Optimizing Photon-to-Photon Latency in MR Headset Video See-Through : Design Guidelines and Tuning Strategies

Wangzan Jin*, Longyun Xiao*, Qining Wang**, Liang Gu*, Qi Gu*, Fan Gong**, Lei Zhao**

***GravityXR Electronics and Technology Co., Ltd., Ningbo, Zhejiang, China**

****Yongjiang Laboratory, Ningbo, Zhejiang, China. Corresponding author: Lei Zhao**

Abstract

In the application of mixed reality (MR) headsets where virtual and real environments are integrated, the Photon-to-Photon (PTP) latency in the Video See-Through (VST) segment is inevitable. This latency causes sensory discomfort and motion mismatch, reducing the immersive experience during use. Therefore, achieving stable and low PTP latency is a critical challenge in the development of MR headsets. This study is based on the implementation of VST technology and analyzes the impacts of frequency and phase changes in camera shooting, image processing, and image rendering on PTP latency through simulation. The findings are validated by empirical measurements on popular MR headsets such as Apple Vision Pro, Pico4, and Meta Quest3. The results indicate that to optimize the latency performance of VST, it is essential to ensure the consistency of frequency and stability of phases across different processing stages. Furthermore, this paper proposes a set of monitoring and adaptive tuning methods and demonstrates the process of tuning PTP latency through case studies.

Author Keywords

Photon-to-photon Latency; Video-see-through; Mixed Reality; Optimization Methods

1. Introduction

As mixed reality (MR) technology advances, the interactive fusion of real and virtual environments has captivated increasing interest. Most current MR headsets employ Video See-Through (VST) display technology. In this setup, cameras capture real-world scenes within the user's field of view. These video streams are then seamlessly integrated with computer-generated virtual imagery. The resulting composite images are displayed on the MR headset's screen, facilitating an immersive mixed reality experience. (1-3)

To mitigate motion feedback lag resulting from delays, mixed reality devices necessitate robust low-latency capabilities to ensure an effective mixed reality experience. (4) (5)The permissible limit for image transmission delay is determined by the system's response rate, which is crucial for maintaining stable gaze. According to human perception assessments, despite complexity in visual content and individual differences in cognitive processing, response times vary from 13 to 80 ms. The perception threshold for delays can differ by several tens of milliseconds among different subjects. (6, 7). Ideally, latency should be low enough so that users are not able to perceive scene motion.It is widely regarded that an image delay under 20 ms does not evoke any abnormal sensations or discomfort.(8-10)

In mixed reality (MR) applications, virtual imagery produced by computer systems can be rendered and presented in designated timeframes and spatial contexts through real-time tracking of user posture and algorithms that anticipate future movements, thereby reducing latency. A critical metric governing the mixed reality experience is the "Photon-to-Photon" (PTP) latency within the VST functionality of MR technology, which is the interval needed for an actual event to be displayed on the MR headset's screen(11, 12), as illustrated in Figure 1.

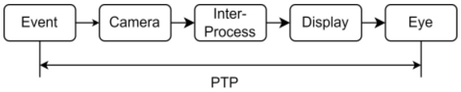


Figure 1. The Correspondence Between PTP and VST

Presently, established Photon-to-Photon (PTP) measurement methods are employed to evaluate the VST latency performance of MR headsets. Test data reveals that the PTP latency for mainstream MR headsets usually lies between 30 to 60 ms, with high-end models reporting an average latency of 14.7 ms and a minimum recorded PTP latency of 9.43 ms. (11)

However, while existing measurement techniques are capable of assessing VST delay performance, they fall short in pinpointing the exact sources of PTP discrepancies. A more detailed examination of the relationship between VST components and latency outcomes, coupled with the development of more precise PTP measurement and optimization techniques, is crucial for improving the MR user experience.

2. Internal Link Decomposition Analysis of VST

To better understand VST delay, both end-to-end system delay and sub-component delays should be measured. (13)The PTP distribution in MR headsets is strongly influenced by the design logic of their VST latency performance and their operational conditions. Therefore, each VST latency link must be analyzed step by step. The camera uses a Rolling Shutter mechanism to capture and display the scene, with its exposure-to-readout sequence shown in Figure 2.

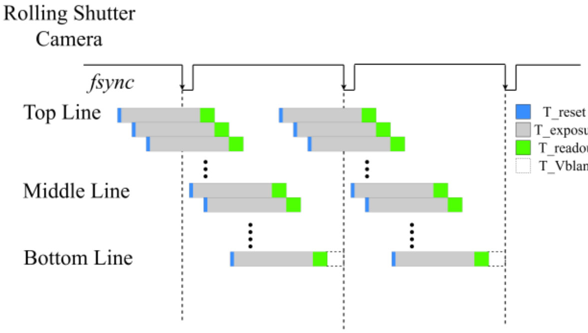


Figure 2. Rolling Shutter Camera Timing Diagram

It is controlled by the clock source f_{sync}, where each row on the sensor is sequentially exposed and read out within the duration of a single frame. Taking the middle line of the sensor as a case study, this extracted process is illustrated in the schematic diagram presented in Figure 3.

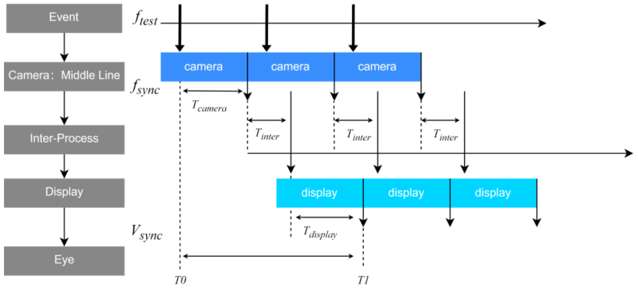


Figure 3.VST Link: From Camera Middle Line to Display

This is a three-clock source system, where the test event is triggered by f_{test}. The camera operates at a frequency of f_{sync}, with a phase φ_{camera} relative to f_{test}. After exposure, the real scene is captured and read out according to f_{sync}. After undergoing intermediate processes such as ISP processing, graphic rendering, and other pre-display processing, the display retrieves and exhibits the screen imagery at a specified frequency V_{sync}, with a phase $\varphi_{display}$ relative to f_{test}. This system involves the coupling of three clock sources, which ultimately affects the final PTP results.

For test events with trigger counts of 1, 2, ..., N, the following can be obtained:

$$T_{test} = \frac{1}{f_{test}} \times N \tag{1}$$

$$T_{camera} = \frac{1}{f_{sync}} - \text{mod}(T_{test} - \varphi_{camera}, \frac{1}{f_{sync}}) \tag{2}$$

$$T_{display} = \frac{1}{v_{sync}} - \text{mod}(T_{test} + T_{camera} + T_{\text{inter}} - \varphi_{display}, \frac{1}{V_{sync}}) \tag{3}$$

$$PTP_N = T_{camera} + T_{\text{inter}} + T_{display} \tag{4}$$

3. Simulation of Principles

A. Measured PTP Standard Definition

PTP Definition and Testing Method: The measurement starts with a real-world triggering event, recorded as the initial time t_0. When the MR headset's camera captures this event, after passing through the intermediate links, the time t_1 is noted. The calculated PTP is given by $PTP\ Latency = \text{t}_1\text{-t}_0$.

The device used for testing PTP is illustrated in Figure 4. This testing apparatus is designed to achieve a measurement accuracy of 0.08 ms(11), ensuring precise evaluation of the PTP latency in various scenarios.

Figure 4.PTP Latency Measurement Device

B Internal Link Decomposition Analysis

A simulation system is established based on the decomposed VST link, with the system parameters shown in Table 1.

Table 1.VST Link Simulation System Parameters

Parameter	f_{test}	f_{sync}	V_{sync}	T_{inter}	φ_{camera}	$\varphi_{display}$
Unit	HZ	HZ	HZ	ms	ms	ms

1) Consistency of Frequency Within the MR Headset

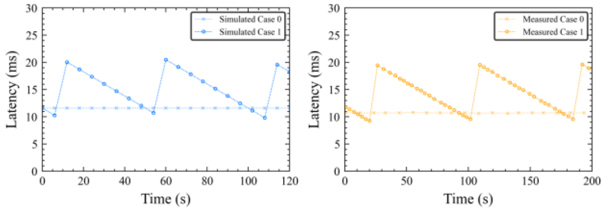
In an ideal scenario, the frequencies of the camera capturing the real scene, processing, and displaying in the VST link of the MR headset #1 are consistent, with a constant phase. For the VST internal frequency-consistent MR headset #1, its $f_{syvc} = \text{V}_{sync}$ =90 Hz, while T_{inter}, φ_{camera} and $\varphi_{display}$ remain constant. By setting different test event frequencies, we obtain the distribution of its PTP. Additionally,a simulation model is used for validation. The parameters are set as shown in Table 2.

Table 2. Test Event and MR headset Simulation Parameters: $f_{sync} = V_{sync}$

Case	f_{test}	f_{sync}	V_{sync}	T_{inter}	φ_{camera}	$\varphi_{display}$
0	0.5	90	90	0	1	0.5
1	0.499722	90	90	0	2	0.5

In case 0, the frequency of the test event is set to a common divisor of the device frequency , with f_{test} set to 0.5 Hz. At this point, each test event corresponds to a fixed position T_{camera} within one frame of the VST camera. In case 1, the frequency of the test event is set to a common divisor of the device frequency with a slight beat frequency,, with f_{test} set to (90 - 0.05) / 180 = 0.499722 Hz. In this case, each test event will move within one frame of the VST camera.

A simulation model is established, and after obtaining the simulation results, the measured latency distribution aligns with the simulation results, as shown in Figure 5.



(a)Simulated PTP (b)Measured PTP

Figure 5. Test Event and MR headset Parameters

case0 : $f_{\text{syvc}} = V_{\text{sync}} = m \times f_{\text{test}}$ case1 : $f_{\text{syvc}} = V_{\text{sync}} \neq m \times f_{\text{test}}$

2)Frequency mismatch within the MR headset.

When there is a frequency mismatch between the camera and the display in the VST link, parameters f_{sync} and V_{sync} have a frequency difference, $f_{sync} \neq \text{V}_{sync}$. A simulation analysis is conducted to examine the impact of this frequency mismatch on PTP within the MR headset's VST link. The parameters are set as shown in Table 3:

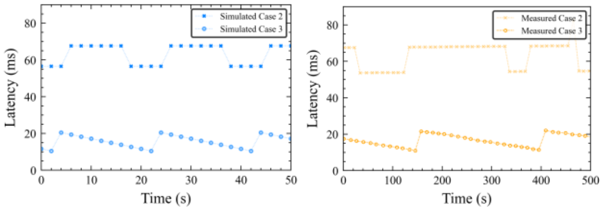
Table 3. Test Event and MR headset Simulation Parameters: $f_{sync} \neq V_{sync}$

Case	f_{test}	f_{sync}	V_{sync}	T_{inter}	φ_{camera}	$\varphi_{display}$
2	0.5	89.95	90	52	2	0.5
3	0.5	90	90.05	3	2	0.5

A simulation model is established, and after obtaining the simulation results, the measured latency distribution aligns with the simulation results, as shown in Figure 6.

To validate the simulation results of case 2, a MR headset #2 with an unstable VST link is used. During testing, the frequency of the test light source is controlled to ensure that the display frequency of the MR headset meets $V_{sync} = m \times f_{test}$, where m is an integer. During the testing process, a slight frequency difference exists between f_{syvc} and V_{sync}, and the measured PTP in this case is consistent with the simulation scenario of case 2.

For validating the simulation results of case 3, the camera's parameter f_{syvc} of the MR headset #3 is introduced. The frequency of the test light source f_{test} is controlled to ensure that the display frequency of the MR headset meets $f_{sync} = m \times f_{test}$, where m is an integer. During the testing process, a slight frequency difference exists between f_{syvc} and V_{sync}, and the measured PTP in this case is consistent with the simulation scenario of case 3.



(a)Simulated PTP (b)Measured PTP

Figure 6. MR headset VST Link Frequency Mismatch

3) Other Frequency Mismatch Scenarios in MR Headsets

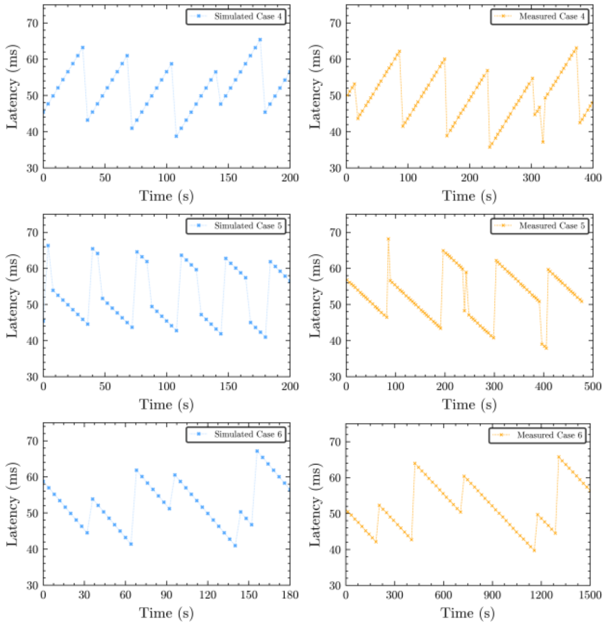
In the previous analysis, we simulated MR headsets with frequency mismatches and aligned the frequency of the test events with the frequency of the camera or display in the link. But what happens when all three clock sources in the system have frequency mismatches? The simulation parameters are shown in Table 4. Under the condition of constant phase, simulations are conducted for the following cases of frequency mismatch.

Table 4. Test Event and MR headset Simulation Parameters: $f_{sync} \neq V_{sync} \neq f_{test}$

Case	f_{test}	f_{sync}	V_{sync}	T_{inter}	φ_{camera}	$\varphi_{display}$
4	0.500277	50	90	37	1	1
5	0.499722	50	89.98	37	1	1
6	0.499722	50	90.02	37	1	1

While obtaining the simulation results, the actual PTP measurements from the MR headsets were also correlated, as shown in Figure 7. The selected MR headset #4 has a significant frequency difference between the camera frequency and the display frequency from the design stage. Additionally, during operation, the frequencies and phases of the clock sources may experience slight drifts.

In the tests, we accumulated a substantial amount of PTP test results under conditions of complete frequency mismatch, which align well with the simulation results. The various forms of PTP variations observed indicate that if the latency performance of VST is not carefully considered during the design phase and if the design guidelines are not followed, the MR headset will not provide users with a satisfactory mixed reality experience.



(a)Simulated PTP (b)Measured PTP

Figure 7. PTP Performance Under Complete Frequency Mismatch (Left: Simulation, Right: Measurement) Case4-6： $f_{sync} \neq V_{sync} \neq m \times f_{test}$

4. VST Overall Debugging

At present, numerous MR headsets implement Monado, an open-source runtime tailored to facilitate the functioning of virtual reality (VR) and augmented reality (AR) technologies. In the VST link, Monado collects data from various sensors, processes the images, and submits the VST frames to the synthesizer. After the synthesis and rendering are complete, it waits for the display. Several moments affect the final time T_{inter} :

1.The time from Camera to entering image algorithm processing.

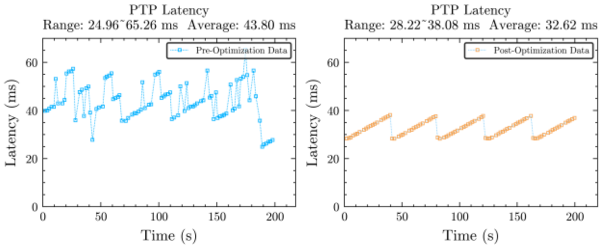
2.The time taken to submit the VST frame to the synthesizer after ISP processing is complete.

3. The time from the completion of synthesis rendering to the presentation at V_{sync}.

By optimizing the characteristic time points of the key processes mentioned above, we can ensure the stability of the VST functionality under different loads and environments, effectively avoiding issues such as frame loss and abnormal delays.

An optimization example is provided for an MR headset with a design frequency of 90 Hz. In its VST link, monitoring functions are set up, allowing for the analysis of Trace data after VST operation. This analysis yields the frame time (i.e., frequency and phase stability) for each clock source per frame, as well as the phase information at key time points within the T_{inter} .

The expected PTP results are as follows: T_{camera} varies according to the frequency of the test event, with a range of $(0, \frac{1}{f_{sync}})$ =0~11.1ms; $T_{display}$ is a fixed value of 5.5ms; and T_{inter} is 23 ms. The total expected PTP is 28.5~39.6ms.The PTP testing result before optimization are shown in Figure 8 (a). At this point, the result does not meet the expectations.



(a) Pre-Optimization (b) Post-Optimization

Figure 8.PTP Test Results Before and After Optimization

By adjusting the parameters of each moments through the tuning system, ensuring that all component times are within ideal values, a subsequent test yields the PTP distribution shown in Figure 8 (b), which now aligns with expectations.

5. Conclusions

This paper combines simulations of the VST link with measured PTP results from typical MR headsets to propose guidelines for the design of VST latency functions in MR headsets. Measurements have shown that currently, average-performing MR headsets have the T_{inter} exceeding 30ms, and there is a frequency mismatch between the camera and display, making it difficult to meet the demands for a satisfactory mixed reality experience. During the design and performance tuning of MR headsets, developers should adhere to the VST latency performance design guidelines proposed in this paper and utilize an adaptive monitoring system for tuning.

The design guidelines are as follows:

1.Stability Requirement: The design frequencies of the camera f_{syvc} and display V_{sync} must be consistent, and there should be no frame fluctuations during operation that would cause phase changes.

2.As the formula (4), to keep the PTP latency within the acceptable threshold for humans (20ms), the frequencies of the camera and display need to be greater than 1/(20- T_{inter}) Hz, considering that the best commercial MR headsets currently have the T_{inter} around 9ms, The minimum allowable frequency for MR headsets can be calculated as 1/(20-9)ms = 90Hz under ideal conditions. At this point, the range of PTP variation for the device would be 9-20ms. If the T_{inter} increases, the frequencies of the camera and display must be further increased to ensure PTP < 20ms.

6. References

1. Ernst JM, Laudien T, Schmerwitz S, editors. Implementation of a mixed-reality flight simulator: blending real and virtual with a video-see-through head-mounted display. Defense + Commercial Sensing; 2023. https://doi.org/10.1117/12.2664848
2. Rolland J, Holloway R, Fuchs H. Comparison of optical and video see-through, head-mounted displays. Proceedings of SPIE - The International Society for Optical Engineering. 1994. https://doi.org/10.1117/12.197322
3. Rolland JP, Fuchs H. Optical Versus Video See-Through Head-Mounted Displays in Medical Visualization. Presence Teleoperators & Virtual Environments. 2000;9(3):287-309. https://doi.org/10.1162/105474600566808
4. Elbamby MS, Perfecto C, Bennis M, Doppler K. Towards Low-Latency and Ultra-Reliable Virtual Reality. IEEE Network. 2018;32(2):78-84. https://doi.org/10.1109/MNET.2018.1700268
5. Jerald JJ. Scene-motion-and latency-perception thresholds for head-mounted displays: The University of North Carolina at Chapel Hill; 2009.
6. Potter MC, Wyble B, Hagmann CE, Mccourt ES. Detecting meaning in RSVP at 13 ms per picture. Attention Perception & Psychophysics. 2014;76(2):270-9. https://doi.org/10.3758/s13414-013-0605-z
7. Mania K, Adelstein BD, Ellis SR, Hill MI, editors. Perceptual sensitivity to head tracking latency in virtual environments with varying degrees of scene complexity. Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization; 2004. https://doi.org/10.1145/1012551.1012559.
8. Berthoz A, Weiss G, Sherwood DE. The Brain's Sense of Movement. Yale Journal of Biology & Medicine. 2003;76(4-6):193-4. https://doi.org/10.1016/S0031-9384(00)00415-7
9. Palmisano S, Allison RS, Teixeira J, Kim J. Differences in virtual and physical head orientation predict sickness during active head-mounted display-based virtual reality. Virtual Reality. 2023;27(2):1293-313. https://doi.org/10.1007/s10055-022-00732-5
10. Yang M, Zhang J, Yu L, editors. Perceptual Tolerance to Motion-To-Photon Latency with Head Movement in Virtual Reality. 2019 Picture Coding Symposium (PCS); 2019 12-15 Nov. 2019. https://doi.org/10.1109/PCS48520.2019.8954518
11. Xiao L, Jin W, Wang Q, Zhao L. 46-5: Photon-to-Photon Latency Test Solution to Video See-Through of Mixed Reality Headset. SID Symposium Digest of Technical Papers. 2024;55. https://doi.org/10.1002/sdtp.17099
12. Gruen R, Ofek E, Steed A, Gal R, Sinclair M, Gonzalez-Franco M, editors. Measuring System Visual Latency through Cognitive Latency on Video See-Through AR devices. 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR); 2020 22-26 March 2020. https://doi.org/10.1109/VR46266.2020.00103
13. Taylor R. OSVR: Sensics Latency-Testing Hardware. 2015.