Call for Proposals: VDC-X

1 Introduction

This document is a call for proposals (CfP) to standardize the algorithm for VESA Display Stream Compression for AR/VR/MR (VDC-X). The document includes requirements that apply to solutions, evaluation criteria, and required items to be submitted. VESA's goal is to standardize a visually lossless coding system to be used for compression of high-bandwidth content in AR/VR.

2 Purpose and Requirements

2.1 Purpose

VDC-X intends to meet these objectives:

- Support AR and VR stereoscopic application-specific requirements with latencies lower than could be supported by a display link with uncompressed, DSC, or VDC-M compressed images and video in real time.
- Reduce the bandwidth requirement compared with uncompressed in the display interface needed to transport a display stream to save power, cost, or both.
  - For example, a higher resolution of the display stream could be supported, or more streams in a multi-stream configuration as required by the system design.
- Enable the use of a lower link rate for applications where high link rates might not be possible.

VDC-X will have the following properties:

- On-the-fly, real-time coding with low complexity hardware and memory in both the encoder and the decoder.
- Encoder and decoder might have different complexity and process node (encoder complexity might be higher than decoder).
- Algorithm is transport agnostic and coded bitstreams could be transported over display links.

2.2 General Requirements

Proposals of a coding system shall include the attributes listed in the following tables:

Table 1: Input

Factors	Description	Agreed Consensus	
		AR	**VR/MR**
Typical Width per eye	Number of possible horizontal sample locations	2048	4096
Typical Height per eye	Number of possible vertical sample locations	2048	4096
Bits/sample	Bit Depth (number of bits per sample)	8, 10	8, 10, 12
Slices	Independently coded set of samples comprising a rectangle in the horizontal and vertical directions (See 2.4.5 Slice Support)	yes	yes
Frame Rate	Number of Frames (updates to all color channels) per second	<=120	<=180
Field Rate	Number of color fields updated per second	<=4x120	<=4x120
Gamma	Non-linear EOTF (Electro-Optical Transfer Function)	sRGB	sRGB, HDR(PQ) with a typical max of 1000 nits
Color Space	Color primaries	Disp. Native	Rec709 (SDR), P3D65 or Rec2020 (HDR)
Maximum PPD	Maximum Pixels per Degree	40	70
Minimum PPD	Minimum Pixels per Degree	10	20
Adjustable Sample Density (ASD)	Method to change the ratio of source samples to display pixels in different regions. (See 2.2.1 Foveation)	Optional*	Yes
Typical ASD factor		{1,2,4}x{1,2,4}	{1,2,4}x{1,2,4}
Maximum ASD factor		16x16	16x16
Chromatic Aberration Correction (CAC)	Pre-applied Chromatic Aberration Correction	Optional*	Optional*
Mura Pre-Compensation	Pre-applied Mura compensation (See 3.1 Test Material)	Optional*	Optional*
Lens Distortion Correction (LDC)	Pre-applied LDC	Yes	Yes
Regional Update	Update only a subset of the display's retained framebuffer, if supported. (See 2.4.5 Slice Support)	Optional*	Optional*

Raster	Inputs arrive in raster scan order	Yes	Yes
Pre-Foveated Raster	Input buffer raster has foveated pre-processing already applied.	Optional*	Optional*
Tiled Render	Inputs arrive via a sparse tile interface (See 2.2.3 Tiled Rendering)	Optional*	Optional*
Slice-synchronized Metadata	Slice-synchronous metadata included in the coded bitstream (uncompressed)	32KB/frame	32KB/frame

*Optional: The standard supports the feature, but it is optional to use in the application.

Table 2: Latency

Factors	**Description**	**Agreed Consensus**	
		AR	VR/MR
End-to-End Latency	Target time between new action and corresponding change to display. (Encoder to decoder latency plus system latency)	4 msec	8 msec

Table 3: Display

Factors	**Description**	**Agreed Consensus**	
		AR	**VR/MR**
FSD	Field Sequential Display (R, G, and B fields sent as planes)	Yes	Optional*
RGB	Interleaved RGB values	Yes	Yes
Rolling/Global illumination	Illumination time interval either depends on or is independent of vertical position	Optional*	Optional*
Fast Skip Blocks	The process in which only coded samples are written to display memory. Uncoded regions could either retain previous values as in partial update or be filled with zeros (auto-cleared) either by a fast-zero-write process or by flagging skipped display memory regions as "treat as zero" which could also allow bypassing pixel reading at display illumination time. A slice contains one or more fast skip blocks. (See 2.2.3 Tiled Rendering)	Optional*	Optional*

Fast Scaling	Pixel replication process which allows foveated samples to be written to display memory pixels at a faster speed than full resolution samples (See 2.2.4 Display Memory Organization for Fast Scaling)	Optional*	Optional*
Scaling method	Interpolation or replication	Replication	Replication
Scaling factors	what x, y scaling factors are supported	{1,2} x {1,2}	{1,2,4} x {1,2,4}
Checkerboard sampling	Pixel subsampling mode (sometimes also called quincunx) by which only the matching row and column parity samples are used. To avoid aliasing, a simple filter can be applied prior to subsampling. (See 2.2.2 Checkerboard)	Optional*	Optional*

*Optional: The standard supports the feature, but it is optional to use in the application.

Table 4: Compression

		Agreed Consensus	
Factors	**Description**	**AR**	**VR/MR**
Lossless	Opportunistically mathematically lossless (reversible)	Yes	Yes
Perceptually Lossless	aka visually lossless, artifacts not visible	Yes	Yes
Foveation	Control parameters and exact designation of how the input samples are derived from the hypothetical full-res source image and how they should be scaled upon write to the display pixels.	Yes	Yes
Sparseness	Signaling mechanism to identify which input pixel regions are coded and written to display memory, and which uncoded regions should not be written (skipped) or implicitly written with zeros.	Yes	Optional*
Compressed bit rate target	pixel rate in bits per pixel	(12 bpp) X 1/Fill Factor	(8 bpp) X 1/Foveation Factor
Buffer Model	Real-time behavior avoids display overflow. (See 2.4.3 Buffer Model)	CBR, UB-VBR	CBR, UB-VBR

*Optional: The standard supports the feature, but it is optional to use in the application.

2.2.1 Foveation

Human vision is most acute in the center of the visual field called the fovea. At larger eccentricity (angular distance from the fovea), humans have reduced spatial acuity. Since AR and VR displays have a fixed display resolution pattern, it is beneficial to reduce the visual resolution sent to different regions of the display. This process depends on many system level aspects outside of the scope of this document, such as:

- Eye tracking
- Gaze prediction
- Reduced resolution rendering

The ratio between the display's sample density (in units of pixels per degree) and the rendered visual sample density (also in units of pixels per degree) is called the Adjustable Sampling Density (ASD) and can be quantified in both X and Y directions. The foveal region will typically have an ASD of 1x1 (full display resolution). Peripheral regions might have, for example, ASD factors of 2x2 or 4x4. The determination of how much to foveate and where to do so impacts the effective compression ratio. Specification and evaluation of perceptual visually lossless foveation requires a formulation of how to compute the ASD across the whole display visually, apply this to input images, transmit this as part of the compressed data, and how to upsample the results to fill in the display.

If the fast scaling capability is lower than the ASD factor, this must be accounted for in the HRD model to prevent bits from overflowing the decoder buffer.

The angular sampling could be handled in different ways. For example:

- The angular sampling could be computed by a rendering system directly.
- The angular sampling could be computed by a display processor based on a full resolution rendered frame.

2.2.2 Checkerboard Sampling

A checkerboard subsampling is sometimes used to provide a further reduction in visual samples. This mechanism reduces the number of samples in a given region by an additional factor of 2. Checkerboard regions will always have an even number of samples horizontally and vertically.
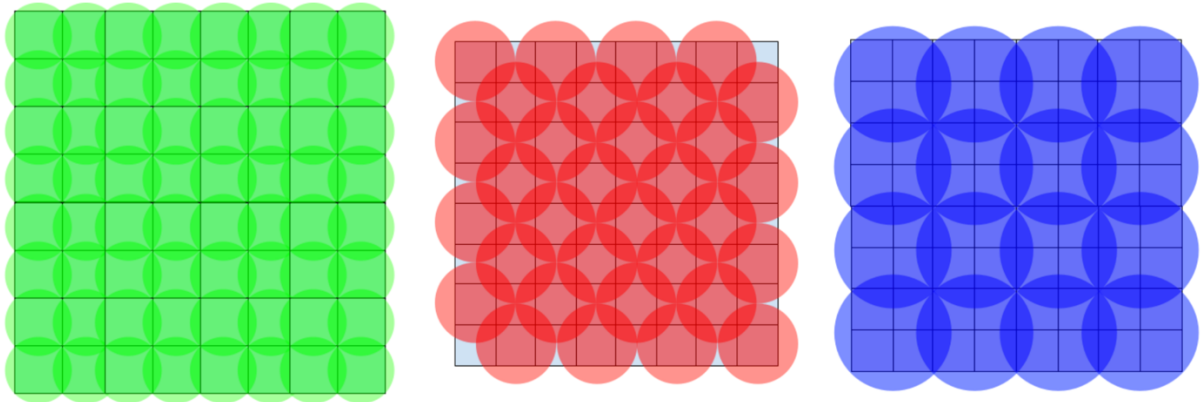
Figure 2: Each of these 8x8 display pixel regions are covered with different sample densities.
1x1 with 64 (8x8) samples, 1x1 Checkerboard with 32 samples (4x8), and 2x2 with 16 (4x4) samples.

2.2.3 Tiled Rendering

Full-Frame rendering systems compose all visual elements into a single unified memory buffer representing the full extent of the display area. These elements might be drawn in an arbitrary order and therefore display transmission cannot begin until the last element is fully rendered. This adds latency to the display pipeline.

Alternatively, Tiled Rendering sequentially renders small regions called tiles (e.g., 64x64 display pixels) of the display area. Only the objects visible in each tile need to be rendered, after which those samples can be transmitted to the display. This significantly reduces the latency of the display process. Additionally, some rendering systems can automatically apply different parameters to the rendering of each tile, including the sampling density (how many samples will cover the displayed area of the tile) and the rendering quality. As horizontally adjacent tiles are rendered, their samples can be organized into sequences of blocks for compression and transmission. Upon reception and decoding, the foveation can be undone by replicated scaling. This allows the rendering sample rate to match the display update rate.

When rendering sparse AR content, there might be tiles which contain no pixel data. The rendering of such tiles could be entirely skipped, further accelerating the rendering process. Ideally, this acceleration could also be leveraged by the encoding, transmission, decoding, and display-write processes. In this way, sparse content could significantly reduce the latency of the display pipeline.

2.2.4 Display Memory Organization for Fast Scaling

High resolution display memories often have custom interfaces to enable both high bandwidth and low clock speed. Such solutions often exhibit very wide buses to write many values in parallel. They might also use "banking" to only drive a subset of neighboring values at the same time (e.g., only the R or G or B pixels in an OLED display each with separate bank-enable lines). Separate pixel rows might not be individually addressed (e.g., with a row address) but

might be grouped together into 4 or 8 rows collectively. Then, separate "row enable" lines might be used to write into single rows at a time. Replicated writes, used to accelerate the writing of foveated samples into the display, could leverage a combination of horizontal "bank enable" and vertical "row enable" lines to write each decoded block into display memory much faster than traditional serial raster writes.

2.3 Picture Quality Requirements

The compression shall be visually lossless, which means that the difference between the original image or image sequence and the same image or image sequence after compression and decompression is not detectable to the eye. Annex A describes the test methodology VESA members will use to evaluate proposals.

2.4 Other Requirements

2.4.1 Design Complexity

- Solutions should exhibit low complexity in terms of gate count, memory sizes, power, and clock rate requirement, and the DSC Task Group prefers low-complexity solutions over high-complexity solutions.
- Proposals shall describe the coding system's complexity from an encoder perspective, a decoder perspective, and a system-level perspective as it relates to the compression layer.
- Proposals shall provide sufficient information to enable others to evaluate complexity.

2.4.2 Display Throughput

The proposal shall support non-uniform color field time per frame (i.e., different compression ratios for different fields (R/G/B) within a frame).

2.4.3 Buffer Model

- The coding system shall guarantee, at minimum, real time operation with no overflow and no underflow of compressed data buffers at the encoder and decoder for any practical implementation in all configurations with all possible content. Proposals shall describe in detail how correct real-time behavior of encoder and decoder buffers is assured by the proposed algorithm and the sizes of those buffers.
- The encoder shall support Constant Bit Rate (CBR) operation. A constant number of bits are removed from the encoder rate buffer per unit time. In CBR mode, this rate buffer is not permitted to overflow nor underflow.
- The encoder shall support Upper-Bounded Variable Bit Rate (UB-VBR). UB-VBR produces bits within a programmed rate while opportunistically saving bits as the coded content complexity allows. In UB-VBR mode, the encoder rate buffer with constant bits removed per unit time is allowed to underflow but not overflow. The DSC Task Group

recommends that the coded bit rate can be programmable. If the rate is programmable, proponents should specify the range and resolution of the programming.

- Proponents should assume that the pixel input to the encoder and pixel output of the decoder represent the passage of time. The "bits per pixel" represents the number of compressed bits that are sent by the encoder and received by the decoder for each pixel time.
- The DSC Task Group requires the inclusion of an HRD (Hypothetical Reference Decoder) model, if applicable, to ensure correct real-time buffer operation in the proposal. See Section 7 References for a description of the HRD.

2.4.4 Transport

The means to transport the bitstream is outside the scope of this CfP. Coded streams need to be transported over a variety of display links and transport schemes. The VESA committee will favor proposals that require minimum changes and constraints to existing transport schemes.

2.4.5 Slice Support

This section discusses slices to support partial refresh (i.e., the ability of a source to send only a portion of a frame and update the corresponding portion of the image in a compressed or uncompressed frame buffer associated with a display). The following definition applies:

Slice: A set of compressed bits that represents a specified set of samples. The set of samples forms a rectangle in the horizontal and vertical dimensions. This set of bits is independently decodable because decoding of any one slice shall not depend on the availability of another slice nor on the decoded result of another slice.

The proposal will be evaluated at a recommended slice size determined by the submitter. The proposal should specify the minimum slice size to be as small as practical. The value of using small slices is apparent in an AR headset, where only a virtual image designed as an irregular polygon could be updated and appear in a portion of the display screen where the remainder of the display allows headset transparency.

In upper-bounded variable bit rate (UB-VBR) operation, the number of bits allocated for each slice will vary. The UB-VBR bitstreams for each slice will start at the beginning location of a compressed slice's boundary; furthermore, the UB-VBR average bpp for any slice shall not exceed a defined maximum bpp for the codec system based on the available decoder frame memory.

In a CBR system, the number of bits allocated for each slice is constant for all content, and shall be equal to the target bits per pixel times the pixels per slice.
In cases of panels containing a frame buffer, coded image data might be stored directly in the frame buffer without passing through a decoder. In this case, a partial frame update with independently coded slices could update a portion of the compressed frame in the frame buffer.

The slice size for evaluation needs to be explicitly specified by the submitter and support visually lossless results. In no case shall the evaluation slice size exceed the picture width horizontally by 8 lines vertically.

2.4.6 Error Recovery

The decoder is not expected to be resilient to bit errors. However, if a bit error occurs during a slice, the next slice received without error shall be processed correctly by the decoder. This implies that no state may be carried from one slice to another slice within a decoder.

3 Test Material, Coding Conditions

3.1 Test Material

Initially, test material will consist of stereoscopic still images from test material developed for previous DSC Task Group virtual reality testing studies; see the references for details of these studies. Augmented reality stereoscopic images with varying fill factors are currently being developed from the virtual reality sets using masking modifications. Stereoscopic images with chromatic aberration correction (CAC), lens distortion correction (LDC), source mura correction, and color field-sequential properties are also under development. Other test material, such as stereoscopic video, will be selected or developed by the DSC TG. Links to available test images will be provided upon request to the VESA Moderator; see Section 8 Contacts and submission process.

The parameters for simulating LDC and CAC will depend on the viewing device (display and optics). Source mura correction will be simulated using a pre-generated mura mask containing independent and identically distributed (iid) noise applied in the gamma-corrected domain. The magnitude of noise for source mura correction will be equal to 5% of the signal dynamic range. The inverse map will be applied to the data prior to compression and the end-to-end picture quality will be assessed.

3.2 Coding Conditions

Initially, only RGB interleaved images will be used for testing. The pixel depth in bits per component, and resolution will vary with each test image set. Testing will be expanded beyond these initial limits by the DSC Task Group as needed.

4 Submission Requirements

Information on file formats can be found in ANNEX B – Model image file types and filename conventions.

There is a three-phase timeline for proposals. The binary encoder and decoder executable models and overview presentations are due for the first phase. The detailed technical documentation is due for the second phase. The source code is due for the selected proposal for the third phase.

Proposals shall include the following (incomplete submissions will not be considered):

4.1 Test Model Binary Encoder and Decoder Executables

Proposers need to submit separate encoder and decoder executable programs (Windows console EXE) with the following attributes or capabilities:
      1) Configurable via command line or configuration file
      2) Can be configured to different compressed bit rates (bpp) by the user (if applicable)
      3) Can be configured to run either with a programmable slice size or with a slice size of the proposer's choosing (if applicable)
      4) Performance consistent with a design that can meet all requirements (throughput, real-time behavior, etc.) Proponents can choose to use executable compression or similar tools to prevent reverse engineering or disassembly of the submitted EXE files.

4.2 Overview Presentation

Each proposal shall include a presentation that provides a general overview of the proposal and associated complexity. This presentation may be either in Microsoft PowerPoint, Word document, or PDF format.

After all submissions of the above are received, proponents may withdraw by notifying the DSC committee before the next phase. Proponents that wish to continue must then submit the following:

4.3 Technical Documentation

A technical description of the coding scheme that includes:
      Theory of operation
      Coded bitstream syntax
      Coding process (encoding and decoding) methodology.

The description shall include all necessary processing (including performance optimizations) that are used to create the bitstreams in Section 4.1 Test Model Binary Encoder and Decoder Executables 4.1 Test Model Binary Encoder and Decoder Executables in a bit-exact manner. The description shall show how the requirements in Section 2 Purpose and Requirements2 Purpose and of this CfP are met. Proponents are encouraged to list other features, benefits and

performance advantages of their architecture or other assumptions explicit or implicit in the proposed architecture.

The technical description shall contain sufficient information for experts to determine an approximate area and power for decoder and encoder implementations. The sizes of any on-chip or off-chip memories required for a hardware implementation need to be specified.

The technical description shall specify how the mandated throughput is achieved, including the relationship of clocks to pixels for all content. If achieving the mandated throughput requires additional complexity, that complexity needs to be documented.

4.4 Source Code

Proponents agree to release source code written in a high-level language, such as C, C++ or Python, within 3 days if their proposal is selected. Source code shall be documented and understandable. Assembly language is not permitted. All libraries used by the source code shall be either public or provided in source code form with an appropriate license permitting unlimited and free re-use. Make files or project files need to support compilation on both Windows and Linux systems. The bitstream output of the encoder and image output of the decoder must match the submitted Windows EXE files. If there are material differences found between the source code and other submitted materials, the selection may be voided at the discretion of the VESA DSC.

4.5 Submission of Technology Requirements

Proposers shall submit a completed and signed "Exhibit A: Submission of Technology Form" from VESA Policy #200D ("Intellectual Property Rights (IPR) Policy"), Section 12, at the same due date as the submission of full technical proposals (see Section 5 Timeline.)

VESA Policy #200D is available for download from the VESA website (http://www.vesa.org/join-vesamemberships/).

5 Timeline

The due dates for responses to this CfP and the DSC selection schedule are provided below. Dates may be subject to change at the discretion of the VESA DSC committee.

CfP Issued: April 15, 2024

Proposals containing the test model binary executables and overview presentations are due: July 29, 2024 (tentative face-to-face meeting to review technology submissions on July 30-August 1, 2024).

Test model technical documentation available: September 15, 2024

Test model evaluation reports: January 2025

Final tool selection and delivery of source code: April 2025

Draft specification: October 2025

Adoption: February 2026

7 References

MacInnis, S., Kerofsky, L. (2012) DSC Hypothetical Reference Decoder (HRD) Buffer Model Introduction, VESA DSC TG, https://vesa.org/wp-content/uploads/2024/03/HRD-Buffer-Model-Intro121026.pdf

"VESA Display Stream Compression (DSC) Standard Version 1.2b." http://www.vesa.org

"VESA Display Compression-M (VDC-M) Standard Version 1.1" http://www.vesa.org

Mohona, S. S., Au, D., Wilcox, L. M., & Allison, R. S. (2022). The effect of chromatic aberration correction on visually lossless compression. Digest of Technical Papers - SID International Symposium, 53(1), 1013–1016.

Mohona, S. S., Wilcox, L. M., & Allison, R. S. (2021). Subjective assessment of display stream compression for stereoscopic imagery. Journal of the Society for Information Display, 29(8), 591–607.

Mohona, S. S., Au, D., Kio, O. G., Robinson, R., Hou, Y., Wilcox, L. M., & Allison, R. S. (2020). Subjective Assessment of Stereoscopic Image Quality: The Impact of Visually Lossless Compression. 12th International Conference on Quality of Multimedia Experience, QoMEX.

Au, D., Mohona, S. S., Cutone, M. D., Hou, Y., Goel, J., Jacobson, N., Allison, R. S., & Wilcox, L. M. (2019). Stereoscopic image quality assessment. Digest of Technical Papers - SID International Symposium, 50 (Book 1), 13–16.
Sudhama A, Cutone M, Hou Y, Goel J, Stolitzka D, Jacobson N, Allison RS, Wilcox LM (2018) Visually lossless compression of high dynamic range images: A large scale evaluation. Invited paper, Society for Information Display Digest, 48:1.

Allison RS, Wilcox LM, Wang W, Hoffman D, Hou Y, Goel J, Deas L, Stolitzka D (2017) Large Scale Subjective Evaluation of Display Stream Compression. Invited paper, Society for Information Display Digest, 48:1, 1101-1104.

Brunnstrom K (Editor) Allison RS, Chandler DM, Collett H, Corriveau P, Daly S, Goel J, Knopf J, Wilcox LM, Yaacob Y, Yang S –N, Zhang Y*. (2017) Industry and business perspectives on the distinctions between visually lossless and lossy video quality: Mobile and large format displays. Invited paper, Human Vision and Electronic Imaging, Burlingame, CA.

8 Contacts and Submission Process

Please submit questions concerning this Call for Proposals to the VESA Moderator at the below email.

Proponents shall contact the VESA Moderator to obtain customized instructions for uploading materials to an HTTPS-secured folder on the VESA Causeway website. After acceptance by the Moderator, the uploaded proposals will be made accessible to VESA members. The DSC Task Group will determine a suitable time to make all proposals simultaneously available for review.

Proponents shall be responsible for annotation of any items that are confidential within the submitted materials. Such annotation shall be subject to the conditions of VESA Policy #221C, "Document Disclosure and Distribution Policy". Any materials marked as confidential to a particular company or legal entity other than VESA will be rejected. Copies of VESA Policy #221C are available from the VESA Moderator for non-members. Moderator email: moderator@vesa.org

ANNEX A – Test methodology

A.1 Methodology

Testing will include both stereoscopic still images and stereoscopic video.   As a minimum, testing using a stereoscope as described in the references will be used to evaluate the compressed images with full reference uncompressed images.

Testing and validation will also include commercial HMDs using stereoscopic still images at minimum for subjective testing.

Objective metrics, for example PSNR, SSIM, S-CIELAB, or those developed specifically for foveated displays, can be quite useful for initial assessment of visually quality of the codec. However, the final evaluation of whether an image is visually lossless or not will be determined by subjective testing as described above.

A.2 Viewing Conditions

The committee does not intend to impose any specific limitations on the viewing conditions for this test. However, evaluators will document the viewing conditions used for the tests.

A.3 Content Types

Many types of stereoscopic images will be evaluated: continuous tone images, landscapes, people portraits, animals, fine text, web pages, graphics, game captures, etc. Test patterns such as noise and zone plates will be evaluated, but some visual loss might be tolerated on certain patterns.

ANNEX B – Model image file types and filename conventions

The submitted Windows EXE models shall support uncompressed portable pixmap format (PPM), P6, (https://en.wikipedia.org/wiki/Netpbm#File_formats) for both image input and output.

Proponents will each be assigned a code (in the examples below, XYZ). The Windows EXE files should be labeled as:

- Encoder_XYZ.exe
- Decoder_XYZ.exe

The encoder and decoder shall be configurable either through a documented command line or configuration file. Each execution shall encode or decode one frame. In CBR operation, the bitstream output file size from the encoder should be approximately the bits per pixel times the number of pixels in the frame. In UB-VBR operation, the bitstream output file size from the encoder is typically less than the CBR case.