# Supplementary Information for Quantum Causal Inference with Extremely Light Touch


Xiangjing Liu[1,2,3,4,5], Yixian Qiu[5], Oscar Dahlsten[3,6,7] and Vlatko Vedral[8]

[1] *CNRS@CREATE, 1 Create Way, 08-01 Create Tower, Singapore 138602, Singapore*
[2] *MajuLab, CNRS-UCA-SU-NUS-NTU International Joint Research Unit, Singapore*
[3] *Department of Physics, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong*
[4] *Department of Physics, Southern University of Science and Technology, Shenzhen 518055, China*
[5] *Centre for Quantum Technologies, National University of Singapore, Singapore 117543, Singapore*
[6] *Shenzhen Institute for Quantum Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China*
[7] *Institute of Nanoscience and Applications, Southern University of Science and Technology, Shenzhen 518055, China*
[8] *Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, United Kingdom*


## I. A METHOD FOR IMPLEMENTING THE COARSE-GRAINED MEASUREMENT

We now describe a method for implementing the coarse grained measurements $P_+$ and $P_-$ associated with the closed form expression of the PDM. As described in the Methods, for a given time point

$$\Big\{P_+ := \frac{\mathbb{1}\otimes\mathbb{1}+\sigma_i\otimes\sigma_j}{2}, P_- := \frac{\mathbb{1}\otimes\mathbb{1}-\sigma_i\otimes\sigma_j}{2}\Big\}.$$

One possible way of obtaining the statistics by coarse-grained measurements is via a quantum scattering circuit [56]. As depicted in FIG. 3, the scattering circuit takes the input state as $\rho = \rho_{\text{ancilla}} \otimes \rho_s = |0\rangle\langle 0| \otimes |\psi\rangle\langle\psi|$. One can calculate that the expectation value $\langle\sigma_3\rangle$ on the ancilla equals to the two-point correlation function of the system:

$$\langle\sigma_3\rangle = \langle\tilde{\sigma}_j(t_1), \tilde{\sigma}_i(t_0)\rangle, \tag{21}$$

where $\tilde{\sigma}_i(t_0) := U_0\tilde{\sigma}_iU_0^\dagger := e^{-iht_0}\tilde{\sigma}_ie^{iht_0}$ and $\tilde{\sigma}_j(t_1) := U_1\tilde{\sigma}_jU_1^\dagger := e^{-iht_1}\tilde{\sigma}_je^{iht_1}$. Eq. (21) can be rewritten as

$$\begin{aligned}\langle\tilde{\sigma}_j(t_1), \tilde{\sigma}_i(t_0)\rangle =&\langle U_1\left(P_+^j - P_-^j\right)U_1^\dagger U_0\left(P_+^i - P_-^i\right)U_0^\dagger\rangle\\ =&\Big\{\langle U_1P_+^j\,U_1^\dagger U_0P_+^i\,U_0^\dagger\rangle + \langle U_1P_-^j\,U_1^\dagger U_0P_-^i\,U_0^\dagger\rangle\Big\}\\ &-\Big\{\langle U_1P_+^j\,U_1^\dagger U_0P_-^i\,U_0^\dagger\rangle + \langle U_1P_+^j\,U_1^\dagger U_0P_-^i\,U_0^\dagger\rangle\Big\}.\end{aligned} \tag{22}$$

Thus, Eq. (21) gives the same value as the expectation value obtained by coarse-grained measurements $P_+^i$ and $P_-^i$ taking place initially and finally.

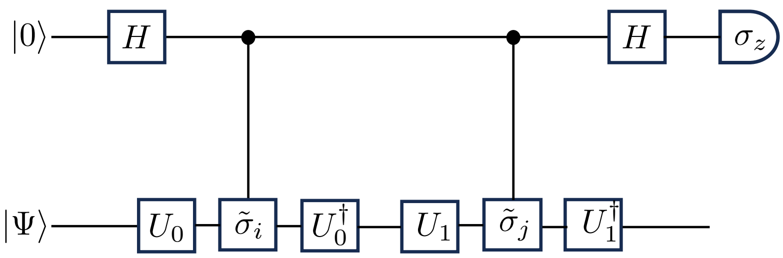


FIG. 3. A quantum circuit for measuring time correlation function $\langle\tilde{\sigma}_j(t_1)\tilde{\sigma}_i(t_0)\rangle$. $\sigma_3 = \sigma_z$ and $\tilde{\sigma}_i$ is the $i$-th Pauli group element on $n$ qubits.

## II. DERIVATION OF THE CLOSED-FORM OF 2-TIME $n$-QUBIT PDM

Having the quantum state $\rho_1$ at time $t_1$, the measurement $\{P_+^\alpha, P_-^\alpha\}$ at each time $t_\alpha$, $\alpha = 1, 2$, and the quantum channel $\mathcal{M}_{2|1}$ with its CJ matrix $M_{12}$, we are in a good position to obtain the 2-time $n$-qubit PDM. Recall that the general form of an $n$-qubit PDM across two times $[t_1, t_2]$ is given by

$$R_{12} = \frac{1}{2^{2n}}\sum_{i_1=0}^{4^n-1}\sum_{i_2=0}^{4^n-1}\langle\tilde{\sigma}_{i_1}, \tilde{\sigma}_{i_2}\rangle\,\tilde{\sigma}_{i_1}\otimes\tilde{\sigma}_{i_2}. \tag{23}$$

Therefore, in order to construct the PDM, it is necessary to calculate the expectation values of the product of the result of the measurements, i.e., $\langle\tilde{\sigma}_{i_1},\tilde{\sigma}_{i_2}\rangle$. According to the definition of the PDM, the initial state collapse to the eigenstates of the $n$-qubit Pauli matrix $\tilde{\sigma}_{i_1}$ after the measurement at time $t_1$. Then the post-measurement states go through the channel followed by a measurement at time $t_2$. Denoting $\{P_+^\alpha = (\mathbb{1}+\tilde{\sigma}_{i_\alpha})/2, P_-^\alpha = (\mathbb{1}-\tilde{\sigma}_{i_\alpha})/2\}$ as the coarse-grained measurement scheme at time $t_\alpha, \alpha = 1,2$, the product of expectation values can be expressed by

$$\langle\tilde{\sigma}_{i_1},\tilde{\sigma}_{i_2}\rangle = \mathrm{Tr}\big[M_{12}(P_+^1\rho_1 P_+^1 \otimes \tilde{\sigma}_{i_2})\big] - \mathrm{Tr}\big[M_{12}(P_-^1\rho_1 P_-^1 \otimes \tilde{\sigma}_{i_2})\big], \tag{24}$$

The equality, $\mathcal{M}_{2|1}(\rho_1) = \mathrm{Tr}_1[M_{12}\rho_1 \otimes \mathbb{1}_2]$, which holds for our choice of convention for the CJ matrix, will be frequently used in the calculations. It is straightforward to calculate that

$$P_+^1\rho_1 P_+^1 - P_-^1\rho_1 P_-^1 = \frac{\mathbb{1}+\tilde{\sigma}_{i_1}}{2}\rho_1\frac{\mathbb{1}+\tilde{\sigma}_{i_1}}{2} - \frac{\mathbb{1}-\tilde{\sigma}_{i_1}}{2}\rho_1\frac{\mathbb{1}-\tilde{\sigma}_{i_1}}{2} = \frac{1}{2}(\rho_1\tilde{\sigma}_{i_1} + \tilde{\sigma}_{i_1}\rho_1). \tag{25}$$

Substituting Eq (25) into Eq (24) leads to

$$\begin{aligned}\langle\tilde{\sigma}_{i_1},\tilde{\sigma}_{i_2}\rangle &= \mathrm{Tr}\left[\frac{1}{2}M_{12}(\rho_1\tilde{\sigma}_{i_1} + \tilde{\sigma}_{i_1}\rho_1)\otimes\tilde{\sigma}_{i_2}\right] \\ &= \mathrm{Tr}\left[\frac{1}{2}(M_{12}\,\rho_1\otimes\mathbb{1}_2 + \rho_1\otimes\mathbb{1}_2\,M_{12})\tilde{\sigma}_{i_1}\otimes\tilde{\sigma}_{i_2}\right] \\ &\equiv \mathrm{Tr}[R_{12}\tilde{\sigma}_{i_1}\otimes\tilde{\sigma}_{i_2}].\end{aligned} \tag{26}$$

where the cyclic property of the trace is used in the second equality. Finally, the two-qubit PDM is expressed by

$$R_{12} = \frac{1}{2}(M_{12}\,\rho + \rho\,M_{12}). \tag{27}$$

where $\rho := \rho_1 \otimes \mathbb{1}_2$.

In the following, we verify that the PDM given by Eq. (27) returns to the normal density matrix when there is one time point left.

1) Tracing out the final time $t_2$, the PDM returns to the initial state,

$$\mathrm{Tr}_2\,R_{12} = \frac{1}{2}\,\mathrm{Tr}_2(M_{12}\,\rho + \rho\,M_{12}) = \rho_1, \tag{28}$$

where $\mathrm{Tr}_2\,M_{12} = \mathbb{1}_1$ is used in the second equality,

2) Tracing out the initial time $t_1$, the PDM returns to the output state,

$$\mathrm{Tr}_1\,R_{12} = \frac{1}{2}\,\mathrm{Tr}_1(M_{12}\,\rho + \rho\,M_{12}) = \mathrm{Tr}_1\,M_{12}\,\rho_1\otimes\mathbb{1}_2 = \mathcal{M}_{2|1}(\rho_1), \tag{29}$$

where the cyclic property of the trace is used in the second equality.

## III. DERIVATION OF THE CLOSED-FORM OF $m$-TIME $n$-QUBIT PDM

In this section, we extend the $n$-qubit PDM formalism to across $m$ times. To do that we first consider an $n$-qubit state $\rho$ at time $t_1$ that undergoes the channel $\mathcal{M}_{2|1}$, arrives at time $t_2$, then experiences channel $\mathcal{M}_{3|2}$ and arrives at time $t_3$. Denote the CJ matrices for channels $\mathcal{M}_{2|1}, \mathcal{M}_{3|2}$ by $M_{12}, M_{23}$ respectively. The tensor product '$\otimes$' is sometimes omitted when there is no confusion.

We first recall that the two-time $n$-qubit PDM takes the form

$$R_{12} = \frac{1}{2}(M_{12}\rho + \rho M_{12}). \tag{30}$$

We now proceed to derive a closed form of the 3-time $n$-qubit PDM. To do that, it is sufficient to calculate the expectation values of measurement events. Consider measurement events at times $t_1, t_2, t_3$ being $\{\tilde{\sigma}_{i_1}\}, \{\tilde{\sigma}_{i_2}\}, \{\tilde{\sigma}_{i_3}\}$, respectively. The initial state $\rho_1$ first collapses to the eigenstates of operator $\tilde{\sigma}_{i_1}$ then goes into the channel $\mathcal{M}_{2|1}$. The central quantity we wish to evaluate is $\langle\tilde{\sigma}_{i_1},\tilde{\sigma}_{i_2},\tilde{\sigma}_{i_3}\rangle$.

Our expression for $\langle\tilde{\sigma}_{i_1},\tilde{\sigma}_{i_2},\tilde{\sigma}_{i_3}\rangle$ will turn out to involve a matrix $\rho_2^{(\tilde{\sigma}_{i_1})}$. We now derive a simplified expression for $\rho_2^{(\tilde{\sigma}_{i_1})}$, which we shall use later.

$$\begin{aligned}
\rho_2^{(\tilde{\sigma}_{i_1})} :=& \,\mathrm{Tr}_1[\frac{1}{2}(M_{12}P_+^1\rho_1P_+^1+P_+^1\rho_1P_+^1M_{12})]-\mathrm{Tr}_1[\frac{1}{2}(M_{12}P_-^1\rho_1P_-^1+P_-^1\rho_1P_-^1M_{12})]\\
=& \,\mathrm{Tr}_1[\frac{1}{2}M_{12}(P_+^1\rho_1P_+^1-P_-^1\rho_1P_-^1)]+\mathrm{Tr}_1[\frac{1}{2}(P_+^1\rho_1P_+^1-P_-^1\rho_1P_-^1)M_{12}]\\
=& \,\mathrm{Tr}_1[\frac{1}{4}M_{12}(\rho_1\tilde{\sigma}_{i_1}+\tilde{\sigma}_{i_1}\rho_1)]+\mathrm{Tr}_1[\frac{1}{4}(\rho_1\tilde{\sigma}_{i_1}+\tilde{\sigma}_{i_1}\rho_1)M_{12}]\\
=& \,\mathrm{Tr}_1[\frac{1}{2}(M_{12}\rho_1+\rho_1M_{12})\tilde{\sigma}_{i_1}]\\
=& \,\mathrm{Tr}_1[R_{12}\tilde{\sigma}_{i_1}],
\end{aligned}\tag{31}$$

where the identity operators $\mathbb{1}_2$ at time $t_2$ are ignored throughout the calculation, Eq. (25) is used in the third equality and the cyclic property of the trace is used in the fourth equality. Then, the expectation value of the product of measurement outcomes at three time points is given by

$$\begin{aligned}
&\langle\tilde{\sigma}_{i_1},\tilde{\sigma}_{i_2},\tilde{\sigma}_{i_3}\rangle\\
=& \,\mathrm{Tr}_{23}[M_{23}\Big(P_+^2\rho_2^{(\tilde{\sigma}_{i_1})}P_+^2-P_-^2\rho_2^{(\tilde{\sigma}_{i_1})}P_-^2\Big)\tilde{\sigma}_{i_3}]\\
=&\frac{1}{2}\,\mathrm{Tr}_{23}[M_{23}\Big((\mathrm{Tr}_1\,R_{12}\tilde{\sigma}_{i_1})\sigma_{i_2}+\tilde{\sigma}_{i_2}(\mathrm{Tr}_1\,R_{12}\tilde{\sigma}_{i_1})\Big)\tilde{\sigma}_{i_3}]\\
=&\frac{1}{2}\,\mathrm{Tr}_{123}[(M_{23}R_{12}\tilde{\sigma}_{i_1}\tilde{\sigma}_{i_2}+R_{12}\tilde{\sigma}_{i_1}M_{23}\tilde{\sigma}_{i_2})\tilde{\sigma}_{i_3}]\\
=&\frac{1}{2}\,\mathrm{Tr}[(M_{23}R_{12}+R_{12}M_{23})\tilde{\sigma}_{i_1}\tilde{\sigma}_{i_2}\tilde{\sigma}_{i_3}],
\end{aligned}\tag{32}$$

where Eq. (25) and Eq. (31) are used in the second equality and the cyclic property of the trace is used in the third equality. Thus, we have the $n$-qubit PDM across 3 times as

$$R_{123}=\frac{1}{2}(R_{12}M_{23}+M_{23}R_{12}).\tag{33}$$

Let us assume that we have the $(m-1)$-time $n$-qubit PDM $R_{12,...,m-1}$. We then consider $m$ events $\{\tilde{\sigma}_{i_1},\tilde{\sigma}_{i_2},...,\tilde{\sigma}_{i_m}\}$ and the channel $\mathcal{M}_{i+1|i}$ between time interval $[t_i,t_{i+1}], i\in\{1,2,3,...,m-1\}$ with the corresponding CJ matrix $M_{i,i+1}$. Similarly, the matrix $\rho_{m-1}^{(\vec{\tilde{\sigma}})}$ just before measurement made at time $t_{m-1}$ can be written as

$$\rho_{m-1}^{(\vec{\tilde{\sigma}})}=\mathrm{Tr}_{12...m-2}[R_{12...m-1}\tilde{\sigma}_{i_1}\tilde{\sigma}_{i_2}...\tilde{\sigma}_{i_{m-2}}].\tag{34}$$

The expectation value of the product of measurement outcomes at those time points is given by

$$\begin{aligned}
&\langle\tilde{\sigma}_{i_1},\tilde{\sigma}_{i_2},...,\tilde{\sigma}_{i_m}\rangle\\
=& \,\mathrm{Tr}_{m-1,m}\Big(M_{m-1,m}\Big(P_+^{m-1}\rho_{m-1}^{(\vec{\tilde{\sigma}})}P_+^{m-1}-P_-^{m-1}\rho_{m-1}^{(\vec{\tilde{\sigma}})}P_-^{m-1}\Big)\tilde{\sigma}_{i_m}\Big)\\
=&\frac{1}{2}\,\mathrm{Tr}_{m-1,m}\Big(M_{m-1,m}\Big((\mathrm{Tr}_{12...m-2}\,R_{12...m-1}\tilde{\sigma}_{i_1}\tilde{\sigma}_{i_2}...\tilde{\sigma}_{i_{m-2}})\tilde{\sigma}_{i_{m-1}}\\
&+\tilde{\sigma}_{i_{m-1}}(\mathrm{Tr}_{12...m-2}\,R_{12...m-1}\tilde{\sigma}_{i_1}\tilde{\sigma}_{i_2}...\tilde{\sigma}_{i_{m-2}})\Big)\tilde{\sigma}_{i_m}\Big)\\
=&\frac{1}{2}\,\mathrm{Tr}_{12...m}\Big(\Big(M_{m-1,m}(R_{12...m-1}\tilde{\sigma}_{i_1}\tilde{\sigma}_{i_2}...\tilde{\sigma}_{i_{m-2}})\tilde{\sigma}_{i_{m-1}}\\
&+(R_{12,...,m-1}\tilde{\sigma}_{i_1}\tilde{\sigma}_{i_2}...\tilde{\sigma}_{i_{m-2}})M_{m-1,m}\tilde{\sigma}_{i_{m-1}}\Big)\tilde{\sigma}_{i_m}\Big)\\
=&\frac{1}{2}\,\mathrm{Tr}_{12...m}\Big((M_{m-1,m}R_{12...m-1}+R_{12...m-1}M_{m-1,m})\tilde{\sigma}_{i_1}\tilde{\sigma}_{i_2}...\tilde{\sigma}_{i_m}\Big),
\end{aligned}\tag{35}$$

Thus, we have $R_{12...m}=\frac{1}{2}(R_{12...m-1}M_{m-1,m}+M_{m-1,m}R_{12...m-1})$.

In summary, the $n$-qubit PDM across $m$ times is given by the following iterative expression

$$R_{12...m}=\frac{1}{2}(R_{12...m-1}M_{m-1,m}+M_{m-1,m}R_{12...m-1})\tag{36}$$

with the initial condition $R_{12}=\frac{1}{2}(\rho M_{12}+M_{12}\rho)$ and with $M_{m-1,m}$ denoting the CJ matrix of the $(m-1)$th channel.

## IV. PARTIAL TRACE PROPERTY OF PDM

Given a pseudo-density matrix $R_{A_1B_1A_2B_2}$ defined over two sets of events $A_1B_1$ and $A_2B_2$ and a pseudo-density matrix $R_{A_1B_2}$ constructed from the sets of events $A_1$ and $B_2$, one can show that the latter can be obtained from $R_{A_1B_1A_2B_2}$ by tracing over the subsystem corresponding to $B_1A_2$, i.e., $R_{A_1B_2} = \mathrm{Tr}_{A_2B_1} R_{A_1B_1A_2B_2}$.

*Proof.* For the simplicity of the notations, we give a proof for the two-time two-qubit PDM case, which can be directly generalized to the $m$-time $n$-qubit PDM case. In the following calculation, the tensor product '$\otimes$' is omitted when there is no risk of confusion.

Based on the definition of pseudo-density matrix $R_{A_1A_2B_1B_2}$, we have

$$R_{A_1B_1A_2B_2} = \frac{1}{2^4}\sum_{i,j,k,l=0}^{3}\langle\sigma_i^{A_1}\sigma_j^{B_1},\sigma_k^{A_2}\sigma_l^{B_2}\rangle\sigma_i^{A_1}\sigma_j^{B_1}\sigma_k^{A_2}\sigma_l^{B_2}. \tag{37}$$

We define

$$R_{A_1B_2} = \frac{1}{4}\sum_{i=0}^{3}\sum_{j=0}^{3}\langle\sigma_i^{A_1}\sigma_0^{B_1},\sigma_0^{A_1}\sigma_j^{B_2}\rangle\sigma_i^{A_1}\sigma_j^{B_2},$$

which we express using shorthand as

$$R_{A_1B_2} = \frac{1}{4}\sum_{i=0}^{3}\sum_{j=0}^{3}\langle\sigma_i^{A_1},\sigma_j^{B_2}\rangle\sigma_i^{A_1}\sigma_j^{B_2}. \tag{38}$$

It is then straightforward to calculate that

$$\begin{aligned}
&\mathrm{Tr}_{B_1A_2} R_{A_1B_1A_2B_2}\\
=&\mathrm{Tr}_{B_1A_2}[\frac{1}{2^4}\sum_{i,j,k,l=0}^{3}\langle\sigma_i^{A_1}\sigma_j^{B_1},\sigma_k^{A_2}\sigma_l^{B_2}\rangle\sigma_i^{A_1}\sigma_j^{B_1}\sigma_k^{A_2}\sigma_l^{B_2}]\\
=&\frac{1}{2^4}\sum_{i,j,k,l=0}^{3}\mathrm{Tr}_{B_1A_2}(\sigma_j^{B_1}\sigma_k^{A_2})\langle\sigma_i^{A_1}\sigma_j^{B_1},\sigma_k^{A_2}\sigma_l^{B_2}\rangle\sigma_i^{A_1}\sigma_l^{B_2}\\
=&\frac{1}{2^2}\sum_{i,j=0}^{3}\langle\sigma_i^{A_1}\sigma_0^{B_1},\sigma_0^{A_2}\sigma_j^{B_2}\rangle\sigma_i^{A_1}\sigma_j^{B_2}\\
=&R_{A_1B_2}.
\end{aligned} \tag{39}$$

This completes the proof. □

## V. PROOF OF THEOREM 3

Theorem 3. *If a quantum channel $\mathcal{P}$ does not allow signaling from $B$ to $A$, then, for any state $\rho_{A_1B_1}$ at time $t_1$, the PDM $R_{B_1A_2}$ is positive semidefinite and the PDM negativity $f(R_{B_1A_2}) = 0$.*

*Proof.* A channel $\mathcal{P} : \mathcal{B}(\mathcal{H}_{A_1B_1}) \to \mathcal{B}(\mathcal{H}_{A_2B_2})$ is called semicausal [47], if,

$$\mathrm{Tr}_{B_2}\mathcal{P}(\rho_{A_1B_1}) = \mathcal{T}(\rho_{A_1}), \tag{40}$$

for all input states $\rho_{A_1B_1}$ and for some completely positive maps $\mathcal{T} : \mathcal{B}(\mathcal{H}_{A_1}) \to \mathcal{B}(\mathcal{H}_{A_2})$. (Ref. [47] moreover shows that semicausal is equivalent to semilocalizable )

For the simplicity of the notations, we give a proof to the two-qubit PDM case, which can be directly generalized to the $n$-qubit PDM case. In the Pauli basis, the CJ matrix of channel $\mathcal{P}$ is

$$M = \sum_{i,j=0}^{2^2-1}|i\rangle\langle j|^T\otimes\mathcal{P}(|i\rangle\langle j|) = \frac{1}{2^2}\sum_{i,j=0}^{3}\sigma_i^{A_1}\sigma_j^{B_1}\mathcal{P}(\sigma_i^{A_1}\sigma_j^{B_1}). \tag{41}$$

where the tensor products '$\otimes$' are omitted. The PDM can be calculated as

$$
\begin{aligned}
R_{B_1A_2} =& \mathrm{Tr}_{A_1B_2}\, R_{A_1B_1A_2B_2} \\
=& \frac{1}{2}\,\mathrm{Tr}_{A_1B_2}[M\rho_{A_1B_1} + \rho_{A_1B_1}M] \\
=& \frac{1}{8}\,\mathrm{Tr}_{A_1B_2}\sum_{i,j=0}^{3}\{\rho_{A_1B_1}, \sigma_i^{A_1}\sigma_j^{B_1}\}\mathcal{P}(\sigma_i^{A_1}\sigma_j^{B_1}) \\
=& \frac{1}{8}\,\mathrm{Tr}_{A_1}\sum_{i=0}^{3}\{\rho_{A_1B_1}, \sigma_i^{A_1}\sigma_0^{B_1}\}\mathcal{T}(\sigma_i^{A_1}) \\
=& \frac{1}{4}\sum_{i,l=0}^{3}\langle\sigma_i^{A_1}\sigma_l^{B_1}\rangle\sigma_l^{B_1}\mathcal{T}(\sigma_i^{A_1}) \\
=& \mathbb{1}_B \otimes \mathcal{T}(S\rho_{A_1B_1}S^\dagger),
\end{aligned}
\tag{42}
$$

where $S$ is the swap operator, $\{\,,\}$ denotes the anticommutator, Eq. (41) is used in the third equality and Eq. (40) is used in the fourth equality. Since $\mathcal{T}$ is completely positive, it is straightforward to see that $R_{B_1A_2}$ is positive semidefinite. Finally, $f(R_{B_1A_2}) = 0$ since the causal monotone $f$ measures the negativity of PDMs. This completes the proof. □

The following Corollary follows immediately from Theorem 3.

**Corollary 1.** *If a quantum channel $\mathcal{P}$ does not allow signaling in both directions, then, for any state $\rho_{A_1B_1}$ at time $t_1$, the two PDMs $R_{B_1A_2}$, $R_{A_1B_2}$ are positive semidefinite and the PDM negativities $f(R_{B_1A_2}) = f(R_{A_1B_2}) = 0$.*

## VI. DETERMINING CAUSAL INFLUENCE OF CHANNEL BY OPTIMIZING THE INPUT STATE

We conjecture that, for the case where the input state is free to choose because one is solely interested in whether the channel is signaling, there is always a family of states such that $f > 0$ provides a necessary and sufficient condition for the channel to allow causal influence in a given direction. We here show the conjecture to hold for a quite general class of two-qubit unitary channels.

We slightly restrict the general two-qubit unitary $U_{AB}$ decomposition from Ref. [57], and demand

$$U_{AB} = U_A \otimes U_B\, e^{i\theta S_{AB}}\, V_A \otimes V_B, \tag{43}$$

where $S$ denotes the swap operator, and the operator $e^{i\theta S_{AB}}$ is the reason for causal influence from both sides. By property (II), the PDM negativity $f(R)$ is invariant under local change of basis. Thus one can ignore the two factors $U_A \otimes U_B$ and $V_A \otimes V_B$ in $U_{AB}$ when analysing causal influence via $f$.

We now provide an example to illustrate that an input state can be found to witness causal influence of the unitary operator $e^{i\theta S_{AB}}$. Consider the state of the compound system $\rho_{A_1B_1} = |0\rangle_{A_1}\langle 0| \otimes |0\rangle_{B_1}\langle 0|$ undergoing a unitary evolution:

$$U_{AB}(\theta) = e^{i\theta S} = \cos(\theta)\,\mathbb{1} + i\sin(\theta)\,S := c\mathbb{1} + isS.$$

Under the evolution $U_{AB}(\theta)$, the initial state $\rho_{A_1}$ of the system cannot influence its final state $\rho_{A_2}$ when $c = 0$, i.e., there is no causal influence from $A_1$ to $A_2$. Moreover, there is causal influence from $A_1$ to $A_2$ when $c \neq 0$. Similar analysis is also applicable to examining causal influences from $A_1$ to $B_2$, $B_1$ to $B_2$, and $B_1$ to $A_2$.

The effective dynamics of the system $A$ can be modeled as a quantum channel with the set of Kraus operators given by $\{c\mathbb{1} + is\,|0\rangle\langle 0|\,, is\,|0\rangle\langle 1|\}$ [21]. The corresponding 2-time PDM of the system $A$ is given by

$$R_{A_1A_2} = |00\rangle\langle 00| + \frac{c(c+is)}{2}\,|01\rangle\langle 10| + \frac{c(c-is)}{2}\,|10\rangle\langle 01|\,. \tag{44}$$

The eigenvalues of $R_{A_1A_2}$ are $\{1, -\frac{c}{2}, \frac{c}{2}, 0\}$. Therefore, $f(R_{A_1A_2}) = |c|$. Therefore, under the choice of the initial state $\rho_{A_1B_1} = |0\rangle_{A_1}\langle 0| \otimes |0\rangle_{B_1}\langle 0|$, $f(R_{A_1A_2}) > 0$ provides a necessary and sufficient condition for the parameterized channel to allow causal influence from $A_1$ to $A_2$.

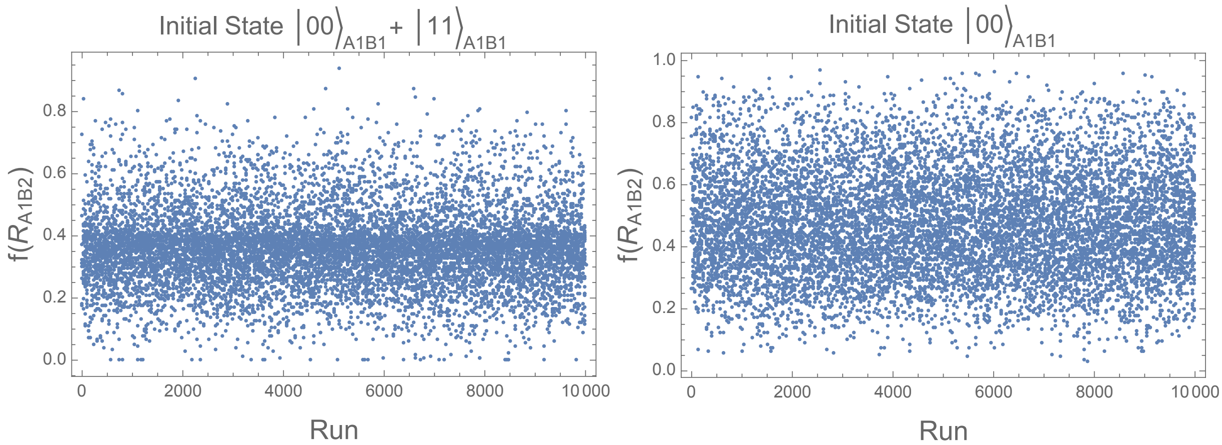


FIG. 4. **Simulations of negativity of PDM with fixed input and random circuits.** We set the channel $\mathcal{N}$ in FIG. 2 to be $\mathcal{N}(\cdot) = S(\cdot)S^\dagger$. $\mathcal{N}$ is here modelled as a unitary channel from $AC$ to $AC$ and $\mathcal{M}$ similarly from $BC$ to $BC$. Then, we simulate the negativity $f(R_{A_1B_2})$ of the PDM $R_{A_1B_2}$ for two pure inputs, $\psi_{A_1B_1} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ and $\psi_{A_1B_1} = |00\rangle$, with $\mathcal{M}$ being a Haar-random 2-qubit unitary channel in FIG. 2. There is a cause-effect mechanism between $A$ and $B$ when $\mathcal{M}$ is not a local unitary channel. We observe that almost all cases have negativity, therefore the negativity rules out the common cause mechanism with a near unit probability in the scenarios.

## VII. PURE CAUSE-EFFECT MECHANISM IS CP

Suppose that the pure cause-effect mechanism shown in case 1 of FIG. 1 is represented by no initial correlation and the semi-causal channel $\mathcal{P} = \mathcal{M}_{BC} \circ \mathcal{N}_{AC}$. $C$ is the intermediate system in the semi-localizable channel. We aim to show that the process $\mathcal{L}_{A\to B}$ induced by $\mathcal{P}$ is a CP map.

First, suppose that the channel $\mathcal{P}$ admits the unitary dilation $U$ such that

$$\mathcal{P}(\rho_{AB}) = \mathrm{Tr}_{CE}\left(U(\rho_{AB} \otimes \rho_C \otimes \rho_E)U^\dagger\right), \tag{45}$$

where $E$ is used for the unitary dilation of $\mathcal{P}$. Second, there is no initial correlation between $A$ and $B$, meaning $\rho_{AB} = \rho_A \otimes \rho_B$ where $\rho_A = \mathrm{Tr}_B\, \rho_{AB}$ and $\rho_B = \mathrm{Tr}_A\, \rho_{AB}$. Combining the two points, the process $\mathcal{L}_{A\to B}$ is given by

$$\mathcal{L}_{A\to B}(\rho_A) = \mathrm{Tr}_{AC}\, \mathcal{P}(\rho_{AB}) = \mathrm{Tr}_{ACE}\left(U(\rho_A \otimes \rho_B \otimes \rho_C \otimes \rho_E)U^\dagger\right). \tag{46}$$

It is not hard to see that the action of $\mathcal{L}_{A\to B}$ can be expressed as the sum of Kraus operators acting on the state. Therefore, the process $\mathcal{L}_{A\to B}$ is CP.

## VIII. SUPPORTING DETAILS FOR QUANTUM CAUSAL INFERENCE

Here we provide detailed reasoning and techniques for the proposed quantum causal inference protocol.

Firstly, if $f(R_{AB}) = 0$, then, according to Proposition 4, $R_{AB}$ can be treated as a density matrix, making it compatible with the common cause mechanism. However, this does not rule out the possibility of $R_{AB}$ representing other causal structures. Nevertheless, we numerically study the behaviour of the negativity of the PDM in FIGs. 4 and 5 and find that $f(R_{AB}) > 0$ with a near unit probability when there is a causal-effect mechanism between $A$ and $B$.

**Proposition 4.** *Given that one only has access to the data that constructs a two-time PDM $R_{AB}$, if $f(R_{AB}) = 0$, then $R_{AB}$ is indistinguishable from the density matrix $\rho_{AB}$ that shares the same entries as $R_{AB}$.*

*Proof.* The proposition is a result of Ref. [12]. According to the definition, any two-time PDM $R_{AB}$ is Hermitian and has a unit trace. These two properties, along with $f(R_{AB}) = 0$, lead to the construction of a density matrix $\rho_{AB}$ with the same entries as $R_{AB}$. Given that we only have access to the data, there is no way to tell whether it is a density matrix or a two-time PDM. □

Secondly, if $f(R_{AB}) > 0$, there must exist temporal correlations, indicating the involvement of the cause-effect mechanism. In this case, one needs the Choi matrix of the process and its time reversal to tell the cause and effect. To extract the CJ matrix $M$ from $R$, we employ the vectorization technique as described in Ref. [49]. For any quantum operator $A = \sum_{ij} A_{ij} |i\rangle\langle j| \in$

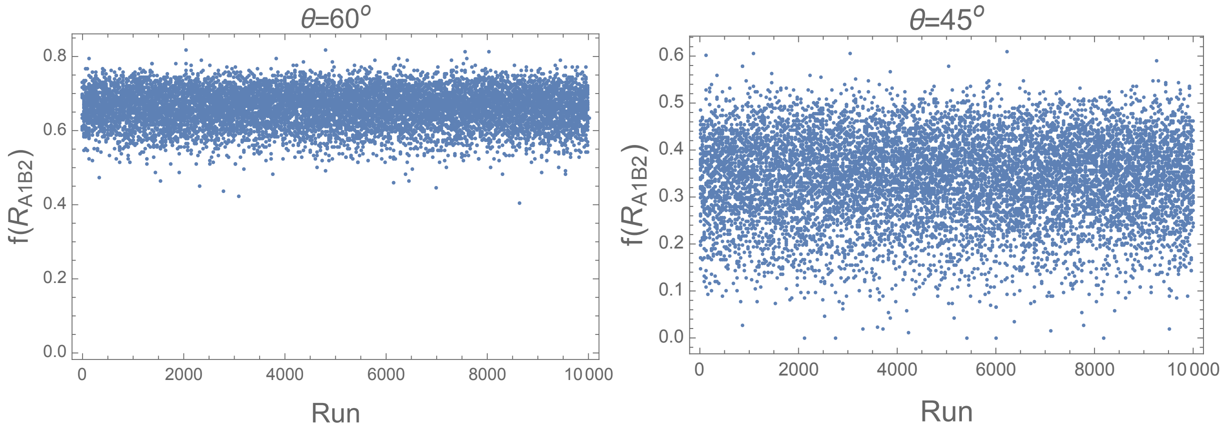


FIG. 5. **Simulations of negativity of PDM with fixed circuit and random inputs.** We set $\mathcal{N}(\cdot) = S(\cdot)S^\dagger$ and $\mathcal{M} = U(\theta)(\cdot)U^\dagger(\theta)$ with $U(\theta) = e^{-i\theta S} := \cos(\theta)\mathbb{1} - i\sin(\theta)S$. In this setting, there is a cause-effect mechanism between $A$ and $B$ when $\sin(\theta) \neq 0$. We simulate the negativity $f(R_{A_1B_2})$ of the PDM $R_{A_1B_2}$ for cases of $\theta = 30^\circ$ and $\theta = 60^\circ$ with the pure input state $|\psi\rangle_{A_1B_1}$ being Haar-random. We observe that almost all cases have negativity, therefore the negativity rules out the common cause mechanism with a near unit probability in the scenarios.

$\mathcal{B}(H)$, its vectorization is given by

$$|A\rangle\rangle = \sum_{ij} A_{ij} |i\rangle \otimes |j\rangle \in H \otimes H. \tag{47}$$

The vectorization of the PDM $R$ is then given by

$$|R\rangle\rangle = \frac{1}{2}|\rho M + M\rho\rangle\rangle = \frac{1}{2}\left(\rho_1 \otimes \mathbb{1}_2 \otimes \mathbb{1}_1 \otimes \mathbb{1}_2 + \mathbb{1}_1 \otimes \mathbb{1}_2 \otimes \rho_1^T \otimes \mathbb{1}_2\right)|M\rangle\rangle =: B|M\rangle\rangle, \tag{48}$$

where $|EFG\rangle\rangle = G \otimes E^T |F\rangle\rangle$ is used in the second equality. When $\rho_1$ is full rank, then the complete information of $M$ is extracted via $|M\rangle\rangle = B^{-1}|R\rangle\rangle$. However, when $\rho_1$ is rank deficient, only the process information on the support of $B$ is obtained via $|M\rangle\rangle = B^\ddagger|R\rangle\rangle$ where $B^\ddagger$ denotes the pseudo-inverse of $B$.

Let us consider an example when $\rho_1$ is rank deficient. Suppose $\rho_1 = |0\rangle\langle 0|$ undergoes an identity channel. The CJ matrix for the identity channel, $\rho(= \rho_1 \otimes \mathbb{1}_2)$ and the corresponding PDM are given by

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \ \rho = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \ R = \frac{1}{2}(\rho M + M\rho) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{49}$$

Working backwards to recover $M$ using $|N\rangle\rangle = B^\ddagger|R\rangle\rangle$, where $N$ denotes the recovered form for $M$, yields

$$N = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & a & b \\ 0 & 0 & c & d \end{pmatrix}, \tag{50}$$

where $a, b, c, d$ are any complex numbers. There are infinitely many solutions. There might be scenarios where some of the Choi matrices $N^T$ corresponding to the solution $N$ are positive, while others are not. We want to examine the compatibility of the data with the pure cause-effect mechanism first, as it is simpler compared to the mixture of the common-cause and cause-effect mechanisms. Therefore, our goal is to find the solutions where $N^T$ is the least negative. We thus design the following semidefinite programming to achieve our goal.

Suppose that the solution $N^{T_1}$ has the decomposition $N^T = N_+^T - N_-^T$ where $N_+^T \geq 0$ and $N_-^T \geq 0$. The semidefinite programming problem is given by

$$\begin{aligned} &\text{minimize } \operatorname{Tr} N_-^T, \\ &\text{subject to } \begin{cases} N \text{ is Hermitian}, \\ N^T = N_+^T - N_-^T \\ \operatorname{Tr}_2 N = \mathbb{1}, \\ N_+^T \geq 0, N_-^T \geq 0. \end{cases} \end{aligned} \tag{51}$$

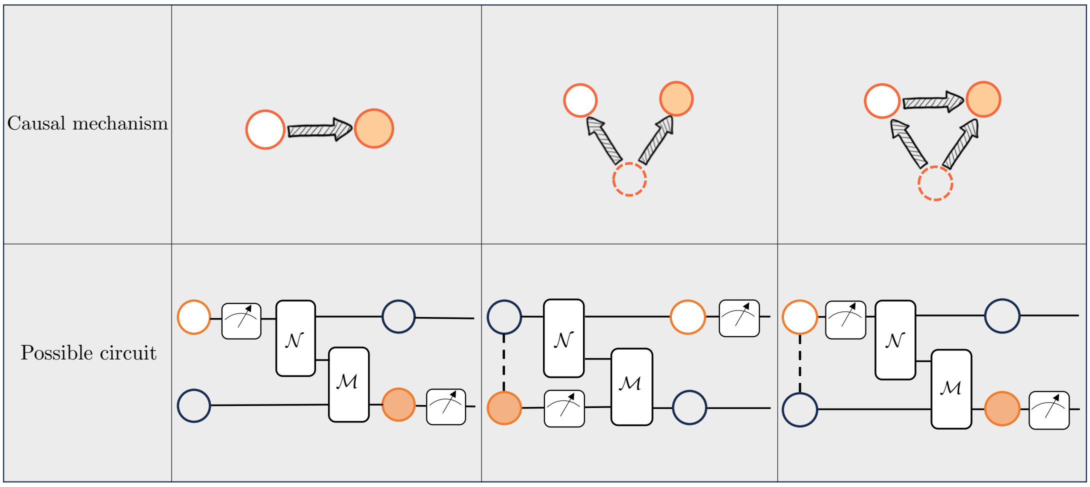


FIG. 6. **Correspondence between causal mechanisms and possible circuit models.** Each column represents the causal mechanism (top row) and an example of a circuit model realising that mechanism (bottom row). From left to right: (i) the cause-effect mechanism which corresponds to case 1 and in inverse direction case 2 of FIG. 1; (ii) The pure common-cause mechanism which corresponds to case 3 in FIG. 1; (iii) The mixture of the cause-effect and common-cause mechanisms, which corresponds to case 5 and in inverse direction case 4 of FIG. 1. The dashed line in the circuit denotes the initial correlation constituting the common case. The measurements used to define the PDM of interest are also depicted.

Finally, detailed justifications for protocol for quantum causal inference are listed in the following.

(1) **Evaluating compatibility with a common-cause mechanism**. Consider the case of no negativity ($f(R_{A_iB_j}) = 0$). Theorem 3 shows that negativity associated with two times ($f(R_{A_iB_j}) \neq 0$) cannot come solely from a common cause ($R_{A_1B_1} \neq R_{A_1} \otimes R_{B_1}$). Thus in this case the data $R_{A_iB_j}$ is compatible with the (purely) common cause mechanism (case 3 in FIG. 1). In particular, by proposition 4, $R_{A_iB_j} = \rho$ where $\rho$ is a single-time standard density matrix so the data generating $R_{A_iB_j}$ can be produced by measurements on two subsystems initially prepared in $\rho$.

(2) **Evaluating compatibility with different cause-effect mechanisms.** Consider the case of negativity ($f(R_{A_iB_j}) > 0$). Theorem 3 rules out the common cause mechanism and we are left to evaluate the compatibility of the data with cases 1, 2, 4, and 5 in FIG. 1. We make use of the time asymmetry results described around Eq. (11) for this evaluation. In particular we extract the two Choi matrices $M^T$, $\bar{M}^T$ associated with $R_{A_iB_j}$ and its time reversal $\bar{R}_{A_iB_j}$. The basic idea is that $M^T > 0$ means there is a CP map on $A$ that gives $B$, indicating that $A$ is the cause and $B$ the effect. More specifically,

- If $M^T \geq 0$ and $\bar{M}^T \ngeq 0$, the data is compatible with $A \to B$ (case 1 in FIG. 1). In particular, case 1 in FIG. 1 corresponds to a semi-causal channel with signalling from $A$ to $B$ acting on a state with no initial correlations. As shown around Eq. (46) the forwards evolution from $A$ to $B$ is then a CP map, such that $M^T \geq 0$. To generate the data $R_{A_iB_j}$ through mechanism 1, one prepares $R_{A_i}$, and applies the CP map associated with $M^T$.

- If $M^T \ngeq 0$ and $\bar{M}^T \geq 0$, the data is compatible with $A \leftarrow B$ (case 2 in FIG. 1). To generate the data $R_{A_iB_j}$ through mechanism 2, one prepares $R_{B_j}$, and applies the CP map associated with $\bar{M}^T$.

- If $M^T \geq 0$ and $\bar{M}^T \geq 0$, the data is compatible with case 1 and/or case 2 in FIG. 1. Now the data can by inspection be generated by either of the immediately above methods.

(3) If none of the above conditions are satisfied, i.e., $f(R_{A_iB_j}) > 0$, $M^T \ngeq 0$ and $\bar{M}^T \ngeq 0$, the causal structure is compatible only with case 4 or 5 in FIG. 1. To generate the data using case 4 or 5 one may use a circuit corresponding to that case (circuit (3) in FIG.6), generate the full PDM and then trace out the undesired parts. We do not give a direct simple way to find a circuit consistent with $R_{A_iB_j}$ but there must be one or more such circuits generating the data since we assume the data is generated by either case 1, 2, 3, 4 or 5 and cases 1, 2 and 3 are inconsistent with this data. More specifically, case 3 is ruled out by $f(R_{AB}) > 0$, and the cases 1 and 2 are incompatible with $M^T < 0$ and $\bar{M}^T < 0$ both holding, since the arguments around Eq. (46) show there is a CP map in at least one direction for those cases.

## IX. EXAMPLE: CAUSE-EFFECT MECHANISM WITH A COMMON CAUSE

Suppose the PDM corresponds to a semi-causal channel as in FIG. 2 but with an initially correlated state. One can see that

$$R_{A_1B_2} = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & c^2 \\ 0 & 0 & s^2 & 0 \\ 0 & s^2 & 0 & 0 \\ c^2 & 0 & 0 & 1 \end{pmatrix}, \tag{52}$$

is the consequence of the initial state $|\psi\rangle_{A_1B_1} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, sequentially undergoing swap channel $\mathcal{N}_{AC} = S$ followed by $\mathcal{M}_{BC} = e^{i\theta S} := c\mathbb{1} + isS$, where $C$ is the intermediate system in the semi-causal channel (see FIG. 2). (We exclude $s = 1$ in which case $A$ replaces $B$ but they do not interact with each other, so the initial correlation does not play a role.)

Following our causal inference protocol, we first determine that $f(R_{A_1B_2}) > 0$, implying the existence of the cause-effect mechanism. We then proceed to calculate the Choi matrix of the process and its time reversal, which can be shown to be

$$M^T = \begin{pmatrix} 1 & 0 & 0 & s^2 \\ 0 & 0 & c^2 & 0 \\ 0 & c^2 & 0 & 0 \\ s^2 & 0 & 0 & 1 \end{pmatrix} = \bar{M}^T. \tag{53}$$

Clearly, $M^T, \bar{M}^T \not\geq 0$. Therefore the protocol would judge the data to be compatible with case 4 or 5 of FIG. 1, corresponding, as desired, to initial correlations and causal influence combined.


References:

(12) Fitzsimons, J., Jones, J. and Vedral, V. Quantum correlations which imply causation. Scientific reports 5, 18281 (2016).

(49) X. Liu, Q. Chen, and O. Dahlsten, Inferring the arrow of time in quantum spatiotemporal correlations, Physical Review A 109, 032219 (2024).

(56) A.Souza, I.Oliveira, and R.Sarthour, A scattering quantum circuit for measuring Bell's time inequality: a nuclear magnetic resonance demonstration using maximally mixed states, New Journal of Physics 13, 053023 (2011).

(57) B. Kraus and J. I. Cirac, Optimal creation of entanglement using a two-qubit gate, Physical Review A 63, 062309 (2001).