**INTERNATIONAL ORGANISATION FOR STANDARDISATION**
**ORGANISATION INTERNATIONALE DE NORMALISATION**
**ISO/IEC JTC1/SC29/WG11**
**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 MPEG2016/JCTVC-W0072**
**February 2016, San Diego, USA**

**Source:** University of Warwick
**Status:** Information and Discussion
**Title:** Highly efficient HDR video compression
**Author:** Jonathan Hatchett, Kurt Debattista, Ratnajit Mukherjee, Thomas Bashford Rogers, Alan Chalmers

**Abstract**

The two current standards for HDR video compression: SMPTE ST2084 and ARIB STD-B67, are relatively computationally complex which may preclude their efficient implementation on time critical ICT infrastructure. Power Transfer Functions (PTFs), on the other hand, are straightforward and well suited to implementation on GPUs. In this paper we present $\text{PTF}_4$, a transfer function that offers significantly improved computational performance, even when compared with LUTs, without any loss in quality.

# 1 Introduction

One-stream HDR video compression methods, including SMPTE ST2084 [1] and ARIB STD-B67[2], take advantage of the higher bit depths of modern encoding standards, such as HEVC [13] and transform a single HDR video input stream into a single compressed stream, using a transfer function [5]. In SMPTE ST2084 this transfer function is the PQ curve [11], while ARIB STD-B67 uses Hybrid Log Gamma (HLG) [4].

A Power Transfer Function (PTF) is a one-stream HDR video compression method that uses a power function, similar to the well-known Gamma function used in SDR video:

$$L = AV^{\gamma} \tag{1}$$

where: A is a constant, $L$ and $V$ are contained by the set $\mathbb{R} \in [0, 1]$, and $\gamma \in \mathbb{R}^+$. In this paper we show that $\gamma = 4$ not only produces high quality HDR video compression for a range of footage, but that this can also be achieved very efficiently.

# 2 $\text{PTF}_4$

The Human Visual System (HVS) perceives a scene in a non-linear manner. A PTF can be chosen which mimics the way in which the HVS has greater sensitivity in relative differences in darker areas of a scene than brighter areas [6, 14]. In particular $\text{PTF}_4$ allocates more values to the dark regions then to the light regions. Before an HDR video is compressed with a PTF it needs to be normalised to the range $[0, 1]$ using a normalisation factor $\mathfrak{N}$. If the footage is of an unknown range then it can be scanned in order to determine the correct $\mathfrak{N}$ for encoding, or for live broadcast, $\mathfrak{N}$ can be set to the peak brightness the camera is capable of capturing or the display can show. The algorithms for encoding and decoding a PTF are shown in Algorithms 1 and 2

**Algorithm 1** Power Transfer Function Encoding

```
procedure ENCODEPTF_4(frames_in, 𝔑)
    for i ← 1,LENGTH(frames_in) do
        S ← frames_in[i]
        L ← S/𝔑
        V ← L^(1/4)
        Q ← QUANTISE(V)
        frames_out[i] ← Q
    end for
    return frames_out, 𝔑
end procedure
```

# 3 Results

To prove the efficacy of $\text{PTF}_4$ we evaluated it against ST2084 (labelled PQ) and STD-B67 (labelled HLG) for a range of footage and bit rates. The metrics used were:

**HDR-VDP-2.2.1** Based on a detailed model of human vision [9], HDR-VDP-2.2.1 estimates

**Algorithm 2** Power Transfer Function Decoding

**procedure** DECODEPTF$_4$($frames_{in}$, $\mathfrak{N}$)
  **for** $i \leftarrow 1$,LENGTH($frames_{in}$) **do**
    $Q \leftarrow frames_{in}[i]$
    $V \leftarrow$DEQUANTISE($Q$)
    $L \leftarrow V^4$
    $S \leftarrow L \cdot \mathfrak{N}$
    $frames_{out}[i] \leftarrow S$
  **end for**
  **return** $frames_{out}$
**end procedure**

the probability at which an average human observer will detect differences between a pair of images in a psychophysical evaluation. This metric has been shown to be the objective metric that correlates most highly with subjective studies [12].

**PSNR** (Peak Signal to Noise Ratio) is one of the most widely used metrics for comparing processed image quality. Although not always analogous to perceived image quality, PSNR can provide an indication of HDR image quality.

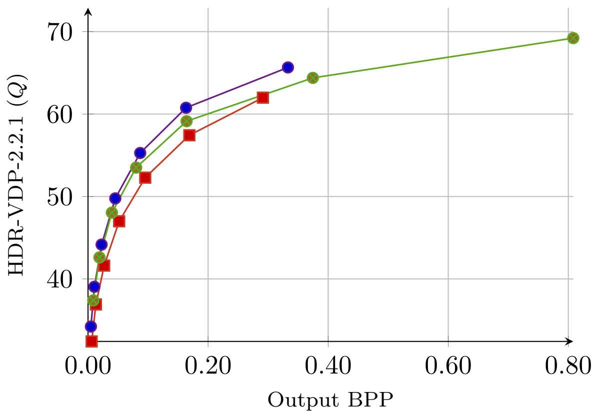
**puPSNR** (perceptually uniform PSNR) is an extension to PSNR such that it is capable of handling real-world luminance levels without affecting the results for existing displays [3].

PSNR and puPSNR were calculated every frame, while HDR-VDP-2.2.1 was calculated every 10th frame (as it is computationally expense to calculate), and all were averaged to produce a final result for the sequence.

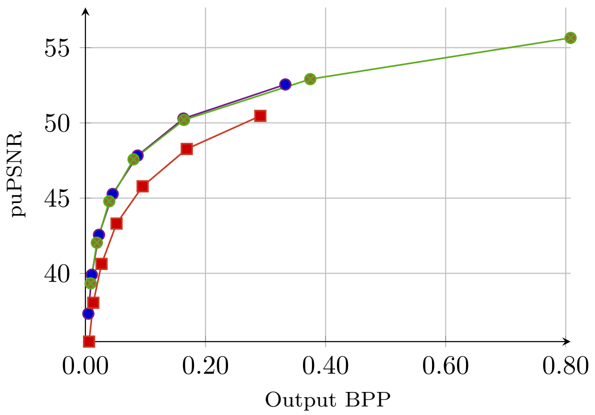
Ten HDR video sequences, described in Table 1, were chosen for the evaluation. These sequences were selected as they cover a wide range of content and dynamic range. Each sequence consisted of 150 frames and was encoded at 24 frames per second.

## 3.1 Quality

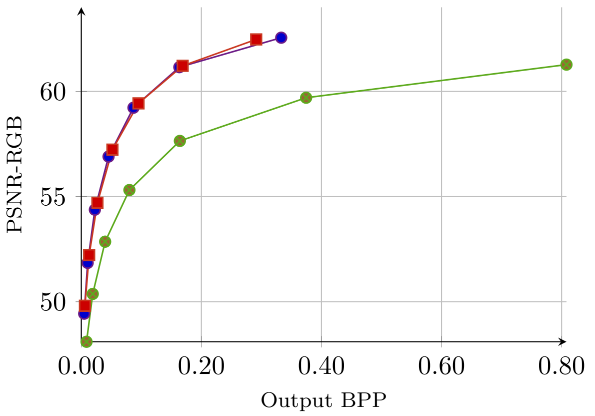
Fig. 1 shows the results of each method averaged across all 10 sequences while Fig. 2 shows results for the individual HDR video sequences. In the figures the HDR-VDP-2.2.1 $Q$ correlates values against the output bits per pixel required to encode the sequences at a requested QP. The results show that PTF$_4$ produces the highest quality for a given bitrate compared to HLG and PQ for HDR-VDP-2.2.1 and puPSNR, while HLG has a higher PSNR than PTF$_4$.




(a) HDR-VDP-2.2.1 Average




(b) puPSNR Average




(c) PSNR-RGB Average

PTF$_4$ HLG PQ

Figure 1: Rate distortion characteristics showing the results of each method averaged over the 10 sequences for the 3 metrics: HDR-VDP-2.2.1, PSNR, and puPSNR.

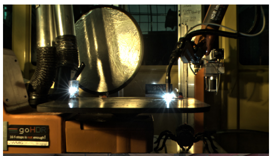
| Preview | Name | Resolution | Dynamic Range (Stops) |
|---|---|---|---|
|  | Welding | 1920×1080 | 20.54 |
|  | CGRoom | 1920×1080 | 19.29 |
|  | Jaguar | 1920×1080 | 25.30 |
|  | Seine | 1920×1080 | 20.54 |
|  | Tears of Steel | 1920×800 | 20.35 |
|  | Cars Fullshot | 1920×1080 | 19.85 |
|  | Beerfest Lightshow 4 | 1900×1060 | 22.11 |
|  | Carousel Fireworks 9 | 1900×1060 | 22.38 |
|  | Bistro 3 | 1900×1060 | 22.13 |
|  | Showgirl 1 | 1900×1060 | 22.73 |

Table 1: The 10 sequences used in the evaluation.

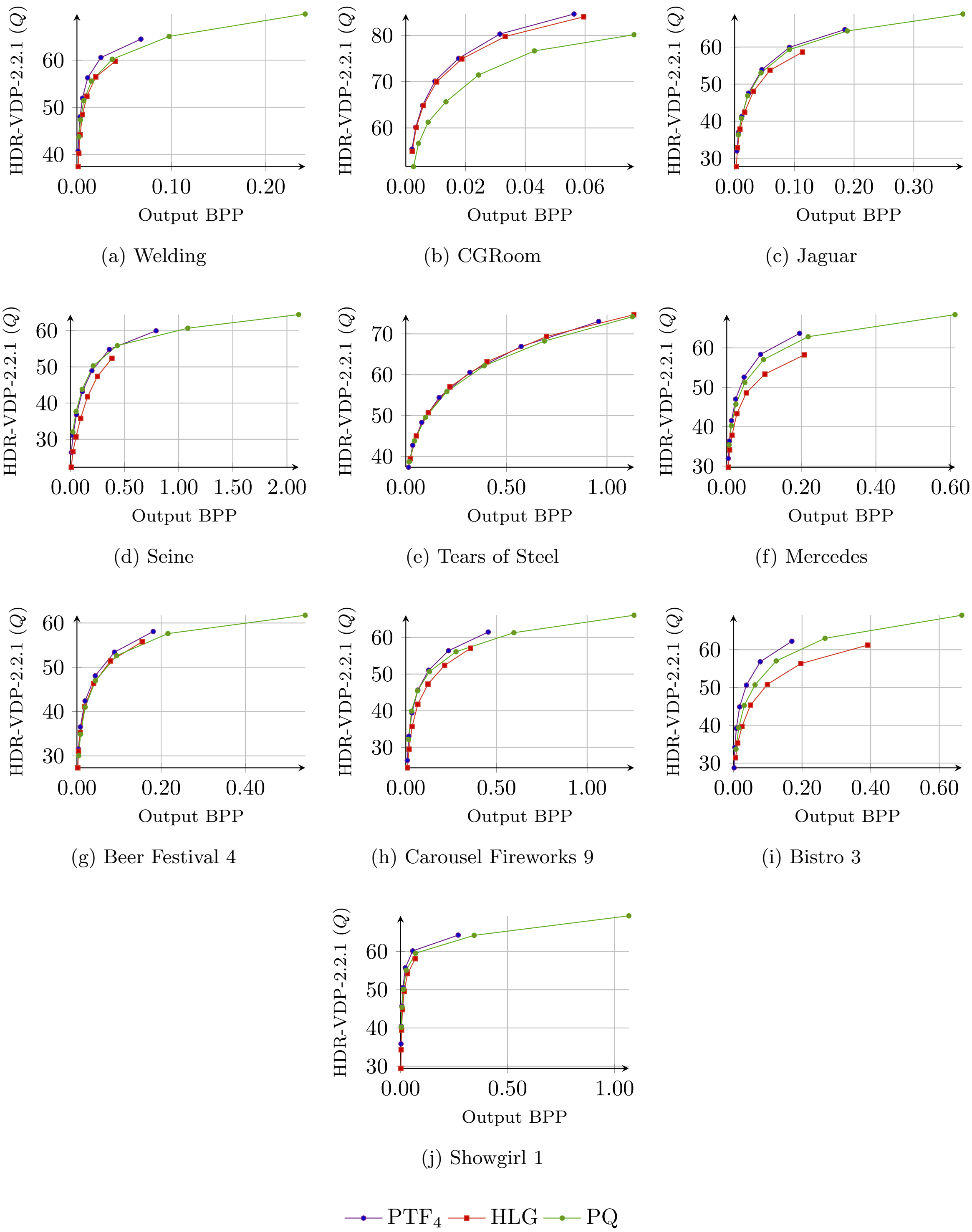



Figure 2: Rate distortion characteristics showing how the different HDR video compression methods perform on a variety of sequences. The rate is measured in output bits per pixel (BPP) and the distortion as a HDR-VDP-2.2.1 $Q$ correlate [9].

| Name | Time per Frame (ms) Analytic $PTF_4$ | PQ | LUT | Speed Up (ratio) $PTF_4$ PQ | LUT |
|---|---|---|---|---|---|
| Welding | 2.57 | 66.37 | 4.13 | 25.85 | 1.61 |
| CGRoom | 2.49 | 66.06 | 3.68 | 26.53 | 1.48 |
| Jaguar | 2.73 | 66.78 | 3.92 | 24.47 | 1.44 |
| Seine | 2.58 | 64.01 | 3.92 | 24.86 | 1.52 |
| Tears of Steel | 2.69 | 98.08 | 3.95 | 36.49 | 1.47 |
| Mercedes | 2.72 | 73.57 | 3.80 | 27.00 | 1.39 |
| Beer Festival 4 | 2.61 | 65.16 | 3.73 | 24.92 | 1.43 |
| Carousel Fireworks 9 | 2.56 | 65.91 | 3.77 | 25.79 | 1.48 |
| Bistro 3 | 2.63 | 65.85 | 3.82 | 25.00 | 1.45 |
| Showgirl 1 | 2.70 | 69.39 | 3.89 | 25.69 | 1.44 |
| Average | 2.63 | 70.12 | 3.86 | 26.66 | 1.47 |

Table 2: Difference in decoding time per frame between $PTF_4$, PQ and a generic LUT ($PTF_4$ in this case) across a range of sequences and averaged over 5 tests per sequence performed on a workstation PC. Speed up is ratio between PQ and $PTF_4$, and between $PTF_4$ and the LUT.

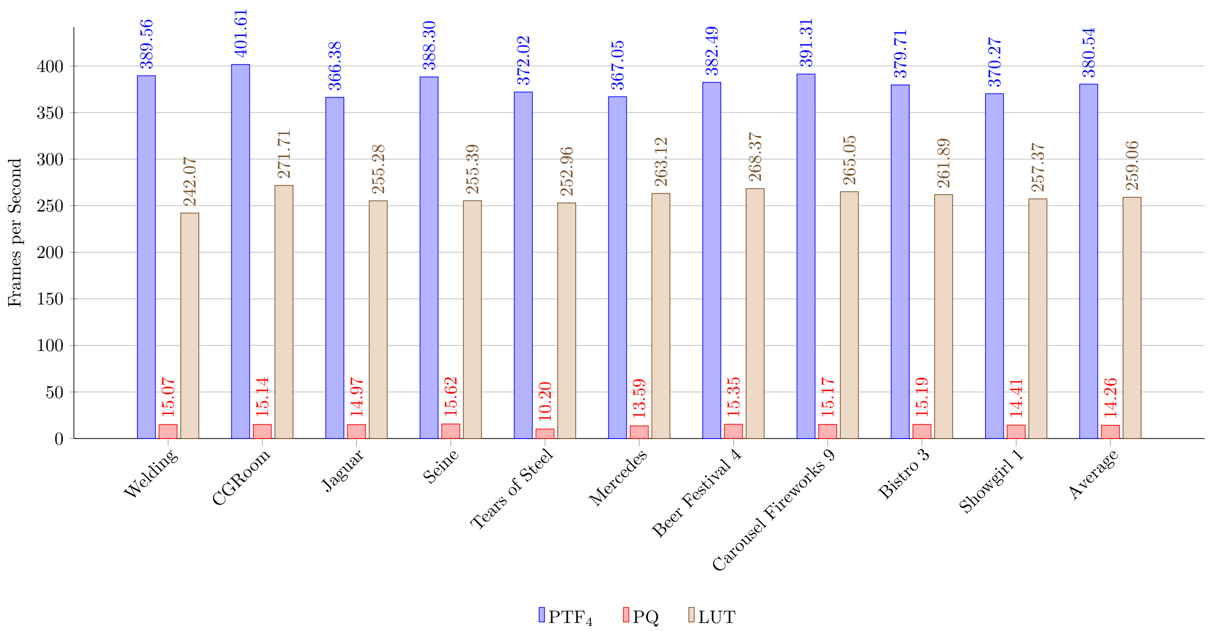



Figure 3: Difference in decoding time in frames per second between $PTF_4$, PQ and a generic LUT ($PTF_4$ in this case) across a range of sequence and averaged over 5 tests per sequence on a workstation PC (Higher is better).

### 3.2 Performance

The results shown in this section were produced by a single-threaded C++ decoding implementation compiled with the Intel C++ Compiler v16.0 hosted in Microsoft Visual Studio 2013 [8, 10]. Only the inner loop was timed so disk read and write speeds are not taken into account. Each result shown is the average of 5 tests per method on each sequence to reduce the variance associated with CPU timing. The software was compiled with the AVX2 instruction set with automatic loop-unrolling, O3 optimisations and fast floating-point calculations. The machine used to run the performance tests was an Intel Xeon E3-1245v3 running at 3.4 GHz with 16 GB of RAM and running the Microsoft Windows 8.1 x86-64 operating system with as few other applications and services running as could reasonably be achieved.

Table 2 and Fig. 3 show the performance of $PTF_4$ and PQ and their look-up table (LUT) equivalents, $PTF_4$ LUT and PQ LUT. Frames of resolution 1920×1080 were decoded from sequences 150 frames in length. The 1D LUTs were generated by storing the result of each transfer function for every 10-bit input value and the result stored in a floating-point array. Scaling was also included in the table to improve performance producing a mapping from 10-bit compressed RGB to full HDR floating-point.

## 4 Conclusion

The quality achieved with $PTF_4$ HDR video compression as evaluated with HDR-VDP-2.2.1 is as good as, and in many cases better, than SMPTE ST2084 and ARIB STD-B67. Furthermore, the performance results clearly show that the straightforward calculations required to decode $PTF_4$ can outperform the calculations required to decode PQ and even the indexing needed to access a look-up table. One reason for the high performance of $PTF_4$ decoding is that it can be implemented using only simple instructions that have high performance SIMD implementations in AVX2.

## 5 Patent Rights Declaration(s)

The University of Warwick may have current or pending patent rights relating to the technology described in this contribution and, conditioned on reciprocity, is prepared to grant licenses under reasonable and non-discriminatory terms as necessary for implementation of the resulting ITU-T Recommendation | ISO/IEC International Standard (per box 2 of the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form).

## 6 Acknowledgements

We would like to thank Edouard Francois for the Seine content which has been kindly provided by Technicolor and Binocle in the context of MPEG. For the Cars Longshot, Beerfest Lightshow, Carousel Fireworks, Bistro and Showgirl sequences we would like to thank the University of Stuttgart [7]. We would like to thank the Blender Foundation for Tears of Steel which is made available under the Creative Commons.

## References

[1] *ST 2084:2014 : High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays.* Society of Motion Picture and Television Engineers (SMPTE), August 2014. doi: 10.5594/s9781614828297. URL `http://dx.doi.org/10.5594/s9781614828297`.

[2] Association of Radio Industries and Businesses (ARIB). STD-B67: Essential parameter values for the extended image dynamic range television (EIDRTV) system for programme production, July 2015.

[3] T. O. Aydın, R. Mantiuk, and H.-P. Seidel. Extending quality metrics to full luminance range images. In *Electronic Imaging 2008*, pages 68060B–68060B. International Society for Optics and Photonics, 2008.

[4] T. Borer. Non-linear opto-electrical transfer functions for high dynamic range television. 2014.

[5] A. Chalmers, J. McNamee, J. Hatchett, R. Mukherjee, I. Olaizola, and K. Debattista. 12 bits is simply not enough for hdr video! *BEC, NAB*, 2015.

[6] G. T. Fechner. Ueber eine scheibe zur erzeugung subjectiver farben. *Annalen der Physik*, 121(10):227–232, 1838. ISSN 1521-3889. doi: 10.1002/andp.18381211004. URL `http://dx.doi.org/10.1002/andp.18381211004`.

[7] J. Froehlich, S. Grandinetti, B. Eberhardt, S. Walter, A. Schilling, and H. Brendel. Creating cinematic wide gamut hdr-video

for the evaluation of tone mapping operators and hdr-displays, 2014. URL http://spiedigitallibrary.org.

[8] Intel Corp. Intel parallel studio xe, January 2016. URL https://software.intel.com/en-us/intel-parallel-studio-xe.

[9] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich. HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions. In *ACM Transactions on Graphics (TOG)*, volume 30, page 40. ACM, 2011.

[10] Microsoft Corp. Microsoft visual studio c++ compiler, January 2016. URL https://www.visualstudio.com/.

[11] S. Miller, M. Nezamabadi, and S. Daly. Perceptual signal coding for more efficient usage of bit codes. In *SMPTE Conferences*, volume 2012, pages 1–9. Society of Motion Picture and Television Engineers, 2012.

[12] R. Mukherjee, K. Debattista, T. Bashford-Rogers, P. Vangorp, R. Mantiuk, M. Bessa, B. Waterfield, and A. Chalmers. Objective and subjective evaluation of high dynamic range video compression. To appear.

[13] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand. Overview of the high efficiency video coding (hevc) standard. *Circuits and Systems for Video Technology, IEEE Transactions on*, 22(12):1649–1668, 2012.

[14] E. H. Weber. *De Pulsu, resorptione, auditu et tactu: Annotationes anatomicae et physiologicae*. CF Koehler, 1834.