STOC03: On the Power of Quantum Fingerprinting (Andrew Yao) , Reading Note

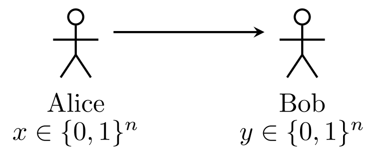
Siyu Liu

January 2026

1 Communication Complexity (Andrew Yao, 1992)

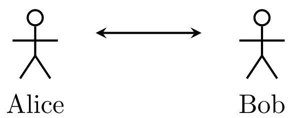
1.1 Classical communication complexity (Andrew Yao, 1978)

1-way model :

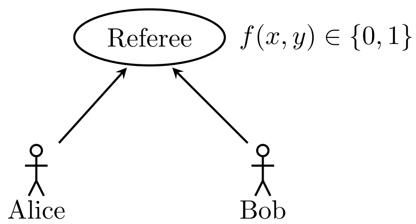


Want to compute Boolean function $f(x, y)$

2-way model :



Referee model: (simultaneous message model)



1.2 Equality function EQ

$f(x, y) = \mathbb{1}[x = y]$

1.2.1 1-way model

Deterministic: $D^{\rightarrow}(EQ) = \Theta(n)$

Randomized: $D^{\rightarrow}(EQ) = \Theta(\log n)$

- Protocol:
  1. Alice chooses a prime p randomly such that $n^2 \leq p < 2n^2$.
  2. Alice sends the pair $(p, x \bmod p)$ to Bob.
  3. Bob checks if $x \bmod p \equiv y \bmod p$.

- Error Rate Analysis:

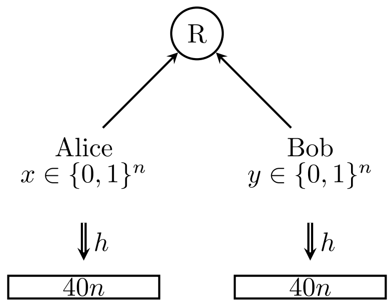
$$
\begin{aligned}
\text{error rate} &= \Pr_p[\, x \equiv y \pmod{p} \mid x \neq y \,] \\
&= \Pr_p[\, p \mid |x-y| \mid x \neq y \,] \\
&\leq \frac{\text{number of prime factors of } |x-y|}{\text{number of primes in } [n^2, 2n^2]} \\
&= O\left(\frac{n}{\pi(2n^2) - \pi(n^2)}\right) = O\left(\frac{n}{n^2/\log n}\right) \\
&= O\left(\frac{\log n}{n}\right)
\end{aligned}
$$

Note: $|x-y|$ *has at most* $\log|x-y| \leq n$ *distinct prime factors.*

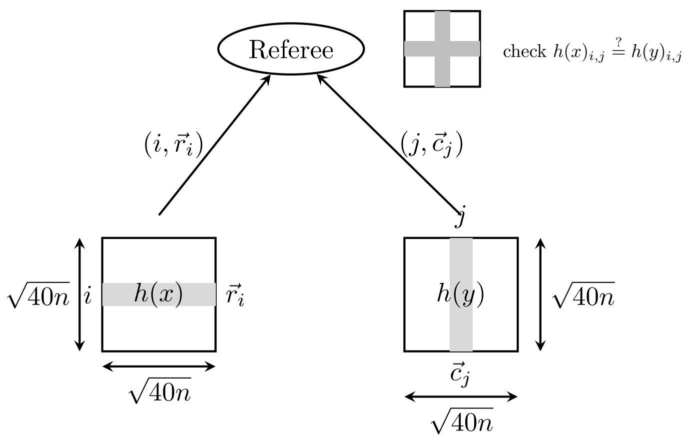
1.2.2 Referee Model

Deterministic: $D^{||}(EQ) = \Theta(n)$

Randomized: $R^{||}(EQ) = \Theta(\sqrt{n})$



We want h to satisfy: if $x \neq y$, then $\text{HammingDist}(h(x), h(y)) \geq 10n$.



- Protocol:
  1. Alice rearrange $h(x)$ into square shape and randomly pick one row r_i.
  2. Alice send the pair $(i, \vec{r}_i)$.
  3. Bob do the same thing except that he chooses column.
  4. Referre check whether $\vec{r}_i(j) = \vec{c}_j(i)$

- Error Rate Analysis:

$$\Pr[\, h(x)_{i,j} = h(y)_{i,j} \mid x \neq y \,] \leq \frac{3}{4}$$

It remains to construct such h. We just make it a linear map:

$$h(x) = Rx, \quad R \in \mathbb{F}_2^{40n \times n}$$

We need to prove $\exists R$ s.t. $\forall x \neq y, \mathrm{HammingDist}(Rx, Ry) \geq 10n$

We construct by probabilistic method, randomly pick R.
For fixed $x \neq y$,

$$\begin{aligned} \Pr[\,\text{Hamming Dist}(Rx, Ry) \leq 10n\,] &= \Pr[\,|R(x-y)| \leq 10n\,] \\ &\leq e^{-\frac{20}{8}n} \quad \text{(Chernoff bound)} \end{aligned}$$

$$\Pr[\,\exists x \neq y, \mathrm{HammingDist}(Rx, Ry) \leq 10n\,] \leq 2^{2n} \cdot e^{-\frac{20}{8}n} < 1$$

So $\exists R$ s.t. $\forall x \neq y, \mathrm{HammingDist}(Rx, Ry) \geq 10n$.
Hard-code this R into Alice and Bob's protocol.

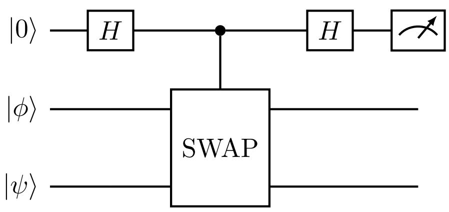
Rmk: Proof of Lower bound $\Omega(\sqrt{n})$ is complicated and omitted.

Quantum: $Q^{||}(EQ) = O(\log n)$

$|h_x\rangle = \frac{1}{\sqrt{40n}} \sum_{i=1}^{40n} |i\rangle \, |h(x)_i\rangle$, where $h(x)_i$ is the i-th bit of $h(x)$
$|h_y\rangle = \frac{1}{\sqrt{40n}} \sum_{i=1}^{40n} |i\rangle \, |h(y)_i\rangle$

Alice sends $|h_x\rangle$, Bob sends $|h_y\rangle$.

Referee: SWAP-test.



test if $|\phi\rangle = |\psi\rangle$ or $|\langle\phi|\psi\rangle| \leq \delta$

$$\begin{aligned} |0\rangle \otimes |\phi\rangle \otimes |\psi\rangle &\longrightarrow \frac{1}{\sqrt{2}} |0\rangle \otimes |\phi\rangle \otimes |\psi\rangle + \frac{1}{\sqrt{2}} |1\rangle \otimes |\psi\rangle \otimes |\phi\rangle \\ &\longrightarrow |0\rangle \otimes \frac{|\phi\rangle \otimes |\psi\rangle + |\psi\rangle \otimes |\phi\rangle}{2} + |1\rangle \otimes \frac{|\phi\rangle \otimes |\psi\rangle - |\psi\rangle \otimes |\phi\rangle}{2} \end{aligned}$$

$\Pr[\text{output} = 1] = \left| \frac{|\phi\rangle \otimes |\psi\rangle - |\psi\rangle \otimes |\phi\rangle}{2} \right|^2 = \frac{1 - \langle\phi|\psi\rangle^2}{2}$
If $|\phi\rangle = |\psi\rangle$, never output 1.
If $\langle\phi|\psi\rangle$ is small, repeat sufficiently large constant times, will output 1 with constant probability.

Note that if $x = y$, $|u_x\rangle = |u_y\rangle$.
If $x \neq y$, $\langle u_x | u_y \rangle = \frac{1}{40n} \sum_{i=1}^{40n} \mathbb{1}[h(x)_i = h(y)_i] \leq \frac{3}{4}$.
Referee can judge with constant probability.

We can use the same method to estimate $\langle\phi|\psi\rangle$.
Repeat swap-test k times, suppose output 1 k' times.
Let $\eta = \begin{cases} \sqrt{1-\frac{2k'}{k}} & \text{if } k' \le \frac{k}{2} \\ 0 & \text{o.w.} \end{cases}$ as our estimate.

Lemma 1. $\Pr\left[|\eta - \langle\phi|\psi\rangle| > \beta\right] < 2e^{-\frac{k\beta^2(\beta+2)^2}{2}}$
Pf. By Chernoff bound:

$$\Pr\left[\left|\frac{k'}{k} - \frac{1-\langle\phi|\psi\rangle^2}{2}\right| > \frac{\beta(\beta+2)}{2}\right] < 2e^{-\frac{k\beta^2(\beta+2)^2}{2}}$$

$$\Pr\left[\left|\eta^2 - \langle\phi|\psi\rangle^2\right| > \beta(\beta+2)\right] < 2e^{-\frac{k\beta^2(\beta+2)^2}{2}}$$

Denote $\Delta = |\eta - \langle\phi|\psi\rangle|$

$$\beta(\beta+2) < |\eta^2 - \langle\phi|\psi\rangle^2| = \Delta|\eta + \langle\phi|\psi\rangle| \le \Delta(\Delta + 2|\langle\phi|\psi\rangle|) \le \Delta(\Delta+2)$$

$$\Delta \ge -|\langle\phi|\psi\rangle| + \sqrt{|\langle\phi|\psi\rangle|^2 + \beta(\beta+2)} > \beta \quad \square$$

1.2.3 Public-coin model

Setting: A, B can share random bits,
i.e. there's an infinite random bit string ξ known to both at first.
A, B both uses random bit in ξ one by one.
A send $a_{x,\xi}$ deterministically. B as well.

Equality function complexity:

$$R^{\|,pub}(EQ) = O(1)$$

Protocol: A send $\langle x, r\rangle \pmod 2$
B send $\langle y, r\rangle \pmod 2$.
r is the shared random bit string.

Analysis: If $x = y$, $\langle x, r\rangle \equiv \langle y, r\rangle \pmod 2$
If $x \ne y$, $\langle x, r\rangle \equiv \langle y, r\rangle \pmod 2$ with probability $\frac{1}{2}$

2 Main results

2.1 Theorem 1

Thm 1. If $R^{\|,pub}(f_n) = O(1)$, then $Q^{\|}(f_n) = O(\log n)$.
More precisely, if $R^{\|,pub}(f_n) \le c$, then $Q^{\|}(f_n) = 2^{O(c)} \cdot \log n$.

Pf. Fix error rate $\epsilon = \frac{1}{10}$.
Suppose a public coin protocol computes f_n using c communication bits.
Let $[M]$ be the message space $M = 2^c$.
$D : [M] \times [M] \to \{0,1\}$ be the referee matrix.
where $\Pr_\xi[f(x,y) \ne D(a_{x,\xi}, b_{y,\xi})] \le \epsilon$.

We first introduce Newman's theorem here.

Thm [Newman] $R^{\|}(f) = O(R^{\|,pub}(f_n) \cdot \log n)$

Pf. Given a public coin protocol Π using random string ξ.
We claim there exists $L = O(n)$ strings $\xi_1, \ldots, \xi_L$ s.t. if ξ is uniformly randomly picked from $\xi_1, \ldots, \xi_L$, the protocol Π still works.

Use probabilistic method:
uniformly randomly generate $\xi_1, \dots, \xi_L$ with same length as ξ.
For a fixed (x, y),

$$\forall i \in [L], \quad \Pr_{\xi_i}[\Pi(x, y; \xi_i) \neq f_n(x, y)] \leq \frac{\epsilon}{2}$$

Let $X_i(x, y) = \mathbb{1}[\Pi(x, y; \xi_i) \neq f_n(x, y)]$
By Chernoff bound,

$$\Pr\left[\frac{1}{L}\sum_{i=1}^{L} X_i(x, y) > \epsilon\right] \leq e^{-\frac{\epsilon^2}{2}L}$$

take $L > \frac{100}{\epsilon^2}n$, then $\Pr\left[\frac{1}{L}\sum_{i=1}^{L} X_i(x, y) > \epsilon\right] < 2^{-2n}$.

$$\Pr\left[\exists (x, y), \frac{1}{L}\sum_{i=1}^{L} X_i(x, y) > \epsilon\right] < 1$$

So there exists L strings $\xi_1, \dots, \xi_L$ s.t.

$$\forall (x, y) \quad \frac{1}{L}\sum_{i=1}^{L} X_i(x, y) \leq \epsilon$$

$$\Pr_{i \in [L]}[\Pi(x, y; \xi_i) \neq f_n(x, y)] \leq \epsilon$$

The claim holds true.

Now we construct a private-coin protocol.
Hard code $\xi_1, \dots, \xi_L$ at first.
Alice uniformly random pick $M \subseteq [L]$, $|M| = \sqrt{L}$.
Bob uniformly random pick $N \subseteq [L]$, $|N| = \sqrt{L}$.
Alice send $\{(i, a_{x,\xi_i}) \mid i \in M\}$.
Bob send $\{(j, b_{y,\xi_j}) \mid j \in N\}$.
Referee check whether $M \cap N = \emptyset$.
If not, take one $t \in M \cap N$ and run protocol as $\Pi(a_{x,\xi_t}, b_{y,\xi_t}; \xi_t)$.
If yes (i.e., $M \cap N = \emptyset$), reject.
By birthday paradox, $M \cap N \neq \emptyset$ with constant probability.
Condition on this, the private coin protocol behaves the same as the public coin one. □

Now using the same technique in Newman's theorem,
we assume ξ is uniformly chosen from $\xi_1, \dots, \xi_L$, $L = O(n)$.
Then $\left|f(x, y) - \frac{1}{L}\sum_{1 \leq i \leq L} D(a_{\xi_i}(x), b_{\xi_i}(y))\right| < \epsilon$.
Let $|u_x\rangle = \frac{1}{\sqrt{L}}\sum_{1 \leq i \leq L} |a_{\xi_i}(x)\rangle |i\rangle$
$|v_y\rangle = \frac{1}{\sqrt{L}}\sum_{1 \leq i \leq L} |b_{\xi_i}(y)\rangle |i\rangle$

Alice send $|u_x\rangle$, Bob send $|v_y\rangle$, which takes $O(\log n) + O(c)$ qubits.
Repeat k times (k to be set).
Referee would like to estimate $\frac{1}{L}\sum_{1 \leq i \leq L} D(a_{\xi_i}(x), b_{\xi_i}(y))$ to approximate $f(x, y)$.

$\frac{1}{L}\sum_{1 \leq i \leq L} D(a_{\xi_i}(x), b_{\xi_i}(y)) = \frac{1}{L}\sum_{1 \leq t, t' \leq M} D(t, t')|A_t(x) \cap B_{t'}(y)|$
where $A_t(x) := \{i : a_{\xi_i}(x) = t\}$, $B_{t'}(y) := \{i : b_{\xi_i}(y) = t'\}$

Let $|u_{x,t}\rangle = \sum_{i \in A_t(x)} |i\rangle$, $\quad |v_{y,t'}\rangle = \sum_{i \in B_{t'}(y)} |i\rangle$.
then $\langle u_{x,t} | v_{y,t'}\rangle = |A_t(x) \cap B_{t'}(y)|$
So $\frac{1}{L}\sum_{1 \leq i \leq L} D(a_{\xi_i}(x), b_{\xi_i}(y)) = \frac{1}{L}\sum_{1 \leq t, t' \leq M} D(t, t') \langle u_{x,t} | v_{y,t'}\rangle$

We're going to estimate $\langle u_{x,t} | v_{y,t'}\rangle$, which we've seen the same thing in Lemma 1.
However, referee only have $|u_x\rangle, |v_y\rangle$ not $|u_{x,t}\rangle, |v_{y,t'}\rangle$.
Need to do some transformation first.

$$|u_x\rangle = \frac{1}{\sqrt{L}}\sum_{1 \leq t \leq M} |t\rangle |u_{x,t}\rangle$$

$$|v_y\rangle = \frac{1}{\sqrt{L}} \sum_{1\le t'\le M} |t'\rangle\,|v_{y,t'}\rangle$$

apply unitary to them and get (with an auxiliary qubit)

$$|u'_x\rangle = \frac{1}{\sqrt{L}}\left(|0\rangle \otimes |t\rangle\,|u_{x,t}\rangle + \sum_{\tau\neq t} |0\rangle \otimes |\tau\rangle\,|u_{x,\tau}\rangle\right)$$

$$\left|v'_y\right\rangle = \frac{1}{\sqrt{L}}\left(|0\rangle \otimes |t'\rangle\,|v_{y,t'}\rangle + \sum_{\tau\neq t'} |1\rangle \otimes |\tau\rangle\,|v_{y,\tau}\rangle\right)$$

(Notice $\|\,\left|v'_y\right\rangle\| = 1$, $\|\,|u'_x\rangle\| = 1$, so such unitary exists.) (Typo here in original paper)
(In fact, Alice can directly send $|u'_x\rangle$. Bob send $\left|v'_y\right\rangle$)

$$\langle u'_x|v'_y\rangle = \frac{1}{L}\langle u_{x,t}|v_{y,t'}\rangle$$

By lemma 1, Repeat the procedure $k = O(M^8 \log M)$ times we get an estimate $\eta(t,t')$ of $\frac{\langle u_{x,t}|v_{y,t'}\rangle}{L}$ s.t.

$$\Pr\left[\left|\eta - \frac{\langle u_{x,t}|v_{y,t'}\rangle}{L}\right| > \frac{\epsilon}{M^2}\right] < \frac{\epsilon}{M^2}$$

Do the same thing for each (t,t'), which multiply $O(M^2)$ to complexity.
We can estimate $f(x,y)$ within 2ϵ by $\sum_{1\le t,t'\le M} D(t,t')\eta(t,t')$.
Referee answers $f(x,y)=1$ iff $\sum_{1\le t,t'\le M} D(t,t')\eta(t,t') > \frac{1}{2}$.

$O(M^{10}\log M(\log n + c))$ communication bits

2.2 Theorem 2

Thm 2. $R^{\|,pub}(HAM_n^{(d)}) = O(d^2)$
where $HAM_n^{(d)}(x,y) = \begin{cases} 1 & \text{if HammingDist}(x,y) \le d \\ 0 & \text{o.w.} \end{cases}$

Pf. We'll construct a protocol with γd^2 communication bits where $\gamma = 10000$.
Public coin consists of $z_1, z_2, \ldots, z_{\gamma d^2}$, each of which is a n-bit string. Every bit is set as 1 with probability $p = \frac{1}{2d}$ independently.
Alice send $a = a_1 a_2 \ldots a_{\gamma d^2}$ where $a_i = \langle x, z_i\rangle \pmod 2$.
Bob send $b = b_1 b_2 \ldots b_{\gamma d^2}$ where $b_i = \langle y, z_i\rangle \pmod 2$.
Referee answers 1 iff $\text{HammingDist}(a,b) \le \frac{\gamma d^2}{2} - q\gamma d^2$
where $q = \frac{1}{4}\left((1-\frac{1}{d})^d + (1-\frac{1}{d})^{d+1}\right)$.

Lemma. Assume $\text{HammingDist}(x,y) = k$. Then each $a_i \oplus b_i$ is an independent $Ber(\alpha_k)$, where

$$\alpha_k = \frac{1}{2} - \frac{1}{2}\left(1 - \frac{1}{d}\right)^k$$

Pf. $a_i \oplus b_i = 1 \iff \langle x\oplus y, z_i\rangle \equiv 1 \pmod 2$.
z_i has odd number of 1s on those k positions where x, y differs.

$$\begin{aligned}\Pr[a_i \oplus b_i = 1] &= \sum_{\substack{0\le i\le k \\ i \text{ is odd}}} \binom{k}{i} p^i (1-p)^{k-i} \\ &= \frac{1}{2}\left(1 - (1-2p)^k\right) \\ &= \frac{1}{2} - \frac{1}{2}\left(1 - \frac{1}{d}\right)^k \square\end{aligned}$$

$$
\begin{aligned}
&\Pr\left[\#1\text{'s in } a \oplus b \le \frac{\gamma d^2}{2} - q\gamma d^2 \,\middle|\, k \ge d+1\right] \\
&= \Pr\left[\frac{1}{\gamma d^2}\sum_{i=1}^{\gamma d^2} a_i \oplus b_i \le \frac{1}{2} - q \,\middle|\, k \ge d+1\right] \\
&\le e^{-2\gamma d(1-\frac{1}{d})^d}
\end{aligned}
$$

Similarly for the other side of error rate.
The error rate is bounded by constant. □

Cor: For constant d, $Q^{\|}(HAM_n^{(d)}) = O(\log n)$.

2.3 Theorem 3

Next, we'd like to improve the constant in Thm 1, and generalize to those f with $R^{\|,pub}(f) \neq O(1)$.
It's natural to ask $Q^{\|}(f) \stackrel{?}{=} O(R^{\|,pub}(f) \cdot \log n)$, since $R^{\|}(f) = O(R^{\|,pub}(f) \cdot \sqrt{n})$.
Next theorem gives a partial result.

Thm 3. $\mathcal{A}$ is a public-coin protocol computing f using $M \times M$ referee matrix D. Then

$$Q^{\|}(f) = O(w(D)^5(1 + \log w(D))(\log M + \log n))$$

where $w(D)$ is "convex width", namely, the smallest integer k s.t. D is the sum of k matrices isomorphic to some real positive semidefinite matrices with only nonnegative entries.
Two matrices are isomorphic if they are equal by permuting rows and columns.

Rmk. Since $w(D) \le M$, Theorem 3 is a generalization of Theorem 1.

Pf. Same as Thm 1, the goal is to send appropriate states to referee, s.t. he can estimate

$$\sum_{1 \le t,t' \le M} D(t,t') \frac{\langle u_{x,t} | v_{y,t'} \rangle}{L}$$

Since $D = \sum_{1 \le \ell \le w(D)} G_\ell$

$$
\begin{aligned}
&\sum_{1 \le t,t' \le M} D(t,t') \frac{\langle u_{x,t} | v_{y,t'} \rangle}{L} \\
&= \sum_{1 \le \ell \le w(D)} \left(\sum_{1 \le t,t' \le M} G_\ell(t,t') \frac{\langle u_{x,t} | v_{y,t'} \rangle}{L} \right)
\end{aligned}
$$

We'd like to estimate $\sum_{1 \le t,t' \le M} G(t,t') \frac{\langle u_{x,t} | v_{y,t'} \rangle}{L}$
WLOG, suppose G is positive semidefinite, otherwise our protocol can adaptively renaming t, t'.
Let $G = R\Lambda R^{-1}$ where $R = (r_{t,s})$ is orthogonal
$\Lambda = \text{diag}(\lambda_s)$

Let $|u'_{x,s}\rangle = \sum_{1 \le t \le M} r_{t,s} |u_{x,t}\rangle$
$|v'_{y,s}\rangle = \sum_{1 \le t \le M} r_{t,s} |v_{y,t}\rangle$

Let $|u'_x\rangle = \frac{1}{\sqrt{L}} \sum_{1 \le s \le M} \sqrt{\lambda_s} |s\rangle |u'_{x,s}\rangle$
$|v'_y\rangle = \frac{1}{\sqrt{L}} \sum_{1 \le s \le M} \sqrt{\lambda_s} |s\rangle |v'_{y,s}\rangle$

Lemma: $\langle u'_x | v'_y \rangle = \sum_{1 \le t,t' \le M} G(t,t') \frac{\langle u_{x,t} | v_{y,t'} \rangle}{L}$
Furthermore $\| \, |u'_x\rangle \, \| \le 1, \quad \| \, |v'_y\rangle \, \| \le 1$

Pf.

$$
\begin{aligned}
\langle u'_x | v'_y \rangle &= \frac{1}{L} \sum_{1\le s\le M} \lambda_s \langle u'_{x,s} | v'_{y,s} \rangle \\
&= \frac{1}{L} \sum_{1\le s\le M} \lambda_s \sum_{1\le t,t'\le M} r_{t,s} r_{t',s} \langle u_{x,t} | v_{y,t'} \rangle \\
&= \frac{1}{L} \sum_{1\le t,t'\le M} (R\Lambda R^T)_{t,t'} \langle u_{x,t} | v_{y,t'} \rangle \\
&= \frac{1}{L} \sum_{1\le t,t'\le M} G(t,t') \langle u_{x,t} | v_{y,t'} \rangle
\end{aligned}
$$

$$
\begin{aligned}
\langle u'_x | u'_x \rangle &= \frac{1}{L} \sum_{1\le t,t'\le M} G(t,t') \langle u_{x,t} | u_{x,t'} \rangle \\
&= \frac{1}{L} \sum_{1\le t\le M} G(t,t) \cdot \| u_{x,t} \|^2 \\
&\le \frac{1}{L} \sum_{1\le t\le M} \| u_{x,t} \|^2 = 1
\end{aligned}
$$

since all entries of G are between 0 and 1.
Similarly $\| \left| v'_y \right\rangle \|^2 \le 1$. $\square$

Now, we have to regularize $|u'_x\rangle, \left|v'_y\right\rangle$ before sending.
Suppose $\cos\theta_x = \|u'_x\|, \quad \cos\phi_y = \|v'_y\|$.
Alice sends $|u''_x\rangle = |0\rangle |u'_x\rangle + \sin\theta_x |1\rangle |\kappa\rangle$
$\left|v''_y\right\rangle = |0\rangle \left|v'_y\right\rangle + \sin\phi_y |1\rangle |\kappa'\rangle$
where $|\kappa\rangle, |\kappa'\rangle$ are two fixed mutually orthonormal vectors.
$\left\langle u''_x \middle| v''_y \right\rangle = \left\langle u'_x \middle| v'_y \right\rangle = \sum_{1\le t,t'\le M} G(t,t') \frac{\left\langle u_{x,t} \middle| v_{y,t'} \right\rangle}{L}$.

By repeating $k = O(w(D)^4(1+\log w(D)))$ times,
we get an estimate η of $\sum_{1\le t,t'\le M} G(t,t') \frac{\left\langle u_{x,t} \middle| v_{y,t'} \middle| u_{x,t} \middle| v_{y,t'} \right\rangle}{L}$
s.t. $\Pr\left[\left|\eta - \sum_{1\le t,t'\le M} G(t,t') \frac{\left\langle u_{x,t} \middle| v_{y,t'} \middle| u_{x,t} \middle| v_{y,t'} \right\rangle}{L}\right| > \frac{\epsilon}{w(D)}\right] < \frac{\epsilon}{w(D)}$.

Then do the same thing for each G in $G_1, \ldots, G_{w(D)}$,
we can estimate $f(x,y)$ within 2ϵ
by $\sum_{1\le \ell\le w(D)} \eta(G_\ell)$.
$O(w(D)^5(1+\log w(D))(\log M + \log n))$ communication bits. $\square$