Quantum Statistical Inference: Theory & Applications

Michalis Skotiniotis

Quantum Thermodynamics and Computation Group
University of Granada

Quantum Matter Summer School

1st-5th September, 2025

Granada, Spain

Universidad de Granada · Gobierno de España, Ministerio de Ciencia, Innovación y Universidades · Instituto Carlos I de Física Teórica y Computacional · Junta de Andalucía · Unión Europea, Fondo Europeo de Desarrollo Regional · FoQaCiA · Quantum Spain

Outline

- Why Statistical Inference is important
- What is Statistical Inference
- Hypothesis Testing and Parameter Estimation
- Quantum Hypothesis Testing and Parameter Estimation
- Applications

Why statistical Inference is important

Why Statistical Inference is Important

- Medical Testing (COVID tests etc..)
- Quality Control (Product testing...)
- Significance of results (Discovery of Higgs boson, p-values...)
- Reporting the value of a physical constant (g, c, μ.....)

Why Statistical Inference is Important

How to cheat on your Tax Return

- Consider a dataset of numbers. The leading digit of all the numbers in the dataset follows **Benford's Law**

$$p(k) = \log_{10}\left(1 + \frac{1}{k}\right)$$

- The taxperson counts the frequency of first digits in your tax return and performs a χ^2-test

$$\chi^2 = \sum_{k=1}^{9} \frac{(n_k - Np(k))^2}{Np(k)}$$

- If χ^2 is too large, you are cheating.

Why Statistical Inference is Important

The German tank Problem

- The allies were particularly worried about the number of Panzer V tanks in the German army (particularly before D-day)
- Intelligence reports put the number of Panzer V tanks produced per month to be around 1400 (from 1940-1042)
- The Statistical branch of the British RAF was tasked with the problem of figuring out the problem.
- All they had to go on were a **small number** of serial numbers of chassis, gearboxes, and road wheels (stupidly the germans numbered them sequentially)

Why Statistical Inference is Important

The German tank Problem

- The boffins came up with an estimate of **246** tanks per month for the period between June 1940- September 1942

- After the war ended, captured german records from the ministry of Albert Speer revealed that the actual number was **245**

What is statistical Inference

What **is** Statistical Inference

- Statistical Inference is a method of analyzing data in order to
    1. Make decisions
    2. Learn something about the process producing the data
    3. Make predictions about future data
- A key ingredient in Statistical inference is the concept of a **random variable**

Definition: Let A be a set (discreet or continuous). A **random variable**, X, is a function that assigns a value, x, to each element in A. Each value, x, occurs with some probability p(x).

What **is** Statistical Inference

- The value, x, of a random variable, X, is often called a **realization** of the random variable (or a **sample** of the random variable)
- $0 \leq p(x) \leq 1 \quad \forall\, x \in X$ and $\displaystyle\sum_{x \in X} p(x) = 1$
- A realization of size N is denoted as $\mathbf{x} := (x_1, \ldots, x_N) \in X^N$
- If for $\mathbf{x} \in X^N, \quad p(x_i) = p(x_j), \quad \forall x_i,\, x_j \in X$, and the value of x_k is independent of all previous realizations then we say that X is **independent and identically distributed** (iid).

What **is** Statistical Inference

Examples of Random Variables

- Tossing a coin: $X = \{0, 1\}$, $X \sim q$, X is iid
- Measuring the period of a pendulum: $X = \mathbb{R}^+$, $X \sim \mathcal{N}(\mu, \sigma)$, X is iid
- Sampling without replacement: $X_1 = \{R, G\}, X_1 \sim 1/2$

  $X_2 = \{R, G\}, X_2 \sim 1/2 - \epsilon$, not iid

Hypothesis Testing and Parameter Estimation

Binary hypothesis testing

- Suppose that we have a Binary random variable $X \in \{0, 1\}$ which is known to be distributed **either** according to $X \sim \text{Bin}(1, p)$ or $X \sim \text{Bin}(1, q)$

- We call $X \sim \text{Bin}(1, p)$ the **null hypothesis** (H_0), and $X \sim \text{Bin}(1, q)$ the **alternative hypothesis** (H_1)

- Given a realization of the iid random variable $X \in \{0, 1\}$ determine which of the two hypothesis is true

Binary hypothesis testing

Definition: A **decision rule** $f : X^{\times N} \to \{H_0, H_1\}$ is a rule that decides whether we accept or reject a hypothesis.

- The outcome $f(\mathbf{x}) \in \{H_0, H_1\}$ is our **decision** as to the underlying hypothesis given we observe $\mathbf{x} \in X^{\times N}$

Guess \ True	H_0	H_1
$\hat{H}_0$	$1-\alpha$	β
$\hat{H}_1$	α	$1-\beta$

Binary hypothesis testing

Definition: A **decision rule** $f : X^{\times N} \to \{H_0, H_1\}$ is a rule that decides whether we accept or reject a hypothesis.

- **Accepting** H_1 when H_0 is true is called a **type-I error** (or a false positive)

Guess \ True	H_0	H_1
$\hat{H}_0$	$1-\alpha$	β
$\hat{H}_1$	α	$1-\beta$

Binary hypothesis testing

Definition: A **decision rule** $f : X^{\times N} \to \{H_0, H_1\}$ is a rule that decides whether we accept or reject a hypothesis.

- **Accepting** H_1 when H_0 is true is called a **type-I error** (or a false positive)
- **Accepting** H_0 when H_1 is true is called a **type-II error** (or a false negative)

Guess \ True	H_0	H_1
$\hat{H}_0$	$1-\alpha$	β
$\hat{H}_1$	α	$1-\beta$

Binary hypothesis testing

Definition: A **decision rule** $f : X^{\times N} \to \{H_0, H_1\}$ is a rule that decides whether we accept or reject a hypothesis.

- **Accepting** H_1 when H_0 is true is called a **type-I error** (or a false positive)
- **Accepting** H_0 when H_1 is true is called a **type-II error** (or a false negative)
- **Rejecting** H_0 when H_1 is true is the **power** of our decision.

Guess \ True	H_0	H_1
$\hat{H}_0$	$1-\alpha$	β
$\hat{H}_1$	α	$1-\beta$

Binary hypothesis testing

- We seek the decision rule $f : X^{\times N} \to \{H_0, H_1\}$ than maximizes the power for a fixed rate of false positives

Guess \ True	H_0	H_1
$\hat{H}_0$	$1-\alpha$	β
$\hat{H}_1$	α	$1-\beta$

Neyman-Pearson: The optimal decision rule is **maximum-likelihood**

$$f(\mathbf{x}) = \begin{cases} H_0 & \text{if } \frac{p(\mathbf{x}|H_0)}{p(\mathbf{x}|H_1)} > \eta \\ H_1 & \text{if } \frac{p(\mathbf{x}|H_0)}{p(\mathbf{x}|H_1)} < \eta \\ \text{either} & \text{if } \frac{p(\mathbf{x}|H_0)}{p(\mathbf{x}|H_1)} = \eta \end{cases}$$

and $1-\alpha = \sum_{\mathbf{x}\in A(\eta)} p(\mathbf{x}|H_0)$ with $A(\eta) := \{\mathbf{x} \in X^{\times N} \,|\, p(\mathbf{x}|H_0) > \eta p(\mathbf{x}|H_1)\}$

Binary hypothesis testing

Tossing a Coin

- Suppose that all €2 coins look identical
- Suppose that all €2 coins are minted either in Spain (H_0) or in Greece (H_1).
- Suppose that 65% of all €2 coins are minted in Spain and 35% in Greece
- A coin coss corresponds to the iid binary random variable $X \in \{0, 1\}$ with

$$p(x|H_0) = a \quad \text{or} \quad p(x|H_1) = b$$

Binary hypothesis testing

Tossing a Coin

Goal: Correctly identify the coin after a **finite** number of tosses

- Mathematically the goal is captured by the **average probability of success**

$$P_S = \Pr(\hat{H}_0|H_0)\,\pi_0 + \Pr(\hat{H}_1|H_1)\,\pi_1$$

- Or **equivalently**

$$P_E = \Pr(\hat{H}_1|H_0)\,\pi_0 + \Pr(\hat{H}_0|H_1)\,\pi_1$$

Binary hypothesis testing

Tossing a Coin

- Observe that for this game type-I and type-II errors are penalised **equally** (symmetric Hypothesis testing)

- Given a realization $\mathrm{x} \in X^{\times N}$ how do you maximize the probability of success?

Binary hypothesis testing

Tossing a Coin

- Observe that for this game type-I and type-II errors are penalised **equally** (symmetric Hypothesis testing)

- Given a realization $\mathbf{x} \in X^{\times N}$ how do you maximize the probability of success?

$$f(\mathbf{x}) = \begin{cases} \hat{H}_0 & \text{if} \pi_0 p(\mathbf{x}|H_0) > \pi_1 p(\mathbf{x}|H_1) \\ \hat{H}_1 & \text{if} \pi_0 p(\mathbf{x}|H_0) < \pi_1 p(\mathbf{x}|H_1) \\ \text{either} & \text{if} \pi_0 p(\mathbf{x}|H_0) = \pi_1 p(\mathbf{x}|H_1) \end{cases}$$

Binary hypothesis testing

Tossing a Coin

- The probability that you win is

$$P_E = \frac{1}{2}\left(1 - \|\mathbf{p}_0\pi_0 - \mathbf{p}_1\pi_1\|\right)$$

where

$$\mathbf{p}_k = \begin{pmatrix} p(0|H_k) \\ p(1|H_k) \end{pmatrix}$$

and

$$\frac{1}{2}\|\mathbf{p}_0\pi_0 - \mathbf{p}_1\pi_1\| = \frac{1}{2}\sum_{\mathbf{x}\in\mathbf{X}^{\times\mathbf{N}}} |p(\mathbf{x}|H_0)\pi_0 - p(\mathbf{x}|H_1)\pi_1|$$

is the **trace-norm** distance

Binary hypothesis testing

- As the number of observations increases the probability of making an error satisfies

$$P_E \sim e^{-rN}$$

- The **rate** r depends on the scenario

1. **Symmetric Hypothesis Testing**

$$r = -\min_{0\leq\lambda\leq 1} \log\left(\sum_{x\in X} p(x|H_0)^{\lambda}\, p(x|H_1)^{1-\lambda}\right)$$

Chernoff rate

2. **Asymmetric Hypothesis Testing**

$$r = \sum_{x\in X} p(x|H_0)\,\log\left(\frac{p(x|H_0)}{p(x|H_1)}\right) := D(\mathbf{p}_0\|\mathbf{p}_1)$$

Stein rate

1. Cover T. & Thomas J. Elements of Information Theory

Parameter Estimation

- What if the number of hypothesis form a continuous set $\Theta \subseteq \mathbb{R}^N$?
  1. $X \in \{0,1\}, \quad X \sim \mathrm{Bin}(1,\theta), \quad \theta \in (0,1)$
  2. $X \in \mathbb{R}, \quad X \sim \mathcal{N}(\mu,\sigma), \quad \mu \in \mathbb{R}, \sigma \in \mathbb{R}_+$
- Our realizations, $\mathbf{x} \in X$ are distributed according to
  1. $p(x|\theta) = \begin{cases} 1-\theta & \text{if } x = 0 \\ \theta & \text{if } x = 1 \end{cases}$
  2. $p(x|\theta) = \frac{1}{\sqrt{2\pi\theta_2^2}} \exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right)$
- Our aim is to determine the parameter(s) $\theta \in \Theta$, based on our observations

Parameter Estimation

The Frequentists

- There is one and only one **true** value $\theta \in \Theta$

The Bayesians

- Θ is itself a random variable
- The distribution of this RV is **subjective**

I will focus mostly on the **frequentist** approach as it is the most widely used

Parameter Estimation

The Frequentists

- There is one and only one **true** value $\theta \in \Theta$

The Bayesians

- Θ is itself a random variable
- The distribution of this RV is **subjective**

I will focus mostly on the **frequentist** approach as it is the most widely used

But I am **not** a frequentist

Parameter Estimation

The Frequentists

- There is one and only one **true** value $\theta \in \Theta$

The Bayesians

- Θ is itself a random variable
- The distribution of this RV is **subjective**

I will focus mostly on the **frequentist** approach as it is the most widely used

But I am **not** a frequentist

And with age.....I am slowly coming round to **not** being a Bayesian either...

Parameter Estimation

Definition: An **estimator** $f : X^N \to \Theta$ is a function that assigns to every $\mathbf{x} \in X^N$ an **estimate** $f(\mathbf{x}) \in \Theta$

A good estimator must satisfy

(i) Unbiasedness: Our estimator must yield the true value **on average**

$$\langle \hat{\theta} \rangle := \int_{X^N} \mathrm{d}^N \mathbf{x}\, p(\mathbf{x}|\theta) f(\mathbf{x}) = \theta$$

2. S. S. Wilks. Mathematical Statistics

Parameter Estimation

Definition: An **estimator** $f : X^N \rightarrow \Theta$ is a function that assigns to every $\mathbf{x} \in X^N$ an **estimate** $f(\mathbf{x}) \in \Theta$

A good estimator must satisfy

(ii) Consistency: It must converge to the true value **in probability.**

For any $\delta > 0$ and sequence of estimates $f^{(k)}(\mathbf{x})$

$$\lim_{k \to \infty} \Pr\left(\left|f^{(k)}(\mathbf{x}) - \theta\right| > \delta\right) = 0$$

2. S. S. Wilks. Mathematical Statistics

Parameter Estimation

Definition: An **estimator** $f : X^N \to \Theta$ is a function that assigns to every $\mathbf{x} \in X^N$ an **estimate** $f(\mathbf{x}) \in \Theta$

A good estimator must satisfy

(iii) Precision

$$\mathrm{Cov}(f)_{jk} := \int_{X^N} \mathrm{d}^N\mathbf{x}(f_j(\mathbf{x}) - \theta_{\mathbf{j}})(\mathbf{f_k}(\mathbf{x}) - \theta_{\mathbf{k}})$$

The **Covariance matrix** must be small (in some norm)

2. S. S. Wilks. Mathematical Statistics

Parameter Estimation

Definition: An **estimator** $f : X^N \to \Theta$ is a function that assigns to every $\mathbf{x} \in X^N$ an **estimate** $f(\mathbf{x}) \in \Theta$

A good estimator must satisfy

(iv) Efficiency: For any other estimator $g : X^N \to \Theta$ it holds

$$\mathrm{Cov}(f)\mathrm{Cov}^{-1}(g) < \mathbb{1}$$

2. S. S. Wilks. Mathematical Statistics

Parameter Estimation

Cramér-Rao: Let $X \sim p(x|\boldsymbol{\theta})$. For **any** unbiased estimator $f : X^N \to \Theta$ it holds

$$\mathrm{Cov}(f) \geq \mathbf{F}^{-1}[p(x|\boldsymbol{\theta})]$$

where $\mathbf{F}[p(x|\boldsymbol{\theta})]$ is the **Fisher Information** matrix

$$\mathbf{F}_{jk}[p(x|\boldsymbol{\theta})] := \int_X \mathrm{d}x \frac{1}{p(x|\boldsymbol{\theta})} \frac{\mathrm{d}p(x|\boldsymbol{\theta})}{\mathrm{d}\theta_j} \frac{\mathrm{d}p(x|\boldsymbol{\theta})}{\mathrm{d}\theta_k}$$

2. S. S. Wilks. Mathematical Statistics

Parameter Estimation

Cramér-Rao: Let $X \sim p(x|\boldsymbol{\theta})$. For **any** unbiased estimator $f : X^N \to \Theta$ it holds

$$\mathrm{Cov}(f) \geq \mathbf{F}^{-1}[p(x|\boldsymbol{\theta})]$$

where $\mathbf{F}[p(x|\boldsymbol{\theta})]$ is the **Fisher Information** matrix

$$\mathbf{F}_{jk}[p(x|\boldsymbol{\theta})] := \int_X \mathrm{d}x \frac{1}{p(x|\boldsymbol{\theta})} \frac{\mathrm{d}p(x|\boldsymbol{\theta})}{\mathrm{d}\theta_j} \frac{\mathrm{d}p(x|\boldsymbol{\theta})}{\mathrm{d}\theta_k}$$

Remarks:

1. The matrix inequality is to be understood as

$$\mathbf{v} \cdot (\mathrm{Cov}(f) - \mathbf{F}[p(x|\boldsymbol{\theta})]) \cdot \mathbf{v} \geq 0 \qquad \forall\, \mathbf{v} \in \mathbb{R}^{|\Theta|}$$

2. S. S. Wilks. Mathematical Statistics

Parameter Estimation

Cramér-Rao: Let $X \sim p(x|\boldsymbol{\theta})$. For **any** unbiased estimator $f : X^N \to \Theta$ it holds

$$\mathrm{Cov}(f) \geq \mathbf{F}^{-1}[p(x|\boldsymbol{\theta})]$$

where $\mathbf{F}[p(x|\boldsymbol{\theta})]$ is the **Fisher Information** matrix

$$\mathbf{F}_{jk}[p(x|\boldsymbol{\theta})] := \int_X \mathrm{d}x \frac{1}{p(x|\boldsymbol{\theta})} \frac{\mathrm{d}p(x|\boldsymbol{\theta})}{\mathrm{d}\theta_j} \frac{\mathrm{d}p(x|\boldsymbol{\theta})}{\mathrm{d}\theta_k}$$

Remarks:

2. The Fisher Information matrix can also be written as the covariance of the score function

$$\mathbf{F}_{jk}[p(x|\boldsymbol{\theta})] := \int_X \mathrm{d}x \, p(x|\boldsymbol{\theta}) \frac{\mathrm{d}\log(p(x|\boldsymbol{\theta}))}{\mathrm{d}\theta_j} \frac{\mathrm{d}\log(p(x|\boldsymbol{\theta}))}{\mathrm{d}\theta_k}$$

2. S. S. Wilks. Mathematical Statistics

Parameter Estimation

Cramér-Rao: Let $X \sim p(x|\boldsymbol{\theta})$. For **any** unbiased estimator $f : X^N \to \Theta$ it holds

$$\mathrm{Cov}(f) \geq \mathbf{F}^{-1}[p(x|\boldsymbol{\theta})]$$

where $\mathbf{F}[p(x|\boldsymbol{\theta})]$ is the **Fisher Information** matrix

$$\mathbf{F}_{jk}[p(x|\boldsymbol{\theta})] := \int_X \mathrm{d}x \frac{1}{p(x|\boldsymbol{\theta})} \frac{\mathrm{d}p(x|\boldsymbol{\theta})}{\mathrm{d}\theta_j} \frac{\mathrm{d}p(x|\boldsymbol{\theta})}{\mathrm{d}\theta_k}$$

Remarks:

2. The Fisher Information matrix quantifies the susceptibility of the score function with respect to the parameter $\boldsymbol{\theta} \in \Theta$

2. S. S. Wilks. Mathematical Statistics

Parameter Estimation

Exercises:

1. Binomial Distribution

$$F[p(x|\theta)] = \frac{1}{\theta(1-\theta)}$$

2. Normal Distribution

$$F[p(x|\boldsymbol{\theta})] = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

Quantum Hypothesis Testing and Parameter Estimation

$$p(x|\boldsymbol{\theta}) = \mathrm{tr}\left(E_x \rho(\boldsymbol{\theta})\right)$$

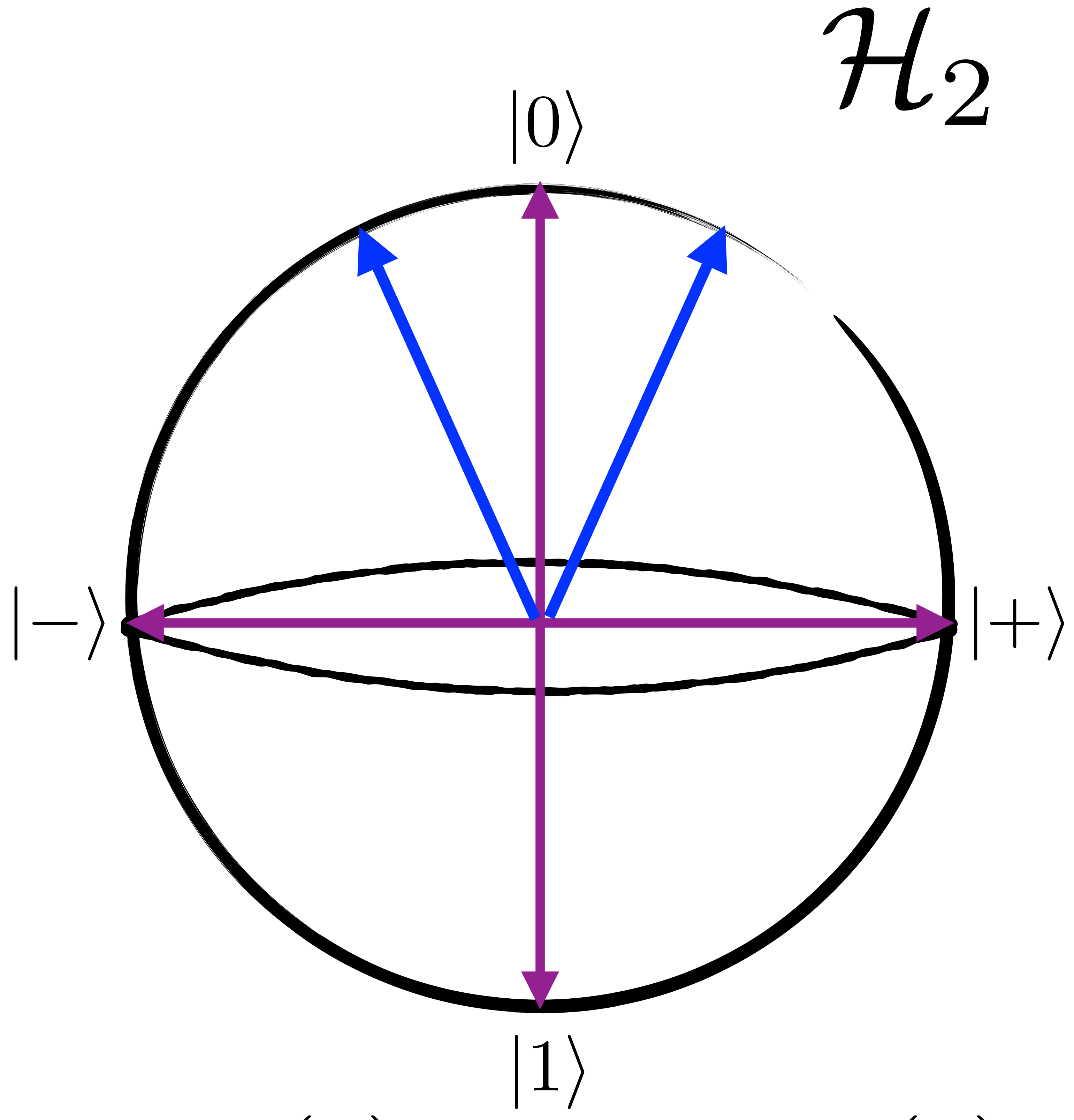
$$\{E_x > 0 \,|\, \sum_x E_x = \mathbb{1}\}$$

In Quantum Theory we can make manifest any probability distribution we so desire

Quantum Hypothesis Testing

State Discrimination

- Suppose that a friend prepares a spin-1/2 particle in one of two possible states $|\psi_0\rangle$, $|\psi_1\rangle$
- But they forgot which state they prepared it in
- All they know is that their equipment prepares $|\psi_0\rangle$ with probability π_0 and $|\psi_1\rangle$ with probability π_1
- Your job is to figure out the identity of the state



$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{-i\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle$$

Quantum Hypothesis Testing

State Discrimination

- The probability of error is still the same as before

$$P_E = \frac{1}{2}\left(1 - \|\mathbf{p}_0\pi_0 - \mathbf{p}_1\pi_1\|\right)$$

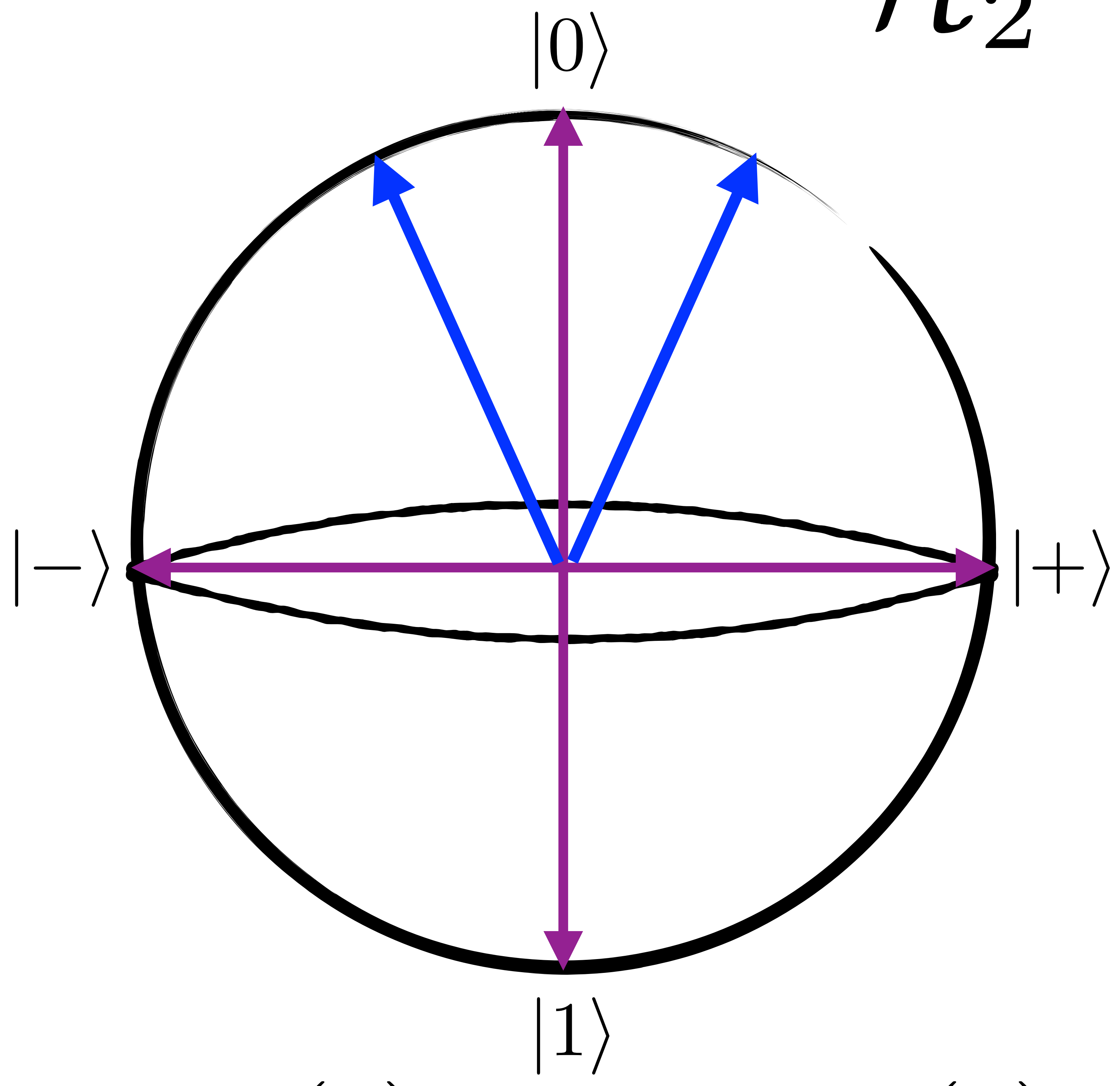
$$\frac{1}{2}\|\mathbf{p}_0\pi_0 - \mathbf{p}_1\pi_1\| = \frac{1}{2}\sum_{\mathbf{x}\in\mathbf{X}^{\times\mathbf{N}}} |p(\mathbf{x}|H_0)\pi_0 - p(\mathbf{x}|H_1)\pi_1|$$

only this time

$$p(x|H_0) = \mathrm{tr}(E_x\,|\psi_0\rangle\langle\psi_0|) = \langle\psi_0|E_x|\psi_0\rangle$$
$$p(x|H_1) = \mathrm{tr}(E_x\,|\psi_1\rangle\langle\psi_1|) = \langle\psi_1|E_x|\psi_1\rangle$$

- The optimal decision rule is again **maximum likelihood**

$\mathcal{H}_2$



$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{-i\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle$$

Quantum Hypothesis Testing

State Discrimination

A little bit of algebra gives

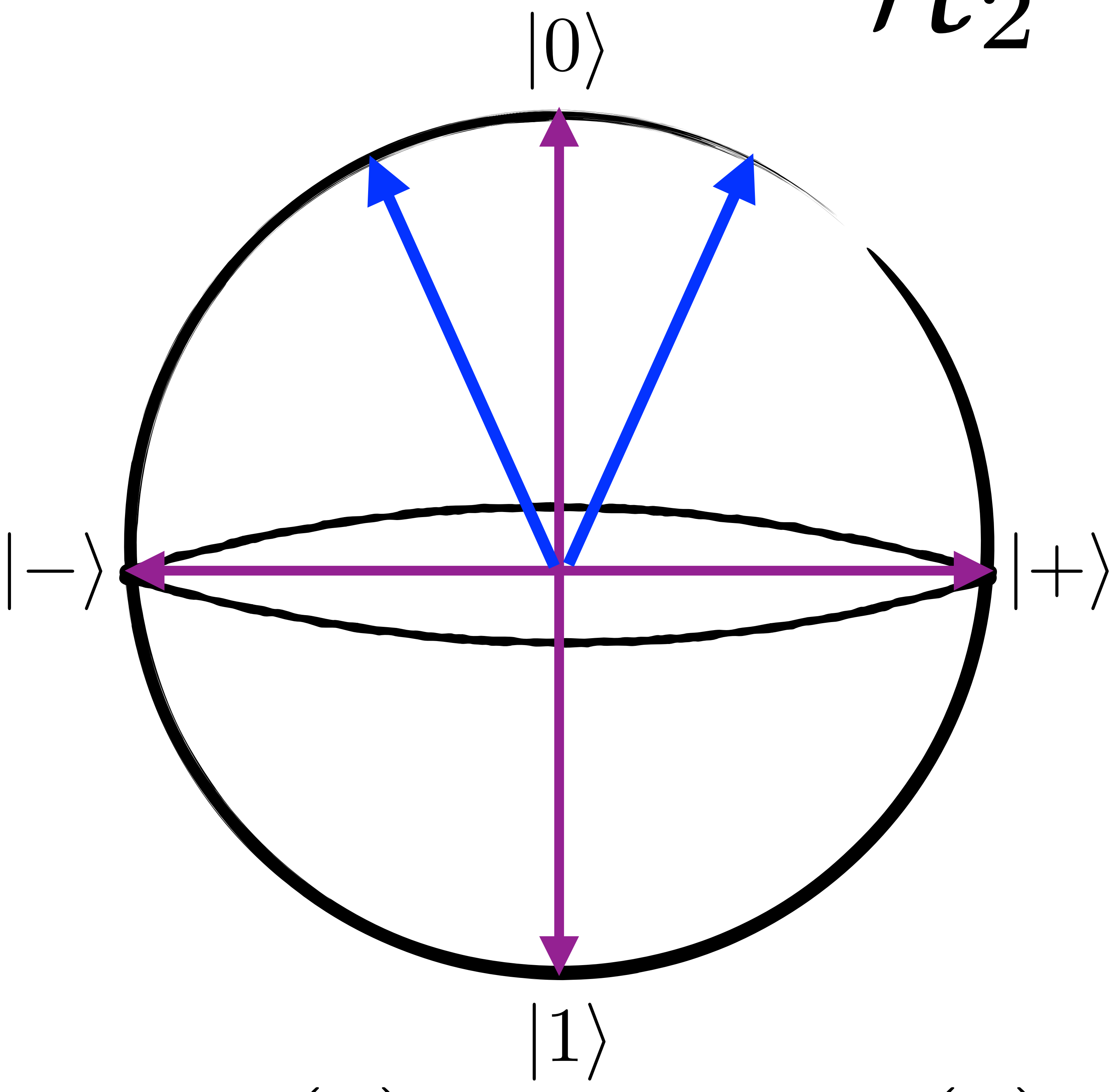
$$P_E = \frac{1}{2}\left(1 - \sum_{x=0}^{1} \left|\mathrm{tr}\left(E_x \left|\psi_0\right\rangle\!\left\langle\psi_0\right| \pi_0\right) - \mathrm{tr}\left(E_x \left|\psi_1\right\rangle\!\left\langle\psi_1\right| \pi_1\right)\right|\right)$$

$$= \frac{1}{2}\left(1 - \sum_{x=0}^{1} \mathrm{tr}\left|E_x \left(\left|\psi_0\right\rangle\!\left\langle\psi_0\right| \pi_0 - \left|\psi_1\right\rangle\!\left\langle\psi_1\right| \pi_1\right)\right|\right)$$

$$= \frac{1}{2}\left(1 - \mathrm{tr}\sum_{x=0}^{1} E_x\left|\left(\left|\psi_0\right\rangle\!\left\langle\psi_0\right| \pi_0 - \left|\psi_1\right\rangle\!\left\langle\psi_1\right| \pi_1\right)\right|\right)$$

$$= \frac{1}{2}\left(1 + \left\| \left|\psi_0\right\rangle\!\left\langle\psi_0\right| \pi_0 - \left|\psi_1\right\rangle\!\left\langle\psi_1\right| \pi_1\right\|\right)$$

$\mathcal{H}_2$



$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{-i\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle$$

Quantum Hypothesis Testing

State Discrimination

$$\frac{1}{2}\left(1 - \mathrm{tr}\sum_{x=0}^{1} E_x \left| \left(|\psi_0\rangle\langle\psi_0|\ \pi_0 - |\psi_1\rangle\langle\psi_1|\ \pi_1 \right) \right| \right)$$

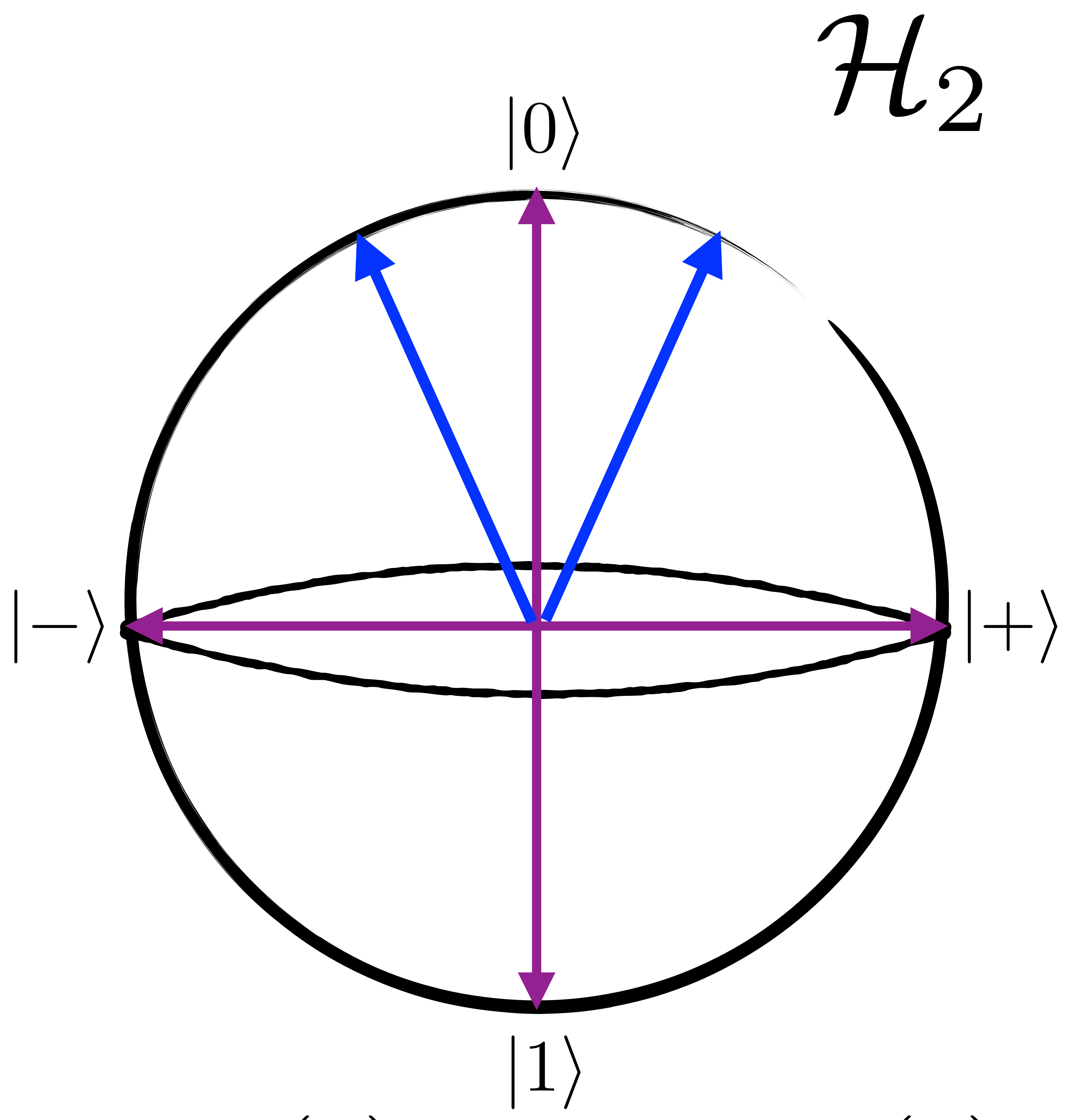
The optimal measurement consists of **projectors** onto the +ve and -ve eigenspaces of

$$\Gamma := |\psi_0\rangle\langle\psi_0|\ \pi_0 - |\psi_1\rangle\langle\psi_1|\ \pi_1$$

and

$$\|\Gamma\| = \frac{1}{2}\mathrm{tr}\left(\sqrt{\Gamma\,\Gamma^\dagger}\right)$$

is the **trace-distance**



$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{-i\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle$$

3. C. W. Helstrom. Quantum Detection and Estimation Theory

Quantum Hypothesis Testing

Holevo's Conditions: Let $\{\pi_k, \rho_k\}_{k=1}^{N}$ be the set of states we wish to discriminate. The **POVM** $\{E_k \geq 0 \,|\, \sum_k E_k = \mathbb{1}\}$ is optimal **if and only if**

$$\Gamma - \pi_k \rho_k \geq 0 \quad \forall\, k \in \{1, \ldots, n\}$$

$$\Gamma = \sum_k \pi_k \rho_k \, E_k$$

The corresponding probability of success is given by

$$P_S = \mathrm{tr}\Gamma$$

Quantum Hypothesis Testing

Holevo's Conditions: Let $\{\pi_k, \rho_k\}_{k=1}^{N}$ be the set of states we wish to discriminate. The **POVM** $\{E_k \geq 0 \,|\, \sum_k E_k = \mathbb{1}\}$ is optimal **if and only if**

$$\Gamma - \pi_k \rho_k \geq 0 \quad \forall\, k \in \{1, \ldots, n\}$$

$$\Gamma = \sum_k \pi_k \rho_k \, E_k$$

The corresponding probability of success is given by

$$P_S = \mathrm{tr}\Gamma$$

Remark: You can very easily prove this theorem yourselves. (Physicsits) This is a Lagrange multipliers problem. (Everyone else) This is a **Semi Definite Program**

4. A. S. Holevo. Probabilistic and Statistical Aspects of Quantum Theory

Quantum Statistical Inference

Quantum Crámer-Rao

$$\mathbf{F}_{jk}[p(x|\boldsymbol{\theta})] := \int_X \mathrm{d}x \frac{1}{p(x|\boldsymbol{\theta})} \frac{\mathrm{d}p(x|\boldsymbol{\theta})}{\mathrm{d}\theta_j} \frac{\mathrm{d}p(x|\boldsymbol{\theta})}{\mathrm{d}\theta_k}$$

Again we substitute

$$p(x|\boldsymbol{\theta}) = \mathrm{tr}\left(E_x \rho(\boldsymbol{\theta})\right)$$

so that

$$\frac{\mathrm{d}p(x|\boldsymbol{\theta})}{\mathrm{d}\theta_j} = \mathrm{tr}\left(E_x \frac{\mathrm{d}\rho(\boldsymbol{\theta})}{\mathrm{d}\theta_j}\right)$$

5. M. G. A. Paris. Int. Jour. Quantum. Info. **7** (supp01), 125
6. J. S. Sidhu & P. Kok. AVS Quantum Science. **2**, 014701

Quantum Statistical Inference

Quantum Crámer-Rao

The **Symmetric Logarithmic Derivative**

$$\frac{\mathrm{d}\rho(\boldsymbol{\theta})}{\mathrm{d}\theta_j} := \frac{L_{\theta_j}\rho(\boldsymbol{\theta}) + \rho(\boldsymbol{\theta})L_{\theta_j}}{2} \qquad L_{\theta_j} \in \mathrm{Herm}(\mathcal{H})$$

Doing a bit of algebra gives

$$\frac{\mathrm{d}p(x|\boldsymbol{\theta})}{\mathrm{d}\theta_j} = \mathrm{Re}\left(\mathrm{tr}\left(E_x L_{\theta_j}\rho(\boldsymbol{\theta})\right)\right)$$

5. M. G. A. Paris. Int. Jour. Quantum. Info. **7** (supp01), 125
6. J. S. Sidhu & P. Kok. AVS Quantum Science. **2**, 014701

Quantum Statistical Inference

Quantum Crámer-Rao

$$\mathbf{F}_{jk}[p(x|\boldsymbol{\theta})] := \int_X \mathrm{d}x \frac{1}{p(x|\boldsymbol{\theta})} \frac{\mathrm{d}p(x|\boldsymbol{\theta})}{\mathrm{d}\theta_j} \frac{\mathrm{d}p(x|\boldsymbol{\theta})}{\mathrm{d}\theta_k}$$

Plugging it all in and using the Schwarz inequality yields the **Quantum Fisher Information** (QFI)

$$\mathcal{F}_{ij}[\rho(\boldsymbol{\theta})] = \mathrm{tr}\left(L_{\theta_i}\rho(\boldsymbol{\theta})L_{\theta_j}\right)$$

and the Quantum Crámer-Rao bound

$$\mathrm{Cov}(f) \geq \mathbf{F}^{-1}[p(x|\boldsymbol{\theta}] \geq \boldsymbol{\mathcal{F}}^{-1}[\rho(\boldsymbol{\theta})]$$

5. M. G. A. Paris. Int. Jour. Quantum. Info. **7** (supp01), 125
6. J. S. Sidhu & P. Kok. AVS Quantum Science. **2**, 014701

Quantum Statistical Inference

$$\mathrm{Cov}(f) \geq \mathbf{F}^{-1}[p(x|\boldsymbol{\theta})] \geq \boldsymbol{\mathcal{F}}^{-1}[\rho(\boldsymbol{\theta})]$$

Remarks:

1. For each parameter $\theta_j \in \Theta$ the measurement that **saturates** the QFI $\mathcal{F}_{jj}[\rho(\boldsymbol{\theta})]$ is a **projective measurement** on the eigenspaces of the SLD L_{θ_j}

2. Unlike the classical case the multi-parameter Quantum Crámer-Rao bound is not always achievable

$$[L_{\theta_i}, L_{\theta_j}] \neq 0$$

3. A necessary and sufficient condition for attainability is

$$\mathrm{tr}\left([L_{\theta_i}, L_{\theta_j}]\rho(\boldsymbol{\theta})\right) = 0$$

Quantum Statistical Inference

Geometric Interpretion of the QFI

1. Maximize the Fisher Information over all allowable measurements

$$\mathcal{F}[\rho(\boldsymbol{\theta})] := \max_{\{E_x \geq 0 \mid \sum_x E_x = \mathbb{1}\}} \mathbf{F}[p(x|\boldsymbol{\theta})]$$

2. It is the **infidelity** between two infinitesimally close states

$$\mathcal{F}[\rho(\boldsymbol{\theta})] := 8\,\frac{1 - F(\rho(\boldsymbol{\theta}), \rho(\boldsymbol{\theta} + \mathrm{d}\boldsymbol{\theta}))}{\mathrm{d}\boldsymbol{\theta}} \qquad F(\rho, \sigma) = \mathrm{tr}\sqrt{\sqrt{\rho}, \sigma\sqrt{\rho}}$$

4. Its square root is proportional to the susceptibility of the **Bures Angle**

$$\mathcal{F}^{\frac{1}{2}}[\rho(\boldsymbol{\theta})] = 2\left(\frac{A(\rho(\boldsymbol{\theta}), \rho(\boldsymbol{\theta} + \mathrm{d}\boldsymbol{\theta}))}{\mathrm{d}\boldsymbol{\theta}}\right) \qquad A(\rho(\boldsymbol{\theta}), \rho(\boldsymbol{\theta} + \mathrm{d}\boldsymbol{\theta})) = \cos^{-1}(F(\rho(\boldsymbol{\theta}), \rho(\boldsymbol{\theta} + \mathrm{d}\boldsymbol{\theta})))$$

Quantum Statistical Inference

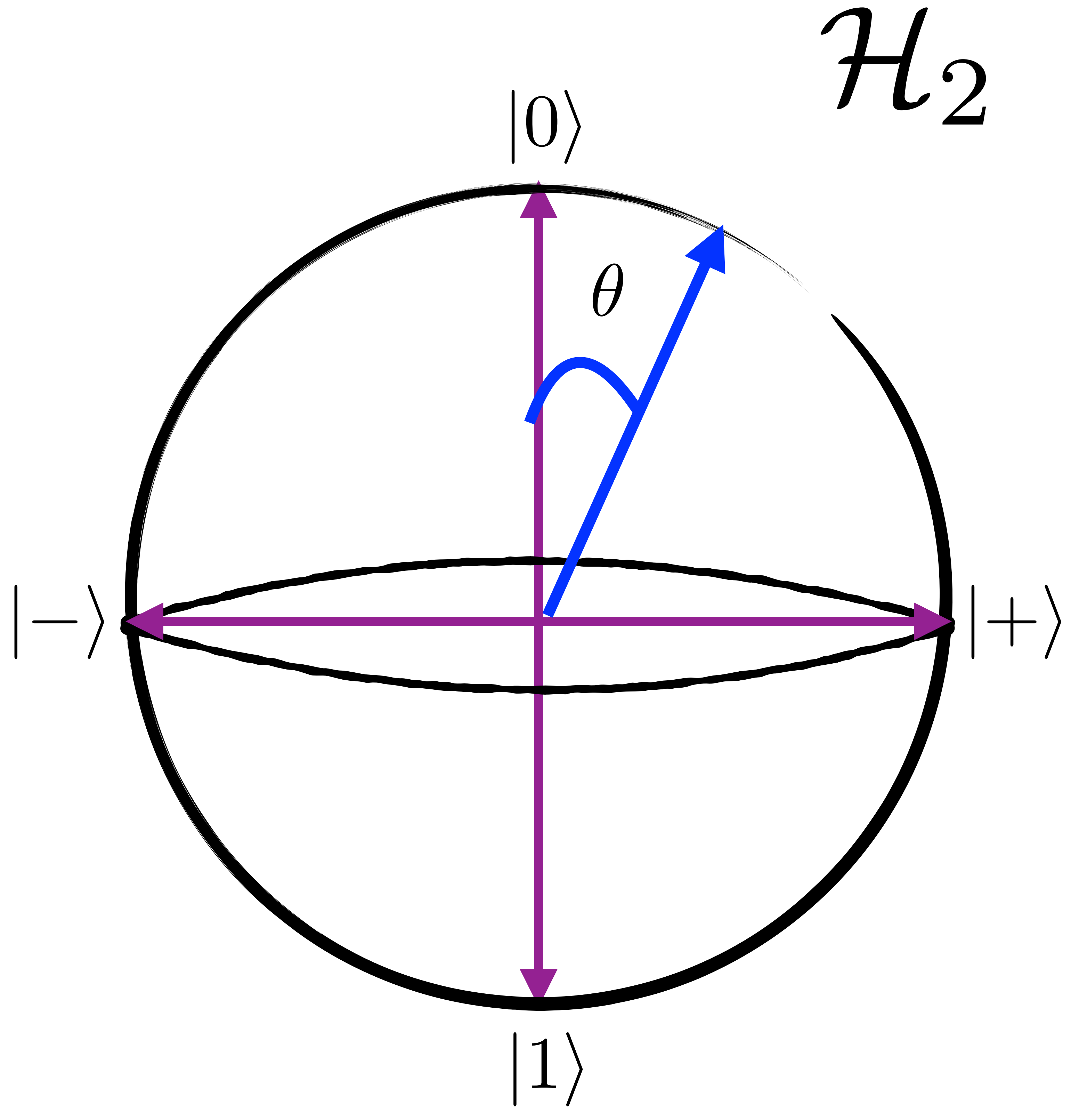
The case of Pure States

- Suppose we wish to estimate the angle $\theta \in (0, \pi)$ encoded in the state of a spin- $^1/_2$ system
- Observe that $\rho(\theta) = |\psi(\theta)\rangle\langle\psi(\theta)| = \rho(\theta)^2$ so that

$$\frac{\mathrm{d}\rho(\theta)}{\mathrm{d}\theta} = \frac{\mathrm{d}}{\mathrm{d}\theta}(\rho(\theta)^2) = \frac{\mathrm{d}\rho(\theta)}{\mathrm{d}\theta}\rho(\theta) + \rho(\theta)\frac{\mathrm{d}\rho(\theta)}{\mathrm{d}\theta}$$

- Recalling that $\frac{\mathrm{d}\rho(\boldsymbol{\theta})}{\mathrm{d}\theta_j} := \frac{L_{\theta_j}\rho(\boldsymbol{\theta}) + \rho(\boldsymbol{\theta})L_{\theta_j}}{2}$ it follows that

$$L_\theta = 2\frac{\mathrm{d}\rho(\theta)}{\mathrm{d}\theta} = 2\left(\frac{\mathrm{d}|\psi(\theta)\rangle}{\mathrm{d}\theta}\langle\psi(\theta)| + |\psi(\theta)\rangle\frac{\mathrm{d}\langle\psi(\theta)|}{\mathrm{d}\theta}\right)$$



$$|\psi(\theta)\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + \sin\left(\frac{\theta}{2}\right)|1\rangle$$

Quantum Statistical Inference

The case of Pure States

- Thus, the QFI is

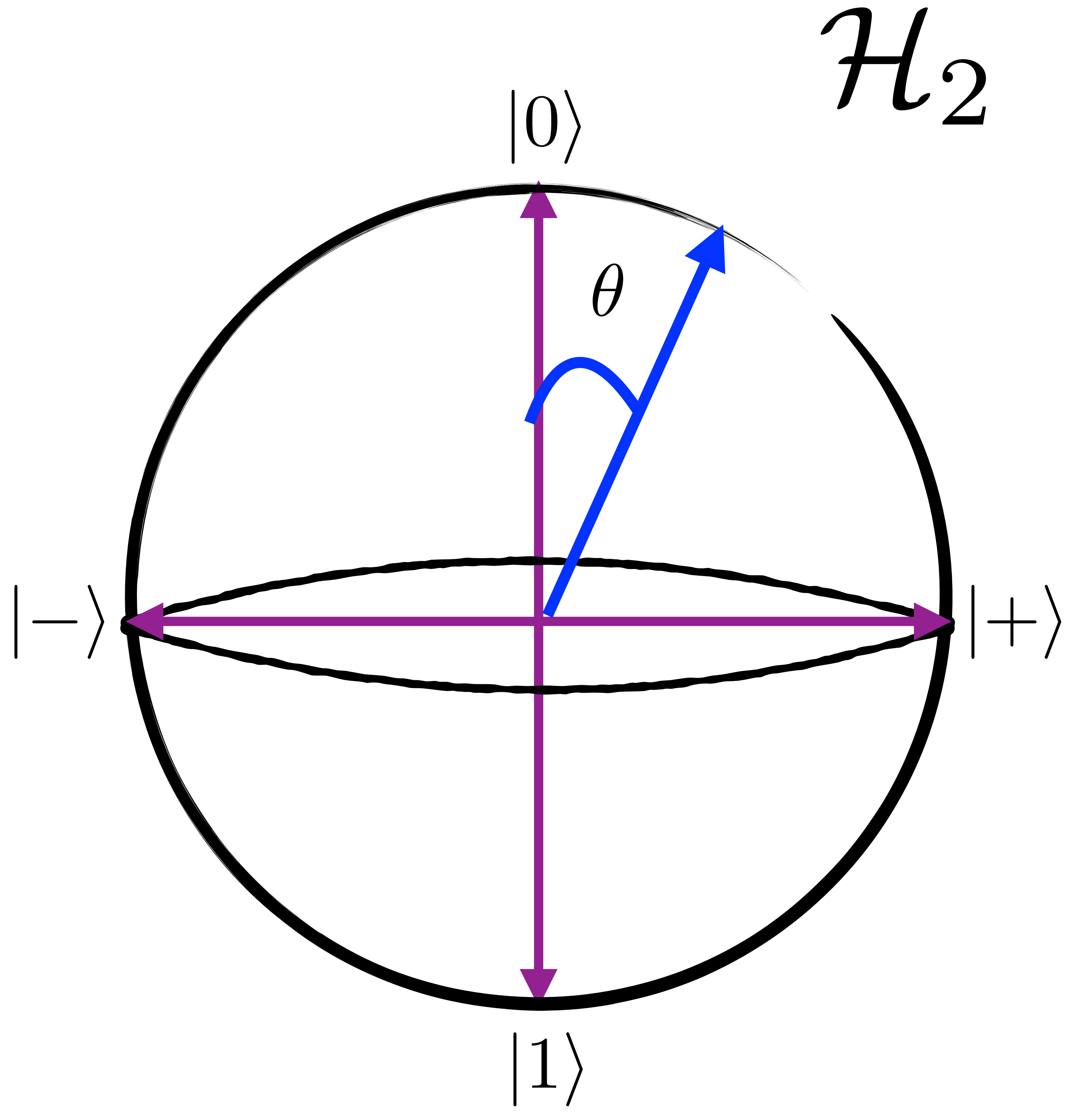
$$\mathcal{F}(\rho(\theta)) = \mathrm{tr}\left(L_\theta \rho(\theta) L_\theta\right)$$
$$= 4\left(\left\langle \dot{\psi(\theta)} | \dot{\psi(\theta)} \right\rangle + (\left\langle \dot{\psi(\theta)} \psi(\theta) | \right\rangle^2\right)$$

- If, furthermore, we notice that

$$|\psi(\theta)\rangle = \begin{pmatrix} \cos \theta/2 & -\sin \theta/2 \\ \sin \theta/2 & \cos \theta/2 \end{pmatrix} |0\rangle = e^{-\mathrm{i}\frac{\theta}{2}\sigma_y} |0\rangle$$

Then

$$\mathcal{F}(\rho(\theta)) = \langle 0|\sigma_y^2|0\rangle - \langle 0|\sigma_y|0\rangle^2 = \Delta^2 \sigma_y$$

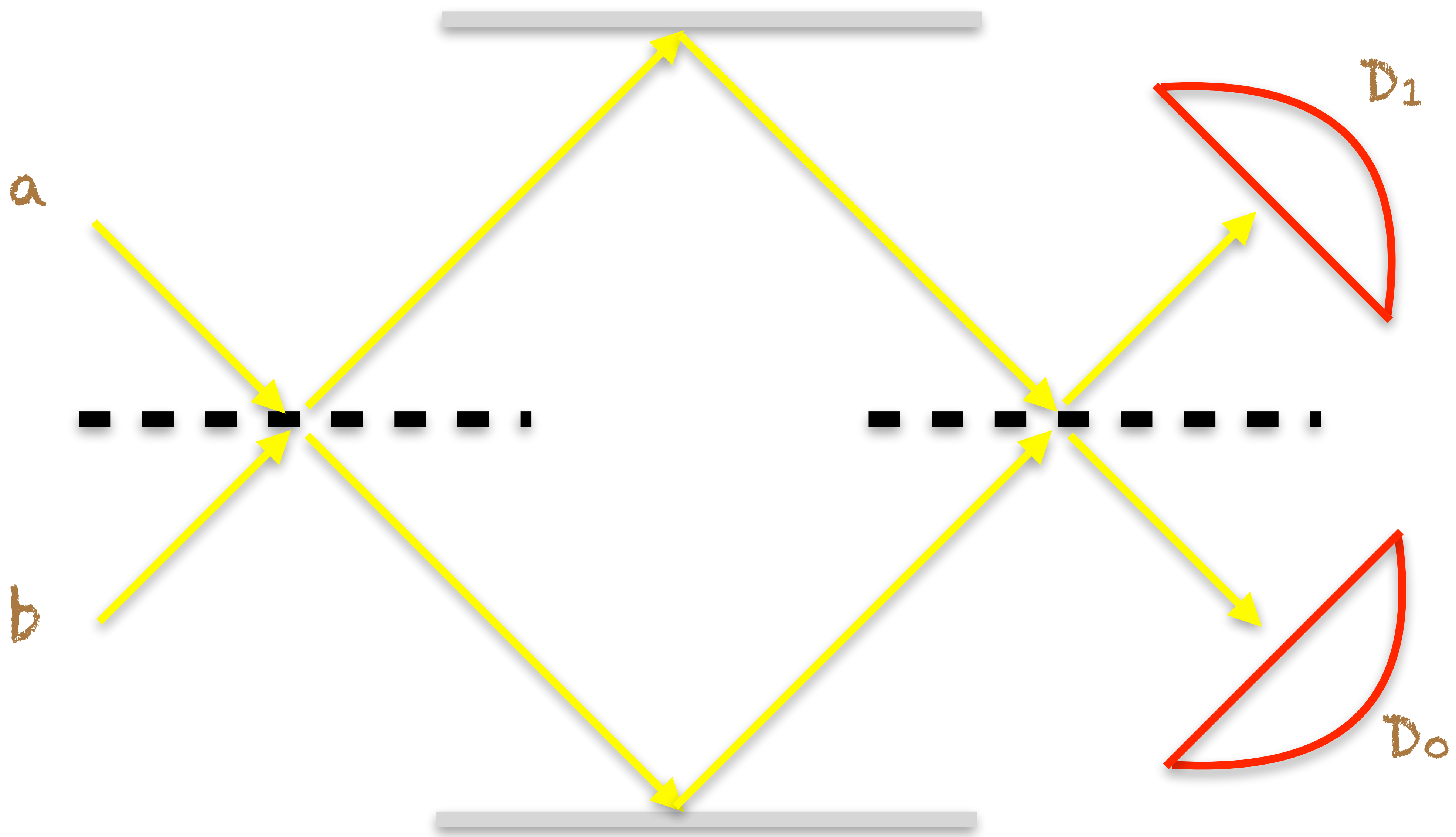


$$|\psi(\theta)\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + \sin\left(\frac{\theta}{2}\right)|1\rangle$$

Applications

Interferometry

- This is the Mach-Zehnder interferometer
- Its the device we use to detect gravitational waves…
- …in fact you can trace the origins of quantum parameter estimation to this device
- Let’s go ahead and analyze it



$$U_{BS} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -\mathrm{i} \\ -\mathrm{i} & 1 \end{pmatrix}$$

7. C. M. Caves. Phys. Rev. D, **23**, 1693

Interferometry

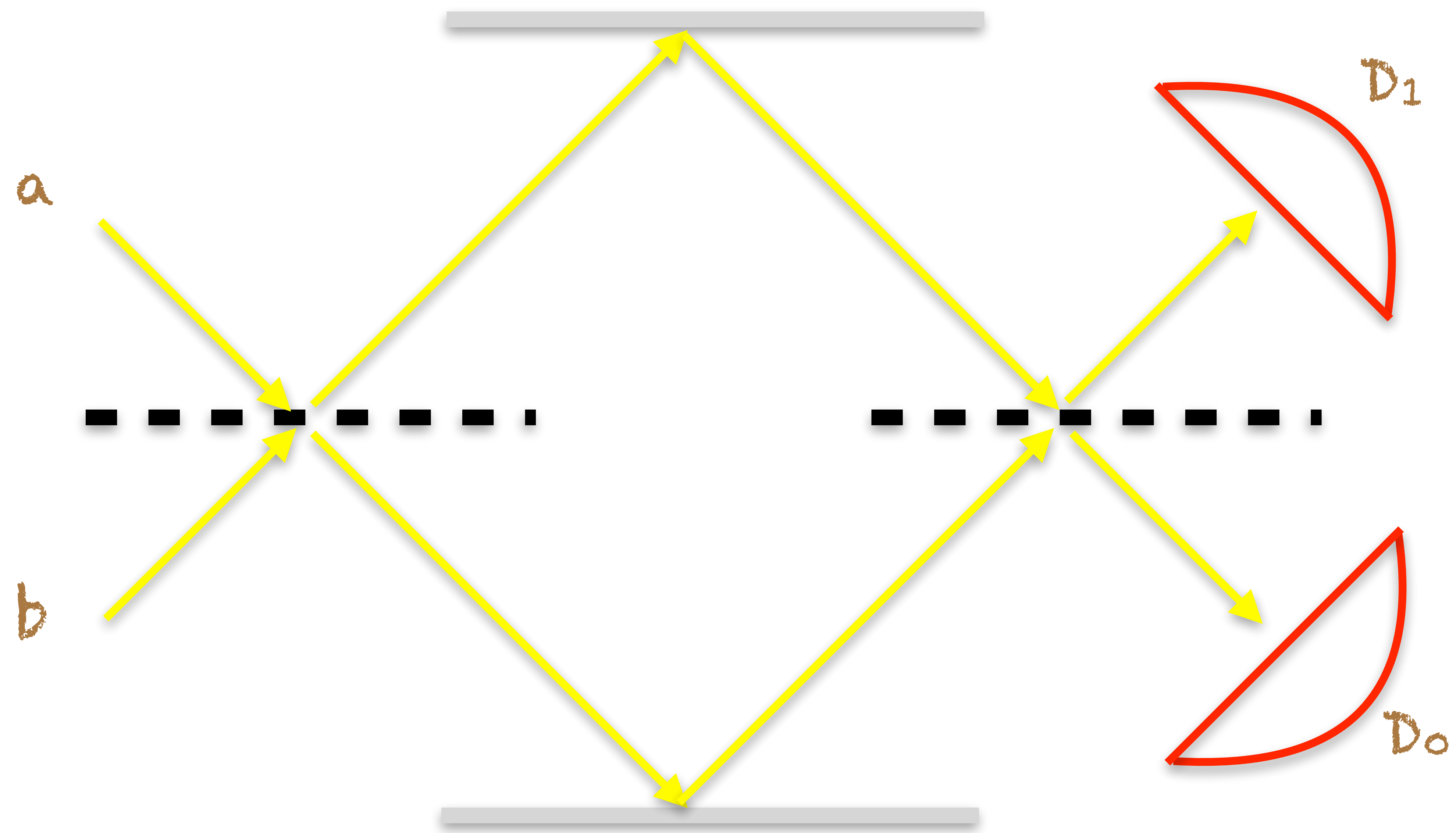
- The device is set up so that the path length between the two arms is identical

- Each **beam-splitter** enacts the transformation

$$U_{BS} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -\mathrm{i} \\ -\mathrm{i} & 1 \end{pmatrix}$$

- Both mirrors together perform the transformation

$$U_M = \mathrm{i} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$



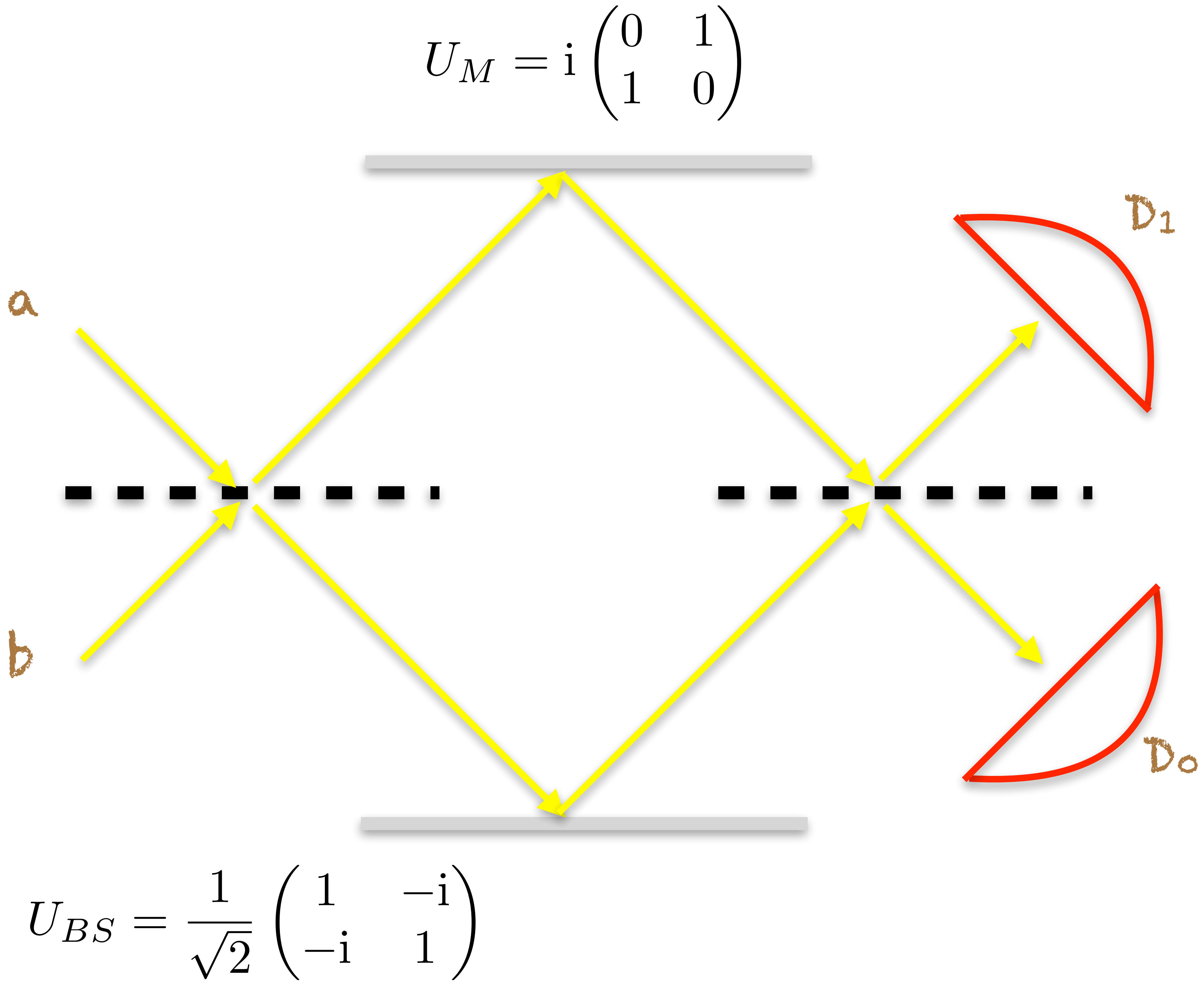
Interferometry

- So the entire device is mathematically described by

$$U_{MZ} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -\mathrm{i} \\ -\mathrm{i} & 1 \end{pmatrix}\begin{pmatrix} 0 & \mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -\mathrm{i} \\ -\mathrm{i} & 1 \end{pmatrix} = \mathbb{1}$$

- If you insert light in port **a** it will **always** come out towards detector D_0

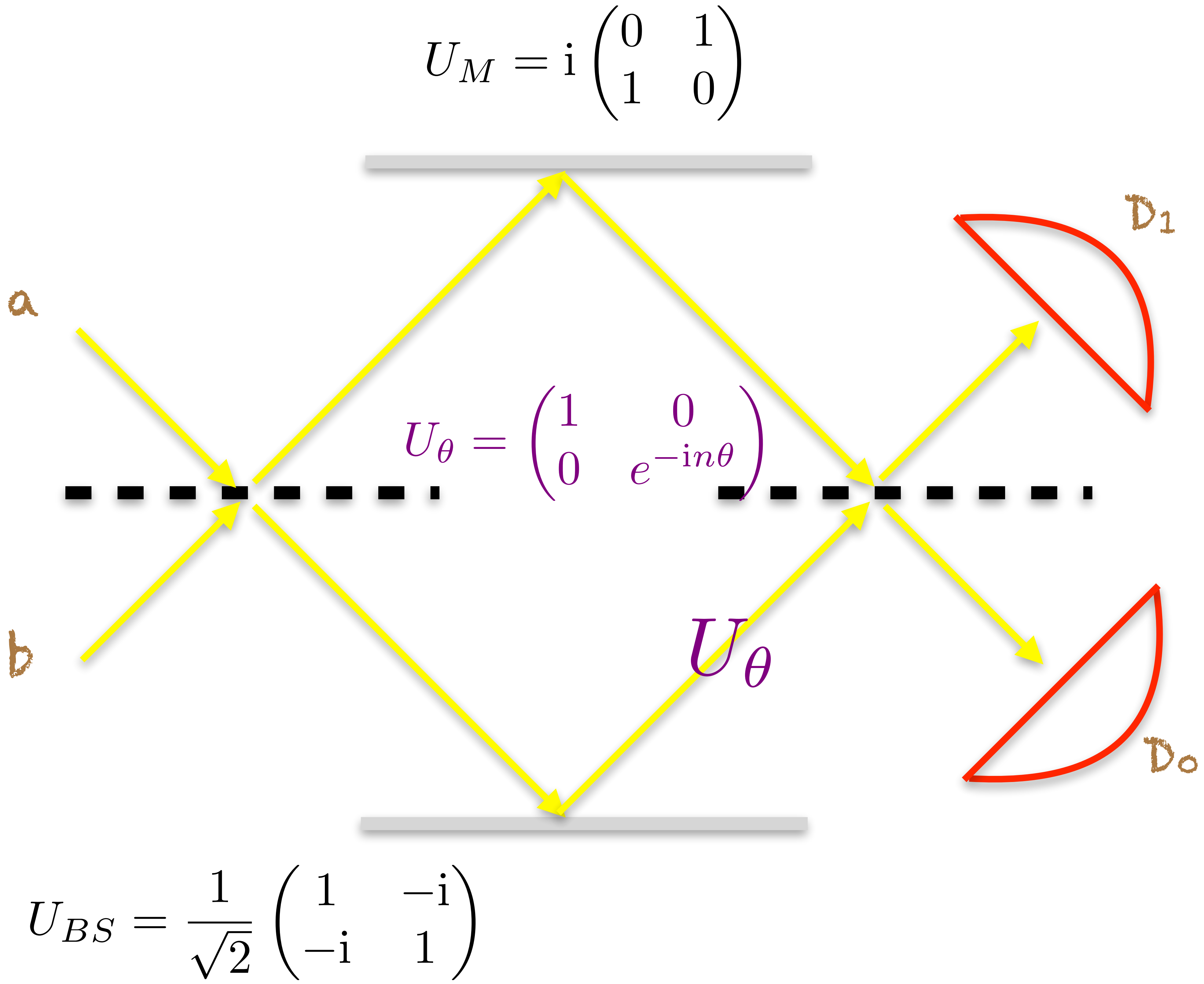
- and if you always insert light in port **b** it will always come out towards detector D_1



Interferometry

- Now suppose we change the length of the lower path
- This change translates to the light going down that path acquiring an additional phase θ
- Mathematically, this is described by the unitary operator

$$U_\theta = \begin{pmatrix} 1 & 0 \\ 0 & e^{-in\theta} \end{pmatrix}$$



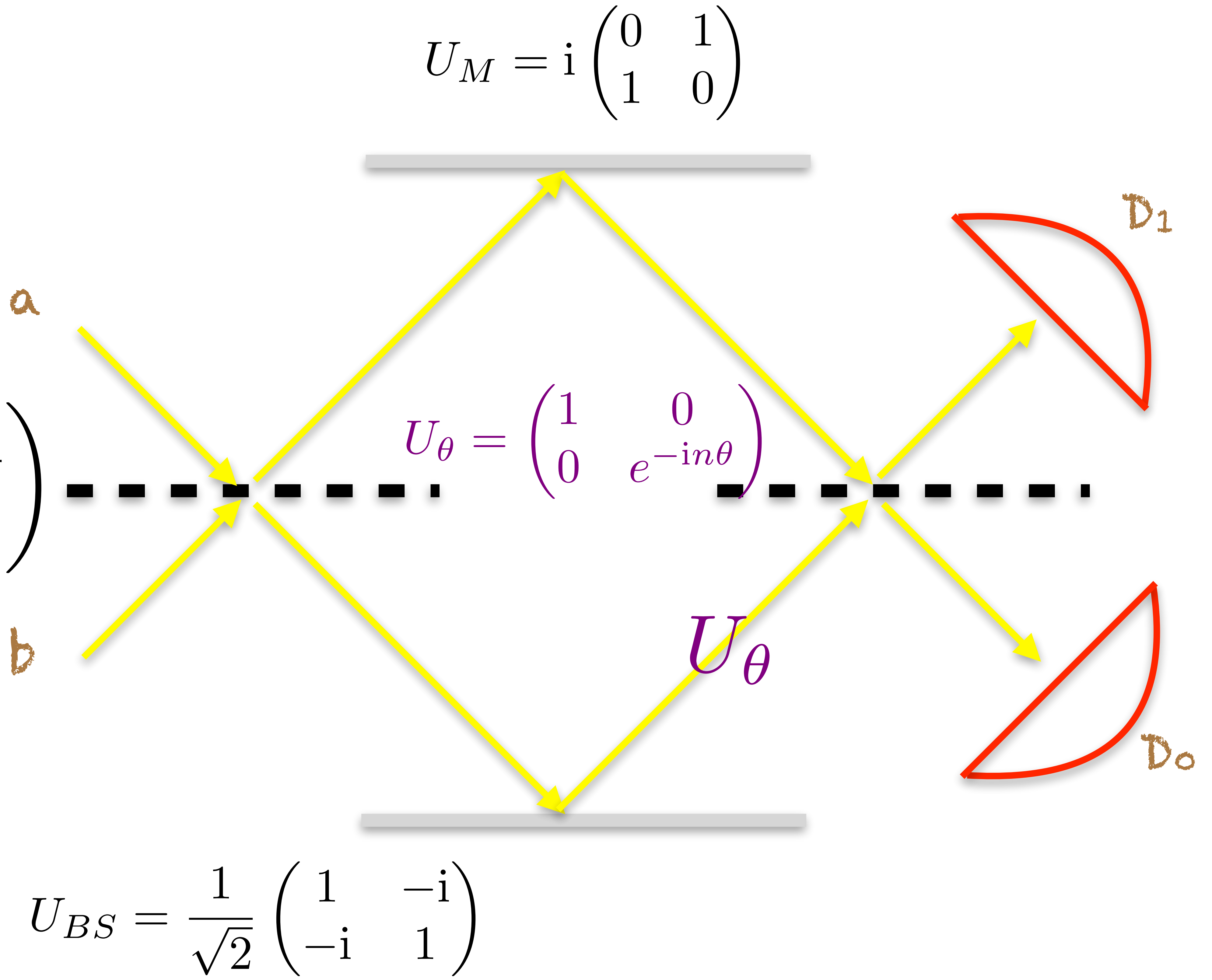
Interferometry

- Now the action of our Mach-Zehnder interferometer is described by

$$U_{MZ} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -\mathrm{i} \\ -\mathrm{i} & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & e^{\mathrm{i}\theta} \end{pmatrix}\begin{pmatrix} 0 & \mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -\mathrm{i} \\ -\mathrm{i} & 1 \end{pmatrix}$$

$$= e^{\mathrm{i}\frac{\theta}{2}}\begin{pmatrix} \cos\frac{\theta}{2} & \sin\frac{\theta}{2} \\ -\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix}$$

$$= e^{\mathrm{i}\frac{\theta}{2}\sigma_y}$$



- We immediately know that for a single photon

$$\mathcal{F}(\rho(\theta)) = \langle 0|\sigma_y^2|0\rangle - \langle 0|\sigma_y|0\rangle^2 = \Delta^2\sigma_y$$

Interferometry

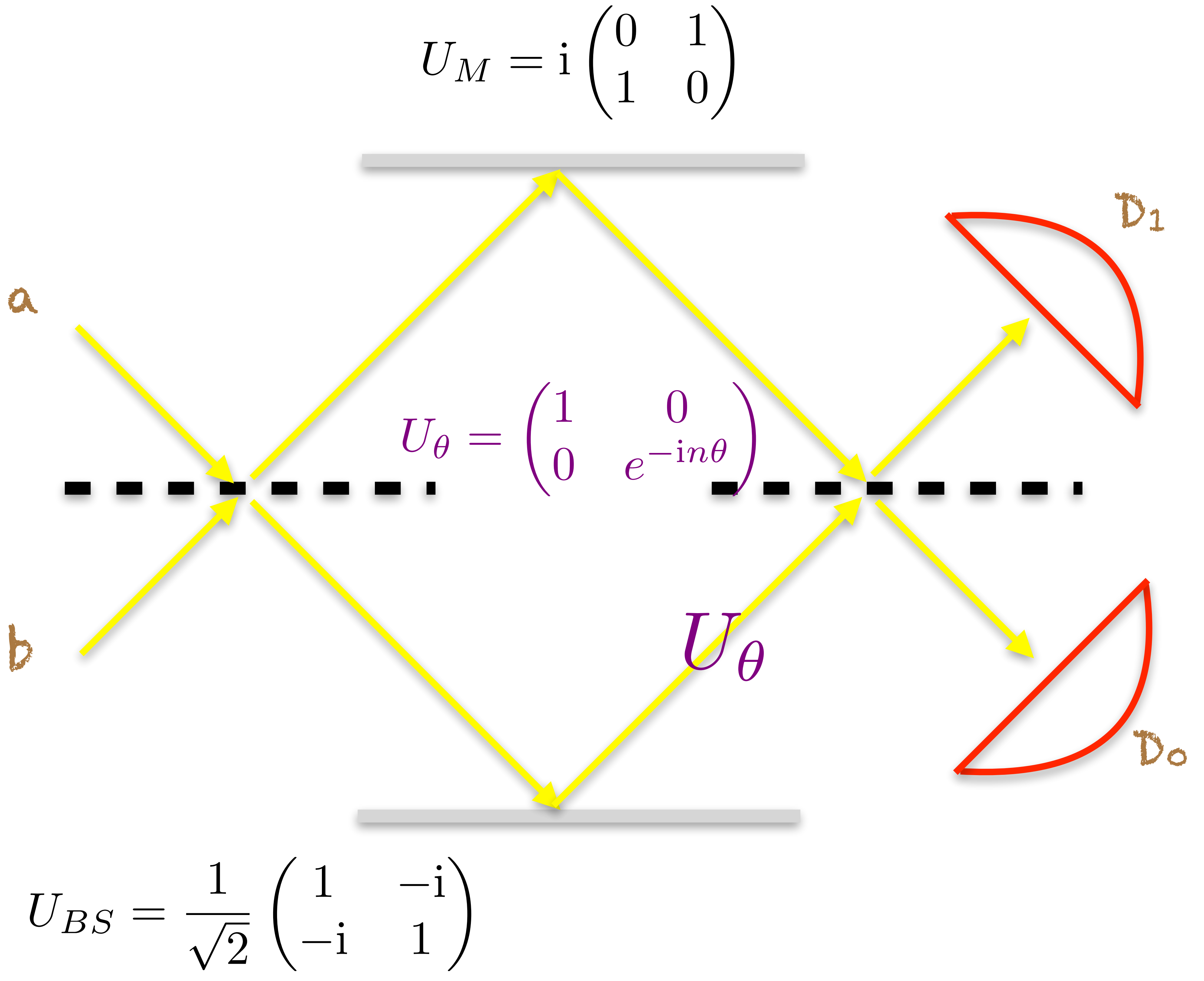
- If we repeat the experiment under identical conditions then

$$\mathcal{F}(\rho(\theta)^{\otimes n}) = n\mathcal{F}(\rho(\theta))$$

- and our error according to Crámer-Rao is

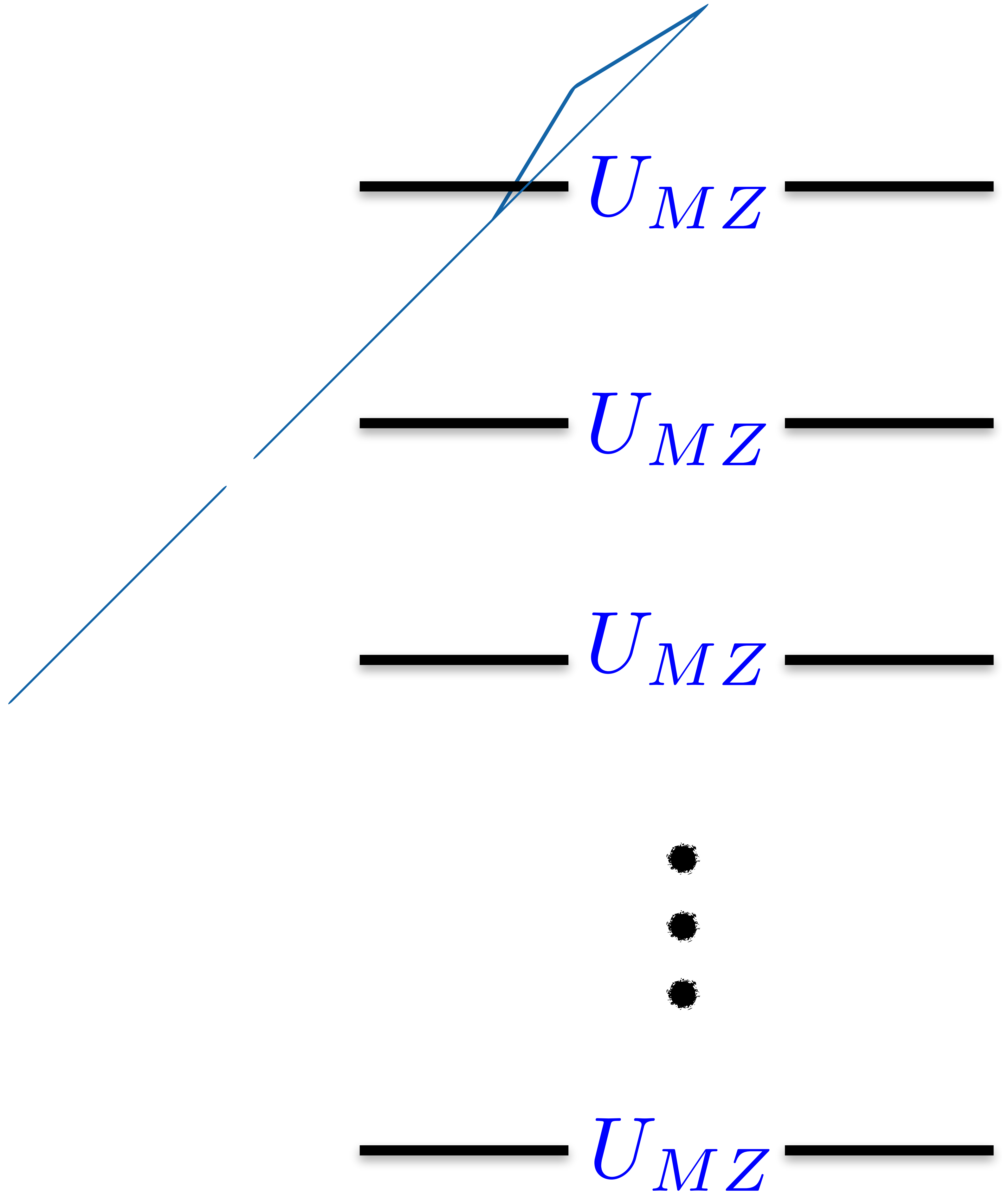
$$\delta\theta \geq \frac{1}{n\mathcal{F}(\rho(\theta))} = \frac{1}{n}$$

This is what is known as the **Standard Quantum Limit**



Interferometry

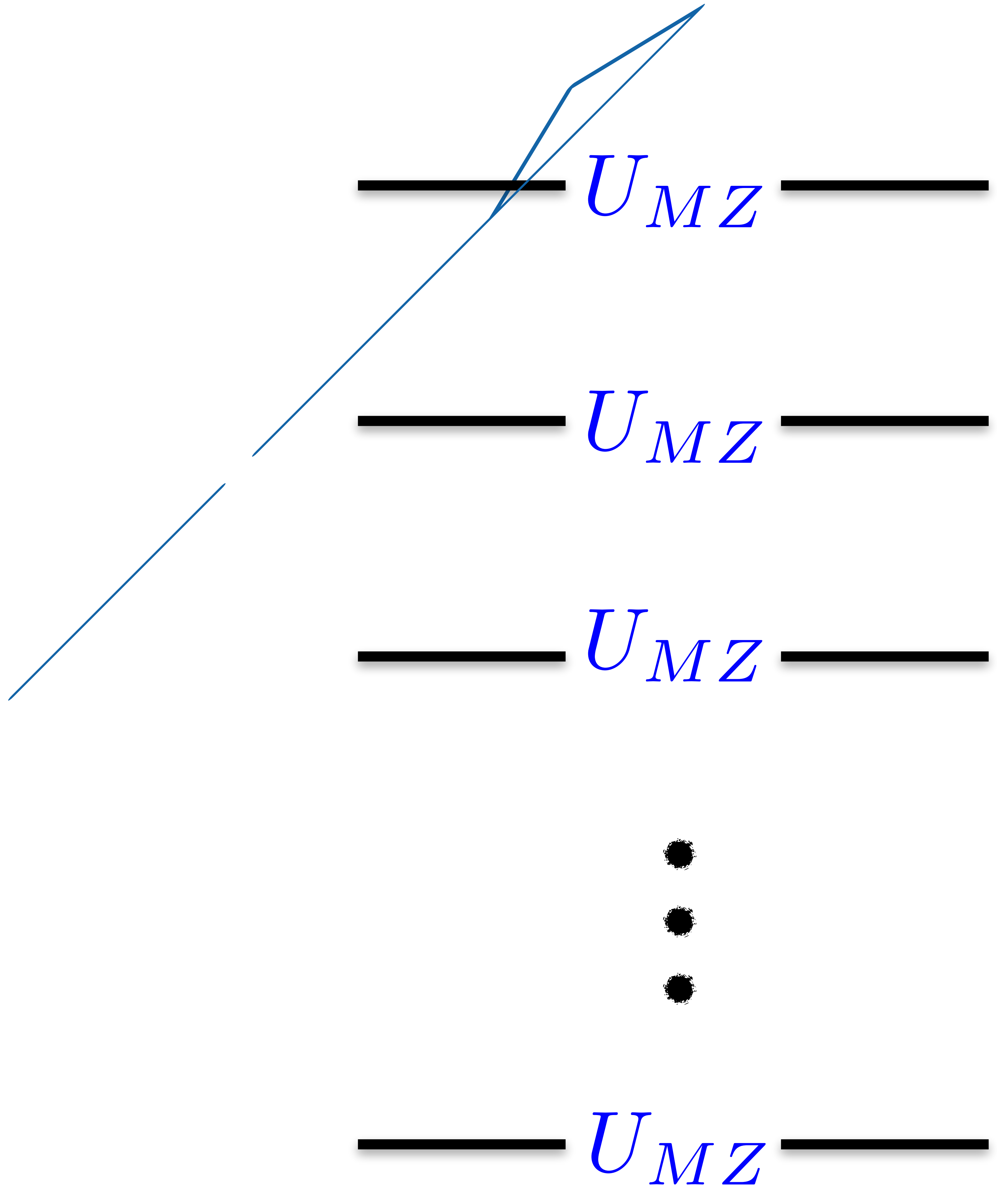
- Now lets use the same n **resources** a bit differently
- Instead of using the MZ device sequentially n times....



Interferometry

- Now lets use the same n **resources** a bit differently
- Instead of using the MZ device sequentially n times....
- Lets use all n times at once

$$U_{MZ}^{\otimes n} = e^{\mathrm{i}\frac{\theta}{2}\sum_{j=1}^{n}\sigma_y^{(j)}}$$



- Observe that the Hamiltonian

$$H = \sum_{j=1}^{n}\sigma_y^{(j)}$$

has n+1 distinct eigenvalues $\lambda_k = n - 2k \;\; k \in (0, \ldots, n)$

Interferometry

- The QFI is

$$\mathcal{F}[\rho(\theta)] = \langle\psi|H^2|\psi\rangle - \langle\psi|H|\psi\rangle^2$$

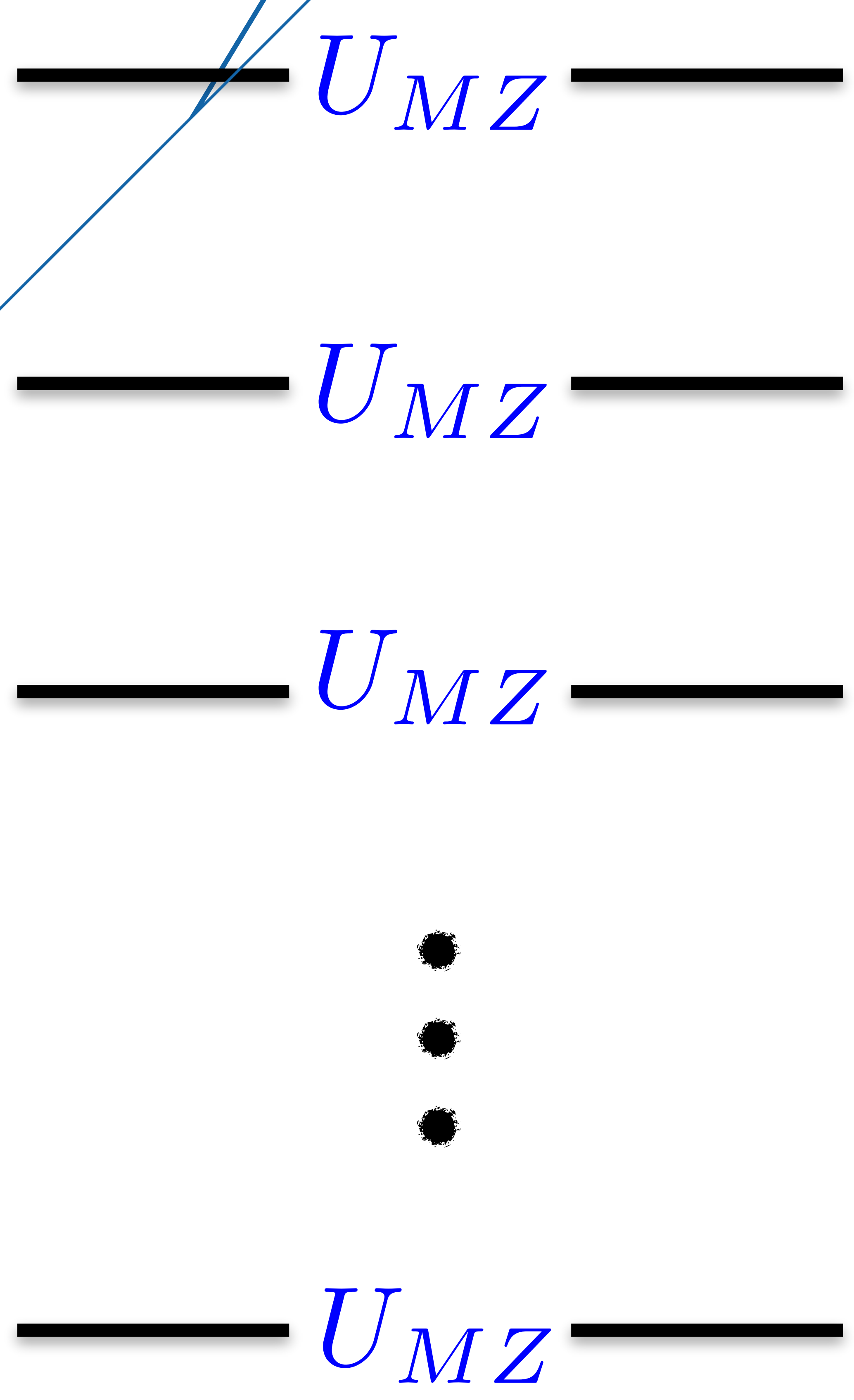
except now have the variance of $H = \sum_{j=1}^{n} \sigma_y^{(j)}$.

- We want to maximize this variance so we best pick

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|\lambda_{\max}\rangle + |\lambda_{\min}\rangle\right)$$

$$|\lambda_{\max}\rangle = |+\mathrm{i}\rangle^{\otimes n}$$

$$|\lambda_{\min}\rangle = |-\mathrm{i}\rangle^{\otimes n}$$



Interferometry

- The QFI now reads

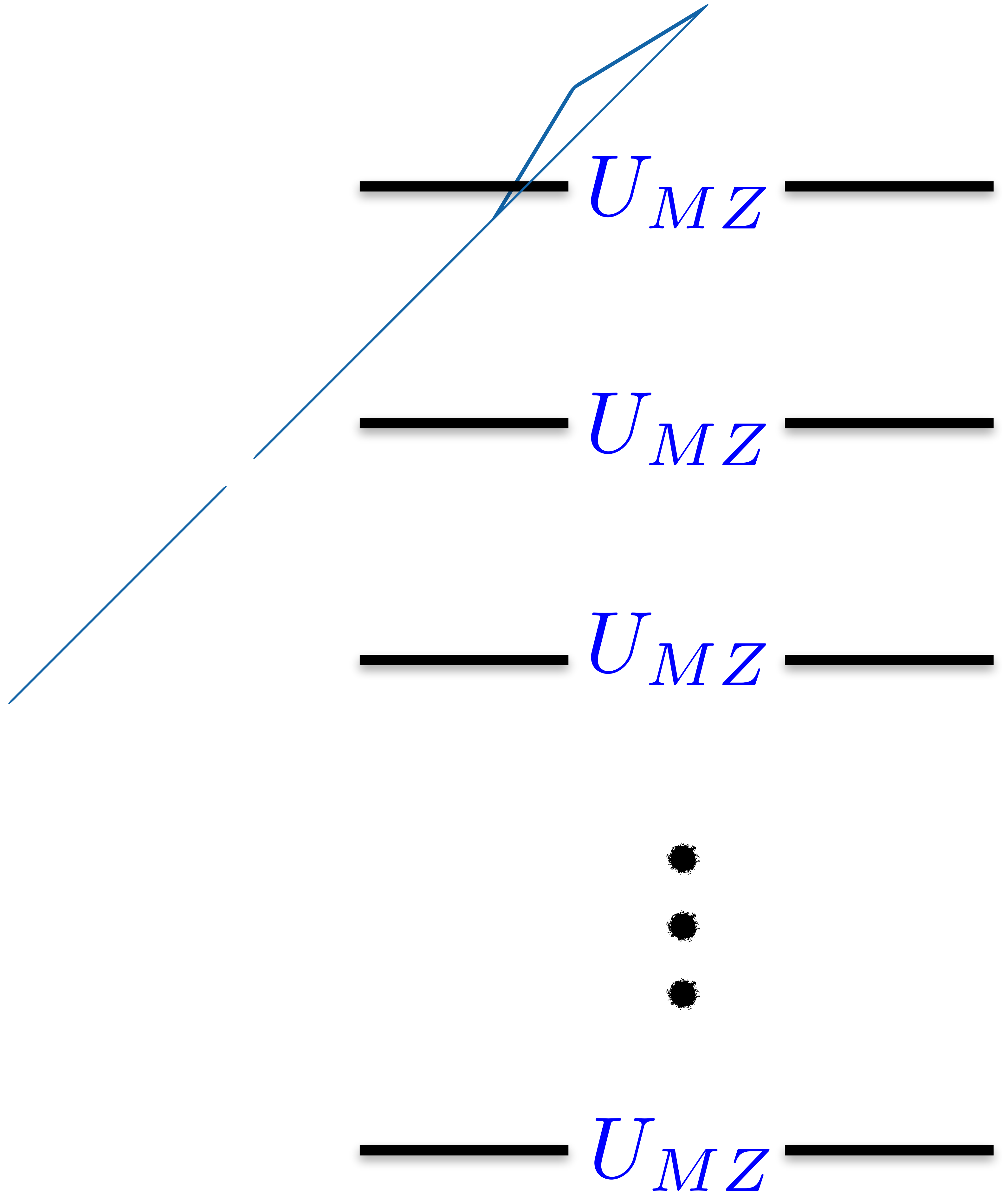
$$\mathcal{F}[\rho(\theta)] = \langle\psi|H^2|\psi\rangle - \langle\psi|H|\psi\rangle^2 = n^2$$

- so our error becomes

$$\delta\theta \geq \frac{1}{\mathcal{F}(\psi(\theta))} = \frac{1}{n^2}$$

quadratically smaller as compared to before

This is known as the **Heisenberg Limit**



Some more applications

Hypothesis Testing

- The capacity of a channel to carry classical (quantum information)
- Security in Quantum Cryptography
- Entanglement detection
- Quantum Radars and Lidars
- Distinguishing ground states of Hamiltonians across a phase transition

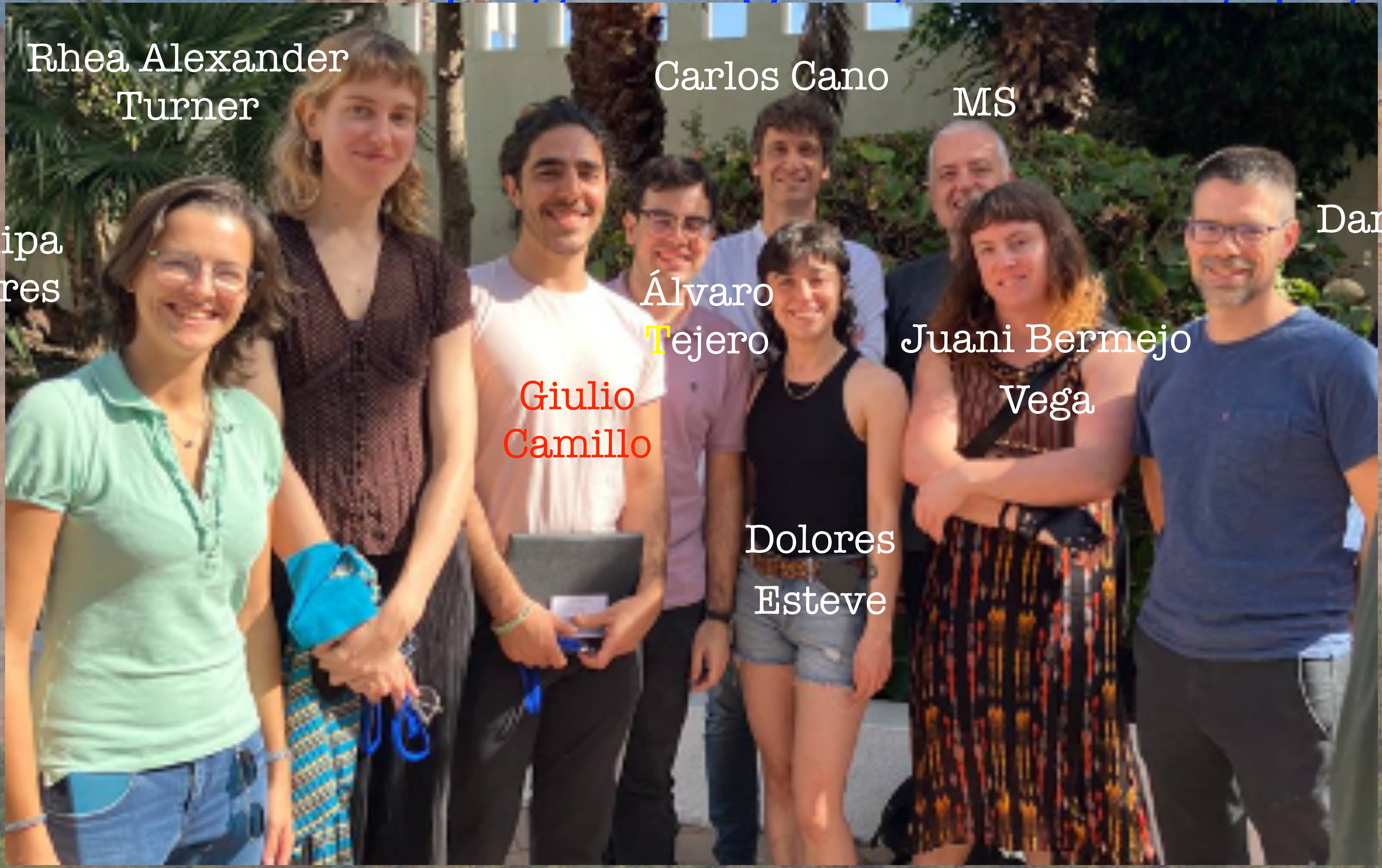
Parameter Estimation

- Atomic clocks
- Magnetimetry
- Accelerometers/gravimetry
- Thermometry
- Quantum Imaging/super-resolution
- Spectrometry

Quantum Thermodynamics and Computation Group

INSTITUTO CARLOS I
DE FÍSICA TEÓRICA Y COMPUTACIONAL

https://ic1.ugr.es/members/qtc/



Thank you for your attention

mskotiniotis@onsager.ugr.es