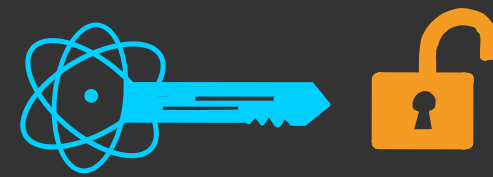
Attacks on QKD: A cautionary tale

Ramona Wolf
University of Siegen



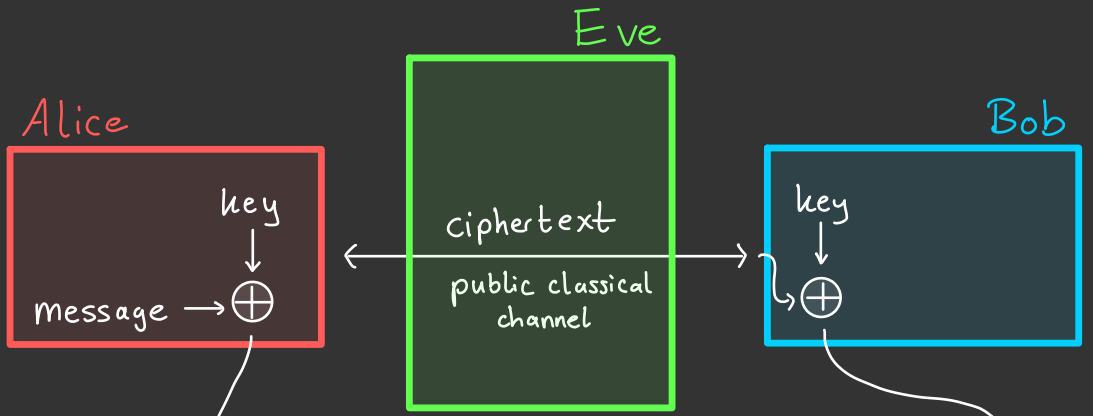
Secure communication

Goal:

* Messages cannot be overheard
* Messages cannot be tampered with

Information-theoretic security

Unbreakable encryption scheme: The one-time pad



message 0 1 0 1 1 0 0 1 ...

random → key 1 0 0 1 0 1 0 0 ...

random → ciphertext 1 1 0 0 1 1 0 1 ...

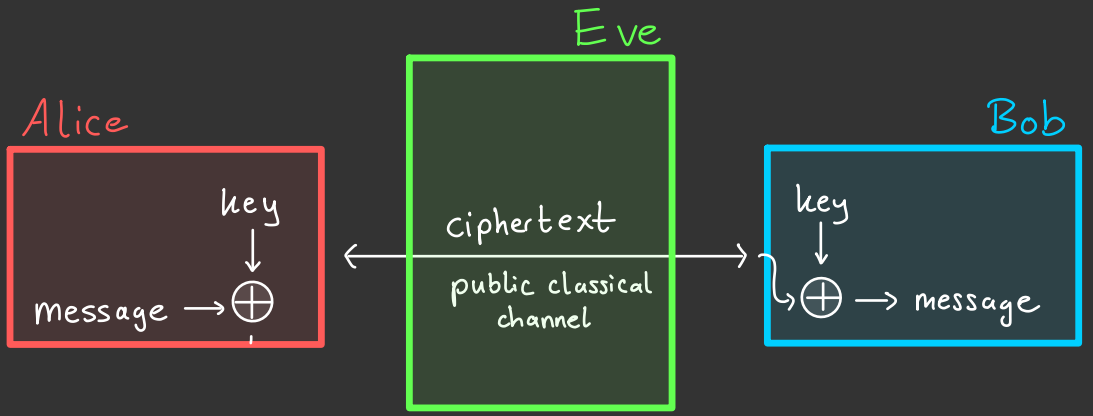
ciphertext 1 1 0 0 1 1 0 1 ...

key 1 0 0 1 0 1 0 0 ...

message 0 1 0 1 1 0 0 1 ...

Information-theoretic security

Unbreakable encryption scheme: The one-time pad



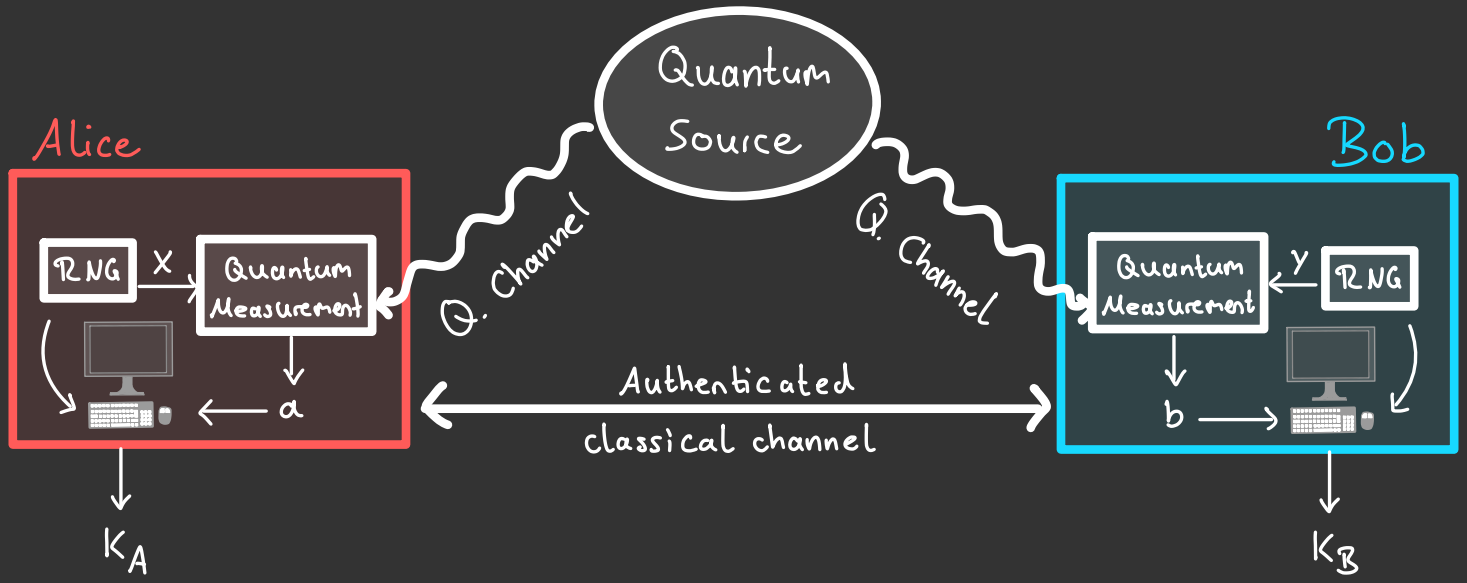
Cryptographic key:

1. Uniformly random
2. Identical for Alice and Bob
3. Private

Quantum physics allows us to generate this!

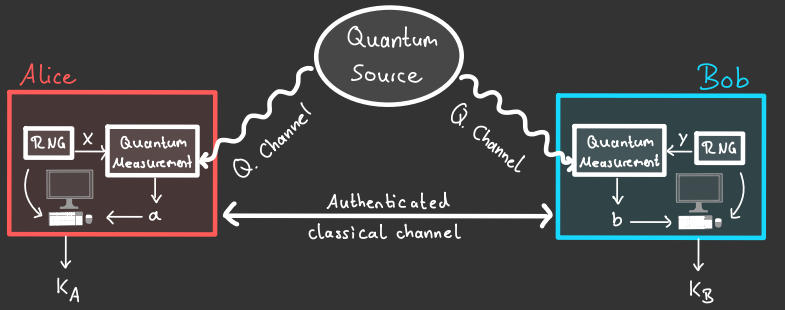
Quantum key distribution

The setup:



Quantum key distribution

Example: $|\Psi_{AB}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$



Alice measures in $\{|0\rangle, |1\rangle\}$ - basis

$p = 50\%$ → $a = 0$

$p = 50\%$ → $a = 1$

Post-measurement state on Bob's side: $|\Psi_B\rangle = |0\rangle$

Post-measurement state on Bob's side: $|\Psi_B\rangle = |1\rangle$

Bob measures in $\{|0\rangle, |1\rangle\}$ - basis → $b = 0$

Bob measures in $\{|+\rangle, |-\rangle\}$ - basis → 50% $b = 0$, 50% $b = 1$

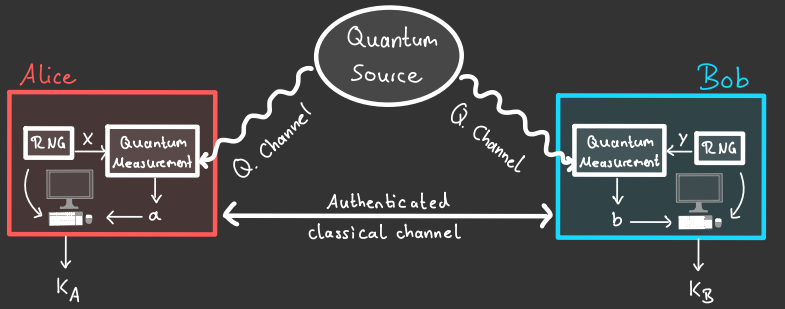
Bob measures in $\{|0\rangle, |1\rangle\}$ - basis → $b = 1$

Bob measures in $\{|+\rangle, |-\rangle\}$ - basis → 50% $b = 0$, 50% $b = 1$

Perfectly correlated outputs if Alice & Bob choose the same basis

Uncorrelated outputs if they choose different bases

Quantum key distribution



The protocol:

I. Quantum phase

n rounds {
1. The source distributes quantum states
2. Alice and Bob measure these states
} Eve introduces errors here

3. Output: raw keys $A_1 \dots A_n$, $B_1 \dots B_n$ → Imperfect cryptographic key

II. Classical post-processing

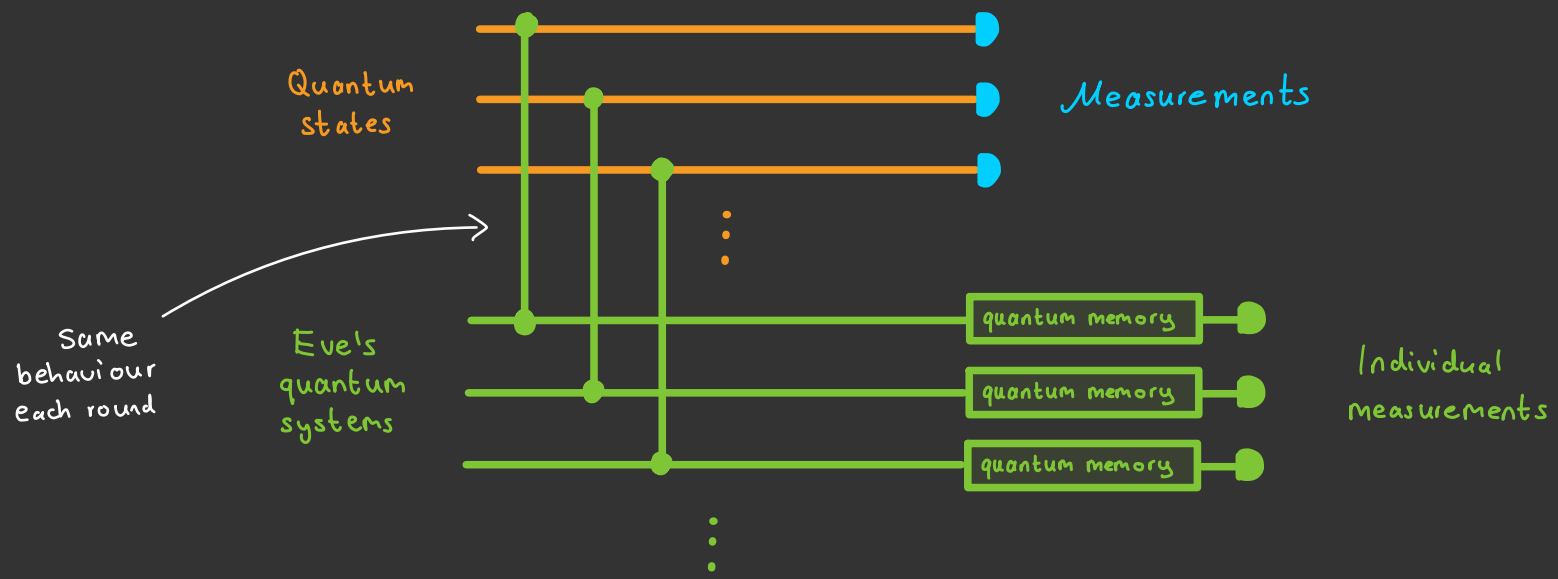
1. Alice and Bob estimate the errors in their raw keys:
   How close are R_A, R_B to a perfect key? → Quantum Mechanics allows us to quantify this!

$$H_{min}^{\varepsilon}(A_1 \dots A_n | E)$$

2. They perform an error-correction protocol → identical bit strings
3. They do privacy amplification → random, private keys K_A, K_B

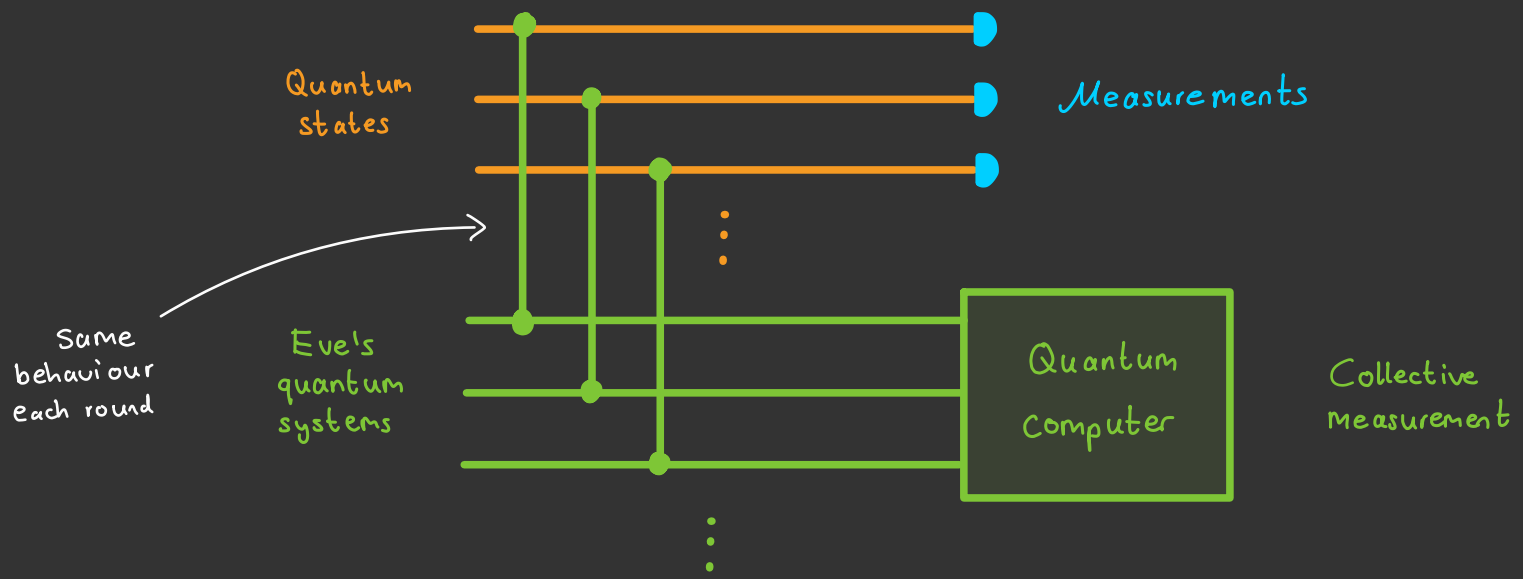
Attacks

1. Individual attacks:



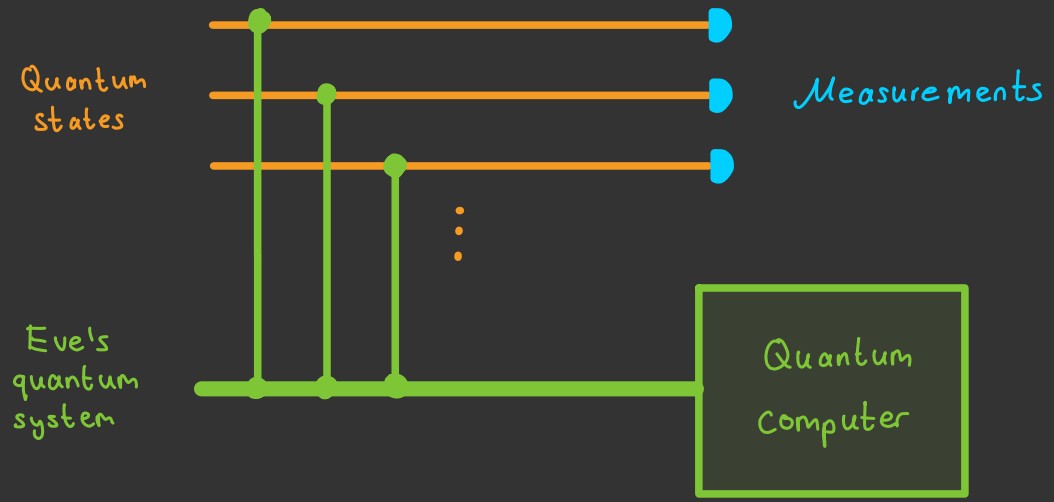
Attacks

2. Collective attacks:



Attacks

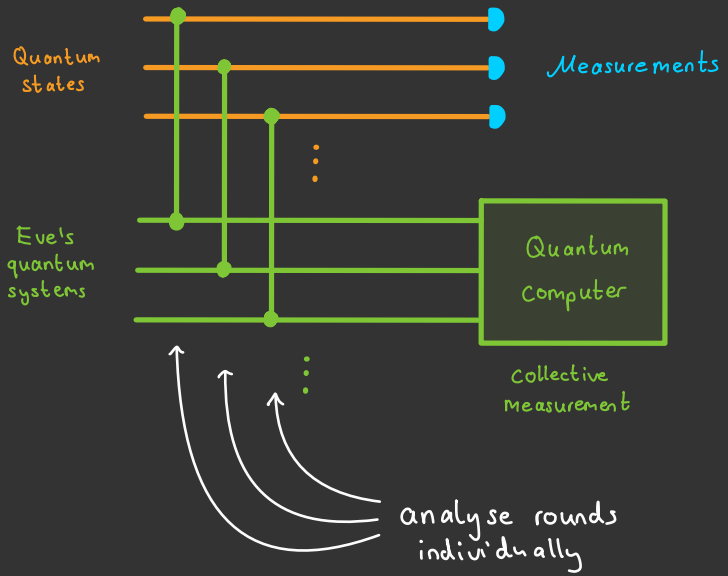
3. Coherent attacks:



Most general attack

Attacks

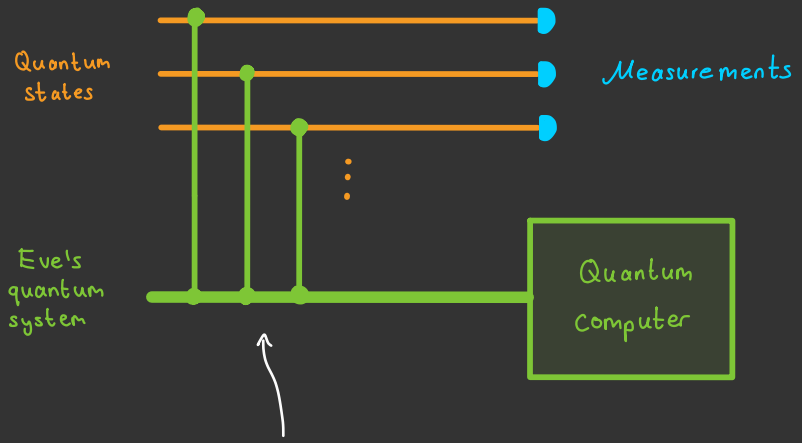
Collective attacks:



$$H_{min}^{\varepsilon}(A_1 \dots A_n | E) \longrightarrow H(A_1 | E_1)$$

(von Neumann entropy)

Coherent attacks:



Have to analyse all of the protocol at once

$$H_{min}^{\varepsilon}(A_1 \dots A_n | E)$$

How to bound H_{min}^{ε} ?

We do not have access to E → Need to bound $H_{min}^{\varepsilon}(A_1 \dots A_n | E)$ based on Alice and Bob's observations

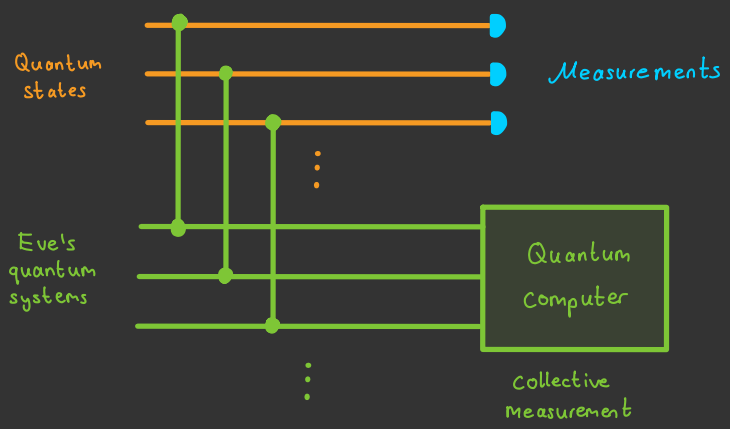
General idea: Show that coherent attacks are <u>not stronger</u> than collective attacks

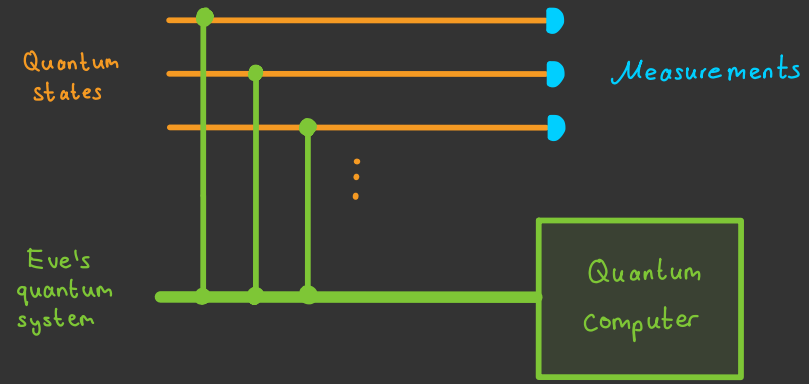
→ bound single-round von Neumann entropy $H(A_1|E)$

we have methods for this

Does not work for all protocols!
Some assumptions have to be fulfilled.

Security proof against collective attacks ⟹ Security proof against coherent attacks

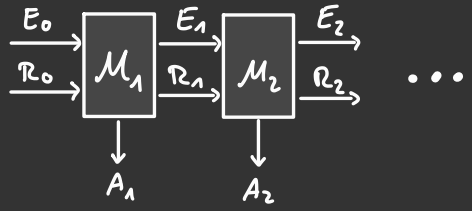




How to bound H_{min}^{ε} ?

1. Quantum de Finetti theorem:
   * Protocol has to be permutation invariant
   * Bound depends on dimension of quantum states

2. Entropy accumulation theorem:
   * Protocol has sequential structure:



   * If information about A_i is supposed to be given to Eve, it has to happen in map $\mathcal{M}_i$;

Almost all known protocols can be analysed with one of these methods

Experiments and Technology

ARTICLE

Received 14 Jan 2011 | Accepted 11 May 2011 | Published 14 Jun 2011

DOI: 10.1038/ncomms1348

Full-field implementation of a perfect eavesdropper on a quantum cryptography system

Ilja Gerhardt[1,*], Qin Liu[2,*], Antía Lamas-Linares[1], Johannes Skaar[2,3], Christian Kurtsiefer[1] & Vadim Makarov[2]

Effects of detector efficiency mismatch on security of quantum cryptosystems

Vadim Makarov, Andrey Anisimov, and Johannes Skaar

Phys. Rev. A **74**, 022313 – Published 17 August 2006; Erratum Phys. Rev. A **78**, 019905 (2008)

Letter | Published: 29 August 2010

Hacking commercial quantum cryptography systems by tailored bright illumination

Lars Lydersen ✉, Carlos Wiechers, Christoffer Wittmann, Dominique Elser, Johannes Skaar & Vadim Makarov

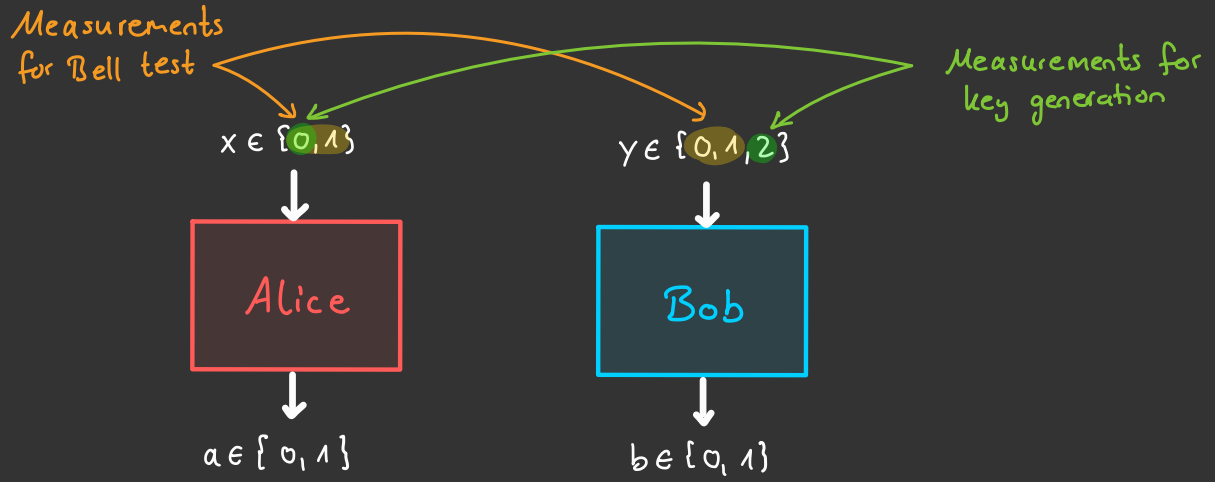
Nature Photonics **4**, 686–689 (2010) | Cite this article

DIQKD with random postselection

DIQKD = Device-Independent QKD ⟶ Quantum devices are treated as black boxes with classical input/output

No assumptions about inner workings

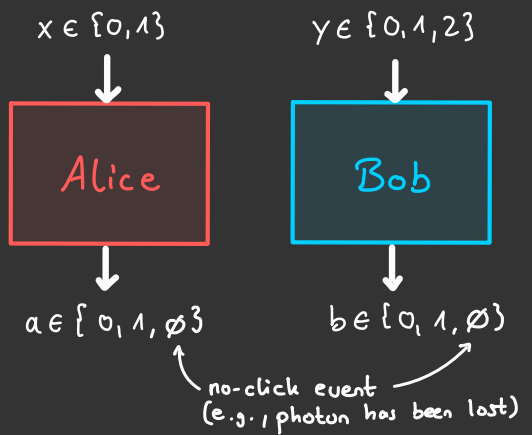
Security is guaranteed by observed Bell violation



DIQKD with random postselection

DIQKD requires a loophole-free Bell test
→ very sensitive to photon loss
→ high requirements on experimental hardware

Random postselection: reduce these requirements by clever classical post-processing



1. Reduce 3-outcome set to 2-outcome set: $\varnothing \longmapsto 1$
   → 1's are now a lot less correlated than 0's

2. With probability p, discard a 1-outcome:
   (Alice & Bob do this independently)

$$0 \longrightarrow 0$$
$$1 \xrightarrow{p} \perp$$
$$\searrow 1$$

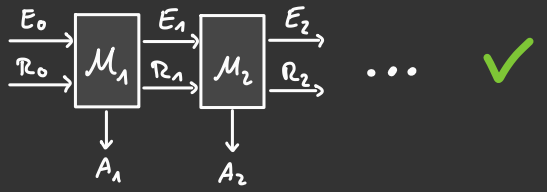
3. Alice and Bob announce publicly which rounds they discard.
   (this will be the problem)

Reminder: How to bound H_{min}^{ε}?

1. Quantum de Finetti theorem:
   * Protocol has to be permutation invariant ✓
   * Bound depends on dimension of quantum states ✗

   In DIQKD, this dimension is generally unbounded

2. Entropy accumulation theorem:
   * Protocol has sequential structure:



   * If information about A_i is supposed to be given to Eve, it has to happen in map $\mathcal{M}_i$ ✗

In DIQKD, we don't know how the quantum devices work
→ outcomes can be correlated
→ the information that round i is discardes can contain information about outcome A_{i-1}

DIQKD with random postselection

Attack: Exploit that we do not assume anything about the quantum devices

Round 1: Device behaves honestly $\longrightarrow$ output $A_1 \in \{0,1\}$

Round 2: If $A_1 = 0$, device behaves honestly again

If $A_1 = 1$, $A_2 = A_1 = 1$. $\xrightarrow{P}$ discarded

↳ Eve knows A_1 and A_2

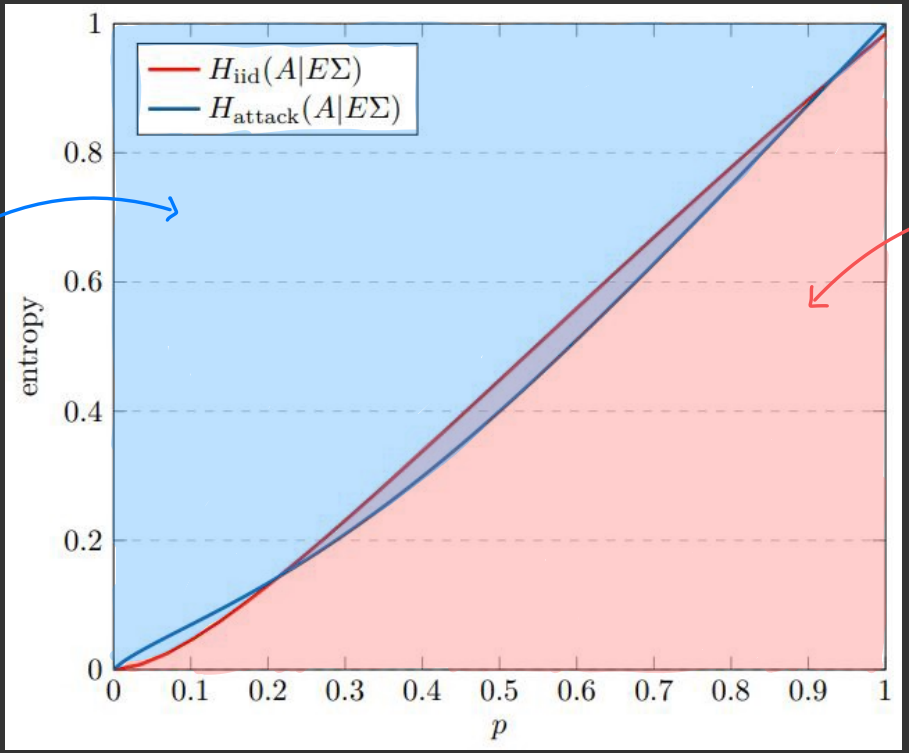
Repeat 1 & 2

It's not a collective attack, because it creates correlation between rounds

But is it stronger than all collective attacks?

DIQKD with random postselection

arXiv: 2306.07364



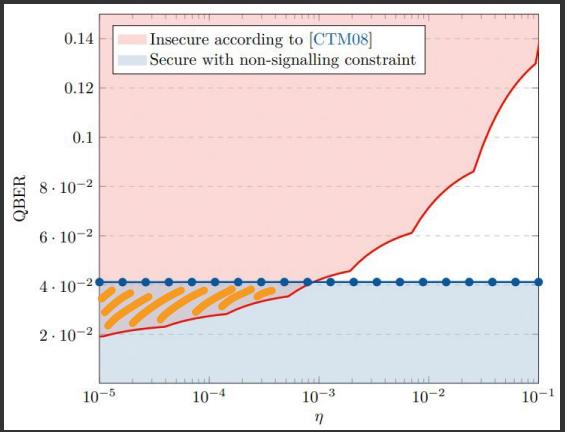
Regime where the protocol is insecure according to our attack

Regime that is secure against collective attacks

Conclusion

* Coherent attacks can be stronger than collective attacks!

* Know another example in device-dependent QKD (where quantum devices are characterized): Differential phase shift QKD (arXiv: 2301.11340)



* We need new security proof techniques that do not reduce coherent to collective attacks!