BEYOND MAGNITUDE: MASSIVE ACTIVATIONS IN DIFFUSION MULTIMODAL LANGUAGE MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion multimodal language models generate responses through repeated masked-token updates, but the computational role of their largest residual activations remains unclear. We investigate these massive values in LaViDa, Dream-VL, and MMaDA through controlled interventions on fixed subsets of five multimodal datasets. Large components occur across image, question, template, and response tokens, yet their functional effects differ across regions, tasks, and models. Peak suppression can disrupt generation, but restoring the original norm does not consistently rescue performance, and retaining the peak alone is insufficient. Amplitude scaling further reveals asymmetric sensitivity. Routing peak access to downstream consumers reveals different dependencies on attention inputs, residual carry, normalized feed-forward reading, and terminal exposure. We develop a conditional theoretical account linking coordinate-concentrated writes and residual propagation to peak formation, and directional perturbations to discrete denoising decisions. These findings characterize massive values through their interaction with surrounding representations and downstream computation, rather than treating magnitude as a direct measure of semantic importance.

1 INTRODUCTION

Diffusion multimodal large language models (dMLLMs) are changing how vision-language responses are generated. Rather than committing to a left-to-right sequence, they repeatedly update a partially masked response conditioned on images and text (Li et al., 2025; Ye et al., 2026; Yang et al., 2025). Multiple positions can be revealed together, and each update becomes context for the next. This creates a distinctive interpretability problem: an internal computation can influence both the present prediction and the evidence available during later denoising. Understanding these models therefore requires identifying not only which representations contain information, but also which structures enable that information to shape a response.

Our investigation begins with a striking feature of those representations. Across network depths, a few residual coordinates rise far above their surroundings (Figure 1). Such *massive values* also occur in image, question, template, and response tokens. Their visibility makes them natural candidates for explaining model behavior, yet magnitude leaves their function ambiguous. A dominant coordinate could support computation shared across inputs, preserve distinctions specific to the current input (Roh et al., 2026), or affect how an answer is expressed. Prior studies connect massive residual activations to computational support and large Q/K values to contextual knowledge (Sun et al., 2024; Jin et al., 2025). In a dMLLM, token roles and repeated masked updates add a further question of where and how the same kind of peak is used.

Deleting a peak is a useful starting point, but it entangles several explanations. The operation removes a coordinate, contracts the vector norm, changes its direction, and alters the inputs received by multiple downstream modules. Even a large performance loss cannot distinguish these effects. Conversely, a peak that matters within a representation need not be sufficient without the surrounding coordinates. A failed answer may reflect loss of task-relevant information, disrupted computation, or premature termination. Separating these outcomes is essential for connecting a visible activation pattern to a specific functional role (Zhang & Nanda, 2024). We therefore organize the study around progressively separating these dependencies, rather than equating activation magnitude with semantic importance.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

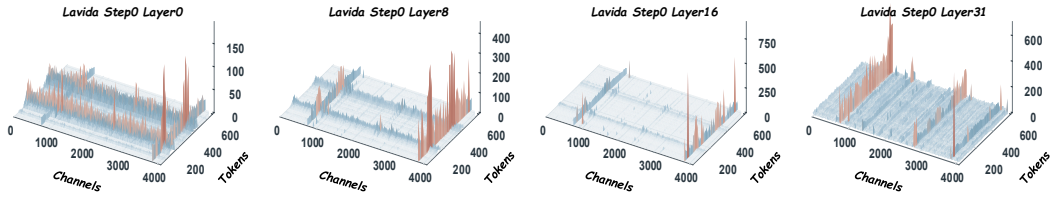


Figure 1: **Massive values across LaViDa layers.** Raw absolute hidden activations at step 0, shown at layers 0, 8, 16, and 31 for the same input. Large peaks appear at multiple depths. Each panel uses an independent vertical scale; no normalization is applied.

We examine LaViDa, Dream-VL, and MMaDA on fixed subsets of five datasets spanning multimodal question answering and captioning. Regional suppression first establishes that widespread occurrence coexists with selective functional dependence. Norm controls then separate scale contraction from peak removal, while retention and amplitude interventions test the representation needed around a peak. Controlled access to QK, V, and the residual continuation localizes downstream dependence. Finally, carry, FFN-reader, and terminal controls distinguish operations that a conventional deletion changes together. This sequence moves from a visible phenomenon to the computations that make it consequential.

The findings support a view of role-dependent computational support. In Dream-VL, suppressing template peaks collapses all four QA scores, while suppressing image peaks largely preserves ChartQA and captioning. Restoring a deleted vector's norm can leave generation impaired, while retaining only its largest coordinate can also destroy useful behavior. Moreover, restoring additive residual carry can recover template performance even when correcting the normalized FFN input does not. These contrasts show why the surrounding representation and downstream use matter alongside the peak itself. Our theoretical account connects concentrated residual writes to peak formation, and directional perturbations to the discrete decisions of masked generation. The resulting explanation makes massive values a means of studying how dMLLMs sustain computation, rather than treating them as isolated stores of answer content.

Contributions.

- We characterize regional peak dependence across three dMLLMs and five fixed multimodal task subsets.
- We disentangle norm, retained support, and amplitude, then separate attention access, residual carry, FFN reading, and terminal exposure.
- We relate these findings through a geometric and temporal account of concentrated writing and its influence on masked denoising.

2 Related Work

Diffusion multimodal models. LaViDa, LLaDA-V, and Dream-VL extend diffusion language modeling to visual instruction following (Li et al., 2025; You et al., 2026; Ye et al., 2026), while MMaDA supports multimodal understanding and generation (Yang et al., 2025). Recent analyses examine sparse features, early visual dependence, and visual attribution (Wang et al., 2026; Kim et al., 2026; Zuo et al., 2026). We focus on the residual components passed between blocks and the operations through which they affect iterative generation.

Massive activations. Prior work studies massive residual activations, large Q/K coordinates, and visual attention sinks (Sun et al., 2024; Gan et al., 2026; Jin et al., 2025; Kang et al., 2025). These are related but distinct computational objects. Subsequent analyses separate massive activations from sinks and study their formation (Sun et al., 2026; Shi et al., 2026). We investigate native-basis residual peaks in dMLLMs through regional and pathway interventions. Appendix I discusses inference design, normalization, and coordinated rotations in greater detail.

3 PRELIMINARIES

Massive values and token roles. Let $\mathbf{h}_{\ell,k,i} \in \mathbb{R}^d$ be the block-output residual at layer ℓ, denoising step k, and token i. Massive values are coordinates with unusually large absolute magnitudes relative to their surroundings (Sun et al., 2024). We define the peak and its tied locations in the model's native basis as

$$m_{\ell,k,i} = \max_c |h_{\ell,k,i,c}|, \qquad \mathcal{C}^\star_{\ell,k,i} = \arg\max_c |h_{\ell,k,i,c}|. \tag{1}$$

Indices are zero-based. We distinguish image, question, template/special, and response tokens. The response state $Z^{(k)}$ begins fully masked and is updated by token predictions and position selection at each step. Cached prefix residuals are computed at prefill, while response residuals evolve during denoising. The full generation notation is in Appendix I.3.

Models and tasks. We use frozen LaViDa-LLaDA v1.0 Instruct, Dream-VL-7B, and MMaDA-8B-MixCoT, retaining native input formats and decoders. ScienceQA-IMG tests scientific multiple-choice reasoning (Lu et al., 2022), GQA compositional visual QA (Hudson & Manning, 2019), TextVQA image-text reading (Singh et al., 2019), ChartQA chart reasoning (Masry et al., 2022), and COCO captioning (Lin et al., 2014). Each task uses 48 fixed inputs shared across backbones and conditions. The targeted mechanistic controls use ChartQA and COCO.

4 CHARACTERIZING MASSIVE VALUES

4.1 FROM WIDESPREAD PEAKS TO REGIONAL DEPENDENCE

Occurrence across token roles. Figure 2 identifies image, question, template, and response regions in the attention map for a representative LaViDa input. The corresponding raw residual slices contain a prominent peak in a sampled token from each region. Massive values are therefore not confined to visual tokens, but their presence alone does not establish a shared computational function. A functional characterization must compare the same intervention across token roles, rather than infer function from peak height alone. We move from this observation to region-specific interventions that test how generation depends on these peaks.

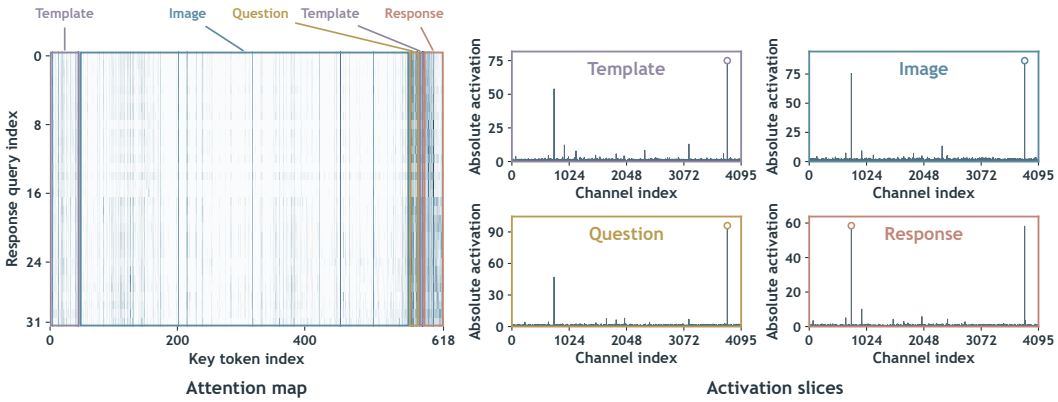


Figure 2: **Massive values across token roles.** LaViDa, layer 16. Left: head-mean attention at step 15 (log colors). Right: raw residual magnitudes with independent vertical scales. Prefix profiles are from prefill; the response profile is from step 15. Matching colors identify token roles.

Regional peak suppression. To test regional dependence, we reselect each target token's largest absolute coordinate at every layer and effective denoising update. *Zero* sets this coordinate to zero. *Mean* and *MinAbs* replace it with the remaining coordinates' signed mean or their minimum magnitude with the original peak sign. *Random* zeros a non-peak coordinate, and *Native* leaves activations unchanged. Model weights, inputs, and generation settings are unchanged.

Table 1: **Regional peak-reset interventions in Dream-VL.** The same 48 inputs per dataset are evaluated with 32 denoising updates and 128 response positions. Native is shared across all four token regions.

		Question answering (%)				COCO (×100)		
Region	Treatment	SQA	GQA	TextVQA	ChartQA	CIDEr	SPICE	B-4
Native	—	**93.75**	**20.83**	**68.75**	**91.67**	**132.96**	**20.85**	**26.17**
Image	Zero	91.67	29.17	60.42	91.67	131.45	19.11	28.45
	Mean	91.67	27.08	61.25	91.67	127.16	19.20	28.06
	MinAbs	93.75	29.17	62.71	89.58	135.08	19.35	30.34
	Random	93.75	18.75	65.63	91.67	133.09	21.35	29.98
Question	Zero	85.42	6.25	62.71	91.67	130.20	20.50	28.57
	Mean	91.67	2.08	62.71	91.67	125.11	19.54	27.30
	MinAbs	89.58	2.08	60.63	91.67	127.87	21.05	27.81
	Random	93.75	18.75	63.33	91.67	134.69	21.73	27.67
Template	Zero	0	0	0	0	0	0	0
	Mean	0	0	0	0	0	0	0
	MinAbs	0	0	0	0	0	0	0
	Random	93.75	20.83	68.75	91.67	131.78	20.77	27.03
Response	Zero	89.58	20.83	28.13	41.67	47.53	14.57	9.22
	Mean	89.58	22.92	30.42	45.83	51.16	16.35	10.78
	MinAbs	89.58	20.83	26.88	39.58	53.62	16.00	10.26
	Random	93.75	18.75	70.00	89.58	129.88	21.05	27.17

Table 1 separates peak occurrence from functional dependence. Dream-VL template suppression collapses all four QA scores, whereas image suppression largely preserves ChartQA and captioning. Response and question effects vary by task. Mean and MinAbs broadly reproduce this pattern, although MinAbs often becomes zero at native precision. Broad template failure points toward generation support rather than selective removal of one visual fact. This separation shifts the focus from which region contains large activations to which computation depends on them. In particular, a template-associated peak can support task performance without itself being a visual-content token. We next separate peak loss from vector contraction. Full results and diagnostics are in Appendices A and H.

4.2 Separating Norm Loss from Peak Removal

Complementary local controls. A peak can dominate vector norm, so its removal also contracts the representation. For a current vector h, let z be its peak-deleted counterpart, $r = \|h\|_2$, and $q = \|z\|_2$. *Zero+Norm* writes $(r/q)z$, restoring the norm without recovering the peak. *Scale only* writes $(q/r)h$, reproducing the contraction while preserving direction. Both controls use their own current states and unchanged inputs and generation settings.

Table 2: **Local norm controls for regional peak suppression.** Results use the same 48 inputs per dataset for all three models.

	ChartQA relaxed accuracy (%)				COCO CIDEr (×100)			
Region	Native	Zero	Zero+Norm	Scale only	Native	Zero	Zero+Norm	Scale only
LaViDa								
Image	70.83	0	0	4.17	72.19	0.02	0.05	0.54
Question	70.83	0	0	0	72.19	0	0.18	2.20
Template	70.83	0	0	4.17	72.19	0.04	0	1.46
Response	70.83	0	0	0	72.19	0	0.81	0
Dream-VL								
Image	91.67	91.67	89.58	87.50	132.96	131.45	125.27	130.19
Question	91.67	91.67	91.67	91.67	132.96	130.20	130.75	130.07
Template	91.67	0	0	91.67	132.96	0	0	128.21
Response	91.67	41.67	56.25	81.25	132.96	47.53	37.63	127.99
MMaDA								
Image	12.50	0	0	2.08	106.05	0.06	0	0.26
Question	12.50	0	0	2.08	106.05	40.80	49.03	40.30
Template	12.50	0	0	0	106.05	0.11	0.24	0
Response	12.50	0	0	0	106.05	0.77	0.15	0.74

Norm restoration is not functional restoration. Dream-VL template Zero+Norm remains at zero while Scale only preserves ChartQA accuracy and most captioning performance (Table 2). The original norm therefore does not recover the deleted representation's function. Yet LaViDa and MMaDA can suffer under Scale only, and Dream-VL response norm restoration helps ChartQA but worsens CIDEr. Scale and representation composition have distinct effects. These outcomes separate two forms of damage: changing the global scale and changing the relative contribution of coordinates. Their differing task effects show why vector norm alone cannot account for the functional pattern. We next reverse suppression, since dependence on a peak does not establish sufficiency without its surroundings.

4.3 Testing Sufficiency and Amplitude Dependence

From a dominant coordinate to a usable representation. We first retain only each target token's peak, comparing one random coordinate with and without norm restoration. Response retention yields zero ChartQA accuracy and near-zero captioning scores across models even after norm restoration (Appendix C). The peak alone is therefore insufficient under sustained regional intervention. We expand retention to the top $1\%, 5\%, 10\%, 20\%$, or 50% of coordinates, alongside equally sized random sets.

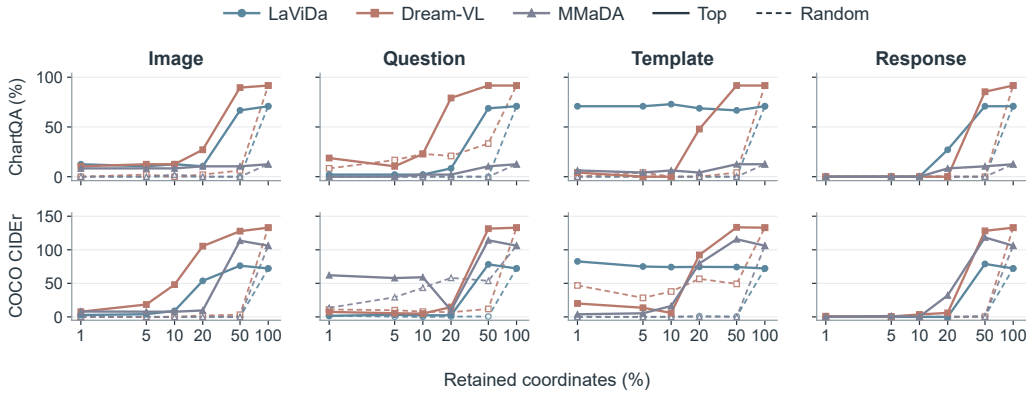


Figure 3: **Regional retention without norm restoration.** Solid and dashed lines denote Top and Random on 48 inputs per task. Scores are point estimates, with CIDEr scaled by 100. The percentage axis is logarithmic and shares Native at 100%.

Distributed support around the peaks. Dream-VL response Top retention reaches 85.42% ChartQA accuracy at 50%, whereas Random retention remains at zero (Figure 3). Magnitude can thus identify useful support even when the single largest coordinate is insufficient. Recovery thresholds differ across regions and backbones, indicating dependence on combinations of components and their surrounding distinctions. The deletion and retention results are therefore complementary. Suppression asks whether the intact system depends on its peak, whereas retention asks whether the peak can replace the rest of the representation. Together, they suggest that a dominant component operates with, rather than substitutes for, a wider set of features. Complete Raw/Norm results are in Appendix H.

Changing amplitude without removing the surrounding coordinates. We next preserve support and multiply only the current peak by $\alpha \in \{0.25, 0.5, 2, 4\}$, with Native at $\alpha = 1$. Other coordinates and signs are unchanged. Figure 4 uses Raw conditions without norm restoration.

Dream-VL tolerates image/question attenuation better than amplification, while template/response tokens are also sensitive to attenuation. Image ChartQA remains 89.58% at $\times 0.25$ and $\times 0.5$, but falls to zero at $\times 2$. A larger peak is not simply a stronger useful signal. Scaling changes its balance with surrounding coordinates and their normalized downstream inputs. The informative comparison is therefore not simply peak versus no peak. It is whether the relative weighting of a peak and its surrounding components remains compatible with the operations that consume them. Fixed-reference controls extend this analysis in Appendix H. We next trace which computations depend on this configuration.

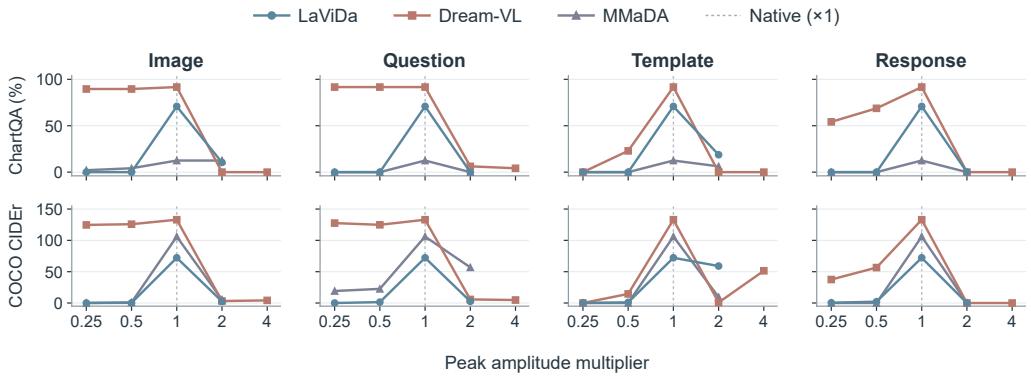


Figure 4: **Regional sensitivity to peak amplitude.** Scores are point estimates on 48 inputs per task without norm restoration, with CIDEr scaled by 100. The logarithmic axis includes Native at $\times 1$. LaViDa and MMaDA $\times 4$ conditions are unscorable, not zero.

4.4 Locating Downstream Peak Dependence

Controlled access to the same peak. The preceding edits combine changes to attention and residual continuation. We separate them by routing the unchanged current block output or its peak-zeroed counterpart independently to the next QK, V, and Direct inputs. Table 3 tests all eight configurations. A cross denies only the selected coordinate, leaving the rest of the branch intact. Direct comprises residual carry, the natural downstream FFN, and the applicable terminal readout. Every configuration follows its own trajectory on the same models, regions, and task subsets. QK access can change how information is selected, while V access changes the features being aggregated. Direct preserves a distinct route through the residual computation. The access matrix therefore distinguishes several uses of the same source coordinate.

Joint use rather than one sufficient route. Removing Direct access sharply reduces template and response performance despite retained QK/V access. Conversely, Direct-only access largely preserves Dream-VL performance but fails for LaViDa and MMaDA template tokens. LaViDa instead sustains much higher scores when QK and Direct retain access together. Thus, residual continuation matters, but its effectiveness can depend on the accompanying attention computation. This contrast narrows the final dissection to operations still bundled inside Direct. Paired effects and secondary metrics are in Appendices E and H.

4.5 Resolving the Direct Continuation

Separating transmission from terminal exposure. We separate Direct's additive carry, normalized FFN reading, and terminal exposure while retaining current-state QK/V access. Base deletes the internal Direct peak but protects terminal readout, unlike D110. R additionally restores carry at the FFN addition without correcting the FFN input. T deletes only terminal peaks, retaining native internal computation.

R substantially recovers template performance beyond Base's preserved attention access, but response recovery is less complete (Figure 5). Dream-VL response CIDEr rises from 43.15 under D110 to 87.94 under Base, while terminal-only deletion degrades Native. Thus, protecting terminal exposure and restoring internal carry resolve different dependencies. We next test whether corrected FFN reading can replace carry restoration.

Correcting the reader without restoring the carry. G corrects Base's FFN normalization denominator, P restores the selected numerator coordinate, and PG corrects both, all without restoring carry. Figure 6 separates prespecified response comparisons from exploratory question results.

Table 3: **Downstream access to regional peaks.** ✓ retains the peak and × zeros it for that consumer, without disabling the branch. ChartQA accuracy (%) and COCO CIDEr (×100) use the same 48 inputs per task. Native is shared across regions. QK, V, and Direct are defined in Section 4.4.

Region	Peak access			LaViDa		Dream-VL		MMaDA	
	QK	V	Direct	ChartQA	CIDEr	ChartQA	CIDEr	ChartQA	CIDEr
Native	✓	✓	✓	**70.83**	**72.19**	**91.67**	**132.96**	**12.50**	**106.05**
Image	×	×	×	0	0.02	91.67	131.45	0	0.06
	×	✓	✓	50.00	53.83	91.67	128.46	8.33	51.80
	✓	×	✓	72.92	73.31	91.67	125.56	12.50	116.81
	✓	✓	×	2.08	0.04	85.42	131.98	0	0.13
	✓	×	×	2.08	0.07	89.58	130.43	0	0.02
	×	✓	×	0	0	87.50	131.44	0	0
	×	×	✓	0	7.01	89.58	135.67	8.33	8.49
Question	×	×	×	0	0	91.67	130.20	0	40.80
	×	✓	✓	2.08	4.04	91.67	126.80	0	61.31
	✓	×	✓	75.00	76.02	91.67	125.61	10.42	109.85
	✓	✓	×	0	3.93	91.67	129.59	0	0
	✓	×	×	2.08	1.42	91.67	129.47	0	0
	×	✓	×	0	0.32	91.67	124.59	0	3.51
	×	×	✓	0	0.11	91.67	121.57	0	57.01
Template	×	×	×	0	0.04	0	0	0	0.11
	×	✓	✓	0	4.68	89.58	121.00	4.17	85.66
	✓	×	✓	68.75	70.63	89.58	138.96	8.33	95.54
	✓	✓	×	0	0.09	6.25	0.12	0	0.34
	✓	×	×	2.08	0	16.67	0.41	0	0.03
	×	✓	×	0	0	14.58	0	0	0.17
	×	×	✓	0	0	89.58	127.53	0	0
Response	×	×	×	0	0	41.67	47.53	0	0.77
	×	✓	✓	0	0.57	91.67	125.47	8.33	88.55
	✓	×	✓	66.67	82.69	85.42	136.26	12.50	101.26
	✓	✓	×	2.08	0	37.50	43.15	0	0.39
	✓	×	×	0	1.45	47.92	40.33	0	1.08
	×	✓	×	0	0.06	45.83	35.37	0	0.09
	×	×	✓	0	0.54	79.17	120.11	2.08	28.21

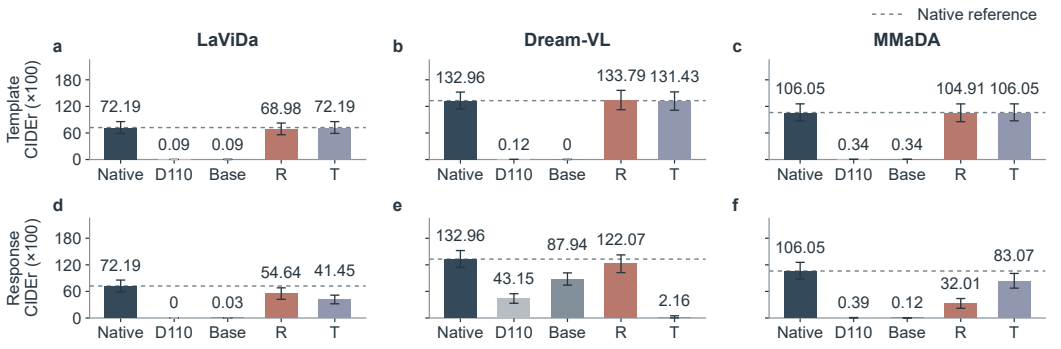


Figure 5: **Carry restoration and terminal exposure.** COCO CIDEr (×100) for the defined policies. Labels are point estimates, whiskers are 95% pointwise image-cluster bootstrap intervals over 48 paired images, and dashed lines mark Native.

Template PG remains near zero despite much higher R scores across backbones, separating normalized reading from additive transmission. Each policy repairs a different operation on the current trajectory. Substantial carry recovery alongside failed reader correction identifies additive transmission as the stronger functional dependence in this template setting. All response G-versus-Base adjusted intervals include zero. Yet MMaDA question CIDEr rises from zero to 63.54 under G, showing a region in which denominator correction matters. Together, these controls locate distinct dependencies in reading, transmission, and exposure. We next connect this operation-dependent use to residual geometry and discrete denoising.

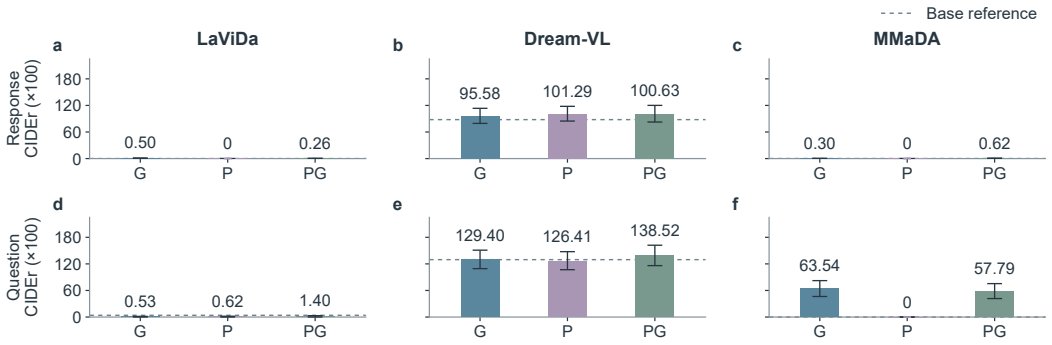


Figure 6: **FFN-reader correction without carry restoration.** G, P, and PG correct the denominator, numerator, or both. Dashed lines mark Base, with scores and whiskers as in Figure 5. Response comparisons are prespecified and Question comparisons are exploratory.

5 Causal Analysis of Massive Values in dMLLMs

The experiments separate magnitude, representation composition, and downstream use. We connect them through concentrated residual writing and directional propagation, then explain how continuous perturbations enter a dMLLM's discrete denoising decisions. This account separates peak formation from peak use: accumulation determines what is available, while downstream geometry determines its consequences.

5.1 Why Massive Values Emerge

Attention writes dominate the selected first-layer common projection in the archived accounting (Appendix G.5), but large writes alone do not imply isolated peaks. At a fixed denoising step, write a token residual as $h_L = a_L u + e_L$, with fixed unit u in the native channel basis and $e_L \perp u$. For actual layer updates $w_\ell = h_{\ell+1} - h_\ell$,

$$a_L = a_0 + \sum_{\ell<L} u^\top w_\ell, \qquad \|h_L\|_\infty \geq |a_L|\|u\|_\infty - \|e_L\|_\infty. \tag{2}$$

Non-cancelling writes can build a_L, while coordinate concentration in u and a controlled remainder yield prominent coordinates. A dense direction can have arbitrarily large amplitude without an isolated maximum. Pre-normalization also leaves the residual bypass intact (Xiong et al., 2020). For an immediate FFN map $\mathcal{F}_0(h) = h + \mathcal{M}(\mathcal{N}_0(h))$, with $\mathcal{N}_0(h) = \sqrt{d}\,\Gamma h/\|h\|_2$ and diagonal gain Γ,

$$\mathcal{F}_0(ch) - \mathcal{F}_0(h) = (c-1)h, \qquad c > 0, \quad h \neq 0. \tag{3}$$

Thus bounded normalized inputs do not impose the same bound on residual magnitude. The masked-generation structure supplies one source of shared writes. On the inspected LaViDa first-layer path, identical MASK values contribute common projected directions across query positions, with softmax-bounded coefficients rather than linear growth with MASK count. This provides a computational route from shared masked inputs to concentrated residual writes. The archived writer measurements support this route at the inspected nodes, while the decomposition also permits later writes to reinforce or cancel it (Appendix G.5).

5.2 Why Massive Values Affect dMLLMs

Peak deletion changes direction, not just scale. Write $h = me_j + b$, $b_j = 0$, $H = \|h\|_2$, $B = \|b\|_2 > 0$, and $\widetilde{h} = Hb/B$. For RMSNorm $\mathcal{N}_\epsilon(h) = \Gamma h/r$, $r^2 = H^2/d + \epsilon$ (Zhang & Sennrich, 2019),

$$\cos\angle(h,b) = \frac{B}{H}, \qquad \mathcal{N}_\epsilon(\widetilde{h}) - \mathcal{N}_\epsilon(h) = \frac{H}{r}\Gamma\left(\frac{b}{B} - \frac{h}{H}\right). \tag{4}$$

Restoring the norm therefore leaves the deletion-induced angular change intact (Figure 7). Conversely, retaining only the peak discards distinctions in b, even when the peak carries most energy.

This explains the separation between energy concentration and usable retained support observed in Section 4.3. Since r depends on m^2, amplifying a peak also attenuates the remaining normalized components (Qiu et al., 2026). Attenuation and amplification therefore need not have symmetric task effects (Appendix G.2).

The reader and carry respond differently. Let $P_{\|} = hh^{\top}/H^2$, $P_{\perp} = I - P_{\|}$, and $\mathcal{F}_{\epsilon}(h) = h + \mathcal{M}(\mathcal{N}_{\epsilon}(h))$. Its exact local derivative is

$$D\mathcal{F}_{\epsilon}(h)[\delta] = \delta + J_{\mathcal{M}} \frac{\Gamma}{r} \left(P_{\perp} + \frac{\epsilon}{r^2} P_{\|} \right) \delta. \tag{5}$$

The first term is residual carry, while the second is the normalized FFN response. At zero epsilon, radial perturbations vanish from the reader but remain in carry, whereas transverse perturbations can affect both. In the tested controls, R restores the additive carry, whereas PG corrects only the FFN reader. Their template contrast shows that correcting normalized reading need not replace residual transmission. Learned projections and attention visibility make these effects region-dependent, consistent with the pathway contrasts. Finite-deletion identities, rather than this local derivative, govern large edits.

In a dMLLM, these continuous changes meet two discrete boundaries: token identity and the confidence ranking of positions selected for revelation. Crossing either can alter $Z^{(k+1)}$ and hence the next forward's context, without requiring the original hidden perturbation to persist. Archived single-pulse state exchanges support this local route, including stopping and formatting changes. A stopping or formatting decision can therefore alter the continuation without directly encoding a visual fact, explaining why answer form and factual content require separate analysis. Identical complete persistent states, future randomness, and remaining intervention policies instead give identical suffixes. Thus discrete state changes can outlast the original perturbation, while hidden changes that leave consequential decisions intact can vanish without changing the answer. Proofs and protocol boundaries are in Appendix G.

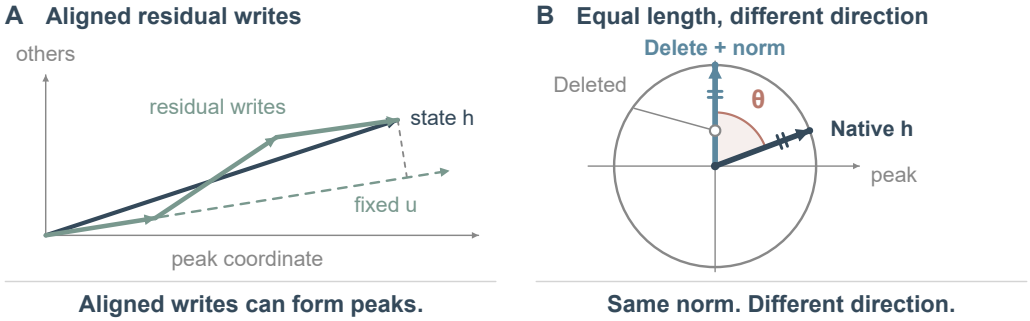


Figure 7: **Directional interpretation of massive values.** Conditional constructions, not fitted model activations. Left: aligned writes from $h_0 = 0$, with a dotted orthogonal remainder. Right: equal-length arrows meet the same circle, while the deletion-induced angular change remains.

6 Conclusion

We investigated massive residual values through a sequence of regional, geometric, and pathway interventions in three dMLLMs. Peaks occur across token roles, but their presence does not imply interchangeable function. Norm restoration separates scale from representation composition, retention reveals the importance of surrounding coordinates, and amplitude controls expose asymmetric sensitivity. Downstream access and carry/reader controls then show why these effects depend on how a component is transmitted and consumed. Our theoretical analysis connects this evidence to concentrated residual writing, directional perturbations, and discrete denoising decisions. The resulting view is that massive values participate in a structured computation rather than acting as isolated containers of semantic importance. The fixed exploratory subsets establish these distinctions within the tested settings, leaving their training origins and broader generalization as directions for future work.

Reproducibility Statement

Section 3 identifies the backbones and task subsets. Sections 4 and 5 define the interventions and theoretical assumptions. Appendices A–F report the full or supporting matrices, generation settings, control semantics, uncertainty procedures, and unavailable conditions. Appendix G provides derivations and separates archival protocols from the main task matrices. Appendix H contains the complete primary score matrices, with full-precision scores and saved intervals supplied as source data. The experiments use fixed exploratory subsets.

AI Use Statement

We used LLM to assist with polishing manuscript text and refining scientific figures. The authors take full responsibility for the claims, citations, figures, and experimental results in the final manuscript, including material prepared with AI assistance.

References

Peter Anderson, Basura Fernando, Mark Johnson, and Stephen Gould. SPICE: Semantic Propositional Image Caption Evaluation. In *Computer Vision – ECCV 2016*, pp. 382–398. Springer International Publishing, 2016. doi: 10.1007/978-3-319-46454-1_24.

Jacob Austin, Daniel D. Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. Structured Denoising Diffusion Models in Discrete State-Spaces. In *Advances in Neural Information Processing Systems*, volume 34, pp. 17981–17993. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/958c530554f78bcd8e97125b70e6973d-Paper.pdf.

Chaofan Gan, Zicheng Zhao, Yuanpeng Tu, Xi Chen, Ziran Qin, Tieyuan Chen, Mehrtash Harandi, and Weiyao Lin. Massive Activations are the Key to Local Detail Synthesis in Diffusion Transformers. In *International Conference on Learning Representations*, pp. 104855–104883, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/ab6e7ad2354f350b451b5a8e14d04f51-Paper-Conference.pdf.

Drew A. Hudson and Christopher D. Manning. GQA: A New Dataset for Real-World Visual Reasoning and Compositional Question Answering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6700–6709, 2019. URL https://openaccess.thecvf.com/content_CVPR_2019/html/Hudson_GQA_A_New_Dataset_for_Real-World_Visual_Reasoning_and_Compositional_CVPR_2019_paper.html.

Mingyu Jin, Kai Mei, Wujiang Xu, Mingjie Sun, Ruixiang Tang, Mengnan Du, Zirui Liu, and Yongfeng Zhang. Massive Values in Self-Attention Modules are the Key to Contextual Knowledge Understanding. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 28063–28096. PMLR, 2025. URL https://proceedings.mlr.press/v267/jin25f.html.

Seil Kang, Jinyeong Kim, Junhyeok Kim, and Seong Jae Hwang. See What You Are Told: Visual Attention Sink in Large Multimodal Models. In *International Conference on Learning Representations*, pp. 87676–87703, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/da8a39bc39ae1c89dd6ebb1e3bcbb3f3-Paper-Conference.pdf.

Keuntae Kim, Mingyu Kang, and Yong Suk Choi. Thinking Diffusion: Penalize and Guide Visual-Grounded Reasoning in Diffusion Multimodal Language Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5154–5164, 2026. URL https://openaccess.thecvf.com/content/CVPR2026/html/Kim_Thinking_Diffusion_Penalize_and_Guide_Visual-Grounded_Reasoning_in_Diffusion_Multimodal_CVPR_2026_paper.html.

Shufan Li, Konstantinos Kallidromitis, Hritik Bansal, Akash Gokul, Yusuke Kato, Kazuki Kozuka, Jason Kuen, Zhe Lin, Kai-Wei Chang, and Aditya Grover. LaViDa: A Large Diffusion Language Model for Multimodal Understanding. In *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 105101–105134. Curran Associates, Inc., 2025. doi: 10.52202/085713-3511.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common Objects in Context. In *Computer Vision – ECCV 2014*, pp. 740–755. Springer International Publishing, 2014. doi: 10.1007/978-3-319-10602-1_48.

Zechun Liu, Changsheng Zhao, Igor Fedorov, Bilge Soran, Dhruv Choudhary, Raghuraman Krishnamoorthi, Vikas Chandra, Yuandong Tian, and Tijmen Blankevoort. SpinQuant: LLM Quantization with Learned Rotations. In *International Conference on Learning Representations*, pp. 92009–92032, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/e5b1c0d4866f72393c522c8a00eed4eb-Paper-Conference.pdf.

Pan Lu, Swaroop Mishra, Tanglin Xia, Liang Qiu, Kai-Wei Chang, Song-Chun Zhu, Oyvind Tafjord, Peter Clark, and Ashwin Kalyan. Learn to Explain: Multimodal Reasoning via Thought Chains for Science Question Answering. In *Advances in Neural Information Processing Systems*, volume 35, pp. 2507–2521. Curran Associates, Inc., 2022. doi: 10.52202/068431-0182.

Ahmed Masry, Do Xuan Long, Jia Qing Tan, Shafiq Joty, and Enamul Hoque. ChartQA: A Benchmark for Question Answering about Charts with Visual and Logical Reasoning. In *Findings of the Association for Computational Linguistics: ACL 2022*, pp. 2263–2279. Association for Computational Linguistics, 2022. doi: 10.18653/v1/2022.findings-acl.177.

Shen Nie, Fengqi Zhu, Zebin You, Xiaolu Zhang, Jingyang Ou, Jun Hu, Jun Zhou, Yankai Lin, Ji-Rong Wen, and Chongxuan LI. Large Language Diffusion Models. In *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 50608–50646. Curran Associates, Inc., 2025. doi: 10.52202/085713-1689.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 311–318. Association for Computational Linguistics, 2002. doi: 10.3115/1073083.1073135.

Zihan Qiu, Zeyu Huang, Kaiyue Wen, Peng Jin, Bo Zheng, Yuxin Zhou, Haofeng Huang, Zekun Wang, Xiao Li, Huaqing Zhang, Yang Xu, Haoran Lian, Siqi Zhang, Rui Men, Jianwei Zhang, Ivan Titov, Dayiheng Liu, Jingren Zhou, and Junyang Lin. A Unified View of Attention and Residual Sinks: Outlier-Driven Rescaling is Essential for Transformer Training. arXiv preprint arXiv:2601.22966, 2026. URL https://arxiv.org/abs/2601.22966.

Youngji Roh, Hyunjin Cho, and Jaehyung Kim. Embracing Anisotropy: Turning Massive Activations into Interpretable Control Knobs for Large Language Models. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 29930–29956. Association for Computational Linguistics, 2026. doi: 10.18653/v1/2026.acl-long.1380.

Subham Sekhar Sahoo, Marianne Arriola, Yair Schiff, Aaron Gokaslan, Edgar Marroquin, Justin T Chiu, Alexander Rush, and Volodymyr Kuleshov. Simple and Effective Masked Diffusion Language Models. In *Advances in Neural Information Processing Systems*, volume 37, pp. 130136–130184. Curran Associates, Inc., 2024. doi: 10.52202/079017-4135.

Zeru Shi, Zhenting Wang, Fan Yang, Qifan Wang, and Ruixiang Tang. A Single Layer to Explain Them All: Understanding Massive Values in Large Language Models. In *Forty-third International Conference on Machine Learning*, 2026. URL https://icml.cc/virtual/2026/poster/61433.

Amanpreet Singh, Vivek Natarajan, Meet Shah, Yu Jiang, Xinlei Chen, Dhruv Batra, Devi Parikh, and Marcus Rohrbach. Towards VQA Models That Can Read. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8317–8326, 2019. URL https://openaccess.thecvf.com/content_CVPR_2019/html/Singh_Towards_VQA_Models_That_Can_Read_CVPR_2019_paper.html.

Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. RoFormer: Enhanced transformer with Rotary Position Embedding. *Neurocomputing*, 568:127063, 2024. doi: 10.1016/j.neucom.2023.127063.

Mingjie Sun, Xinlei Chen, J. Zico Kolter, and Zhuang Liu. Massive Activations in Large Language Models. In *Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=F7aAhfitX6.

Shangwen Sun, Alfredo Canziani, Yann LeCun, and Jiachen Zhu. The Spike, the Sparse and the Sink: Anatomy of Massive Activations and Attention Sinks. arXiv preprint arXiv:2603.05498,

2026. URL https://arxiv.org/abs/2603.05498.

Ramakrishna Vedantam, C. Lawrence Zitnick, and Devi Parikh. CIDEr: Consensus-Based Image Description Evaluation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4566–4575, 2015. doi: 10.1109/CVPR.2015.7299087.

Xu Wang, Bingqing Jiang, Yu Wan, Baosong Yang, Lingpeng Kong, and Difan Zou. DLM-Scope: Mechanistic Interpretability of Diffusion Language Models via Sparse Autoencoders. In *Forty-third International Conference on Machine Learning*, 2026. URL https://icml.cc/virtual/2026/poster/63199.

Ruibin Xiong, Yunchang Yang, Di He, Kai Zheng, Shuxin Zheng, Chen Xing, Huishuai Zhang, Yanyan Lan, Liwei Wang, and Tieyan Liu. On Layer Normalization in the Transformer Architecture. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 10524–10533. PMLR, 2020. URL https://proceedings.mlr.press/v119/xiong20b.html.

Ling Yang, Ye Tian, Bowen Li, Xinchen Zhang, Ke Shen, Yunhai Tong, and Mengdi Wang. MMaDA: Multimodal Large Diffusion Language Models. In *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 138867–138907. Curran Associates, Inc., 2025. doi: 10.52202/085713-4636.

Jiacheng Ye, Shansan Gong, Jiahui Gao, Junming Fan, Shuang Wu, Wei Bi, Haoli Bai, Lifeng Shang, and Lingpeng Kong. Dream-VL & Dream-VLA: Open Vision-Language and Vision-Language-Action Models with Diffusion Language Model Backbone. arXiv preprint arXiv:2512.22615, 2026. URL https://arxiv.org/abs/2512.22615.

Zebin You, Shen Nie, Xiaolu Zhang, Jun Zhou, Zhiwu Lu, Ji-Rong Wen, and Chongxuan Li. LLaDA-V: Large Language Diffusion Models with Visual Instruction Tuning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10093–10105, 2026. URL https://openaccess.thecvf.com/content/CVPR2026/html/You_LLaDA-V_Large_Language_Diffusion_Models_with_Visual_Instruction_Tuning_CVPR_2026_paper.html.

Biao Zhang and Rico Sennrich. Root Mean Square Layer Normalization. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/1e8a19426224ca89e83cef47f1e7f53b-Paper.pdf.

Fred Zhang and Neel Nanda. Towards Best Practices of Activation Patching in Language Models: Metrics and Methods. In *International Conference on Learning Representations*, pp. 1651–1678, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/06a52a54c8ee03cd86771136bc91eb1f-Paper-Conference.pdf.

Haomin Zuo, Yidi Li, Luoxiao Yang, and Xiaofeng Zhang. Diffusion-CAM: Faithful Visual Explanations for dMLLMs. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 12085–12101. Association for Computational Linguistics, 2026. doi: 10.18653/v1/2026.acl-long.553.

A Complete Regional Intervention Scores

Table 4 extends the Dream-VL results in Table 1 to all three backbones, retaining all seven task metrics. The scores are descriptive results on the fixed exploratory subsets; Random matches coordinate count, not intervention energy. SQA denotes ScienceQA-IMG accuracy, GQA uses raw exact-answer accuracy, TextVQA uses soft accuracy, and ChartQA uses relaxed accuracy. Captioning uses CIDEr (Vedantam et al., 2015), SPICE (Anderson et al., 2016), and BLEU-4 (B-4) (Papineni et al., 2002), all displayed on the $\times 100$ scale. Values below 0.01 are displayed as zero at table precision, without changing the underlying scores.

Table 4: **Regional peak-reset interventions at 32 denoising updates.** Each panel targets one token role, using 128 response positions and the same 48 inputs per dataset.

(a) Image

	Question answering				COCO		
Treatment	SQA	GQA	TVQA	CQA	CIDEr	SPICE	B-4
LaViDa							
Native	79.17	0.00	42.92	70.83	72.19	13.04	15.66
Zero	2.08	0.00	0.00	0.00	0.02	0.00	< 0.01
Mean	2.08	0.00	0.00	$0.00 < 0.01$		0.00	< 0.01
MinAbs	2.08	0.00	0.00	0.00	0.00	0.00	0.00
Random	79.17	0.00	40.83	70.83	70.30	13.07	15.57
Dream-VL							
Native	93.75	20.83	68.75	91.67	132.96	20.85	26.17
Zero	91.67	29.17	60.42	91.67	131.45	19.11	28.45
Mean	91.67	27.08	61.25	91.67	127.16	19.20	28.06
MinAbs	93.75	29.17	62.71	89.58	135.08	19.35	30.34
Random	93.75	18.75	65.63	91.67	133.09	21.35	29.98
MMaDA							
Native	64.58	0.00	2.08	12.50	106.05	17.86	24.57
Zero	0.00	0.00	0.00	0.00	0.06	$0.11 < 0.01$	
Mean	0.00	0.00	0.00	0.00	0.04	$0.00 < 0.01$	
MinAbs	0.00	0.00	0.00	0.00	0.02	$0.00 < 0.01$	
Random	64.58	0.00	2.08	14.58	104.65	17.70	26.43

(b) Question

	Question answering				COCO		
Treatment	SQA	GQA	TVQA	CQA	CIDEr	SPICE	B-4
LaViDa							
Native	79.17	0.00	42.92	70.83	72.19	13.04	15.66
Zero	0.00	0.00	0.00	0.00	0.00	$0.00 < 0.01$	
Mean	0.00	0.00	0.00	0.00	0.00	$0.00 < 0.01$	
MinAbs	0.00	0.00	0.00	0.00	0.00	$0.00 < 0.01$	
Random	81.25	0.00	42.92	68.75	71.04	13.05	15.34
Dream-VL							
Native	93.75	20.83	68.75	91.67	132.96	20.85	26.17
Zero	85.42	6.25	62.71	91.67	130.20	20.50	28.57
Mean	91.67	2.08	62.71	91.67	125.11	19.54	27.30
MinAbs	89.58	2.08	60.63	91.67	127.87	21.05	27.81
Random	93.75	18.75	63.33	91.67	134.69	21.73	27.67
MMaDA							
Native	64.58	0.00	2.08	12.50	106.05	17.86	24.57
Zero	2.08	0.00	0.00	0.00	40.80	16.08	12.19
Mean	0.00	0.00	0.00	0.00	40.87	15.34	12.57
MinAbs	0.00	0.00	0.00	0.00	42.80	16.10	12.11
Random	64.58	0.00	2.08	12.50	107.36	17.51	26.49

(c) Template/special

	Question answering				COCO		
Treatment	SQA	GQA	TVQA	CQA	CIDEr	SPICE	B-4
LaViDa							
Native	79.17	0.00	42.92	70.83	72.19	13.04	15.66
Zero	0.00	0.00	0.00	0.00	0.04	$0.00 < 0.01$	
Mean	0.00	0.00	0.00	0.00	0.04	$0.00 < 0.01$	
MinAbs	0.00	0.00	0.00	0.00	0.02	$0.00 < 0.01$	
Random	79.17	0.00	42.92	68.75	71.20	13.27	13.84
Dream-VL							
Native	93.75	20.83	68.75	91.67	132.96	20.85	26.17
Zero	0.00	0.00	0.00	0.00	0.00	$0.00 < 0.01$	
Mean	0.00	0.00	0.00	0.00	0.00	$0.00 < 0.01$	
MinAbs	0.00	0.00	0.00	0.00	0.00	$0.00 < 0.01$	
Random	93.75	20.83	68.75	91.67	131.78	20.77	27.03
MMaDA							
Native	64.58	0.00	2.08	12.50	106.05	17.86	24.57
Zero	0.00	0.00	0.00	0.00	0.11	$0.00 < 0.01$	
Mean	0.00	0.00	0.00	0.00	0.35	$0.00 < 0.01$	
MinAbs	0.00	0.00	0.00	0.00	0.12	$0.00 < 0.01$	
Random	64.58	0.00	2.08	12.50	103.61	17.70	24.33

(d) Response

	Question answering				COCO		
Treatment	SQA	GQA	TVQA	CQA	CIDEr	SPICE	B-4
LaViDa							
Native	79.17	0.00	42.92	70.83	72.19	13.04	15.66
Zero	4.17	0.00	2.08	0.00	0.00	0.00	0.00
Mean	4.17	0.00	2.08	0.00	0.00	$0.00 < 0.01$	
MinAbs	2.08	0.00	2.08	0.00	0.41	$0.13 < 0.01$	
Random	81.25	0.00	42.92	70.83	70.95	13.73	13.90
Dream-VL							
Native	93.75	20.83	68.75	91.67	132.96	20.85	26.17
Zero	89.58	20.83	28.13	41.67	47.53	14.57	9.22
Mean	89.58	22.92	30.42	45.83	51.16	16.35	10.78
MinAbs	89.58	20.83	26.88	39.58	53.62	16.00	10.26
Random	93.75	18.75	70.00	89.58	129.88	21.05	27.17
MMaDA							
Native	64.58	0.00	2.08	12.50	106.05	17.86	24.57
Zero	0.00	0.00	0.00	0.00	0.77	$0.00 < 0.01$	
Mean	0.00	0.00	0.00	0.00	1.05	$0.00 < 0.01$	
MinAbs	0.00	0.00	0.00	0.00	0.26	$0.00 < 0.01$	
Random	64.58	0.00	2.08	12.50	108.51	18.05	25.70

Metrics: SQA: ScienceQA-IMG accuracy; GQA: raw exact-answer accuracy; TVQA: TextVQA soft accuracy; CQA: ChartQA relaxed accuracy (all in %). COCO reports CIDEr, SPICE, and BLEU-4 (B-4), each displayed $\times 100$. *Controls:* Native is shared across regions. Random resets a non-peak coordinate and is not energy matched. GQA uses format-sensitive full-answer matching. Positive values below 0.01 are shown as < 0.01; 0.00 denotes exact zero.

B Local Norm Control Details

The controls in Section 4.2 use the same 48 fixed exploratory inputs per dataset, 32 denoising updates, and 128 response positions. Peak coordinates are reselected at each targeted decoder block output. Zero+Norm rescales the remaining coordinates, whereas Scale only rescales the entire token vector. Norms are computed from each condition's own current hidden state, so the controls do not hold norms equal across diverging trajectories. Original model caches, samplers, and precision are retained, and matching region names does not establish equal intervention exposure across models. Table 2 reports descriptive scores, not full benchmark results. The 2,304 added condition records and 1,440 reused, identity-validated records do not constitute independent replications.

Existing paired intervals use 2,000 bootstrap resamples of image clusters with seed 20260925. These are pointwise exploratory 95% percentile intervals, without multiplicity correction or an equivalence test. For Dream-VL response tokens, Zero+Norm relative to Zero changes ChartQA accuracy by $+14.58$ percentage points with interval $[0, 29.17]$, and CIDEr by -9.90 with interval $[-21.84, 1.71]$. Neither comparison establishes a reliable rescue across tasks, and the opposite point estimates are not a formal interaction test. For template tokens, Scale only relative to Zero changes ChartQA accuracy by $+91.67$ percentage points with interval $[83.33, 97.92]$, and CIDEr by $+128.21$ with interval $[107.50, 150.29]$. These contrasts do not identify semantic content in the deleted coordinate. Dream-VL template interventions with Zero and Zero+Norm produce empty visible answers on all 48 ChartQA inputs. On COCO, the corresponding empty outputs number 45 and 47 of 48. Scale only produces no empty outputs on either task. These diagnostics show that task-score collapse includes broad output failure and should not be read directly as selective loss of specific visual facts. MMaDA's Native ChartQA accuracy is 12.50% (6 of 48), limiting interpretation of differences near the performance floor.

C Regional Retention Details

The two retention tests in Section 4.3 use the same 48 fixed exploratory inputs per task, 32 denoising updates, and 128 response positions. Each token region is intervened on separately at the available decoder block output hooks. For the current vector h, Top retains the coordinates with the largest absolute values, preserving their signs and values. The retained count is either one or $k = \lceil pd \rceil$, where d is the hidden dimension and $p \in \{0.01, 0.05, 0.10, 0.20, 0.50\}$. The support is recomputed at each intervention, rather than fixed across inputs or updates. Random uses a permutation independent of the generation random state, with nested supports across retained counts and one random draw per site. It can include the peak and is not matched to Top for retained energy. Raw retention sets unselected coordinates to zero without rescaling. The norm-restored variant rescales the retained vector to its own current pre-edit norm, not to a separate Native trajectory. Other regions and subsequent computations remain available, so coordinates may be regenerated. The models retain their original caches, samplers, and precision, and the interventions do not establish equal exposure across backbones.

The completed matrix contains 27,648 added condition records, of which 26,499 generated successfully and 1,149 encountered numerical failures. These are repeated conditions on the fixed input sets, not independent samples. All 24 single-coordinate Random+Norm groups contain at least one input for which a zero retained vector cannot be rescaled to a nonzero norm. They are therefore reported as unscorable groups rather than scored on the surviving outputs. Table 5 omits these uniformly unavailable columns and rounds displayed scores below 0.01 to zero without modifying the stored scores. In particular, a displayed zero in CIDEr does not always denote an exact zero. The 100% endpoints in Figure 3 reuse Native scores and are not additional generations. The figure shows task-level point estimates with no smoothing or confidence bands. The accompanying experimental records retain 2,000 image-cluster paired bootstrap intervals for differences from Native, which must not be treated as confidence intervals for absolute scores. No new scoring or resampling was performed for this presentation.

The results support regional and task dependence, not a universal sparsity threshold. LaViDa template retention at 1% matches its Native ChartQA point estimate of 70.83%, whereas Dream-VL response retention remains at zero through 20%. Top also need not exceed Random at every fraction. For Dream-VL template tokens on COCO, retaining 10% without norm restoration yields 5.92

Table 5: **Single-coordinate retention in each token region.** Top retains one peak, not 1% of coordinates. Top+Norm restores the local norm. All Random+Norm groups are unscorable due to numerical failures, as detailed in Appendix C.

	ChartQA relaxed accuracy (%)				COCO CIDEr ($\times 100$)			
Region	Native	Top	Top+Norm	Random	Native	Top	Top+Norm	Random
LaViDa								
Image	70.83	4.17	10.42	0	72.19	0.11	2.88	0
Question	70.83	0	0	0	72.19	2.67	1.26	2.35
Template	70.83	0	0	10.42	72.19	29.57	25.50	12.21
Response	70.83	0	0	0	72.19	0	0	0
Dream-VL								
Image	91.67	0	0	0	132.96	1.68	2.22	0
Question	91.67	12.50	6.25	10.42	132.96	7.04	3.92	14.93
Template	91.67	0	0	0	132.96	26.17	44.89	31.46
Response	91.67	0	0	0	132.96	0	0	0
MMaDA								
Image	12.50	8.33	10.42	0	106.05	0.36	0.35	0.61
Question	12.50	0	0	0	106.05	62.76	65.31	21.62
Template	12.50	4.17	4.17	0	106.05	3.18	3.18	0.02
Response	12.50	0	0	0	106.05	0	0	0.02

CIDEr for Top and 37.85 for Random. Conversely, Top at 50% has higher point estimates than Random in all model, task, and region combinations in this matrix. These comparisons do not establish statistical equivalence, a precise transition between sampled fractions, or an equal-strength intervention advantage. MMaDA's Native ChartQA accuracy of 12.50% also limits interpretation near the performance floor. The retained supports concern conditional local computation, not a compressed model or measured inference acceleration.

D Peak-Amplitude Scaling Details

The relative-scaling experiment in Section 4.3 uses the same fixed exploratory sets of 48 ChartQA inputs and 48 COCO images. There are 32 denoising updates and 128 response positions. Each region is treated separately at the available decoder block output hooks, retaining each model's original sampler, cache, and precision. The peak is dynamically reselected from the current vector at every intervention, so neither the channel nor its pre-edit value is taken from an additional Native trajectory. Multiplication by a positive factor preserves the selected sign but changes both the vector norm and its direction relative to other coordinates. The figure uses the Raw condition, with no subsequent rescaling of the full vector. The tested positive multipliers are $0.25, 0.5, 2, 4$, with Native reused at 1. The earlier Zero intervention provides the $\alpha = 0$ comparison, which is not placed on the positive logarithmic axis. Norm-restored scaling is retained in the experimental records but is not mixed into this figure.

The full relative-scaling matrix comprises 9,216 added terminal records across Raw and norm-restored conditions. Of 192 condition groups, 176 are scorable and 16 are unavailable, all belonging to LaViDa or MMaDA at $\alpha = 4$ without norm restoration. All 48 inputs in each such group encounter numerical-validation failure, for 768 records in total. These failures primarily concern aggregate audits or requested-norm checks in FP32 and should not be described as directly observed nonfinite BF16 hidden states or logits. The unavailable conditions have no task score and are not plotted as zero. The main figure therefore contains 104 scored points and 16 unavailable model, task, region, and multiplier combinations, including reused Native points. Generation uses a fixed seed. The existing 2,000 image-cluster paired bootstrap intervals describe differences from Native, not confidence bands for absolute scores or variability across seeds. The figure connects observed task-level point estimates without smoothing or interpolation across unavailable conditions.

For Dream-VL on ChartQA, image attenuation by 0.25 and 0.5 yields 89.58% accuracy, compared with 91.67% for Native, while amplification by 2 and 4 yields zero. Question attenuation retains the Native point estimate, while amplification yields 6.25% and 4.17%. These contrasts do not establish statistical equivalence under attenuation or a continuous safe interval between sampled multipliers. On COCO, the Dream-VL template curve rises from 0.82 CIDEr at $\alpha = 2$ to 51.47 at $\alpha = 4$, still

below its Native score of 132.96. This nonmonotonic result is retained in the main figure rather than suppressed to imply steadily increasing damage. MMaDA's Native ChartQA accuracy of 12.50% limits interpretation near the floor. Different cache behavior and intervention exposure also prevent ranking intrinsic regional importance across models from loss magnitudes alone.

E Downstream Peak-Access Details

Intervention semantics. At each available decoder block output, the largest absolute coordinate is selected once for each targeted token. An unchanged copy and a copy with only that signed coordinate set to zero are then routed independently to the next QK, V, and Direct consumers. The three access bits are ordered QK/V/Direct. For example, 110 denies the peak to Direct while retaining it for QK and V, whereas 001 retains the peak only for Direct. Neither policy disables an entire module. The input embedding is unchanged. The final block has no successor QK/V edge and is handled through its applicable Direct/readout path. The original cache, RoPE, sampler, RNG, precision, and Dream-VL shifted readout are retained. Direct includes residual carry and the natural downstream FFN, rather than an isolated skip connection or FFN-only intervention. Each policy uses its own current state, not a separately replayed Native trajectory.

Coverage and reporting. Each model and task has a shared Native condition and seven non-Native policies for each of four regions. The resulting 174 unique groups contain 8,352 condition records across the three models and two tasks, with 48 inputs per group and no numerical failures. The four-region extension comprises 3,456 new generations and 4,896 strictly reused records. These are repeated conditions on exploratory subsets, not independent images or a full benchmark. All runs use 32 updates and 128 response positions. Table 3 reports ChartQA relaxed accuracy in percent and COCO CIDEr multiplied by 100. Values below 0.01 are displayed as zero at this precision, while the source data retain their original values. MMaDA's Native ChartQA score is only $6/48$, which limits interpretation near the floor.

Paired effects and limits. Table 6 contrasts removing Direct access with Native and retaining only Direct access with Zero. Intervals are the existing 95% pointwise image-cluster paired percentile bootstrap intervals with 2,000 draws and a fixed reference IDF for CIDEr. They are exploratory and not adjusted for multiple comparisons. No new scores or intervals are computed for this presentation. Similar point estimates do not establish equivalence, and the 101 results are not a dedicated QK-by-Direct interaction test. Low task scores also have different causes. Under template 110, COCO empty-output rates are 89.58%, 95.83%, and 0% for LaViDa, Dream-VL, and MMaDA, respectively, despite low CIDEr in all three. Task failure therefore cannot uniformly be identified with empty output or removal of specific visual facts.

Table 6: **Paired effects of Direct peak access.** Entries are differences [95% pointwise intervals]. ChartQA differences are percentage points and CIDEr differences use the $\times 100$ scale.

Model	Region	ChartQA Δ [95% CI]	CIDEr Δ [95% CI]
Remove Direct access: 110 − Native			
LaViDa	Template	−70.83 [−83.33, −58.33]	−72.11 [−85.22, −59.26]
LaViDa	Response	−68.75 [−81.25, −54.17]	−72.19 [−85.26, −59.36]
Dream-VL	Template	−85.42 [−95.83, −75.00]	−132.84 [−152.21, −114.12]
Dream-VL	Response	−54.17 [−68.75, −39.58]	−89.81 [−109.91, −68.96]
MMaDA	Template	−12.50 [−22.92, −4.17]	−105.71 [−126.50, −87.91]
MMaDA	Response	−12.50 [−22.92, −4.17]	−105.66 [−126.59, −87.69]
Retain Direct access only: 001 − 000			
LaViDa	Template	+0.00 [+0.00, +0.00]	−0.04 [−0.08, +0.00]
LaViDa	Response	+0.00 [+0.00, +0.00]	+0.54 [+0.00, +1.63]
Dream-VL	Template	+89.58 [+79.17, +97.92]	+127.53 [+106.35, +151.35]
Dream-VL	Response	+37.50 [+22.86, +54.17]	+72.58 [+55.96, +89.74]
MMaDA	Template	+0.00 [+0.00, +0.00]	−0.11 [−0.28, −0.01]
MMaDA	Response	+2.08 [+0.00, +6.25]	+27.44 [+19.39, +36.18]

F CARRY, FFN-READER, AND TERMINAL CONTROLS

Current-state interventions. E and G are complementary analyses of one experiment, not independent replications. Let h^+ be the current source-block output and h^- its counterpart with the token's signed largest-absolute coordinate j set to zero. QK and V receive the native-normalized current h^+. Given that condition's attention contribution a, define $z^+ = h^+ + a$, $z^- = h^- + a$, and $r(z) = \sqrt{\text{mean}(z^2) + \epsilon}$. The intended FFN-reader operators are

$$\begin{aligned} f_{\text{Base}} = f_{\text{R}} = \gamma \odot z^-/r(z^-), \quad f_{\text{G}} = \gamma \odot z^-/r(z^+), \\ f_{\text{P}} = \gamma \odot z^+/r(z^-), \quad f_{\text{PG}} = \gamma \odot z^+/r(z^+). \end{aligned} \tag{6}$$

Base, G, P, and PG use z^- in the FFN residual addition, whereas R uses z^+ with the Base reader. The coordinate is not reselected at the FFN, and h^+ belongs to the current trajectory, not an externally replayed Native run. Native dtype, cast and addition order, cache, RoPE, sampler, RNG, and Dream-VL shifted readout are preserved. Base/G/P/PG/R protect the terminal output, T deletes only at that output, and D110 retains its earlier terminal policy.

Coverage and metrics. The matrix retains three models, four regions, 48 inputs per exploratory task subset, 32 updates, and 128 response positions. Its 174 groups comprise 6,912 new generations and 1,440 identity-verified references without numerical failures, not 8,352 independent images. The figures show unnormalized COCO CIDEr multiplied by 100 and saved 95% pointwise percentile intervals from 20,000 image-cluster draws, seed 20260925, with fixed reference IDF. Labels below 0.01 display zero only at that precision. No scores or intervals are recomputed.

Primary paired comparisons. Table 7 gives the six prespecified COCO response comparisons with Bonferroni-adjusted 99.1667% intervals. These paired intervals, not absolute error-bar overlap, govern the primary inference. All G-versus-Base intervals include zero, which does not establish zero effects. MMaDA's G-versus-P interval excludes zero, but G's absolute 0.30 versus Native's 106.05 is not substantive task recovery.

Table 7: **Prespecified response-region reader comparisons.** CIDEr differences use the $\times 100$ scale. Intervals are adjusted for the six-comparison family, unlike the figures' pointwise whiskers.

Model	Contrast	CIDEr difference	Adjusted interval
LaViDa	G-Base	$+0.47$	$[-0.08, +1.93]$
LaViDa	G-P	$+0.50$	$[+0.00, +1.96]$
Dream-VL	G-Base	$+7.64$	$[-15.23, +34.98]$
Dream-VL	G-P	-5.71	$[-27.30, +16.84]$
MMaDA	G-Base	$+0.18$	$[-0.07, +0.46]$
MMaDA	G-P	$+0.30$	$[+0.12, +0.54]$

Exploratory contrasts and regional boundaries. For COCO template tokens, R yields 68.98, 133.79, and 104.91 CIDEr for LaViDa, Dream-VL, and MMaDA, respectively, while PG yields only 0.01, 0, and 0.04. The corresponding R-minus-PG pointwise 95% paired intervals are $68.97\,[55.97, 82.33]$, $133.79\,[112.43, 156.19]$, and $104.87\,[85.45, 125.56]$. These template contrasts distinguish carry from reader restoration without establishing Native equivalence. ChartQA template R scores are 56.25%, 91.67%, and 8.33%, compared with Native's 70.83%, 91.67%, and 12.50%. MMaDA's $6/48$ Native score limits inference near the floor. The exploratory MMaDA question G-minus-Base COCO difference is $63.54\,[46.40, 82.24]$. Conversely, LaViDa image R reaches only 3.30 versus Native's 72.19, and MMaDA response R reaches 32.01 versus 106.05.

Terminal-policy interpretation. Dream-VL response Base-minus-D110 is $44.79\,[27.44, 61.16]$, and T-minus-Native is $-130.80\,[-149.91, -112.25]$, using exploratory 95% paired intervals. These different trajectories do not yield additive mediation fractions. Prefix-region T may be structurally null when targeted rows do not intersect the actual readout. Native-equivalent behavior then does not demonstrate that an effective terminal intervention was harmless.

G Derivations and Evidence Boundaries for the Theoretical Analysis

All algebra below uses real arithmetic and fixed parameters. The identities describe conditional local computations, not training dynamics or exact BF16 equivalence. Experimental controls preserve the model's actual casts and update order and are interpreted using their recorded numerical audits.

G.1 Shared Writes, Coordinate Concentration, and RoPE

At a fixed step and token row, define the actual update $w_\ell = h_{\ell+1} - h_\ell$. Telescoping gives $h_L = h_0 + \sum_{\ell<L} w_\ell$, and projection onto any *fixed* u preserves the equality. The direction must be held fixed for this cross-layer statement. Our layer-specific C4 projections use different references and cannot be summed as a single cross-layer attribution. A large positive contribution can be followed by a negative one, so neither increasing peaks nor monotonic recovery follows from residual architecture.

Origin bound and its nontrivial conditions. Define $a_L = u^\top h_L$ and $e_L = (I - uu^\top)h_L$, for a fixed native-basis unit u. At $j^\star \in \arg\max_j |u_j|$, the reverse triangle inequality gives $\|h_L\|_\infty \geq |a_L||u_{j^\star}| - |e_{L,j^\star}| \geq |a_L|\|u\|_\infty - \|e_L\|_\infty$. This proves Equation 2. For a unique dominant channel $j^\star$, the stronger separation bound is

$$|h_{L,j^\star}| - \max_{c \neq j^\star} |h_{L,c}| \geq |a_L| \left(|u_{j^\star}| - \max_{c \neq j^\star} |u_c| \right) - 2\|e_L\|_\infty. \tag{7}$$

A large positive right-hand side is a sufficient condition for a separated peak. A negative bound is uninformative. Large amplitude alone is insufficient, as $u = (1, \ldots, 1)/\sqrt{d}$ has no isolated peak regardless of a_L. One large write followed by residual retention can suffice, without repeated same-sign accumulation. None of these conditions is asserted to hold at all layers or across every token role.

For a pre-normalized block at fixed denoising time,

$$U = H + \mathcal{A}(\mathcal{N}_a(H)), \qquad H_{\text{next}} = U + \mathcal{M}(\mathcal{N}_f(U)). \tag{8}$$

Branch normalization does not directly normalize the residual sum. For the immediate FFN map in Equation 3, positive-scale invariance of zero-epsilon RMSNorm gives $\mathcal{N}_0(ch) = \mathcal{N}_0(h)$, so subtracting the two outputs leaves exactly $(c-1)h$. This is an inference-time local identity, not a claim of a global training instability or a recurrent attractor.

For a group of first-layer positions with identical value vectors, the head-wise attention sum is $\sum_{j \in \mathcal{R}} A_{ij}^{(r)} v_j^{(r)} = s_{\mathcal{R}}^{(r)}(i) v_{\text{MASK}}^{(r)}$. Linearity of the output projection gives

$$o_{\mathcal{R}}(i) = \sum_{r=1}^{n_h} s_{\mathcal{R}}^{(r)}(i) W_O^{(r)} v_{\text{MASK}}^{(r)}, \qquad s_{\mathcal{R}}^{(r)}(i) = \sum_{j \in \mathcal{R}} A_{ij}^{(r)} \in [0, 1]. \tag{9}$$

Thus repeated masks supply shared directions without multiplying each head's value by the number of mask positions. Any output bias is included once, outside the group decomposition. The resulting matrix over queries lies in the span of the projected head directions, whose rank is at most the number of independent such directions. It is rank one only under additional alignment or coefficient constraints. The derivation requires identical values, evaluation-time dropout zero, and no value-side position-dependent transform. It applies to the inspected first-layer LaViDa MASK path, not automatically to Dream-VL layer 23 or to already contextualized response positions. Sharing across positions within one input and a common mean across different inputs are distinct properties. The source-group experiment tests the former, while the shared-background analysis measures the latter.

For a dominant component $h = au$, $\|u\|_2 = 1$, the equalities $\max_c |h_c| = |a|\|u\|_\infty$ and $\|h\|_2 = |a|$ separate magnitude from coordinate concentration. An orthogonal change of coordinates preserves the latter norm but can change the maximum. This is a statement about representation geometry, not a claim that rotating a hidden state while keeping all weights fixed preserves the model. It also does not assign semantic importance by amplitude.

For standard RoPE (Su et al., 2024), R_p is block diagonal with two-dimensional rotations, giving

$$R_p^\top R_p = I, \qquad \|(R_p q)^{(j)}\|_2 = \|q^{(j)}\|_2, \qquad \|R_p(q_k - q_{k'})\|_2 = \|q_k - q_{k'}\|_2. \tag{10}$$

The superscript j indexes a rotated pair. Rotation changes individual coordinates but neither creates pair energy nor transfers it between frequency pairs. At a fixed sequence position it also does not itself amplify differences between denoising steps. Moreover, $q = R_p W_Q \mathcal{N}(h)$ contains a learned projection, so a residual channel is not itself a RoPE frequency pair. This is the representation boundary between our study and the Q/K concentration reported by Jin et al. (2025).

G.2 Peak Scaling, Norm Restoration, and Information Loss

For $h = me_j + b$, $b_j = 0$, scaling the source coordinate gives

$$\mathcal{N}_\epsilon(h_\alpha) = \frac{\Gamma(\alpha m e_j + b)}{r_\alpha}, \qquad r_\alpha^2 = \frac{\alpha^2 m^2 + \|b\|_2^2}{d} + \epsilon. \tag{11}$$

Direct subtraction and differentiation give

$$\mathcal{N}(h_\alpha) - \mathcal{N}(h) = \Gamma\left[m\left(\frac{\alpha}{r_\alpha} - \frac{1}{r_1}\right)e_j + \left(\frac{1}{r_\alpha} - \frac{1}{r_1}\right)b\right], \tag{12}$$

$$\frac{\partial \mathcal{N}(h_\alpha)}{\partial \alpha} = \frac{\Gamma}{d r_\alpha^3}\left[m(\|b\|_2^2 + d\epsilon)e_j - \alpha m^2 b\right]. \tag{13}$$

All denominators must be nonzero. Deletion increases the gain on every remaining reader component by $r_1/r_0 = \sqrt{(m^2 + \|b\|_2^2 + d\epsilon)/(\|b\|_2^2 + d\epsilon)}$. Positive amplification eventually saturates the peak reader but diminishes the other coordinates. The learned diagonal Γ and subsequent projections still determine the resulting feature sensitivity. The equations therefore do not assert monotonic task damage.

For $b \neq 0$ and $c = \|h\|_2/\|b\|_2 > 0$,

$$\mathcal{N}_\epsilon(cb) = \frac{\Gamma b}{\sqrt{\|b\|_2^2/d + \epsilon/c^2}}, \qquad \mathcal{N}_0(cb) = \mathcal{N}_0(b). \tag{14}$$

For the immediate map $\mathcal{F}_0(z) = z + \mathcal{M}(\mathcal{N}_0(z))$, $\mathcal{F}_0(cb) - \mathcal{F}_0(b) = (c-1)b$. This is a conditional single-map identity, not an equivalence of full trajectories or of a block containing intervening attention additions. At nonzero epsilon the reader ratio differs by $\sqrt{(\|b\|_2^2/d + \epsilon)/(\|b\|_2^2/d + \epsilon/c^2)}$. BF16 rounding and casts add further deviations.

High retained energy is also insufficient as an information criterion. For a fixed coordinate projector $P_j = e_j e_j^\top$, let $\rho = m^2/\|h\|_2^2$. Then $\cos(h, P_j h) = \sqrt{\rho}$. Nevertheless, $h_+ = (A, 1)$ and $h_- = (A, -1)$, with $A > 1$, have opposite labels under a classifier using the second coordinate and become identical after retaining only the peak. Their retained energy tends to one as A grows. Norm restoration cannot reconstruct the removed sign. This is a counterexample to energy-based sufficiency, not a fitted account of our models' labels. Dynamic top-coordinate selection is a state-dependent, piecewise projection rather than a single global linear projector.

Finite angular change under deletion. For $H = \|h\|_2$, $B = \|b\|_2 > 0$, $h^\top b = B^2$, so $\cos\angle(h, b) = B/H = \sqrt{1-\rho}$. Norm restoration $\widetilde{h} = Hb/B$ remains on the same ray as b, hence leaves this angle unchanged. Since $r(\widetilde{h}) = r(h)$, direct subtraction gives Equation 4 for arbitrary nonnegative epsilon. This is an exact finite relation, not a first-order extrapolation to deletion. At zero epsilon, for any nonzero h, h', $\mathcal{N}_0(h') - \mathcal{N}_0(h) = \sqrt{d}\,\Gamma(\widehat{h}' - \widehat{h})$. Therefore direction determines the normalized reader before its learned consumer, but neither angle nor energy predicts the task score without that consumer.

Radial and transverse sensitivity. With $\widehat{h} = h/H$, $P_\parallel = \widehat{h}\widehat{h}^\top$, and $P_\perp = I - P_\parallel$, differentiation yields

$$J_{\mathcal{N}_\epsilon}(h) = \Gamma\left(\frac{I}{r} - \frac{hh^\top}{dr^3}\right) = \frac{\Gamma}{r}\left(P_\perp + \frac{\epsilon}{r^2}P_\parallel\right). \tag{15}$$

The last step uses $H^2/d = r^2 - \epsilon$. The chain rule for the local residual map yields Equation 5, with $J_{\mathcal{M}}$ evaluated at $\mathcal{N}_\epsilon(h)$. This derivative is exact wherever the FFN is differentiable. At zero epsilon the radial reader derivative is zero while the carry derivative is the identity. At nonzero epsilon the

radial factor is ϵ/r^2, but a large learned $J_{\mathcal{M}}\Gamma$ precludes an unconditional bound on the final FFN change. For coordinate deletion, the perturbation $-me_j$ generally has a transverse component of norm $|m|B/H$. Calling this deletion a purely radial rescaling is therefore incorrect. Large interventions are interpreted using the finite identities and actual experiments, not by extrapolating the derivative.

G.3 Carry, Reader, and Attention Depend on Different Operators

For a common local input, let $r_\pm = r(z^\pm)$, $\delta = z^+ - z^-$, and $\lambda = r_-/r_+$. Equation 6 gives

$$\begin{aligned} f_{\mathrm{G}} &= \lambda f_{\mathrm{Base}}, & f_{\mathrm{P}} &= f_{\mathrm{Base}} + \Gamma\delta/r_-, \quad f_{\mathrm{PG}} = \lambda f_{\mathrm{P}}, \\ f_{\mathrm{PG}} - f_{\mathrm{P}} - f_{\mathrm{G}} + f_{\mathrm{Base}} &= \Gamma\delta(1/r_+ - 1/r_-). \end{aligned} \tag{16}$$

Even the reader corrections are generally nonadditive. Also, $z^\pm = h^\pm + a_{\mathrm{att}}$, so deletion at the source does not imply $z_j^- = 0$ after attention. In fact, $r_+^2 - r_-^2 = (2\langle z^-, \delta\rangle + \|\delta\|_2^2)/d$ can have either sign. The simple peak-deletion gain inequality cannot be transferred to the FFN reader without checking this condition.

For the four conditional local outputs,

$$\begin{aligned} Y_{\mathrm{Base}} &= z^- + F(f_{\mathrm{Base}}), & Y_{\mathrm{R}} &= z^+ + F(f_{\mathrm{Base}}), \\ Y_{\mathrm{PG}} &= z^- + F(f_{\mathrm{PG}}), & Y_+ &= z^+ + F(f_{\mathrm{PG}}), \\ Y_+ - Y_{\mathrm{Base}} &= (Y_{\mathrm{R}} - Y_{\mathrm{Base}}) + (Y_{\mathrm{PG}} - Y_{\mathrm{Base}}). \end{aligned} \tag{17}$$

Y_+ means jointly restored at this local state, not the output of an independent Native trajectory. The identity is conditional on a shared attention contribution and shared inputs. In particular, $Y_{\mathrm{R}} - Y_{\mathrm{Base}} = \delta$, whereas $Y_{\mathrm{PG}} - Y_{\mathrm{Base}} = F(\mathcal{N}(z^+)) - F(\mathcal{N}(z^-))$. These exact conditional contrasts connect the local geometry to the tested R and PG operators. Different intervention runs later develop different states, and neither accuracy nor CIDEr is linear in these outputs. Consequently the empirical R, PG, and G contrasts identify computational dependencies without yielding additive mediation fractions.

For attention, a first-order change has the form

$$\delta o_i = \sum_j A_{ij}\,\delta v_j + \sum_j A_{ij}(v_j - o_i)\,\delta s_{ij}, \qquad \delta s_{ij} = \frac{\delta q_i^\top k_j + q_i^\top \delta k_j}{\sqrt{d_h}}. \tag{18}$$

Here $o_i = \sum_j A_{ij} v_j$, Q/K include their position transforms, and sums include only allowed keys. Second-order terms are omitted, so this is a sensitivity relation rather than an exact formula for large deletions. The V path changes message content, while Q/K change its weights. For example, adding the same scalar to all allowed logits leaves their softmax unchanged, whereas adding it to only one token region generally changes relative attention mass. Thus raw peak magnitude alone does not determine downstream influence. Visibility, readout support, projection directions, and residual bypasses permit different regional effects without assigning an intrinsic universal function to a token-role label.

G.4 Conditional State Handoff and Decision Boundaries

Let $X^{(k)} = (Z^{(k)}, C^{(k)})$ include the complete persistent state, including any caches, retained embeddings, and scheduler bookkeeping. The forward and sampling update is

$$X^{(k+1)} = \mathcal{T}_{\theta,k}(I, Q, X^{(k)}; \xi_k). \tag{19}$$

Conditional state-handoff proposition. Consider two runs with the same image, question, frozen parameters, remaining intervention policy, and step index. Assume deterministic execution conditional on the supplied random variates. If their complete persistent states $X^{(k+1)}$ agree and every subsequent random variate agrees, then every later state and final output agrees.

Proof. Apply the same map $\mathcal{T}_{\theta,k+1}$ to equal inputs and randomness. The resulting states at $k+2$ are equal. Induction establishes equality at all later boundaries, and a common final readout gives

equal output. The same argument applies to a one-time intervention that has ended, as used in our pulse controls. It does not permit one run to retain a changed prefix cache or continue receiving an intervention absent from the other run. Equality of Z alone suffices only if the other persistent state agrees or is a fixed common function of (I, Q, k, Z). The state includes any persistent embeddings, retained confidences, scheduler history, and stopping bookkeeping actually used by the implementation.

Token and position margins. For greedy token selection at a position, write the logit gap between the top two candidates as $g > 0$. If $\|\delta\ell\|_\infty < g/2$, the preferred token cannot change, because the gap can shrink by at most $2\|\delta\ell\|_\infty$. This alone does not guarantee the same denoising update. If a scheduler commits the B most confident positions, a sufficient additional condition is that all confidence changes have magnitude below half the positive gap between ranks B and $B + 1$. Any thresholds, eligibility rules, and tie handling must also remain unchanged. Under these conditions the same positions receive the same tokens, so $Z^{(k+1)}$ agrees. For stochastic sampling the realized decision boundaries depend on coupled random variates, rather than solely on a greedy logit gap.

A visible-answer extractor that truncates at the first stop token can change discontinuously even when the total number of network evaluations is fixed. A stop committed at position zero can therefore erase the scored visible prefix without demonstrating loss of every encoded visual fact or reduced compute. At fixed positions RoPE does not supply an additional denoising-time state. Cache reuse can also explain an unchanged prefix peak across steps, so cached values must be distinguished from freshly computed response activations.

G.5 Supporting Archival Evidence and Protocol Separation

The following completed analyses support limited parts of the theory and are not additional runs of the three-backbone, 48-input task matrices in Section 4. They use LaViDa and Dream-VL only and retain their original exploratory scope.

Formation evidence. The archived geometry analysis of native source states found that an across-input mean accounted for 87.948–96.838% of raw squared energy across the two models and two modalities. These means were estimated on the same reused development inputs and are descriptive, not held-out reconstruction scores. After centering, the leading singular direction accounted for 23.972–73.645% of the remaining energy. The denominators differ, so these ranges do not partition semantic information. The archived Why analysis decomposed actual residual writes and evaluated source-group perturbations before the first response update, without generating final answers. Its first-layer accounting found small incoming projections of approximately 0.12–0.13, compared with attention-write projections of 86.61/92.16 for LaViDa and 85.47/85.88 for Dream-VL on CLEVR/MMVP, respectively. These are percentages of layer-specific calibration-reference projections, not raw amplitudes or shares of semantics. The later source experiment used different fixed nodes, LaViDa layer 0 and Dream-VL layer 23, with 896 scientific pre-answer conditions and 28 technical checks per model. The signed prefix/MASK message projections were $0.671/0.326$ and $0.707/0.300$ for LaViDa, versus $-0.090/0.964$ and $-0.127/1.072$ for Dream-VL. Negative projections denote cancellation, not negative information. Late MASK values may already incorporate image and question context, so key position does not identify ultimate information origin. Eight Dream-VL prefix-random records exceeded the 1% write-matching tolerance and qualify those control comparisons. These observations support local shared writes without identifying their training cause or a common source across unequal depths.

Temporal evidence. The independent pulse protocol used eight reused open CLEVR-style development scenes, each expanded into 128 inputs per model with 64 image-conditioned and 64 text-conditioned cases. It used 32 response slots and 32 evaluations, rather than the 128-slot task matrices in Section 4. The frozen interventions targeted LaViDa layer 27 at its first response forward and Dream-VL layer 23 at its first effective response update. The prediction-support rows included Dream-VL's shifted predecessor. The source per-token top-10% set was sign-reversed once, followed by exchange of the complete 32-slot response state. This set contains hundreds of dimensions and is not synonymous with the largest four coordinates.

Table 8: **Archived response-state exchange.** Strict correct answers out of 64 per model and modality. Matched aggregate counts are accompanied by exact input-level final-token and suffix-state checks for all 512 exchange records.

Model	Modality	Native	Treated	Treated + clean Z	Native + treated Z
LaViDa	Image	46	4	46	4
	Text	59	1	59	1
Dream-VL	Image	60	13	60	13
	Text	64	7	64	7

The 1,024 scientific trajectories include 256 clean-state restorations and 256 treated-state transplants. All exchange records matched their intended final tokens and suffix states, with next-forward input, cache/embedding, RNG, hidden, and logit checks supporting the boundary control. LaViDa wrote a stop at response position zero in 57 of 64 treated image inputs, accounting for 57 of 58 empty visible outputs. For all 13 legal Dream-VL image-conditioned rubber-to-metal flips, the first native commitment was a stop at position five, replaced in the treated state by a space in 12 cases or comma in one case at position one. Whole-response-state exchange does not separate token identity from position or establish that whitespace encodes a material.

The accompanying next-layer recut experiment also did not establish useful semantic regeneration. Source deletion, adjacent-layer deletion, and their combination preserved material performance broadly. LaViDa text strict accuracy changed from 59/64 to 57/64 under the combination, but both changes were expansions retaining the correct material word. Its scene-level 95% interval included zero. Consequently reappearing energy alone is not evidence of necessary compensation. These archived results support a local hidden-to-decision-to-output route, not universal visual-fact transport, an autoregressive comparison, or a demonstrated model improvement.

H Complete Experimental Matrices and Supplemental Diagnostics

This appendix provides the complete scored matrices underlying the main-text study, including conditions and metrics omitted from the figures. The tables preserve the saved scores, rather than recomputing inference, evaluation, or uncertainty. Each task uses 48 fixed inputs shared across conditions. Repeated conditions are not independent samples. The regional-reset study includes 16 and 32 updates, while A, B, C1, C2, D, and E/G use 32 updates and 128 response positions. The archived single-pulse evidence in Appendix G.5 follows a separate protocol and is not pooled with these matrices.

Reading the tables. QA scores are percentages. COCO CIDEr, SPICE, and BLEU-4 are displayed on the $\times 100$ scale. Raw denotes the intervention without norm restoration, while Norm restores each edited vector's current pre-edit norm. NA denotes an unscorable condition and is never substituted with zero. Positive scores below 0.01 are displayed as 0.00 at table precision. Machine-readable files retain full precision and status fields. Native is repeated for readability, not counted as a new experiment.

Uncertainty and source data. The accompanying `source_data` directory contains every saved score and paired contrast from the regional-reset and A/B/C analyses, alongside the complete D and E/G records. Regional-reset, A/B/C, and D paired intervals use the saved 2,000-draw image-cluster percentile analysis. E/G uses 20,000 draws with fixed reference IDF for CIDEr. The six prespecified response-reader contrasts use the adjusted intervals in Table 7. Other reported intervals are exploratory and pointwise. Corpus BLEU-4 scores are retained without inventing unavailable uncertainty intervals. These files include the separately labeled GQA format companion. The primary tables retain raw exact-answer scoring.

H.1 Regional suppression at both denoising budgets

The following tables give all regional-reset conditions at 16 and 32 updates. QA and captioning are separated to keep the metrics legible. The treatment changes one selected coordinate per token in the chosen region, not the entire token representation.

Table 9: LaViDa: regional suppression, 16 updates.

Region	Treatment	SQA	GQA	TextVQA	ChartQA
All	Native	81.25	0.00	40.83	72.92
Image	Zero	2.08	0.00	0.00	0.00
Image	Mean	4.17	0.00	0.00	0.00
Image	MinAbs	2.08	0.00	0.00	0.00
Image	Random	81.25	0.00	42.92	72.92
Question	Zero	0.00	0.00	0.00	0.00
Question	Mean	0.00	0.00	0.00	0.00
Question	MinAbs	0.00	0.00	0.00	0.00
Question	Random	81.25	0.00	40.83	72.92
Template	Zero	0.00	0.00	0.00	0.00
Template	Mean	0.00	0.00	0.00	0.00
Template	MinAbs	0.00	0.00	0.00	0.00
Template	Random	79.17	0.00	42.92	68.75
Response	Zero	6.25	0.00	0.00	0.00
Response	Mean	4.17	2.08	0.00	0.00
Response	MinAbs	2.08	0.00	4.17	0.00
Response	Random	81.25	0.00	42.92	75.00

Table 10: LaViDa: regional suppression, 16 updates.

Region	Treatment	CIDEr	SPICE	B-4
All	Native	50.63	10.75	9.32
Image	Zero	0.00	0.00	0.00
Image	Mean	0.00	0.00	0.00
Image	MinAbs	0.00	0.00	0.00
Image	Random	50.29	10.98	8.35
Question	Zero	0.00	0.00	0.00
Question	Mean	0.00	0.00	0.00
Question	MinAbs	0.00	0.00	0.00
Question	Random	52.81	10.91	9.31
Template	Zero	0.02	0.00	0.00
Template	Mean	0.06	0.00	0.00
Template	MinAbs	0.31	0.20	0.00
Template	Random	49.81	11.06	9.70
Response	Zero	0.00	0.00	0.00
Response	Mean	1.22	0.22	0.00
Response	MinAbs	0.23	0.00	0.00
Response	Random	53.41	10.99	9.79

Table 11: LaViDa: regional suppression, 32 updates.

Region	Treatment	SQA	GQA	TextVQA	ChartQA
All	Native	79.17	0.00	42.92	70.83

Continued on next page

Continued: LaViDa: regional suppression, 32 updates.

Region	Treatment	SQA	GQA	TextVQA	ChartQA
Image	Zero	2.08	0.00	0.00	0.00
Image	Mean	2.08	0.00	0.00	0.00
Image	MinAbs	2.08	0.00	0.00	0.00
Image	Random	79.17	0.00	40.83	70.83
Question	Zero	0.00	0.00	0.00	0.00
Question	Mean	0.00	0.00	0.00	0.00
Question	MinAbs	0.00	0.00	0.00	0.00
Question	Random	81.25	0.00	42.92	68.75
Template	Zero	0.00	0.00	0.00	0.00
Template	Mean	0.00	0.00	0.00	0.00
Template	MinAbs	0.00	0.00	0.00	0.00
Template	Random	79.17	0.00	42.92	68.75
Response	Zero	4.17	0.00	2.08	0.00
Response	Mean	4.17	0.00	2.08	0.00
Response	MinAbs	2.08	0.00	2.08	0.00
Response	Random	81.25	0.00	42.92	70.83

Table 12: LaViDa: regional suppression, 32 updates.

Region	Treatment	CIDEr	SPICE	B-4
All	Native	72.19	13.04	15.66
Image	Zero	0.02	0.00	0.00
Image	Mean	0.00	0.00	0.00
Image	MinAbs	0.00	0.00	0.00
Image	Random	70.30	13.07	15.57
Question	Zero	0.00	0.00	0.00
Question	Mean	0.00	0.00	0.00
Question	MinAbs	0.00	0.00	0.00
Question	Random	71.04	13.05	15.34
Template	Zero	0.04	0.00	0.00
Template	Mean	0.04	0.00	0.00
Template	MinAbs	0.02	0.00	0.00
Template	Random	71.20	13.27	13.84
Response	Zero	0.00	0.00	0.00
Response	Mean	0.00	0.00	0.00
Response	MinAbs	0.41	0.13	0.00
Response	Random	70.95	13.73	13.90

Table 13: Dream-VL: regional suppression, 16 updates.

Region	Treatment	SQA	GQA	TextVQA	ChartQA
All	Native	93.75	20.83	58.12	91.67
Image	Zero	91.67	27.08	47.50	83.33
Image	Mean	91.67	29.17	51.88	79.17
Image	MinAbs	91.67	31.25	47.50	83.33
Image	Random	93.75	20.83	60.21	91.67
Question	Zero	87.50	6.25	64.17	91.67
Question	Mean	83.33	8.33	58.54	91.67
Question	MinAbs	85.42	4.17	60.00	91.67
Question	Random	89.58	20.83	56.04	91.67

Continued on next page

Continued: Dream-VL: regional suppression, 16 updates.

Region	Treatment	SQA	GQA	TextVQA	ChartQA
Template	Zero	0.00	0.00	0.00	0.00
Template	Mean	0.00	0.00	0.00	0.00
Template	MinAbs	0.00	0.00	0.00	0.00
Template	Random	93.75	18.75	58.12	91.67
Response	Zero	89.58	20.83	20.62	35.42
Response	Mean	93.75	22.92	30.21	47.92
Response	MinAbs	91.67	22.92	20.83	43.75
Response	Random	93.75	20.83	58.12	91.67

Table 14: Dream-VL: regional suppression, 16 updates.

Region	Treatment	CIDEr	SPICE	B-4
All	Native	110.22	16.85	21.27
Image	Zero	112.07	17.19	22.41
Image	Mean	108.39	16.52	24.01
Image	MinAbs	114.14	17.58	24.45
Image	Random	105.84	16.31	21.59
Question	Zero	109.33	16.90	22.38
Question	Mean	110.45	17.65	22.99
Question	MinAbs	112.23	17.19	22.38
Question	Random	102.00	16.07	19.22
Template	Zero	1.64	0.34	0.00
Template	Mean	3.14	0.71	0.00
Template	MinAbs	1.19	0.14	0.00
Template	Random	109.71	16.49	22.21
Response	Zero	45.80	13.30	7.50
Response	Mean	49.46	12.87	11.49
Response	MinAbs	45.22	12.35	7.07
Response	Random	107.49	16.57	21.65

Table 15: Dream-VL: regional suppression, 32 updates.

Region	Treatment	SQA	GQA	TextVQA	ChartQA
All	Native	93.75	20.83	68.75	91.67
Image	Zero	91.67	29.17	60.42	91.67
Image	Mean	91.67	27.08	61.25	91.67
Image	MinAbs	93.75	29.17	62.71	89.58
Image	Random	93.75	18.75	65.62	91.67
Question	Zero	85.42	6.25	62.71	91.67
Question	Mean	91.67	2.08	62.71	91.67
Question	MinAbs	89.58	2.08	60.62	91.67
Question	Random	93.75	18.75	63.33	91.67
Template	Zero	0.00	0.00	0.00	0.00
Template	Mean	0.00	0.00	0.00	0.00
Template	MinAbs	0.00	0.00	0.00	0.00
Template	Random	93.75	20.83	68.75	91.67
Response	Zero	89.58	20.83	28.12	41.67
Response	Mean	89.58	22.92	30.42	45.83
Response	MinAbs	89.58	20.83	26.88	39.58
Response	Random	93.75	18.75	70.00	89.58

Table 16: Dream-VL: regional suppression, 32 updates.

Region	Treatment	CIDEr	SPICE	B-4
All	Native	132.96	20.85	26.17
Image	Zero	131.45	19.11	28.45
Image	Mean	127.16	19.20	28.06
Image	MinAbs	135.08	19.35	30.34
Image	Random	133.09	21.35	29.98
Question	Zero	130.20	20.50	28.57
Question	Mean	125.11	19.54	27.30
Question	MinAbs	127.87	21.05	27.81
Question	Random	134.69	21.73	27.67
Template	Zero	0.00	0.00	0.00
Template	Mean	0.00	0.00	0.00
Template	MinAbs	0.00	0.00	0.00
Template	Random	131.78	20.77	27.03
Response	Zero	47.53	14.57	9.22
Response	Mean	51.16	16.35	10.78
Response	MinAbs	53.62	16.00	10.26
Response	Random	129.88	21.05	27.17

Table 17: MMaDA: regional suppression, 16 updates.

Region	Treatment	SQA	GQA	TextVQA	ChartQA
All	Native	64.58	0.00	2.08	12.50
Image	Zero	0.00	0.00	0.00	0.00
Image	Mean	0.00	0.00	0.00	0.00
Image	MinAbs	0.00	0.00	0.00	0.00
Image	Random	62.50	0.00	2.08	12.50
Question	Zero	0.00	0.00	0.00	0.00
Question	Mean	0.00	0.00	0.00	0.00
Question	MinAbs	2.08	0.00	0.00	0.00
Question	Random	64.58	0.00	0.00	12.50
Template	Zero	0.00	0.00	0.00	0.00
Template	Mean	0.00	0.00	0.00	0.00
Template	MinAbs	0.00	0.00	0.00	0.00
Template	Random	66.67	0.00	2.08	12.50
Response	Zero	2.08	0.00	0.00	0.00
Response	Mean	0.00	0.00	0.00	0.00
Response	MinAbs	0.00	0.00	0.00	0.00
Response	Random	64.58	0.00	4.17	12.50

Table 18: MMaDA: regional suppression, 16 updates.

Region	Treatment	CIDEr	SPICE	B-4
All	Native	96.08	15.20	21.94
Image	Zero	0.03	0.00	0.00
Image	Mean	0.06	0.00	0.00
Image	MinAbs	0.23	0.00	0.00
Image	Random	97.88	15.57	22.16
Question	Zero	45.05	15.78	10.56
Question	Mean	45.13	16.09	10.65

Continued on next page

Continued: MMaDA: regional suppression, 16 updates.

Region	Treatment	CIDEr	SPICE	B-4
Question	MinAbs	44.51	16.54	9.28
Question	Random	93.79	15.38	20.85
Template	Zero	0.12	0.00	0.00
Template	Mean	0.27	0.00	0.00
Template	MinAbs	0.27	0.00	0.00
Template	Random	94.11	15.53	21.92
Response	Zero	0.47	0.00	0.00
Response	Mean	0.48	0.00	0.00
Response	MinAbs	0.54	0.00	0.00
Response	Random	92.80	15.66	21.15

Table 19: MMaDA: regional suppression, 32 updates.

Region	Treatment	SQA	GQA	TextVQA	ChartQA
All	Native	64.58	0.00	2.08	12.50
Image	Zero	0.00	0.00	0.00	0.00
Image	Mean	0.00	0.00	0.00	0.00
Image	MinAbs	0.00	0.00	0.00	0.00
Image	Random	64.58	0.00	2.08	14.58
Question	Zero	2.08	0.00	0.00	0.00
Question	Mean	0.00	0.00	0.00	0.00
Question	MinAbs	0.00	0.00	0.00	0.00
Question	Random	64.58	0.00	2.08	12.50
Template	Zero	0.00	0.00	0.00	0.00
Template	Mean	0.00	0.00	0.00	0.00
Template	MinAbs	0.00	0.00	0.00	0.00
Template	Random	64.58	0.00	2.08	12.50
Response	Zero	0.00	0.00	0.00	0.00
Response	Mean	0.00	0.00	0.00	0.00
Response	MinAbs	0.00	0.00	0.00	0.00
Response	Random	64.58	0.00	2.08	12.50

Table 20: MMaDA: regional suppression, 32 updates.

Region	Treatment	CIDEr	SPICE	B-4
All	Native	106.05	17.86	24.57
Image	Zero	0.06	0.11	0.00
Image	Mean	0.04	0.00	0.00
Image	MinAbs	0.02	0.00	0.00
Image	Random	104.65	17.70	26.43
Question	Zero	40.80	16.08	12.19
Question	Mean	40.87	15.34	12.57
Question	MinAbs	42.80	16.10	12.11
Question	Random	107.36	17.51	26.49
Template	Zero	0.11	0.00	0.00
Template	Mean	0.35	0.00	0.00
Template	MinAbs	0.12	0.00	0.00
Template	Random	103.61	17.70	24.33
Response	Zero	0.77	0.00	0.00
Response	Mean	1.05	0.00	0.00

Continued on next page

Continued: MMaDA: regional suppression, 32 updates.

Region	Treatment	CIDEr	SPICE	B-4
Response	MinAbs	0.26	0.00	0.00
Response	Random	108.51	18.05	25.70

H.2 LOCAL NORM CONTROLS

All three models and four regions are retained. Zero+Norm removes the peak and restores the current vector norm. Scale only reproduces the contraction while retaining direction.

Table 21: LaViDa: complete A matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
All	Native	70.83	72.19	13.04	15.66
Image	Zero	0.00	0.02	0.00	0.00
Image	Zero+Norm	0.00	0.05	0.00	0.00
Image	Scale only	4.17	0.54	0.15	0.00
Question	Zero	0.00	0.00	0.00	0.00
Question	Zero+Norm	0.00	0.18	4.70	0.00
Question	Scale only	0.00	2.20	0.17	0.00
Template	Zero	0.00	0.04	0.00	0.00
Template	Zero+Norm	0.00	0.00	0.00	0.00
Template	Scale only	4.17	1.46	0.69	0.00
Response	Zero	0.00	0.00	0.00	0.00
Response	Zero+Norm	0.00	0.81	0.39	0.00
Response	Scale only	0.00	0.00	0.00	0.00

Table 22: Dream-VL: complete A matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
All	Native	91.67	132.96	20.85	26.17
Image	Zero	91.67	131.45	19.11	28.45
Image	Zero+Norm	89.58	125.27	19.33	25.00
Image	Scale only	87.50	130.19	20.49	28.08
Question	Zero	91.67	130.20	20.50	28.57
Question	Zero+Norm	91.67	130.75	21.68	27.29
Question	Scale only	91.67	130.07	20.51	26.61
Template	Zero	0.00	0.00	0.00	0.00
Template	Zero+Norm	0.00	0.00	0.00	0.00
Template	Scale only	91.67	128.21	19.32	28.36
Response	Zero	41.67	47.53	14.57	9.22
Response	Zero+Norm	56.25	37.63	14.91	7.97
Response	Scale only	81.25	127.99	22.58	29.08

Table 23: MMaDA: complete A matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
All	Native	12.50	106.05	17.86	24.57
Image	Zero	0.00	0.06	0.11	0.00
Image	Zero+Norm	0.00	0.00	0.00	0.00

Continued on next page

Continued: MMaDA: complete A matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
Image	Scale only	2.08	0.26	0.00	0.00
Question	Zero	0.00	40.80	16.08	12.19
Question	Zero+Norm	0.00	49.03	17.33	11.07
Question	Scale only	2.08	40.30	14.17	9.82
Template	Zero	0.00	0.11	0.00	0.00
Template	Zero+Norm	0.00	0.24	0.00	0.00
Template	Scale only	0.00	0.00	0.00	0.00
Response	Zero	0.00	0.77	0.00	0.00
Response	Zero+Norm	0.00	0.15	0.00	0.00
Response	Scale only	0.00	0.74	0.00	0.00

H.3 Complete sparse-retention matrix

The $k = 1$ and percentage conditions test retained support at each target node. Top and Random select equal numbers of coordinates. Raw and Norm remain separate. All Random $k = 1$ Norm groups are unscorable, and partially successful conditions are not scored using surviving inputs alone.

Table 24: LaViDa: complete B matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
All	Native	70.83	72.19	13.04	15.66
Image	Top $k = 1$, Raw	4.17	0.11	0.00	0.00
Image	Top $k = 1$, Norm	10.42	2.88	0.37	0.00
Image	Random $k = 1$, Raw	0.00	0.00	0.00	0.00
Image	Random $k = 1$, Norm	NA	NA	NA	NA
Image	Top 1%, Raw	12.50	2.77	1.10	0.00
Image	Top 1%, Norm	10.42	2.81	1.32	0.00
Image	Random 1%, Raw	0.00	0.00	0.00	0.00
Image	Random 1%, Norm	0.00	0.00	0.00	0.00
Image	Top 5%, Raw	10.42	4.03	2.19	0.00
Image	Top 5%, Norm	10.42	4.82	2.47	0.00
Image	Random 5%, Raw	0.00	0.11	0.00	0.00
Image	Random 5%, Norm	0.00	0.00	0.00	0.00
Image	Top 10%, Raw	12.50	8.99	2.97	3.92
Image	Top 10%, Norm	12.50	14.37	3.80	3.74
Image	Random 10%, Raw	0.00	0.00	0.08	0.00
Image	Random 10%, Norm	4.17	0.13	0.00	0.00
Image	Top 20%, Raw	10.42	53.70	9.40	9.96
Image	Top 20%, Norm	8.33	50.88	8.75	9.42
Image	Random 20%, Raw	0.00	0.40	0.16	0.00
Image	Random 20%, Norm	6.25	0.07	0.00	0.00
Image	Top 50%, Raw	66.67	76.30	13.04	13.24
Image	Top 50%, Norm	66.67	76.88	13.58	13.95
Image	Random 50%, Raw	0.00	0.02	0.00	0.00
Image	Random 50%, Norm	6.25	1.01	0.36	0.00
Question	Top $k = 1$, Raw	0.00	2.67	1.05	0.00
Question	Top $k = 1$, Norm	0.00	1.26	0.42	0.00
Question	Random $k = 1$, Raw	0.00	2.35	0.52	0.00
Question	Random $k = 1$, Norm	NA	NA	NA	NA
Question	Top 1%, Raw	2.08	1.56	0.42	0.00
Question	Top 1%, Norm	2.08	1.84	0.75	0.00
Question	Random 1%, Raw	0.00	1.80	0.17	0.00
Question	Random 1%, Norm	2.08	2.50	0.51	0.00

Continued on next page

Continued: LaViDa: complete B matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
Question	Top 5%, Raw	2.08	3.22	0.75	0.00
Question	Top 5%, Norm	2.08	2.44	0.61	0.00
Question	Random 5%, Raw	0.00	0.36	0.14	0.00
Question	Random 5%, Norm	0.00	1.86	0.52	0.00
Question	Top 10%, Raw	2.08	2.37	0.59	0.00
Question	Top 10%, Norm	4.17	3.00	0.94	0.00
Question	Random 10%, Raw	0.00	0.02	0.00	0.00
Question	Random 10%, Norm	2.08	1.82	0.38	0.00
Question	Top 20%, Raw	8.33	2.41	0.42	0.00
Question	Top 20%, Norm	6.25	1.57	0.42	0.00
Question	Random 20%, Raw	0.00	0.45	0.49	0.00
Question	Random 20%, Norm	2.08	1.44	0.14	0.00
Question	Top 50%, Raw	68.75	78.26	14.53	15.44
Question	Top 50%, Norm	75.00	77.32	14.52	15.49
Question	Random 50%, Raw	0.00	0.61	0.00	0.00
Question	Random 50%, Norm	0.00	2.44	0.50	0.00
Template	Top $k = 1$, Raw	0.00	29.57	5.42	3.11
Template	Top $k = 1$, Norm	0.00	25.50	4.01	2.35
Template	Random $k = 1$, Raw	10.42	12.21	3.51	0.00
Template	Random $k = 1$, Norm	NA	NA	NA	NA
Template	Top 1%, Raw	70.83	82.72	12.81	16.35
Template	Top 1%, Norm	75.00	71.34	12.21	13.73
Template	Random 1%, Raw	0.00	0.00	0.00	0.00
Template	Random 1%, Norm	6.25	3.56	0.94	0.00
Template	Top 5%, Raw	70.83	75.05	12.27	13.91
Template	Top 5%, Norm	75.00	72.03	11.48	12.59
Template	Random 5%, Raw	0.00	0.00	0.00	0.00
Template	Random 5%, Norm	6.25	5.74	2.11	0.00
Template	Top 10%, Raw	72.92	74.20	12.21	13.45
Template	Top 10%, Norm	70.83	73.47	12.37	13.54
Template	Random 10%, Raw	0.00	0.27	0.14	0.00
Template	Random 10%, Norm	2.08	3.59	1.05	0.00
Template	Top 20%, Raw	68.75	74.52	13.26	14.27
Template	Top 20%, Norm	68.75	77.05	12.67	14.82
Template	Random 20%, Raw	0.00	0.95	0.53	0.00
Template	Random 20%, Norm	6.25	7.38	1.08	0.00
Template	Top 50%, Raw	66.67	74.29	13.60	16.08
Template	Top 50%, Norm	68.75	75.70	13.55	16.19
Template	Random 50%, Raw	0.00	0.63	0.36	0.00
Template	Random 50%, Norm	2.08	2.05	0.63	0.00
Response	Top $k = 1$, Raw	0.00	0.00	0.00	0.00
Response	Top $k = 1$, Norm	0.00	0.00	0.00	0.00
Response	Random $k = 1$, Raw	0.00	0.00	0.00	0.00
Response	Random $k = 1$, Norm	NA	NA	NA	NA
Response	Top 1%, Raw	0.00	0.00	0.00	0.00
Response	Top 1%, Norm	0.00	0.00	0.00	0.00
Response	Random 1%, Raw	0.00	0.00	0.05	0.00
Response	Random 1%, Norm	0.00	0.00	0.00	0.00
Response	Top 5%, Raw	0.00	0.00	0.00	0.00
Response	Top 5%, Norm	0.00	0.00	0.00	0.00
Response	Random 5%, Raw	0.00	0.01	0.00	0.00
Response	Random 5%, Norm	2.08	0.20	0.00	0.00
Response	Top 10%, Raw	0.00	0.00	0.00	0.00
Response	Top 10%, Norm	0.00	0.03	0.00	0.00
Response	Random 10%, Raw	0.00	0.00	0.00	0.00
Response	Random 10%, Norm	0.00	0.26	0.00	0.00
Response	Top 20%, Raw	27.08	0.02	0.00	0.00

Continued on next page

Continued: LaViDa: complete B matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
Response	Top 20%, Norm	20.83	0.05	0.00	0.00
Response	Random 20%, Raw	0.00	0.07	0.00	0.00
Response	Random 20%, Norm	0.00	0.03	0.00	0.00
Response	Top 50%, Raw	70.83	78.91	12.52	16.10
Response	Top 50%, Norm	68.75	82.59	11.94	15.27
Response	Random 50%, Raw	0.00	1.22	0.22	0.00
Response	Random 50%, Norm	0.00	0.00	0.00	0.00

Table 25: Dream-VL: complete B matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
All	Native	91.67	132.96	20.85	26.17
Image	Top $k = 1$, Raw	0.00	1.68	0.53	0.60
Image	Top $k = 1$, Norm	0.00	2.22	1.66	0.00
Image	Random $k = 1$, Raw	0.00	0.00	0.23	0.00
Image	Random $k = 1$, Norm	NA	NA	NA	NA
Image	Top 1%, Raw	10.42	7.77	2.86	0.00
Image	Top 1%, Norm	14.58	12.80	4.47	4.86
Image	Random 1%, Raw	0.00	0.00	0.34	0.00
Image	Random 1%, Norm	0.00	0.07	1.08	0.00
Image	Top 5%, Raw	12.50	18.47	4.47	9.07
Image	Top 5%, Norm	10.42	54.29	9.47	9.75
Image	Random 5%, Raw	2.08	0.00	0.00	0.00
Image	Random 5%, Norm	0.00	0.30	0.68	0.00
Image	Top 10%, Raw	12.50	48.07	9.46	11.69
Image	Top 10%, Norm	27.08	73.56	13.68	11.04
Image	Random 10%, Raw	0.00	0.14	0.15	0.00
Image	Random 10%, Norm	0.00	0.61	0.79	0.00
Image	Top 20%, Raw	27.08	105.40	16.86	20.71
Image	Top 20%, Norm	47.92	105.20	17.70	23.84
Image	Random 20%, Raw	2.08	2.02	0.86	0.00
Image	Random 20%, Norm	0.00	1.74	0.39	0.00
Image	Top 50%, Raw	89.58	127.72	19.36	28.46
Image	Top 50%, Norm	91.67	124.93	20.05	23.77
Image	Random 50%, Raw	6.25	3.19	1.41	0.00
Image	Random 50%, Norm	10.42	4.87	1.13	1.85
Question	Top $k = 1$, Raw	12.50	7.04	2.02	0.00
Question	Top $k = 1$, Norm	6.25	3.92	6.82	2.13
Question	Random $k = 1$, Raw	10.42	14.93	4.34	4.03
Question	Random $k = 1$, Norm	NA	NA	NA	NA
Question	Top 1%, Raw	18.75	7.43	1.64	0.00
Question	Top 1%, Norm	20.83	8.08	1.83	0.00
Question	Random 1%, Raw	8.33	11.09	2.96	0.00
Question	Random 1%, Norm	14.58	14.84	3.63	0.00
Question	Top 5%, Raw	10.42	5.44	1.15	0.00
Question	Top 5%, Norm	12.50	4.95	0.92	0.00
Question	Random 5%, Raw	16.67	9.96	3.23	0.00
Question	Random 5%, Norm	12.50	11.11	3.48	0.00
Question	Top 10%, Raw	22.92	4.88	0.97	0.00
Question	Top 10%, Norm	14.58	8.52	1.45	0.07
Question	Random 10%, Raw	22.92	7.93	2.21	0.00
Question	Random 10%, Norm	10.42	12.68	3.82	0.00
Question	Top 20%, Raw	79.17	14.68	2.22	0.44
Question	Top 20%, Norm	70.83	10.95	2.49	0.00

Continued on next page

Continued: Dream-VL: complete B matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
Question	Random 20%, Raw	20.83	6.97	2.78	0.00
Question	Random 20%, Norm	16.67	19.87	4.93	0.00
Question	Top 50%, Raw	91.67	131.45	21.26	28.11
Question	Top 50%, Norm	91.67	128.56	19.98	27.16
Question	Random 50%, Raw	33.33	11.88	3.48	0.00
Question	Random 50%, Norm	16.67	19.23	4.71	0.00
Template	Top $k = 1$, Raw	0.00	26.17	5.24	4.73
Template	Top $k = 1$, Norm	0.00	44.89	9.11	14.85
Template	Random $k = 1$, Raw	0.00	31.46	7.70	7.06
Template	Random $k = 1$, Norm	NA	NA	NA	NA
Template	Top 1%, Raw	4.17	19.99	3.96	0.33
Template	Top 1%, Norm	2.08	26.44	5.87	1.24
Template	Random 1%, Raw	0.00	46.83	8.68	12.06
Template	Random 1%, Norm	0.00	56.65	10.12	14.53
Template	Top 5%, Raw	0.00	13.61	3.38	0.26
Template	Top 5%, Norm	0.00	6.08	0.98	0.00
Template	Random 5%, Raw	4.17	28.36	5.90	5.35
Template	Random 5%, Norm	0.00	50.41	9.06	11.82
Template	Top 10%, Raw	0.00	5.92	1.66	0.00
Template	Top 10%, Norm	6.25	6.77	1.71	0.00
Template	Random 10%, Raw	0.00	37.85	6.60	5.94
Template	Random 10%, Norm	0.00	62.65	11.40	14.70
Template	Top 20%, Raw	47.92	92.31	14.56	15.84
Template	Top 20%, Norm	58.33	105.42	16.54	21.11
Template	Random 20%, Raw	0.00	56.66	10.44	11.44
Template	Random 20%, Norm	0.00	54.18	10.70	14.13
Template	Top 50%, Raw	91.67	133.44	20.94	28.40
Template	Top 50%, Norm	91.67	138.37	20.82	32.18
Template	Random 50%, Raw	4.17	49.22	8.87	9.08
Template	Random 50%, Norm	8.33	47.68	7.88	8.10
Response	Top $k = 1$, Raw	0.00	0.00	0.00	0.00
Response	Top $k = 1$, Norm	0.00	0.00	0.00	0.00
Response	Random $k = 1$, Raw	0.00	0.00	0.07	0.00
Response	Random $k = 1$, Norm	NA	NA	NA	NA
Response	Top 1%, Raw	0.00	1.02	0.00	0.00
Response	Top 1%, Norm	0.00	0.18	0.00	0.00
Response	Random 1%, Raw	0.00	0.00	0.15	0.00
Response	Random 1%, Norm	0.00	0.00	0.13	0.00
Response	Top 5%, Raw	0.00	0.57	0.17	0.00
Response	Top 5%, Norm	0.00	0.21	0.00	0.00
Response	Random 5%, Raw	0.00	0.00	0.14	0.00
Response	Random 5%, Norm	0.00	0.00	0.18	0.00
Response	Top 10%, Raw	0.00	3.64	1.65	0.00
Response	Top 10%, Norm	0.00	0.38	1.47	0.00
Response	Random 10%, Raw	0.00	0.00	0.07	0.00
Response	Random 10%, Norm	0.00	0.00	0.16	0.00
Response	Top 20%, Raw	0.00	5.98	8.48	0.97
Response	Top 20%, Norm	0.00	3.07	6.18	0.00
Response	Random 20%, Raw	0.00	0.02	0.11	0.00
Response	Random 20%, Norm	0.00	0.01	0.00	0.00
Response	Top 50%, Raw	85.42	128.19	20.84	28.42
Response	Top 50%, Norm	87.50	127.15	22.99	25.81
Response	Random 50%, Raw	0.00	0.35	0.00	0.00
Response	Random 50%, Norm	0.00	0.92	0.40	0.00

Table 26: MMaDA: complete B matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
All	Native	12.50	106.05	17.86	24.57
Image	Top $k = 1$, Raw	8.33	0.36	0.00	0.00
Image	Top $k = 1$, Norm	10.42	0.35	0.79	0.00
Image	Random $k = 1$, Raw	0.00	0.61	0.00	0.00
Image	Random $k = 1$, Norm	NA	NA	NA	NA
Image	Top 1%, Raw	8.33	8.05	2.91	0.00
Image	Top 1%, Norm	10.42	4.73	2.32	0.00
Image	Random 1%, Raw	0.00	0.08	0.00	0.00
Image	Random 1%, Norm	0.00	0.51	0.00	0.00
Image	Top 5%, Raw	8.33	7.79	2.84	0.00
Image	Top 5%, Norm	8.33	7.91	2.85	0.00
Image	Random 5%, Raw	0.00	0.00	0.00	0.00
Image	Random 5%, Norm	0.00	0.23	0.00	0.00
Image	Top 10%, Raw	8.33	7.70	2.85	0.00
Image	Top 10%, Norm	8.33	7.56	2.83	0.00
Image	Random 10%, Raw	2.08	0.00	0.00	0.00
Image	Random 10%, Norm	0.00	0.27	0.00	0.00
Image	Top 20%, Raw	10.42	9.59	2.28	4.51
Image	Top 20%, Norm	10.42	6.20	1.44	3.22
Image	Random 20%, Raw	0.00	0.00	0.00	0.00
Image	Random 20%, Norm	0.00	0.11	0.00	0.00
Image	Top 50%, Raw	10.42	113.54	19.37	28.11
Image	Top 50%, Norm	10.42	113.94	19.51	28.37
Image	Random 50%, Raw	0.00	0.00	0.00	0.00
Image	Random 50%, Norm	0.00	0.08	0.00	0.00
Question	Top $k = 1$, Raw	0.00	62.76	18.99	12.67
Question	Top $k = 1$, Norm	0.00	65.31	19.48	12.72
Question	Random $k = 1$, Raw	0.00	21.62	11.23	9.17
Question	Random $k = 1$, Norm	NA	NA	NA	NA
Question	Top 1%, Raw	0.00	62.07	19.57	13.45
Question	Top 1%, Norm	0.00	63.93	19.26	12.97
Question	Random 1%, Raw	0.00	13.49	3.07	0.00
Question	Random 1%, Norm	0.00	56.53	13.55	14.63
Question	Top 5%, Raw	0.00	57.87	19.57	13.30
Question	Top 5%, Norm	0.00	60.55	19.28	12.63
Question	Random 5%, Raw	0.00	29.28	8.18	11.27
Question	Random 5%, Norm	0.00	51.89	14.66	13.19
Question	Top 10%, Raw	2.08	59.07	18.46	13.30
Question	Top 10%, Norm	0.00	60.20	18.64	13.38
Question	Random 10%, Raw	0.00	43.41	12.01	10.77
Question	Random 10%, Norm	0.00	52.99	17.58	10.67
Question	Top 20%, Raw	2.08	9.71	1.76	0.49
Question	Top 20%, Norm	2.08	4.38	1.09	0.25
Question	Random 20%, Raw	0.00	58.16	13.91	14.11
Question	Random 20%, Norm	0.00	52.90	17.93	12.47
Question	Top 50%, Raw	10.42	114.17	18.45	26.93
Question	Top 50%, Norm	12.50	106.22	17.81	24.38
Question	Random 50%, Raw	0.00	53.77	15.49	13.85
Question	Random 50%, Norm	0.00	60.93	18.67	11.88
Template	Top $k = 1$, Raw	4.17	3.18	0.28	0.00
Template	Top $k = 1$, Norm	4.17	3.18	0.28	0.00
Template	Random $k = 1$, Raw	0.00	0.02	0.00	0.00
Template	Random $k = 1$, Norm	NA	NA	NA	NA
Template	Top 1%, Raw	6.25	3.87	0.78	0.00
Template	Top 1%, Norm	6.25	4.41	0.89	0.00

Continued on next page

Continued: MMaDA: complete B matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
Template	Random 1%, Raw	0.00	0.21	0.00	0.00
Template	Random 1%, Norm	0.00	0.00	0.00	0.00
Template	Top 5%, Raw	4.17	5.38	1.16	0.00
Template	Top 5%, Norm	4.17	5.55	0.98	0.00
Template	Random 5%, Raw	0.00	0.07	0.00	0.00
Template	Random 5%, Norm	2.08	0.18	0.20	0.00
Template	Top 10%, Raw	6.25	16.22	3.30	0.10
Template	Top 10%, Norm	6.25	13.91	2.38	0.00
Template	Random 10%, Raw	0.00	0.10	0.00	0.00
Template	Random 10%, Norm	0.00	0.16	0.29	0.00
Template	Top 20%, Raw	4.17	79.57	13.05	19.86
Template	Top 20%, Norm	4.17	85.47	14.26	22.53
Template	Random 20%, Raw	0.00	0.40	0.14	0.00
Template	Random 20%, Norm	0.00	1.59	0.67	0.00
Template	Top 50%, Raw	12.50	115.71	19.17	29.95
Template	Top 50%, Norm	12.50	107.38	18.26	26.35
Template	Random 50%, Raw	0.00	0.00	0.00	0.00
Template	Random 50%, Norm	0.00	0.30	0.18	0.00
Response	Top $k = 1$, Raw	0.00	0.00	0.00	0.00
Response	Top $k = 1$, Norm	0.00	0.00	0.00	0.00
Response	Random $k = 1$, Raw	0.00	0.02	0.10	0.00
Response	Random $k = 1$, Norm	NA	NA	NA	NA
Response	Top 1%, Raw	0.00	0.31	0.00	0.00
Response	Top 1%, Norm	0.00	0.08	0.00	0.00
Response	Random 1%, Raw	0.00	0.00	0.38	0.00
Response	Random 1%, Norm	0.00	0.03	0.29	0.00
Response	Top 5%, Raw	0.00	1.27	0.00	0.00
Response	Top 5%, Norm	0.00	0.93	0.00	0.00
Response	Random 5%, Raw	0.00	0.15	0.30	0.00
Response	Random 5%, Norm	0.00	0.09	0.12	0.00
Response	Top 10%, Raw	0.00	0.30	0.00	0.00
Response	Top 10%, Norm	0.00	0.96	0.17	0.00
Response	Random 10%, Raw	0.00	0.29	0.13	0.00
Response	Random 10%, Norm	0.00	0.26	0.33	0.00
Response	Top 20%, Raw	8.33	32.09	9.66	7.29
Response	Top 20%, Norm	6.25	40.48	10.08	7.42
Response	Random 20%, Raw	0.00	0.17	0.00	0.00
Response	Random 20%, Norm	0.00	0.43	0.00	0.00
Response	Top 50%, Raw	10.42	118.37	19.56	29.19
Response	Top 50%, Norm	12.50	122.72	21.35	27.88
Response	Random 50%, Raw	0.00	0.28	0.17	0.00
Response	Random 50%, Norm	0.00	0.77	0.17	0.00

H.4 PEAK-AMPLITUDE SCALING WITH AND WITHOUT NORM RESTORATION

Only the current peak is scaled. The full matrix includes Raw and Norm versions for all four multipliers. The LaViDa and MMaDA Raw $\times 4$ conditions remain NA because their required numerical audit did not complete successfully. This status is not a zero task score or a general assertion of nonfinite hidden states.

Table 27: LaViDa: complete C1 matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
All	Native	70.83	72.19	13.04	15.66
Image	$\times 0.25$, Raw	0.00	0.09	0.00	0.00

Continued on next page

Continued: LaViDa: complete C1 matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
Image	$\times 0.25$, Norm	0.00	0.07	0.00	0.00
Image	$\times 0.5$, Raw	0.00	0.27	0.00	0.00
Image	$\times 0.5$, Norm	0.00	0.08	0.00	0.00
Image	$\times 2$, Raw	10.42	2.59	0.53	0.00
Image	$\times 2$, Norm	10.42	2.18	0.15	0.00
Image	$\times 4$, Raw	NA	NA	NA	NA
Image	$\times 4$, Norm	12.50	2.05	0.33	0.00
Question	$\times 0.25$, Raw	0.00	0.00	0.00	0.00
Question	$\times 0.25$, Norm	0.00	0.04	0.00	0.00
Question	$\times 0.5$, Raw	0.00	1.43	0.36	0.00
Question	$\times 0.5$, Norm	0.00	10.06	1.06	0.14
Question	$\times 2$, Raw	0.00	2.64	1.11	0.00
Question	$\times 2$, Norm	4.17	1.23	0.17	0.00
Question	$\times 4$, Raw	NA	NA	NA	NA
Question	$\times 4$, Norm	2.08	2.29	0.54	0.00
Template	$\times 0.25$, Raw	0.00	0.05	0.00	0.00
Template	$\times 0.25$, Norm	0.00	0.00	0.00	0.00
Template	$\times 0.5$, Raw	0.00	1.03	0.35	0.00
Template	$\times 0.5$, Norm	0.00	0.00	0.00	0.00
Template	$\times 2$, Raw	18.75	59.02	11.21	8.84
Template	$\times 2$, Norm	6.25	42.73	6.82	9.13
Template	$\times 4$, Raw	NA	NA	NA	NA
Template	$\times 4$, Norm	0.00	29.79	5.32	4.61
Response	$\times 0.25$, Raw	0.00	0.43	0.20	0.00
Response	$\times 0.25$, Norm	0.00	3.00	1.35	0.00
Response	$\times 0.5$, Raw	0.00	2.14	0.20	0.00
Response	$\times 0.5$, Norm	0.00	2.83	2.58	0.00
Response	$\times 2$, Raw	0.00	0.00	0.00	0.00
Response	$\times 2$, Norm	0.00	0.00	0.00	0.00
Response	$\times 4$, Raw	NA	NA	NA	NA
Response	$\times 4$, Norm	0.00	0.00	0.00	0.00

Table 28: Dream-VL: complete C1 matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
All	Native	91.67	132.96	20.85	26.17
Image	$\times 0.25$, Raw	89.58	124.65	20.13	25.37
Image	$\times 0.25$, Norm	89.58	125.77	20.03	27.56
Image	$\times 0.5$, Raw	89.58	125.83	19.73	27.13
Image	$\times 0.5$, Norm	89.58	139.38	20.38	31.03
Image	$\times 2$, Raw	0.00	3.17	1.45	1.02
Image	$\times 2$, Norm	6.25	4.99	1.41	0.00
Image	$\times 4$, Raw	0.00	4.19	1.40	1.11
Image	$\times 4$, Norm	0.00	2.35	1.61	0.00
Question	$\times 0.25$, Raw	91.67	127.58	20.61	25.60
Question	$\times 0.25$, Norm	91.67	128.84	20.87	26.99
Question	$\times 0.5$, Raw	91.67	124.84	20.20	27.58
Question	$\times 0.5$, Norm	91.67	135.53	21.99	29.25
Question	$\times 2$, Raw	6.25	5.71	7.95	2.93
Question	$\times 2$, Norm	16.67	13.61	3.91	0.38
Question	$\times 4$, Raw	4.17	4.77	6.34	2.34
Question	$\times 4$, Norm	16.67	20.27	8.76	5.09
Template	$\times 0.25$, Raw	0.00	0.00	0.00	0.00

Continued on next page

Continued: Dream-VL: complete C1 matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
Template	$\times 0.25$, Norm	12.50	4.82	1.14	0.00
Template	$\times 0.5$, Raw	22.92	14.52	3.06	0.00
Template	$\times 0.5$, Norm	54.17	39.79	6.24	7.15
Template	$\times 2$, Raw	0.00	0.82	0.19	0.00
Template	$\times 2$, Norm	4.17	0.00	0.00	0.00
Template	$\times 4$, Raw	0.00	51.47	11.59	13.80
Template	$\times 4$, Norm	0.00	105.34	18.05	24.37
Response	$\times 0.25$, Raw	54.17	37.50	16.55	8.43
Response	$\times 0.25$, Norm	64.58	37.79	15.79	6.79
Response	$\times 0.5$, Raw	68.75	56.54	17.23	10.60
Response	$\times 0.5$, Norm	66.67	42.91	18.30	11.42
Response	$\times 2$, Raw	0.00	0.00	0.00	0.00
Response	$\times 2$, Norm	0.00	0.78	0.00	0.00
Response	$\times 4$, Raw	0.00	0.00	0.00	0.00
Response	$\times 4$, Norm	0.00	0.02	0.13	0.00

Table 29: MMaDA: complete C1 matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
All	Native	12.50	106.05	17.86	24.57
Image	$\times 0.25$, Raw	2.08	0.00	0.21	0.00
Image	$\times 0.25$, Norm	0.00	0.67	0.39	0.00
Image	$\times 0.5$, Raw	4.17	0.97	0.32	0.00
Image	$\times 0.5$, Norm	4.17	9.05	2.41	4.02
Image	$\times 2$, Raw	12.50	5.35	1.92	0.00
Image	$\times 2$, Norm	0.00	0.00	0.00	0.00
Image	$\times 4$, Raw	NA	NA	NA	NA
Image	$\times 4$, Norm	4.17	0.00	0.00	0.00
Question	$\times 0.25$, Raw	0.00	19.17	4.66	3.89
Question	$\times 0.25$, Norm	4.17	69.08	15.55	17.60
Question	$\times 0.5$, Raw	0.00	22.58	4.43	4.48
Question	$\times 0.5$, Norm	0.00	105.49	18.06	23.74
Question	$\times 2$, Raw	0.00	56.88	18.74	11.45
Question	$\times 2$, Norm	0.00	63.60	19.07	13.20
Question	$\times 4$, Raw	NA	NA	NA	NA
Question	$\times 4$, Norm	0.00	62.99	18.78	13.77
Template	$\times 0.25$, Raw	0.00	0.02	0.00	0.00
Template	$\times 0.25$, Norm	0.00	0.42	0.00	0.00
Template	$\times 0.5$, Raw	0.00	0.00	0.00	0.00
Template	$\times 0.5$, Norm	0.00	0.16	0.00	0.00
Template	$\times 2$, Raw	6.25	9.48	1.48	0.00
Template	$\times 2$, Norm	0.00	0.00	0.00	0.00
Template	$\times 4$, Raw	NA	NA	NA	NA
Template	$\times 4$, Norm	2.08	0.00	0.00	0.00
Response	$\times 0.25$, Raw	0.00	0.02	0.00	0.00
Response	$\times 0.25$, Norm	0.00	0.00	0.00	0.00
Response	$\times 0.5$, Raw	0.00	0.02	0.00	0.00
Response	$\times 0.5$, Norm	0.00	0.14	0.00	0.00
Response	$\times 2$, Raw	0.00	0.00	0.00	0.00
Response	$\times 2$, Norm	0.00	0.00	0.00	0.00
Response	$\times 4$, Raw	NA	NA	NA	NA
Response	$\times 4$, Norm	0.00	0.00	0.00	0.00

H.5 Replacing current amplitudes with calibrated references

C2 asks whether the current peak amplitude can be replaced by a magnitude reference. For each model and task, eight open calibration images are separate from the scientific and technical inputs. For each layer, update, and region, a within-image median of token peak magnitudes is followed by a median over calibration images, yielding $q_{\ell kR}$. Define

$$\bar{q}_{\ell R} = \left(K^{-1} \sum_{k=1}^{K} q_{\ell kR}^2 \right)^{1/2}. \tag{20}$$

The current peak is replaced, with its current sign, by $q_{\ell kR}$, $\bar{q}_{\ell R}$, or $4\bar{q}_{\ell R}$. The constant reference is constant across updates within a model, task, layer, and region, not across the whole network. Peak location, sign, and all other coordinates remain input-dependent. Norm versions additionally restore the current vector norm. This test therefore isolates amplitude replacement while preserving other channels of input dependence.

Dream-VL image and question ChartQA scores remain at 91.67 under the typical and constant Raw references, whereas template scores are zero and response scores differ. On COCO response tokens, constant-minus-typical differences are 46.93 [33.94,61.46], 5.57 [-7.17,18.32], and 33.41 [18.89,49.78] for LaViDa, Dream-VL, and MMaDA, respectively. These existing exploratory paired intervals show that reference scheduling can matter, without yielding one universally effective magnitude schedule. The fixed-reference experiment comprises 6,912 scientific generation conditions plus 48 calibration captures.

Table 30: LaViDa: complete C2 matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
All	Native	70.83	72.19	13.04	15.66
Image	q_k, Raw	50.00	72.38	12.72	14.98
Image	q_k, Norm	10.42	0.41	0.31	0.00
Image	$\bar{q}$, Raw	50.00	72.38	12.72	14.98
Image	$\bar{q}$, Norm	10.42	0.41	0.31	0.00
Image	$4\bar{q}$, Raw	8.33	29.80	6.41	3.97
Image	$4\bar{q}$, Norm	10.42	3.62	1.24	0.00
Question	q_k, Raw	58.33	70.14	13.29	14.99
Question	q_k, Norm	2.08	74.75	13.31	14.91
Question	$\bar{q}$, Raw	58.33	70.14	13.29	14.99
Question	$\bar{q}$, Norm	2.08	74.75	13.31	14.91
Question	$4\bar{q}$, Raw	2.08	4.45	1.41	0.00
Question	$4\bar{q}$, Norm	0.00	2.21	0.73	0.00
Template	q_k, Raw	35.42	57.96	11.10	10.61
Template	q_k, Norm	0.00	0.00	0.00	0.00
Template	$\bar{q}$, Raw	35.42	57.96	11.10	10.61
Template	$\bar{q}$, Norm	0.00	0.00	0.00	0.00
Template	$4\bar{q}$, Raw	39.58	60.92	11.07	12.54
Template	$4\bar{q}$, Norm	0.00	34.44	5.16	4.57
Response	q_k, Raw	70.83	37.63	10.55	5.76
Response	q_k, Norm	4.17	0.00	0.00	0.00
Response	$\bar{q}$, Raw	66.67	84.56	13.70	18.25
Response	$\bar{q}$, Norm	6.25	5.44	0.90	0.00
Response	$4\bar{q}$, Raw	0.00	0.00	0.00	0.00
Response	$4\bar{q}$, Norm	0.00	0.00	0.00	0.00

Table 31: Dream-VL: complete C2 matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
All	Native	91.67	132.96	20.85	26.17
Image	q_k, Raw	91.67	129.12	20.85	26.28
Image	q_k, Norm	91.67	122.88	20.34	25.09
Image	$\bar{q}$, Raw	91.67	125.01	21.31	24.71
Image	$\bar{q}$, Norm	91.67	124.79	19.80	26.14
Image	$4\bar{q}$, Raw	12.50	108.47	16.32	23.30
Image	$4\bar{q}$, Norm	4.17	83.17	15.84	18.72
Question	q_k, Raw	91.67	133.05	21.01	28.71
Question	q_k, Norm	91.67	128.10	21.01	26.43
Question	$\bar{q}$, Raw	91.67	138.26	21.62	29.91
Question	$\bar{q}$, Norm	91.67	128.40	21.04	25.36
Question	$4\bar{q}$, Raw	75.00	135.24	20.15	26.64
Question	$4\bar{q}$, Norm	29.17	106.06	15.93	20.60
Template	q_k, Raw	0.00	0.00	0.00	0.00
Template	q_k, Norm	0.00	0.00	0.00	0.00
Template	$\bar{q}$, Raw	0.00	0.00	0.00	0.00
Template	$\bar{q}$, Norm	0.00	0.00	0.00	0.00
Template	$4\bar{q}$, Raw	0.00	0.00	0.00	0.00
Template	$4\bar{q}$, Norm	0.00	0.00	0.00	0.00
Response	q_k, Raw	62.50	11.46	12.90	4.60
Response	q_k, Norm	60.42	2.51	11.95	3.78
Response	$\bar{q}$, Raw	52.08	17.02	14.34	5.59
Response	$\bar{q}$, Norm	54.17	9.46	15.03	5.54
Response	$4\bar{q}$, Raw	2.08	0.00	0.00	0.00
Response	$4\bar{q}$, Norm	0.00	0.00	0.00	0.00

Table 32: MMaDA: complete C2 matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
All	Native	12.50	106.05	17.86	24.57
Image	q_k, Raw	12.50	113.47	18.95	29.04
Image	q_k, Norm	10.42	64.81	10.93	14.69
Image	$\bar{q}$, Raw	12.50	110.64	18.50	26.73
Image	$\bar{q}$, Norm	10.42	65.90	11.20	14.60
Image	$4\bar{q}$, Raw	6.25	0.00	0.00	0.00
Image	$4\bar{q}$, Norm	8.33	0.00	0.00	0.00
Question	q_k, Raw	12.50	102.52	17.34	23.60
Question	q_k, Norm	10.42	112.35	18.16	26.45
Question	$\bar{q}$, Raw	10.42	108.95	17.89	27.09
Question	$\bar{q}$, Norm	10.42	107.49	17.84	26.09
Question	$4\bar{q}$, Raw	0.00	67.43	19.48	15.50
Question	$4\bar{q}$, Norm	0.00	64.41	19.42	13.49
Template	q_k, Raw	0.00	0.00	0.00	0.00
Template	q_k, Norm	0.00	0.03	0.00	0.00
Template	$\bar{q}$, Raw	0.00	0.00	0.00	0.00
Template	$\bar{q}$, Norm	0.00	0.04	0.00	0.00
Template	$4\bar{q}$, Raw	2.08	1.46	0.15	0.00
Template	$4\bar{q}$, Norm	4.17	1.60	0.15	0.00
Response	q_k, Raw	2.08	39.78	10.09	11.44
Response	q_k, Norm	0.00	0.03	0.00	0.00
Response	$\bar{q}$, Raw	6.25	73.20	13.62	15.80
Response	$\bar{q}$, Norm	2.08	2.87	0.60	0.00

Continued on next page

Continued: MMaDA: complete C2 matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
Response	$4\bar{q}$, Raw	0.00	0.00	0.00	0.00
Response	$4\bar{q}$, Norm	0.00	0.00	0.00	0.00

H.6 ALL DOWNSTREAM-ACCESS CONFIGURATIONS

Access codes are ordered QK/V/Direct. D000 equals Zero and Native corresponds to D111. Each zero bit removes only the selected coordinate from that consumer.

Table 33: LaViDa: complete D matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
All	Native	70.83	72.19	13.04	15.66
Image	D000	0.00	0.02	0.00	0.00
Image	D001	0.00	7.01	1.33	0.00
Image	D010	0.00	0.00	0.00	0.00
Image	D011	50.00	53.83	11.77	10.43
Image	D100	2.08	0.07	0.00	0.00
Image	D101	72.92	73.31	14.59	15.73
Image	D110	2.08	0.04	0.00	0.00
Question	D000	0.00	0.00	0.00	0.00
Question	D001	0.00	0.11	0.00	0.00
Question	D010	0.00	0.32	0.00	0.00
Question	D011	2.08	4.04	1.03	0.00
Question	D100	2.08	1.42	0.48	0.00
Question	D101	75.00	76.02	13.81	14.15
Question	D110	0.00	3.93	0.32	0.00
Template	D000	0.00	0.04	0.00	0.00
Template	D001	0.00	0.00	0.00	0.00
Template	D010	0.00	0.00	0.00	0.00
Template	D011	0.00	4.68	1.04	0.00
Template	D100	2.08	0.00	0.23	0.00
Template	D101	68.75	70.63	12.75	14.02
Template	D110	0.00	0.09	0.00	0.00
Response	D000	0.00	0.00	0.00	0.00
Response	D001	0.00	0.54	0.00	0.00
Response	D010	0.00	0.06	0.00	0.00
Response	D011	0.00	0.57	0.14	0.00
Response	D100	0.00	1.45	0.17	0.00
Response	D101	66.67	82.69	15.54	17.83
Response	D110	2.08	0.00	0.00	0.00

Table 34: Dream-VL: complete D matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
All	Native	91.67	132.96	20.85	26.17
Image	D000	91.67	131.45	19.11	28.45
Image	D001	89.58	135.67	21.11	28.81
Image	D010	87.50	131.44	19.24	28.79
Image	D011	91.67	128.46	20.03	27.67
Image	D100	89.58	130.43	19.33	27.53
Image	D101	91.67	125.56	21.24	25.37

Continued on next page

Continued: Dream-VL: complete D matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
Image	D110	85.42	131.98	19.22	27.27
Question	D000	91.67	130.20	20.50	28.57
Question	D001	91.67	121.57	19.92	25.52
Question	D010	91.67	124.59	19.79	26.51
Question	D011	91.67	126.80	21.06	27.48
Question	D100	91.67	129.47	20.76	27.08
Question	D101	91.67	125.61	20.44	24.35
Question	D110	91.67	129.59	20.05	28.37
Template	D000	0.00	0.00	0.00	0.00
Template	D001	89.58	127.53	20.58	27.59
Template	D010	14.58	0.00	0.00	0.00
Template	D011	89.58	121.00	20.19	22.80
Template	D100	16.67	0.41	0.23	0.00
Template	D101	89.58	138.96	20.54	32.10
Template	D110	6.25	0.12	0.12	0.00
Response	D000	41.67	47.53	14.57	9.22
Response	D001	79.17	120.11	19.14	24.30
Response	D010	45.83	35.37	13.67	7.01
Response	D011	91.67	125.47	19.67	27.66
Response	D100	47.92	40.33	15.86	8.27
Response	D101	85.42	136.26	20.67	28.51
Response	D110	37.50	43.15	15.33	11.23

Table 35: MMaDA: complete D matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
All	Native	12.50	106.05	17.86	24.57
Image	D000	0.00	0.06	0.11	0.00
Image	D001	8.33	8.49	1.59	0.00
Image	D010	0.00	0.00	0.00	0.00
Image	D011	8.33	51.80	9.80	12.33
Image	D100	0.00	0.02	0.00	0.00
Image	D101	12.50	116.81	19.44	26.47
Image	D110	0.00	0.13	0.00	0.00
Question	D000	0.00	40.80	16.08	12.19
Question	D001	0.00	57.01	18.84	11.99
Question	D010	0.00	3.51	0.65	0.00
Question	D011	0.00	61.31	19.74	12.81
Question	D100	0.00	0.00	0.00	0.00
Question	D101	10.42	109.85	17.76	25.73
Question	D110	0.00	0.00	0.00	0.00
Template	D000	0.00	0.11	0.00	0.00
Template	D001	0.00	0.00	0.00	0.00
Template	D010	0.00	0.17	0.00	0.00
Template	D011	4.17	85.66	14.29	17.31
Template	D100	0.00	0.03	0.00	0.00
Template	D101	8.33	95.54	15.18	21.79
Template	D110	0.00	0.34	0.00	0.00
Response	D000	0.00	0.77	0.00	0.00
Response	D001	2.08	28.21	6.75	3.37
Response	D010	0.00	0.09	0.00	0.00
Response	D011	8.33	88.55	13.81	20.35
Response	D100	0.00	1.08	0.00	0.00

Continued on next page

Continued: MMaDA: complete D matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
Response	D101	12.50	101.26	16.38	22.33
Response	D110	0.00	0.39	0.00	0.00

H.7 ALL CARRY, READER, AND TERMINAL CONFIGURATIONS

D110, Base, G, P, PG, R, and T follow the current-state definitions in Appendix F. Secondary caption metrics and all four regions are included. E and G are views of this same experiment, not independent replications.

Table 36: LaViDa: complete EG matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
All	Native	70.83	72.19	13.04	15.66
Image	Base	2.08	0.04	0.00	0.00
Image	G	0.00	0.02	0.00	0.00
Image	P	0.00	0.14	0.00	0.00
Image	PG	0.00	0.09	0.00	0.00
Image	R	16.67	3.30	0.86	0.00
Image	T	70.83	72.19	13.04	15.66
Image	D110	2.08	0.04	0.00	0.00
Question	Base	0.00	3.93	0.32	0.00
Question	G	0.00	0.53	0.17	0.00
Question	P	2.08	0.62	0.17	0.00
Question	PG	0.00	1.40	0.53	0.00
Question	R	52.08	78.82	12.70	13.79
Question	T	70.83	72.19	13.04	15.66
Question	D110	0.00	3.93	0.32	0.00
Template	Base	0.00	0.09	0.00	0.00
Template	G	0.00	0.00	0.00	0.00
Template	P	0.00	0.20	0.00	0.00
Template	PG	0.00	0.01	0.00	0.00
Template	R	56.25	68.98	11.40	14.26
Template	T	70.83	72.19	13.04	15.66
Template	D110	0.00	0.09	0.00	0.00
Response	Base	0.00	0.03	0.00	0.00
Response	G	2.08	0.50	0.00	0.00
Response	P	0.00	0.00	0.00	0.00
Response	PG	0.00	0.26	0.00	0.00
Response	R	43.75	54.64	9.41	12.70
Response	T	72.92	41.45	11.42	6.99
Response	D110	2.08	0.00	0.00	0.00

Table 37: Dream-VL: complete EG matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
All	Native	91.67	132.96	20.85	26.17
Image	Base	85.42	131.98	19.22	27.27
Image	G	87.50	119.93	20.31	25.71
Image	P	87.50	131.14	19.75	26.30
Image	PG	89.58	126.97	19.53	25.53
Image	R	93.75	136.10	21.84	30.63

Continued on next page

Continued: Dream-VL: complete EG matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
Image	T	91.67	132.96	20.85	26.17
Image	D110	85.42	131.98	19.22	27.27
Question	Base	91.67	129.59	20.05	28.37
Question	G	91.67	129.40	20.27	28.94
Question	P	91.67	126.41	20.61	26.49
Question	PG	91.67	138.52	21.67	32.04
Question	R	89.58	134.72	21.00	28.68
Question	T	91.67	132.96	20.85	26.17
Question	D110	91.67	129.59	20.05	28.37
Template	Base	8.33	0.00	0.00	0.00
Template	G	22.92	0.48	0.12	0.00
Template	P	10.42	0.00	0.00	0.00
Template	PG	6.25	0.00	0.00	0.00
Template	R	91.67	133.79	21.12	34.07
Template	T	87.50	131.43	20.84	27.84
Template	D110	6.25	0.12	0.12	0.00
Response	Base	52.08	87.94	20.36	20.59
Response	G	64.58	95.58	20.29	20.54
Response	P	60.42	101.29	19.34	20.13
Response	PG	68.75	100.63	19.84	21.97
Response	R	87.50	122.07	21.11	25.22
Response	T	54.17	2.16	11.11	2.71
Response	D110	37.50	43.15	15.33	11.23

Table 38: MMaDA: complete EG matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
All	Native	12.50	106.05	17.86	24.57
Image	Base	0.00	0.13	0.00	0.00
Image	G	2.08	0.00	0.00	0.00
Image	P	0.00	0.04	0.00	0.00
Image	PG	2.08	0.04	0.00	0.00
Image	R	10.42	89.31	15.47	24.29
Image	T	12.50	106.05	17.86	24.57
Image	D110	0.00	0.13	0.00	0.00
Question	Base	0.00	0.00	0.00	0.00
Question	G	0.00	63.54	18.13	13.84
Question	P	0.00	0.00	0.00	0.00
Question	PG	2.08	57.79	17.92	14.07
Question	R	4.17	65.36	18.08	11.73
Question	T	12.50	106.05	17.86	24.57
Question	D110	0.00	0.00	0.00	0.00
Template	Base	0.00	0.34	0.00	0.00
Template	G	0.00	0.00	0.00	0.00
Template	P	0.00	0.04	0.00	0.00
Template	PG	0.00	0.04	0.00	0.00
Template	R	8.33	104.91	20.19	25.08
Template	T	12.50	106.05	17.86	24.57
Template	D110	0.00	0.34	0.00	0.00
Response	Base	0.00	0.12	0.00	0.00
Response	G	0.00	0.30	0.00	0.00
Response	P	0.00	0.00	0.00	0.00
Response	PG	0.00	0.62	0.00	0.00

Continued on next page

Continued: MMaDA: complete EG matrix.

Region	Condition	ChartQA	CIDEr	SPICE	B-4
Response	R	12.50	32.01	6.73	6.63
Response	T	12.50	83.07	20.12	20.43
Response	D110	0.00	0.39	0.00	0.00

H.8 Task subgroups and output diagnostics

Subgroups provide task-level context for the regional effects. ScienceQA grade groups contain 34 and 14 inputs, while GQA binary/open and ChartQA human/augmented each contain 24 inputs. These are subdivisions of the original subsets. The accompanying diagnostic records also distinguish changes in visible answers, final tokens, and denoising trajectories. For example, Dream-VL ChartQA image Zero preserves the aggregate 32-update accuracy, yet changes 6 of 48 visible answers, 9 final token sequences, and 47 trajectories, with one gain and one loss. By contrast, template suppression can produce 48 empty answers. Unchanged aggregate accuracy and empty-output failure therefore represent different outcomes. The detailed trajectory audit covers its explicitly listed Dream-VL ScienceQA and ChartQA conditions, not all backbones. MinAbs frequently writes actual zero in native precision and is not a guaranteed nonzero control.

Table 39: Regional-reset subgroups: SCIENCEQA, 32 updates.

Region	Model	Treatment	All	Grades 1–6	Grades 7–12
Image	LaViDa	Native	79.17	82.35	71.43
Image	Dream-VL	Native	93.75	97.06	85.71
Image	MMaDA	Native	64.58	61.76	71.43
Image	LaViDa	Zero	2.08	2.94	0.00
Image	Dream-VL	Zero	91.67	97.06	78.57
Image	MMaDA	Zero	0.00	0.00	0.00
Image	LaViDa	Mean	2.08	2.94	0.00
Image	Dream-VL	Mean	91.67	97.06	78.57
Image	MMaDA	Mean	0.00	0.00	0.00
Image	LaViDa	MinAbs	2.08	2.94	0.00
Image	Dream-VL	MinAbs	93.75	97.06	85.71
Image	MMaDA	MinAbs	0.00	0.00	0.00
Image	LaViDa	Random	79.17	82.35	71.43
Image	Dream-VL	Random	93.75	97.06	85.71
Image	MMaDA	Random	64.58	61.76	71.43
Question	LaViDa	Native	79.17	82.35	71.43
Question	Dream-VL	Native	93.75	97.06	85.71
Question	MMaDA	Native	64.58	61.76	71.43
Question	LaViDa	Zero	0.00	0.00	0.00
Question	Dream-VL	Zero	85.42	88.24	78.57
Question	MMaDA	Zero	2.08	2.94	0.00
Question	LaViDa	Mean	0.00	0.00	0.00
Question	Dream-VL	Mean	91.67	94.12	85.71
Question	MMaDA	Mean	0.00	0.00	0.00
Question	LaViDa	MinAbs	0.00	0.00	0.00
Question	Dream-VL	MinAbs	89.58	94.12	78.57
Question	MMaDA	MinAbs	0.00	0.00	0.00
Question	LaViDa	Random	81.25	85.29	71.43
Question	Dream-VL	Random	93.75	97.06	85.71
Question	MMaDA	Random	64.58	61.76	71.43
Template	LaViDa	Native	79.17	82.35	71.43
Template	Dream-VL	Native	93.75	97.06	85.71
Template	MMaDA	Native	64.58	61.76	71.43
Template	LaViDa	Zero	0.00	0.00	0.00

Continued on next page

Continued: Regional-reset subgroups: SCIENCEQA, 32 updates.

Region	Model	Treatment	All	Grades 1–6	Grades 7–12
Template	Dream-VL	Zero	0.00	0.00	0.00
Template	MMaDA	Zero	0.00	0.00	0.00
Template	LaViDa	Mean	0.00	0.00	0.00
Template	Dream-VL	Mean	0.00	0.00	0.00
Template	MMaDA	Mean	0.00	0.00	0.00
Template	LaViDa	MinAbs	0.00	0.00	0.00
Template	Dream-VL	MinAbs	0.00	0.00	0.00
Template	MMaDA	MinAbs	0.00	0.00	0.00
Template	LaViDa	Random	79.17	82.35	71.43
Template	Dream-VL	Random	93.75	97.06	85.71
Template	MMaDA	Random	64.58	61.76	71.43
Response	LaViDa	Native	79.17	82.35	71.43
Response	Dream-VL	Native	93.75	97.06	85.71
Response	MMaDA	Native	64.58	61.76	71.43
Response	LaViDa	Zero	4.17	5.88	0.00
Response	Dream-VL	Zero	89.58	97.06	71.43
Response	MMaDA	Zero	0.00	0.00	0.00
Response	LaViDa	Mean	4.17	2.94	7.14
Response	Dream-VL	Mean	89.58	97.06	71.43
Response	MMaDA	Mean	0.00	0.00	0.00
Response	LaViDa	MinAbs	2.08	2.94	0.00
Response	Dream-VL	MinAbs	89.58	97.06	71.43
Response	MMaDA	MinAbs	0.00	0.00	0.00
Response	LaViDa	Random	81.25	85.29	71.43
Response	Dream-VL	Random	93.75	97.06	85.71
Response	MMaDA	Random	64.58	61.76	71.43

Table 40: Regional-reset subgroups: GQA, 32 updates.

Region	Model	Treatment	All	Binary	Open
Image	LaViDa	Native	0.00	0.00	0.00
Image	Dream-VL	Native	20.83	37.50	4.17
Image	MMaDA	Native	0.00	0.00	0.00
Image	LaViDa	Zero	0.00	0.00	0.00
Image	Dream-VL	Zero	29.17	50.00	8.33
Image	MMaDA	Zero	0.00	0.00	0.00
Image	LaViDa	Mean	0.00	0.00	0.00
Image	Dream-VL	Mean	27.08	45.83	8.33
Image	MMaDA	Mean	0.00	0.00	0.00
Image	LaViDa	MinAbs	0.00	0.00	0.00
Image	Dream-VL	MinAbs	29.17	50.00	8.33
Image	MMaDA	MinAbs	0.00	0.00	0.00
Image	LaViDa	Random	0.00	0.00	0.00
Image	Dream-VL	Random	18.75	33.33	4.17
Image	MMaDA	Random	0.00	0.00	0.00
Question	LaViDa	Native	0.00	0.00	0.00
Question	Dream-VL	Native	20.83	37.50	4.17
Question	MMaDA	Native	0.00	0.00	0.00
Question	LaViDa	Zero	0.00	0.00	0.00
Question	Dream-VL	Zero	6.25	12.50	0.00
Question	MMaDA	Zero	0.00	0.00	0.00
Question	LaViDa	Mean	0.00	0.00	0.00
Question	Dream-VL	Mean	2.08	4.17	0.00
Question	MMaDA	Mean	0.00	0.00	0.00

Continued on next page

Continued: Regional-reset subgroups: GQA, 32 updates.

Region	Model	Treatment	All	Binary	Open
Question	LaViDa	MinAbs	0.00	0.00	0.00
Question	Dream-VL	MinAbs	2.08	4.17	0.00
Question	MMaDA	MinAbs	0.00	0.00	0.00
Question	LaViDa	Random	0.00	0.00	0.00
Question	Dream-VL	Random	18.75	33.33	4.17
Question	MMaDA	Random	0.00	0.00	0.00
Template	LaViDa	Native	0.00	0.00	0.00
Template	Dream-VL	Native	20.83	37.50	4.17
Template	MMaDA	Native	0.00	0.00	0.00
Template	LaViDa	Zero	0.00	0.00	0.00
Template	Dream-VL	Zero	0.00	0.00	0.00
Template	MMaDA	Zero	0.00	0.00	0.00
Template	LaViDa	Mean	0.00	0.00	0.00
Template	Dream-VL	Mean	0.00	0.00	0.00
Template	MMaDA	Mean	0.00	0.00	0.00
Template	LaViDa	MinAbs	0.00	0.00	0.00
Template	Dream-VL	MinAbs	0.00	0.00	0.00
Template	MMaDA	MinAbs	0.00	0.00	0.00
Template	LaViDa	Random	0.00	0.00	0.00
Template	Dream-VL	Random	20.83	37.50	4.17
Template	MMaDA	Random	0.00	0.00	0.00
Response	LaViDa	Native	0.00	0.00	0.00
Response	Dream-VL	Native	20.83	37.50	4.17
Response	MMaDA	Native	0.00	0.00	0.00
Response	LaViDa	Zero	0.00	0.00	0.00
Response	Dream-VL	Zero	20.83	41.67	0.00
Response	MMaDA	Zero	0.00	0.00	0.00
Response	LaViDa	Mean	0.00	0.00	0.00
Response	Dream-VL	Mean	22.92	41.67	4.17
Response	MMaDA	Mean	0.00	0.00	0.00
Response	LaViDa	MinAbs	0.00	0.00	0.00
Response	Dream-VL	MinAbs	20.83	41.67	0.00
Response	MMaDA	MinAbs	0.00	0.00	0.00
Response	LaViDa	Random	0.00	0.00	0.00
Response	Dream-VL	Random	18.75	33.33	4.17
Response	MMaDA	Random	0.00	0.00	0.00

Table 41: Regional-reset subgroups: CHARTQA, 32 updates.

Region	Model	Treatment	All	Human	Augmented
Image	LaViDa	Native	70.83	58.33	83.33
Image	Dream-VL	Native	91.67	87.50	95.83
Image	MMaDA	Native	12.50	8.33	16.67
Image	LaViDa	Zero	0.00	0.00	0.00
Image	Dream-VL	Zero	91.67	87.50	95.83
Image	MMaDA	Zero	0.00	0.00	0.00
Image	LaViDa	Mean	0.00	0.00	0.00
Image	Dream-VL	Mean	91.67	87.50	95.83
Image	MMaDA	Mean	0.00	0.00	0.00
Image	LaViDa	MinAbs	0.00	0.00	0.00
Image	Dream-VL	MinAbs	89.58	87.50	91.67
Image	MMaDA	MinAbs	0.00	0.00	0.00
Image	LaViDa	Random	70.83	58.33	83.33
Image	Dream-VL	Random	91.67	87.50	95.83

Continued on next page

Continued: Regional-reset subgroups: CHARTQA, 32 updates.

Region	Model	Treatment	All	Human	Augmented
Image	MMaDA	Random	14.58	12.50	16.67
Question	LaViDa	Native	70.83	58.33	83.33
Question	Dream-VL	Native	91.67	87.50	95.83
Question	MMaDA	Native	12.50	8.33	16.67
Question	LaViDa	Zero	0.00	0.00	0.00
Question	Dream-VL	Zero	91.67	87.50	95.83
Question	MMaDA	Zero	0.00	0.00	0.00
Question	LaViDa	Mean	0.00	0.00	0.00
Question	Dream-VL	Mean	91.67	87.50	95.83
Question	MMaDA	Mean	0.00	0.00	0.00
Question	LaViDa	MinAbs	0.00	0.00	0.00
Question	Dream-VL	MinAbs	91.67	87.50	95.83
Question	MMaDA	MinAbs	0.00	0.00	0.00
Question	LaViDa	Random	68.75	54.17	83.33
Question	Dream-VL	Random	91.67	87.50	95.83
Question	MMaDA	Random	12.50	8.33	16.67
Template	LaViDa	Native	70.83	58.33	83.33
Template	Dream-VL	Native	91.67	87.50	95.83
Template	MMaDA	Native	12.50	8.33	16.67
Template	LaViDa	Zero	0.00	0.00	0.00
Template	Dream-VL	Zero	0.00	0.00	0.00
Template	MMaDA	Zero	0.00	0.00	0.00
Template	LaViDa	Mean	0.00	0.00	0.00
Template	Dream-VL	Mean	0.00	0.00	0.00
Template	MMaDA	Mean	0.00	0.00	0.00
Template	LaViDa	MinAbs	0.00	0.00	0.00
Template	Dream-VL	MinAbs	0.00	0.00	0.00
Template	MMaDA	MinAbs	0.00	0.00	0.00
Template	LaViDa	Random	68.75	54.17	83.33
Template	Dream-VL	Random	91.67	87.50	95.83
Template	MMaDA	Random	12.50	8.33	16.67
Response	LaViDa	Native	70.83	58.33	83.33
Response	Dream-VL	Native	91.67	87.50	95.83
Response	MMaDA	Native	12.50	8.33	16.67
Response	LaViDa	Zero	0.00	0.00	0.00
Response	Dream-VL	Zero	41.67	41.67	41.67
Response	MMaDA	Zero	0.00	0.00	0.00
Response	LaViDa	Mean	0.00	0.00	0.00
Response	Dream-VL	Mean	45.83	45.83	45.83
Response	MMaDA	Mean	0.00	0.00	0.00
Response	LaViDa	MinAbs	0.00	0.00	0.00
Response	Dream-VL	MinAbs	39.58	45.83	33.33
Response	MMaDA	MinAbs	0.00	0.00	0.00
Response	LaViDa	Random	70.83	58.33	83.33
Response	Dream-VL	Random	89.58	87.50	91.67
Response	MMaDA	Random	12.50	8.33	16.67

I Extended Related Work and Generation Notation

I.1 Diffusion Multimodal Language Models

Diffusion multimodal language models extend masked denoising to visually conditioned generation (Austin et al., 2021; Sahoo et al., 2024; Li et al., 2025). Building on large-scale diffusion language modeling (Nie et al., 2025), LaViDa, LLaDA-V, and Dream-VL adapt diffusion backbones for visual instruction following (Li et al., 2025; You et al., 2026; Ye et al., 2026), while MMaDA unifies textual reasoning, multimodal understanding, and image generation (Yang et al., 2025). LaViDa

also uses prefix caching and timestep shifting, illustrating how inference design affects information reuse and the schedule of denoising updates (Li et al., 2025). Recent work examines the denoising process itself. DLM-Scope extracts interpretable features in text diffusion models (Wang et al., 2026), Thinking Diffusion studies early answer commitment and visual dependence in multimodal reasoning (Kim et al., 2026), and Diffusion-CAM provides gradient-based visual explanations (Zuo et al., 2026). Our study focuses on residual components shared across token roles and the operations through which they affect generation. This complements feature- and attention-level analyses with interventions on the representation passed between network blocks.

I.2 Massive Activations and Representation Geometry

Massive activations are unusually large hidden-state components that can remain comparatively stable across inputs and provide bias-like computational support (Sun et al., 2024). Related studies connect large Q/K coordinates to contextual knowledge use (Jin et al., 2025) and large residual components to visual attention sinks in autoregressive multimodal models (Kang et al., 2025). Residual activations, projected Q/K coordinates, and attention probabilities occupy different stages of the computation. Their relationship motivates tracing downstream use rather than assigning the same function to every large value. Recent work distinguishes massive activations from attention sinks (Sun et al., 2026) and links their emergence in studied LLMs to interactions between RMSNorm and feed-forward layers (Shi et al., 2026). SpinQuant uses coordinated orthogonal rotations of activations and weights to redistribute outliers while preserving full-precision computation (Liu et al., 2025), highlighting the basis dependence of coordinate peaks. We investigate the fixed native basis of dMLLMs, separating regional dependence, magnitude and direction, and residual transmission from normalized reading.

I.3 Diffusion Generation and Token Roles

Given an image I and a question Q, a masked dMLLM initializes M response positions as an all-mask state $Z^{(0)}$ (Sahoo et al., 2024; Nie et al., 2025). At step k, it predicts token distributions $P^{(k)}$ and updates the response:

$$P^{(k)} = f_\theta(I, Q, Z^{(k)}), \qquad Z^{(k+1)} = \mathcal{S}_k(Z^{(k)}, P^{(k)}; \xi_k). \tag{21}$$

The update rule $\mathcal{S}_k$ selects the positions and token values to change, with randomness ξ_k when applicable. The multimodal sequence comprises image, question, template/special, and response tokens. Template/special tokens specify prompt format, while response positions can be masked or revealed. In models with prefix key/value caching, prefix residuals are computed during prefill and response residuals during denoising. The main experiments retain each backbone's native multimodal input format and decoding implementation. Differences in cache behavior imply differences in intervention exposure across backbones, even when the same named token region is targeted.