A. Detailed Results

Here we present detailed experimental results in three tables for each dataset (CIFAR10-C, CIFAR100-C and ImageNet-C). The results for AQR and its baselines are specified for two different batch sizes and three corruption severity levels.

Table 1. Summary of CIFAR10 experiments. Average accuracy (%) for each batch size and severity level. Values show mean±std across corruptions and seeds.

Method	Batch Size 128			Batch Size 512		
	Sev 1	Sev 3	Sev 5	Sev 1	Sev 3	Sev 5
ResNet18 (BN)	59.5±4.4	55.4±8.0	46.8±8.3	59.5±4.4	55.4±8.0	46.8±8.3
TTN	**78.0**±2.0	75.2±2.4	70.2±4.4	**78.3**±2.0	75.5±2.4	70.6±4.4
TENT	**78.0**±2.0	75.2±2.4	70.2±4.4	**78.3**±2.0	75.5±2.4	70.6±4.4
SAR	**78.0**±2.0	75.2±2.4	70.2±4.4	**78.3**±2.0	75.5±2.4	70.6±4.4
AQR	77.9±2.0	**75.3**±2.4	**70.6**±4.2	**78.3**±2.0	**75.6**±2.4	**70.9**±4.2
ResNet26 (GN)	**81.8**±4.3	77.3±5.0	68.2±7.0	81.8±4.3	77.3±5.0	68.2±7.0
TENT	**81.8**±4.3	77.3±5.0	68.2±7.0	81.8±4.3	77.3±5.0	68.2±7.0
SAR	**81.8**±4.3	77.3±5.0	68.2±7.0	81.8±4.3	77.3±5.0	68.2±7.0
AQR	81.7±3.7	**78.6**±4.4	**73.2**±5.7	**81.9**±3.6	**78.9**±4.4	**73.6**±5.6
ViT Base (FT)	92.9±5.1	86.9±8.8	74.0±15.8	92.9±5.1	86.9±8.8	74.0±15.8
TENT	89.9±6.9	80.8±15.1	66.4±22.4	90.0±6.9	80.8±15.1	66.4±22.4
SAR	90.0±6.9	80.9±15.0	66.5±22.4	90.0±6.9	80.9±15.0	66.5±22.4
AQR	**93.2**±3.9	**88.4**±7.0	**78.6**±11.2	**93.4**±3.9	**88.6**±7.0	**78.9**±11.3
ViT Small (TS)	75.7±3.5	68.9±7.4	58.7±14.9	75.7±3.5	68.9±7.4	58.7±14.9
TENT	74.3±3.1	67.2±6.5	57.7±12.9	74.3±3.2	67.1±6.6	57.6±12.9
SAR	74.5±3.1	67.3±6.6	57.7±13.0	74.6±3.2	67.3±6.6	57.7±13.0
AQR	**77.4**±1.7	**74.0**±2.5	**67.3**±9.2	**77.8**±1.7	**74.4**±2.5	**67.7**±9.3

Table 2. Summary of CIFAR100 experiments. Average accuracy (%) for each batch size and severity level. Values show mean±std across corruptions and seeds.

Method	Batch Size 128			Batch Size 512		
	Sev 1	Sev 3	Sev 5	Sev 1	Sev 3	Sev 5
ResNet50 (BN)	63.3±13.6	47.2±16.4	31.1±17.3	63.3±13.6	47.2±16.4	31.1±17.3
TTN	68.5±4.9	63.5±8.0	56.4±10.1	69.5±4.9	64.5±8.0	57.2±10.2
TENT	68.4±4.8	63.5±7.8	56.3±10.0	69.5±4.8	64.5±8.0	57.2±10.1
SAR	**69.7**±4.4	**65.2**±6.9	**58.7**±8.6	**69.7**±4.8	**64.8**±7.8	57.6±9.9
AQR	67.2±4.6	62.4±7.0	55.7±9.0	69.3±4.6	64.6±7.0	**57.7**±9.2
ResNet50 (GN)	62.1±10.1	52.6±14.2	39.5±14.4	62.1±10.1	52.6±14.2	39.5±14.4
TENT	62.2±10.1	52.7±14.2	39.6±14.4	62.2±10.1	52.7±14.2	39.6±14.4
SAR	63.3±9.5	54.2±14.3	40.7±15.5	62.4±10.0	53.0±14.1	39.8±14.5
AQR	**66.0**±4.2	**61.5**±6.5	**54.8**±8.2	**66.9**±4.1	**62.4**±6.5	**55.8**±8.2
ViT Base (FT)	81.2±7.6	72.8±10.5	58.3±14.6	81.2±7.6	72.8±10.5	58.3±14.6
TENT	78.8±9.2	68.9±14.1	53.3±17.9	78.9±9.2	69.0±14.1	53.4±18.0
SAR	79.0±9.1	69.3±13.8	53.8±17.5	78.9±9.1	69.2±13.9	53.7±17.6
AQR	**81.6**±6.3	**74.3**±9.3	**61.6**±12.7	**81.8**±6.3	**74.5**±9.2	**61.8**±12.7
ViT Small (TS)	44.8±3.9	38.7±6.2	31.1±10.3	44.8±3.9	38.7±6.2	31.1±10.3
TENT	44.6±3.5	38.8±5.4	31.2±10.4	44.5±3.6	38.8±5.3	31.2±10.4
SAR	44.5±3.6	38.7±5.4	31.1±10.3	44.4±3.5	38.7±5.3	31.1±10.3
AQR	**45.8**±2.4	**42.9**±3.1	**37.6**±8.0	**46.1**±2.4	**43.2**±3.1	**37.9**±8.1

Table 3. Summary of ImageNet experiments. Average accuracy (%) for each batch size and severity level. Values show mean±std across corruptions and seeds.

Method	Batch Size 128			Batch Size 512		
	Sev 1	Sev 3	Sev 5	Sev 1	Sev 3	Sev 5
ResNet50 (BN)	61.9±6.6	41.6±14.6	19.4±14.4	61.9±6.6	41.6±14.6	19.4±14.4
TTN	67.3±4.4	52.7±12.0	32.5±16.0	67.7±4.4	53.1±12.0	32.9±16.1
TENT	67.4±4.4	52.7±12.0	32.6±16.0	67.7±4.4	53.1±11.9	32.9±16.1
SAR	67.4±4.4	52.8±11.9	32.7±16.0	67.7±4.4	53.2±11.9	33.0±16.1
AQR	**67.4**±3.7	**53.9**±11.9	**33.7**±16.4	**67.8**±3.7	**54.4**±11.8	**34.2**±16.5
ResNet50 (GN)	**70.6**±4.8	55.7±12.5	32.6±16.3	**70.6**±4.8	55.7±12.5	32.6±16.3
TENT	**70.6**±4.8	55.7±12.5	32.6±16.3	70.6±4.8	55.7±12.5	32.6±16.3
SAR	**70.6**±4.8	55.7±12.5	32.6±16.3	70.6±4.8	55.7±12.5	32.6±16.3
AQR	69.0±3.6	**56.6**±11.5	**36.2**±16.7	**70.8**±3.6	**58.1**±11.6	**37.3**±17.0
ViT Base (FT)	71.3±5.0	60.0±12.2	40.6±12.6	71.3±5.0	60.0±12.2	40.6±12.6
TENT	68.0±5.6	54.9±12.4	34.1±11.1	68.0±5.6	54.9±12.4	34.1±11.1
SAR	68.0±5.5	55.0±12.4	34.1±11.1	71.3±4.9	59.8±12.1	40.7±12.5
AQR	**73.9**±3.9	**63.7**±10.3	**45.6**±12.4	**73.9**±3.9	**63.8**±10.3	**45.9**±12.2
ViT Base (TS)	66.6±5.4	49.7±13.2	24.6±13.4	66.6±5.4	49.7±13.2	24.6±13.4
TENT	66.6±5.4	49.7±13.2	24.6±13.4	66.6±5.4	49.7±13.2	24.6±13.4
SAR	66.6±5.4	49.8±13.2	24.6±13.4	66.6±5.4	49.8±13.2	24.6±13.4
AQR	**67.6**±4.7	**52.8**±11.5	**29.0**±13.6	**67.5**±4.7	**52.7**±11.5	**28.9**±13.6

B. Complementary Results for Performance on Model Architecture Variants

Here we include results analogous to Figure **??**, for ResNet (BN) architectures. This allows comparison of TTN against other approaches under the same architecture.
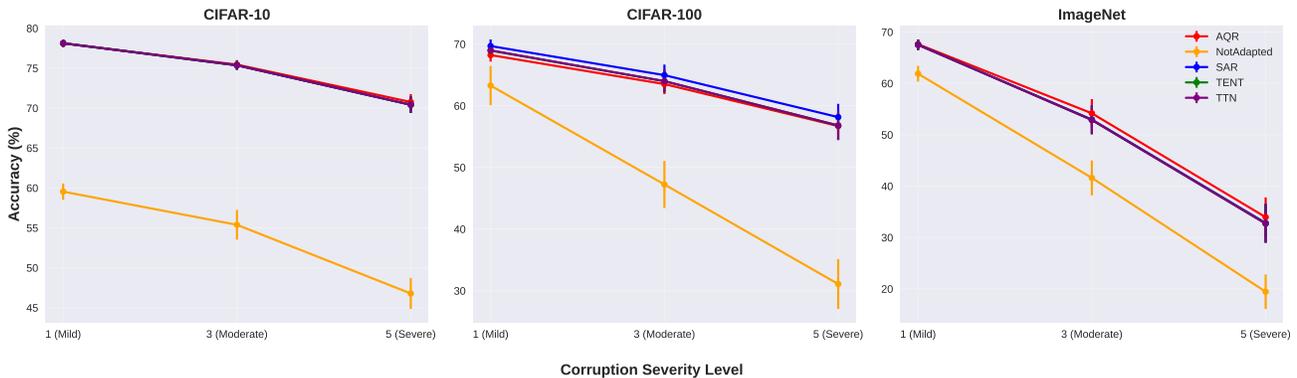

Figure 1. Performance comparison across corruption severity levels for ResNet (BN) models. Results averaged across all corruption types and batch sizes. Error bars represent the standard error of the mean for different experimental conditions.

C. Hyper-parameter Details in Experiments

Table 4. Baseline methods hyperparameters used in our experiments.

Method	Hyperparameter	Value
TENT & SAR	Learning rate	2.5×10^{-4} (SGD with momentum)
	Learning rate (ViT)	$\frac{10^{-3}}{64} \times$ batch size (SGD with momentum)
	Episodic setting	Offline/episodic (2 steps)
SAR	E_0 threshold	$\ln(1000) \times 0.40$
	Layer freezing strategy	Freeze top layers
	ResNet50 (GN) frozen layers	layer 4
	ViT-Base (LN) frozen layers	blocks 9, 10, and 11

Table 5. AQR hyperparameters used in our experiments.

Hyperparameter	Value
Source Data Configuration	
Source samples for computing percentiles (n_s)	10,000
Quantile granularity	101 (percentiles $p_0 \ldots p_{100}$)
Layer Selection Strategy	
ViT-Base-patch16-224 (ImageNet-pre-trained/trained)	Top half of transformer blocks
ViT-Patch4-32 (CIFAR-10/100 from scratch)	All layers
ResNets	All layers
Tail Calibration Strategy	
Tail calibration batch size	100
Tail calibration samples (repeats)	1,000

D. Ablation on Blocks to Adapt

The ablation study evaluates our method across different ViT architectures, where ViT-Base (FT) refers to models pretrained on ImageNet21k and finetuned on ImageNet1k or CIFAR-10 respectively, while ViT-Small (TS) was trained from scratch on CIFAR-10, with results averaged across multiple random seeds and corruption types.

Table 6. Ablation study of applying AQR to different blocks in ViT architecture.

Model	Adapted Blocks	Fine-tuning/Training Dataset	Pretraining Dataset	Accuracy (%)
ViT-Base (FT)	All Blocks	CIFAR-10	ImageNet-21k	87.47±7.78
ViT-Base (FT)	Bottom Half Blocks	CIFAR-10	ImageNet-21k	82.4±16.01
ViT-Base (FT)	Top Half Blocks	CIFAR-10	ImageNet-21k	88.35±7.0
ViT-Base (FT)	All Blocks	ImageNet-1k	ImageNet-21k	36.25±24.44
ViT-Base (FT)	Bottom Half Blocks	ImageNet-1k	ImageNet-21k	22.6±25.28
ViT-Base (FT)	Top Half Blocks	ImageNet-1k	ImageNet-21k	63.7±10.33
ViT-Small (TS)	All Blocks	CIFAR-10	-	74.05±2.46
ViT-Small (TS)	Bottom Half Blocks	CIFAR-10	-	68.93±7.41
ViT-Small (TS)	Top Half Blocks	CIFAR-10	-	68.93±7.41

E. Finite-Sample and Finite Quantile Discretization Analysis of AQR

In this appendix, we provide a finite-sample analysis of the AQR method under practical constraints: finite sample sizes and finite quantile discretization. Our main objective is to establish theoretical guarantees showing that AQR's error converges to zero while TTN's error is lower bounded by a constant bias, therefore providing theoretical justification for AQR's superior empirical performance.

Main Theoretical Goal. For analytical clarity, we focus on a single neuron i and establish per-neuron bounds. The analysis extends to the full network by summing over all neurons or applying union bounds across neurons.

We aim to prove that under finite-sample conditions with quantile discretization, AQR achieves (per neuron):

- **Convergent error**: $\mathrm{MSE}_i(T_{i,K,n}^{\mathrm{AQR}}) \to 0$ as sample sizes n_S, n_T and quantile resolution K increase

- **Explicit convergence rates**: $O(K^{-4}) + O(n_S^{-1}) + O(n_T^{-1})$ with logarithmic factors

- **Superiority over TTN**: For sufficiently large K, n_S, n_T, we have $\mathrm{MSE}_i(T_{i,K,n}^{\mathrm{AQR}}) < \mathrm{MSE}_i(T_i^{\mathrm{TTN}})$

This analysis extends the idealized proof from the main text (Section **??**) to more realistic settings with empirical estimation and computational constraints.

Setting and Notation. We adopt the one-hidden-layer MLP framework from the main text. Consider a network operating on the source domain with input $x \sim P$ on $\mathbb{R}^d$, weight matrix $W \in \mathbb{R}^{m \times d}$, bias $b \in \mathbb{R}^m$, and strictly increasing activation $\phi : \mathbb{R} \to \mathbb{R}$. The hidden representations are:

$$h^S = \phi(Wx + b) \in \mathbb{R}^m, \quad h^T = k(h^S)$$

where $k = (k_1, \ldots, k_m)$ with each $k_i : \mathbb{R} \to \mathbb{R}$ being strictly increasing (representing per-neuron corruptions).

For each neuron $i \in \{1, \ldots, m\}$, we denote:

- P_i: source marginal of h_i^S with CDF F_{P_i}, density f_{P_i}, quantile function $H_i = F_{P_i}^{-1}$

- Q_i: target marginal of h_i^T with CDF F_{Q_i}

- Corruption relationship: $F_{Q_i}(u) = F_{P_i}(k_i^{-1}(u))$

The oracle AQR transformation is: $T_{i,*}^{\mathrm{AQR}}(z) := F_{P_i}^{-1}(F_{Q_i}(z)) = H_i(F_{Q_i}(z))$ which satisfies $T_{i,*}^{\mathrm{AQR}}(h_i^T) = h_i^S$ (perfect recovery). We measure adaptation quality by the per-neuron mean squared error: $\mathrm{MSE}_i(T) := \mathbb{E}[(T_i(h_i^T) - h_i^S)^2]$

Regularity Assumptions. For theoretical tractability, we impose these conditions on each source density f_{P_i}:

$$0 < \underline{f}_i \leq f_{P_i}(x) \leq M_i < \infty \quad \text{for all } x \in \mathrm{supp}(P_i) \tag{1}$$

$$f_{P_i} \in C^1 \text{ with } \|f_{P_i}'\|_\infty \leq L_i \tag{2}$$

These conditions ensure that the quantile function H_i is well-behaved with bounded derivatives:

$$\|H_i'\|_\infty \leq \frac{1}{\underline{f}_i}, \quad \|H_i''\|_\infty \leq \frac{L_i}{\underline{f}_i^3} \tag{3}$$

These bounds follow from the identities $H_i'(u) = 1/f_{P_i}(H_i(u))$ and $H_i''(u) = -f_{P_i}'(H_i(u))/f_{P_i}(H_i(u))^3$.

Practical AQR with Finite Samples and Quantiles. In practice, we only observe finite samples from both domains:

- Source samples: $\{h_j^S\}_{j=1}^{n_S} \overset{\mathrm{iid}}{\sim} P$ with marginals P_i

- Target samples: $\{h_j^T\}_{j=1}^{n_T} \overset{\mathrm{iid}}{\sim} Q$ with marginals Q_i

From these samples, we construct empirical CDFs $\widehat{F}_{P_i}$ and $\widehat{F}_{Q_i}$ for each neuron i. For computational efficiency, we discretize the quantile transformation using K uniform knots:

$$u_j := \frac{j}{K}, \quad j = 0, 1, \ldots, K.$$

The empirical source quantiles are:

$$\widehat{q}_{i,j} := \widehat{F}_{P_i}^{-1}(u_j).$$

We then define $\widetilde{H}_{i,K}$ as the piecewise-linear interpolant satisfying $\widetilde{H}_{i,K}(u_j) = \widehat{q}_{i,j}$ and linear interpolation on each interval $[u_{j-1}, u_j]$. The practical AQR map becomes:

$$T_{i,K,n}^{\text{AQR}}(z) := \widetilde{H}_{i,K}(\widehat{F}_{Q_i}(z))$$

where $n = (n_S, n_T)$ represents the sample sizes.

Main Theoretical Result. Our main result provides explicit finite-sample error bounds for the practical AQR method:

Theorem 1 (Finite-Sample AQR Error Bound). *Under the regularity conditions* (1)-(2), *for any neuron i and any $\delta \in (0, 1)$, with probability at least $1 - 2\delta$:*

$$MSE_i(T_{i,K,n}^{AQR}) \leq 3 \left(\frac{L_i}{8 \underline{f}_i^3} \right)^2 K^{-4} + \frac{3}{\underline{f}_i^2} \varepsilon_S(\delta, n_S)^2 + \frac{3}{\underline{f}_i^2} \varepsilon_T(\delta, n_T)^2 \tag{4}$$

where $\varepsilon_\bullet(\delta, n) := \sqrt{\frac{1}{2n} \log \frac{2}{\delta}}$.
In particular, $MSE_i(T_{i,K,n}^{AQR}) \to 0$ as $K, n_S, n_T \to \infty$ at rates $O(K^{-4})$, $O(n_S^{-1})$, and $O(n_T^{-1})$.

This result establishes that the error for AQR's mapping converges to zero, while the error for TTN's mapping maintains a constant positive bias whenever corruptions are non-affine.

Proof Strategy and Key Lemmas. Our proof decomposes the total error into three manageable components and bounds each using concentration inequalities and approximation theory. We present the analysis through a sequence of lemmas that build towards the main result.

Lemma 1 (Error Decomposition). *For any neuron i and input $z \in \mathbb{R}$:*

$$T_{i,K,n}^{AQR}(z) - T_{i,*}^{AQR}(z) = \underbrace{(\widetilde{H}_{i,K} - H_i)(\widehat{F}_{Q_i}(z))}_{\text{quantile estimation error}} + \underbrace{H_i(\widehat{F}_{Q_i}(z)) - H_i(F_{Q_i}(z))}_{\text{CDF estimation error}} \tag{5}$$

Proof. Considering practical AQR as $T_{i,K,n}^{\text{AQR}}(z) = \widetilde{H}_{i,K}(\widehat{F}_{Q_i}(z))$ and oracle AQR as $T_{i,*}^{\text{AQR}}(z) = H_i(F_{Q_i}(z))$; We add and subtract the intermediate term $H_i(\widehat{F}_{Q_i}(z))$:

$$T_{i,K,n}^{\text{AQR}}(z) - T_{i,*}^{\text{AQR}}(z) = \widetilde{H}_{i,K}(\widehat{F}_{Q_i}(z)) - H_i(F_{Q_i}(z)) \tag{6}$$

$$= \widetilde{H}_{i,K}(\widehat{F}_{Q_i}(z)) - H_i(\widehat{F}_{Q_i}(z)) + H_i(\widehat{F}_{Q_i}(z)) - H_i(F_{Q_i}(z)) \tag{7}$$

$$= (\widetilde{H}_{i,K} - H_i)(\widehat{F}_{Q_i}(z)) + H_i(\widehat{F}_{Q_i}(z)) - H_i(F_{Q_i}(z)) \tag{8}$$

$$\square$$

Intuition: This decomposition separates two distinct error sources: (1) *quantile estimation error* from approximating the true quantile function H_i with the empirical piecewise-linear $\widetilde{H}_{i,K}$, and (2) *CDF estimation error* from using the empirical target CDF $\widehat{F}_{Q_i}$ instead of the true F_{Q_i}.
The next lemma provides concentration bounds for empirical CDF estimation:

Lemma 2 (Concentration via Dvoretzky-Kiefer-Wolfowitz Inequality). *For any $\delta \in (0,1)$, with probability at least $1 - \delta$:*

$$\|\widehat{F}_{Q_i} - F_{Q_i}\|_\infty \leq \varepsilon_T(\delta, n_T) \tag{9}$$

$$\|\widehat{F}_{P_i} - F_{P_i}\|_\infty \leq \varepsilon_S(\delta, n_S) \tag{10}$$

where $\varepsilon_\bullet(\delta, n) = \sqrt{\frac{1}{2n} \log \frac{2}{\delta}}$.

Explanation: The DKW inequality is a fundamental result in empirical process theory that provides uniform concentration bounds for empirical distribution functions. The bound decreases at the optimal parametric rate $O(n^{-1/2})$ with logarithmic dependence on the confidence level.

We next establish how CDF estimation errors transfer to quantile estimation errors:

Lemma 3 (CDF-to-Quantile Error Transfer). *Under the density bounds* (1)*:*

$$\sup_{u \in [0,1]} |\widehat{F}_{P_i}^{-1}(u) - F_{P_i}^{-1}(u)| \leq \frac{\|\widehat{F}_{P_i} - F_{P_i}\|_\infty}{\underline{f}_i} \leq \frac{\varepsilon_S(\delta, n_S)}{\underline{f}_i}$$

Key Insight: The density lower bound $\underline{f}_i$ controls how CDF errors amplify into quantile errors. When the density is bounded away from zero, quantile estimation remains stable.

The following lemma bounds the discretization error from using finite quantiles:

Lemma 4 (Finite-Quantile Discretization Error). *Let $H_{i,K}$ be the piecewise-linear interpolant of the true quantile function H_i on the uniform knots $\{u_j\}_{j=0}^K$. Then:*

$$\|H_i - H_{i,K}\|_\infty \leq \frac{\|H_i''\|_\infty}{8} K^{-2} \leq \frac{L_i}{8\underline{f}_i^3} K^{-2}$$

Explanation: This is a classical result from approximation theory. Linear interpolation of a twice-differentiable function achieves quadratic convergence in the mesh size. The bound depends on the second derivative of the quantile function, which we control through our smoothness assumptions.

Finally, we bound how empirical estimation affects the interpolant:

Lemma 5 (Knot Stability). *The empirical and theoretical piecewise-linear interpolants satisfy:*

$$\|\widetilde{H}_{i,K} - H_{i,K}\|_\infty \leq \max_{0 \leq j \leq K} |\widehat{q}_{i,j} - H_i(u_j)| \leq \frac{\varepsilon_S(\delta, n_S)}{\underline{f}_i}$$

Reasoning: Since both interpolants use the same piecewise-linear basis, their difference is controlled by the maximum knot error. The knot errors are bounded using Lemma 3.

Proof of Main Theorem. **Step 1: Uniform Error Bound.** We combine all lemmas to bound the uniform error for each neuron. From Lemma 1 and the triangle inequality:

$$\sup_{z \in \mathbb{R}} |T_{i,K,n}^{\text{AQR}}(z) - T_{i,*}^{\text{AQR}}(z)| \leq \|\widetilde{H}_{i,K} - H_i\|_\infty + \|H_i'\|_\infty \|\widehat{F}_{Q_i} - F_{Q_i}\|_\infty \tag{11}$$

We further decompose the quantile error using the triangle inequality:

$$\|\widetilde{H}_{i,K} - H_i\|_\infty \leq \|\widetilde{H}_{i,K} - H_{i,K}\|_\infty + \|H_{i,K} - H_i\|_\infty$$

Applying Lemmas 4 and 5, along with $\|H_i'\|_\infty \leq 1/\underline{f}_i$ from (3), we obtain with probability at least $1 - 2\delta$:

$$\sup_{z \in \mathbb{R}} |T_{i,K,n}^{\text{AQR}}(z) - T_{i,*}^{\text{AQR}}(z)| \leq \frac{L_i}{8\underline{f}_i^3} K^{-2} + \frac{\varepsilon_S(\delta, n_S)}{\underline{f}_i} + \frac{\varepsilon_T(\delta, n_T)}{\underline{f}_i} \tag{12}$$

Step 2: Per-Neuron MSE Bound. Since the oracle AQR achieves perfect recovery ($T_{i,*}^{\text{AQR}}(h_i^T) = h_i^S$), we have:
$\text{MSE}_i(T_{i,K,n}^{\text{AQR}}) = \mathbb{E}[(T_{i,K,n}^{\text{AQR}}(h_i^T) - h_i^S)^2] = \mathbb{E}[(T_{i,K,n}^{\text{AQR}}(h_i^T) - T_{i,*}^{\text{AQR}}(h_i^T))^2]$

This MSE is bounded by the squared uniform error: $\text{MSE}_i(T_{i,K,n}^{\text{AQR}}) \leq \left(\sup_z |T_{i,K,n}^{\text{AQR}}(z) - T_{i,*}^{\text{AQR}}(z)|\right)^2$

Using the inequality $(a + b + c)^2 \leq 3(a^2 + b^2 + c^2)$ and (12), with probability at least $1 - 2\delta$:

$$\text{MSE}_i(T_{i,K,n}^{\text{AQR}}) \leq 3\left(\frac{L_i}{8\underline{f}_i^3}\right)^2 K^{-4} + \frac{3}{\underline{f}_i^2}\varepsilon_S(\delta, n_S)^2 + \frac{3}{\underline{f}_i^2}\varepsilon_T(\delta, n_T)^2 \tag{13}$$

This completes the proof of the main theorem. $\qquad\square$

Comparison with TTN and Practical Implications. Our finite-sample analysis reveals a fundamental difference between AQR and TTN:

TTN Limitation: From the main text analysis, TTN uses affine transformations $T_i^{\text{TTN}}(z) = \mu_i^S + \sigma_i^S \frac{z - \mu_i^T}{\sigma_i^T}$ that match first and second moments. However, $\text{MSE}(T^{\text{TTN}}) > 0$ whenever some corruption k_i is non-affine, representing a *constant bias* that does not decrease with sample size or computational resources.

AQR Advantage: Our Theorem 1 shows that for each neuron, AQR's error decreases as:

- $O(K^{-4})$ with quantile resolution (quartic convergence from linear interpolation)

- $O(n_S^{-1/2})$ and $O(n_T^{-1/2})$ with sample sizes (optimal parametric rates)

- Additional logarithmic factors $\log(1/\delta)$ accounting for confidence level

Practical Consequence: For sufficiently large values of K, n_S, and n_T, we have $\text{MSE}_i(T_{i,K,n}^{\text{AQR}}) < \text{MSE}_i(T_i^{\text{TTN}})$ for each neuron, providing theoretical justification for AQR's superior empirical performance in domain adaptation tasks.

Extension to Full Network: The total network MSE is simply $\text{MSE}(T_{K,n}^{\text{AQR}}) = \sum_{i=1}^m \text{MSE}_i(T_{i,K,n}^{\text{AQR}})$. Since each neuron achieves the convergence rates established in our theorem, the network-level performance inherits the same asymptotic behavior, making AQR superior to TTN at the network level as well.

F. Tail Calibration Strategy: Gaussian Estimation

Here we present additional details on the *Gaussian Estimation* strategy described in **??**. Rather than using empirical extrema, this approach models both source and target distributions as Gaussian, and estimates their theoretical tail values, as shown in Equation 14.

$$\text{AQR}(x) = \begin{cases} \left(\frac{x - Q(0)^T}{Q(1)^T - Q(0)^T} \cdot \left(Q(1)^S - Q(0)^S\right)\right) + Q(0)^S & x < p_1^T \\ \left(\frac{x - Q(99)^T}{Q(100)^T - Q(99)^T} \cdot \left(Q(100)^S - Q(99)^S\right)\right) + p_{99}^S & x \geq p_{99}^T \end{cases} \tag{14}$$

Here the source and target quantile functions are

$$Q(p)^S = \beta + \gamma\,\Phi^{-1}(p), \qquad Q(p)^T = \mu(X) + \sigma(X)\,\Phi^{-1}(p),$$

where $\Phi^{-1}(p) = \sqrt{2}\,\text{erf}^{-1}(2p - 1)$ is the probit transform. X denotes all activations from a given channel of a layer for the test input.