and contributed to manuscript review and revision.

## Appendix A Proof of the device-independent no-go

*Proof of Theorem 1* Let $\mathcal{A}$ be the qubit depolarizing channel $\mathcal{A}(\rho) = (1-q)\rho + q\,\frac{\mathbb{I}}{2}$ with $0 < q < 1$. It is unital, $\mathcal{A}(\mathbb{I}) = \mathbb{I}$, trace-preserving, and self-adjoint with respect to the Hilbert–Schmidt inner product, $\mathcal{A}^\dagger = \mathcal{A}$; its QLDP value is $\varepsilon(\mathcal{A}) = \ln\big((2-q)/q\big)$ by (2), which tends to 0 as $q \to 1$, so $\mathcal{A}$ can be made arbitrarily private.

Fix any POVM $\{M'_b\}$ on the output space, $M'_b \geq 0$, $\sum_b M'_b = \mathbb{I}$, and define $M_b \coloneqq \mathcal{A}^\dagger(M'_b) = \mathcal{A}(M'_b)$. Each $M_b$ is positive because $\mathcal{A}^\dagger$ is completely positive, and $\sum_b M_b = \mathcal{A}^\dagger\big(\sum_b M'_b\big) = \mathcal{A}^\dagger(\mathbb{I}) = \mathbb{I}$ since $\mathcal{A}$ is unital; hence $\{M_b\}$ is a valid POVM. By the definition of the adjoint, for every input $\tau_x$,

$$\operatorname{Tr}\big[M_b\,\mathrm{id}(\tau_x)\big] = \operatorname{Tr}\big[\mathcal{A}^\dagger(M'_b)\,\tau_x\big] = \operatorname{Tr}\big[M'_b\,\mathcal{A}(\tau_x)\big],$$

which is (4). Thus the identity channel—for which $\varepsilon(\mathrm{id}) = \infty$, since it maps orthogonal inputs to perfectly distinguishable outputs—reproduces, under the POVM $\{M_b\}$, exactly the statistics produced by the private mechanism $\mathcal{A}$ under $\{M'_b\}$. Any privacy parameter assigned on the basis of device-independent statistics alone must therefore be consistent with $\varepsilon = \infty$, giving $\sup \varepsilon = \infty$ and establishing vacuousness. □

The construction uses only self-duality and unitality, so it applies verbatim to any self-dual unital mechanism, not merely the qubit depolarizing channel; the depolarizing case is singled out only because it is the QLDP-optimal mechanism and so makes the gap between certified and true privacy as large as possible.

# Appendix B Worst-case certificate and the analytic witness

This appendix gives the convex program underlying the certified parameter Equation (6) and derives the closed form Equation (10) on the depolarizing family.

## B.1 The worst-case program.

Write the mechanism through its Choi operator $J = (\mathcal{I} \otimes \mathcal{A})(|\Omega\rangle\langle\Omega|)$, with $|\Omega\rangle = \sum_i |ii\rangle$ the unnormalized maximally entangled vector. Complete positivity of $\mathcal{A}$ is $J \succeq 0$ and trace preservation is the partial-trace condition $\mathrm{Tr}_{\mathrm{out}}\, J = \mathbb{I}$. The trusted, informationally complete probes $\{\tau_x\}$ and the trusted measurement $\{E_b\}$ produce, by Equation (5), the statistics $p(b\,|\,x) = \mathrm{Tr}[(\tau_x^{\mathsf{T}} \otimes E_b)J]$, which are linear in $J$. The certified parameter is the largest privacy value over all Choi operators reproducing the observed statistics,

$$\varepsilon_{\mathrm{cert}} = \sup_{J \succeq 0} \Big\{ \varepsilon(J) : \mathrm{Tr}_{\mathrm{out}}\, J = \mathbb{I}, \quad \mathrm{Tr}\left[(\tau_x^{\mathsf{T}} \otimes E_b)J\right] = p(b\,|\,x) \quad \forall b, x \Big\}. \tag{B.1}$$

The feasible set is the intersection of the positive-semidefinite cone with affine constraints, hence convex and compact. The objective $\varepsilon(J)$ is the QLDP value: by Section 3.1 it is the supremum over neighboring inputs $\rho, \sigma$ of $\ln \lambda_{\max}\big[\mathcal{A}(\sigma)^{-1/2}\mathcal{A}(\rho)\,\mathcal{A}(\sigma)^{-1/2}\big]$, the measured hockey-stick divergence of the output states. This objective is *convex* in $J$: the hockey-stick divergence is jointly convex [5] and $\mathcal{A}(\rho)$ is linear in $J$. Consequently Equation (B.1) is a convex *maximization*—a convex objective over a convex set—whose optimum is attained at an extreme point of the feasible spectrahedron and which is NP-hard in general; it is *not* reducible to a single semidefinite program. This is the source of both the strength and the computational care required by the certificate. For a *fixed* $J$ the inner divergence $\varepsilon(J)$ is itself a semidefinite program [44], with a dual whose every feasible point is a rigorous upper bound (below). The outer supremum over a nontrivial consistent set is not asserted to be efficiently solvable in full generality. In the regimes treated in this paper, it is evaluated either because the consistent set collapses to a single $J$, or because the remaining ambiguity has a low-dimensional parametrization, as in the coarse depolarizing example. In any deployment beyond these regimes, a numerical relaxation must be conservative from above to preserve the never-undercounts guarantee. Because the bound is a supremum over the consistent set, $\varepsilon_{\mathrm{cert}} \geq \varepsilon(\mathcal{A})$ for every $\mathcal{A}$ consistent with

the data, which is the never-undercounts property of Theorem 2. When the only consistent mechanisms are near-unitary, the supremum is large and the certificate correctly declines to certify.

## B.2 The closed form on the depolarizing family.

Specialize to the qubit case with SIC probes $\{\tau_x\}_{x=1}^4$, Bloch vectors $\{\boldsymbol{n}_x\}$ forming a tetrahedron ($\boldsymbol{n}_b \cdot \boldsymbol{n}_x = 1$ for $b = x$ and $-1/3$ otherwise), and the trusted SIC POVM $E_b = \tau_b/2$. For a depolarizing mechanism $\mathcal{A}(\rho) = (1-q)\rho + q\,\frac{\mathbb{I}}{2}$,

$$p(b\,|\,x) = \mathrm{Tr}[E_b\,\mathcal{A}(\tau_x)] = \tfrac{1}{2}\Big[(1-q)\,\tfrac{1+\boldsymbol{n}_b\cdot\boldsymbol{n}_x}{2} + \tfrac{q}{2}\Big],$$

so the diagonal terms are $p(x\,|\,x) = \frac{1}{2}(1-\frac{q}{2})$ and the agreement statistic of Equation (9) is

$$A = \sum_x p(x\,|\,x) = 4\cdot\tfrac{1}{2}\big(1-\tfrac{q}{2}\big) = 2-q. \tag{B.2}$$

The depolarizing Bloch map is the isotropic contraction $T = (1-q)\mathbb{I}$, $\boldsymbol{t} = 0$, so every pure input has output Bloch radius $s = 1-q$ and, by Equation (2), $\varepsilon(\mathcal{A}) = 2\ \mathrm{artanh}(s)$. Eliminating $q$ between $s = 1-q$ and Equation (B.2) gives $s = A-1$ and hence the closed form

$$\varepsilon_{\mathrm{cert}} = 2\ \mathrm{artanh}(A-1),$$

which is Equation (10). On this family it coincides with the worst-case privacy certificate Equation (B.1) and equals the true privacy parameter.

## B.3 Identifiability, and when the certificate reduces to an evaluation.

What the observed statistics determine depends on what is trusted about the measurement, and it is worth being precise. With trusted informationally complete inputs $\{\tau_x\}$ but a measurement left *untrusted*, the data $p(b \mid x) = \mathrm{Tr}[N_b\,\mathcal{A}(\tau_x)]$ fix, for each outcome, only the *effective* operator $F_b = \mathcal{A}^\dagger(N_b)$—the composition of the channel's adjoint with the unknown POVM—and never the channel and the measurement separately. This is exactly the degeneracy of Theorem 1: the pair $(\mathcal{A}, \{N_b\})$ is identifiable only up to the gauge $\mathcal{A}^\dagger(N_b) = \mathcal{A}'^\dagger(N'_b)$, so the Choi operator is not determined and the privacy functional is not certifiable. Trusting the *measurement*—taking the device to implement the characterized informationally complete POVM $\{E_b\}$, as in our model (Section 3.3)—is what removes the gauge: then $F_b = \mathcal{A}^\dagger(E_b)$ is known for a spanning set of effects, the linear map $\mathcal{A}^\dagger$ is determined on a complete operator basis, and $J$ is fixed uniquely. We confirmed this numerically: with the trusted IC measurement, maximizing and minimizing many random linear functionals $\langle C, J\rangle$ over the feasible set of Equation (B.1) gives a spread below $10^{-13}$ across all noise levels $q \in [0.1, 0.7]$, so the consistent set is a single point and the certificate reduces to the evaluation $\varepsilon(J_{\mathrm{data}})$. We are explicit that this identifiability is a consequence of the trusted measurement, and that in this regime the reconstruction of $J$ coincides with quantum process estimation; the content of the certificate is not the reconstruction but the operational

privacy guarantee read from it, and its soundness against an adversarial mechanism (Theorem 2), which holds for every channel consistent with the data. The worst-case program Equation (6) is the appropriate object precisely when the statistics do *not* pin $J$—when the trusted measurement is coarse-grained or only partially characterized, or under finite statistics, so that a nontrivial set of channels remains consistent—and the certified value is then the supremum of $\varepsilon(J)$ over that set; the pinned case is its tight limit.

## B.4 Convexity and the worked coarse-grained example.

The worst-case program is a convex *maximization*, and we verified the fact that makes it so: that $\varepsilon(J)$ is convex in $J$. Sampling random pairs of Choi operators $J_0, J_1$ on the depolarizing line and evaluating the QLDP value of the midpoint against the average, the Jensen inequality $\varepsilon\big(\frac{J_0+J_1}{2}\big) \le \frac{1}{2}[\varepsilon(J_0) + \varepsilon(J_1)]$ held in every trial, with the midpoint value strictly below the average—confirming convexity and ruling out concavity. Convex maximization attains its optimum at an extreme point of the feasible set and is NP-hard in general; the outer certificate is therefore not a single semidefinite program. In the cases used here it is evaluated by a low-dimensional parametrization or by a fixed-$J$ SDP/dual certificate. In more general settings, relaxations must be explicitly conservative from above; otherwise they would not constitute sound privacy certificates. The coarse-grained example of Section 5.2 is the latter case. A measurement recording only output populations fixes the depolarizing strength to an interval, say $p \in [0.30, 0.50]$, of positive width independent of the number of repetitions. Because $\varepsilon = 2\,\mathrm{artanh}(1-p)$ is decreasing in $p$, the worst case is at $p_{\mathrm{lo}} = 0.30$: the certificate is $\bar{\varepsilon} = \varepsilon(0.30) \approx 1.735$, whereas the interval-midpoint point estimate is $\varepsilon(0.40) \approx 1.386$. The certificate exceeds the point estimate by $\approx 0.35$, the difference being precisely the worst-case margin a reconstruction estimate would omit; and since no single $J$ is identified, no reconstruction estimate is even well defined. This is the sense in which the certificate is a worst-case guarantee and not process estimation followed by evaluation.

Second, the evaluation $\varepsilon(J_{\mathrm{data}})$ is itself an SDP with a rigorous dual. Fixing a neighboring pair, the measured hockey-stick divergence $E_\gamma^M\big(\mathcal{A}(\rho)\,\|\,\mathcal{A}(\sigma)\big) = \max_{0\preceq M\preceq \mathbb{I}} \mathrm{Tr}\big[M\big(\mathcal{A}(\rho) - \gamma\,\mathcal{A}(\sigma)\big)\big]$ is linear in the measurement operator $M$ and hence a semidefinite program [44]; the duality of such channel and correlation programs is standard [57, 58]. Its Lagrangian dual is a minimization whose every feasible point is, by weak duality, a rigorous *upper* bound on the divergence for the given $J$. The certified parameter is the smallest $\varepsilon$ with $E_{e^\varepsilon}^M \le 0$ on every neighboring pair, so a dual-feasible point at $\gamma = e^\varepsilon$ constitutes a machine-checkable certificate that the mechanism has privacy parameter no larger than $\varepsilon$—no appeal to solver convergence is needed. We verified that the SDP value matches the closed-form positive-eigenvalue sum for $d = 3$ to $10^{-4}$, so the dual certificate is available in the dimensions where the closed form Equation (10) does not apply.

## B.5 Scope.

The derivation used isotropy of the depolarizing Bloch map. For an anisotropic mechanism the single statistic $A$ no longer determines the extremal output radius, and

2 artanh$(A-1)$ can fall below the true $\varepsilon(\mathcal{A})$—it undercounts, as warned in Remark 1. The worst-case certificate Equation (B.1), which optimizes over the full feasible set rather than a single statistic, never does, and is the object used for any mechanism outside the depolarizing family.

# Appendix C Finite-statistics soundness via martingale concentration

This appendix carries out the finite-statistics argument behind Equation (12). We identify the quantity that must be controlled, state and apply the Azuma–Hoeffding bound for the memoryful adaptive adversary, give the variance-adapted refinement, record the explicit constants and sample complexity, and explain why entropy accumulation is not the appropriate tool for this certificate.

## C.1 The object of the soundness analysis

The certified parameter Equation (10) is the deterministic monotone map $\varepsilon_{\mathrm{cert}}(A) = 2\,\mathrm{artanh}(A-1)$ of the agreement statistic Equation (9). Operationally the verifier records, in round $i$, a binary *match indicator* $m_i \in \{0,1\}$ that equals one when the trusted informationally complete measurement returns the outcome matched to the round's probe input, and estimates the agreement by the empirical mean $\widehat{A} = 1 + \frac{1}{n}\sum_{i=1}^{n} m_i$. Because the certified privacy parameter is a function only of this bounded classical average, soundness requires nothing more than a one-sided confidence statement: that the agreement realized by the mechanism does not exceed $\widehat{A}$ by more than the sampling fluctuation, except with the allowed probability $\delta$. No secret is extracted and none must be hidden from the adversary, so there is no conditional min-entropy in the problem.

## C.2 Azuma–Hoeffding bound for a memoryful adaptive adversary

The adversary consists of the untrusted server together with any controller, ancilla, or memory carried by the mechanism across rounds; its action in round $i$ may depend arbitrarily on the full prior transcript. Let $\mathcal{F}_{i-1}$ be the $\sigma$-algebra generated by everything observable before round $i$, and define the centred increments

$$\xi_i \;=\; m_i - \mathbb{E}[m_i \mid \mathcal{F}_{i-1}]\,. \tag{C.1}$$

Two properties hold regardless of the device's memory or strategy: $\xi_i$ is a martingale difference, $\mathbb{E}[\xi_i \mid \mathcal{F}_{i-1}] = 0$, and it is bounded, $|\xi_i| \le 1$, because $m_i \in \{0,1\}$. More is true, and it is what fixes the constant: conditioned on $\mathcal{F}_{i-1}$ the increment $\xi_i$ lies in an interval of width exactly one, namely $\big[-\mathbb{E}[m_i \mid \mathcal{F}_{i-1}],\, 1-\mathbb{E}[m_i \mid \mathcal{F}_{i-1}]\big]$. It is this conditional *range* of unity, rather than the looser two-sided bound $|\xi_i| \le 1$, that supplies the factor 2 in the exponent of Equation (C.2): the bounded-difference (McDiarmid) form of the Azuma–Hoeffding inequality [59] gives, for a sum of martingale differences with conditional ranges $d_i$, $\Pr[\sum_i \xi_i \le -x] \le \exp(-2x^2/\sum_i d_i^2)$,

and with $d_i = 1$ and $x = nt$ this is exactly $e^{-2nt^2}$. The adaptive memory may change the conditional mean $\mathbb{E}[m_i \mid \mathcal{F}_{i-1}]$ from round to round, but it cannot violate the bounded-difference property. Writing the time-averaged agreement actually realized as $\bar{A}_n = 1 + \frac{1}{n}\sum_i \mathbb{E}[m_i \mid \mathcal{F}_{i-1}]$, the partial sums of $\xi_i$ form a bounded martingale, and the one-sided Azuma–Hoeffding inequality [59–61] gives

$$\Pr\Big[\, \bar{A}_n - \widehat{A} \;\geq\; t \,\Big] \;\leq\; e^{-2nt^2}, \tag{C.2}$$

for every $t > 0$. Setting the right-hand side to $\delta$ yields the half-width $t(\delta, n) = \sqrt{\ln(1/\delta)/(2n)}$ of Equation (11), and hence the finite-sample certificate Equation (12). No independence assumption enters: conditioning on $\mathcal{F}_{i-1}$ absorbs the adversary's memory, which is precisely why the martingale bound is sound against the fully adaptive, memoryful device. The certified quantity is the realized time-averaged agreement $\bar{A}_n$; for a fixed untrusted channel probed repeatedly, the per-round conditional means coincide with the channel's agreement, so $\bar{A}_n$ certifies that channel's privacy parameter directly.

## C.3 Variance-adapted refinement

Azuma–Hoeffding uses only the worst-case range $|\xi_i| \leq 1$. When the honest device operates at high agreement the conditional variance $V_i = \mathbb{E}[m_i \mid \mathcal{F}_{i-1}]\big(1 - \mathbb{E}[m_i \mid \mathcal{F}_{i-1}]\big)$ is small, and Freedman's martingale inequality [62]—the martingale form of Bernstein's inequality—sharpens Equation (C.2) to

$$\Pr\Big[\, \textstyle\sum_i \xi_i \geq x,\; \sum_i V_i \leq v \,\Big] \;\leq\; \exp\Big( - \frac{x^2}{2(v + x/3)} \Big), \tag{C.3}$$

giving a half-width of order $\sqrt{2\bar{V}\ln(1/\delta)/n} + \ln(1/\delta)/(3n)$ with $\bar{V}$ the average conditional variance. In the high-privacy regime where $A \to 2$ this is materially tighter than Hoeffding; an empirical-Bernstein variant lets $\bar{V}$ be estimated from the data. We use the Hoeffding form for the transparent statement in the main text and note that the variance-adapted bound is available when the tightest constants are wanted.

## C.4 Constants, coverage, and sample complexity

The finite-size coefficient of the Hoeffding half-width, $t(\delta, n) = c/\sqrt{n}$, is $c = \sqrt{\ln(1/\delta)/2}$, equal to 1.86 at $\delta = 10^{-3}$ and 2.63 at $\delta = 10^{-6}$. Certifying the privacy parameter to within a few percent of its true value therefore takes of order $10^4$ probe rounds: for a depolarizing mechanism with true agreement $A = 1.7$, reaching a certified value within 5% of the truth requires a half-width $t \lesssim 0.022$, hence $n \gtrsim 1.5 \times 10^4$ at $\delta = 10^{-6}$. We confirmed numerically, simulating an adaptive memoryful device whose per-round match probability is driven by the running deficit of past outcomes, that the empirical coverage of the event $\varepsilon^{(n)}_{\mathrm{cert}} \geq \varepsilon(\mathcal{A})$ meets the nominal $1 - \delta$ across the tested range of $n$ and $\delta$, with the bound conservative as expected of a worst-case martingale tail.

### C.5 Why not entropy accumulation

The entropy accumulation theorem [27] and its generalized form [28] lower-bound the smooth conditional *min-entropy* of an $n$-round register against an adversary's quantum side information, and are the right tool when a secret key or random string is extracted and must be made unpredictable to a quantum adversary. The present certificate extracts no secret: it reports a deterministic function of a bounded classical statistic, and its soundness is one-sided concentration of that statistic under adaptive sampling. Invoking entropy accumulation here would bound a quantity the protocol does not use and would import entropic second-order and smoothing penalties—a finite-size coefficient roughly an order of magnitude larger, and a sample complexity roughly two orders of magnitude larger—that the martingale analysis correctly avoids. The same choice is made in device-independent randomness certification, where the estimated test statistic is controlled by an Azuma–Hoeffding argument against a memoryful device [32], and in composable device-independent *certification* of a property from an estimated statistic, where soundness rests on a Hoeffding bound rather than entropy accumulation [4]. We reserve entropy-accumulation language for the genuinely different and open problem of accumulating the privacy divergence directly across rounds, which we do not attempt here.

## Appendix D The Holevo instrument and advantage survival

This appendix details the optimal mechanisms behind the advantage of Section 6, the direct computation of the utility ratio Equation (13), and the survival thresholds.

### D.1 The EITFF mechanisms.

For data of alphabet size $n$ the optimal high-privacy QLDP mechanism [8] encodes input $x$ into the state

$$\sigma_x = \frac{\mu}{d}\,\mathbb{I}_d + \frac{1-\mu}{r}\,P_x, \tag{D.1}$$

where $\{P_x\}_{x=1}^n$ is an equi-isoclinic tight fusion frame (EITFF) of rank-$r$ projectors in dimension $d$, and the mixing parameter is fixed by the target privacy through $1-\mu = \big(1+(1-c)/\sinh^2(\varepsilon/2)\big)^{-1/2}$, with $c$ the isoclinic angle. The frame condition $\sum_x P_x = (nr/d)\mathbb{I}_d$ makes the encoding maximally symmetric. For $n = 3, 4$ the optimal frame is rank one ($r = 1$): an equiangular tight frame of $n$ lines, namely the real simplex in $d = 2$ for $n = 3$ and the qubit SIC tetrahedron for $n = 4$, recovering the mechanisms of Nam *et al.* [7]. For $n \geq 5$ the optimum uses genuine rank-$r \geq 2$ isoclinic subspaces, which we construct numerically by a multiple-restart projected-gradient search enforcing tightness and equi-isocliny to a residual below $10^{-6}$. Each constructed $\{\sigma_x\}$ is verified to be $\varepsilon$-QLDP by checking $e^{\varepsilon}\sigma_x - \sigma_y \succeq 0$ for all pairs.

### D.2 Direct computation of the utility ratio.

The utility is the Holevo information $\chi = S\big(\frac{1}{n}\sum_x \sigma_x\big) - \frac{1}{n}\sum_x S(\sigma_x)$ of the encoded ensemble, with $S$ the von Neumann entropy; the classical baseline is the optimal classical locally private mechanism. We compute $\chi$ *directly* from the states Equation (D.1)—not by substitution into a formula—and extract the small-$\varepsilon$ quadratic coefficient. The quantum-to-classical ratio so obtained reproduces

$$\frac{Q}{C} = \frac{n(n-1)}{2\lfloor n/2\rfloor\lceil n/2\rceil} = 2 - \frac{1}{\lceil n/2\rceil}$$

of Equation (13), giving $1, \frac{3}{2}, \frac{3}{2}, \frac{5}{3}$ for $n = 2, 3, 4, 5$ to a residual of order $10^{-4}$, the residual being the finite-$\varepsilon$ fit and vanishing as $\varepsilon \to 0$. The ratio is the same for any admissible utility sharing the leading quadratic structure, so the choice of Holevo information is without loss of generality.

### D.3 Survival under noise.

To model certification and device overhead we apply an additional depolarizing layer of strength $p$ to the encoded ensemble, $\sigma_x \mapsto (1-p)\sigma_x + p\,\frac{\mathbb{I}}{d}$, while holding the classical baseline at its optimum—the least favorable accounting for the quantum side—and recompute $Q/C$ directly from the Holevo information, locating the threshold $p^\star$ at which the computed ratio crosses unity. This survival analysis is our own; it is not inherited from the optimal-mechanism literature, which fixes the noiseless ratio Equation (13) but does not address robustness to a depolarizing certification overhead. The directly computed thresholds are $p^\star \approx 0.18$ for $n = 3, 4$, $p^\star \approx 0.23$ for $n = 5$, and $p^\star \approx 0.25$ for $n = 7$, obtained on genuine equi-isoclinic tight fusion frames (constructed to a tightness residual below $10^{-6}$) and stable across the high-privacy range $\varepsilon \in [0.1, 0.8]$; thus the genuine advantage $Q/C > 1$ persists up to roughly 18–25% depolarizing noise, rising with the alphabet size. The strict ratio $Q/C \geq 3/2$, being the noiseless value for $n = 3, 4$, is eroded by any $p$ there and acquires a noise margin only at $n \geq 5$. These are the thresholds quoted in Section 6.2 and plotted in Figure 6.

## Appendix E Composable-security proof

This appendix proves that the certification protocol $\pi$ constructs the ideal certified-private-aggregation resource $S_{\mathrm{CPA}}$ within the error Equation (15), in the sense of abstract cryptography [51, 52]. We give the ideal resource and the simulator explicitly, prove completeness and soundness as separate lemmas, and assemble the composable error term by term over the abort and accept branches. The argument follows the structure of composable device-independent state certification [4] but is carried out in full for the certified-privacy resource: the soundness step uses the Fuchs–van de Graaf inequality [63] together with the martingale tail of Section C, and carries no privacy-amplification term.

## E.1 Resources, protocol, and simulator

We work with classical–quantum (cq) states carrying an explicit flag register that records acceptance ($|\mathrm{acc}\rangle$) or abort ($|\perp\rangle$), following the sub-normalized-branch convention standard in composable finite-size security [51].

### E.1.1 Ideal resource.

$S_{\mathrm{CPA}}$ has a client interface and an adversarial interface. On input it either raises the abort flag $\perp$ on the client interface, or outputs the pair $(\mathrm{aggregate}, \varepsilon_{\mathrm{cert}})$ together with an accepted branch that satisfies the reported $\varepsilon_{\mathrm{cert}}$-QLDP specification by construction. Its output cq-state is

$$\rho^{\mathrm{ideal}} = (1 - p_{\mathrm{acc}})\ |\perp\rangle\langle\perp| + p_{\mathrm{acc}}\ |\mathrm{acc}\rangle\langle\mathrm{acc}| \otimes \rho^{\star}_{\varepsilon_{\mathrm{cert}}}, \tag{E.1}$$

where $\rho^{\star}_{\varepsilon_{\mathrm{cert}}}$ is an ideal reference state inside the $\varepsilon_{\mathrm{cert}}$-QLDP specification and $p_{\mathrm{acc}}$ is the acceptance probability fixed by the simulator below. The ideal branch represents the certified privacy specification rather than a claim that the real accepted mechanism is identical to a unique reconstructed Choi state.

### E.1.2 Real protocol.

The protocol converter $\pi$ applied to the real (untrusted) resource produces

$$\rho^{\mathrm{real}} = (1 - p_{\mathrm{acc}})\ |\perp\rangle\langle\perp| + p_{\mathrm{acc}}\ |\mathrm{acc}\rangle\langle\mathrm{acc}| \otimes \rho^{\mathrm{real}}_{|\mathrm{acc}}, \tag{E.2}$$

with $\rho^{\mathrm{real}}_{|\mathrm{acc}}$ the Choi state of the accept-conditioned mechanism actually realized by the device. The two states share the abort term $(1 - p_{\mathrm{acc}})\,|\perp\rangle\langle\perp|$ verbatim; this is what makes the abort branch cancel in the trace-distance bound below.

### E.1.3 Simulator.

The simulator is the explicit converter $\sigma$ that attaches to the adversarial interface of $S_{\mathrm{CPA}}$ and acts as follows.

1. It runs the device's own black-box strategy on the trusted probe rounds that the real protocol would have used, obtaining the same per-round transcript the real adversary produces.
2. From that transcript it computes the empirical agreement statistic $\widehat{A}$ and the accept/abort decision exactly as $\pi$ does, and forwards the resulting flag to $S_{\mathrm{CPA}}$, so that $S_{\mathrm{CPA}}$ aborts iff the real protocol aborts. This fixes $p_{\mathrm{acc}}$ in Equation (E.1) to equal the real acceptance probability identically.
3. On acceptance it outputs the public pair $(\widehat{A}, \varepsilon_{\mathrm{cert}})$ on the adversarial interface, with $\varepsilon_{\mathrm{cert}} = 2\,\mathrm{artanh}(\widehat{A} + t - 1)$ of Equation (12).

The construction is well defined because every quantity $\sigma$ needs is public: the certified value is a function of the agreement statistic and the trusted probe identities alone, and the verifier's probe labels are drawn from private trusted randomness and revealed to

the adversary only through the published agreement. Hence $\sigma$ uses *nothing* from the clients' private inputs and has exactly the information available to the real adversary—the requirement that makes it an admissible simulator at the adversarial interface.

## E.2 Completeness

**Lemma 1** (Completeness) *There is an honest implementation of the device—the prescribed $\varepsilon_{\mathrm{cert}}$-QLDP mechanism, realized for instance by the depolarizing mechanism of the design strength—for which $\pi$ aborts with probability at most $\varepsilon_{\mathrm{compl}} = 2\,e^{-2n\kappa^2}$, which vanishes as the number of rounds $n \to \infty$.*

*Proof* Under the honest mechanism the per-round agreement indicators $m_i \in \{0,1\}$ have conditional mean exceeding the acceptance threshold by a margin $\kappa > 0$, and acceptance is the event that the empirical agreement $\widehat{A}$ stays within $\kappa$ of its mean. The indicators are bounded, so the two-sided Azuma–Hoeffding bound of Section C gives $\Pr[\,|\widehat{A} - \mathbb{E}\widehat{A}| \geq \kappa\,] \leq 2\,e^{-2n\kappa^2}$, whence the honest run accepts except with probability $\varepsilon_{\mathrm{compl}} = 2\,e^{-2n\kappa^2}$. Choosing $\kappa$ below the certificate's tolerance keeps the honest certified value within specification. We confirmed numerically that $\varepsilon_{\mathrm{compl}}$ falls below $10^{-3}$ well within the sample sizes required for soundness, so completeness is not the binding constraint. □

## E.3 Soundness

**Lemma 2** (Soundness) *For an arbitrary adversarial mechanism, the trace distance between the real output Equation (E.2) and the simulated ideal output $\sigma S_{\mathrm{CPA}}$ of Equation (E.1) satisfies*

$$\tfrac{1}{2}\big\|\rho^{\mathrm{real}} - \sigma\, S_{\mathrm{CPA}}\big\|_1 \;\leq\; \delta_{\mathrm{sound}} + \varepsilon_{\mathrm{est}}, \qquad \delta_{\mathrm{sound}} = e^{-2nt^2}, \tag{E.3}$$

*with no privacy-amplification term.*

*Proof* Because the simulator fixes the abort flag and acceptance probability to those of the real protocol, the abort terms of Equations (E.1) and (E.2) coincide and cancel under subtraction, and the joint trace distance reduces to the accept branch weighted by its probability,

$$\tfrac{1}{2}\big\|\rho^{\mathrm{real}} - \sigma\, S_{\mathrm{CPA}}\big\|_1 = p_{\mathrm{acc}} \cdot \tfrac{1}{2}\big\|\rho^{\mathrm{real}}_{|\mathrm{acc}} - \rho^{\star}_{\varepsilon_{\mathrm{cert}}}\big\|_1. \tag{E.4}$$

We emphasize that the left-hand side is the *joint* quantity Pr[accept $\wedge$ certificate wrong]-weighted distance, not the accept-conditioned distance; this is the quantity composable security requires, and it is bounded by the joint event probability rather than a conditional one. Let $\Omega$ be the event that the finite-statistics bound Equation (12) holds, i.e. that the true agreement does not exceed $\widehat{A}+t$. By the one-sided Azuma–Hoeffding bound of Equation (11), $\Pr[\neg\Omega] \leq e^{-2nt^2} = \delta_{\mathrm{sound}}$, and on $\Omega$, Theorem 2 guarantees that the accept-conditioned mechanism satisfies the reported $\varepsilon_{\mathrm{cert}}$-QLDP specification, so its Choi state lies in the QLDP specification ball $\mathcal{B}_{\varepsilon_{\mathrm{cert}}} = \{\rho : \rho \text{ is the Choi state of an } \varepsilon_{\mathrm{cert}}\text{-QLDP mechanism}\}$. We take the ideal reference $\rho^{\star}_{\varepsilon_{\mathrm{cert}}}$ to be the element of $\mathcal{B}_{\varepsilon_{\mathrm{cert}}}$ closest to the accepted state in trace distance—the projection of the accepted Choi state onto the specification ball—so that the metric bridge is explicit: the distance from the accepted state to the ideal target is exactly the distance from the accepted state to the specification it is certified to meet. By the Fuchs–van de Graaf inequality [63] this radius is controlled by the fidelity to that target,

$\frac{1}{2}\|\rho^{\text{real}}_{|\text{acc}} - \rho^{\star}_{\varepsilon_{\text{cert}}}\|_1 \le \sqrt{1-F} \le \varepsilon_{\text{est}}$, where $\varepsilon_{\text{est}}$ is the residual error in estimating the certified parameter from the accepted agreement and equals zero when the certificate is the exact closed-form map of the target family (Section 5.5). Splitting Equation (E.4) over $\Omega$ and its complement, bounding $p_{\text{acc}} \le 1$, and using $\frac{1}{2}\|\cdot\|_1 \le 1$ on $\neg\Omega$, gives Equation (E.3). No smoothing parameter enters: the protocol extracts no key and performs no privacy amplification, so the only contributions are the martingale failure probability $\delta_{\text{sound}}$ and the estimation radius $\varepsilon_{\text{est}}$. □

## E.4 Assembling the composable error

**Proposition 4** (Construction within $\varepsilon_{\text{comp}}$) *The protocol $\pi$ constructs $S_{\text{CPA}}$ within $\varepsilon_{\text{comp}} = \varepsilon_{\text{compl}} + \delta_{\text{sound}} + \varepsilon_{\text{est}}$.*

*Proof* By Helstrom's theorem Equation (14), the optimal distinguishing advantage between the real system $\pi R$ and the simulated ideal system $\sigma S_{\text{CPA}}$, over all distinguishers granted the protocol output and the adversary's side information but not the clients' private data, equals the trace distance $\frac{1}{2}\|\pi R - \sigma S_{\text{CPA}}\|_1$. The availability (completeness) condition contributes the honest-abort term $\varepsilon_{\text{compl}}$ of Lemma 1; the soundness condition contributes the distance Equation (E.3) of Lemma 2. Adding the two,

$$d(\pi R, \sigma S_{\text{CPA}}) \;\le\; \varepsilon_{\text{compl}} + \delta_{\text{sound}} + \varepsilon_{\text{est}} \;=\; \varepsilon_{\text{comp}},$$

which is Equation (15). We verified the combined inequality numerically on the certified EITFF ensemble for $n = 3, 4, 5$: with the finite-statistics perturbation set at the allocated budget, the joint trace distance to the ideal resource (after shared-abort cancellation, equal to $p_{\text{acc}}$ times the accept-branch distance) stays within $\varepsilon_{\text{comp}}$ and scales linearly with the perturbation, as the proof predicts. □

## E.5 Composition and the two budgets

The construction composes sequentially. Writing $P_r$ for the channel that the $r$-th invocation applies and $\sigma_r S^{(r)}_{\text{CPA}}$ for its ideal counterpart, the triangle inequality together with the contractivity of the trace distance under the completely positive trace-preserving inter-round maps [64, Exercise 9.1.9] gives, for $R$ rounds,

$$\tfrac{1}{2}\Big\|\bigcirc_r P_r(\rho_{\text{init}}) - \bigotimes_r \sigma_r S^{(r)}_{\text{CPA}}\Big\|_1 \;\le\; \sum_{r=1}^{R} \varepsilon^{(r)}_{\text{comp}} \;=:\; \varepsilon^{\text{tot}}_{\text{comp}}, \tag{E.5}$$

the standard additive accumulation of construction errors in abstract cryptography [51, 52]. Independently, the QLDP privacy budget of the delivered mechanisms composes additively to $\varepsilon^{\text{tot}}_{\text{cert}} = \sum_r \varepsilon^{(r)}_{\text{cert}}$ by the composition theorem for quantum local differential privacy [6], whose additivity holds irrespective of whether the per-round inputs are independent, classically correlated, or entangled.

The two budgets accumulate on independent axes and are reported as the pair $\big(\varepsilon^{\text{tot}}_{\text{cert}}; \varepsilon^{\text{tot}}_{\text{comp}}\big)$, never summed: the first is a privacy level carried in the resource specification, the second a probability that the specification is not met—distinct quantities with distinct units, exactly as a key-distribution protocol reports a key length and a

security error separately. Under this additive accounting the privacy budget grows linearly in $R$, so the number of composable rounds at a fixed target privacy is bounded; a tighter advanced-composition analysis is left to future work. We note that Guan's additivity is established for mechanisms composed across distributed systems; for federated rounds whose joint implementation correlates the per-round channels in ways outside that structure, the additive QLDP accounting should be read as the guarantee for the tensor-product (independently implemented) regime, which is the setting of our protocol.

# References


[1] Gilbert, A.C., McMillan, A.: Property testing for differential privacy. In: 56th Annual Allerton Conference on Communication, Control, and Computing, pp. 249–258 (2018). arXiv:1806.06427

[2] Biswas, A., Cormode, G.: Verifiable differential privacy. arXiv preprint (2022) arXiv:2208.09011 [cs.CR]

[3] Bell, Z.R., Goldwasser, S., Kim, M.P., Watson, J.-L.: Certifying private probabilistic mechanisms. In: Advances in Cryptology – CRYPTO 2024. Lecture Notes in Computer Science, vol. 14925, pp. 348–386. Springer, Berlin, Heidelberg (2024). https://doi.org/10.1007/978-3-031-68391-6_11 . Cryptology ePrint Archive, Paper 2024/938

[4] Bhavsar, R., Wooltorton, L., Bae, J.: Composable framework for device-independent state certification. PRX Quantum **6**, 040356 (2025) https://doi.org/10.1103/1j5f-qjhx . arXiv:2505.06833

[5] Hirche, C., Rouzé, C., França, D.S.: Quantum differential privacy: An information theory perspective. arXiv preprint (2023) arXiv:2202.10717 [quant-ph]

[6] Guan, J.: Optimal mechanisms for quantum local differential privacy. arXiv preprint (2024) arXiv:2407.13516 [quant-ph]. ACM CCS 2025

[7] Nam, S.-H., Park, H.-Y., Lee, S.-H., Bae, J.: Quantum advantage in locally differentially private hypothesis testing. arXiv preprint (2025) arXiv:2501.10152 [quant-ph]. IEEE ISIT 2025

[8] Yoshida, Y.: Optimal quantum locally differentially private mechanisms in the high-privacy regime. arXiv preprint (2026) arXiv:2605.27278 [quant-ph]

[9] Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: Theory of Cryptography (TCC 2006). Lecture Notes in Computer Science, vol. 3876, pp. 265–284. Springer, Berlin, Heidelberg (2006). https://doi.org/10.1007/11681878_14

[10] Dwork, C., Roth, A.: The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science **9**(3–4), 211–407 (2014) https://doi.org/10.1561/0400000042

[11] Warner, S.L.: Randomized response: A survey technique for eliminating evasive answer bias. Journal of the American Statistical Association **60**(309), 63–69 (1965) https://doi.org/10.1080/01621459.1965.10480775

[12] Kasiviswanathan, S.P., Lee, H.K., Nissim, K., Raskhodnikova, S., Smith, A.: What can we learn privately? SIAM Journal on Computing **40**(3), 793–826 (2011) https://doi.org/10.1137/090756090

[13] Duchi, J.C., Jordan, M.I., Wainwright, M.J.: Local privacy and statistical minimax rates. In: 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pp. 429–438 (2013). https://doi.org/10.1109/FOCS.2013.53

[14] Aaronson, S., Rothblum, G.N.: Gentle measurement of quantum states and differential privacy. In: 51st Annual ACM SIGACT Symposium on Theory of Computing (STOC), pp. 322–333 (2019). https://doi.org/10.1145/3313276.3316378 . arXiv:1904.08747

[15] Zhou, L., Ying, M.: Differential privacy in quantum computation. In: 2017 IEEE 30th Computer Security Foundations Symposium (CSF), pp. 249–262 (2017). https://doi.org/10.1109/CSF.2017.23

[16] Angrisani, A., Kashefi, E.: Quantum differential privacy in the local model. IEEE Transactions on Information Theory **71**(5), 3675–3692 (2025) https://doi.org/10.1109/TIT.2025.3552671 . Preprint arXiv:2203.03591

[17] Angrisani, A., Doosti, M., Kashefi, E.: A unifying framework for differentially private quantum algorithms. arXiv preprint (2023). arXiv:2307.04733

[18] Du, Y., Hsieh, M.-H., Liu, T., You, S., Tao, D.: Quantum differentially private sparse regression learning. IEEE Transactions on Information Theory **68**(8), 5217–5233 (2022) https://doi.org/10.1109/TIT.2022.3164726

[19] Šupić, I., Bowles, J.: Self-testing of quantum systems: a review. Quantum **4**, 337 (2020) https://doi.org/10.22331/q-2020-09-30-337

[20] Brunner, N., Cavalcanti, D., Pironio, S., Scarani, V., Wehner, S.: Bell nonlocality. Reviews of Modern Physics **86**, 419–478 (2014) https://doi.org/10.1103/RevModPhys.86.419

[21] Mayers, D., Yao, A.: Self testing quantum apparatus. Quantum Information and Computation **4**(4), 273–286 (2004). arXiv:quant-ph/0307205

[22] Kaniewski, J.: Analytic and nearly optimal self-testing bounds for the clauser-horne-shimony-holt and mermin inequalities. Physical Review Letters **117**, 070402 (2016) https://doi.org/10.1103/PhysRevLett.117.070402

[23] Reichardt, B.W., Unger, F., Vazirani, U.: Classical command of quantum systems. Nature **496**, 456–460 (2013) https://doi.org/10.1038/nature12035

[24] Bancal, J.-D., Sangouard, N., Sekatski, P.: Noise-resistant device-independent certification of bell state measurements. Physical Review Letters **121**, 250506 (2018) https://doi.org/10.1103/PhysRevLett.121.250506

[25] Lo, H.-K., Curty, M., Qi, B.: Measurement-device-independent quantum key distribution. Physical Review Letters **108**, 130503 (2012) https://doi.org/10.1103/PhysRevLett.108.130503

[26] Branciard, C., Rosset, D., Liang, Y.-C., Gisin, N.: Measurement-device-independent entanglement witnesses for all entangled quantum states. Physical Review Letters **110**, 060405 (2013) https://doi.org/10.1103/PhysRevLett.110.060405

[27] Dupuis, F., Fawzi, O., Renner, R.: Entropy accumulation. Communications in Mathematical Physics **379**, 1–47 (2020) https://doi.org/10.1007/s00220-020-03839-5

[28] Metger, T., Fawzi, O., Sutter, D., Renner, R.: Generalised entropy accumulation. Communications in Mathematical Physics **405**(11), 261 (2024) https://doi.org/10.1007/s00220-024-05121-4 . Preprint arXiv:2203.04989; conference version FOCS 2022

[29] Arnon-Friedman, R., Dupuis, F., Fawzi, O., Renner, R., Vidick, T.: Practical device-independent quantum cryptography via entropy accumulation. Nature Communications **9**, 459 (2018) https://doi.org/10.1038/s41467-017-02307-4

[30] Metger, T., Renner, R.: Security of quantum key distribution from generalised entropy accumulation. Nature Communications **14**, 5272 (2023) https://doi.org/10.1038/s41467-023-40920-8

[31] Tomamichel, M., Leverrier, A.: A largely self-contained and complete security proof for quantum key distribution. Quantum **1**, 14 (2017) https://doi.org/10.22331/q-2017-07-14-14

[32] Pironio, S., Acín, A., Massar, S., Giroday, A., Matsukevich, D.N., Maunz, P., Olmschenk, S., Hayes, D., Luo, L., Manning, T.A., Monroe, C.: Random numbers certified by Bell's theorem. Nature **464**(7291), 1021–1024 (2010) https://doi.org/10.1038/nature09008

[33] Hayashi, M., Koshiba, T.: Quantum verifiable protocol for secure modulo zero-sum randomness. Quantum Information Processing **21**, 291 (2022) https://doi.org/10.1007/s11128-022-03639-x . arXiv:1910.05976

[34] Canetti, R.: Universally composable security: A new paradigm for cryptographic protocols. In: 42nd Annual IEEE Symposium on Foundations of Computer Science (FOCS), pp. 136–145 (2001). https://doi.org/10.1109/SFCS.2001.959888

[35] Unruh, D.: Universally composable quantum multi-party computation. In: Advances in Cryptology – EUROCRYPT 2010. Lecture Notes in Computer Science, vol. 6110, pp. 486–505. Springer, Berlin, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13190-5_25 . arXiv:0910.2912

[36] Ben-Or, M., Mayers, D.: General security definition and composability for quantum and classical protocols. arXiv preprint (2004). arXiv:quant-ph/0409062

[37] Ambainis, A., Jakobsson, M., Lipmaa, H.: Cryptographic randomized response techniques. In: Public Key Cryptography (PKC 2004). Lecture Notes in Computer Science, vol. 2947, pp. 425–438. Springer, Berlin, Heidelberg (2004). https://doi.org/10.1007/978-3-540-24632-9_31

[38] Narayan, A., Feldman, A., Papadimitriou, A., Haeberlen, A.: Verifiable differential privacy. In: Proceedings of the Tenth European Conference on Computer Systems (EuroSys) (2015). https://doi.org/10.1145/2741948.2741978

[39] Munilla Garrido, G., Babel, M., Sedlmeir, J.: Towards verifiable differentially-private polling. In: Proceedings of the 17th International Conference on Availability, Reliability and Security (ARES) (2022). https://doi.org/10.1145/3538969.3538992 . arXiv:2206.07220

[40] Shamsabadi, A.S., Tan, G., Cebere, T.I., Bellet, A., Haddadi, H., Papernot, N., Wang, X., Weller, A.: Confidential-DPproof: Confidential proof of differentially private training. In: International Conference on Learning Representations (ICLR) (2024)

[41] Bontekoe, T., Asghar, H.J., Turkmen, F.: Efficient verifiable differential privacy with input authenticity in the local and shuffle model. Proceedings on Privacy Enhancing Technologies (PoPETs) **2025**(2), 543–563 (2025) https://doi.org/10.56553/popets-2025-0076

[42] Shi, K., Ma, T.: Auditing differential privacy in the black-box setting. arXiv preprint (2025). arXiv:2503.12045

[43] Eisert, J., Hangleiter, D., Walk, N., Roth, I., Markham, D., Parekh, R., Chabaud, U., Kashefi, E.: Quantum certification and benchmarking. Nature Reviews Physics **2**(7), 382–390 (2020) https://doi.org/10.1038/s42254-020-0186-4

[44] Nuradha, T., Singh, V., Wilde, M.M.: Measured hockey-stick divergence and its applications to quantum pufferfish privacy. arXiv preprint (2025) arXiv:2501.12359 [quant-ph]. IEEE ISIT 2025

[45] Renes, J.M., Blume-Kohout, R., Scott, A.J., Caves, C.M.: Symmetric informationally complete quantum measurements. Journal of Mathematical Physics **45**(6), 2171–2180 (2004) https://doi.org/10.1063/1.1737053

[46] Scott, A.J., Grassl, M.: Symmetric informationally complete positive-operator-valued measures: A new computer study. Journal of Mathematical Physics **51**(4), 042203 (2010) https://doi.org/10.1063/1.3374022

[47] Holevo, A.S.: Bounds for the quantity of information transmitted by a quantum communication channel. Problems of Information Transmission **9**(3), 177–183 (1973)

[48] Audenaert, K.M.R., Calsamiglia, J., Muñoz-Tapia, R., Bagan, E., Masanes, L., Acin, A., Verstraete, F.: Discriminating states: The quantum chernoff bound. Physical Review Letters **98**, 160501 (2007) https://doi.org/10.1103/PhysRevLett.98.160501

[49] Hiai, F., Petz, D.: The proper formula for relative entropy and its asymptotics in quantum probability. Communications in Mathematical Physics **143**(1), 99–114 (1991) https://doi.org/10.1007/BF02100287

[50] Ogawa, T., Nagaoka, H.: Strong converse and stein's lemma in quantum hypothesis testing. IEEE Transactions on Information Theory **46**(7), 2428–2433 (2000) https://doi.org/10.1109/18.887855

[51] Portmann, C., Renner, R.: Security in quantum cryptography. Rev. Mod. Phys. **94**, 025008 (2022) https://doi.org/10.1103/RevModPhys.94.025008

[52] Maurer, U., Renner, R.: Abstract cryptography. In: Innovations in Computer Science (ICS), pp. 1–21. Tsinghua University Press, Beijing (2011)

[53] Solomons, N.R., Markham, D.: Composable privacy of networked quantum sensing. arXiv preprint (2026) arXiv:2510.06326 [quant-ph]

[54] Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H.B., Patel, S., Ramage, D., Segal, A., Seth, K.: Practical secure aggregation for privacy-preserving machine learning. In: 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS), pp. 1175–1191 (2017). https://doi.org/10.1145/3133956.3133982

[55] Broadbent, A., Fitzsimons, J., Kashefi, E.: Universal blind quantum computation. In: 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pp. 517–526 (2009). https://doi.org/10.1109/FOCS.2009.36 . arXiv:0807.4154

[56] Chehimi, M., Saad, W.: Quantum federated learning with quantum data. In: 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8617–8621 (2022). https://doi.org/10.1109/ICASSP43922.2022.9746622 . arXiv:2106.00005

[57] Watrous, J.: Semidefinite programs for completely bounded norms. Theory of Computing **5**, 217–238 (2009) https://doi.org/10.4086/toc.2009.v005a011 . arXiv:0901.4709

[58] Navascués, M., Pironio, S., Acin, A.: A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. New Journal of Physics **10**(7), 073013 (2008) https://doi.org/10.1088/1367-2630/10/7/073013

[59] McDiarmid, C.: On the method of bounded differences. In: Surveys in Combinatorics 1989. London Mathematical Society Lecture Note Series, vol. 141, pp. 148–188. Cambridge University Press, Cambridge (1989). https://doi.org/10.1017/CBO9781107359949.008

[60] Hoeffding, W.: Probability inequalities for sums of bounded random variables. Journal of the American Statistical Association **58**(301), 13–30 (1963) https://doi.org/10.1080/01621459.1963.10500830

[61] Azuma, K.: Weighted sums of certain dependent random variables. Tohoku Mathematical Journal **19**(3), 357–367 (1967) https://doi.org/10.2748/tmj/1178243286

[62] Freedman, D.A.: On tail probabilities for martingales. The Annals of Probability **3**(1), 100–118 (1975) https://doi.org/10.1214/aop/1176996452

[63] Fuchs, C.A., Graaf, J.: Cryptographic distinguishability measures for quantum-mechanical states. IEEE Transactions on Information Theory **45**(4), 1216–1227 (1999) https://doi.org/10.1109/18.761271

[64] Wilde, M.M.: Quantum Information Theory. Cambridge University Press, Cambridge (2013). https://doi.org/10.1017/CBO9781139525343