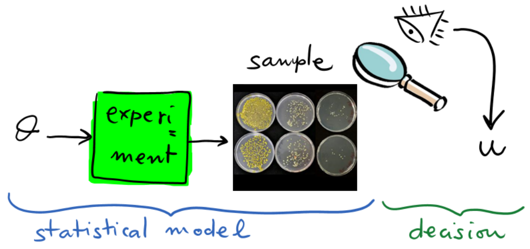
# The Theory of Quantum Statistical Comparison and Some Applications in Quantum Information Sciences

Francesco Buscemi (Nagoya U)

*The 39th Quantum Information Technologies Symposium (QIT)*.
RCAST, The University of Tokyo, 27 November 2018

# The Original Formulation

# Statistical Models and Decision Problems




$$\begin{tikzcd} \Theta \arrow[r, "\text{experiment}"] \arrow[d, squiggly] & \mathcal{X} \arrow[r, "\text{decision}"] \arrow[d, squiggly] & \mathcal{U} \arrow[d, squiggly] \\ \theta \arrow[r, "w(x|\theta)"'] & x \arrow[r, "d(u|x)"'] & u \end{tikzcd}$$

## Definition

- The statistical model is given by: the parameter set $\Theta$, the sample set $\mathcal{X}$, and the PDs $w(x|\theta)$.
- The statistical decision problem: is given by the parameter set $\Theta$, the action set $\mathcal{U}$, and the payoff function $\ell : \Theta \times \mathcal{U} \to \mathbb{R}$.

# How Much Is an Experiment Worth?

- the experiment *is given*, i.e., it is the "resource"
- the decision instead *can be optimized*

$$\begin{tikzcd}
\Theta \arrow[r, "\text{experiment}"] \arrow[d, squiggly] & \mathcal{X} \arrow[r, "\text{decision}"] \arrow[d, squiggly] & \mathcal{U} \arrow[d, squiggly] \\
\theta \arrow[r, "w(x|\theta)"'] & x \arrow[r, "d(u|x)"'] & u
\end{tikzcd}$$

## Definition (Expected Payoff)

The expected payoff of statistical model $\mathbf{w} = \langle \Theta, \mathcal{X}, w(x|\theta) \rangle$ w.r.t. a decision problem $\boldsymbol{\ell} = \langle \Theta, \mathcal{U}, \ell(\theta, u) \rangle$ is given by

$$\mathbb{E}_{\boldsymbol{\ell}}[\mathbf{w}] \stackrel{\text{def}}{=} \max_{d(u|x)} \sum_{u,x,\theta} \ell(\theta, u) d(u|x) w(x|\theta) |\Theta|^{-1} .$$

# Comparing Statistical Models 1/2

First model: $\mathbf{w} = \langle \Theta, \mathcal{X}, w(x|\theta) \rangle$

$$\begin{tikzcd}
\Theta \arrow[r, "\text{experiment}"] \arrow[d, squiggly] & \mathcal{X} \arrow[r, "\text{decision}"] \arrow[d, squiggly] & \mathcal{U} \arrow[d, squiggly] \\
\theta \arrow[r, "w(x|\theta)"'] & x \arrow[r, "d(u|x)"'] & u
\end{tikzcd}$$

Second model: $\mathbf{w}' = \langle \Theta, \mathcal{Y}, w'(y|\theta) \rangle$

$$\begin{tikzcd}
\Theta \arrow[r, "\text{experiment}"] \arrow[d, squiggly] & \mathcal{Y} \arrow[r, "\text{decision}"] \arrow[d, squiggly] & \mathcal{U} \arrow[d, squiggly] \\
\theta \arrow[r, "w'(y|\theta)"'] & y \arrow[r, "d'(u|y)"'] & u
\end{tikzcd}$$

Given a statistical decision problem $\boldsymbol{\ell} = \langle \Theta, \mathcal{U}, \ell(\theta, u) \rangle$, if $\mathbb{E}_{\boldsymbol{\ell}}[\mathbf{w}] \geq \mathbb{E}_{\boldsymbol{\ell}}[\mathbf{w}']$, then one says that model $\mathbf{w}$ is *more informative than* model $\mathbf{w}'$ with respect to problem $\boldsymbol{\ell}$.

# Comparing Statistical Models 2/2

**Definition (Information Preorder)**

If model $\mathbf{w} = \langle \Theta, \mathcal{X}, w(x|\theta) \rangle$ is more informative than model $\mathbf{w}' = \langle \Theta, \mathcal{Y}, w'(y|\theta) \rangle$ **for all** decision problems $\boldsymbol{\ell} = \langle \Theta, \mathcal{U}, \ell(\theta, u) \rangle$, then we say that $\mathbf{w}$ is *(always) more informative* than $\mathbf{w}'$, and write

$$\mathbf{w} \succeq \mathbf{w}' \,.$$

**Problem.** The information preorder is operational, but not really "concrete". Can we visualize this better?

# A Fundamental Result

**Blackwell-Sherman-Stein (1948-1953)**

Given two models with the same parameter space, $\mathbf{w} = \langle \Theta, \mathcal{X}, w(x|\theta) \rangle$ and $\mathbf{w}' = \langle \Theta, \mathcal{Y}, w'(y|\theta) \rangle$, the condition $\mathbf{w} \succeq \mathbf{w}'$ holds *iff* $\mathbf{w}$ is **sufficent** for $\mathbf{w}'$, that is, iff there exists a conditional PD $\varphi(y|x)$ such that $w'(y|\theta) = \sum_x \varphi(y|x) w(x|\theta)$.

$$\begin{array}{ccccccccc}
\Theta & \longrightarrow & \mathcal{Y} & & \Theta & \longrightarrow & \mathcal{X} & \xrightarrow{\text{noise}} & \mathcal{Y} \\
\wr & & \wr & = & \wr & & \wr & & \wr \\
\theta & \xrightarrow[w'(y|\theta)]{} & y & & \theta & \xrightarrow[w(x|\theta)]{} & x & \xrightarrow[\varphi(y|x)]{} & y
\end{array}$$

David H. Blackwell (1919-2010)

# The Precursor: Majorization

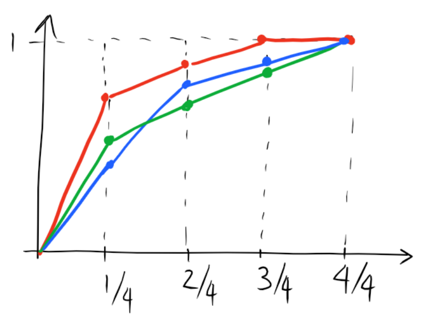
# Lorenz Curves and Majorization

- two probability distributions, $\boldsymbol{p}$ and $\boldsymbol{q}$, of the same dimension $n$
- truncated sums $P(k) = \sum_{i=1}^{k} p_i^{\downarrow}$ and $Q(k) = \sum_{i=1}^{k} q_i^{\downarrow}$, for all $k = 1, \ldots, n$
- $\boldsymbol{p}$ majorizes $\boldsymbol{q}$, i.e., $\boldsymbol{p} \succeq \boldsymbol{q}$, whenever $P(k) \geq Q(k)$, for all $k$
- minimal element: uniform distribution $\boldsymbol{e} = n^{-1}(1, 1, \cdots, 1)$

**Hardy, Littlewood, and Pólya (1934)**

$\boldsymbol{p} \succeq \boldsymbol{q} \iff \boldsymbol{q} = M\boldsymbol{p}$, for some bistochastic matrix $M$.

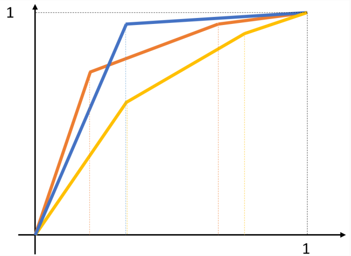
Lorenz curve for probability distribution $\boldsymbol{p} = (p_1, \cdots, p_n)$:




$$(x_k, y_k) = (k/n, P(k)), \quad 1 \leq k \leq n$$

# Generalization: Relative Majorization

- two *pairs* of probability distributions, $(\boldsymbol{p}_1, \boldsymbol{p}_2)$ and $(\boldsymbol{q}_1, \boldsymbol{q}_2)$, of dimension $m$ and $n$, respectively
- relabel entries such that ratios $p_1^i/p_2^i$ and $q_1^j/q_2^j$ are nonincreasing
- construct the truncated sums $P_{1,2}(k) = \sum_{i=1}^{k} p_{1,2}^i$ and $Q_{1,2}(k)$
- $(\boldsymbol{p}_1, \boldsymbol{p}_2) \succeq (\boldsymbol{q}_1, \boldsymbol{q}_2)$ iff the relative Lorenz curve of the former is never below that of the latter




Relative Lorenz curves:

$(x_k, y_k) = (P_2(k), P_1(k))$

**Blackwell (Theorem for Dichotomies), 1953**

$(\boldsymbol{p}_1, \boldsymbol{p}_2) \succeq (\boldsymbol{q}_1, \boldsymbol{q}_2) \iff \boldsymbol{q}_i = M\boldsymbol{p}_i$, for some stochastic matrix $M$.

# Formulation in Terms of Channels

# Statistics vs Information Theory

**Statistical theory**

Nature does not bother with coding

$$\begin{tikzcd}
\Theta \arrow[r, "\text{experiment}"] \arrow[d, squiggly] & \mathcal{X} \arrow[r, "\text{decision}"] \arrow[d, squiggly] & \mathcal{U} \arrow[d, squiggly] \\
\theta \arrow[r, "w(x|\theta)"'] & x \arrow[r, "d(u|x)"'] & u
\end{tikzcd}$$

**Communication theory**

a sender, instead, *does code*

$$\begin{tikzcd}
\mathcal{M} \arrow[r, "\text{encoding}"] \arrow[d, squiggly] & \Theta \arrow[r, "\text{channel}"] \arrow[d, squiggly] & \mathcal{X} \arrow[r, "\text{decoding}"] \arrow[d, squiggly] & \mathcal{U} \arrow[d, squiggly] \\
m \arrow[r, "e(\theta|m)"'] & \theta \arrow[r, "w(x|\theta)"'] & x \arrow[r, "d(u|x)"'] & u
\end{tikzcd}$$

# Statistics vs Information Theory

**Statistical theory**

Nature does not bother with coding

$$
\begin{tikzcd}
\Theta \arrow[r, "\text{experiment}"] \arrow[d, squiggly] & \mathcal{X} \arrow[r, "\text{decision}"] \arrow[d, squiggly] & \mathcal{U} \arrow[d, squiggly] \\
\theta \arrow[r, "w(x|\theta)"'] & x \arrow[r, "d(u|x)"'] & u
\end{tikzcd}
$$

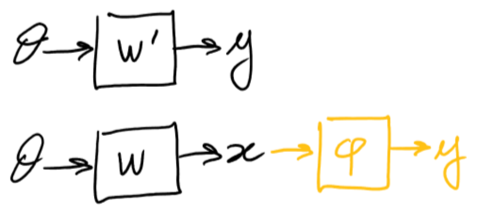
**Communication theory**

a sender, instead, *does code*

$$
\begin{tikzcd}
\mathcal{M} \arrow[r, "\text{encoding}"] \arrow[d, squiggly] & \mathcal{X} \arrow[r, "\text{channel}"] \arrow[d, squiggly] & \mathcal{Y} \arrow[r, "\text{decoding}"] \arrow[d, squiggly] & \mathcal{M} \arrow[d, squiggly] \\
m \arrow[r, "e(x|m)"'] & x \arrow[r, "w(y|x)"'] & y \arrow[r, "d(m'|y)"'] & m'
\end{tikzcd}
$$

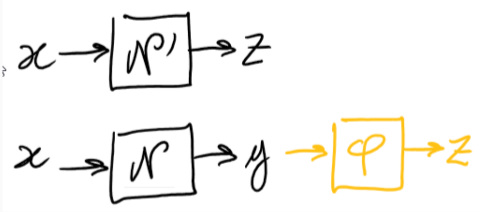
# Sufficiency vs Degradability

**Sufficiency relation**
for statistical experiments




$$w'(y|\theta) = \sum_x \varphi(y|x) w(x|\theta)$$

**Degradability relation**
for noisy channels




$$w'(z|x) = \sum_y \varphi(z|y) w(y|x)$$

Only the labeling convention changes, but the two conditions are absolutely equivalent.

# Decoding Problems and Codes Fidelities

When dealing with communication channels, it is natural to restrict to particular decision problems that we name "decoding problems".


$\mathcal{E} = \{e(x|m)\}$, $\mathcal{N} = \{w(y|x)\}$, $\mathcal{D} = \{d(m'|y)\}$

### Code Fidelity

Given a noisy channel $\mathcal{N} : \mathcal{X} \to \mathcal{Y}$, its *code fidelity*, for any set $\mathcal{M}$ and any coding channel $\mathcal{E} : \mathcal{M} \to \mathcal{X}$, is defined as

$$\mathbb{E}_{\langle \mathcal{E} \rangle}[\mathcal{N}] \stackrel{\text{def}}{=} \max_{\mathcal{D}:\mathcal{Y}\to\mathcal{M}} \frac{1}{|\mathcal{M}|} \sum_{m,x,y,m'} e(x|m)w(y|x)d(m'|y)\delta_{m,m'}$$

# Comparison of Noisy Channels

**Theorem (Coding Problems Are Complete)**

*Given two noisy channels $\mathcal{N} : \mathcal{X} \to \mathcal{Y}$ and $\mathcal{N}' : \mathcal{X} \to \mathcal{Y}'$, $\mathcal{N}$ is degradable into $\mathcal{N}'$ if and only if*

$$\mathbb{E}_{\langle \mathcal{E} \rangle}[\mathcal{N}] \geq \mathbb{E}_{\langle \mathcal{E} \rangle}[\mathcal{N}'] \,,$$

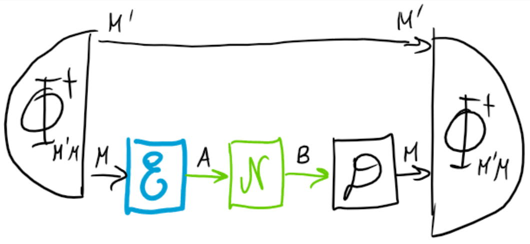
*for all codes $\mathcal{E} : \mathcal{M} \to \mathcal{X}$, with $\mathcal{M} \cong \mathcal{Y}'$.*

# Extensions to the Quantum Case

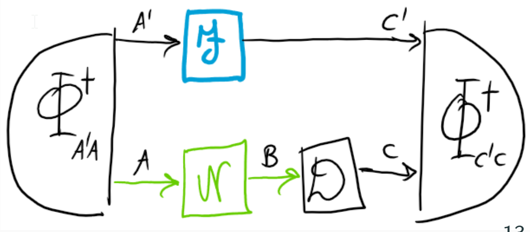
# Extending Decoding Problems


decoding problems

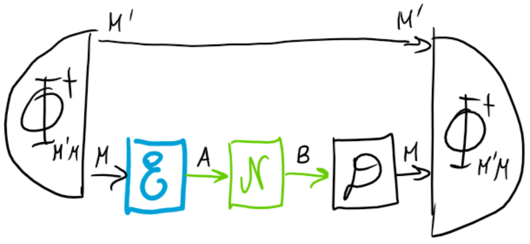
quantum decoding problems




quantum "realignment" problems



# Quantum Decoding Problems



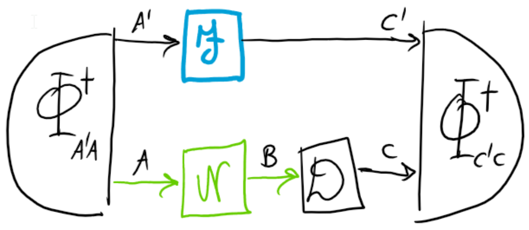

$$|\Phi^+_{M'M}\rangle = \frac{1}{\sqrt{d_M}} \sum_{i=1}^{d_M} |i\rangle_M |i\rangle_M$$

## Quantum Code Fidelity

Given a quantum channel (i.e., CPTP linear map) $\mathcal{N} : A \to B$, its *quantum code fidelity*, for any Hilbert space $\mathcal{H}_M \cong \mathcal{H}_{M'}$ and any quantum coding channel $\mathcal{E} : M \to A$, is defined as

$$\mathbb{E}^q_{\langle \mathcal{E} \rangle}[\mathcal{N}] \stackrel{\text{def}}{=} \max_{\mathcal{D}:B\to M} \langle \Phi^+_{M'M} | (\mathrm{id}_{M'} \otimes \mathcal{D} \circ \mathcal{N} \circ \mathcal{E})(\Phi^+_{M'M}) | \Phi^+_{M'M} \rangle = d_M^{-1} 2^{-H_{\min}(M'|B)}$$

# Quantum Realignment Problems




## Quantum Realignment Fidelity

Given a quantum channel $\mathcal{N} : A \to B$, for any Hilbert space $\mathcal{H}_C \cong \mathcal{H}_{C'}$ and any "misaligning" channel $\mathcal{F} : A' \to C'$, its *quantum realignment fidelity* is defined as

$$\mathbb{F}^q_{\langle \mathcal{F} \rangle}[\mathcal{N}] \stackrel{\text{def}}{=} \max_{\mathcal{D}:B\to C} \langle \Phi^+_{C'C} | (\mathcal{F}_{A'} \otimes \mathcal{D} \circ \mathcal{N})(\Phi^+_{A'A}) | \Phi^+_{C'C} \rangle = d_C^{-1} 2^{-H_{\min}(C'|B)}$$

# Comparison of Quantum Channels

**Theorem (Quantum Coding and Realignment Problems Are Complete)**

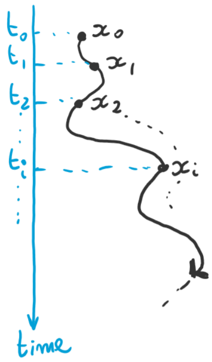
*Given two quantum channels $\mathcal{N} : A \to B$ and $\mathcal{N}' : A \to B'$, the following are equivalent:*

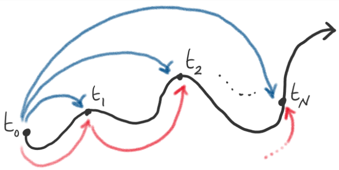
1. *$\mathcal{N}$ is degradable into $\mathcal{N}'$;*
2. *for any quantum coding channel $\mathcal{E} : M \to A$, with $\mathcal{H}_M \cong \mathcal{H}_{B'}$, one has $\mathbb{E}^q_{\mathcal{E}}[\mathcal{N}] \geq \mathbb{E}^q_{\mathcal{E}}[\mathcal{N}']$, or, equivalently, $H_{\min}(M'|B) \leq H_{\min}(M'|B')$;*
3. *for any quantum misaligning channel $\mathcal{F} : A' \to C'$, with $\mathcal{H}_{C'} \cong \mathcal{H}_{B'}$, one has $\mathbb{F}^q_{\mathcal{F}}[\mathcal{N}] \geq \mathbb{F}^q_{\mathcal{F}}[\mathcal{N}']$, or, equivalently, $H_{\min}(C'|B) \leq H_{\min}(C'|B')$.*

# Application to Open Quantum Systems Dynamics

# Discrete-Time Stochastic Processes

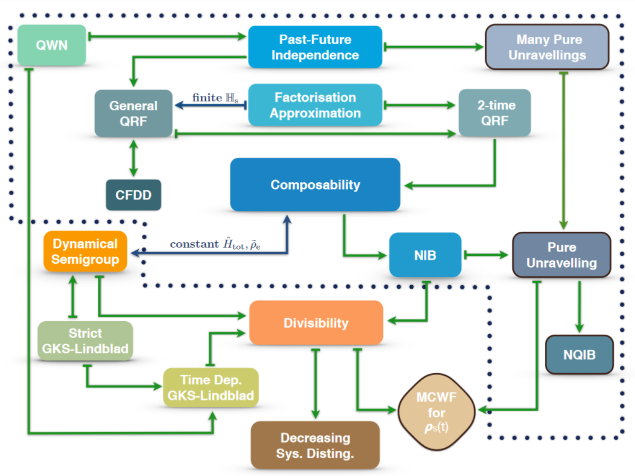
- Let $x_i$, for $i = 0, 1, \dots,$ index the state of a system at time $t = t_i$
- if the system can be initialized at time $t = t_0$, the process is fully described by the conditional distribution $p(x_N, \dots, x_1 | x_0)$
- if the system evolving is quantum, we only have a quantum dynamical mapping $\left\{ \mathcal{N}^{(i)}_{Q_0 \to Q_i} \right\}_{i=1,\dots,N}$
- the process is divisible if there exist channels $\mathcal{D}^{(i)}$ such that $\mathcal{N}^{(i+1)} = \mathcal{D}^{(i)} \circ \mathcal{N}^{(i)}$ for all $i$



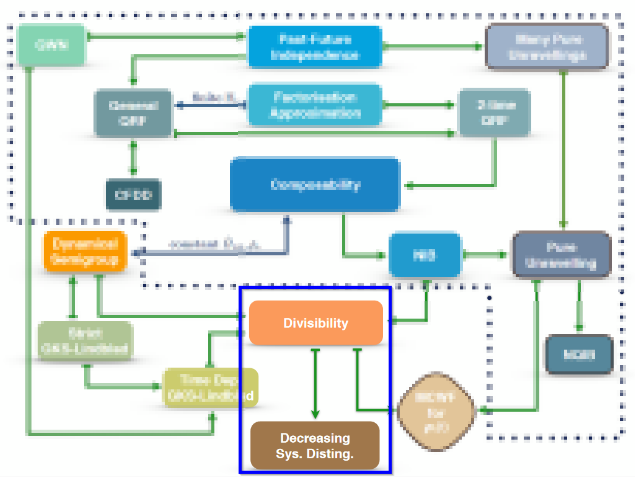
# A “Zoo of Quantum Markovianities”

From: Li Li, Michael J. W. Hall, Howard M. Wiseman. *Concepts of quantum non-Markovianity: a hierarchy*. (arXiv:1712.08879 [quant-ph])



# A “Zoo of Quantum Markovianities”

From: Li Li, Michael J. W. Hall, Howard M. Wiseman. *Concepts of quantum non-Markovianity: a hierarchy*. (arXiv:1712.08879 [quant-ph])

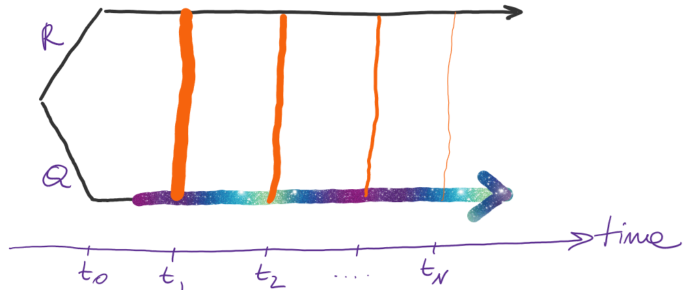


# Divisibility as "Information Flow"

**Theorem**

*Given a mapping* $\left\{\mathcal{N}^{(i)}_{Q_0\to Q_i}\right\}_{i\geq 1}$, *the following are equivalent to divisibility*

1. *for any quantum code, its fidelity is monotonically non-increasing in time*
2. *for any misaligning channel, its quantum realignment fidelity is monotonically non-increasing in time*



# Application to Quantum Thermodynamics

3.5 years ago I presented some ideas (arXiv:1505.00535)

**Second laws as statistical comparisons**

Francesco Buscemi (Nagoya U)

QIT32, Osaka, 26 May 2015

that eventually led to (arXiv:1708.04302)

**Quantum majorization and a complete set of entropic conditions for quantum thermodynamics**

Gilad Gour,[1,2,*] David Jennings,[3,4,†] Francesco Buscemi,[5] Runyao Duan,[6] and Iman Marvian[7]

# Thermal Processes

Formulation of quantum thermodynamics as a "resource theory of out-of-thermal-equilibrium–*ness*"

- thermal (or Gibbs) distribution: $\gamma = \gamma(\hat{H}, T) = Z^{-1} e^{-\hat{H}/kT}$
- thermal processes (Janzing et al, 2000; Horodecki and Oppenheim, 2013) use free thermal ancillas, total energy-preserving interactions, partial traces: $\mathcal{E}_{th}(\rho_S) = \mathrm{Tr}\left[U(\rho_S \otimes \gamma_E)U^\dagger\right]$
- thermal $\implies$ Gibbs-preserving (but $\not\Longleftarrow$ )
- thermal accessibility: $\rho \xrightarrow{th} \sigma$ whenever there exists thermal process $\mathcal{E}_{th}$ such that $\mathcal{E}_{th}(\rho) = \sigma$
- thermal monotone: any function $f(\rho)$ such that $f(\rho) \geq f(\mathcal{E}_{th}(\rho))$ for any thermal process $\mathcal{E}_{th}$ (e.g., the free energy)

# Free Energy as Statistical Distinguishability

- **fact**: the free energy $F(\rho) = \mathrm{Tr}[\rho\ H] - kTS(\rho)$ can be expressed in terms of the quantum relative entropy as $(kT)^{-1}F(\rho) + \log Z = D(\rho\|\gamma)$
- hence, $F(\rho)$ measures the statistical distinguishability of $\rho$ from a thermal background...
- ...and the Second Law is nothing but a data-processing inequality: since $\mathcal{E}_{th}(\gamma) = \gamma$, one has

$$\rho \xrightarrow{th} \sigma \implies D(\rho\|\gamma) \geq D(\sigma\|\gamma) \iff F(\rho) \geq F(\sigma)$$

- **problem**: to find a complete set of generalized "free energies" $F_f(\rho)$ (i.e., generalized divergences $D_f(\rho\|\gamma)$) such that

$$\rho \xrightarrow{th} \sigma \iff D_f(\rho\|\gamma) \geq D_f(\sigma\|\gamma), \forall D_f$$

# Second Laws as Quantum Relative Majorization

- **idea**: to characterize "all" statistical distinguishability measures *at once*...
- ...in terms of one relative majorization relation: $(\rho, \gamma) \succeq (\sigma, \gamma)$
- as the majorization preorder captures the statistical distinguishability of a given PD from a uniform background...
- ...so that thermo-majorization preorder captures thermal accessibility
- however: known results only concern classical PDs
- our result: extension of thermo-majorization to deal with non-commuting density operators

# Conclusions

## Conclusions

- statistical comparison was formulated as a generalization of the "majorization" order
- it constitutes an important foundational tool in mathematical statistics
- in this talk, I argue that it can play an important role also in other areas where statistical predictions are involved
- most importantly, information theory and quantum mechanics
- applications found in quantum statistical mechanics and quantum foundations, but more are waiting!

**Thank you**