1. Quantum Feature Spaces & Kernels

Background:

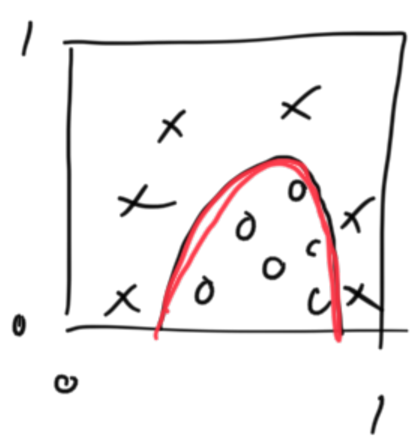
Def: feature map $\qquad x\in\Omega$ data space

$\mathcal{H}\in\mathbb{R}^N$ feature space

$$\phi:\Omega\longrightarrow\mathcal{H}\subseteq\mathbb{R}^N$$

$$\phi: x\longmapsto\phi(x)$$ "ϕ non-linear"

$$\left[\phi(\lambda x+\mu y)\neq\lambda\phi(x)+\mu\phi(y)\right]$$



$\Omega\subseteq[0,1]\times[0,1]$

$x\in\Omega \qquad (x_1,x_2)$

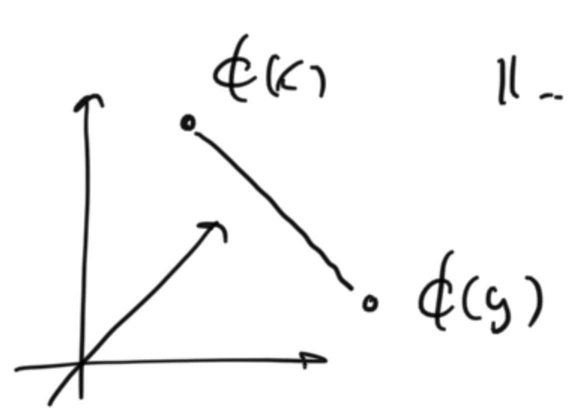
$$\phi_2:(x_1,x_2)\longmapsto\begin{pmatrix}1\\ \sqrt{2}x_1\\ \sqrt{2}x_2\\ x_1^2\\ \sqrt{2}x_1x_2\\ x_2^2\end{pmatrix}\in\mathbb{R}^6$$

"features"

Def: kernel-function $(\phi,\mathcal{H})$

$$K:\Omega\times\Omega\longrightarrow\mathbb{R}$$

$$K:(x,y)\longmapsto K(x,y)=\langle\phi(x),\phi(y)\rangle$$



$\|\cdot\|_2^2$

$$\|\phi(x)-\phi(y)\|_2^2=K(x,x)+K(y,y)-2K(x,y)$$

Example: ϕ_2:

$$\langle\phi_2(x),\phi_2(y)\rangle=K(x,y)=(\langle x,y\rangle+1)^2$$

Converse possible:

Start with $K(x,y) \hookrightarrow (\phi, \mathcal{H})$
[Mercer's Condition]

Examples:

$$K_{poly}(x,y) = (\langle x,y\rangle + 1)^d$$
$$K_{RBF}(x,y) = e^{-\frac{1}{2}\|x-y\|_2^2}$$

Quantum feature spaces:

[Nature vol. 567 p.p. 209-212 (2019)]
[arXiv: 1804.11136]

- Vector space of matrice:

$A, B \in \mathbb{C}^{N\times N}$ $\quad \alpha, \beta \in \mathbb{C}$

$\alpha A + \beta B \in \mathbb{C}^{N\times N}$

$$\begin{bmatrix} \alpha A_{11} & \alpha A_{12} \\ \alpha A_{21} & \alpha A_{22} \end{bmatrix} + \begin{bmatrix} \beta B_{11} & \cdots \\ \cdots & \beta B_{22} \end{bmatrix}$$
$$= \begin{bmatrix} \alpha A_{11} + \beta B_{11} & \vdots \\ \cdots & \cdot \end{bmatrix}$$

- inner product: (Hilbert Schmidt / trace) inner product.

$$\langle A, B\rangle_{HS} = \mathrm{tr}[A^\dagger B]$$
$$= \sum_{ij} A^*_{ij} B_{ij}$$

$$\mathrm{tr}[A] = \sum_i A_{ii}$$

- $\dim(\mathbb{C}^{N\times N}) = N^2$

- Single qubit example:

Recall:

$$f_w(x) = \mathrm{sign}\left(\langle \phi(x) | \hat{W} | \phi(x) \rangle \right)$$

- $\hat{W} = W(\theta)\ \mathrm{Diag}(+1, -1)\ W^\dagger(\theta)$

$$= W(\theta) \begin{bmatrix} 1 & \\ & -1 \end{bmatrix} W^\dagger(\theta)$$

- $|\phi(x)\rangle = U(x)\,|0\rangle$

Using $\langle A, B\rangle_{HS}$

$$f_w(x) = \mathrm{sign}\left(\mathrm{tr}\left[|\phi(x)\rangle\langle\phi(x)| \cdot \hat{W} \right] \right)$$

$$= \mathrm{sign}\left(\langle \underline{\Phi}(x), \hat{W} \rangle_{HS} \right)$$

$\underline{\Phi}(x) = |\phi(x)\rangle\langle\phi(x)|$ ~ density matrix ~ "feature vec."

$\hat{W}$ ~ operator ~ "normal"

- Single Qubit Pauli Basis

$$\mathbb{1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$$

$$Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

check:

$$\mathrm{tr}[X^\dagger X] = \langle X, X\rangle_{HS}$$

$$= \mathrm{tr}[\mathbb{1}] = 2$$

$$\mathrm{tr}[X^\dagger Z] = \mathrm{tr}\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = 0$$

$\vdots$

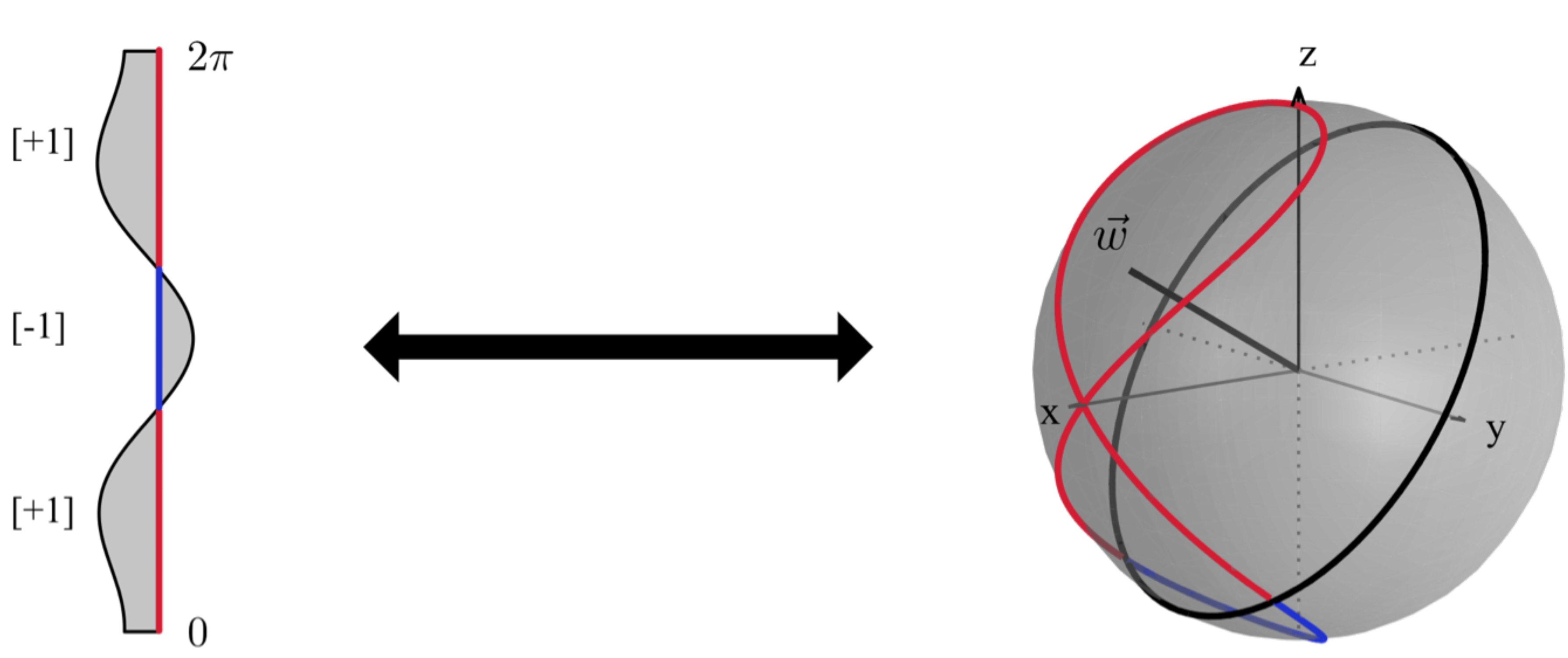
Expand:

$$\underline{\Phi}(x) = \frac{1}{2}\Big(\underbrace{\mathrm{tr}[\underline{\Phi}(x)]}_{1}\,\underline{1} + \mathrm{tr}[\underline{\Phi}(x)X]\,X$$
$$+ \ldots + \mathrm{tr}[\underline{\Phi}(x)Z]\,Z \Big)$$
$$= \frac{1}{2}\Big(\underline{1} + r_x(x)X + r_y(x)Y + r_z(x)Z \Big)$$

$$\mathrm{tr}[\hat{w}] = 0$$

$$\hat{w} = \frac{1}{2}\left(w_x X + w_y Y + w_z Z \right)$$

$$w_x = \mathrm{tr}[X\hat{w}] \quad \ldots$$



$$\Omega \xrightarrow{U(x)}$$
$[0, 2\pi]$

$$\underbrace{- \boxed{H} - \boxed{Z_x} - \boxed{H} - \boxed{Z_x}}_{U(x)}$$

$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ $\quad e^{i\frac{x}{2}Z}$

$$\Phi(x) = \boxed{U(x)} |0\rangle\langle 0| U^{\dagger}(x)$$

$$\mathrm{tr}[\hat{W} \Phi(x)]$$

- What about $|\phi(x)\rangle \in \mathbb{C}^{2^n}$

[Phys. Rev. Lett. 122 040504 (2019)]

$$\left.\begin{matrix} |\phi(x)\rangle \\ |w\rangle \end{matrix}\right\} \in \mathbb{C}^{2^n}$$

$$f(x) = \mathrm{sign}(\langle w | \phi(x)\rangle)$$

$$\langle w | \phi(x)\rangle \in \mathbb{C}$$

- $|\phi(x)\rangle \sim \underbrace{e^{i\eta} |\phi(x)\rangle}_{|\tilde{\phi}(x)\rangle} \quad \eta \in \mathbb{R}$

$$\langle \phi(x) | A | \phi(x)\rangle = \langle \tilde{\phi}(x) | A | \tilde{\phi}(x)\rangle$$

$\langle w | \phi(x)\rangle$ positive number

$\langle w | (-|\phi(x)\rangle) = -\langle w | \phi(x)\rangle$ neg ?

- $K(x,y) = \langle \phi(x) | \phi(y)\rangle$

Summarize: (n-qubit)

Featur maps:

- $$\Phi : x \in \Omega \longmapsto \Phi(x) = |\phi(x)\rangle\langle\phi(x)| = \underline{U(x)\,|0\rangle\langle 0|\,U^{+}(x)}$$

- $$K(x,y) = \langle \Phi(x), \Phi(y) \rangle_{HS} = |\langle \phi(x)|\phi(y)\rangle|^2 \in \mathbb{R}^{+} = |\langle 0|\, U^{+}(x)\, U(y)\, |0\rangle|^2$$

- Featurs: Matrix Basis

$$P_\alpha^{+} = P_\alpha \qquad \alpha = 1 \ldots 4^n$$

$$[\Phi(x)]_\alpha = tr[P_\alpha \Phi(x)] = \langle \phi(x)| P_\alpha |\phi(x)\rangle$$

- Linear Classifier:

$$f(x) = sign\left(tr[\hat{W} \Phi(x)] \right) = sign\left(\sum_{\alpha=1}^{4^n} w_\alpha [\Phi]_\alpha(x) \right)$$

$$w_\alpha = tr[\hat{W} P_\alpha]$$

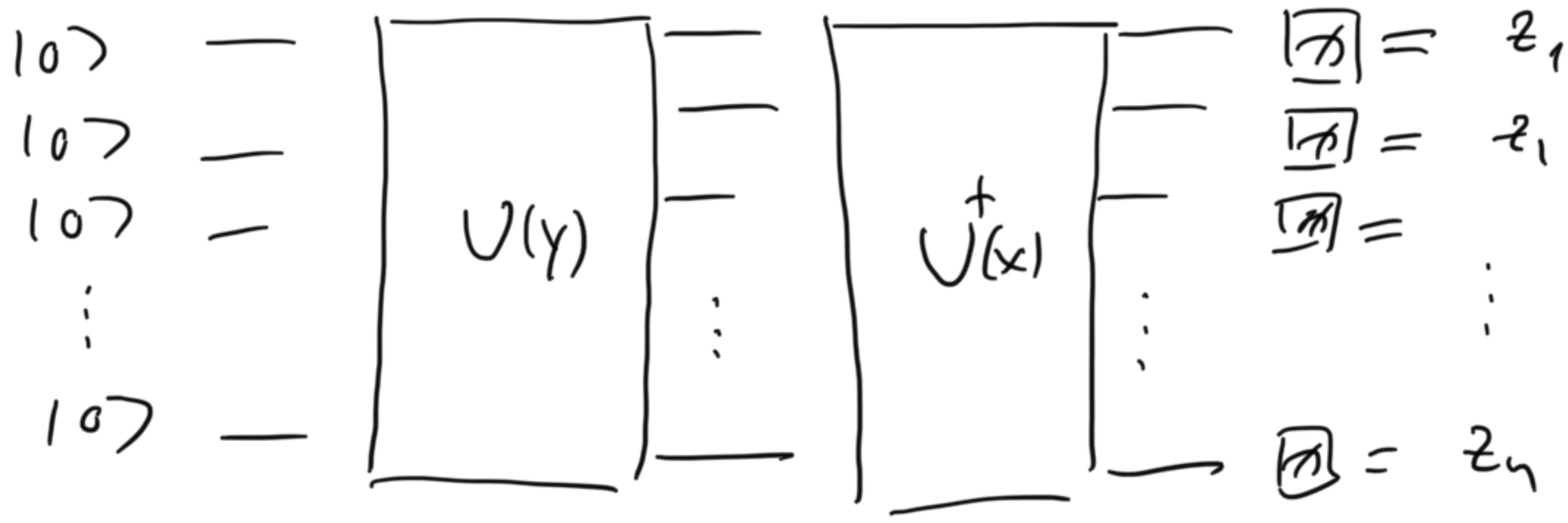
$\boxed{U(x)}$ feature map Circuit:

2. Quantum Kernel Estimation:

Use Quantum Computer to estimate:

$$K(x,y) = |\langle \phi(x)|\underbrace{\phi(y)}_{}\rangle|^2 = |\langle 0|\overbrace{U^{+}(x)\,U(y)}|0\rangle|^2$$

transition amplitude



Set: $r = 0$

Repeat R - times:

1) Perp $|0^n\rangle$

2) apply $U(y)$

3) apply $U^\dagger(x)$

measure in Z-basis

if we see $x = (0, \dots, 0)$

set $r \mapsto r+1$

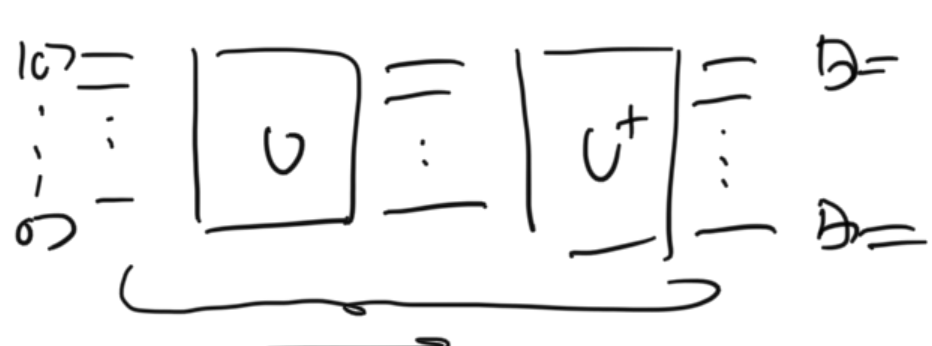
$$\hat{K}(x,y) = \frac{r}{R}$$

with high probability

$$\hat{K}(x,y) = K(x,y) + \frac{1}{\sqrt{R}}$$

additive sampling error

- "near-term" application



$\hat{K}$ is an expectation value

$\Rightarrow$ "Error mitigation" can be applied

Algorithms that use Quantum kernel estimation

Example: Classification (SVM)

$$T = \{ (\underline{x}_1, y_1), \dots, (\underline{x}_n, y_n) \} \qquad \underline{x}_i \in \Omega$$

$$y_i \in \{+1, -1\}$$

Training:

$$\underset{\{\alpha_i\}}{\text{maximize}} \quad \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \underbrace{\hat{K}(x_i, x_j)}_{\text{QKE}}$$

subject to: $\quad 0 \leq \alpha_i \leq C$

$$\sum_i \alpha_i y_i = 0$$

- m^2- many times $\hat{K}$
- α_i^*

Use classifier:

- $z \in \Omega$

$$f(z) = \text{sign}\left(\sum_i \alpha_i^* y_i \hat{K}(x_i, z) \right)$$

QKE

3. Choice of $U(x)$:

- $U(x)$ unitary circuit
  depends on x non-linearly

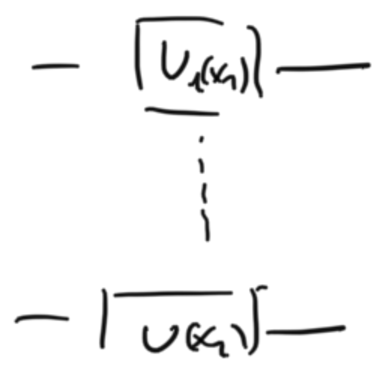
- $K(x,y) = \underbrace{|\langle 0| U^\dagger(x) U(y) |0\rangle|^2}$

We want that estimation of this is "difficult" for a classical computer

$\underline{\hat{K} = K + \varepsilon}$ [even with an additive error]

a) **Trivial Examples:** (CT) [Classically tractible]

- product states: $\underline{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$



$$U(x) = \bigotimes_{i=1}^{n} U_i(x_i)$$

$$\underline{U(x_i) = e^{i \frac{x_i}{2} Y}}$$

$$K(x,z) = |\langle 0| U^\dagger(x) U(y) |0\rangle|^2$$

$$= \prod_{i=1}^{n} \underbrace{|\langle 0| U_i^\dagger(x_i) U_i(z_i) |0\rangle|^2}_{\text{indiviually}}$$

$O(n)$ $\quad |\langle 0| U^\dagger(x_i) U(y_i) |0\rangle|^2 = \cos\left(\frac{x_i - y_i}{2}\right)^2$

- Amplititude encoding

$(x_1 \dots x_n) \hookrightarrow |\vec{x}\rangle$ polynomial Kernels

- Coherent states $\to$ Gaussian RBFs

$$|\langle 0| D^\dagger(\alpha) D(\beta) |0\rangle|^2 = e^{-\frac{1}{2}\|\alpha - \beta\|_2^2}$$

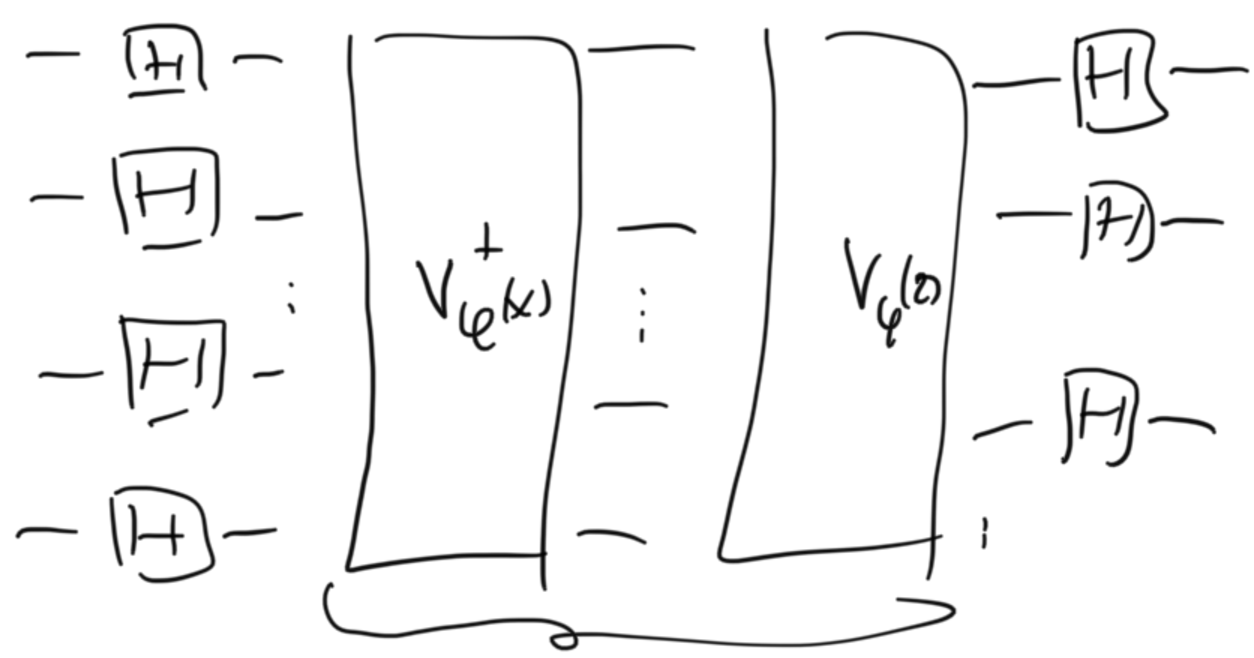
- **IQP Circuits:** [Classically tractible amplitude] (up to ε-error)

$U(x)$

$$V_{\varphi}(x) = e^{i\left(\sum_{i=1}^{n} \varphi_i(x) Z_i + \sum_{i,j} \varphi_{ij}(x) Z_i Z_j\right)}$$

$$V_{\varphi}(x) = \begin{bmatrix} e^{\theta_1(x)} & & \\ & \ddots & \\ & & e^{\theta_n} \end{bmatrix}$$

$K(x,y) =$



Dum

$$K(x,y) = \left| \frac{1}{2^n} \sum_{z \in \{+1,-1\}^n} \underbrace{e^{i\left(\sum_i [\varphi_i(x) - \varphi_i(y)] Z_i + \sum_{ab} [\varphi_{ab}(x) - \varphi_{ab}(y)] Z_a Z_b\right)}}_{e^{i\theta(x,y)}} \right|^2$$

↑ uniform sum. $\qquad$ $e^{i\theta(x,y)}$ ← phase

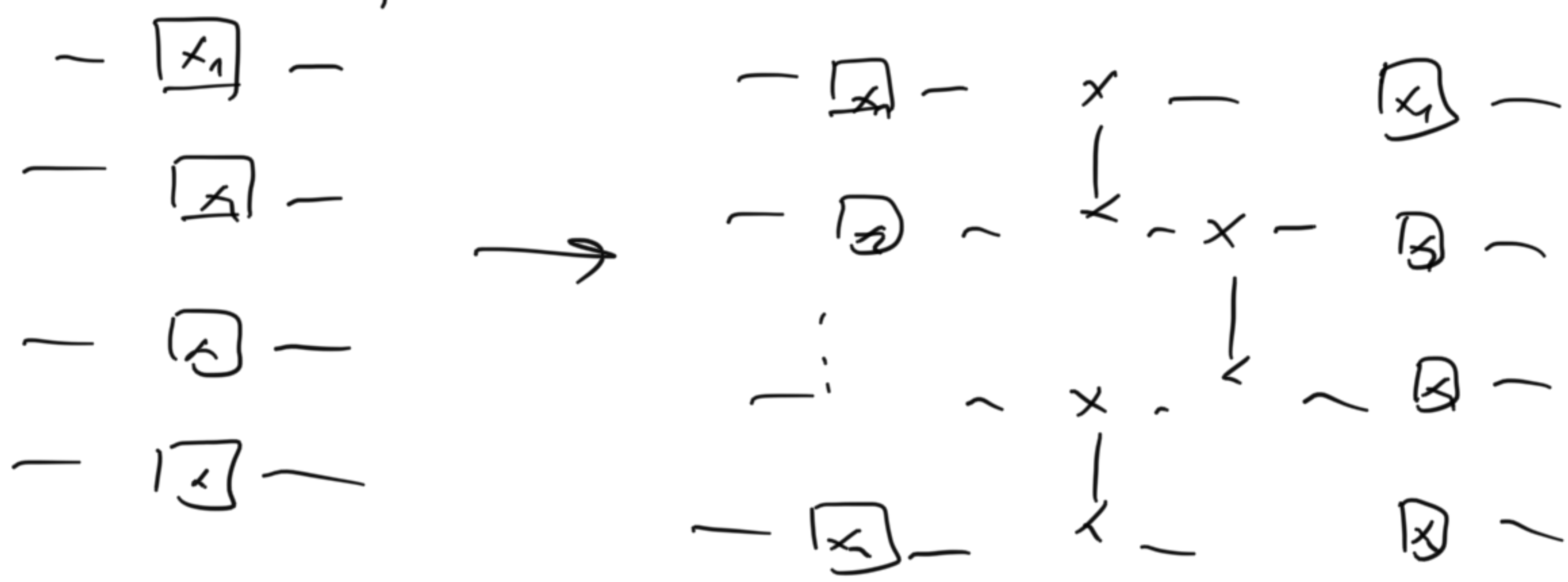
Estimation (classically):

- draw $z \in \{+1,-1\}^n$ uniformly at random w.p. $\frac{1}{2^n}$
- add phase $e^{i\theta(x,y)}$ [bounded variance]

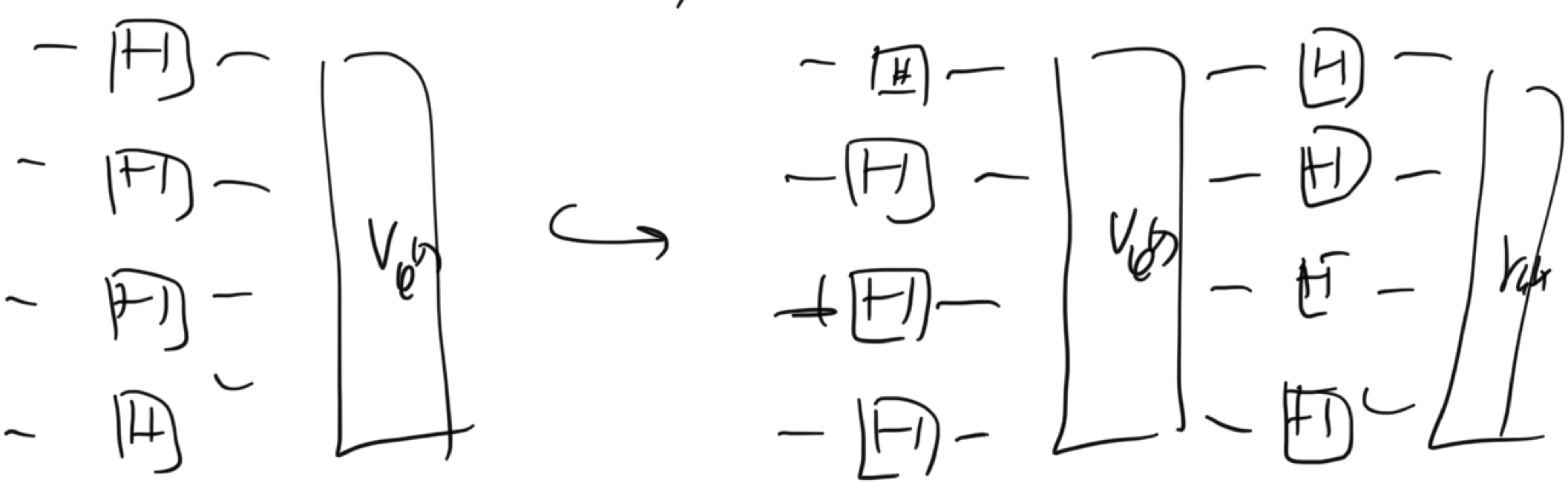
$\Rightarrow$ $\hat{K} = K + \varepsilon$ with $O(\varepsilon^{-2})$ samples.

b) Suggestions for non-trivial circuits

product → entangled

- 

- Repeated depth →



$$K \approx \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$$

- DLOG example: [Seperating Example]

arxiv: 2010.02174.