Appendix

A. Proof of Proposition 1

We divide the proof into two parts: approximation error and effective dimension.

Part I: Approximation error. Let $g^* : \mathcal{X} \to R$ be the target function. By assumption, there exist latent maps:

$$\mathbf{z}^* = g_0(\mathbf{x}), \quad \mathbf{x}_b^* = g_1(\mathbf{z}^*), \quad \mathbf{h}^* = g_2(\mathbf{z}^*), \quad \boldsymbol{\phi}^* = g_3(\mathbf{h}^*),$$

such that

$$\sup_{\mathbf{x}\in\mathcal{X}} \left| g^*(\mathbf{x}) - Q_L(\widehat{\mathbf{x}}^*, \boldsymbol{\phi}^*) \right| \leq \epsilon_q(U, L), \tag{A.1}$$

which is exactly Eq. (9) in the main text.

Now let $f_{\boldsymbol{\omega}} \in F_{\boldsymbol{\omega}}^{(r)}$ be a TensorDual-VQC model with TT-rank at most r, producing the outputs as follows:

$$\mathbf{z} = f_{\boldsymbol{\theta}}^{(0)}(\mathbf{x}), \quad (\widehat{\mathbf{x}}, \mathbf{h}) = \left(f_{\boldsymbol{\theta}}^{(1)}(\mathbf{z}), f_{\boldsymbol{\theta}}^{(2)}(\mathbf{z})\right), \quad \boldsymbol{\phi} = f_{\psi}(\mathbf{h}),$$

and

$$f_{\boldsymbol{\omega}}(\mathbf{x}) = \mathcal{Q}_L(\widehat{\mathbf{x}}, \boldsymbol{\phi}).$$

By the shared low-rank approximation assumption, the joint mapping $\mathbf{x} \mapsto (\widehat{\mathbf{x}}^*, \boldsymbol{\phi}^*)$ admits an approximate TT representation of rank r, so there exists a model in $\mathcal{F}_{\boldsymbol{\omega}}^{(r)}$ such that

$$\|(\widehat{\mathbf{x}}, \boldsymbol{\phi}) - (\widehat{\mathbf{x}}^*, \boldsymbol{\phi}^*)\|_2 \leq C_f r^{-\alpha} \tag{A.2}$$

uniformly over $\mathbf{x} \in \mathcal{X}$. This corresponds to Eq. (11) in the main text.

Further, by the joint Lipschitz property of $\mathcal{Q}_L$ in $(\widehat{\mathbf{x}}, \boldsymbol{\phi})$,

$$\left| \mathcal{Q}_L(\widehat{\mathbf{x}}^*, \boldsymbol{\phi}^*) - \mathcal{Q}_L(\widehat{\mathbf{x}}, \boldsymbol{\phi}) \right| \leq L_{\mathcal{Q}} \left\| (\widehat{\mathbf{x}}, \boldsymbol{\phi}) - (\widehat{\mathbf{x}}^*, \boldsymbol{\phi}^*) \right\|_2, \tag{A.3}$$

which is Eq. (12).

We now decompose the approximation error:

$$\begin{aligned} \left| g^*(\mathbf{x}) - f_{\boldsymbol{\omega}}(\mathbf{x}) \right| &= \left| g^*(\mathbf{x}) - \mathcal{Q}_L(\widehat{\mathbf{x}}, \boldsymbol{\phi}) \right| \\ &\leq \left| g^*(\mathbf{x}) - \mathcal{Q}_L(\widehat{\mathbf{x}}^*, \boldsymbol{\phi}^*) \right| + \left| \mathcal{Q}_L(\widehat{\mathbf{x}}^*, \boldsymbol{\phi}^*) - \mathcal{Q}_L(\widehat{\mathbf{x}}, \boldsymbol{\phi}) \right|. \end{aligned} \tag{A.4}$$

Applying (A.1), (A.3), and then (A.2), we obtain

$$\left| g^*(\mathbf{x}) - f_{\boldsymbol{\omega}}(\mathbf{x}) \right| \leq \epsilon_q(U, L) + L_{\mathcal{Q}} C r^{-\alpha}. \tag{A.5}$$

Taking the supremum over $\mathbf{x} \in \mathcal{X}$ gives

$$\|f_{\boldsymbol{\omega}} - g^*\|_\infty \leq \epsilon_q(U, L) + L_{\mathcal{Q}} C r^{-\alpha}. \tag{A.6}$$

Since $L_{\mathcal{Q}}$ and C are constants independent of r, this yields

$$\inf_{f_{\boldsymbol{\omega}} \in \mathcal{F}_{\boldsymbol{\omega}}^{(r)}} \|f_{\boldsymbol{\omega}} - g^*\|_\infty \leq \mathcal{O}(r^{-\alpha}) + \epsilon_q(U, L), \tag{A.7}$$

which completes the proof.

B. Proof of Proposition 3

We exploit the fact that $\mathcal{K}_{\boldsymbol{\theta}}$ and $\mathcal{K}_{\boldsymbol{\psi}}$ are positive semidefinite (PSD) Gram matrices.

Step1: Kernel decomposition. From the definition of the empirical NTK,

$$\mathcal{K}_{\mathrm{td}} = \mathbf{G}_{\boldsymbol{\theta}}\mathbf{G}_{\boldsymbol{\theta}}^{\top} + \mathbf{G}_{\boldsymbol{\psi}}\mathbf{G}_{\boldsymbol{\psi}}^{\top} = \mathcal{K}_{\boldsymbol{\theta}} + \mathcal{K}_{\boldsymbol{\psi}}, \tag{B.1}$$

where both $\mathcal{K}_{\boldsymbol{\theta}}$ and $\mathcal{K}_{\boldsymbol{\psi}}$ are symmetric PSD matrices.

Step 2: Nullspace characterization. Let $\mathbf{x} \in \mathbb{R}^N$ (where N is the number of training samples). Then

$$\mathbf{x} \in \ker(\mathcal{K}_{\mathrm{td}}) \quad \Longleftrightarrow \quad \mathbf{x}^{\top}\mathcal{K}_{\mathrm{td}}\mathbf{x} = 0. \tag{B.2}$$

Using the decomposition,

$$\mathbf{x}^{\top}\mathcal{K}_{\mathrm{td}}\mathbf{x} = \mathbf{x}^{\top}\mathcal{K}_{\boldsymbol{\theta}}\mathbf{x} + \mathbf{x}^{\top}\mathcal{K}_{\boldsymbol{\psi}}\mathbf{x}.$$

Since both $\mathcal{K}_{\boldsymbol{\theta}}$ and $\mathcal{K}_{\boldsymbol{\psi}}$ are PSD,

$$\mathbf{x}^{\top}\mathcal{K}_{\boldsymbol{\theta}}\mathbf{x} \geq 0, \quad \mathbf{x}^{\top}\mathcal{K}_{\boldsymbol{\psi}}\mathbf{x} \geq 0.$$

Thus,

$$\mathbf{x}^{\top}\mathcal{K}_{\mathrm{td}}\mathbf{x} = 0 \quad \Longrightarrow \quad \mathbf{x}^{\top}\mathcal{K}_{\boldsymbol{\theta}}\mathbf{x} = 0 \quad \text{and} \quad \mathbf{x}^{\top}\mathcal{K}_{\boldsymbol{\psi}}\mathbf{x} = 0. \tag{B.3}$$

For PSD matrices, $\mathbf{x}^{\top}\mathcal{K}\mathbf{x} = 0$ implies $\mathcal{K}\mathbf{x} = 0$. Therefore,

$$\mathbf{x} \in \ker(\mathcal{K}_{\boldsymbol{\theta}}) \cap \ker(\mathcal{K}_{\boldsymbol{\psi}}). \tag{B.4}$$

Conversely, if $\mathbf{x} \in \ker(\mathcal{K}_{\boldsymbol{\theta}}) \cap \ker(\mathcal{K}_{\boldsymbol{\psi}})$, then clearly $\mathcal{K}_{\mathrm{td}}\mathbf{x} = 0$. Hence,

$$\ker(\mathcal{K}_{\mathrm{td}}) = \ker(\mathcal{K}_{\theta}) \cap \ker(\mathcal{K}_{\boldsymbol{\psi}}). \tag{B.5}$$

Step 3: Rank inequality. Taking dimensions,

$$\dim\ker(\mathcal{K}_{\mathrm{td}}) = \dim\left(\ker(\mathcal{K}_{\boldsymbol{\theta}}) \cap \ker(\mathcal{K}_{\boldsymbol{\psi}})\right) \leq \min\{\dim\ker(\mathcal{K}_{\boldsymbol{\theta}}), \dim\ker(\mathcal{K}_{\boldsymbol{\psi}})\}.$$

Applying the rank–nullity theorem,

$$\mathrm{rank}(\mathcal{K}) = N - \dim\ker(\mathcal{K}),$$

we obtain

$$\mathrm{rank}(\mathcal{K}_{\mathrm{td}}) = N - \dim\ker(\mathcal{K}_{\mathrm{td}}) \geq N - \min\{\dim\ker(\mathcal{K}_{\boldsymbol{\theta}}), \dim\ker(\mathcal{K}_{\boldsymbol{\psi}})\}. \tag{B.6}$$

Thus,

$$\mathrm{rank}(\mathcal{K}_{\mathrm{td}}) \geq \max\{\mathrm{rank}(\mathcal{K}_{\boldsymbol{\theta}}), \mathrm{rank}(\mathcal{K}_{\boldsymbol{\psi}})\}, \tag{B.7}$$

which proves the claimed inequality.

C. Proof of Variance Scaling under Isotropic Perturbations

Step 1: Bound the parameter-induced output fluctuation. Define

$$\delta_{\boldsymbol{\omega}}(\mathbf{x}) := f_{\boldsymbol{\omega}+\Delta\boldsymbol{\omega}}(\mathbf{x}) - f_{\boldsymbol{\omega}}(\mathbf{x}).$$

By the local Lipschitz property,

$$|\delta_{\boldsymbol{\omega}}(\mathbf{x})| \leq L_{\boldsymbol{\omega}} \|\Delta\boldsymbol{\omega}\|_2.$$

Squaring both sides gives:

$$|\delta_{\boldsymbol{\omega}}(\mathbf{x})|^2 \leq L_{\boldsymbol{\omega}}^2 \|\Delta\boldsymbol{\omega}\|_2^2.$$

Taking expectations yields:

$$\mathbb{E}\big[|\delta_{\boldsymbol{\omega}}(\mathbf{x})|^2\big] \leq L_{\boldsymbol{\omega}}^2\, \mathbb{E}\big[\|\Delta\boldsymbol{\omega}\|_2^2\big]\,. \tag{C.1}$$

Since for any random variable Z,

$$\mathrm{Var}\big[Z\big] \leq \mathbb{E}\big[Z^2\big],$$

we obtain

$$\mathrm{Var}[\delta_{\boldsymbol{\omega}}(\mathbf{x})] \leq \mathbb{E}\big[|\delta_{\boldsymbol{\omega}}(\mathbf{x})|^2\big] \leq L_{\boldsymbol{\omega}}^2\, \mathbb{E}\big[\|\Delta\boldsymbol{\omega}\|_2^2\big]\,. \tag{C.2}$$

Using the isotropic perturbation assumption,

$$\mathbb{E}\big[\|\Delta\boldsymbol{\omega}\|_2^2\big] \leq C_{\boldsymbol{\omega}}^2 d_{\mathrm{eff}}\varepsilon^2,$$

we conclude

$$\mathrm{Var}\big[\delta_{\boldsymbol{\omega}}(\mathbf{x})\big] \leq L_{\boldsymbol{\omega}}^2 C_{\boldsymbol{\omega}}^2 d_{\mathrm{eff}}\varepsilon^2. \tag{C.3}$$

This is the desired variance scaling for the parameter-perturbation part. In big-$\mathcal{O}$ form,

$$\mathrm{Var}\big[\delta_{\boldsymbol{\omega}}(\mathbf{x})\big] = \mathcal{O}\big(d_{\mathrm{eff}}\varepsilon^2\big). \tag{C.4}$$

Step 2: Add the measurement-noise contribution.

$$\widetilde{f}_{\boldsymbol{\omega}}(\mathbf{x}) = f_{\boldsymbol{\omega}}(\mathbf{x}) + \delta_{\boldsymbol{\omega}}(\mathbf{x}) + \xi(\mathbf{x}).$$

Since $f_{\boldsymbol{\omega}}(\mathbf{x})$ is deterministic for fixed $\mathbf{x}$ and $\boldsymbol{\omega}$,

$$\mathrm{Var}\big[\widetilde{f}_{\boldsymbol{\omega}}(\mathbf{x})\big] = \mathrm{Var}\big[\delta_{\boldsymbol{\omega}}(\mathbf{x}) + \xi(\mathbf{x})\big].$$

If $\delta_{\boldsymbol{\omega}}(\mathbf{x})$ and $\xi(\mathbf{x})$ are independent, then

$$\mathrm{Var}[\widetilde{f}_{\boldsymbol{\omega}}(\mathbf{x})] = \mathrm{Var}[\delta_{\boldsymbol{\omega}}(\mathbf{x})] + \mathrm{Var}[\xi(\mathbf{x})].$$

Using (C.3) and $\mathrm{Var}[\xi(\mathbf{x})] \leq \sigma^2$, we obtain

$$\mathrm{Var}[\widetilde{f}_{\boldsymbol{\omega}}(\mathbf{x})] \leq L_{\boldsymbol{\omega}}^2 C^2 d_{\mathrm{eff}}\varepsilon^2 + \sigma^2. \tag{C.5}$$

Hence,

$$\mathrm{Var}[\widetilde{f}_{\boldsymbol{\omega}}(\mathbf{x})] = \mathcal{O}(d_{\mathrm{eff}}\varepsilon^2) + \mathcal{O}(\sigma^2), \tag{C.6}$$

which is exactly the scaling stated in the robustness section. This proves the variance scaling under isotropic perturbations.

D. Proof of The Generalization Bound (Eq. (33))

Fix a sample set $S = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$. The empirical Rademacher complexity is

$$\widehat{\mathfrak{R}}_S(\mathcal{F}_{\boldsymbol{\omega}}^{(r)}) := \mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{\mathbf{u}\in\mathcal{U}} \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\sigma}_i f_{\mathbf{u}}(\mathbf{x}_i)\right], \tag{D.1}$$

where $\boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_N$ are i.i.d. Rademacher variables.

Define the empirical L_2-norm metric on the function class by

$$d_S(f_{\mathbf{u}}, f_{\mathbf{v}}) := \left(\frac{1}{N}\sum_{i=1}^{N}\big(f_{\mathbf{u}}(\mathbf{x}_i) - f_{\mathbf{v}}(\mathbf{x}_i)\big)^2\right)^{1/2}. \tag{D.2}$$

By the assumed Lipschitz dependence on the effective coordinates,

$$d_S(f_{\mathbf{u}}, f_{\mathbf{v}}) \le \left(\frac{1}{N}\sum_{i=1}^{N} L_{\mathbf{u}}^2 \|\mathbf{u} - \mathbf{v}\|_2^2\right)^{1/2} = L_{\mathbf{u}}\|\mathbf{u} - \mathbf{v}\|_2.$$

So the map $\mathbf{u} \mapsto f_{\mathbf{u}}$ is $L_{\mathbf{u}}$-Lipschitz from $(\mathcal{U}, \|\cdot\|_2)$ into $(\mathcal{F}_{\boldsymbol{\omega}}^{(r)}, d_S)$. Because $\mathcal{U} \subseteq B_2(A) \subset \mathbb{R}^M$, an $\varepsilon/L_{\mathbf{u}}$-cover of $B_2(A)$ in Euclidean norm induces an ε-cover of $\mathcal{F}_{\boldsymbol{\omega}}^{(r)}$ in d_S. Hence, the empirical covering number satisfies:

$$\mathcal{N}\big(\varepsilon, \mathcal{F}_{\omega}^{(r)}, d_S\big) \le \mathcal{N}\left(\frac{\varepsilon}{L_{\mathbf{u}}}, B_2(A), \|\cdot\|_2\right) \le \left(1 + \frac{2AL_u}{\varepsilon}\right)^M,$$

where $M = d_{\text{eff}}$. Therefore,

$$\log \mathcal{N}\big(\varepsilon, \mathcal{F}_{\boldsymbol{\omega}}^{(r)}, d_S\big) \le M \log\left(1 + \frac{2AL_u}{\varepsilon}\right). \tag{D.3}$$

Next, applying Dudley's entropy integral for empirical Rademacher complexity:

$$\widehat{\mathfrak{R}}_S(\mathcal{F}_{\boldsymbol{\omega}}^{(r)}) \le \frac{4\alpha}{\sqrt{N}} + \frac{12}{\sqrt{N}}\int_{\alpha}^{\operatorname{diam}(\mathcal{F}_{\boldsymbol{\omega}}^{(r)}, d_S)} \sqrt{\log \mathcal{N}(\varepsilon, \mathcal{F}_{\boldsymbol{\omega}}^{(r)}, d_S)}\, d\varepsilon,$$

for any $\alpha > 0$. Since $d_S(f_{\mathbf{u}}, f_{\mathbf{v}}) \le L_{\mathbf{u}}\|\mathbf{u} - \mathbf{v}\|_2$ and $\mathcal{U} \subseteq B_2(A)$,

$$\operatorname{diam}(\mathcal{F}_{\boldsymbol{\omega}}^{(r)}, d_S) \le 2AL_u.$$

Substituting the covering-number bound gives

$$\widehat{\mathfrak{R}}_S(\mathcal{F}_{\boldsymbol{\omega}}^{(r)}) \le \frac{4\alpha}{\sqrt{N}} + \frac{12\sqrt{M}}{\sqrt{N}}\int_{\alpha}^{2AL_u} \sqrt{\log\left(1 + \frac{2AL_u}{\varepsilon}\right)}\, d\varepsilon. \tag{D.4}$$

Taking the limit $\alpha \downarrow 0$, and using the change of variables

$$t = \frac{\varepsilon}{2AL_u},$$

we obtain

$$\widehat{\mathfrak{R}}_S(\mathcal{F}_{\boldsymbol{\omega}}^{(r)}) \leq \frac{24AL_{\mathbf{u}}\sqrt{M}}{\sqrt{N}} \int_0^1 \sqrt{\log(1+1/t)}\, dt. \tag{D.5}$$

The integral is finite and bounded by a universal constant, so it can be absorbed into a constant C. Therefore,

$$\widehat{\mathfrak{R}}_S(\mathcal{F}_{\boldsymbol{\omega}}^{(r)}) \leq CAL_u\sqrt{\frac{M}{N}} = CAL_{\mathbf{u}}\sqrt{\frac{d_{\text{eff}}}{N}}. \tag{D.6}$$

Using the manuscript's intrinsic-dimension scaling

$$d_{\text{eff}} = \mathcal{O}(rKd),$$

we conclude

$$\widehat{\mathfrak{R}}_S(\mathcal{F}_{\boldsymbol{\omega}}^{(r)}) = \mathcal{O}\left(\sqrt{\frac{rKd}{N}}\right), \tag{D.7}$$

which proves the claimed scaling.