Supplemental Material: Z++: Temporal, Non-Uniform, and 3D Z Sampling

ABDALLA G. M. AHMED, Shenzhen University, China
MATT PHARR, NVIDIA, USA
HUI HUANG*, Shenzhen University, China

Fig. S1. Layout of a representative video frame. The four temporal models occupy the top row, while the four benchmarking references occupy the bottom row. Color-coded insets magnify the corresponding regions marked in the reference panel.

S-1 SUPPLEMENTARY VIDEO

Unlike conventional optimization-based screen-space spatiotemporal samplers, which target preset configurations, temporal Z++ is a generic construction that can be readily incorporated into a broad range of Z-indexed samplers, typically amounting to only a few additional bit-manipulation operations in the sampler code. The accompanying video demonstrates this flexibility. The remainder of this section briefly explains the video composition and serves as a guide to its interpretation.

*Corresponding author: Hui Huang (hhzhiyan@gmail.com)

Authors' addresses: Abdalla G. M. Ahmed, Visual Computing Research Center (VCC), College of Computer Science and Software Engineering (CSSE), Shenzhen University, China, abdalla_gafar@hotmail.com; Matt Pharr, NVIDIA, USA, matt@pharr.org; Hui Huang, VCC, CSSE, Shenzhen University, China, hhzhiyan@gmail.com.

S-1.1 Scene Design

The scene is based on the coffeemaker model distributed through Bitterli's Rendering Resources [Bitterli 2016]. We adapt it for our purposes as follows:

- animating the coffeemaker to rotate about its vertical axis;
- placing a stationary Utah teapot from the same collection beside it; and
- replacing the original area lights with a realistic café environment map.

We additionally assign a subsurface material to the teapot. This compact setup combines moving and stationary regions with nontrivial light transport arising from glossy interreflections and subsurface scattering within the teapot. In practice, this combination proved challenging for the tested state-of-the-art samplers at the evaluated sampling rates.

S-1.2 Temporal Averaging

A practical temporal-antialiasing (TAA) implementation, typically integrated into a real-time renderer, uses motion vectors computed from the current and previous camera and object transformations to reproject the accumulated history into the current frame. For this demonstration, we avoid this implementation-dependent complexity and instead apply a 16-frame box moving average using FFmpeg's `tmix` filter. The scene is designed to contain both stationary objects and rotationally invariant regions of the moving coffeemaker, whose image-space appearance remains aligned throughout the averaging window. The filter produces a motion-blur artifact on moving regions, but provides the intended temporal-accumulation comparison on static or invariant pixels.

S-1.3 Layout

Each video frame is arranged into eight panels, as illustrated in Fig. S1. The top row presents the four evaluated temporal models: PerPixel, TZ, STZ, and ReShuffle. The bottom row presents four benchmark panels:

Reference: a sufficiently converged frame rendered at 16k spp per frame.
No TAA: raw frames rendered at the specified per-frame sampling rate without temporal averaging.
Ideal: each individual frame rendered with the tested sampler at the aggregate sampling rate of the 16-frame window, without temporal averaging, serving as the best hoped-for temporal-accumulation benchmark.
Ideal ZSS: similar to ideal, but rendered using the classic subspace Z sampler (ZSS), intended primarily as a comparison reference for ReShuffle.

Four color-coded insets magnify regions exhibiting different forms of sampling error. We selected the four inset regions to favor PerPixel, TZ, STZ, and ReShuffle, respectively, in reading order, although this correspondence cannot be guaranteed throughout the video. Indeed, rendering behavior is highly dynamic and varies substantially across scene poses and configurations.

S-1.4 Progression

Beyond its spatial composition, the video progresses through nine configuration segments over a half-turn of the coffeemaker, covering:

- Five Z-indexed samplers: (1) the classic subspace Z sampler (ZSS), adapted to use O2m3 for subsurface sampling; (2) the native high-dimensional SZ sequence sampler [Ahmed et al. 2025]; (3) StART and (4) O2 from the recently introduced NILE framework [Ahmed et al. 2026]; and (5) an embedding of O2m3 under SZ.
- Two importance-sampling approaches: the native *pbrt*-4 `Infinite` sampler and our ZEnv approach.
- Four per-frame sampling rates: 16, 32, 64, and 128 spp.

The central text bar (Fig. S1) identifies the configuration active in each displayed frame; whenever a setting changes, its corresponding field label is briefly highlighted.

Although the video employs all three Z++ extensions introduced in this paper, it focuses primarily on the temporal component. The other two are included to demonstrate the compatibility and architectural orthogonality of the three components, rather than to provide a dedicated evaluation of either non-temporal extension. Indeed, in this particular scene, ZEnv slightly underperforms the native *pbrt*-4 `Infinite` sampler, whereas O2m3 substantially outperforms O2, which effectively embeds a three-dimensional Sobol sequence.

S-1.5 Analysis

Although the video is primarily intended as a visual demonstration, the 900 rendered frames in each sampler sequence form a modest data set from which we distilled several insights into the behavior of the temporal models. We developed a Python script that emulates the temporal moving average implemented using FFmpeg's `tmix` filter, computes the root-mean-square error (RMSE) of the frame displayed in each panel against the corresponding reference frame, and generates several tabulated and plotted summaries, as illustrated in Fig. S2.

We first plot the raw full-frame RMSE measurements across all rendered frames in Fig. S2(A). Over frames 0–15, the errors of the four temporally averaged models rapidly drop from the NoTAA level toward the corresponding Ideal level as the accumulation window fills. This transition can also be observed directly in the video by slowly stepping through these frames.

A subtle distinction is already visible at frame 0: the PerPixel and TZ frames are identical to NoTAA, whereas the STZ and ReShuffle frames are not. Indeed, PerPixel and TZ preserve individual-frame quality while seeking negative correlation between frames. At the other extreme, ReShuffle also preserves individual-frame quality but enforces independence between frames. STZ lies between these extremes, compromising individual-frame quality to create a mixture of independent and negatively correlated per-pixel sequences across frames.

Despite their initial descent toward Ideal, the full-frame RMSE curves subsequently become dominated by the unintended motion-blur artifact caused by temporally averaging the moving coffeemaker, as reflected in their gradual upward trend. A motion-aware real-time TAA system would mitigate this effect by reprojecting samples using renderer-provided motion vectors. In our simplified demonstration, we instead circumvent this issue by restricting the analysis to the right half of the scene, which contains no moving objects. As shown in Fig. S2(B), the RMSE curves then become nearly flat and approximately parallel to their corresponding Ideal curves.

Next, we omit NoTAA and the first 15 partially accumulated frames to obtain a closer view of the four models together with Ideal ZSS. In Fig. S2(C), we further normalize each curve by the RMSE of the corresponding Ideal frame rendered using the same tested sampler. This normalization largely factors out residual frame-to-frame variation in rendering difficulty, including changes caused by reflections of the moving coffeemaker within the stationary half of the scene. The curves become clearly distinguishable, with ReShuffle exhibiting the largest relative error, as expected, while the ranking of the remaining models varies with the tested sampler and scene pose.

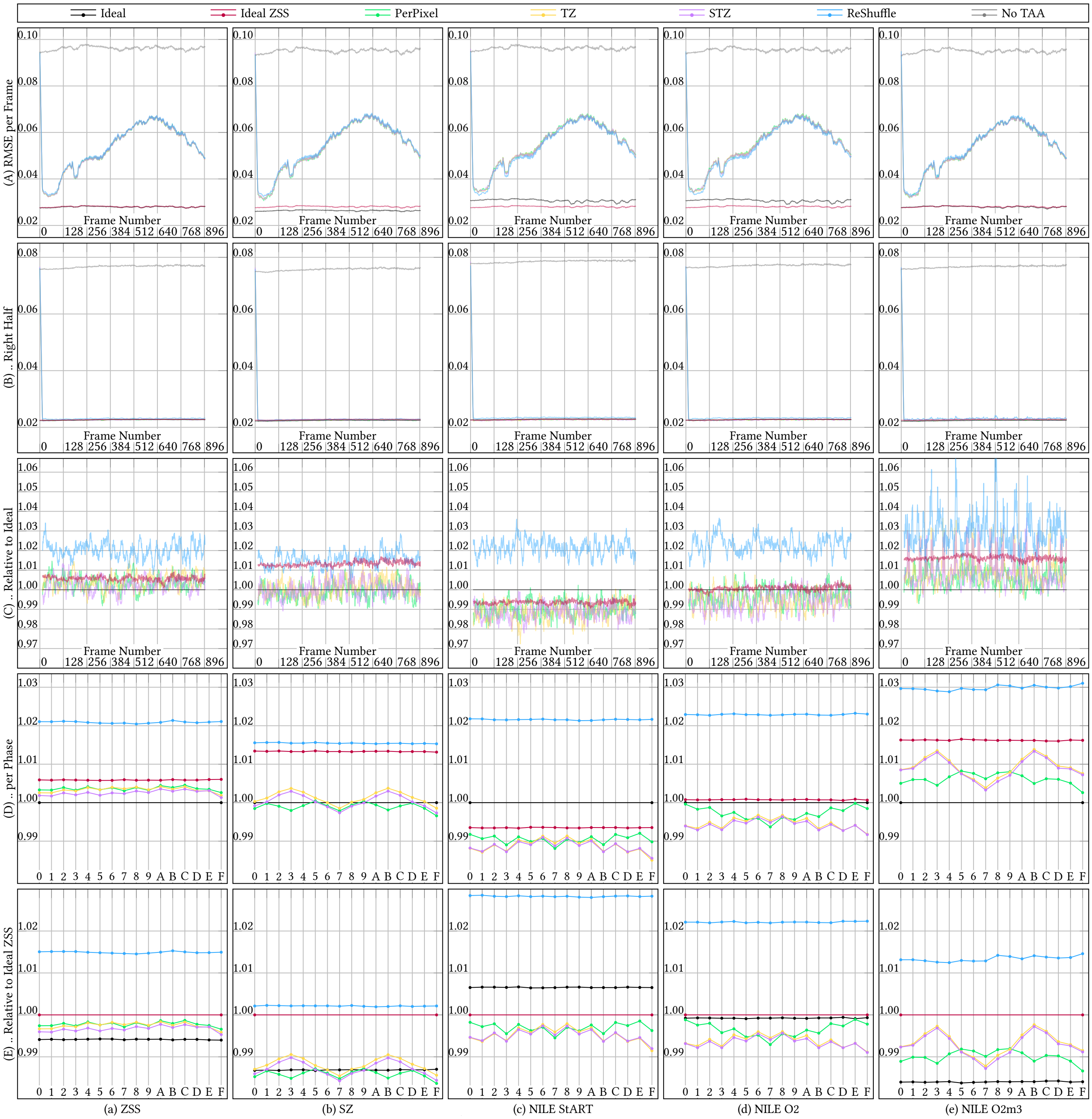


Fig. S2. RMSE data for the rendered frames across five different samplers (columns) over a 900-frame sequence covering a half-turn of the coffeemaker. (A) Raw full-frame RMSE. (B) RMSE restricted to the stationary right half (RH) of the scene. (C) Relative right-half RMSE, normalized by the corresponding *Ideal* RMSE; NoTAA and the first 15 frames are omitted from this and subsequent plots to better resolve the remaining closely spaced curves. (D) The same relative RMSE averaged by frame phase within the 16-frame accumulation window. (E) The corresponding phase-wise analysis relative to *Ideal ZSS*, providing a common baseline across the tested samplers. Values below one in (C)–(E) indicate lower error than the respective baseline.

Although the curves appear noisy, careful inspection of the three negative-correlation models reveals a recurring pattern across 16-frame cycles. We expose this behavior by grouping the full-window frames according to their phase within each cycle and averaging the normalized RMSE within each group, as shown in Fig. S2(D). The resulting phase profiles are smooth and readily comparable. TZ and STZ are shifted by approximately half a cycle relative to PerPixel, consistent with their half-pixel-block exchange construction. We exploit this cyclic cancellation in the running RMSE readouts displayed in each video panel. For both the full image and its right half, each readout is a 16-frame moving average of normalized RMSE and therefore contains one observation from every phase, thereby suppressing phase-dependent fluctuations.

Finally, because Ideal ZSS is invariant across the five sampler configurations in this experiment, we use it as a common baseline in Fig. S2(E). Normalizing all curves by the corresponding Ideal ZSS RMSE places the five plots on a common scale and aligns their shared baseline at one. In column (a), Ideal differs slightly from Ideal ZSS because incorporating temporal support into ZSS required revising its index shuffling, which incidentally reduced the RMSE of the resulting Ideal frames in this experiment. The StART and O2 configurations are comparatively weak in this example; more strikingly, their negative-correlation models actually outperform their respective Ideal baselines. This initially suspicious result is readily explained by inspecting the corresponding frames in the video. For this particular scene and sampling configuration, these samplers exhibit substantial pixelation not only at the designated rate of 16 spp, but also in their Ideal frames rendered at 256 spp. The perturbations introduced by the negative-correlation models slightly weaken this structured aliasing, allowing them to outperform Ideal, with STZ achieving the lowest error among the three.

This shared aliasing behavior suggests strong correlations among dimensions of the underlying orchestrating SZ sequence, which are consequently inherited by different nested NILE samplers. O2m3 is substantially less affected despite sharing the same nested structure [Ahmed et al. 2026]. One plausible explanation is that O2m3 makes better use of the dimensions of the orchestrating SZ sequence by nesting not two but three subspace dimensions within each global dimension, thereby keeping three path bounces within the first compact 16D subspace of SZ. Notably, native SZ itself outperforms the nested variants, although it still appears to suffer from the same phenomenon. Indeed, the tight clustering of its temporal models around the corresponding Ideal might initially suggest strong performance. Comparison with O2m3, however, indicates that the SZ Ideal is itself suboptimal; this apparent agreement therefore reflects the weakness of the target rather than the strength of the temporal models. The scene is particularly revealing in this respect because its multiple glossy reflections exercise several consecutive sampling dimensions along each path, allowing correlations among them to manifest as coherent image-space aliasing.

When the targeted Ideal is itself pixelated, the three negative-correlation models offer only limited protection, and this is where ReShuffle excels. Our tests so far indicate that reshuffling can successfully disrupt coherent sampler correlations even when both the individual frames and the aggregate-rate target are pixelated. High-dimensional low-discrepancy samplers can all be vulnerable to such aliasing at very low sampling rates, with no simple general rule predicting which sampler and scene configuration will produce a visible pattern. Thus, although ReShuffle may exhibit the largest relative RMSE, it is arguably the safest model against coherent aliasing under severe undersampling. In this demonstration, for example, reaching the sampling rate of the converged reference would require increasing the per-frame sample count by a factor of 1024.

After the sampler configurations, all of which use the native *pbrt*-4 `Infinite` importance sampler, the video switches to ZEnv and then progresses through four per-frame sampling rates. Fig. S3 presents the corresponding plots, first comparing `Infinite` and ZEnv at 16 spp, and then increasing the ZEnv rate to 32, 64, and 128 spp.

For this scene, switching to ZEnv slightly degrades the Ideal, increasing its mean RMSE by 0.24% and 0.39% over the full frame and the stationary right half, respectively. More importantly, ZEnv notably amplifies the phase-dependent variations, revealing an interaction between the importance and temporal constructions that warrants further investigation. Increasing the sampling rate reveals further complexity in the phase behavior, with visibly different patterns between even and odd dyadic exponents. We avoid drawing conclusions from this limited evidence, since the rendering pipeline combines several entangled base-4-aware components, including SZ, O2m3, and ZEnv. Furthermore, as sampling error decreases, dynamic lighting effects become relatively more prominent in the stationary right half, making it a less reliable basis for this analysis. We therefore leave a systematic investigation to future work.

S-1.6 Bundled Content

To facilitate reproduction and further experimentation, we provide the scene model and all scripts used to render, analyze, and assemble the video, enabling readers to adjust the experimental parameters or adapt the workflow to their own needs. The package contains the following items:

`zcoffeeteapot.zip`: The animated scene used throughout the video.

`zframes.sh`: Renders the video frames under the selected configurations.

`zrmse.py`: Analyzes and plots the RMSE of the rendered frames.

`zspatial.sh`: Composes the eight-panel comparison for each configuration.

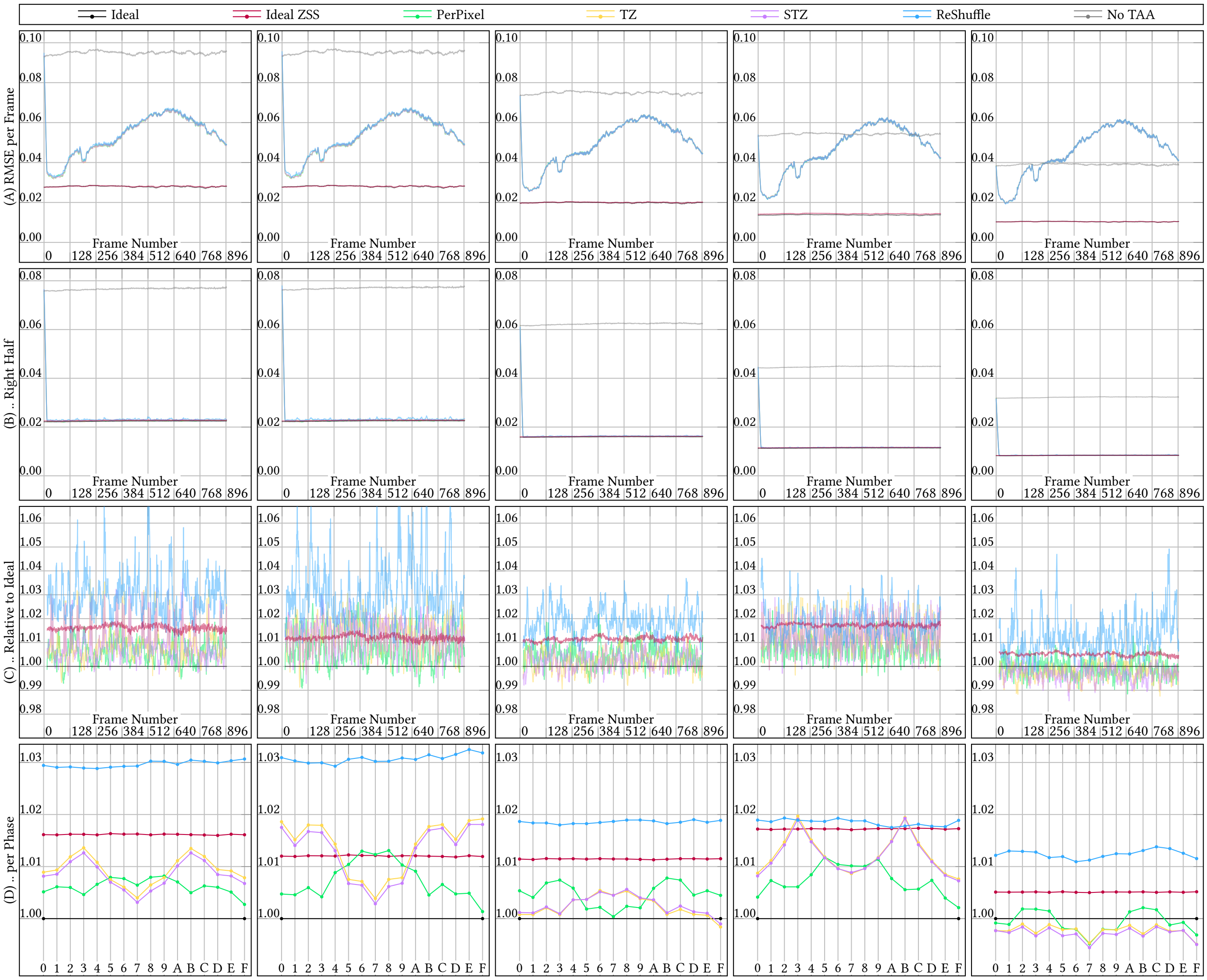
`ztemporal.sh`: Combines the configurations into the final video progression.

REFERENCES

Abdalla G. M. Ahmed, Matt Pharr, Victor Ostromoukhov, and Hui Huang. 2025. SZ Sequences: Binary-Constructed $(0, 2^q)$-Sequences. *ACM Trans. Graph.* 44, 6, Article 206 (Dec. 2025). https://doi.org/10.1145/3763272

Abdalla G. M. Ahmed, Matt Pharr, Victor Ostromoukhov, and Hui Huang. 2026. NILE: Nested Interleaving of Low-Dimensional Elements. *ACM Trans. Graph.* 45, 4, Article 105 (July 2026). https://doi.org/10.1145/3811283

Benedikt Bitterli. 2016. Rendering Resources. https://benedikt-bitterli.me/resources/ Accessed: 2025-11-08.



(a) O2m3, `Infinite`, 16 spp (b) O2m3, ZEnv, 16 spp (c) O2m3, ZEnv, 32 spp (d) O2m3, ZEnv, 64 spp (e) O2m3, ZEnv, 128 spp

Fig. S3. RMSE plots for the rendered 900 frames across two importance samplers and four sampling rates (columns). The rows display similar successive refinements to Fig. S2.