

# Temporal Frequency Probing for 5D Transient Analysis of Global Light Transport

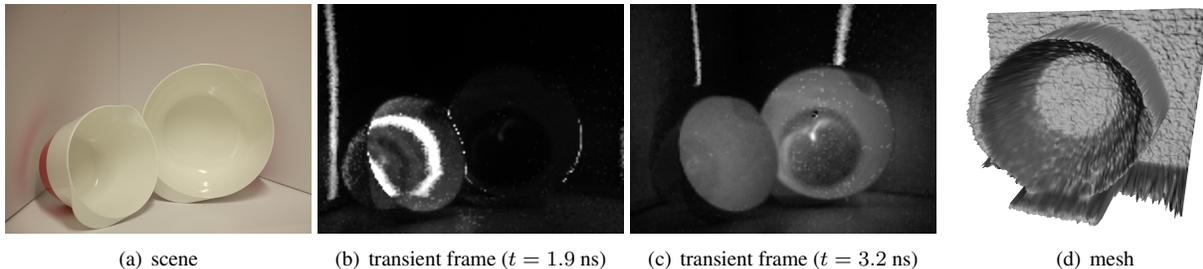
Matthew O’Toole<sup>1</sup> Felix Heide<sup>2</sup> Lei Xiao<sup>2</sup> Matthias B. Hullin<sup>3</sup> Wolfgang Heidrich<sup>2,4</sup> Kiriakos N. Kutulakos<sup>1</sup>

<sup>1</sup>University of Toronto

<sup>2</sup>University of British Columbia

<sup>3</sup>University of Bonn

<sup>4</sup>KAUST



**Figure 1:** (a) We use a photonic mixer device (PMD) coupled with a 2D projector to acquire “light-in-flight” images and depth maps for scenes with complex indirect transport. (b)-(c) The sharp, evolving wavefronts of light travelling along direct and caustic light paths can be seen clearly. (d) We also recover time-of-flight depth maps in a way that is not affected by strong indirect light. This indirect light mainly consists of diffuse inter-reflections for the scene in (a) but we also show experimental results where caustic light paths dominate.

## Abstract

We analyze light propagation in an unknown scene using projectors and cameras that operate at transient timescales. In this new photography regime, the projector emits a spatio-temporal 3D signal and the camera receives a transformed version of it, determined by the set of all light transport paths through the scene and the time delays they induce. The underlying 3D-to-3D transformation encodes scene geometry and global transport in great detail, but individual transport components (*e.g.*, direct reflections, inter-reflections, caustics, *etc.*) are coupled nontrivially in both space and time.

To overcome this complexity, we observe that transient light transport is always *separable in the temporal frequency domain*. This makes it possible to analyze transient transport one temporal frequency at a time by trivially adapting techniques from conventional projector-to-camera transport. We use this idea in a prototype that offers three never-seen-before abilities: (1) acquiring time-of-flight depth images that are robust to general indirect transport, such as interreflections and caustics; (2) distinguishing between direct views of objects and their mirror reflection; and (3) using a photonic mixer device to capture sharp, evolving wavefronts of “light-in-flight”.

**CR Categories:** I.4.1 [Image Processing and Computer Vision]: Digitization and Image Capture—Imaging geometry, radiometry

**Keywords:** coded illumination, computational photography, light transport, time-of-flight imaging, photonic mixer devices

**Links:** [DL](#) [PDF](#) [WEB](#)

## 1 Introduction

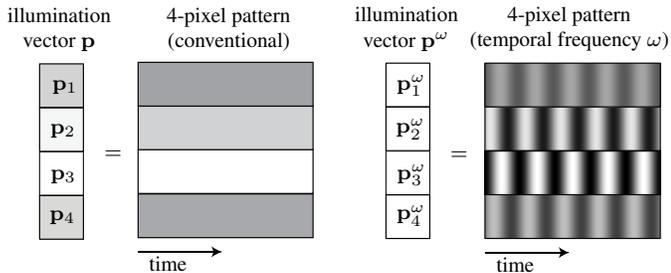
Recent years have seen a strong interest in using computational imaging methods to probe and analyze light transport in complex environments. A cornerstone of this body of work is the light transport equation, which describes the interaction of light with a scene in terms of a simple linear relation [Ng et al. 2003; Debevec et al. 2000; Goral et al. 1984; Kajiya 1986]:

$$\mathbf{i} = \mathbf{T} \mathbf{p} \quad (1)$$

where  $\mathbf{i}$  is a 2D image represented as a column vector of pixels,  $\mathbf{T}$  is the scene’s  $I \times P$  transport matrix, and  $\mathbf{p}$  is a vector that represents the scene’s spatially-varying illumination (*e.g.*, a 2D pattern projected onto the scene by a conventional video projector).

The transport equation governs a broad range of optical phenomena but the matrix  $\mathbf{T}$  is often unknown and must be inferred from images. This inference problem—after more than a decade of graphics and vision research—has produced a large repertoire of analysis techniques and is now quite mature. Examples include methods for acquiring [Peers et al. 2009], decomposing [Bai et al. 2010], transposing [Sen et al. 2005], approximating [Garg et al. 2006; O’Toole and Kutulakos 2010], or inverting [Seitz et al. 2005] the matrix; methods that use its properties for image-based rendering [Debevec et al. 2000] or transport-robust shape acquisition [Gupta and Nayar 2012]; and imaging techniques that enhance the contribution of specific light transport components [Nayar et al. 2006; O’Toole et al. 2012; Reddy et al. 2012]. All these techniques make the problem tractable by capturing images under many different illuminations at rates limited by current projection technology (30kHz or less).

Against this backdrop, transient imaging has emerged as an alternative paradigm for transport analysis that exploits light’s finite speed. Instead of illuminating a scene with spatially-varying patterns they rely on temporally-varying ones, using MHz to THz lasers and sensors sensitive to these rates (*e.g.*, streak cameras [Velten et al. 2013] and photonic mixer devices [Heide et al. 2013; Kadambi et al. 2013]). These techniques have opened new frontiers—looking around corners [Kirmani et al. 2011], time-of-flight depth imaging [Velten et al. 2012; Kirmani et al. 2013], lensless imaging [Wu et al. 2013], and capturing propagating optical wavefronts [Velten et al. 2013]—but are fundamentally limited in their ability to analyze complex global transport.



**Figure 2:** *Spatial vs. single-frequency spatio-temporal patterns. On the right, pixel intensities vary sinusoidally with a common temporal frequency  $\omega$  but their amplitudes and phases differ.*

In this paper we combine both paradigms by considering, for the first time, the generation and acquisition of transient space-time patterns for scene analysis. In this new analysis regime, the projector emits a 3D signal (2D space  $\times$  1D time) and the camera receives a transformed 3D version of it. This brings two sets of constraints to bear on the same problem: constraints on the *spatial layout* of light paths arriving at a given pixel and constraints on their *travel time*. These constraints are complementary but *not* orthogonal; thus, by considering them jointly we can draw far stronger conclusions about light transport than when spatial or temporal light patterns are used in isolation, or sequentially.

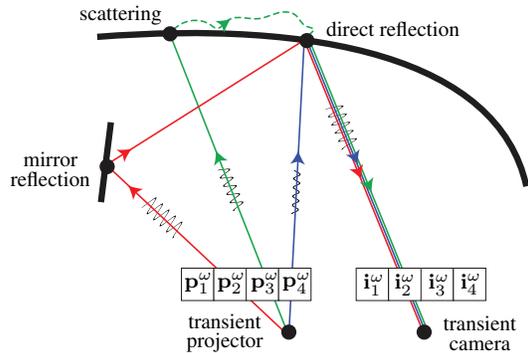
To demonstrate the practical advantages of this approach, we apply it to the problem of decomposing a scene’s transient appearance into transport-specific components. This very basic imaging task has received considerable attention in conventional light transport but is poorly understood in transient settings. Here we give a solution to this problem for general scenes and three transport components: (1) direct reflections and retro-reflections, (2) caustic light paths and (3) non-caustic indirect paths. We then implement this basic imaging ability in a functional prototype to

- improve the robustness of time-of-flight sensors against indirect transport by acquiring and processing only the direct/retro-reflective time-of-flight component;
- capture sharp, evolving wavefronts of “light-in-flight” that so far have been directly observed only with very expensive streak camera technology; and
- conduct space-time light path analysis to separate “true” scene points from their mirror reflections.

Toward these goals, our key contribution is a new form of the light transport equation that makes many classical transport analysis problems easy to formulate and solve in the transient domain. This new equation, which we call the *transient frequency transport equation*, assumes that we “probe” the scene by projecting a special spatio-temporal signal onto it—a signal whose pixel intensities vary sinusoidally in time with a common frequency  $\omega$  (Figure 2):

$$\mathbf{i}^\omega = \mathbf{T}^\omega \mathbf{p}^\omega, \quad (2)$$

where  $\mathbf{p}^\omega$  is a column vector of  $P$  complex numbers, each representing the sinusoid’s amplitude and phase for a specific projector pixel;  $\mathbf{i}^\omega$  represents the per-pixel sinusoids received at the camera; and  $\mathbf{T}^\omega$  is the scene’s  $I \times P$  *transient frequency transport matrix* for emission frequency  $\omega$ . Intuitively, this matrix tells us how the temporal sinusoid received at a specific camera pixel is affected by temporal sinusoids emitted by different projector pixels, after accounting for all global light transport paths and the delays they induce (Figure 3). In this sense, the transient frequency transport matrix describes light transport exactly like the conventional matrix does, except that it deals with per-pixel temporal sinusoids instead of per-pixel intensities. This matrix is different for different emission frequencies and reduces to the conventional transport matrix



**Figure 3:** *Visualizing the transient frequency transport equation. The projector emits sinusoids of frequency  $\omega$  from pixels 2, 3 and 4 and these signals propagate to camera pixel 2 along three paths: a direct path (blue), where light bounces only once; a caustic path (red), where light enters and exits the scene along distinct rays and undergoes several bounces in between, at most one of which is non-specular; and a non-caustic indirect path (green). Since all sinusoids have the same frequency, their superposition, recorded by camera pixel 2, will have the same frequency as well. Pixel 2’s amplitude and phase depend on path-specific attenuations and delays, specified in the elements of  $\mathbf{T}^\omega$  (e.g., for the green path it is  $\mathbf{T}_{23}^\omega$ ). For an example of a retro-reflective path, which can occur only when the projector and camera are coaxial, see Figure 10.*

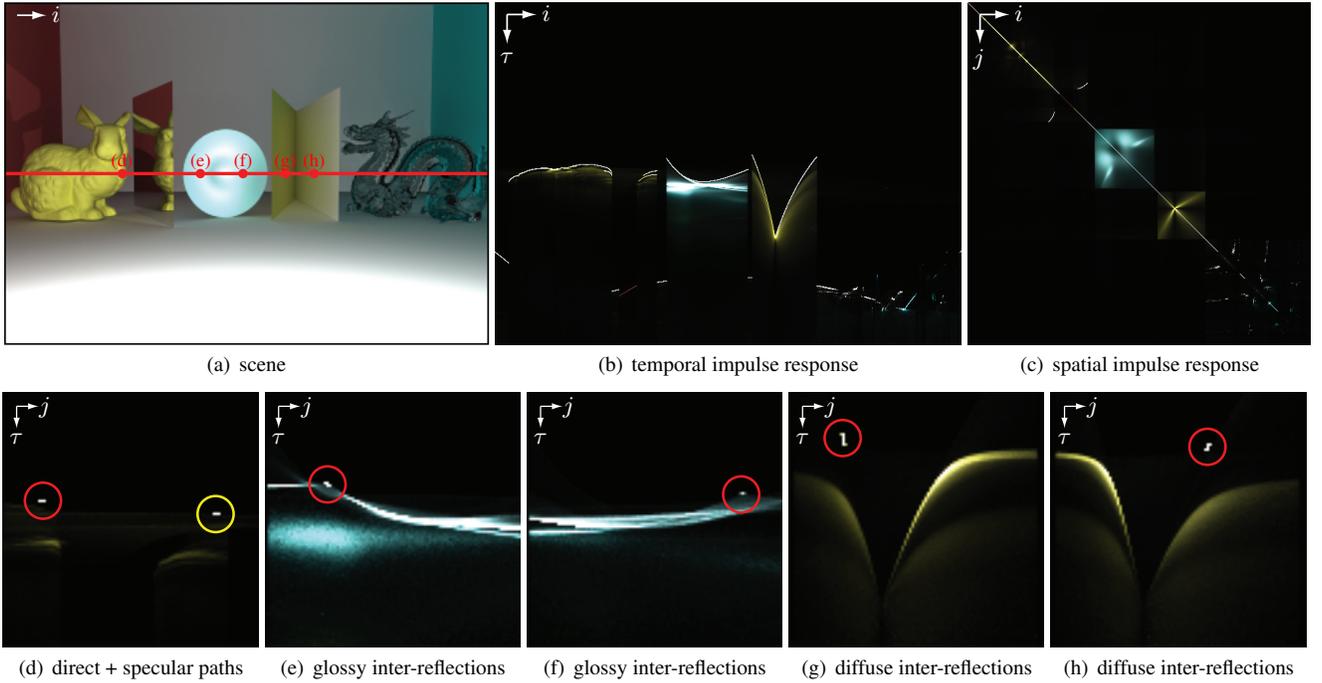
for the DC frequency, where illumination is a constant projection pattern. Taken together, this continuous family of discrete transport matrices is a five-dimensional structure that fully describes global light transport at transient timescales—from a projector plane to a camera plane via a 1D set of emission frequencies.

The fact that Equation 2 exists has two implications for light transport analysis. First and foremost, we can directly apply conventional light transport techniques to the transient case—without compromising their effectiveness or adding extra assumptions. This involves simply replacing conventional projectors and cameras with transient ones, and “probing” the scene by emitting from every projector pixel a temporal sinusoid of a common frequency  $\omega$ . We use this idea extensively in our prototype to acquire specific components of light transport with such probing patterns. These components cannot be captured robustly with existing techniques because temporal signals may propagate through the scene along many different paths and combine upon arrival at a pixel, making it impossible to separate them without strong assumptions about the actual signal received (e.g., diffuse one-[Kirmani et al. 2012] or three-bounce transport [Kirmani et al. 2011], dominant peak [Wu et al. 2012], parametric [Heide et al. 2013], and/or sparse [Dorrington et al. 2011; Kadambi et al. 2013]).

Second, probing the scene with a specific temporal frequency is relatively easy to implement with photonic mixer devices (PMDs). These devices offer affordable spatio-temporal imaging and can be configured to operate at a single emission frequency. Our prototype is built around such a camera, with spatio-temporal projection made possible by replacing the light source of a conventional projector with the PMD’s laser diode. In this respect, our prototype can be thought of as generalizing the point-source/transient-camera system of Heide et al. [2013] and the transient-projector/single-pixel-camera system of Kirmani et al. [2012].

## 2 Global Light Transport in Space and Time

We begin by deriving the transient frequency transport equation from first principles by considering propagation in space and time.



**Figure 4:** Simulated transient light transport, rendered using a modified path tracing integrator for non-caustic transport and a photon mapping integrator for caustic transport. For each camera pixel  $i$ , we simulate the travel time  $\tau$  of all light paths from the scene lit by a projector pixel  $j$ . (a) The scene contains, from left to right, a diffuse bunny, a mirror casting caustic light on the bunny, a glossy bowl, a diffuse v-shaped wedge, and a glass dragon. (b) A transient image of pixels  $i$  across times  $\tau$ . (c) A conventional light transport matrix, tabulating transport for each camera pixel  $i$  in response to projector pixel  $j$ . (d-h) Space-time impulse response functions for distinct camera pixels, as labelled in (a). Both the direct light paths, highlighted by the red circles, and caustic light paths, highlighted in yellow, are impulses in space-time. Note that (g) and (h) show the individual bounces of light occur between the two faces of the v-shaped wedge, whereas these multi-bounce light paths overlap in (b) and (c).

**The space-time impulse response** The conventional light transport equation ignores time completely. One way to interpret that equation is to think of the projector as emitting a time-invariant pattern; the camera then captures a photo only after light propagation through the scene has reached a steady state.

To take transient phenomena into account, we must consider the case where light has not yet reached a steady state. The most straightforward way to do this is to add a time dimension to the basic light transport equation:

$$\tilde{\mathbf{i}}(\tau) = \tilde{\mathbf{T}}(\tau) \tilde{\mathbf{p}}(0) \quad (3)$$

Here  $\tilde{\mathbf{p}}(0)$  is an illumination pattern that is emitted for an infinitely-brief interval at time zero, and  $\tilde{\mathbf{i}}(\tau)$  records only light that arrived at the camera exactly at time  $\tau$ . The time-varying matrix  $\tilde{\mathbf{T}}(\tau)$  therefore describes the part of the light transport in the scene that has a travel time of exactly  $\tau$ . From a signal processing perspective, this non-negative matrix-valued function can be thought of as the scene’s space-time impulse response [Kirmani et al. 2011].

The relationship between the space-time impulse response and the conventional light transport matrix is a simple time integral:

$$\mathbf{T} = \int_0^\infty \tilde{\mathbf{T}}(\tau) d\tau \quad (4)$$

**Transport of general spatio-temporal patterns** Now consider a general pattern  $\tilde{\mathbf{p}}(t)$  that varies in both space and time. If the projector emits  $\tilde{\mathbf{p}}(t_0)$  for an infinitely-brief interval at time instant

$t = t_0$ , the time sequence of camera pixel intensities that results from that emission will be

$$\tilde{\mathbf{i}}(t_0 + \tau) = \tilde{\mathbf{T}}(\tau) \tilde{\mathbf{p}}(t_0) \quad (5)$$

Emitting the full spatio-temporal pattern  $\tilde{\mathbf{p}}(t)$  will produce a time-varying image that takes into account all possible travel times:<sup>1</sup>

$$\tilde{\mathbf{i}}(t) = \int_{-\infty}^\infty \tilde{\mathbf{T}}(\tau) \tilde{\mathbf{p}}(t - \tau) d\tau \quad (6)$$

$$\stackrel{\text{def}}{=} (\tilde{\mathbf{T}} * \tilde{\mathbf{p}})(t) \quad (7)$$

where the operator  $*$  convolves a matrix function and a vector function over time.<sup>2</sup> More information on this form of the convolution operator and a brief overview of algebra and notation on matrix and vector functions can be found in the supplemental material.

When the scene’s space-time impulse response is known, we can use the convolution integral of Equation 7 to render the scene under any space-time illumination pattern (Figure 4). Applying this equation in practice, however, is difficult for several reasons. First, rendering even a single transient snapshot  $\tilde{\mathbf{i}}(t_0)$  requires the full 5D space-time impulse response because of the convolution integral involved. Second, this function can be extremely large because of the extra time dimension, compared to the conventional light transport matrix, making it even more challenging to measure, store, and

<sup>1</sup>Since light cannot travel back in time,  $\tilde{\mathbf{T}}(\tau)$  is zero for  $\tau < 0$ .

<sup>2</sup>Our definition of convolution here is consistent with convolution operators on tensor fields [Hlawitschka et al. 2004].

#	Description	Reference(s)	Conventional Light Transport	Single-Frequency Transient Light Transport
1	transport equation	[Debevec et al. 2000; Ng et al. 2003]	$\mathbf{i} = \mathbf{T} \mathbf{p}$	$\mathbf{i}^\omega = \mathbf{T}^\omega \mathbf{p}^\omega$
2	dual equation	[Sen et al. 2005; Sen and Darabi 2009]	$\mathbf{i} = \mathbf{T}^\top \mathbf{p}$	$\mathbf{i}^\omega = (\mathbf{T}^\omega)^\top \mathbf{p}^\omega$
3	inverse equation	[Wetzstein and Bimber 2007]	$\mathbf{i} = \mathbf{T}^\dagger \mathbf{p}$	$\mathbf{i}^\omega = (\mathbf{T}^\omega)^\dagger \mathbf{p}^\omega$
4	radiosity equation	[Goral et al. 1984]	$\mathbf{i} = \mathbf{p} + \mathbf{A} \mathbf{F} \mathbf{i}$	$\mathbf{i}^\omega = \mathbf{p}^\omega + \mathbf{A} \mathbf{F}^\omega \mathbf{i}^\omega$
5	radiosity solution	[Goral et al. 1984]	$\mathbf{i} = (\mathbf{I} - \mathbf{A} \mathbf{F})^{-1} \mathbf{p}$	$\mathbf{i}^\omega = (\mathbf{I} - \mathbf{A} \mathbf{F}^\omega)^{-1} \mathbf{p}^\omega$
6	inverse transport	[Seitz et al. 2005; Bai et al. 2010]	$\mathbf{T}^{-1} = \mathbf{A}^{-1} - \mathbf{F}$	$(\mathbf{T}^\omega)^{-1} = (\mathbf{D}^\omega)^{-1} [\mathbf{A}^{-1} - \mathbf{F}^\omega] (\mathbf{D}^\omega)^{-1}$
7	transport eigenvectors	[O’Toole and Kutulakos 2010]	$\lambda \mathbf{v} = \mathbf{T} \mathbf{v}$	$\lambda \mathbf{v} = \mathbf{T}^\omega \mathbf{v}$
8	probing equation	[O’Toole et al. 2012; O’Toole et al. 2014]	$\mathbf{i} = (\mathbf{T} \odot \mathbf{\Pi}) \mathbf{1}$	$\mathbf{i}^\omega = (\mathbf{T}^\omega \odot \mathbf{\Pi}) \mathbf{1}$
9	low/high-frequency transport separation	[Nayar et al. 2006]	$\mathbf{i}_{\text{low}} = \frac{1}{\alpha} \min_k \mathbf{T} \mathbf{p}_k$ $\mathbf{i}_{\text{high}} = \max_k \mathbf{T} \mathbf{p}_k - \alpha \mathbf{i}_{\text{low}}$	$\mathbf{i}_{\text{low}}^\omega = \frac{1}{\alpha} \min_k \mathbf{T}^\omega \mathbf{p}_k^\omega$ $\mathbf{i}_{\text{high}}^\omega = \max_k \mathbf{T}^\omega \mathbf{p}_k^\omega - \alpha \mathbf{i}_{\text{low}}^\omega$

**Table 1:** Related works on light transport analysis that have simple extensions to the single-frequency transient domain. In each instance, the transient formulation becomes the conventional (steady state) formulation at  $\omega = 0$ . **Rows 1-7:** Refer to the supplemental materials for more details on the notation and formulation. **Rows 8 and 9:** We implement the probing equation and fast transport separation to separate an image into three parts: a direct/retro-reflective component, a caustic component, and a non-caustic indirect component.

analyze directly. Third, representing this function as a discrete 5D array makes it difficult to infer properties of transient light transport because light travels along continuous and unbounded path lengths, and the function’s dynamic range can be extremely high (*e.g.*, direct light paths consist of Dirac peaks). Fourth, using impulse-like illumination patterns to analyze transient transport typically requires expensive equipment, long capture times, and exotic techniques to overcome signal-to-noise ratio issues.

**The transient frequency transport equation** Observe that the convolution in Equation 7 is only in the temporal domain. To derive the transient frequency transport equation we apply the convolution theorem to the time axis only, independently for each element of matrix  $\tilde{\mathbf{T}}(\tau)$ :

$$\mathcal{F}\{\tilde{\mathbf{i}}\}(\omega) = \mathcal{F}\{\tilde{\mathbf{T}} * \tilde{\mathbf{p}}\}(\omega) \quad (8)$$

$$\underbrace{\mathcal{F}\{\tilde{\mathbf{i}}\}(\omega)}_{\text{image } \mathbf{i}^\omega} = \underbrace{\mathcal{F}\{\tilde{\mathbf{T}}\}(\omega)}_{\text{matrix } \mathbf{T}^\omega} \underbrace{\mathcal{F}\{\tilde{\mathbf{p}}\}(\omega)}_{\text{pattern } \mathbf{p}^\omega} \quad (9)$$

where  $\mathcal{F}\{\}$  denotes the element-wise Fourier transform along the time axis and  $\omega$  denotes temporal frequency. For a fixed frequency  $\omega$ , Equation 9 is a matrix-vector product whose factors we denote as  $\mathbf{T}^\omega$  and  $\mathbf{p}^\omega$  for notational convenience. This brings it into the form shown in Equation 2.

The transient frequency transport equation can be interpreted as an image formation model for patterns like those shown in Figures 2 and 3, which contain just one temporal frequency. The most important advantage of this model is that it is separable in three domains simultaneously: the temporal frequency domain and the domains of the projector and the camera pixels, respectively. This leads to a mathematically simpler representation for light transport analysis. In particular, the contribution of all light paths from a specific projector pixel to a specific camera pixel—including all attenuations and time delays—is captured in just one complex number per temporal frequency (Figure 3).

Analyzing light transport one frequency at a time offers computational advantages as well: in contrast to the scene’s space-time impulse response which is 5D, the transient frequency transport matrix is a 4D object that has exactly the same size as the conventional transport matrix. Moreover, inexpensive and light-efficient PMD cameras can be configured to follow this single-frequency image formation model exactly, making it possible to perform transient light transport analysis *directly* in the temporal frequency domain.

### 3 Analysis by Temporal Frequency Probing

Much of the theory established for conventional light transport analysis applies to the transient case, as long as we restrict illumination to patterns that are single-frequency temporal sinusoids. Below we consider transient versions of two techniques we have implemented, as well as of the rendering equation. Table 1 summarizes several more, with details provided in the supplemental material.

**Probing the transport matrix** The techniques of O’Toole et al. [2012; 2014] make it possible to capture direct/retro-reflective and indirect-only photos of a scene. Their main feature is that they can be applied to scenes with high-spatial-frequency indirect transport paths (*e.g.*, containing mirror reflections, refractions, *etc.*). We use transient versions of these techniques to perform similar transport decompositions for images acquired with a PMD camera.

Specifically, transport probing implements a generalized image formation model that can be used to block specific light paths from contributing to a photo:

$$\mathbf{i} = (\mathbf{T} \odot \mathbf{\Pi}) \mathbf{1} \quad (10)$$

Here  $\mathbf{\Pi}$  is a known matrix that is multiplied element-wise with the transport matrix of the scene, which is considered unknown. Light paths are blocked by setting some of  $\mathbf{\Pi}$ ’s elements to zero. This has the effect of “zeroing-out” the corresponding elements of the transport matrix, preventing their associated light paths from affecting the photo. For instance, O’Toole et al. [2012] use a coaxial projector-camera arrangement to acquire direct/retro-reflective by choosing  $\mathbf{\Pi}$  to be diagonal. This is because direct light transport occurs only along retro-reflective paths, which always correspond to the diagonal elements of  $\mathbf{\Pi}$  and  $\mathbf{T}$ . Similarly, the binary complement of  $\mathbf{\Pi}$  yields an indirect-only photo.

We can readily use the transport probing model of Equation 10 for transient imaging because matrix elements  $\mathbf{T}_{ij}$  and  $\mathbf{T}_{ij}^\omega$  represent the exact same set of light paths for any  $i, j$ . As a result, matrix  $\mathbf{\Pi}$  has the same effect in both imaging regimes:

$$\mathbf{i}^\omega = (\mathbf{T}^\omega \odot \mathbf{\Pi}) \mathbf{1} \quad (11)$$

**The transient rendering equation** This equation represents all light transport as a function of time  $t$  [Kirmani et al. 2011]:

$$l_{ij}(t) = q_{ij}(t) + \sum_{k=1}^M f_{kij} l_{ki}(t - \tau_{ki}) \quad (12)$$

where  $l_{ij}(t)$  captures radiance along a ray,  $q_{ij}(t)$  is the emitted radiance,  $f_{kij}$  represents the form factors, and  $\tau_{ki}$  corresponds to the flight time between two points  $k$  and  $i$ . Note that this spatially-discrete expression is based on a continuous-valued (i.e. non-discretized) set of travel times  $\tau_{ki}$  between pairwise points.

We now prove that the transient rendering equation is itself separable in the temporal frequency domain:

$$\begin{aligned} l_{ij}^\omega &= q_{ij}^\omega + \int_{-\infty}^{\infty} \left( \sum_{k=1}^M f_{kij} l_{ki}(t - \tau_{ki}) \right) e^{-2\pi i t \omega} dt \\ &= q_{ij}^\omega + \sum_{k=1}^M f_{kij} e^{-2\pi i \tau_{ki} \omega} \int_{-\infty}^{\infty} l_{ki}(t) e^{-2\pi i t \omega} dt \\ &= q_{ij}^\omega + \sum_{k=1}^M f_{kij} e^{-2\pi i \tau_{ki} \omega} l_{ki}^\omega \end{aligned} \quad (13)$$

where  $l_{ij}^\omega$  denotes the Fourier coefficient of temporal function  $l_{ij}(t)$  for frequency  $\omega$ . The key difference between Equation 13 and the conventional rendering equation [Kajiya 1986] are the complex exponential factors that represent phase delay between pairs of scene points. Most of the equations in Table 1 follow as direct consequences of this equation. Note that the equation remains valid for discrete temporal frequencies as well.

#### Fast separation of low/high-frequency transport components

Another way to capture individual components of light transport is to conventionally project onto the scene a sequence of patterns that have high spatial frequencies [Nayar et al. 2006]. This takes advantage of the spatial-frequency-preserving properties of different components of the light transport matrix. The typical application is for acquiring the direct-only and indirect-only images of a scene. In particular, sub-surface scattering and diffuse inter-reflections typically act as low-pass filters on incident illumination because the form factors  $f_{kij}$  of the rendering equation are smooth relative to the spatial frequency of the illumination patterns. Direct transport, on the other hand, preserves high spatial frequencies and thus can be distinguished from low-spatial-frequency transport by analyzing a scene’s response to various high-spatial-frequency 2D patterns.

Nayar et al.’s separation technique can be applied to the transient domain too. This is because the transient form factors  $f_{kij} e^{-2\pi i \tau_{ki} \omega}$  in Equation 13 will also be smooth for space-time signals of a sufficiently low temporal frequency  $\omega$ .

In general, of course, indirect transport may also preserve high spatial frequencies because of caustic light paths. We follow the approach suggested by O’Toole et al. [2012] to handle this case: we capture the direct/retro-reflective and indirect-only components by probing the transient transport matrix at a single frequency; then we apply Nayar et al.’s method to decompose the indirect-only component further, into caustic and non-caustic indirect components.

**Relation between matrices  $\mathbf{T}^\omega$  for different frequencies  $\omega$**  Per-frequency analysis is most effective when transport matrices at different temporal frequencies are related, so that analyzing one tells us something about the others. Fortunately strong correlations do exist, and we use three of them here. First and foremost, element  $\mathbf{T}_{ij}^\omega$  represents the same physical 3D transport path(s) from projector pixel  $j$  to camera pixel  $i$  regardless of the frequency  $\omega$ . Thus, if it represents direct or caustic transport at one frequency, it will do so at all others.<sup>3</sup> Second, a direct path between these pixels contributes a Dirac peak to pixel  $i$ ’s temporal profile, which has a

<sup>3</sup>See O’Toole et al. [2012; 2014] for an analysis of the block-diagonal structure of  $\mathbf{T}$  and its relation to individual transport components.

---

#### Algorithm 1 Acquire a PMD photo for illumination pattern $\mathbf{p}$ .

---

**In:** frequency  $\omega$  and real-valued spatial illumination pattern  $\mathbf{p}$

**Out:** photo equal to  $\mathbf{i}^\omega = \mathbf{T}^\omega \mathbf{p}^\omega$

- 1: given frequency  $\omega$ , set hardware-defined modulation functions  $f(t)$ ,  $g(t)$  such that  $h(\tau) = (f * g)(\tau) = \cos(\omega\tau) + b$  for some arbitrary constant offset  $b$
  - 2: define phase delay vector  $\phi = [0, \frac{\pi}{2\omega}, \frac{\pi}{\omega}, \frac{3\pi}{2\omega}]$
  - 3: **for**  $d = 1$  to 4 **do**
  - 4: display pattern  $\mathbf{p}$
  - 5: modulate sensor and source with  $f(t)$ ,  $g(t - \phi_d)$  so that  $h(\tau) = \cos(\omega(\tau + \phi_d)) + b$
  - 6: capture image  $\mathbf{i}_d^\omega$  satisfying Equation 14
  - 7: **end for**
  - 8: set  $\mathbf{i}^\omega = (\mathbf{i}_1^\omega - \mathbf{i}_3^\omega) + i(\mathbf{i}_2^\omega - \mathbf{i}_4^\omega)$
  - 9: **return** captured PMD photo  $\mathbf{i}^\omega$
- 

flat spectrum. We can therefore use element  $\mathbf{T}_{ij}^\omega$  to predict amplitude at all frequencies, and to predict phase up to a discrete (phase-unwrapping) ambiguity. Third, a similarly strong correlation occurs for elements representing caustic paths, which also produce Dirac temporal peaks in typical settings.

In contrast, correlations are much weaker when the transport between pixels  $j$  and  $i$  is non-caustic indirect. Such transport often involves a broad distribution of path lengths and contributes temporal profiles with complex shape and small spectral support. This makes it hard to predict phase and amplitude far from  $\omega$ .

## 4 Implementation: Hardware & Algorithms

**Basic imaging procedure** The working principle behind photonic mixer devices (PMDs) is the synchronous modulation of both the incident and outgoing light at high frequencies. Modulating the incident light by function  $f(t)$  and projecting pattern  $\tilde{\mathbf{p}}(t) = g(t)\mathbf{p}$ , with temporal modulation function  $g(t)$  and 2D spatial pattern  $\mathbf{p}$ , yields the following image formation model:

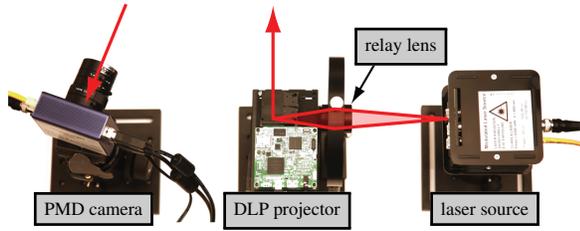
$$\begin{aligned} \mathbf{i} &= \int_0^{NT} f(t) \int_{-\infty}^{\infty} \tilde{\mathbf{T}}(\tau) \tilde{\mathbf{p}}(t - \tau) d\tau dt \\ &= \left[ \int_{-\infty}^{\infty} \tilde{\mathbf{T}}(\tau) \left( N \int_0^T f(t)g(t - \tau) dt \right) d\tau \right] \mathbf{p} \\ &= \left[ N \int_{-\infty}^{\infty} \tilde{\mathbf{T}}(\tau) h(\tau) d\tau \right] \mathbf{p} \end{aligned} \quad (14)$$

where the function  $h(\tau)$  is the convolution between the two modulation functions  $f(t)$  and  $g(t)$ ,  $T = \frac{1}{\omega}$  is the modulation period, and the integer  $N$  is the number of periods captured during a single exposure. This becomes exactly Equation 2 when the convolution function is chosen to be the complex exponential  $h(\tau) = e^{-2\pi i \tau \omega}$ .

In practice, the modulation functions and their convolution have non-negative, real values. The basic imaging procedure of a PMD camera synthesizes images for mean-zero, complex-valued  $h(\tau)$  by capturing and linearly combining four images: two for the real term and two for the imaginary term. See Algorithm 1 for the basic imaging procedure when the illumination pattern  $\mathbf{p}$  is real valued,<sup>4</sup> and refer to Figures 12 and 13 of the supplemental material for illustrations of the procedure.

In real PMD cameras, the specific modulation functions  $g(t)$  and  $f(t)$  are usually determined by hardware constraints, and cannot

<sup>4</sup>Illuminating the scene with a complex pattern  $\mathbf{p}^\omega$  is not physically realizable. It can be simulated, however, by capturing four PMD photos in a way similar to Algorithm 1. We do not use such patterns in our experiments.



**Figure 5:** Overhead view of our prototype. The modulated laser source (right) emits light that passes through a lens and the existing optics of a DLP projector (middle) on its way to the scene. The PMD camera (left) captures the light returning from the scene.

be chosen arbitrarily. In particular,  $h(\tau)$  is not always an exact sinusoid. We note, however, that, if  $g(t)$  and  $f(t)$  are periodic with frequency  $\omega$ , then so is  $h(\tau)$ . Therefore  $h(\tau)$  is a superposition of the base frequency  $\omega$  and its harmonics. In this case, the transport matrices are simply a weighted sum of the transport matrices for the base frequency and its harmonics. In practice, we conduct the bulk of our analysis at frequency  $\omega_1 = 100\text{MHz}$  (i.e., depth acquisition and phase estimation for direct and caustic paths), where our prototype’s deviation from a perfect sinusoid is negligible.

**Hardware** Figure 5 shows an overhead photo of our prototype in its non-coaxial configuration, in which the projector and camera can be thought of as forming a stereo pair. We modified the  $160 \times 120$ -resolution sensor of a PMD PhotonICs 19k-S3 by replacing the internal modulation signal with our own external modulation signals, outputting frequencies that range from 12 MHz to 140 MHz. We illuminate the scene with a custom laser projector, built by replacing the RGB light source of an off-the-shelf DLP LightCrafter projector with the light emitted from six 650 nm laser diodes. A single 40 mm lens (Thorlabs TRS254-040-A-ML) directs the modulated laser illumination through the existing optics of the projector, from which the RGB LEDs and dichroic mirrors were removed. For coaxial camera and projector arrangements, we add a 50/50 beamsplitter (Edmund Optics #46-583) to optically align the projector and the camera.

The exposure time of our PMD camera was strictly limited to the range of 1 to 8 ms. We used a 1 ms exposure time (Step 5 of Algorithm 1) when operating the camera in a stereo arrangement as shown in Figure 5; in coaxial arrangements we increased it to 8 ms to compensate for the system’s 25% ideal light efficiency and for beamsplitter imperfections. Thus, capturing one PMD photo with Algorithm 1 takes 4 or 32 ms, depending on the arrangement.

**Calibration** The sensing behavior of individual pixels is not perfectly uniform over the entire PMD sensor. We model these deviations as an element-wise product between two complex images—the PMD photo of the scene and a 2D “noise” pattern  $\mathbf{w}^\omega$ . The pattern’s amplitude,  $\|\mathbf{w}^\omega\|$ , models the non-uniformity of pixel sensitivities and represents *fixed pattern noise* (FPN). Its phase,  $\arg \mathbf{w}^\omega$ , models pixel-specific offsets in the phase of the sensor’s modulation function, and thus represents *fixed phase-pattern noise* (FPPN). Removing this deterministic noise pattern involves two steps: (1) pre-compute  $\mathbf{w}^\omega$  by capturing a photo of a scene that is constant in both amplitude and phase, and (2) multiply its reciprocal  $(\mathbf{w}^\omega)^{-1}$  with every captured PMD photo.

**Transport matrix probing with a PMD camera** The key to probing comes from approximating Equation 11 with a sum of bilinear matrix-vector products [O’Toole et al. 2012]:

$$(\mathbf{T}^\omega \odot \mathbf{\Pi}) \mathbf{1} \approx \sum_{k=1}^K \mathbf{m}_k \odot \mathbf{T}^\omega \mathbf{p}_k \quad (15)$$

---

**Algorithm 2** Combine PMD imaging and matrix probing.

---

**In:** phase delay  $\phi_d$ ; spatial binary patterns  $\mathbf{p}_1, \dots, \mathbf{p}_K$  and masks  $\mathbf{m}_1, \dots, \mathbf{m}_K$  such that  $\mathbf{\Pi} \approx \sum_1^K \mathbf{m}_k (\mathbf{p}_k)^\top$

**Out:** component  $i_d^\omega$  of PMD photo  $(\mathbf{T}^\omega \odot \mathbf{\Pi}) \mathbf{1}$

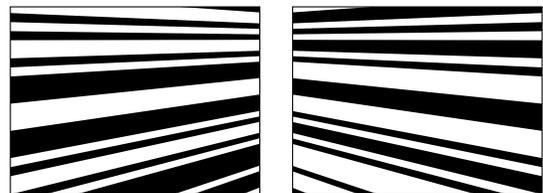
- 1: modulate sensor and source according to  $f(t), g(t - \phi_d)$
  - 2: open camera shutter
  - 3: **for**  $k = 1$  to  $K$  **do**
  - 4:   apply pixel mask  $\mathbf{m}_k$
  - 5:   display pattern  $\mathbf{p}_k$  for  $\frac{1}{K}$ -th the exposure time
  - 6: **end for**
  - 7: close camera shutter
  - 8: **return** image  $i_d^\omega$
- 

where the sequence of vectors  $\mathbf{p}_k$  and  $\mathbf{m}_k$  defines a rank- $K$  approximation of the probing matrix:  $\mathbf{\Pi} \approx \sum_1^K \mathbf{m}_k (\mathbf{p}_k)^\top$ . When  $\omega$  is the DC frequency, O’Toole et al. [2012] showed how to implement Equation 15 directly in the optical domain by capturing just one conventional photo. Briefly, the idea is to treat vectors  $\mathbf{p}_k$  as illumination patterns that are projected onto the scene one by one, each for  $1/K$  of the photo’s exposure. At the same time, the light arriving at individual sensor pixels is modulated by a 2D pixel mask, defined by the vectors  $\mathbf{m}_k$ , which changes in lockstep with the projection patterns. A video-rate implementation of this procedure was recently described in [O’Toole et al. 2014], using a pair of synchronized DLP LightCrafter kits for projection and masking. Their implementation approximates  $\mathbf{T}$  with a sequence of binary patterns and masks and allows up to  $K = 96$  projection/masking operations within the 36 ms exposure time of one video frame.

Generalizing this procedure to arbitrary temporal frequencies  $\omega$  and to complex-valued PMD photos is straightforward: we simply replace Steps 4 and 5 of Algorithm 1 (which acquire images without a mask under a fixed illumination pattern) with Steps 1-8 of Algorithm 2 (which change masks and illumination patterns  $K$  times). Note that this modification of Algorithm 1 does not change the total number of images captured, which remains equal to four. Unfortunately, hardware constraints prevented us from implementing Algorithm 2 exactly as shown, impacting the number of images we capture in experiments. We return to this at the end of Section 4.

**Mask/pattern sequences for indirect-only imaging** The imaging behavior of a matrix probing operation is critically dependent on the precise definition of the mask and pattern sequences. We refer interested readers to [O’Toole et al. 2012; O’Toole et al. 2014] for detailed derivations and in-depth discussion of specific cases. Here we concentrate on the binary mask/pattern sequences for indirect-only imaging, which are used extensively in our PMD experiments.

In particular, let us first consider the case of a stereo projector-camera arrangement. Following [O’Toole et al. 2014], we compute the epipolar geometry between the projector and the PMD camera and construct each mask  $\mathbf{m}_k$  by randomly turning each of its epipo-



**Figure 6:** Mask/pattern pairs for indirect-only imaging. Corresponding epipolar lines are never “on” at the same time in the randomly-generated mask (left) and its projection pattern (right).

---

**Algorithm 3** *Acquire a direct/retro-reflective PMD photo.*

---

**In:** frequency  $\omega$  and sequence length  $K$

**Out:** direct/retro-reflective PMD photo  $\mathbf{i}_{\text{direct}}^\omega$

- 1: acquire conventional PMD photo  $\mathbf{i}^\omega$  using Algorithm 1 and an all-white pattern
  - 2: construct indirect-only binary sequences  $\mathbf{p}_1, \dots, \mathbf{p}_K$  and  $\mathbf{m}_1, \dots, \mathbf{m}_K$
  - 3: acquire indirect-only PMD photo  $\mathbf{i}_{\text{indirect}}^\omega$  using Algorithms 1 and 2
  - 4: **return** image  $\mathbf{i}^\omega - \mathbf{i}_{\text{indirect}}^\omega$
- 

---

**Algorithm 4** *Acquire low-/high-frequency indirect PMD photos.*

---

**In:** frequency  $\omega$  and sequence lengths  $J$  and  $K$

**Out:** photos  $\mathbf{i}_{\text{low}}^\omega$  and  $\mathbf{i}_{\text{high}}^\omega$  containing low- and high-frequency indirect transport components, respectively

- 1: construct indirect-only binary sequences  $\mathbf{p}_1, \dots, \mathbf{p}_K$  and  $\mathbf{m}_1, \dots, \mathbf{m}_K$
  - 2: construct high-frequency pattern sequence,  $\mathbf{q}_1, \dots, \mathbf{q}_J$
  - 3: set  $\mathbf{i}_{\text{low}}^\omega = \mathbf{i}_{\text{high}}^\omega = 0$
  - 4: **for**  $j = 1$  **to**  $J$  **do**
  - 5: acquire a PMD photo  $\mathbf{i}^\omega$  using Algorithms 1 and 2, the mask sequence  $\mathbf{m}_1, \dots, \mathbf{m}_K$  and the pattern sequence  $\mathbf{p}_1 \odot \mathbf{q}_j, \dots, \mathbf{p}_K \odot \mathbf{q}_j$
  - 6: set  $\mathbf{i}_{\text{low}}^\omega = \min(\mathbf{i}_{\text{low}}^\omega, \mathbf{i}^\omega)$
  - 7: set  $\mathbf{i}_{\text{high}}^\omega = \max(\mathbf{i}_{\text{high}}^\omega, \mathbf{i}^\omega)$
  - 8: **end for**
  - 9: set  $\mathbf{i}_{\text{high}}^\omega = \mathbf{i}_{\text{high}}^\omega - \mathbf{i}_{\text{low}}^\omega$
  - 10: set  $\mathbf{i}_{\text{low}}^\omega = \frac{1}{\alpha} \mathbf{i}_{\text{low}}^\omega$
  - 11: **return** separated components  $\mathbf{i}_{\text{low}}^\omega, \mathbf{i}_{\text{high}}^\omega$
- 

lar lines “on” or “off” with probability 0.5. Given mask  $\mathbf{m}_k$ , we construct the corresponding pattern  $\mathbf{p}_k$  by turning “on” all epipolar lines that were turned “off” in the mask, and vice-versa. Intuitively, such a mask/pattern pair is guaranteed to block all direct light paths from projector to camera because (1) direct paths always satisfy the epipolar constraint and (2) corresponding epipolar lines are never “on” simultaneously on the projector and camera planes. Figure 6 shows such an example. In coaxial projector-camera arrangements, where epipolar geometry is degenerate, we apply this construction to the rows of an image instead of its epipolar lines.

**Acquiring direct/retro-reflective PMD photos** The basic procedure is shown in Algorithm 3. It amounts to capturing a conventional PMD photo and then subtracting its indirect-only component.

**Fast separation of low/high-frequency transport components** We further decompose the indirect component of a PMD photo into its low- and high-frequency components, corresponding to caustic indirect and non-caustic indirect paths, respectively. We do this with the approach suggested by O’Toole et al. [2012]. The idea is to modify the indirect-only mask/pattern sequences in a way that combines transport matrix probing with the frequency-based separation technique of Nayar et al. [2006]. Algorithm 4 shows the basic steps, adapted to the case of PMD imaging.

**Depth acquisition from direct/retro-reflective PMD photos** The phase component of a PMD pixel encodes the depth of each scene point as a value that ranges from 0 to  $2\pi$ . Specifically, a coaxial system produces pixel values of the following form:

$$\mathbf{i}^\omega = ae^{-2\pi id \frac{2\omega}{c}} \quad (16)$$

where  $a$  is the albedo of a scene point,  $d$  is its depth,  $2d$  is the round-trip distance travelled by light to the camera, and  $c$  is the speed of light. This produces ambiguities in the relation between phase and depth. For example, frequency  $\omega_1 = 100$  MHz only encodes depth for a maximum unambiguous range of  $\frac{c}{2\omega_1} \approx 1.5$  m. For a greater depth range, we acquire direct/retro-reflective PMD photos for two frequencies  $\omega_1$  and  $\omega_2 = \omega_1/2$  and use phase-unwrapping [Mei et al. 2013] to calculate depth. Specifically, given photos  $\mathbf{i}^{\omega_1}$  and

PMD imaging task	Optical masking	Computational masking	Experiments
illumination pattern $\mathbf{p}$	4	4	4
indirect-only	4	$4K$	512
direct/retro-reflective	8	$4 + 4K$	516
low-/high-freq. indirect	$4J$	$4KJ$	3072
depth acquisition	8	$4 + 4K$	516
transport decomposition	$8 + 4J$	$4 + 4K + 4KJ$	3588
light-in-flight imaging	$8 + 4J + 4F$	$4 + 4K + 4KJ + 4F$	3718

**Table 2:** *Images required for transient transport analysis. We use  $K = 128, J = 6, F = 65$  in our experiments, with  $\omega_1 = 100$  MHz and no phase unwrapping. See Sections 4 and 5 for explanations.*

$\mathbf{i}^{\omega_2}$ , the phase-unwrapped depth is

$$d = \frac{c}{2\omega_1} \left\{ 2 - \frac{\arg \mathbf{i}^{\omega_1}}{2\pi} - \left\lfloor \frac{\pi + 2 \arg \mathbf{i}^{\omega_2} - \arg \mathbf{i}^{\omega_1}}{2\pi} \right\rfloor \right\} \quad (17)$$

Since the scenes in our experiments were well within 1.5 m, imaging at a single frequency  $\omega_1$  with no unwrapping was sufficient.

**Hardware limitations and computational masking** The 8 ms maximum exposure time of our PMD camera prevented us from implementing Algorithm 2 in one shot. This is because our DLP kits can perform at most  $K = 21$  projection/masking operations in that interval, leading to much poorer approximations of Equation 15 compared to the 96 patterns that fit in a 36 ms video frame. To overcome this limitation we mask images computationally, by pushing Steps 2 and 7 of Algorithm 2 inside the loop. In particular, we capture  $K = 128$  images individually, each with a 1 ms exposure; we multiply element-wise the image captured in the  $k$ -th iteration with the associated mask  $\mathbf{m}_k$ ; and accumulate the result. This increased significantly the number of images we had to capture for the experiments in Section 5. Table 2 gives full details.

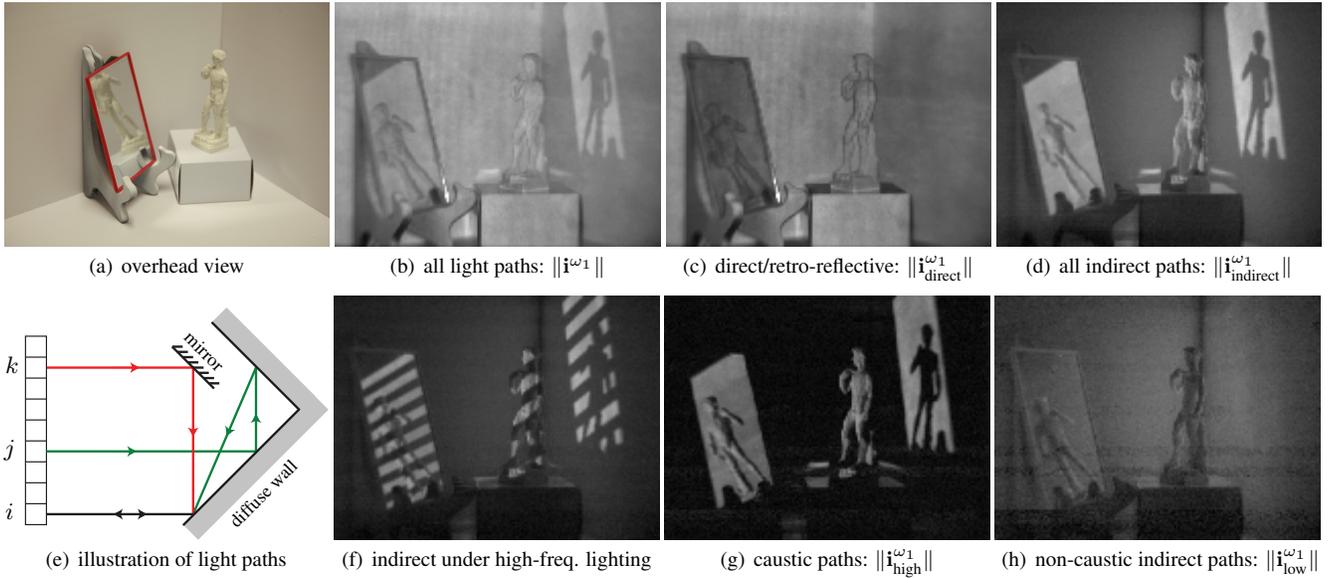
## 5 Results

**Transport decomposition for time-of-flight imaging** We start with separating a scene’s transient appearance into its three transport components—direct/retro-reflective, caustic, and non-caustic indirect. Since both the direct/retro-reflective and the caustic components are due to distinct temporally-isolated reflection events, they correspond to Dirac peaks in the time domain. We run Algorithm 3 for frequency  $\omega_1$  to localize the former and Algorithm 4 for the same frequency to identify and localize the latter. This also gives us the non-caustic indirect contributions for frequency  $\omega_1$ .

Figure 7 shows this decomposition for a scene containing a mirror and a miniature statue of Michelangelo’s David positioned near the corner of a room. We used a coaxial projector-camera arrangement for this example.

**Time-of-flight depth images robust to indirect transport** PMD cameras compute depth by acquiring PMD photos for one or more frequencies with a co-located light source, and then using Equation 17 to turn phases into depth values. An unfortunate consequence of this approach is that indirect light has a pronounced influence on the measurements. Though methods exist for removing the influence of indirect light from a PMD image computationally [Fuchs 2010; Godbaz et al. 2012; Jiménez et al. 2014], these methods rely on predictive models that do not generalize for handling all forms of indirect light.

We demonstrate the ability to recover accurate depth images that are robust to indirect light transport using PMD cameras. Specifically, we use Algorithm 3 to capture the direct/retro-reflective PMD photo of a scene for frequency  $\omega_1$ . This photo is by definition invariant to indirect transport so its phase yields transport-robust depth maps.



**Figure 7:** Capturing PMD photos corresponding to individual transport components. Since these photos are complex, we only show their amplitudes in this figure. (a) A scene containing a mirror and a miniature statue of David. (b) PMD photo returned by Algorithm 1 for an all-white illumination pattern. (c) PMD photo returned by Algorithm 3. (d) PMD photo returned by Algorithm 2 for mask/pattern pairs similar to those in Figure 6. (e) Basic path geometries for a camera pixel  $i$ : direct/retroreflective light path  $i \rightarrow i$  (black), caustic path  $k \rightarrow i$  (red), and non-caustic indirect path  $j \rightarrow i$  (green). (f) One of six PMD photos acquired in Step 5 of Algorithm 4. This photo is an indirect-only transient view of the scene under a projection pattern consisting of binary stripes. (g-h) PMD photos returned by Algorithm 4.

Figure 8 shows depth acquisition results for three scenes with significant low- and high-frequency indirect transport. The scene in row 1 is a bowl pressed flush against a planar wall (Figure 9). The deep concavity of the bowl produces a significant amount of indirect light through diffuse inter-reflections. As a result, the “raw” time-of-flight measurements place the base of the bowl 4 to 5 cm behind the wall itself (columns (a) and (c)). Our approach reconstructs a physically-valid solution, with the bowl’s base coinciding with the expected position of the wall (columns (b), (d) and (e)). Note that the reconstructions of the convex handle of the bowl are the same in both cases (lower-left region of the slice in column (e)). The concave lip, on the other hand, in the conventionally-acquired time-of-flight image is offset from its correct position (lower-right region of the slice in column (e)).

Row 2 of Figure 8 shows reconstruction results for an open book whose pages were kept as flat as possible. The significant diffuse inter-reflections between these pages distort their conventionally-acquired shape (columns (a) and (c)) but leave direct/retro-reflective shape measurements unaffected: their flat shape is evident in columns (b), (d) and (e). Row 3 shows results for the David scene. The caustic light paths, specularly reflected by the mirror onto the right wall, produce an embossed silhouette on the wall, indicated by the yellow arrow in column (c). In contrast, the direct/retro-reflective component is not influenced by caustic paths—and also recovers the depth of objects viewed within the mirror itself.

### Distinguishing between direct views and mirror reflections

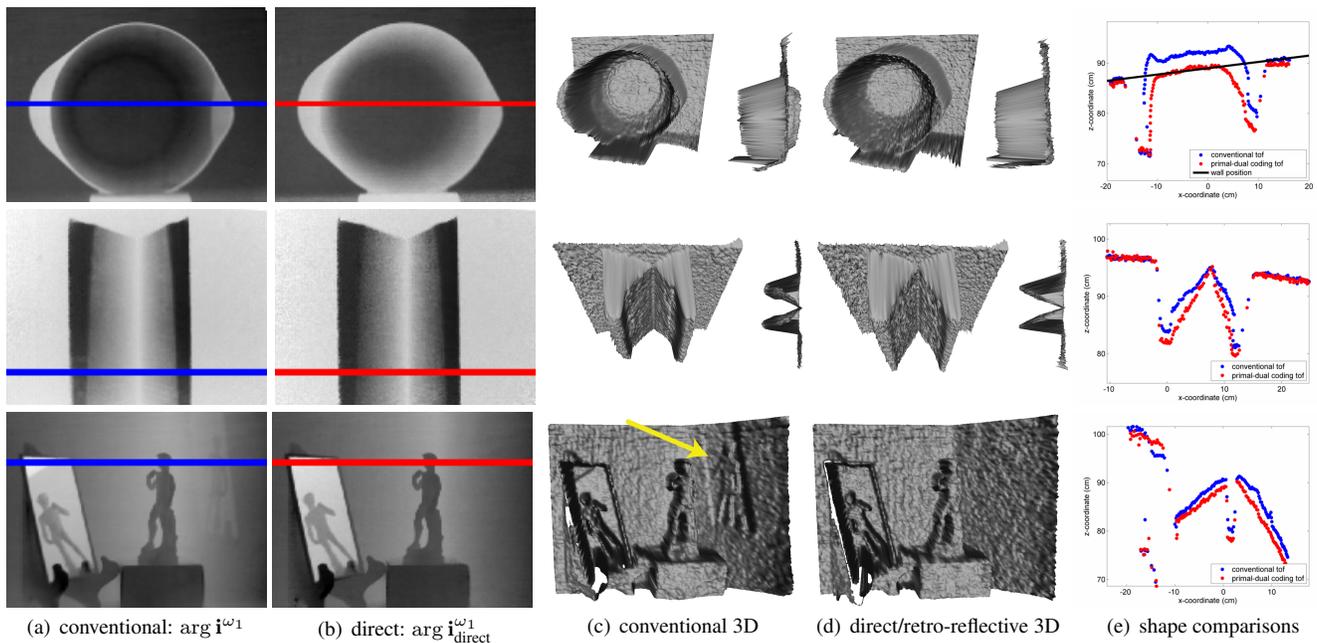
We can use the phase of PMD photos to classify pixels according to whether they receive light directly or via a retro-reflective path (e.g., as would occur if we viewed a diffuse point through a mirror). In particular, when the projector and camera are coaxial, direct light paths are always the shortest paths received by each camera pixel. Retro-reflective paths, on the other hand, do not have this property: if a pixel receives light from both retro-reflective and caustic light paths, the caustic paths always take the shortest route to the camera. This turns pixel classification into a simple pixel-wise comparison

of phase values in direct/retro-reflective and caustic PMD photos. See Figure 10 for a detailed demonstration.

**Capturing evolving wavefronts of “light-in-flight”** PMD cameras provide a cheap and light-efficient alternative to capturing transient images [Heide et al. 2013]. Unfortunately, the maximum modulation frequency is limited by PMD hardware, making reconstructing transient images a highly ill-conditioned problem, and requiring strong regularizers to perform the reconstruction.

To overcome these limitations, we propose a transport-specific reconstruction for each of the direct/retro-reflective, caustic, and non-caustic transport components of a scene. In particular, we reconstruct the temporal intensity profiles for direct/retro-reflective and caustic light paths by running Algorithms 3 and 4 at frequency  $\omega_1$ , and fitting a Dirac peak to each pixel in  $i_{\text{direct}}^{\omega_1}$  and  $i_{\text{high}}^{\omega_1}$ . To reconstruct the non-caustic indirect wavefront of a scene, we (1) capture conventional PMD photos at  $F = 65$  frequencies from 12 to 140 MHz in 2 MHz increments, (2) subtract the predicted contribution of the direct and caustic Dirac peaks from these photos, and (3) apply the reconstruction method of Heide et al. [2013] to recover the temporal intensity profile due to non-caustic indirect light. Please refer to Figure 14 in the supplemental material for more details on why subtracting the predicted direct/caustic contributions yields more accurate estimates of this profile.

Figure 11 compares our approach to that of Heide et al. [2013], which does not perform transport decomposition. The first scene (rows 1 and 2) contains transparent objects that refract the wavefront traveling through them. The second scene (rows 3 and 4) produces strong diffuse inter-reflections from two large bowls that are positioned near a corner. The third scene (rows 5 and 6) includes a mirror and a jug filled with milky water; these produce caustic and retro-reflected light paths via the mirror, as well as volumetric scattering through the water. Note that the direct and caustic wavefronts propagating through the scene are well-resolved in both space and time in our reconstructions, whereas they appear broad and poorly-localized in the absence of transport decomposition.



**Figure 8:** Geometry results for the three scenes in Figure 7(a) and 9. (a) The phase of conventionally-acquired PMD photos. (b) The phase of direct/retro-reflective PMD photos returned by Algorithm 3. (c-d) Views of the 3D meshes computed from (a) and (b), respectively. (e) Plots of the  $x$ - and  $z$ -coordinates for a slice of each scene, computed from the conventional (blue) and the direct/retro-reflective (red) phases. Observe that the base of the conventionally-acquired bowl protrudes through the back wall by about 5 cm; the pages of the conventionally-acquired book appear curved; the corner of the room in the conventionally-acquired David scene is rounded, and the caustic paths illuminating the room’s right wall produce a 2 to 3 cm offset in depth values. None of these artifacts appear in (b) or (d).

## 6 Conclusion

In this paper we uncovered a key mathematical link between time-of-flight and conventional photography, making it possible to readily transfer computational illumination techniques from one domain to the other. The approach hinges on an ability to “probe” scenes by illuminating them with coded space-time patterns that vary sinusoidally in time, with the same frequency at every pixel. Technology with this capability built-in is already entering the consumer market in the form of off-the-shelf devices such as PMD cameras and the Kinect 2.

On the practical side, we expect the biggest immediate impact of our work to be in the design of time-of-flight depth cameras that are robust to indirect light. Although time-of-flight imaging is already proving to be a superior technology for depth acquisition, this is one area where conventional imaging still has the edge. Our experimental results, although preliminary, suggest that this does not have to be the case.

At a more fundamental level, the most exciting avenue for future work is the design of new 3D scene analysis techniques that integrate both geometric and time-of-flight constraints on light paths. Distinguishing between direct views and mirror reflections of an object—which is impossible to do from just stereo or time-of-flight constraints—suggests that perhaps a great deal of untapped scene information is lurking around the corner.

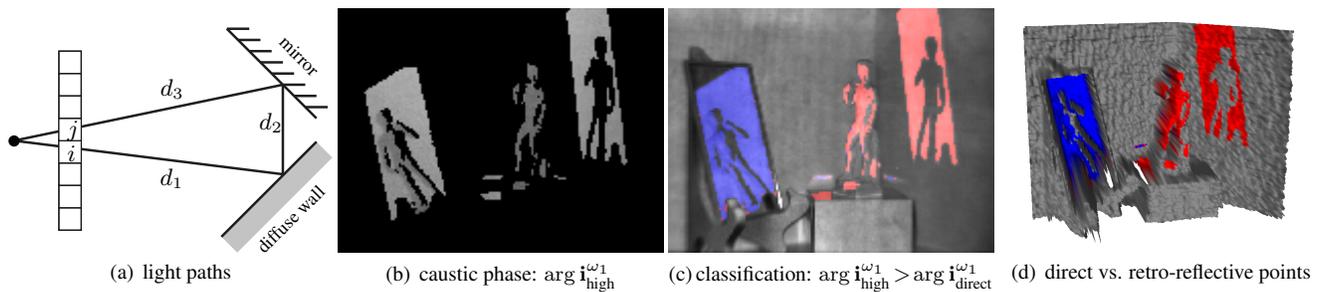
**Acknowledgements** The authors gratefully acknowledge the support of the Natural Sciences and Engineering Research Council of Canada under the PGS-D, RGPIN, RTI, SGP and GRAND NCE programs. The authors also thank all anonymous reviewers for their helpful comments.



**Figure 9:** Conventional photos of scenes reconstructed in Figure 8. We used a coaxial projector-camera arrangement for these scenes.

## References

- BAI, J., CHANDRAKER, M., NG, T.-T., AND RAMAMOORTHI, R. 2010. A dual theory of inverse and forward light transport. In *Proc. ECCV*, 294–307.
- DEBEVEC, P., HAWKINS, T., TCHOU, C., DUIKER, H.-P., SAROKIN, W., AND SAGAR, M. 2000. Acquiring the reflectance field of a human face. *ACM SIGGRAPH*, 145–156.
- DORRINGTON, A., GODBAZ, J., CREE, M., PAYNE, A., AND STREETER, L. 2011. Separating true range measurements from multi-path and scattering interference in commercial range cameras. In *Proc. SPIE*, vol. 7864.
- FUCHS, S. 2010. Multipath interference compensation in time-of-flight camera images. In *Proc. ICPR*, 3583–3586.
- GARG, G., TALVALA, E.-V., LEVOY, M., AND LENSCH, H. P. 2006. Symmetric photography: exploiting data-sparseness in reflectance fields. In *Proc. EGSR*, 251–262.
- GODBAZ, J. P., CREE, M. J., AND DORRINGTON, A. A. 2012. Closed-form inverses for the mixed pixel/multipath interference problem in amcw lidar. In *Proc. SPIE*, vol. 8296.



**Figure 10:** Distinguishing between direct views and mirror reflections. (a) For a co-located camera and projector, the direct light path  $i \rightarrow i$  has length  $2d_1$ , the caustic paths  $i \rightarrow j$  and  $j \rightarrow i$  have lengths  $d_1 + d_2 + d_3$ , and the retroreflective path  $j \rightarrow j$  has length  $2(d_2 + d_3)$ . It follows that caustic paths are always longer than direct paths and shorter than retro-reflective paths. (b) The phase of the caustic component of the David scene, with phases zeroed-out at pixels where the amplitude  $\|i_{high}^{\omega_1}\|$  is low, and thus phase is uninformative. (c) The result of a pixel-wise “greater than” operator between the caustic PMD photo shown in (b) and the direct/retro-reflective PMD photo in row 3 of Figure 8(b). Retro-reflective light paths are shown in blue and direct light paths in red. Note that we cannot distinguish between direct and retro-reflective pixels in the absence of a caustic component. (d) Color-coded mesh using the conventions in (c).

- GORAL, C. M., TORRANCE, K. E., GREENBERG, D. P., AND BATTAILE, B. 1984. Modeling the interaction of light between diffuse surfaces. *ACM SIGGRAPH*, 213–222.
- GUPTA, M., AND NAYAR, S. K. 2012. Micro phase shifting. In *Proc. ICCV*, 813–820.
- HEIDE, F., HULLIN, M. B., GREGSON, J., AND HEIDRICH, W. 2013. Low-budget transient imaging using photonic mixer devices. *ACM SIGGRAPH*, 45:1–45:10.
- HLAWITSCHKA, M., EBLING, J., AND SCHEUERMANN, G. 2004. Convolution and fourier transform of second order tensor fields. In *Proc. IASTED VIIP*, 78–83.
- JIMÉNEZ, D., PIZARRO, D., MAZO, M., AND PALAZUELOS, S. 2014. Modeling and correction of multipath interference in time of flight cameras. *Image and Vision Computing* 32, 1, 1–13.
- KADAMBI, A., WHYTE, R., BHANDARI, A., STREETER, L., BARSİ, C., DORRINGTON, A., AND RASKAR, R. 2013. Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles. *ACM Trans. Graph.* 32, 6, 167:1–167:10.
- KAJIYA, J. T. 1986. The rendering equation. *ACM SIGGRAPH*, 143–150.
- KIRMANI, A., HUTCHISON, T., DAVIS, J., AND RASKAR, R. 2011. Looking around the corner using ultrafast transient imaging. *Int. J. of Computer Vision* 95, 13–28.
- KIRMANI, A., COLACO, A., WONG, F. N. C., AND GOYAL, V. K. 2012. Codac: a compressive depth acquisition camera framework. In *Proc. ICASSP*, 5425–5428.
- KIRMANI, A., VENKATRAMAN, D., SHIN, D., COLACO, A., WONG, F. N. C., SHAPIRO, J. H., AND GOYAL, V. K. 2013. First-photon imaging. *Science*.
- MEI, J., KIRMANI, A., COLACO, A., AND GOYAL, V. 2013. Phase unwrapping and denoising for time-of-flight imaging using generalized approximate message passing. In *Proc. ICIP*, 364–368.
- NAYAR, S. K., KRISHNAN, G., GROSSBERG, M. D., AND RASKAR, R. 2006. Fast separation of direct and global components of a scene using high frequency illumination. *ACM SIGGRAPH*, 935–944.
- NG, R., RAMAMOORTHY, R., AND HANRAHAN, P. 2003. All-frequency shadows using non-linear wavelet lighting approximation. *ACM SIGGRAPH*, 376–381.
- O’TOOLE, M., AND KUTULAKOS, K. N. 2010. Optical computing for fast light transport analysis. *ACM Trans. Graph.*, 164:1–164:12.
- O’TOOLE, M., RASKAR, R., AND KUTULAKOS, K. N. 2012. Primal-dual coding to probe light transport.
- O’TOOLE, M., MATHER, J., AND KUTULAKOS, K. N. 2014. 3D shape and indirect appearance by structured light transport. In *Proc. CVPR*.
- PEERS, P., MAHAJAN, D. K., LAMOND, B., GHOSH, A., MATUSIK, W., RAMAMOORTHY, R., AND DEBEVEC, P. 2009. Compressive light transport sensing. *ACM Trans. Graph.* 28, 1.
- REDDY, D., RAMAMOORTHY, R., AND CURLESS, B. 2012. Frequency-space decomposition and acquisition of light transport under spatially varying illumination. In *Proc. ECCV*, 596–610.
- SEITZ, S. M., MATSUSHITA, Y., AND KUTULAKOS, K. N. 2005. A theory of inverse light transport. In *Proc. ICCV*, 1440–1447.
- SEN, P., AND DARABI, S. 2009. Compressive dual photography. *Computer Graphics Forum* 28, 2, 609–618.
- SEN, P., CHEN, B., GARG, G., MARSCHNER, S., HOROWITZ, M., LEVOY, M., AND LENSCH, H. P. A. 2005. Dual photography. *ACM SIGGRAPH*, 745–755.
- VELTEN, A., WILLWACHER, T., GUPTA, O., VEERARAGHAVAN, A., BAWENDI, M. G., AND RASKAR, R. 2012. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications*.
- VELTEN, A., WU, D., JARABO, A., MASIA, B., BARSİ, C., JOSHI, C., LAWSON, E., BAWENDI, M., GUTIERREZ, D., AND RASKAR, R. 2013. Femto-photography: capturing and visualizing the propagation of light. *ACM Trans. Graph.* 32, 4.
- WETZSTEIN, G., AND BIMBER, O. 2007. Radiometric compensation through inverse light transport. 391–399.
- WU, D., O’TOOLE, M., VELTEN, A., AGRAWAL, A., AND RASKAR, R. 2012. Decomposing global light transport using time of flight imaging. In *Proc. CVPR*, 366–373.
- WU, D., WETZSTEIN, G., BARSİ, C., WILLWACHER, T., DAI, Q., AND RASKAR, R. 2013. Ultra-fast lensless computational imaging through 5d frequency analysis of time-resolved light transport. *Int. J. of Computer Vision*, 1–13.



(a) PMD image (b) transient frame ( $t = 1.9$  ns) (c) transient frame ( $t = 3.2$  ns) (d) transient frame ( $t = 5.0$  ns)

**Figure 11:** Transient image comparisons between Heide et al. [2013] and our approach. These scenes were all imaged with a stereo projector-camera arrangement. (a) Steady state images of the scene captured with a normal camera under ambient illumination (rows 1, 3, and 5) and our PMD camera with Algorithm 1 under white projector illumination (rows 2, 4, and 6). (b-d) Frames from the temporal intensity profile reconstructed using the conventional approach (rows 1, 3, and 5) and our approach (rows 2, 4, and 6). Note the sharp Dirac impulses travelling along the walls in our reconstructions, which meet in the corner of the scene. These correspond to direct transport, although sharp caustic wavefronts also occur in some cases (rows 2 and 6). Moreover, even though reconstructing the non-caustic indirect time profile remains highly ill-conditioned, reconstructing the direct and caustic contributions separately simplifies the reconstruction process for the non-caustic indirect component, and improves its accuracy. This is most evident in column (b), where contributions from diffuse scattering and inter-reflections appear to occur throughout the three scenes in the conventional reconstructions. This is physically impossible, however, since the elapsed time is too brief for light to have actually reached those regions. In contrast, non-caustic indirect components are dark in our reconstructions and appear to trail the direct wavefronts.