Title: Quantum speedup in testing causal hypotheses

Date: Apr 09, 2018 11:30 AM

URL: http://pirsa.org/18040122

Abstract: An important ingredient of the scientific method is the ability to test alternative hypotheses on the causal relations relating a given set of variables. In the classical world, this task can be achieved with a variety of statistical, information-theoretic, and computational techniques. In this talk I will address the extension from the classical scenario to the quantum scenario, and, more generally, to general probabilistic theories. After introducing the basic hypothesis testing framework, I will focus on a concrete example, where the task is to identify the causal intermediary of a given variable, under the promise that the causal intermediary belongs to a given set of candidate variables. In this problem, I will show that quantum physics offers an exponential advantage over the best classical strategies, with a doubling of the exponential decay of the error probability. The source of the advantage can be found in the combination of two quantum features: the complementarity between the information on the causal structure and other properties of the cause effect relation, and the ability to perform multiple tests in a quantum superposition. An interesting possibility is that one of the "hidden principles" of quantum theory could be on our ability to test alternative causal hypotheses.

QUANTUM SPEEDUP IN TESTING CAUSAL HYPOTHESES

Giulio Chiribella
Department of Computer Science, University of Oxford,
CIFAR-Azrieli Global Scholars Program
joint work with
Daniel Ebler
Department of Computer Science, The University of Hong Kong

Algorithmic Information, Induction, and Observers in Physics, April 9-13 2018
Perimeter Institute

CIFAR CANADIAN INSTITUTE FOR ADVANCED RESEARCH

JTF

RGC

Croucher Foundation
裘槎基金會

National Natural Science Foundation of China
NSFC

FQXi FOUNDATIONAL QUESTIONS INSTITUTE

CAUSAL INFERENCE (CLASSICAL)

The problem: discovering causal relations among a set of variables
(cf. Pearl, Spirtes-Glymour-Scheines)

Basic idea: A is *a cause* for B iff
intervening on A *has an effect on the statistics of* B

Caveat: "correlation does not imply causation":

no way to infer a causal relation
from a *single* probability distribution $p(a,b)$.
It is necessary to probe different settings for a

CAUSAL INFERENCE (GENERAL)

Recently, various extensions of the notions of
"causal relation" and "causal network"
to quantum theory and beyond.

Basic idea (modulo variations across frameworks):
Variables: physical systems.
Causal relations: variable A is *a cause* for variable B iff
changing the state of A *induces a change of the state of* B

Leifer (2006), GC-D'Ariano-Perinotti (2008),
Coecke-Spekkens (2012), Leifer-Spekkens (2013),
Henson-Lal-Pusey (2014), Pienaar-Brukner (2015), Costa-Shrapnel (2016),
Portmann-Matt-Maurer-Renner-Tackmann (2017),
Allen-Barrett-Horsman-Lee-Spekkens (2017), MacLean-Ried-Spekkens-Resch (2017).

MOTIVATIONS FOR QUANTUM EXTENSION

- **Foundational:**
  -understanding interplay between causality and quantum features
  -gain insights into future theories that will combine both
  -find new principles for quantum theory

- **Practical:**
  -identifying new quantum advantages
  -identifying working principles for new quantum devices,
  develop a "technology" of quantum causality.

PLAN OF THIS TALK

Formulate and analyze the quantum version of the task of **testing causal hypotheses.**

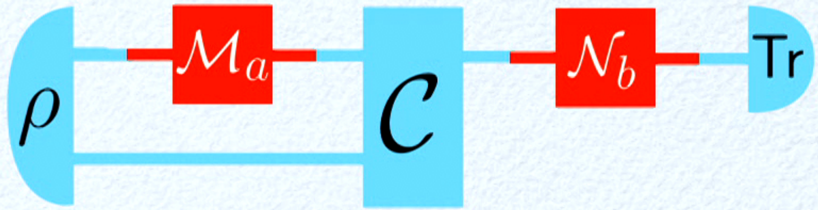
In this task, one has a **set of candidate hypotheses** on the causal relations occurring in a process and the goal is to **identify the correct hypothesis.**

PROLOGUE

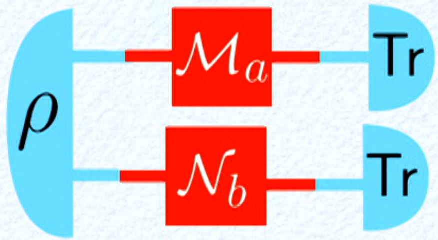
AN INTRIGUING EXAMPLE

Task: distinguish between

- Situation (1): A causes B



- Situation (2): A and B have a common cause



Fact: *for some specific* ρ and $\mathcal{C}$ it is possible to distinguish between (1) and (2) using *only projective measurements.*

Fitzsimons, Jones, and Vedral, Scientific Reports 5, 18281 (2015).
Ried, Agnew, Vermeyden, Janzing, Spekkens, and Resch, Nature Physics 11, 414 (2015).

QUESTION

This is an intriguing and stimulating observation.
Still, the type of advantage here
is contingent on a restriction on the allowed measurements,
which in classical theory is equivalent to a restriction to
passive observational strategies, where ***no intervention*** is allowed.

Question:
Can we find advantages in the situation where
arbitrary interventions are allowed?

QUESTION

This is an intriguing and stimulating observation.
Still, the type of advantage here
is contingent on a restriction on the allowed measurements,
which in classical theory is equivalent to a restriction to
passive observational strategies, where ***no intervention*** is allowed.

Question:
Can we find advantages in the situation where
arbitrary interventions are allowed?

TESTING
CAUSAL HYPOTHESES:

A THEORY-INDEPENDENT FRAMEWORK

CAUSAL DISCOVERY
VS CAUSAL HYPOTHESIS TESTING

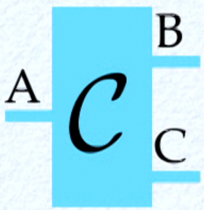
Causal discovery. *Input:* variables A, B, C, …
Output: the causal relations among them.

Causal hypothesis testing: *Input:* variables A, B, C, …
and a set of hypotheses on the causal relations among them.
Output: the correct hypothesis

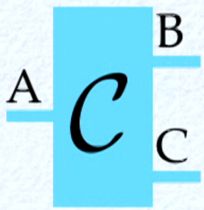
CAUSAL HYPOTHESES

Causal Hypothesis: an hypothesis **on the causal structure** of the process connecting the variables.

e.g.



(H1) A causes B but not C



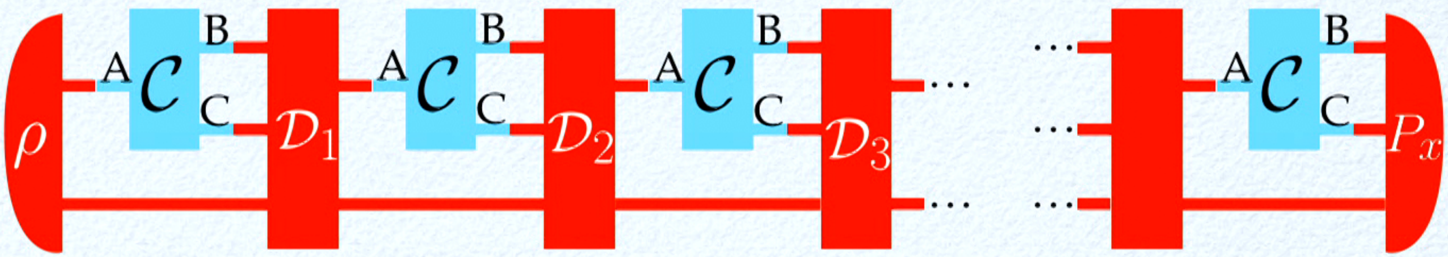
(H2) A causes B but not C

NB: causal hypotheses can be formulated **independently of the underlying theory.**

TESTING CAUSAL HYPOTHESES

The experimenter can probe the **same process** for a **finite number of times**, performing **arbitrary interventions.**

Most general intervention:



$x=$ guess for the correct hypothesis

Special cases: process tomography, parallel queries, etc…

DISCRIMINATION RATE

Goal of causal hypothesis testing:
minimize the probability of choosing the wrong hypothesis.

Worst-case approach: since the process $\mathcal{C}$ is **unknown**
(a part from the fact that it is compatible with one and only one of the given hypotheses)
we will consider the
worst-case error probability $p_{\mathrm{err}}(N)$

Discrimination rate: $$R = \lim_{N\to\infty} \frac{-\log p_{\mathrm{err}}(N)}{N}$$

quantifies the distinguishability of the hypotheses

DISCRIMINATION RATE

Goal of causal hypothesis testing:
minimize the probability of choosing the wrong hypothesis.

Worst-case approach: since the process $\mathcal{C}$ is **unknown**
(a part from the fact that it is compatible with one and only one of the given hypotheses)
we will consider the
worst-case error probability $p_{\rm err}(N)$

Discrimination rate: $$R = \lim_{N\to\infty} \frac{-\log p_{\rm err}(N)}{N}$$
quantifies the distinguishability of the hypotheses

EXAMPLE:

IDENTIFYING
THE CAUSAL INTERMEDIARY

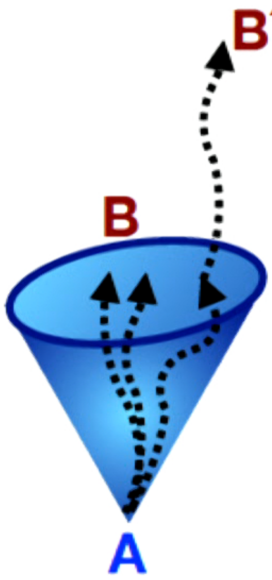
CAUSAL INTERMEDIARIES

Variable B is a **(complete) causal intermediary** for variable A, if “all the causal influences of A” propagate through B.

More formally:
Variable B is a *causal intermediary* for variable A if

- B is an effect of A
- every other effect of A, say B′, is an effect of B (assuming that B′ takes place after B)

Example: A localized at a spacetime point and B localized at a section of the forward lightcone based at A.



IDENTIFYING THE CAUSAL INTERMEDIARY

Variables: A, B, and C

Hypothesis (1): B is a causal intermediary of A,
while C fluctuates uniformly at random.

Hypothesis (2): C is a causal intermediary of A,
while B fluctuates uniformly at random.

Problem: decide which hypothesis is correct.

IDENTIFYING THE CAUSAL INTERMEDIARY

Variables: A, B, and C

Hypothesis (1): B is a causal intermediary of A,
while C fluctuates uniformly at random.

Hypothesis (2): C is a causal intermediary of A,
while B fluctuates uniformly at random.

Problem: decide which hypothesis is correct.

CLASSICAL SOLUTION

SETTINGS

Assume that the random variables A, B, and C have all the **same dimension d.**

With this assumption, Hypotheses (1) and (2) become:

Hypothesis (1): b is a permutation of a,
and c is uniformly random

Hypothesis (2): c is a permutation of a,
and b is uniformly random

NAIVE CLASSICAL STRATEGY

Initialize the input variable A to a certain value,
and observe the values taken by the output variables B and C.
Repeat for N times, possibly trying different values of A.

NAIVE CLASSICAL STRATEGY

Initialize the input variable A to a certain value,
and observe the values taken by the output variables B and C.
Repeat for N times, possibly trying different values of A.

Example for N=8, d=2

	1	2	3	4	5	6	7	8
A	0	0	1	1	0	0	0	1
B	1	1	0	0	1	1	1	0
C	0	0	1	1	1	0	0	1

PROBABILITY OF ERROR (NAIVE STRATEGY)

Error occurs when both variables B and C
take values that are compatible with permutations.

In that unlucky case, the probability of error is 1/2.

PROBABILITY OF ERROR (NAIVE STRATEGY)

Error occurs when both variables B and C
take values that are compatible with permutations.

In that unlucky case, the probability of error is 1/2.

If we try v different values for A,
the probability to be unlucky is

$$p_{\text{unlucky}} = \frac{\left|\{\text{injective functions from } v \text{ element set to } d \text{ element set}\}\right|}{d^N}$$

$$= \frac{d(d-1)(d-2)\cdots(d-v+1)}{d^N}$$

DISCRIMINATION RATE (NAIVE STRATEGY)

Choosing v=1, the errpr probability of the naive strategy is minimal:

$$p_{\text{err}}(N) = \frac{p_{\text{unlucky}}}{2} = \frac{1}{2d^{N-1}}$$

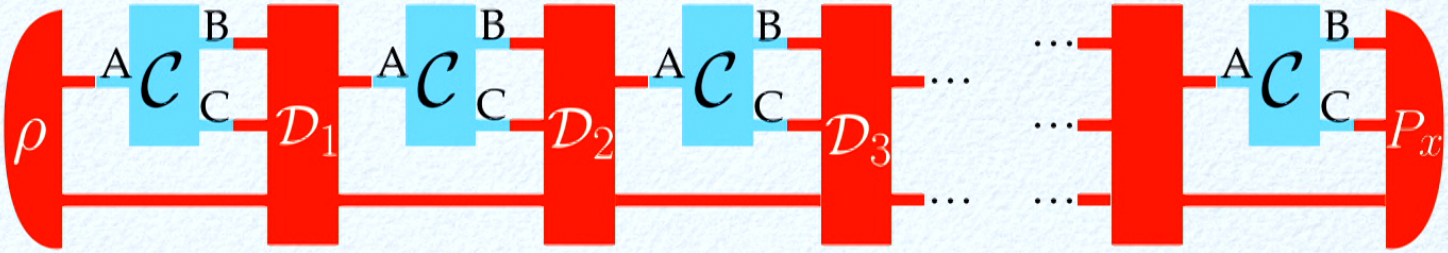
Discrimination rate:

$$R = \lim_{N\to\infty} \frac{-\log p_{\text{err}}(N)}{N}$$

$$= \log d$$

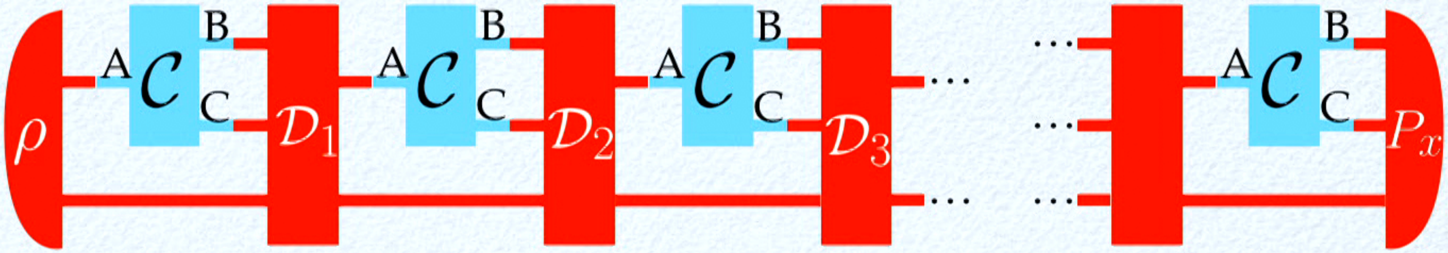
GENERAL CLASSICAL STRATEGIES?

We have found the rate of the naive classical strategy.
What about general strategies?



GENERAL CLASSICAL STRATEGIES?

We have found the rate of the naive classical strategy.
What about general strategies?



Theorem [Hayashi, IEEE TIT, 55, 3807 (2009)]:
The optimal *asymptotic rate* in distinguishing *two classical channels* can be attained by a parallel strategy.

Applying this theorem to a fixed pair of permutations, we obtain that $\log d$ is an **upper bound to the rate.**

IN SUMMARY

For classical variables of dimension *d*,
the optimal rate in identifying a complete causal intermediary is

$$R_{\mathrm{C}} = \log d$$

Attained by the naive strategy
"initialize variable A for N times to the same value"

QUANTUM SOLUTION

SETTINGS

Assume that the quantum systems A, B, and C have all the **same dimension *d*.**

With this assumption, Hypotheses (1) and (2) become:

Hypothesis (1): $$\mathcal{C}_{A\to BC}(\rho_A) = (U\rho U^\dagger)_B \otimes \left(\frac{I}{d}\right)_C$$

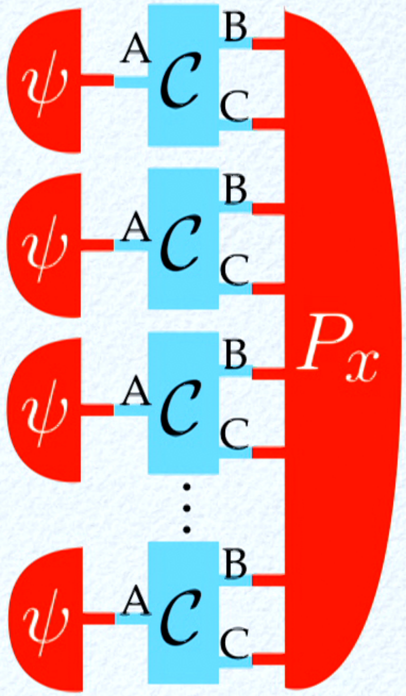
for some unknown unitary U

Hypothesis (2): $$\mathcal{C}_{A\to BC}(\rho_A) = \left(\frac{I}{d}\right)_B \otimes (V\rho V^\dagger)_C$$

for some unknown unitary V

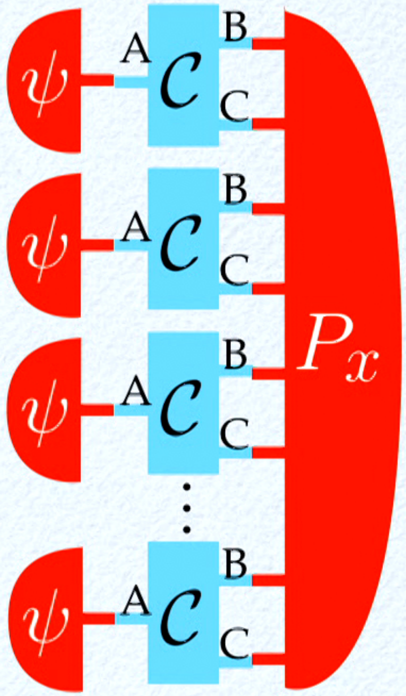
NAIVE QUANTUM STRATEGY

Initialize the input system A
in a fixed state,
repeat for N times,
measure the output state.



NAIVE QUANTUM STRATEGY

Initialize the input system A
in a fixed state,
repeat for N times,
measure the output state.



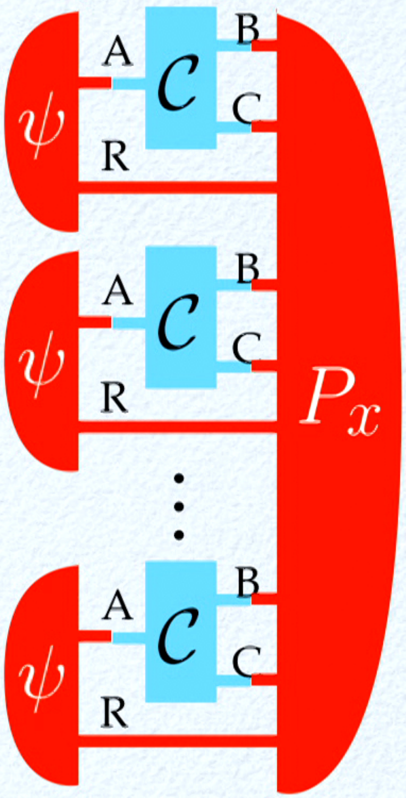
Error probability:

$$p_{\text{err}}(N) = \frac{\binom{N+d-1}{d-1}}{2d^N}$$

Worse than the classical error probability.
But at least, same rate: log *d*

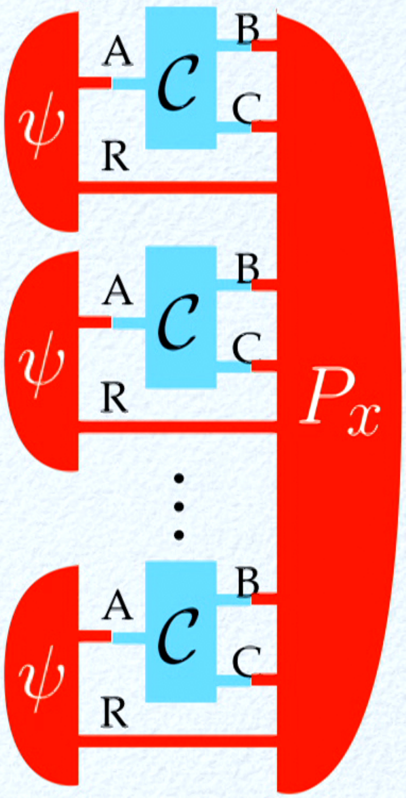
QUANTUM TOMOGRAPHY?

Initialize the input system A
together with a reference system R
in a fixed state,
repeat for N times,
measure the output state.



QUANTUM TOMOGRAPHY?

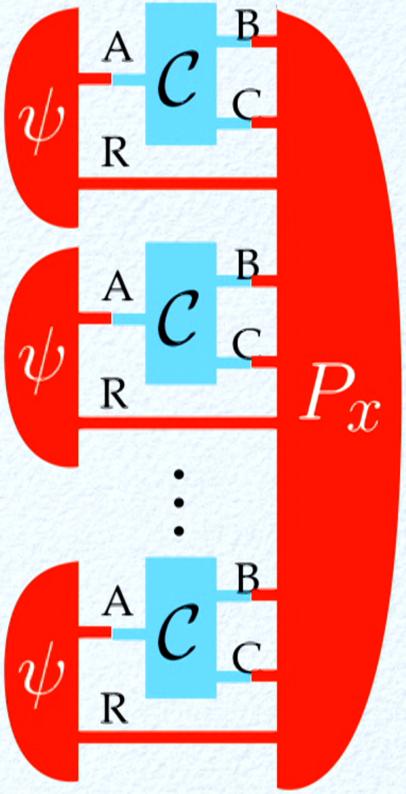
Initialize the input system A
together with a reference system R
in a fixed state,
repeat for N times,
measure the output state.



Rate
same as the classical rate: **log *d***

QUANTUM TOMOGRAPHY?

Initialize the input system A
together with a reference system R
in a fixed state,
repeat for N times,
measure the output state.

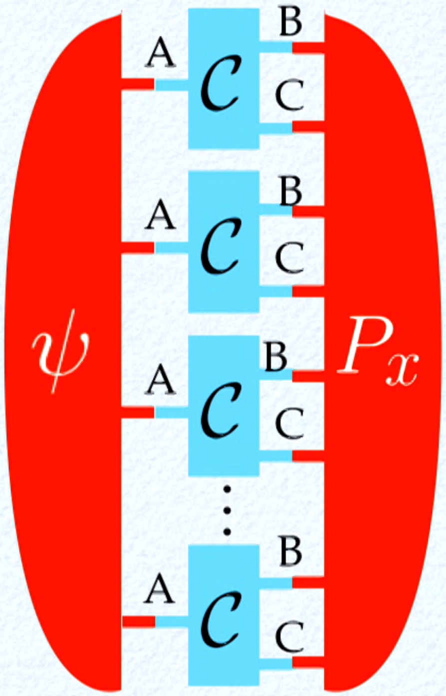


OPTIMAL PARALLEL STRATEGIES

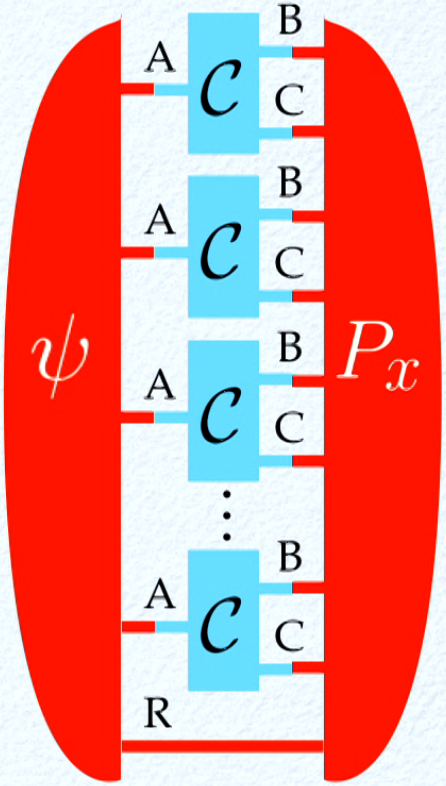
OPTIMAL
PARALLEL STRATEGIES

PARALLEL STRATEGIES

Without reference:



With reference:



OPTIMAL STRATEGY WITHOUT REFERENCE

For simplicity, assume $d = 2$ and N even, say N=2p.

Divide the N input variables in p pairs.
Prepare each group in the singlet state $|\Psi^-\rangle = \dfrac{|0\rangle \otimes |1\rangle - |1\rangle \otimes |0\rangle}{\sqrt{2}}$

OPTIMAL STRATEGY WITHOUT REFERENCE

For simplicity, assume $d = 2$ and N even, say N=2p.

Divide the N input variables in p pairs.
Prepare each group in the singlet state $|\Psi^-\rangle = \frac{|0\rangle \otimes |1\rangle - |1\rangle \otimes |0\rangle}{\sqrt{2}}$

Key intuition: invariance of the singlet

$$(U \otimes U)|\Psi^-\rangle = |\Psi^-\rangle \qquad \forall U$$

we can test the causal structure without extracting any information about the functional dependence between cause and effect.

ERROR PROBABILITY

For general dimension d,
divide the N input variables in groups of d
and prepare each group in the SU(d) singlet

$$|S_d\rangle = \frac{1}{\sqrt{d!}} \sum_{k_1, k_2, \cdots, k_d} \epsilon_{k_1 k_2 \dots k_d} |k_1\rangle |k_2\rangle \cdots |k_d\rangle$$

Perform the Helstrom measurement on the output.

ERROR PROBABILITY

For general dimension d,
divide the N input variables in groups of d
and prepare each group in the SU(d) singlet

$$|S_d\rangle = \frac{1}{\sqrt{d!}} \sum_{k_1, k_2, \cdots, k_d} \epsilon_{k_1 k_2 \dots k_d} |k_1\rangle |k_2\rangle \cdots |k_d\rangle$$

Perform the Helstrom measurement on the output.

Error probability: $p_{\mathrm{err}}(N) = \frac{1}{2d^N}$

Better than classical value $p_{\mathrm{err}}(N) = \frac{1}{2d^{N-1}}$

but rate is still log *d*

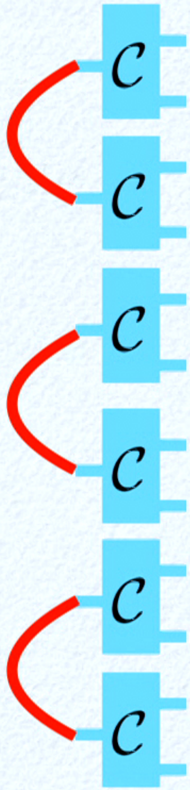
OPTIMAL
PARALLEL STRATEGIES
WITH
REFERENCE

EQUIVALENT STRATEGIES

d=2, N even. Many ways to partition the inputs into pairs:

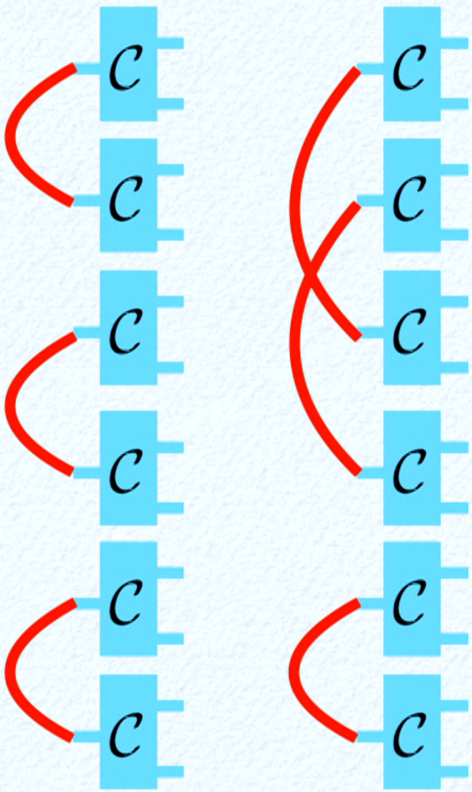
EQUIVALENT STRATEGIES

d=2, N even. Many ways to partition the inputs into pairs:



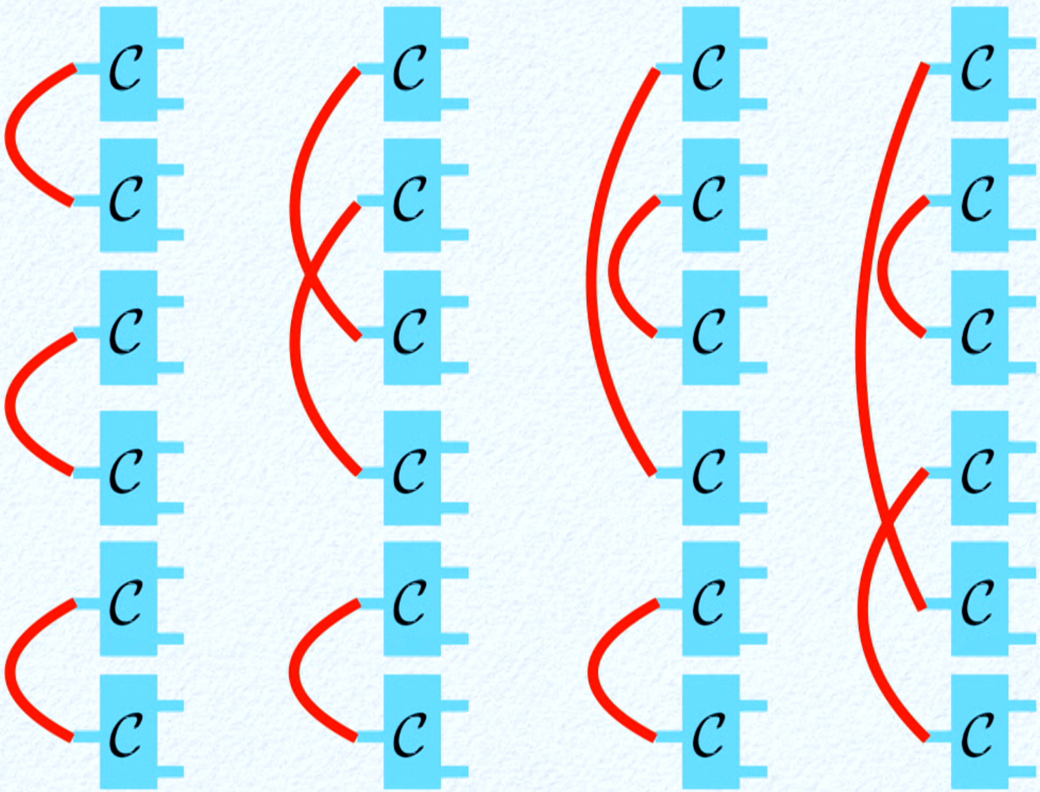
EQUIVALENT STRATEGIES

d=2, N even. Many ways to partition the inputs into pairs:



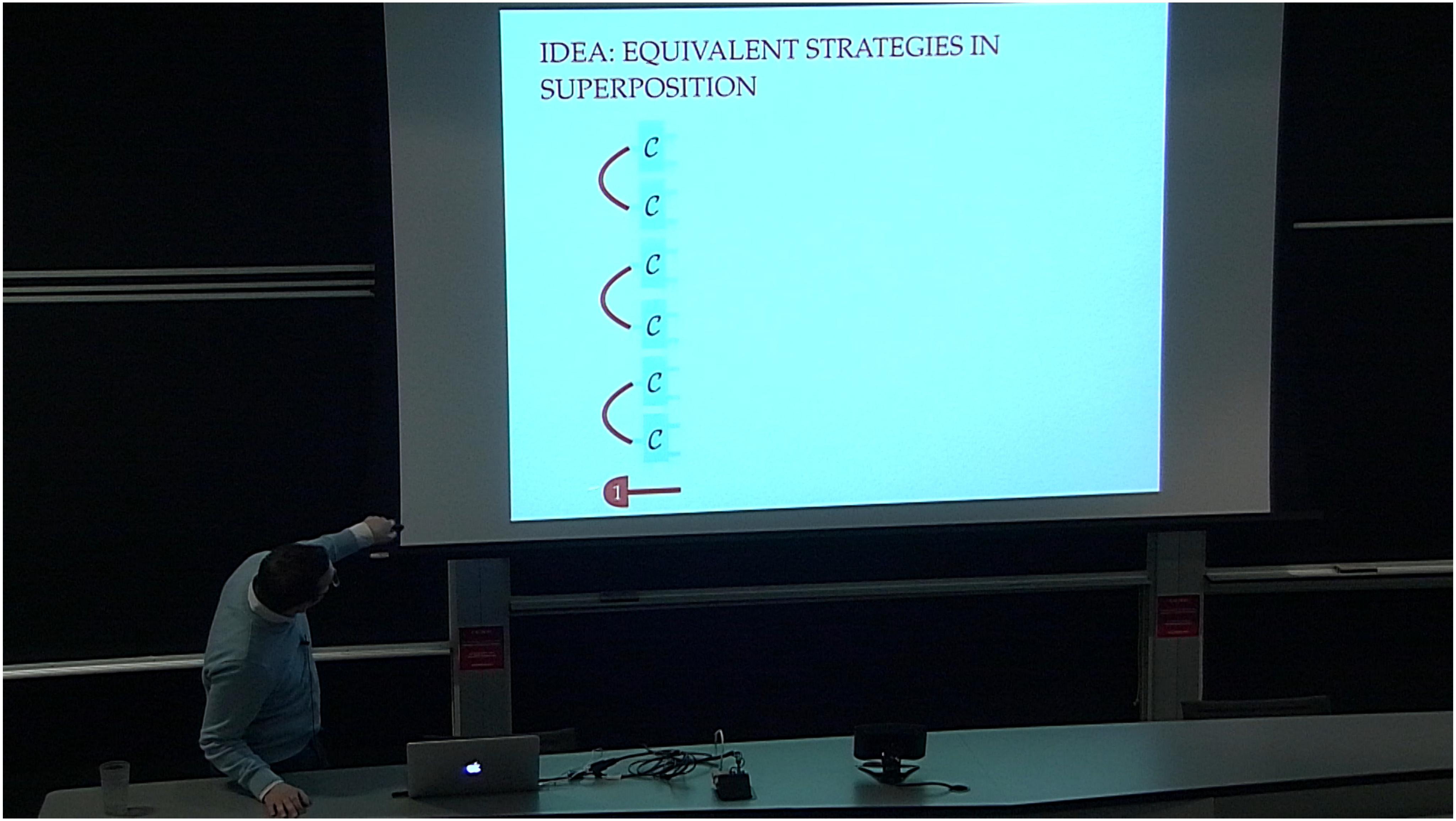
EQUIVALENT STRATEGIES

d=2, N even. Many ways to partition the inputs into pairs:

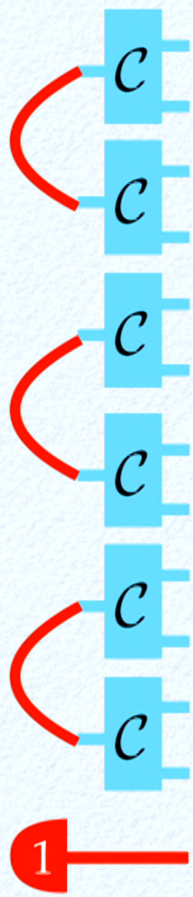


etc,
etc…

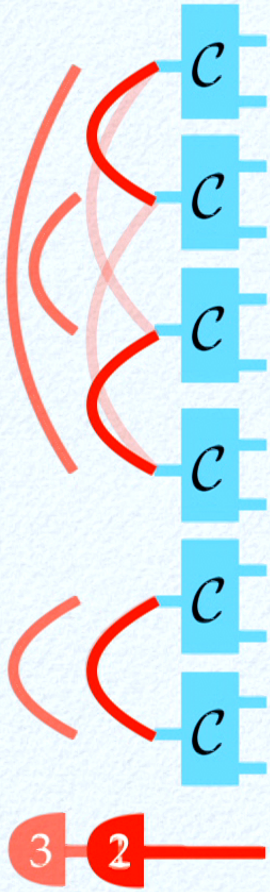
IDEA: EQUIVALENT STRATEGIES IN SUPERPOSITION


IDEA: EQUIVALENT STRATEGIES IN SUPERPOSITION
C
C
C
C
C
C
1

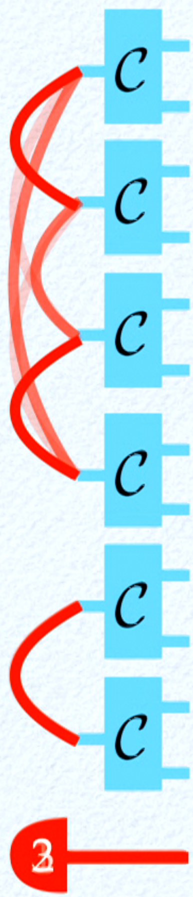
IDEA: EQUIVALENT STRATEGIES IN SUPERPOSITION



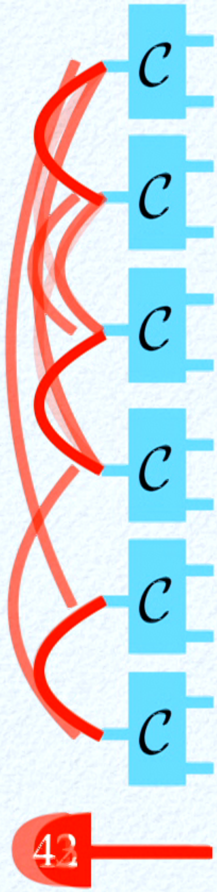
IDEA: EQUIVALENT STRATEGIES IN SUPERPOSITION



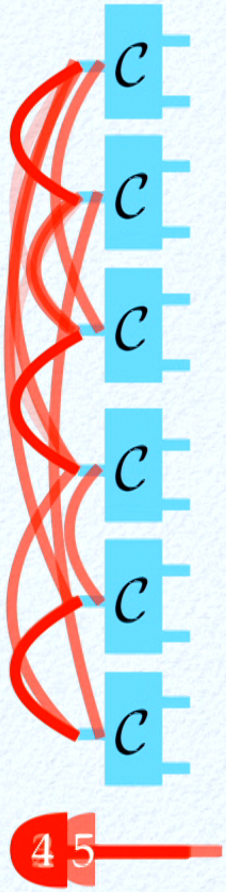
IDEA: EQUIVALENT STRATEGIES IN SUPERPOSITION



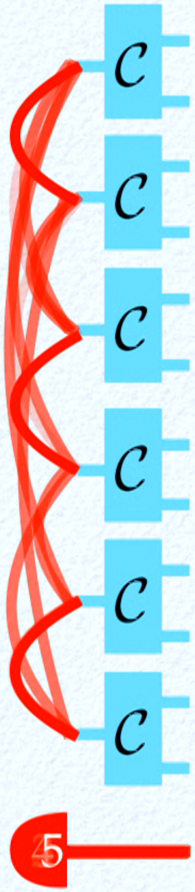
IDEA: EQUIVALENT STRATEGIES IN SUPERPOSITION



IDEA: EQUIVALENT STRATEGIES IN SUPERPOSITION



IDEA: EQUIVALENT STRATEGIES IN SUPERPOSITION



In dimension d:

$$|\Psi\rangle = \frac{1}{\sqrt{L}} \sum_{i=1}^{L} \left(|S_d\rangle^{\otimes N/d} \right)_i \otimes |i\rangle_R$$

where i labels the way to group the systems and L is the number of groupings in the superposition

ERROR PROBABILITY

When there are r linearly independent groupings, the error probability is

$$p_{\mathrm{err}}^{\mathrm{Q}}(r) = \frac{r}{2d^N}\left(1 - \sqrt{1 - r^{-2}}\right) \xrightarrow{r \gg 1} \frac{1}{4rd^N}$$

ERROR PROBABILITY

When there are *r* linearly independent groupings, the error probability is

$$p_{\mathrm{err}}^{\mathrm{Q}}(r) = \frac{r}{2d^N}\left(1 - \sqrt{1 - r^{-2}}\right) \xrightarrow{r \gg 1} \frac{1}{4rd^N}$$

Picking the maximum *r*, we obtain the rate

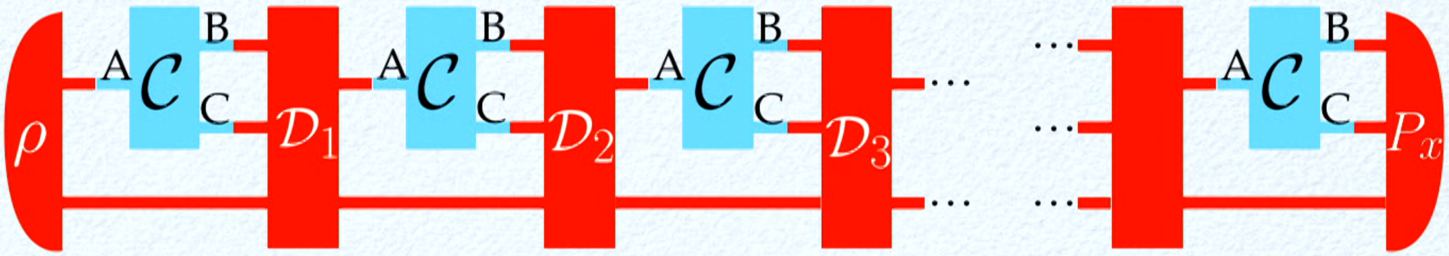
$$R_{\mathrm{Q}} = -\lim_{N\to\infty} \frac{\log p_{\mathrm{err}}^{\mathrm{Q}}}{N} = 2\log d$$

Twice the classical rate!

GENERAL
QUANTUM STRATEGIES

GENERAL CLASSICAL STRATEGIES?

We have found the rate of the best parallel strategies.
What about general strategies?

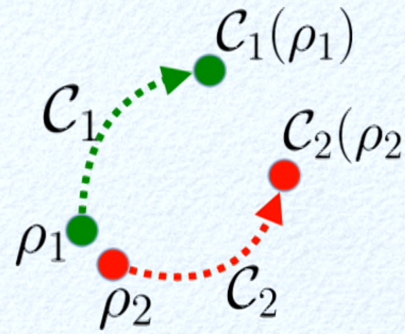


In principle, we should optimize over all **quantum testers**

GC-D'Ariano-Perinotti, PRL 101, 180501 (2008)
Gutoski-Watrous, Proc. STOC, p. 565-574 (2007).

However, the optimization is hard.

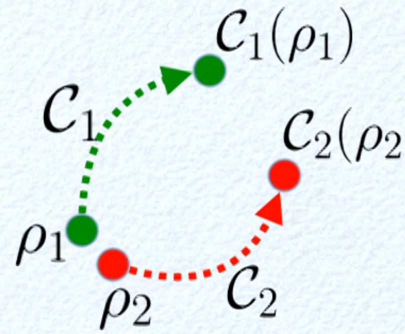
A TRICK



Define the **fidelity divergence** of two channels

$$\partial F(\mathcal{C}_1, \mathcal{C}_2) = \inf_R \inf_{\rho_1, \rho_2} \frac{F\Big[(\mathcal{C}_1 \otimes \mathcal{I}_R)(\rho_1)\,,(\mathcal{C}_2 \otimes \mathcal{I}_R)(\rho_2)\Big]}{F(\rho_1, \rho_2)}$$

A TRICK



Define the **fidelity divergence** of two channels

$$\partial F(\mathcal{C}_1, \mathcal{C}_2) = \inf_R \inf_{\rho_1, \rho_2} \frac{F\Big[(\mathcal{C}_1 \otimes \mathcal{I}_R)(\rho_1)\,, (\mathcal{C}_2 \otimes \mathcal{I}_R)(\rho_2)\Big]}{F(\rho_1, \rho_2)}$$

Fact: if we try to distinguish between two channels with N queries, the error probability satisfies

$$p^{\mathrm{seq}}_{\mathrm{err}}(\mathcal{C}_1, \mathcal{C}_2; N) \geq \frac{\partial F(\mathcal{C}_1, \mathcal{C}_2)^N}{4}$$

Upper bound on the rate: $R^{\mathrm{seq}}_{\mathrm{Q}}(\mathcal{C}_1, \mathcal{C}_2) \leq -\log \partial F(\mathcal{C}_1, \mathcal{C}_2)$

OPTIMAL RATE

The **fidelity divergence** between the channels

$$\mathcal{C}_{A\to BC}(\rho_A) = (U\rho U^\dagger)_B \otimes \left(\frac{I}{d}\right)_C$$

and

$$\mathcal{C}_{A\to BC}(\rho_A) = \left(\frac{I}{d}\right)_B \otimes (V\rho V^\dagger)_C$$

is $\partial F = \dfrac{1}{d^2}$

Hence, we have the bound $R_{\mathrm{Q}} \leq 2\log d$

IN SUMMARY

For quantum variables of dimension *d*,
the rate

$$R_{\mathrm{Q}} = 2 \log d$$

is optimal, and it is attained by preparing
singlets in a superposition of different groupings.

EXTENSION
TO
K CAUSAL HYPOTHESES

CAUSAL INTERMEDIARY: *K* CANDIDATES

Variables: A, B_1, B_2 … B_k

Hypothesis (i): B_i is a causal intermediary of A,
i=1, …, *k* and all the other variables fluctuate uniformly at random.

Problem: decide which hypothesis is correct.

OPTIMAL RATES

Classical: log d

Quantum without reference: log d
(attained with singlets)

Quantum with reference: 2 log d
(attained with superposition of singlets, optimal among all quantum strategies)

Note: rates are independent of the number of hypotheses k

CONCLUSIONS

- Theory-independent framework for **testing causal hypotheses**
- Instance of the problem: **identifying the causal intermediary.**
- Classical solution: rate log *d*
- Quantum solution: rate 2 log *d*, achieved by **superposition of singlet states in equivalent configurations**

SUMMARY AND OUTLOOK

CONCLUSIONS

- Theory-independent framework for **testing causal hypotheses**
- Instance of the problem: **identifying the causal intermediary.**
- Classical solution: rate log d
- Quantum solution: rate 2 log d, achieved by **superposition of singlet states in equivalent configurations**

OUTLOOK

- Is it always true that quantum theory does better (or at least, not worse) than classical theory in the task of causal hypothesis testing?
- Is quantum theory optimal for causal hypothesis testing?
- If not, which physical principles determine the power in identifying causal hypotheses?
- What about indefinite causal order?
  How well can we test non-standard hypotheses on the causal structure?