Lemma 5.6. *For any $\rho_{WE} \in \mathcal{D}(\mathcal{H}_W \otimes \mathcal{H}_E)$ with classical W, and for any classical $\rho_V \in \mathcal{D}(\mathcal{H}_W)$:*

$$\operatorname{rank}(\rho_{WE} - \rho_V \otimes \rho_E) \leq \dim(\mathcal{H}_W) \cdot \operatorname{rank}(\rho_E)\,.$$

Proof. By the block-diagonal structure of the density matrices $\rho_{WE} = \sum_w P_W(w)|w\rangle\langle w| \otimes \rho_E^w$ and $\rho_V \otimes \rho_E = \sum_w P_V(w)|w\rangle\langle w| \otimes \rho_E$, it is sufficient to show that $\operatorname{rank}\big(P_W(w)\rho_E^w - P_V(w)\rho_E\big) \leq \operatorname{rank}(\rho_E)$ for any w. First, we note that for an arbitrary positive semi-definite matrix $\sigma \geq 0$, its kernel $\ker(\sigma)$ coincides with $\{|\varphi\rangle \,|\, \langle\varphi|\sigma|\varphi\rangle = 0\}$; this can easily be seen using the spectral decomposition of σ. From this, and since $\rho_E = \sum_w P_W(w)\rho_E^w$, it follows that $\ker(\rho_E) \subseteq \ker(\rho_E^w)$ for any w with $P_W(w) > 0$. From this it now follows that $\ker\big(P_W(w)\rho_E^w - P_V(w)\rho_E\big) \supseteq \ker(\rho_E)$ for all w, which implies the relation on the ranks. □

Proof (of Theorem 5.3). By Lemma 4.8 and using the fact that S, U and E are all independent,

$$\begin{aligned}
&\delta\big(\rho_{f(S,X)SE}, \rho_U \otimes \rho_{SE}\big) = E_{s\leftarrow S}\big[\delta\big(\rho_{f(S,X)E|S=s}, \rho_U \otimes \rho_{E|S=s}\big)\big] = E_{s\leftarrow S}\big[\delta\big(\rho_{f(s,X)E}, \rho_U \otimes \rho_E\big)\big] \\
&\leq \frac{1}{2}2^{\frac{1}{2}(\ell+\mathrm{H}_0(\rho_E))} E_{s\leftarrow S}\left[\sqrt{\Delta\big(\rho_{f(s,X)E}, \rho_U \otimes \rho_E\big)}\right] \leq \frac{1}{2}2^{\frac{1}{2}(\ell+\mathrm{H}_0(\rho_E))}\sqrt{E_{s\leftarrow S}\big[\Delta\big(\rho_{f(s,X)E}, \rho_U \otimes \rho_E\big)\big]}\,,
\end{aligned}$$

where the first inequality follows from Lemmas 5.5 and 5.6, and the second from Jensen's inequality (Proposition B.4). It thus suffices to show that $E_{s\leftarrow S}\big[\Delta\big(\rho_{f(s,X)E}, \rho_U \otimes \rho_E\big)\big] \leq 2^{-\mathrm{H}_2(X)}$.

Let s be a fixed value in $\mathcal{S}$. Writing $W = f(s, X)$, we can re-write the square of the Hilbert-Schmidt distance as follows.

$$\begin{aligned}
\Delta\big(\rho_{WE}, \rho_U \otimes \rho_E\big) &= \operatorname{tr}\Big(\big(\sum_w P_W(w)|w\rangle\langle w| \otimes \rho_E^w - 2^{-\ell}\sum_w |w\rangle\langle w| \otimes \rho_E\big)^2\Big) \\
&= \operatorname{tr}\Big(\big(\sum_w |w\rangle\langle w| \otimes (P_W(w)\rho_E^w - 2^{-\ell}\rho_E)\big)^2\Big) \\
&= \operatorname{tr}\Big(\sum_w \big(P_W(w)\rho_E^w - 2^{-\ell}\rho_E\big)^2\Big) \\
&= \operatorname{tr}\Big(\sum_w \big(P_W(w)^2(\rho_E^w)^2 - 2P_W(w)2^{-\ell}\rho_E^w\rho_E + 2^{-2\ell}(\rho_E)^2\big)\Big) \\
&= \operatorname{tr}\Big(\sum_w P_W(w)^2(\rho_E^w)^2 - 2^{-\ell}(\rho_E)^2\Big)
\end{aligned}$$

where the third equality follows from the fact that the matrix at hand is in block-diagonal form. Plugging in $f(s, X)$ for W and using linearity of the trace, we thus get

$$\Delta\big(\rho_{f(s,X)E}, \rho_U \otimes \rho_E\big) = \operatorname{tr}\Big(\sum_w P_{f(s,X)}(w)^2(\rho_E^w)^2\Big) - \operatorname{tr}\Big(2^{-\ell}(\rho_E)^2\Big)\,.$$

Now note that

$$\begin{aligned}
\operatorname{tr}\Big(\sum_w P_{f(s,X)}(w)^2(\rho_E^w)^2\Big) &= \operatorname{tr}\Big(\sum_w \big(\sum_{x:f(s,x)=w} P_X(x)\rho_E^x\big)\big(\sum_{x':f(s,x')=w} P_X(x')\rho_E^{x'}\big)\Big) \\
&= \operatorname{tr}\Big(\sum_{x,x'} P_X(x)P_X(x')\delta_{f(s,x),f(s,x')}\rho_E^x\rho_E^{x'}\Big) = \sum_{x,x'} P_X(x)P_X(x')\delta_{f(s,x),f(s,x')}\operatorname{tr}\big(\rho_E^x\rho_E^{x'}\big)
\end{aligned}$$

where $\delta_{w,w'}$ is the Kronecker Delta with $\delta_{w,w'} = 1$ if $w = w'$ and 0 otherwise, and

$$\operatorname{tr}\big(2^{-\ell}\rho_E^2\big) = \operatorname{tr}\Big(2^{-\ell}\big(\sum_x P_X(x)\rho_E^x\big)^2\Big) = \sum_{x,x'} P_X(x)P_X(x')2^{-\ell}\operatorname{tr}\big(\rho_E^x\rho_E^{x'}\big)$$

Using that due to the universal property $E_{s\leftarrow S}[\delta_{f(s,x),f(s,x')}] = P[f(S,x)=f(S,x')] \leq 2^{-\ell}$ for all $x \neq x'$, and using that $0 \leq \mathrm{tr}\big((\rho_E^x)^2\big) \leq 1$ and $0 \leq \mathrm{tr}\big(\rho_E^x \rho_E^{x'}\big)$, it follows that

$$\begin{aligned} E_{s\leftarrow S}\big[\Delta\big(\rho_{f(s,X)E}, \rho_U \otimes \rho_E\big)\big] &= \sum_{x,x'} P_X(x) P_X(x') \big(E_{s\leftarrow S}[\delta_{f(s,x),f(s,x')}] - 2^{-\ell}\big) \mathrm{tr}\big(\rho_E^x \rho_E^{x'}\big) \\ &\leq \sum_x P_X(x)^2 \mathrm{tr}\big(\rho_x^2\big) \leq \sum_x P_X(x)^2 = 2^{-\mathrm{H}_2(X)}\,, \end{aligned}$$

which was to be shown. □

The proof of the following follows easily from Theorem 5.3 and Jensen's inequality.

Corollary 5.7. *Let X be a random variable over $\mathcal{X}$, let Y be a random variable over $\mathcal{Y}$, and let E be a quantum system whose state may depend on X and Y. Let $f : \mathcal{S} \times \mathcal{X} \to \{0,1\}^\ell$ be a universal hash function and let S be uniformly distributed over $\mathcal{S}$ (independent of X, Y and E). Then*

$$\delta\big(\rho_{f(S,X)SYE}, \rho_U \otimes \rho_{SYE}\big) \leq \frac{1}{2}\, 2^{-\frac{1}{2}(\mathrm{H}_2(X|Y) - \mathrm{H}_0(\rho_E) - \ell)} \leq \frac{1}{2}\, 2^{-\frac{1}{2}(\mathrm{H}_\infty(X|Y) - \mathrm{H}_0(\rho_E) - \ell)}\,,$$

where U is uniformly distributed over $\{0,1\}^\ell$ (independent of X, S and E).

6 Subset Sampling

6.1 The Classical Case

The **Hamming weight** $W(x)$ of a bit-string $x = (x_1, \ldots, x_n) \in \{0,1\}^n$ is defined to be the number of 1's occurring within x. Similarly, the **relative** Hamming weight $\omega(x)$ of x is given by its Hamming weight divided by its bit-length n: $\omega(x) = W(x)/n$. We say that the relative Hamming weight of x is *ε-close* to $\beta \in \mathbb{R}$, denoted as $\omega(x) \approx_\varepsilon \beta$, if $|\omega(x) - \beta| \leq \varepsilon$. For any subset $T \subseteq \{1, \ldots, n\}$ of size k, we write x_T for the restriction of x to the positions in T: $x_T = (x_i)_{i \in T} \in \{0,1\}^k$.

Consider the following problem: we want to *estimate* the (relative) Hamming weight of an unknown but fixed string $x \in \{0,1\}^n$ (of known bit-length n) by only looking at a small number of positions in x. A canonical way to do so is as follows: choose at random a sample subset $T \subset \{1, \ldots, n\}$ of linear size (i.e. size αn for some constant $0 < \alpha < 1$), and take $\omega(x_T)$ as estimate for $\omega(x)$. Very generally, we allow the following kinds of sampling strategies: choose a sample subset $T \subset \{1, \ldots, n\}$ according to *some* fixed probability distribution P_T, and compute the estimate for $\omega(x)$ as *some* function of x_T.

Definition 6.1. *A* **sampling strategy** *consists of a tuple $\Sigma = (P_T, \hat{\omega})$, where P_T is a distribution over the subsets of $\{1, \ldots, n\}$, and $\hat{\omega}$ is a function*

$$\hat{\omega} : \{(t, v) \,|\, t \subseteq \{1, \ldots, n\}, v \in \{0,1\}^{|t|}\} \to \mathbb{R}\,.$$

We want to measure the reliability of such a general estimation strategy, i.e., how well it predicts the (relative) Hamming weight of the string x. Actually, for technical reasons (and because the positions within the sample subset T are anyway revealed), we want to measure how well such a general strategy predicts the (relative) Hamming weight of $x_{\bar{T}} \in \{0,1\}^m$ (where $m = n - |T|$), i.e., of x restricted to the positions $\bar{T} = \{1, \ldots, n\} \setminus T$ outside of the sample T (see Figure 10, top).

Definition 6.2. *Let $\Sigma = (P_T, \hat{\omega})$ be a sampling strategy. For any $t \subseteq \{1, \ldots, n\}$ and $\varepsilon > 0$, we define*

$$B_t^\varepsilon(\Sigma) := \{x \in \{0,1\}^n \,|\, |\hat{\omega}(t, x_t) - \omega(x_{\bar{t}})| \leq \varepsilon\}\,.$$

Informally, $B_t^\varepsilon(\Sigma)$ collects all strings x for which the estimate is ε-close to the real value, assuming that subset t has been used for computing the estimate. If the sampling strategy is clear from the context, we simply write B_t^ε instead of $B_t^\varepsilon(\Sigma)$.

Definition 6.3. *For any $\varepsilon > 0$, we define the* **error probability**

$$err_\varepsilon(\Sigma) := \max_{x \in \{0,1\}^n} P[x \notin B_T^\varepsilon] = \max_{x \in \{0,1\}^n} \sum_{t : x \notin B_t^\varepsilon} P_T(t)\,.$$

By definition, for any choice of $x \in \{0,1\}^n$: $\omega(x_{\bar{T}}) \approx_\varepsilon \hat{\omega}(T, x_T)$ except with probability at most $err_\varepsilon(\Sigma)$. Using classical sampling theory (like Theorem B.3), one can for instance show that for the above canonical example with a random T of size αn with $\alpha \leq \frac{1}{2}$: $err_\varepsilon(\Sigma) \leq 2e^{-\varepsilon^2 \alpha n/2}$, and as such is exponentially small in n for fixed $\varepsilon, \alpha > 0$.

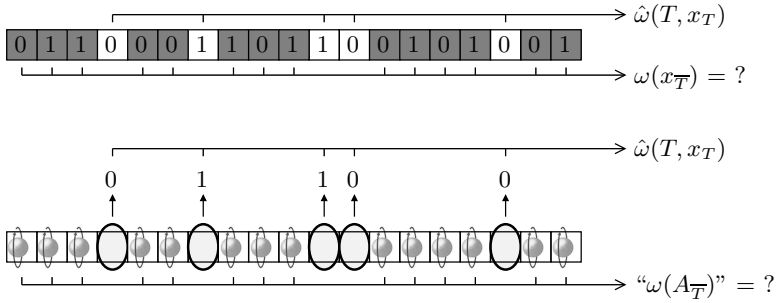

Fig. 10. Estimating ω for a string (top) and a quantum state (bottom).

6.2 The Quantum Case

We now turn to the corresponding quantum problem. Consider a n-qubit system $\mathsf{A} = \mathsf{A}_1 \cdots \mathsf{A}_n$, possibly entangled with another system E, and we want to estimate in a similar way as above how far away the state of A is from the all-zero state $|0 \cdots 0\rangle$, by *measuring* only a small number of the qubits. An estimation strategy as used above in the classical setting, defined by the distribution P_T and the function $\hat{\omega}$, can also be applied here: choose $T \subset \{1, \ldots, m\}$ according to P_T, measure the qubits $\mathsf{A}_T = (\mathsf{A}_i)_{i \in T}$ qubit-wise in the computational basis to obtain x_T, and compute $\hat{\omega}(T, x_T)$ as estimate for the "relative Hamming weight" of the remaining system $\mathsf{A}_{\bar{T}}$ (see Figure 10, bottom). It remains to discuss what it should mean to have an estimation of the Hamming weight of a multi-qubit quantum system, and how to define and how to compute the reliability of a general strategy in this context.

Definition 6.4. *Let $\Sigma = (P_T, \hat{\omega})$ be a sampling strategy. For any $t \subseteq \{1, \ldots, n\}$ and $\varepsilon > 0$, we define*

$$\mathrm{span}(B_t^\varepsilon) := \mathrm{span}\{|x\rangle \,|\, x \in B_t^\varepsilon\} \subset \mathcal{H}_\mathsf{A} = \mathbb{C}^2 \otimes \ldots \otimes \mathbb{C}^2\,.$$

Let the (pure) state of AE be given by $|\varphi_{\mathsf{AE}}\rangle$. We can describe this state together with the (independent) choice of the subset T by means of the hybrid state

$$\rho_{TAE} = \rho_T \otimes \rho_{AE} = \sum_t P_T(t) |t\rangle\langle t| \otimes |\varphi_{AE}\rangle\langle\varphi_{AE}|\,.$$

Definition 6.5. *For any $\varepsilon > 0$, we define the* **quantum error probability**

$$err^*_\varepsilon(\Sigma) := \sup_{\mathcal{H}_E} \sup_{|\varphi_{AE}\rangle} \inf_{\tilde{\rho}_{TAE}} \delta(\rho_{TAE}, \tilde{\rho}_{TAE}),$$

where the first sup *is over all finite-dimensional $\mathcal{H}_E$, the second* sup *is over all state vectors $|\varphi_{AE}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_E$, and the* inf *is over all hybrid states $\tilde{\rho}_{TAE}$ of the form*

$$\tilde{\rho}_{TAE} = \sum_t P_T(t) |t\rangle\langle t| \otimes |\tilde{\varphi}^t_{AE}\rangle\langle\tilde{\varphi}^t_{AE}|$$

with $|\tilde{\varphi}^t_{AE}\rangle \in \mathrm{span}(B^\varepsilon_t) \otimes \mathcal{H}_E$ for any t.

The following follows immediately from the above definition.

Corollary 6.6. *Let the hybrid state $\rho_{TSA'E}$ be obtained by applying a sampling strategy Σ to a state $|\varphi_{AE}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_E$, where $S = X_T$ is the observed sample and $A' = A_{\bar{T}}$ the remaining qubits. Then, for any $\varepsilon > 0$, there exists a hybrid state $\tilde{\rho}_{TSA'E}$ with $\delta(\rho_{TSA'E}, \tilde{\rho}_{TSA'E}) \leq err^*_\varepsilon(\Sigma)$ and such that for every $t \subseteq \{1, \ldots, n\}$ and $s \in \{0,1\}^{|t|}$:*

$$\tilde{\rho}_{A'E|T=t,S=s} = |\tilde{\varphi}^{ts}_{A'E}\rangle\langle\tilde{\varphi}^{ts}_{A'E}| \qquad \text{with} \qquad |\tilde{\varphi}^{ts}_{A'E}\rangle = \sum_{z:\, \omega(z) \approx_\varepsilon \beta} \alpha^{ts}_z |z\rangle \otimes |\tilde{\varphi}^{tsz}_E\rangle$$

where $\beta = \hat{\omega}(t, s)$.

What will be important for us is that β allows us to bound the number of z's occurring in the sum. Indeed, it is known that for any $\beta \leq \frac{1}{2}$, the number of $z \in \{0,1\}^m$ with $\omega(z) \leq \beta + \varepsilon$ is upper bounded by $2^{\mathrm{h}(\beta+\varepsilon)m}$, where h is the binary entropy function (see Lemma B.12).

Proposition 6.7. *For any sampling strategy Σ and for any $\varepsilon > 0$:*

$$err^*_\varepsilon(\Sigma) \leq \sqrt{err_\varepsilon(\Sigma)}\,.$$

Proof. We show that for any state vector $|\varphi_{AE}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_E$, with arbitrary $\mathcal{H}_E$, there exists a suitable ideal state $\tilde{\rho}_{TAE}$ with $\delta(\rho_{TAE}, \tilde{\rho}_{TAE}) \leq \sqrt{err_\varepsilon(\Sigma)}$. We construct $\tilde{\rho}_{TAE}$ as required by Definition 6.5, where the $|\tilde{\varphi}^t_{AE}\rangle$'s are defined by the following decomposition into orthogonal components:

$$|\varphi_{AE}\rangle = \Pi_t|\varphi_{AE}\rangle + \Pi_t^\perp|\varphi_{AE}\rangle = \langle\tilde{\varphi}^t_{AE}|\varphi_{AE}\rangle|\tilde{\varphi}^t_{AE}\rangle + \langle\tilde{\varphi}^{t\perp}_{AE}|\varphi_{AE}\rangle|\tilde{\varphi}^{t\perp}_{AE}\rangle,$$

where Π_t is the orthogonal projection into $\mathrm{span}(B^\varepsilon_t) \otimes \mathcal{H}_E$,

$$\Pi_t^\perp = \sum_{x \notin B^\varepsilon_t} |x\rangle\langle x| \otimes \mathbb{I}$$

the projection into the orthogonal complement $\mathrm{span}(B^\varepsilon_t)^\perp \otimes \mathcal{H}_E$, and $|\tilde{\varphi}^t_{AE}\rangle$ and $|\tilde{\varphi}^{t\perp}_{AE}\rangle$ are the renormalized projections. It follows that $\langle\varphi_{AE}|\Pi_t^\perp|\varphi_{AE}\rangle = |\langle\tilde{\varphi}^{t\perp}_{AE}|\varphi_{AE}\rangle|^2$, and hence that

$$\begin{aligned}\sum_t P_T(t)\, |\langle\tilde{\varphi}^{t\perp}_{AE}|\varphi_{AE}\rangle|^2 &= \sum_t P_T(t) \sum_{x \notin B^\varepsilon_t} \langle\varphi_{AE}|(|x\rangle\langle x| \otimes \mathbb{I})|\varphi_{AE}\rangle \\ &= \sum_{x \in \{0,1\}^n} \langle\varphi_{AE}|(|x\rangle\langle x| \otimes \mathbb{I})|\varphi_{AE}\rangle \sum_{t: x \notin B^\varepsilon_t} P_T(t) = \sum_{x \in \{0,1\}^n} \langle\varphi_{AE}|(|x\rangle\langle x| \otimes \mathbb{I})|\varphi_{AE}\rangle\, P[x \notin B^\varepsilon_T] \leq err_\varepsilon(\Sigma)\,,\end{aligned}$$

where the inequality follows by definition of $err_\varepsilon(\Sigma)$ and since $\sum_x |\langle\varphi_{AE}|(|x\rangle\langle x|\otimes\mathbb{I})|\varphi_{AE}\rangle = 1$. From elementary properties of the trace distance, and using Jensen's inequality, we conclude that

$$\delta\big(\rho_{TAE},\tilde\rho_{TAE}\big) = \sum_t P_T(t)\delta\big(|\varphi_{AE}\rangle\langle\varphi_{AE}|,|\tilde\varphi^t_{AE}\rangle\langle\tilde\varphi^t_{AE}|\big) = \sum_t P_T(t)\sqrt{1-|\langle\tilde\varphi^t_{AE}|\varphi_{AE}\rangle|^2}$$
$$= \sum_t P_T(t)|\langle\tilde\varphi^{t\perp}_{AE}|\varphi_{AE}\rangle| \le \sqrt{\sum_t P_T(t)|\langle\tilde\varphi^{t\perp}_{AE}|\varphi_{AE}\rangle|^2} \le \sqrt{err_\varepsilon(\Sigma)}\,,$$

which was to be shown. □

As mentioned above, what will be important for us is that if a small value of $\beta = \hat\omega(T,x_T)$ is observed, then the remaining state is (close to) a superposition of a *small* number of orthogonal states. This allows us to apply the following lemma.

Lemma 6.8. *For arbitrary $\mathcal{H}_A$ and $\mathcal{H}_E$, let $|\varphi_{AE}\rangle\in\mathcal{H}_A\otimes\mathcal{H}_E$ be a state of the form $|\varphi_{AE}\rangle = \sum_{i\in J}\alpha_i|i\rangle|\varphi^i_E\rangle$, where $\{|i\rangle\}_{i\in I}$ is a basis of $\mathcal{H}_A$ and $J\subseteq I$. Then, the following holds.*

1. *Let $\tilde\rho_{AE} = \sum_{i\in J}|\alpha_i|^2|i\rangle\langle i|\otimes|\varphi^i_E\rangle\langle\varphi^i_E|$ be the corresponding mixture, and let W and $\tilde W$ be the outcome of measuring A of $|\varphi_{AE}\rangle$ respectively of $\tilde\rho_{AE}$ in some basis $\{|w\rangle\}_{w\in\mathcal{W}}$. Then,*

$$\mathrm{H}_\infty(W) \ge \mathrm{H}_\infty(\tilde W) - \log|J|\,.$$

2. *The reduced density matrix $\rho_E = \mathrm{tr}_A(|\varphi_{AE}\rangle\langle\varphi_{AE}|)$ has max-entropy*

$$\mathrm{H}_0(\rho_E) \le \log|J|\,.$$

Proof. For 1., we write $\mathbb{I}\in\mathcal{E}nd(\mathcal{H}_E)$ as $\mathbb{I} = \sum_z|z\rangle\langle z|$ for a basis of $\mathcal{H}_E$. We may understand $\tilde\rho_{AE}$ as being in state $|i\rangle|\varphi^i_E\rangle$ with probability $|\alpha_i|^2$, so that

$$P_{\tilde W}(w) = \sum_{i\in J}|\alpha_i|^2(\langle i|\otimes\langle\varphi^i_E|)(|w\rangle\langle w|\otimes\mathbb{I})(|i\rangle\otimes|\varphi^i_E\rangle) = \sum_z\sum_{i\in J}|\alpha_i\langle w|i\rangle\langle z|\varphi^i_E\rangle|^2\cdot\sum_{i\in J}1^2\cdot\frac{1}{|J|}$$
$$\ge \frac{1}{|J|}\cdot\sum_z\Big|\sum_{i\in J}\alpha_i\langle w|i\rangle\langle z|\varphi^i_E\rangle\Big|^2 = \frac{1}{|J|}\cdot\sum_z\sum_{i,j\in J}\bar\alpha_j\alpha_i\langle j|w\rangle\langle w|i\rangle\langle\varphi^j_E|z\rangle\langle z|\varphi^i_E\rangle$$
$$= \frac{1}{|J|}\cdot\langle\varphi_{AE}|(|w\rangle\langle w|\otimes\mathbb{I})|\varphi_{AE}\rangle = \frac{1}{|J|}\cdot P_W(w)\,,$$

where the inequality is Cauchy-Schwarz. This proves 1.

For 2., note that $\rho_E = \mathrm{tr}_A(\rho_{AE}) = \sum_{i\in J}|\alpha_i|^2|\varphi^i_E\rangle\langle\varphi^i_E|$. The claim follows immediately from the sub-additivity of the rank:

$$\mathrm{rank}(\rho_E) \le \sum_{i\in J}\mathrm{rank}\big(|\alpha_i|^2|\varphi^i_E\rangle\langle\varphi^i_E|\big) = \sum_{i\in J}1 = |J|\,,$$

where we use that the $|\varphi^i_E\rangle\langle\varphi^i_E|$'s have rank 1. □

6.3 An Example Application

In the following, let $\mathcal{H}_A = \mathcal{H}_{A_1}\otimes\cdots\otimes\mathcal{H}_{A_n}$ and $\mathcal{H}_B = \mathcal{H}_{B_1}\otimes\cdots\otimes\mathcal{H}_{B_n}$ be n-qubit state spaces (i.e. $\mathcal{H}_{A_i} = \mathcal{H}_{B_i} = \mathbb{C}^2$), and let $\mathcal{H}_E$ be an arbitrary state space. Consider a pure state $|\psi_{ABE}\rangle$ in $\mathcal{H}_{ABE} = \mathcal{H}_A\otimes\mathcal{H}_B\otimes\mathcal{H}_E$ which is of the very special form

$$|\psi_{ABE}\rangle = 2^{-n/2}\sum_{x\in\{0,1\}^n}|x_1\rangle\cdots|x_n\rangle|x_1\rangle\cdots|x_n\rangle|\psi_E\rangle\,,$$

i.e., every $\mathsf{A}_i\mathsf{B}_i$ forms an EPR pair, and E is in some arbitrary (pure) state $|\psi_{\mathsf{E}}\rangle \in \mathcal{H}_{\mathsf{E}}$. Let $\rho_{\Theta XYE}$ be the state obtained by choosing $\Theta \in \{0,1\}^n$ at random and measuring each 2-qubit system $\mathsf{A}_i\mathsf{B}_i$ for $i \in \{1,\ldots,n\}$ qubit-wise in basis $\{H^{\Theta_i}|0\rangle, H^{\Theta_i}|1\rangle\}$ to obtain $(X_i, Y_i) \in \{0,1\}^2$. Formally, $\rho_{\Theta XYE}$ can be written as

$$\rho_{\Theta XYE} = \mathcal{T}_{\Theta\mathsf{AB}}(\rho_\Theta \otimes |\psi_{\mathsf{ABE}}\rangle\langle\psi_{\mathsf{ABE}}|)$$

where $\mathcal{T}_{\Theta\mathsf{AB}}$ decomposes into $\mathcal{T}_{\Theta\mathsf{AB}} = \mathcal{T}_{\Theta_n\mathsf{B}_n} \circ \mathcal{T}_{\Theta_n\mathsf{A}_n} \circ \ldots \circ \mathcal{T}_{\Theta_1\mathsf{B}_1} \circ \mathcal{T}_{\Theta_1\mathsf{A}_1}$, and where the quantum operation $\mathcal{T}_{\Theta_i\mathsf{A}_i}$ implements the action of "looking up Θ_i and measuring A_i accordingly", and similarly for $\mathcal{T}_{\Theta_i\mathsf{B}_i}$.[18] It is not too hard to see that all these components commute with each other. Note that by the properties of an EPR pair, it holds that $X = Y$ and $\mathrm{H}_\infty(X|\Theta) = n$. Furthermore, since E is not entangled with the other systems, $\rho_{\mathsf{E}} = |\psi_{\mathsf{E}}\rangle\langle\psi_{\mathsf{E}}|$ and as such $\mathrm{H}_0(\rho_{\mathsf{E}}) = 0$.

Of course if the initial state $|\psi_{\mathsf{ABE}}\rangle$ has a different form, say, may be arbitrary, then these values for the min- and max-entropies are not guaranteed anymore. For instance some of the subsystems A_i may be in state $|0\rangle$ (and thus not entangled with B_i), so that measuring A_i in basis $\{|0\rangle, |1\rangle\}$ produces no uncertainty. Or, A_i may be entangled with E (instead of with B_i), so that ρ_{E} is not pure anymore and thus $\mathrm{H}_0(\rho_{\mathsf{E}}) > 0$. However, in both cases, since some of the $\mathsf{A}_i\mathsf{B}_i$'s are not EPR pairs anymore, an additional effect is that also $X = Y$ does not hold anymore with certainty. In fact, one might expect the Hamming distance between X and Y, i.e., the Hamming weight $W(X \oplus Y)$, to give an estimation on how far away the initial state $|\psi_{\mathsf{ABE}}\rangle$ was from the ideal case considered in the beginning, and thus how far away the values for $\mathrm{H}_\infty(X|\Theta)$ and $\mathrm{H}_0(\rho_{\mathsf{E}})$ are from the ideal values n and 0, respectively.

The following theorem proves this intuition correct, and makes the statement precise.

Proposition 6.9. *Let $\mathsf{A} = \mathsf{A}_1 \cdots \mathsf{A}_n$ and $\mathsf{B} = \mathsf{B}_1 \cdots \mathsf{B}_n$ be n-qubit systems, and E is arbitrary. Let $|\psi_{\mathsf{ABE}}\rangle \in \mathcal{H}_{\mathsf{ABE}}$ be a pure state, and let $\rho_{\Theta XY\mathsf{E}} = \mathcal{T}_{\Theta\mathsf{AB}}(\rho_\Theta \otimes |\psi_{\mathsf{ABE}}\rangle\langle\psi_{\mathsf{ABE}}|)$ be obtained as described above. I.e., Θ is random in $\{0,1\}^n$ and X_i is obtained by measuring A_i in basis $\{H^{\Theta_i}|0\rangle, H^{\Theta_i}|1\rangle\}$, and similarly Y_i. Then, for any $0 < \delta < \frac{1}{2}$ there exists $\tilde{\rho}_{\Theta XY\mathsf{E}} \in \mathcal{D}(\mathcal{H}_{\mathsf{ABE}})$ with classical Θ, X and Y, so that*

$$\delta\big(\rho_{\Theta XY\mathsf{E}}, \tilde{\rho}_{\Theta XY\mathsf{E}}\big) \leq e^{-\frac{1}{4}\delta^2 n}$$

and such that for every $\theta, s \in \{0,1\}^n$, where s has relative Hamming weight $\omega(s) < (\frac{1}{2} - \delta)$,

$$\mathrm{H}_\infty(\tilde{P}_{X|\Theta=\theta, X\oplus Y=s}) \geq n - \mathrm{h}\big(\omega(s)+\delta\big)n \quad \textit{and} \quad \mathrm{H}_0(\tilde{\rho}_{\mathsf{E}|\Theta=\theta, X\oplus Y=s}) \leq \mathrm{h}\big(\omega(s)+\delta\big)n\,.$$

Remark 6.10. Actually, with the appropriate definition of min-entropy conditioned on quantum information (see Remark 5.4), one can show that $\mathrm{H}_\infty(\tilde{\rho}_{X\mathsf{E}|\Theta=\theta, X\oplus Y=s}|\mathsf{E}) \geq n - \mathrm{h}\big(\omega(s)+\delta\big)n$.

The proof of Proposition 6.9 relies on the unitary transformation $U \in \mathcal{U}(\mathbb{C}^2 \otimes \mathbb{C}^2)$ defined by

$$U|x\rangle|y\rangle = |x \oplus y\rangle H|x\rangle\,,$$

and which similarly satisfies that $U(H|x\rangle H|y\rangle) = H|y\rangle|x \oplus y\rangle$, for $x, y \in \{0,1\}$. Now, we are fully equipped for the proof of Proposition 6.9.

Proof (of Proposition 6.9). Let $\rho_{\Theta WS\mathsf{E}}$ be the state determined by (and which itself determines) $\rho_{\Theta XY\mathsf{E}}$ by letting

$$W_i = \begin{cases} X_i \text{ if } \Theta_i = 0 \\ Y_i \text{ if } \Theta_i = 1 \end{cases} \qquad \text{and} \qquad S_i = X_i \oplus Y_i\,. \tag{1}$$

[18] Here, and similarly later, the quantum operation $\mathcal{T}_{\Theta_i\mathsf{A}_i} : \mathcal{D}(\mathcal{H}_{\Theta_i\mathsf{A}_i}) \to \mathcal{D}(\mathcal{H}_{\Theta_i X_i})$ can be understood as quantum operation $\mathcal{T}_{\Theta_i\mathsf{A}_i} : \mathcal{D}(\mathcal{H}_{\Theta\mathsf{ABE}}) \to \mathcal{D}(\mathcal{H}_{\Theta X\mathsf{BE}})$ that acts as identity on the subsystems different to Θ_i and A_i.

From the above mentioned properties of the unitary U, it follows that $\rho_{\Theta W S E}$ may equivalently be obtained by applying U to each subsystem $A_i B_i$ within $|\psi_{ABE}\rangle$ to obtain $|\varphi_{ABE}\rangle \in \mathcal{H}_{ABE}$, and then measuring each subsystem $A_i B_i$ of the state $|\varphi_{ABE}\rangle$ as follows: A_i in basis $\{|0\rangle, |1\rangle\}$ to obtain S_i and B_i in basis $\{H|0\rangle, H|1\rangle\}$ to obtain W_i if $\Theta_i = 0$, and A_i in basis $\{H|0\rangle, H|1\rangle\}$ to obtain W_i and B_i in basis $\{|0\rangle, |1\rangle\}$ to obtain S_i if $\Theta_i = 1$; see Figure 11.

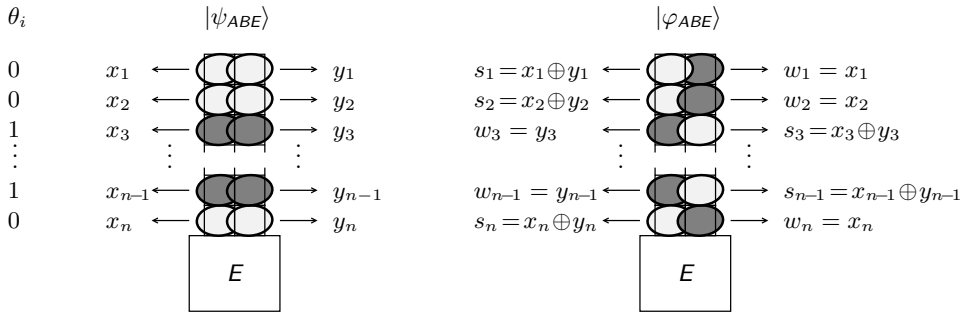

Fig. 11. Two equivalent ways to obtain the state $\rho_{\Theta W S E}$.

Let $\rho_{\Theta S C E}$ be the hybrid state (with classical Θ and S), obtained from the state $|\varphi_{ABE}\rangle$ by measuring, for every i, either A_i or B_i, depending on Θ_i, in the computational basis to obtain S_i, and leaving the other qubit (B_i or A_i) untouched, but calling it C_i now. Due to the commutativity of measuring different subsystems, we can now understand $\rho_{\Theta W S E}$ as being obtained from $\rho_{\Theta S C E}$ by measuring every qubit C_i in the Hadamard basis.

The crucial observation now is that we may understand $\rho_{\Theta S C E}$ as being obtained by applying a sampling strategy Σ to $|\varphi_{ABE}\rangle$. Indeed, it is obtained by first choosing a random and independent string Θ, which specifies a size-n subset of the $2n$ qubits AB (namely it selects A_i or B_i depending on Θ_i), and then $\rho_{\Theta S C E}$ is obtained by measuring the selected qubits in the computational basis (and renaming the unmeasured qubits). It thus follows from Corollary 6.6 that there exists a state $\tilde{\rho}_{\Theta S C E}$ that is $err^*_\delta(\Sigma)$-close to $\rho_{\Theta S C E}$, and for which $\tilde{\rho}_{C E|\Theta=\theta, S=s} = |\tilde{\varphi}^{\theta s}_{CE}\rangle\langle\tilde{\varphi}^{\theta s}_{CE}|$ with $|\tilde{\varphi}^{\theta s}_{CE}\rangle = \sum_z \alpha^{\theta s}_z |z\rangle \otimes |\tilde{\varphi}^{\theta s z}_E\rangle$, where the sum is over all $z \in \{0,1\}^n$ with $\omega(z) \approx_\delta \hat{\omega}(\theta, s) = \omega(s)$. We recall that there are at most $2^{\mathrm{h}(\omega(s)+\delta)}$ such strings z. It now follows from Lemma 6.8 that the state $\tilde{\rho}_{W E|\Theta=\theta, S=s}$, obtained from $|\tilde{\varphi}^{\theta s}_{CE}\rangle$ by measuring C in the Hadamard basis, satisfies

$$\mathrm{H}_\infty(\tilde{P}_{W|\Theta=\theta,S=s}) \geq n - \mathrm{h}\big(\omega(s)+\delta\big)n \quad \text{and} \quad \mathrm{H}_0(\tilde{\rho}_{E|\Theta=\theta,S=s}) \leq \mathrm{h}\big(\omega(s)+\delta\big)n\,,$$

where we used that, for any $z \in \{0,1\}^n$, measuring the state $|z\rangle$ in the Hadamard basis produces a uniformly distributed measurement outcome with n bits of min-entropy, and the same holds when measuring a mixture of such states. Note that since, when given Θ and S, W uniquely determines X and vice versa, the same lower bound also applies to $\mathrm{H}_\infty(\tilde{P}_{W|\Theta=\theta,S=s})$. It remains to argue that $err^*_\delta(\Sigma) \leq e^{-\frac{1}{4}\delta^2 n}$, but this follows from Proposition 6.7, and from the analysis of the classical error probability $err_\delta(\Sigma)$ given below.

Let $x^0 = (x^0_1, \ldots, x^0_n)$ and $x^1 = (x^1_1, \ldots, x^1_n)$ be two arbitrary but fixed n-bit strings. For any $\theta \in \{0,1\}^n$, we let x^θ denote the n-bit string $x^\theta = (x^{\theta_1}_1, \ldots, x^{\theta_n}_n)$ and $x^{\bar{\theta}}$ the n-bit string $x^{\bar{\theta}} = (x^{\theta_1 \oplus 1}_1, \ldots, x^{\theta_n \oplus 1}_n)$. We want to bound the probability that $W(x^{\bar{\Theta}}) > W(x^\Theta) + \delta n$ for a uniformly distributed Θ. Let d be the number of positions i with $x^0_i \neq x^1_i$, and let y^Θ and $y^{\bar{\Theta}}$ be the restriction of x^Θ and $x^{\bar{\Theta}}$ to those positions (so that $W(y^\Theta) + W(y^{\bar{\Theta}}) = d$). Then,

$$\begin{aligned} P\big[W(x^{\bar{\Theta}}) > W(x^\Theta) + \delta n\big] &= P\big[W(y^{\bar{\Theta}}) > W(y^\Theta) + \delta n\big] \\ &= P\big[W(y^{\bar{\Theta}}) > \tfrac{1}{2}d + \tfrac{1}{2}\delta n\big] \leq e^{-2(\frac{\delta n}{2d})^2 d} = e^{-\frac{1}{2}\delta^2 n \frac{n}{d}} \leq e^{-\frac{1}{2}\delta^2 n}\,, \end{aligned}$$

where the first inequality is Hoeffding's inequality (Theorem B.3). This concludes the proof. □

7 Quantum Key-Distribution

7.1 Setting

We consider two parties, *Alice* and *Bob*, which want to agree on a secret key in the presence of an adversary *Eve*. We assume that Alice and Bob may communicate over a quantum channel as well as over a classical communication channel. However, Eve may have full control over the quantum channel, and she may fully eavesdrop on the classical communication. See Figure 12. Still, the goal for Alice and Bob is to produce a secret key k, which is known to both Alice and Bob but about which Eve has essentially no information, merely by communicating over these "insecure" channels. In contrast to computationally-secure key-distribution schemes (like the Diffie-Hellman scheme) we aim here for *information-theoretic* security, i.e., security against a computationally unbounded Eve.

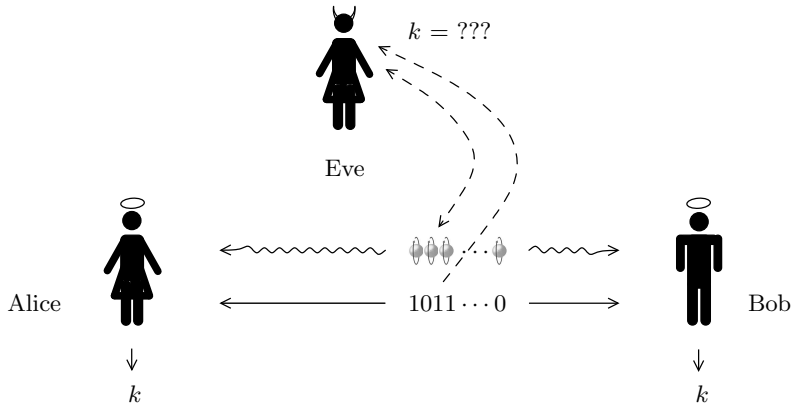

Fig. 12. Quantum key distribution by public communication.

We stress that we allow Eve to *read* the classical communication between Alice and Bob, but we do *not* allow her to *modify* it (without being noticed). In cryptographic terms, we require the classical communication to be authentic. This could for instance be achieved by using an information-theoretically secure authentication code (assuming that Alice and Bob share a short authentication key), but we assume here that it is a physical property of the communication channel (like a phone line where one can recognize the other person's voice). Indeed, without such an authentication mechanism, there is no way to prevent Eve from simply impersonating, say, Alice and letting Bob believe that he is actually producing a common key with Alice.

7.2 The Protocol

Below in Figure 13 is the quantum key distribution (QKD) protocol that allows Alice and Bob to agree on a key k in the above setting. The protocol is parametrized by parameters n, ℓ, and κ, where n is the number of qubits transmitted, ℓ is the size of the key k to agree upon, κ is some security parameter (n also acts as security parameter). The protocol makes use of two universal hash functions $f : \mathcal{S} \times \{0,1\}^n \to \{0,1\}^\ell$ and $g : \mathcal{T} \times \{0,1\}^n \to \{0,1\}^\kappa$.

This protocol $\mathcal{QKD}$ requires Bob to *store* the received qubits for a short while. However, reliably storing quantum states turns out to be technologically extremely challenging. In the following variant (see Figure 14), Alice and Bob only need to prepare, communicate and measure-upon-arrival qubits; they do not need to store qubits or apply any quantum operations (whereas we allow Eve to do arbitrary quantum operations). These are operations that can be done with current technology (though note the discussion in Section 7.4).

PROTOCOL $\mathcal{QKD}$

Preparation: Alice picks random $x, \theta \in \{0,1\}^n$ and sends $H^{\theta_1}|x_1\rangle, \ldots, H^{\theta_n}|x_n\rangle$ to Bob. Once Bob has confirmed receipt of the qubits, Alice sends θ to Bob, and Bob measures the received qubits in respective bases $H^{\theta_1}\{|0\rangle, |1\rangle\}, \ldots, H^{\theta_n}\{|0\rangle, |1\rangle\}$ to obtain $y \in \{0,1\}^n$.

Verification: Alice chooses a random seed $t \in \mathcal{T}$ and sends t and $test := g(t,x)$ to Bob. If $test = g(t,y)$ then Bob sends "`okay`" to Alice; else, he sends "`reject`" and both abort.

Key extraction: Alice chooses a random seed $s \in \mathcal{S}$ and sends it to Bob. Alice and Bob compute their keys as $k := f(s,x)$ and $k' := f(s,y)$, respectively.

Fig. 13. Protocol $\mathcal{QKD}$

It is not hard to see that for any strategy of Eve in $\mathcal{QKD}_{BB84}$, resulting in a hybrid state $\rho_{\Theta XY\mathsf{E}}$ after the sifting step, there exists a strategy for Eve in $\mathcal{QKD}$ that results in the very same hybrid state $\rho_{\Theta XY\mathsf{E}}$ after the preparation. As such, security of $\mathcal{QKD}$ implies security of $\mathcal{QKD}_{BB84}$. In the remainder, we thus focus on $\mathcal{QKD}$ and prove the security of $\mathcal{QKD}$.

PROTOCOL $\mathcal{QKD}_{BB84}$

Preparation: Alice picks random $x, \theta \in \{0,1\}^N$ and sends $H^{\theta_1}|x_1\rangle, \ldots, H^{\theta_N}|x_N\rangle$ to Bob. Bob picks a random $\theta' \in \{0,1\}^N$ and measures the received qubits in respective bases $H^{\theta'_1}\{|0\rangle, |1\rangle\}, \ldots, H^{\theta'_N}\{|0\rangle, |1\rangle\}$ to obtain $y \in \{0,1\}^N$, and he sends "`done`" to Alice.

Sifting: Alice and Bob exchange θ and θ'. If $|\{i \,|\, \theta_i = \theta'_i\}| < n$, then Alice and Bob restart; else they update x and y by keeping only the first n x_i's respectively y_i's with $\theta_i = \theta'_i$.

Verification: Alice chooses a random seed $t \in \mathcal{T}$ and sends t and $test := g(t,x)$ to Bob. If $test = g(t,y)$ then Bob sends "`okay`" to Alice; else, he sends "`reject`" and both abort.

Key extraction: Alice chooses a random seed $s \in \mathcal{S}$ and sends it to Bob. Alice and Bob compute their keys as $k := f(s,x)$ and $k' := f(s,y)$, respectively.

Fig. 14. Protocol $\mathcal{QKD}_{BB84}$

The intuition behind the security of the protocol is that if Eve does not interfere with the quantum communication then she has no information on x and y and thus on the final key $k = k'$. On the other hand, if she does interfere with the quantum communication to obtain some information, then this will disturb the quantum communication, having the effect that (with high probability) after the sifting step $x \neq y$, and therefore by the universal property that $g(t,x) = g(t,y)$ only with probability $2^{-\kappa}$; thus, in this case Alice and Bob abort (with high probability). Formally, protocol $\mathcal{QKD}$ is secure in the following sense.

Theorem 7.1. *Consider an execution of $\mathcal{QKD}$ with an eavesdropper Eve, resulting in key K for Alice (equal to "$\perp$" in case she aborts) and quantum system E^+ for Eve, where the latter includes all classical communication. Let $\rho_{K\mathsf{E}^+} \in \mathcal{D}(\mathcal{H}_{K\mathsf{E}^+})$ be the corresponding state. Then, for any $\delta > 0$, there exists $\bar{\rho}_{K\mathsf{E}^+} \in \mathcal{D}(\mathcal{H}_{K\mathsf{E}^+})$ with classical K so that*

$$\delta\big(\rho_{K\mathsf{E}^+}, \bar{\rho}_{K\mathsf{E}^+}\big) \leq 2e^{-\frac{1}{4}\delta^2 n} + 2 \cdot 2^{-\kappa} + \frac{1}{2}\, 2^{-\frac{1}{2}\left((1-2\mathrm{h}(\delta))n - \kappa - \ell\right)}$$

and $\bar{\rho}_{K\mathsf{E}^+|K\neq\perp} = \rho_U \otimes \bar{\rho}_{\mathsf{E}^+|K\neq\perp}$ with U uniform over $\{0,1\}^\ell$.

Note that there is no way to prevent Eve from making Alice and Bob abort the protocol, the only thing we can hope for is that in case they do not abort, the agreed-upon key is secure. This is what the above indeed shows (or so far claims).

7.3 Proof of Theorem 7.1

First, we note that the hybrid states $\rho_{\Theta_i X_i B_i} = \frac{1}{4}\sum_{\theta_i,x_i} |\theta_i\rangle\langle\theta_i| \otimes |x_i\rangle\langle x_i| \otimes H^{\theta_i}|x_i\rangle\langle x_i|H^{\theta_i}$ that Alice prepares can also be obtained as

$$\rho_{\Theta_i X_i B_i} = \mathcal{T}_{\Theta_i A_i}(\rho_{\Theta_i} \otimes |\Phi_{A_i B_i}\rangle\langle\Phi_{A_i B_i}|)\,,$$

where $\rho_{\Theta_i} = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$ and $|\Phi_{A_i B_i}\rangle$ is an EPR pair, and $\mathcal{T}_{\Theta_i A_i}$ is the quantum operation that "looks up Θ_i and measures A_i in basis $\{H^{\Theta_i}|0\rangle, H^{\Theta_i}|1\rangle\}$". This follows from the basic properties of an EPR pair. We emphasize that $\mathcal{T}_{\Theta_i A_i}$ only acts on systems Θ_i and A_i, whereas Eve acts on the B_i's and her local quantum system E. As such $\mathcal{T}_{\Theta A} = \mathcal{T}_{\Theta_n A_n} \circ \ldots \circ \mathcal{T}_{\Theta_1 A_1}$ and Eve's action $\mathcal{T}_{BE}$ commute, so that

$$\rho_{\Theta X B E} = \mathcal{T}_{BE}(\rho_{\Theta X B}) = \mathcal{T}_{\Theta A}\big(\mathcal{T}_{BE}(\rho_\Theta \otimes |\Phi_{AB}\rangle\langle\Phi_{AB}|)\big) = \mathcal{T}_{\Theta A}\big(\rho_\Theta \otimes \mathcal{T}_{BE}(|\Phi_{AB}\rangle\langle\Phi_{AB}|)\big)\,,$$

where we may assume that $\mathcal{T}_{BE}(|\Phi_{AB}\rangle\langle\Phi_{AB}|) = |\psi_{ABE}\rangle\langle\psi_{ABE}|$ for some pure state $|\psi_{ABE}\rangle$. Note that a priori, $\mathcal{T}_{BE}(|\Phi_{AB}\rangle\langle\Phi_{AB}|)$ may be mixed, since a quantum operation in general involves tracing out a subsystem; however, tracing out some subsystem is only to Eve's disadvantage.

Finally, the Y_i's are obtained by measuring the B_i's:

$$\rho_{\Theta X Y E} = \mathcal{T}_{\Theta B} \circ \mathcal{T}_{\Theta A}\big(\rho_\Theta \otimes |\psi_{ABE}\rangle\langle\psi_{ABE}|\big) = \mathcal{T}_{\Theta AB}\big(\rho_\Theta \otimes |\psi_{ABE}\rangle\langle\psi_{ABE}|\big)\,.$$

Note that $\rho_{\Theta X Y E}$ is in the right form to apply Proposition 6.9. Hence, we can conclude that there exists $\tilde{\rho}_{\Theta X Y E} \in \mathcal{D}(\mathcal{H}_{ABE})$ with classical Θ, X and Y, so that

$$\delta\big(\rho_{\Theta X Y E}, \tilde{\rho}_{\Theta X Y E}\big) \leq e^{-\frac{1}{4}\delta^2 n}$$

and such that for every $\theta, s \in \{0,1\}^n$, where s has relative Hamming weight $\omega(s) < (\frac{1}{2} - \delta)$,

$$\mathrm{H}_\infty(\tilde{P}_{X|\Theta=\theta, X\oplus Y=s}) \geq n - \mathrm{h}\big(\omega(s)+\delta\big)n \quad \text{and} \quad \mathrm{H}_0(\tilde{\rho}_{E|\Theta=\theta, X\oplus Y=s}) \leq \mathrm{h}\big(\omega(s)+\delta\big)n\,. \qquad (2)$$

Since the trace distance does not increase under quantum operations (Theorem 4.6), it is sufficient to show existence of $\bar{\rho}_{KE^+}$ that is close enough to $\tilde{\rho}_{KE^+}$. For simplicity, we will omit the tilde and write $\rho_{\Theta X Y E}$ instead of $\tilde{\rho}_{\Theta X Y E}$, etc. Furthermore, we write T for the randomly chosen seed during the verification, and set $Test = g(T, X)$. Finally, we write S for the randomly chosen seed during key extraction, so that $K = f(S, X)$. This results in the state $\rho_{K S T Test \Theta X Y E}$, and ρ_{KE^+} is given by $\rho_{KE^+} = \mathrm{tr}_{XY}(\rho_{K S T Test \Theta X Y E})$, with $S\,T\,Test\,\Theta E$ being collected into E^+. We can write

$$\begin{aligned}\rho_{XYE^+} = {} & P[Test \neq g(T,Y)] \cdot \rho_{XYE^+|Test\neq g(T,Y)} + P[X=Y] \cdot \rho_{XYE^+|X=Y} \\ & + P[Test = g(T,Y) \wedge X \neq Y] \cdot \rho_{XYE^+|Test=g(T,Y)\wedge X\neq Y}\,,\end{aligned}$$

where $P[Test = g(T,Y) \wedge X \neq Y] \leq 2^{-\kappa}$ by the universality of g.

We now set

$$\begin{aligned}\bar{\rho}_{KXYE^+} = {} & P[Test \neq g(T,Y)] \cdot |\perp\rangle\langle\perp| \otimes \rho_{XYE^+|Test\neq g(T,Y)} + P[X=Y] \cdot \rho_U \otimes \rho_{XYE^+|X=Y} \\ & + P[Test = g(T,Y) \wedge X \neq Y] \cdot |\perp\rangle\langle\perp| \otimes \rho_{XYE^+|Test=g(T,Y)\wedge X\neq Y}\,,\end{aligned}$$

and $\bar{\rho}_{K\boldsymbol{E}^+} = \mathrm{tr}_{XY}(\bar{\rho}_{KXY\boldsymbol{E}^+})$. It follows immediately from triangle inequality that

$$\delta\big(\rho_{K\boldsymbol{E}^+}, \bar{\rho}_{K\boldsymbol{E}^+}\big) \leq 2^{-\kappa} + \delta(\rho_{K\boldsymbol{E}^+|X=Y}, \rho_U \otimes \rho_{\boldsymbol{E}^+|X=Y}),$$

and thus to finish the proof it is sufficient to show that $\delta(\rho_{K\boldsymbol{E}^+|X=Y}, \rho_U \otimes \rho_{\boldsymbol{E}^+|X=Y})$ is upper bounded by $\frac{1}{2}2^{-\frac{1}{2}(n-2\mathrm{h}(\delta)n-\kappa-\ell)}$, but this follows from privacy amplification, as we now outline.

For an arbitrary but fixed choice of θ in $\{0,1\}^n$, we use Ω as a short hand for the event $\Omega = [\Theta = \theta \wedge X = Y]$, and we consider the hybrid state $\rho_{XTTest\boldsymbol{E}|\Omega}$. From (2), it follows that $\mathrm{H}_\infty(X|\Omega) \geq n - \mathrm{h}\big(\delta)n$ and $\mathrm{H}_0(\rho_{\boldsymbol{E}|\Omega}) \geq \mathrm{h}(\delta)n$. By the independence of T, the first bound also holds conditioned on T, and from the chain rule Lemma B.16 we obtain $\mathrm{H}_\infty(X|TTest, \Omega) \geq n - \mathrm{h}(\delta)n - \kappa$. From privacy amplification (Corollary 5.7), it now follows that

$$\delta\big(\rho_{f(S,X)STTest\boldsymbol{E}|\Theta=\theta,X=Y}, \rho_U \otimes \rho_{STTest\boldsymbol{E}|\Theta=\theta,X=Y}\big) \leq \frac{1}{2}2^{-\frac{1}{2}(n-2\mathrm{h}(\delta)n-\kappa-\ell)}.$$

By Lemma 4.8, the same bound also holds for $\delta\big(\rho_{f(S,X)STTest\Theta\boldsymbol{E}|X=Y}, \rho_U \otimes \rho_{STTest\Theta\boldsymbol{E}|X=Y}\big)$, which concludes the proof. □

7.4 Dealing With Noisy Quantum Communication

From a practical point of view, it is not realistic to assume that if Alice sends qubit $|x\rangle$ and Bob measures it in "correct" basis $\{|0\rangle, |1\rangle\}$, that then Bob observes x with certainty, as the rule predicts. In reality, this only holds with non-perfect probability $1-\eta$, for some (hopefully small) η. This is due to imperfection of the devices and due to noise. Protocol $\mathcal{QKD}$ (and $\mathcal{QKD}_{BB84}$) can be modified in order to take care of such noisy quantum communication by replacing the *verification* step, which verifies if $x = y$, by an *error-estimation* step, which estimates the number of bits in which x and y differ, and by adding an *error-correction* step.

In the *error-estimation* step, Alice and Bob compare x and y on κ randomly chosen positions, and use this to estimate the number of positions in which x and y differ. Hoeffding's inequality ensures that the obtained estimate β for the relative Hamming weight $\omega(x \oplus y)$ is off by at most δ, except with probability $2e^{-2\delta^2\kappa}$. In the analysis, this then affects the bounds on $\mathrm{H}_\infty(X|\Omega)$ and $\mathrm{H}_0(\rho_{\boldsymbol{E}|\Omega})$ in that $\mathrm{h}(\delta)$ is replaced by $\mathrm{h}(\beta + 2\delta)$.

In the *error-estimation* step, Alice sends to Bob the *syndrome* of x with respect to an error correcting code that is capable of correcting a $(\beta+\delta)$-fraction of errors. This allows Bob to recover the correct x from y. Using a "good" error correcting code, the syndrome is $\mathrm{h}(\beta+2\delta)n$ bits long, reducing the min-entropy by an additional $\mathrm{h}(\beta+2\delta)n$ bits.

As a result, with now $(1-3h(\beta+2\delta))n-\kappa-\ell$ in the exponent of the error, QKD is possible as long as η, the noise level, satisfies $h(\eta) < \frac{1}{3}$, which approximately evaluates to $\eta < 0.061$. Using the approach mentioned in Remarks 5.4 and 6.10 allows for $(1-2h(\gamma+2\delta))n-\kappa-\ell$ in the exponent, which means that $h(\eta) < \frac{1}{2}$ needs to hold, which approximately evaluates to $\eta < 0.11$.