NEURIPS 2026 · POSTER

# A Dual-Domain Vision Transformer with Spectral Positional Bias



Learn the attention position bias as Fourier coefficients: 3.25× fewer bias parameters and an exact resize rule.

**Abdulaziz Alshamsi · Abdulla Alghfeli · Chenghua Lin · Hujun Yin**

The University of Manchester

MANCHESTER 1824

The University of Manchester

THE PROBLEM

# Today’s position bias is tied to one grid, or absent

## Swin table

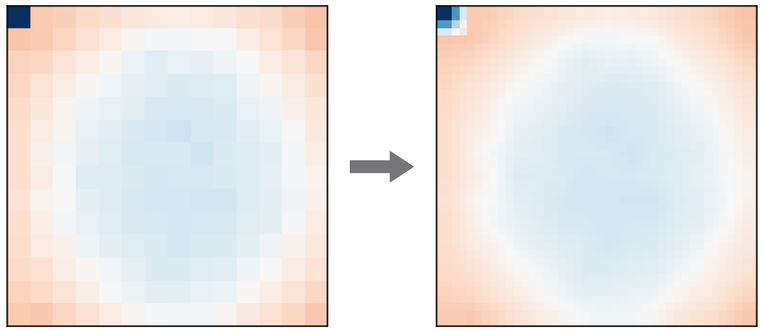
**2,916**

parameters per block, at 4 heads

- One value for each displacement
- A new image size needs bilinear resizing

A learned kernel
14 × 14

Bilinear resize
42 × 42



Bilinear resizing of a learned kernel (SCQB, from our model): the dark corner, each patch’s bias to itself, spreads

## No position bias

**0**

parameters

- No relative position term in attention
- 2.74 pp below SCQB (Tiny, 10.7M params)

OUR IDEA · SCQB, A SPECTRAL-CONDITIONED QUERY-KEY BIAS

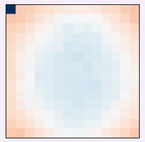
# Learn the bias as Fourier coefficients

**Coefficients**
4 heads × 14 × 8, complex

→

**Inverse FFT**
real, 2D

→

**Spatial kernel**

14 × 14 per head

→

**Wrap-around shift**
196 × 196 bias for all patch pairs, added to the attention logits

**896**
parameters per block, against 2,916 for the table

**3.25×**
fewer bias parameters

**Exact**
resize rule: zero-pad to go finer, truncate to go coarser

Error < $10^{-6}$ at shared grid points; bilinear deviates by up to 1.57

THE BACKBONE · DDR, THE DUAL-DOMAIN RESIDUAL BLOCK

# Every block filters in frequency, then attends

BRANCH 1

## Spectral

FFT filter over all patches

Global mixing in the frequency domain · not gated

→

BRANCH 2

## Attention

$QK^T/\sqrt{d}$ + SCQB bias

Scaled by a learned gate σ(α), one per block

→

BRANCH 3

## Feed-forward

SwiGLU network

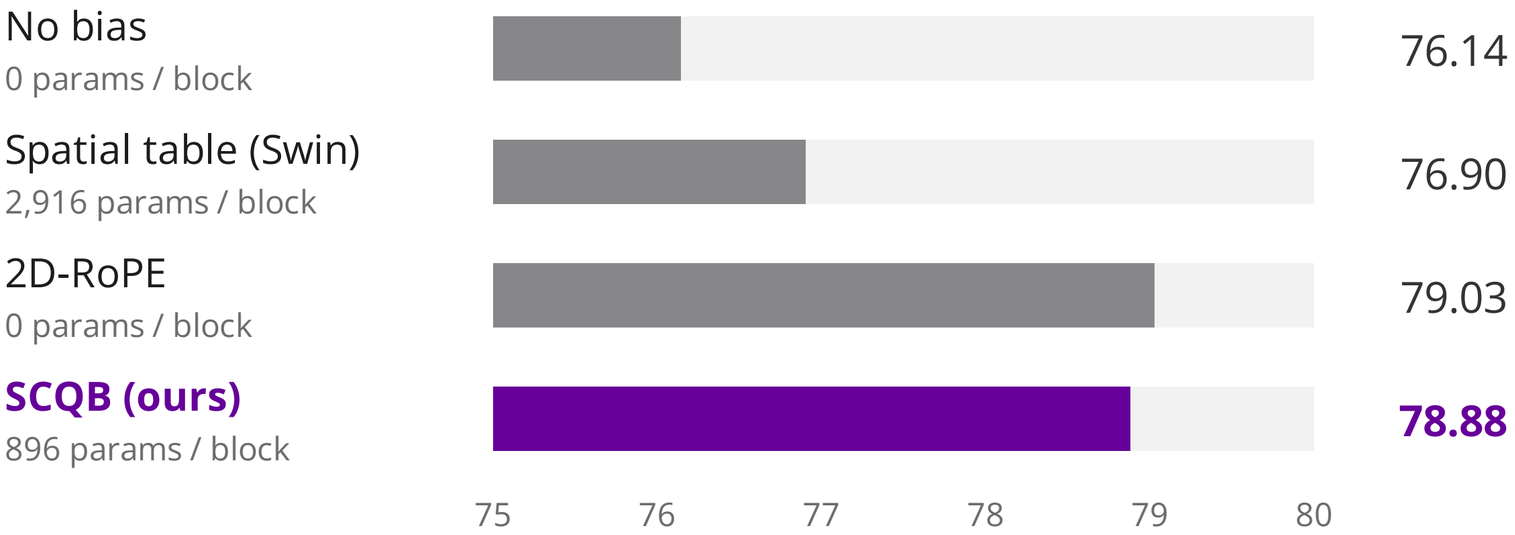
Applied to each patch token

Each branch is a residual update. The three updates run in this order in every block.

The bias and the spectral filter are both Fourier coefficients: the same frequency-bin rule resizes both.

CONTROLLED RESULTS · SAME BACKBONE, RECIPE AND SEED

# SCQB beats the table at Tiny scale and ties it at Small




Top-1 (%); the axis starts at 75%. DDR-Ti-SF (Tiny, 10.7M params), ImageNet-1K, 300 epochs; one run per arm.

**+1.98 pp**

over the spatial table, at Tiny scale

**+2.74 pp**

over no bias, at Tiny scale

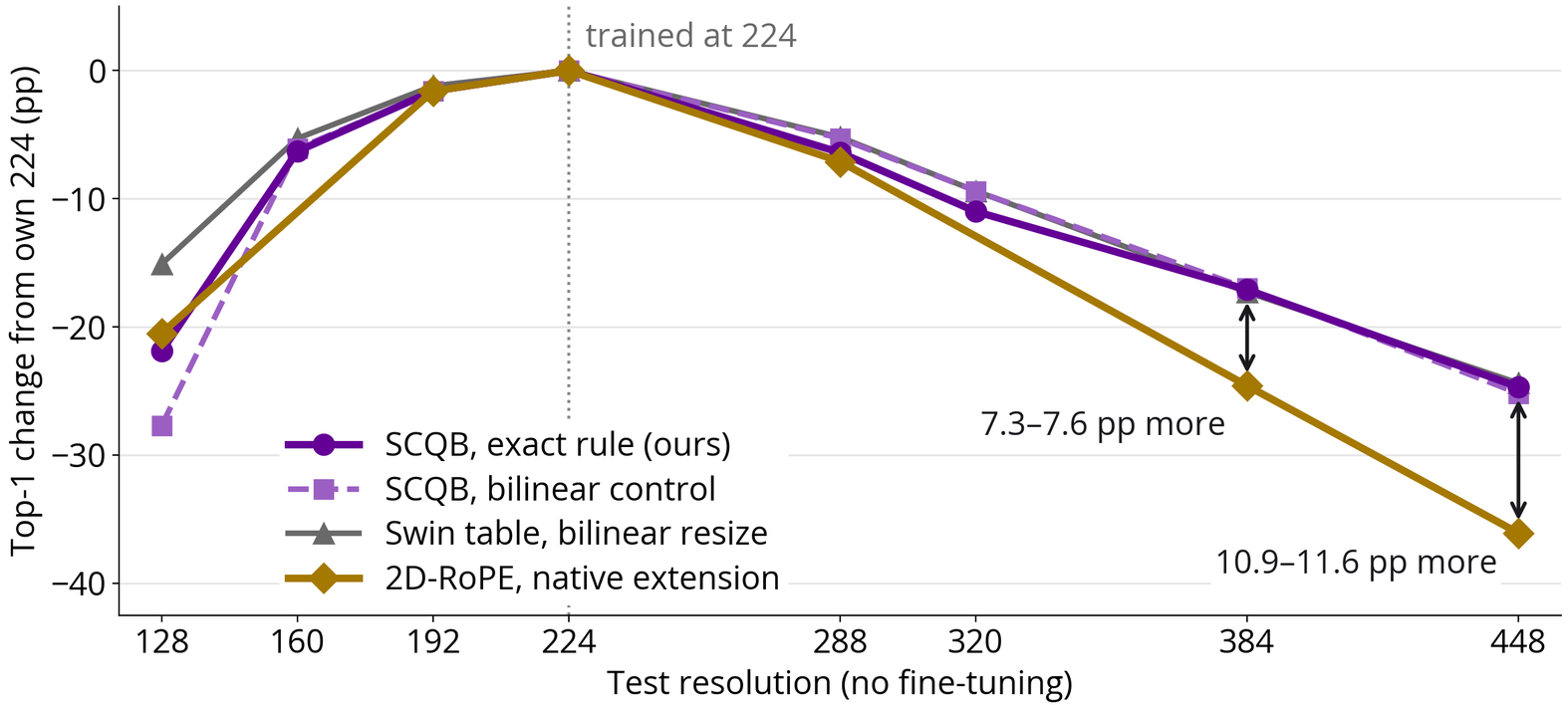
Parameter-free 2D-RoPE ties SCQB at $224^2$ (2D-RoPE +0.15 pp). Next slide: other resolutions.

Small scale (DDR-S, 23.1M params): SCQB 81.2 · spatial table 81.3 · no bias 80.7

**SCQB and the table tie at Small scale.**

NEW RESOLUTIONS · NO FINE-TUNING

# Above the training size, 2D-RoPE drops most




## 2D-RoPE (gold)

Drops 7.3 to 11.6 pp more than the learned biases at $384^2$ and $448^2$

## SCQB and table

Close above $224^2$; the table is the most robust at $128^2$

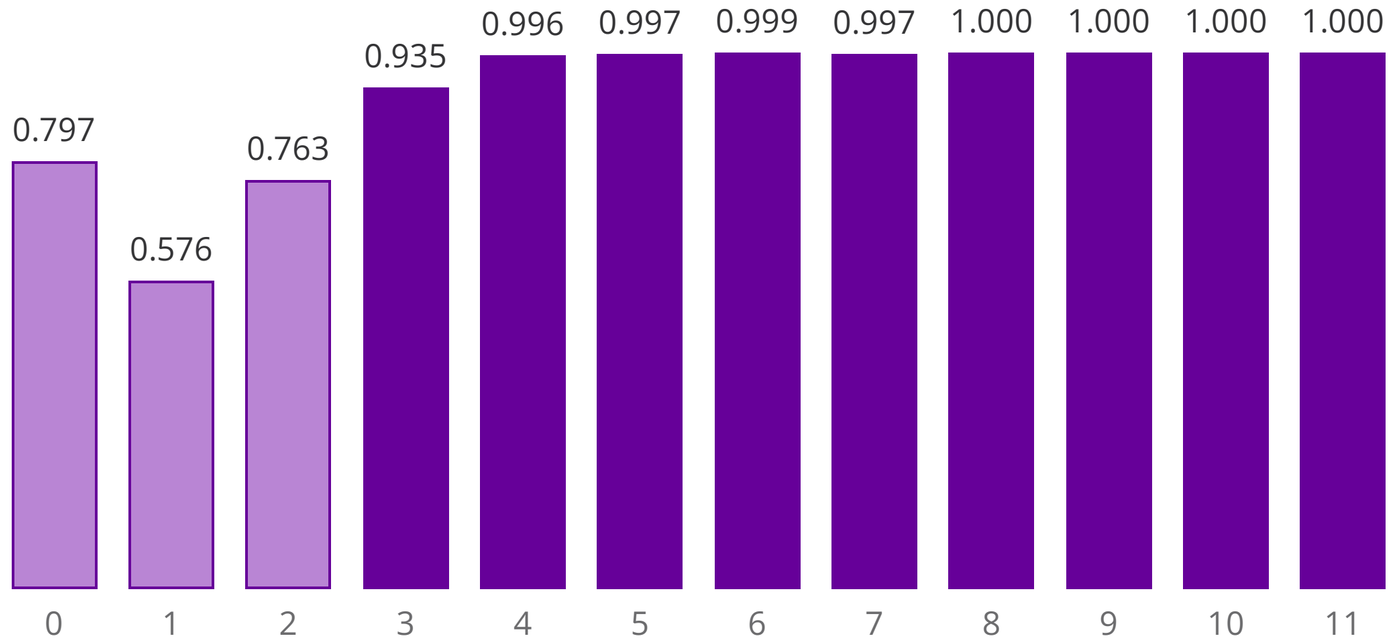
## Exact rule vs bilinear

Dashed line: the same SCQB kernel, resized by bilinear interpolation. The exact rule gives no general accuracy gain.

DDR-Ti-SF models trained at $224^2$. SCQB and table: a matched 150-epoch pair. 2D-RoPE: the 300-epoch model, tested at six of the eight sizes.

WHAT THE MODEL LEARNED

# The gates find a schedule on their own




Attention gate σ(α) per block (0 to 11), DDR-Ti-SF, ImageNet-1K. 1.0 = attention at full weight.

## Blocks 0 to 2 (light)

Gate 0.58 to 0.80: attention has less weight; the spectral branch has more relative weight

## Blocks 3 to 11 (dark)

Gate 0.93 or more: attention near full weight

## ImageNet-100 control

Same DDR model: learned gates 88.3% vs gates frozen to SpectFormer's 4 + 8 schedule 88.0% (one seed; supporting evidence)

SUMMARY

# What holds, and what does not

## What holds

- 3.25× fewer bias parameters than a spatial table
- Exact, closed-form transfer to a new patch grid
- Like the table, a far smaller drop than 2D-RoPE at $384^2$ and $448^2$
- Context only: 78.9% at 10.7M, 82.3% at 36.1M, 82.5% at 63.1M params; consistent with published spectral hybrids

## Limits

- Gain over the table: clear at Tiny, a tie at Small
- 2D-RoPE ties SCQB at $224^2$, with no parameters
- Exact rule: no general accuracy gain over bilinear
- Bias cost: about 2% throughput at $224^2$, 4% at $384^2$
- FFT branch: about 3× slower than a plain ViT of similar size

**Keep position in the frequency domain: a smaller bias with an exact resize rule.**

NEURIPS 2026 · POSTER

# Thank you

Keep position in the frequency domain: a smaller bias with an exact resize rule.

Paper page: neurips.cc/virtual/2026/poster/151216

SCQB reference code: supplementary material of the paper



**Abdulaziz Alshamsi · Abdulla Alghfeli · Chenghua Lin · Hujun Yin**

The University of Manchester

MANCHESTER 1824

The University of Manchester