Scalable Imaging VLC Receivers with Token-Based Pixel Selection for Spatial Multiplexing*

Jimmy C. Chau and Thomas D.C. Little
Department of Electrical and Computer Engineering
Boston University, Boston, Massachusetts
{*jchau,tdcl*}*@bu.edu*

September 18, 2014

MCL Technical Report No. 09-18-2014

Abstract– Visible light communications (VLC) can be adopted in lighting infrastructure to provide ubiquitous wireless access in spaces where light is consumed by humans. Unfortunately, the requirement to provide high-quality diffuse illumination reduces the potential capacity of the free space links and we seek ways to accommodate both the lighting and data rate goals. We investigate the combination of multiple VLC transmissions through spatial multiplexing (SM), a MIMO technique, to meet our data rate goals. Specifically, this paper deals with receiver designs intended to receive and decode increasing numbers of SM/MIMO VLC data streams.

Conventional imaging (camera) sensors have been used as VLC receivers; however, they are intended to capture frames at relatively low speeds and their architectures do not translate well for receiving multiple high-rate VLC streams. Thus we consider new techniques to optimize imaging receivers to meet the capacity requirements of multiple SM streams.

In this paper, we propose token-based pixel selection (TBPS) for CMOS image sensors as a scalable alternative to mitigate this decrease in sampling rate. We show that in many cases, TBPS-based image sensors sample transmissions several times more frequently than windowing image sensors, yielding higher VLC data rates. Assuming the same reset, integration, and read times, TBPS always performs as well as, and often better than, windowing.

*In *Proceedings of the 1st ACM MobiCom Workshop on Visible Light Communication Systems*, Maui, Hawaii, USA, September 2014, DOI: 10.1145/2643164.2643172. This work was supported primarily by the Engineering Research Centers Program of the National Science Foundation under NSF Cooperative Agreement No. EEC-0812056.

1 Introduction

As mobile devices and networked appliances become more prevalent, we face a growing need for ubiquitous network connectivity and wireless communication capacity. To meet this demand, we aim to deliver high-speed networking through indoor lighting with visible light communication (VLC) [3]. In VLC systems, data is conveyed wirelessly from an optical transmitter to an optical receiver as modulated light. Light-emitting diodes (LEDs), which are becoming common as an energy-efficient light source, are able to modulate light (turn on and off or change intensity) quickly (with bandwidths as high as $20\ \mathrm{MHz}$ [4]), allowing them to serve as high-speed VLC transmitters in addition to providing illumination.

Although the achievable throughput of a single VLC transmitter is substantial, it remains a bottleneck in VLC systems. In comparison, VLC receivers are much faster: able to achieve bandwidths of several GHz [5]. Furthermore, in dual-use VLC-lighting systems, lighting quality and energy-efficiency requirements can constrain VLC transmitters, reducing their individual throughput. Fortunately, the abundance of indoor lights (and hence, the abundance of VLC transmitters) provides opportunities to increase throughput many times through spatial multiplexing (SM). To better utilize the throughput of VLC receivers and to achieve our goal of providing gigabit-per-second VLC links, we propose that each VLC receiver should simultaneously receive from multiple transmitters.

The remainder of this paper is organized as follows: Section 2 explains spatial multiplexing and its receiver requirements in mobile use cases. Section 3 examines image sensor architectures to highlight the challenges of creating scalable imaging VLC receivers for spatial multiplexing. To address this scalability challenge, we present a token-based runtime pixel selection mechanism in Section 4. The performance of the proposed architecture is analyzed in Section 5. Section 6 concludes the paper.

2 Spatial Multiplexing with Imaging VLC Receivers

Assuming a linear and time-invariant channel with additive and white Gaussian noise, a single-input and single-output link's capacity is limited by Shannon's capacity theorem:

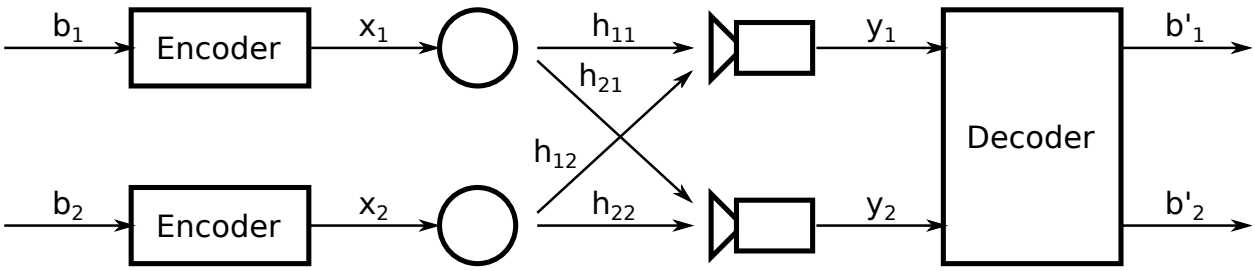
$$C = B \log_2(1 + S/N) \tag{1}$$

where B is the bandwidth and S/N is the signal-to-noise ratio.

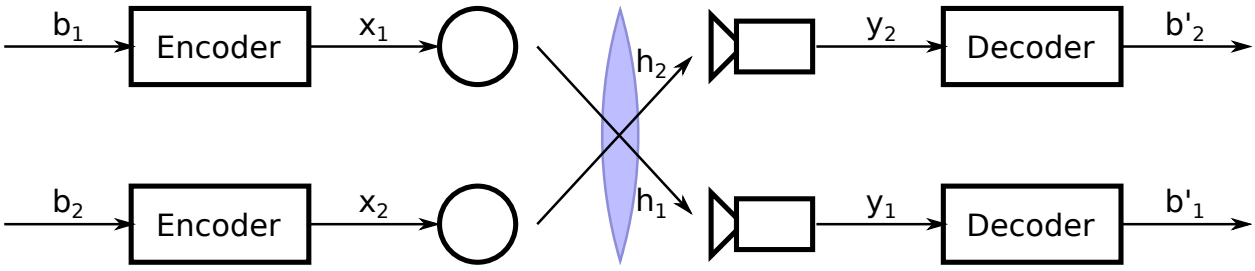
Ideally, as with wired links, the capacity of this link can be multiplied through replication, resulting in a link capacity, C_M, equal to the sum of each individual link capacity, c_i:

$$C_M = \sum_{i=1}^{M} c_i = \sum_{i=1}^{M} B_i \log_2(1 + S_i/N_i) \tag{2}$$

Although this capacity can be realized when the M links are independent, crosstalk can limit the number of independent channels despite the number of transmitter-receiver pairs. For example, if $h_{11} = h_{22} = h_{12} = h_{21}$ in the 2x2 MIMO system shown in Figure 1a, we effectively only have one channel: $y_1 = y_2 = h_{11}(x_1 + x_2)$. In this case, using two transmitter-receiver pairs did not double the link capacity. Fortunately, as illustrated in Figure 1b, imaging optics (e.g., commonly available



(a) Crosstalk between transmitter-receiver pairs



(b) Imaging lens maintains prevents crosstalk

Figure 1: An imaging lens can focus light to prevent crosstalk between transmitter-receiver pairs.

camera lenses) provide a convenient mechanism to prevent crosstalk between transmitter-receiver pairs.

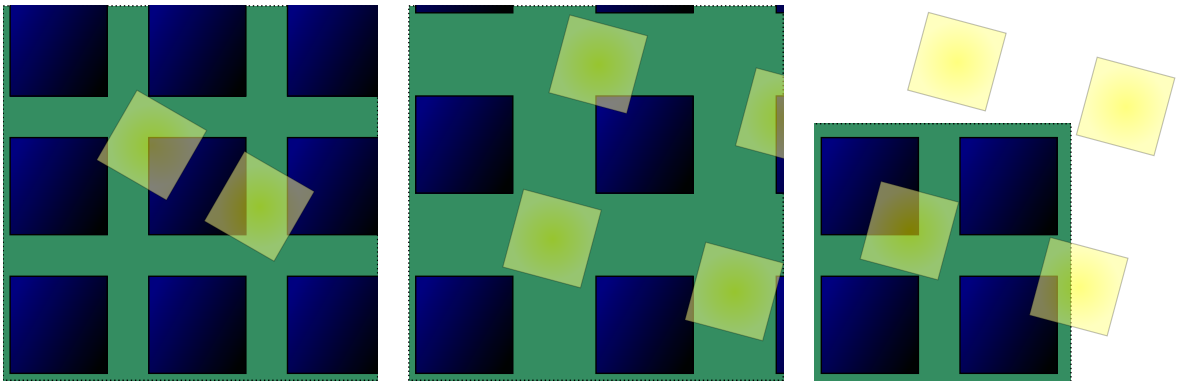
To receive many data streams (independent channels), an imaging VLC receiver needs a high-resolution image sensor. Although a statically positioned VLC system needs only one receiver pixel per transmitter to maximize the number of independent channels (by aligning each receiver pixel to a separate transmitter image), this static alignment is not possible for mobile devices. As illustrated in Figure 2, imaging VLC receivers need many more pixels to support mobility. Otherwise, the imaging VLC receiver may suffer from crosstalk (despite the imaging optics) or may altogether miss certain transmissions, which will reduce the throughput of the resulting link.

Unfortunately, the resolution of existing VLC receiver architectures do not scale up well at the high sampling rates required for VLC; as the image resolution increases, the frame rate of image sensors tend to decrease.

3 Imaging Receiver Architectures

Many imaging VLC receivers share the physical structure illustrated in Figure 3, consisting of an image sensor, optics, and packaging. Their differences are primarily in the design of the image sensor. High-resolution image sensors are typically manufactured as an array of photodetectors on an integrated circuit (IC) with amplifiers, analog-to-digital converters (ADC), control circuitry, and other supporting electronics.

When the number of pixels is small, each pixel can be equipped with its own signal output chain, allowing all pixels to be read in parallel [1]. Unfortunately, high-resolution image sensors tend to have many more pixels than the image sensor can simultaneously read out due to practical limitations, such as the throughput of later signal processing stages (e.g., ADCs and digital signal processors), the number of output pads and pins available on the image sensor's IC and packaging,



(a) Small pixels (b) Dense pitch (c) Wide FOV

Figure 2: To maximize the resulting number of independent channels, an imaging VLC receiver needs: (a) small pixels to prevent interference between transmitter images that land on the same receiver pixel, (b) a dense pixel pitch to receive transmitter images that would otherwise land between pixels, and (c) a large area to cover a wide field of view (FOV) so that transmitter images do not miss the image sensor altogether.

the chip area available for read-out electronics, and the number of signal routing layers available to route signals from every pixel to an output.

For these reasons, this fully parallel approach does not scale well and limits the imaging receiver to very low resolutions. CMOS image sensors typically use row scanning to cope with these constraints.

3.1 Line Scanning and Windowing

Row, or line, scanning image sensors, illustrated in Figure 4, have one shared output per column of pixels, allowing higher resolutions to be supported by fewer outputs. However, these image sensors can only read one row of pixels at a time. As a consequence, the sampling rate for each pixel (i.e., the frame rate), f_s, is inversely proportional to the number of rows, D_R:

$$f_s = f_p / D_R \tag{3}$$

when the bottleneck is the output and each output can read pixels at rate f_p.

Windowing can increase this frame rate by scanning only a subset of rows, reducing the denominator, D_R, in Equation 3 to D_{RW}: the number of rows in the window. For example, the image sensor illustrated in Figure 4 can double its frame rate by scanning only two out of four rows.

However, reading just one pixel from each row would still require scanning the entire image. Due to mobility, transmitter images are often scattered across all (or most) rows of the image sensor. In such cases, reading just a few pixels would be as slow as reading the entire image.

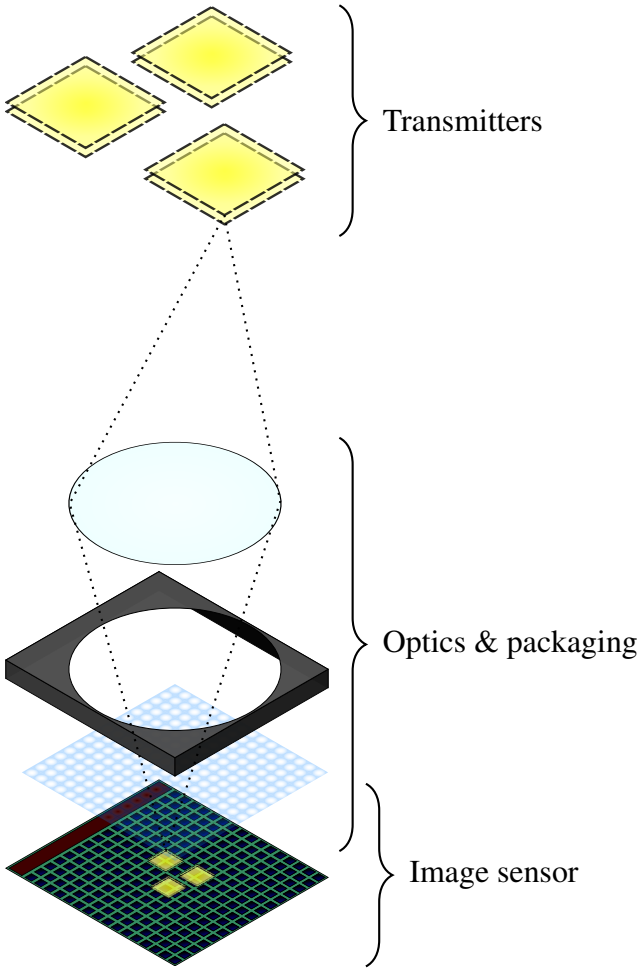


Figure 3: An imaging VLC receiver primarily consists of an image sensor (that has an array of photosensitive pixels) and optics. Here, an imaging VLC receiver is shown with three transmitters and their corresponding transmitter images projected onto the image sensor.

4 Token-Based Pixel Selection

We propose token-based pixel selection (TBPS) as a more flexible alternative to windowing. Like windowing, TBPS aims to skip pixels that do not receive VLC signals[1]. Unlike windowing though, TBPS allows each column to independently choose which pixels to read.

To configure which pixels are read and which are skipped, each pixel is programmed to be enabled (which are those that receive VLC signals) or disabled. At initialization, each column of pixels is given one token to arbitrate access to the shared output. As illustrated in Figure 5, this token is passed from pixel to pixel, skipping over any disabled pixel. When an enabled pixel receives the token, it holds the token for one read cycle and outputs its sample. At the end of the read cycle, the pixel passes the token to the next enabled pixel.

This mechanism allows TBPS image sensors to skip any pixels without VLC signals and to

[1]The VLC signal detection or tracking necessary to determine which pixel receives a VLC signal is outside the scope of this paper.

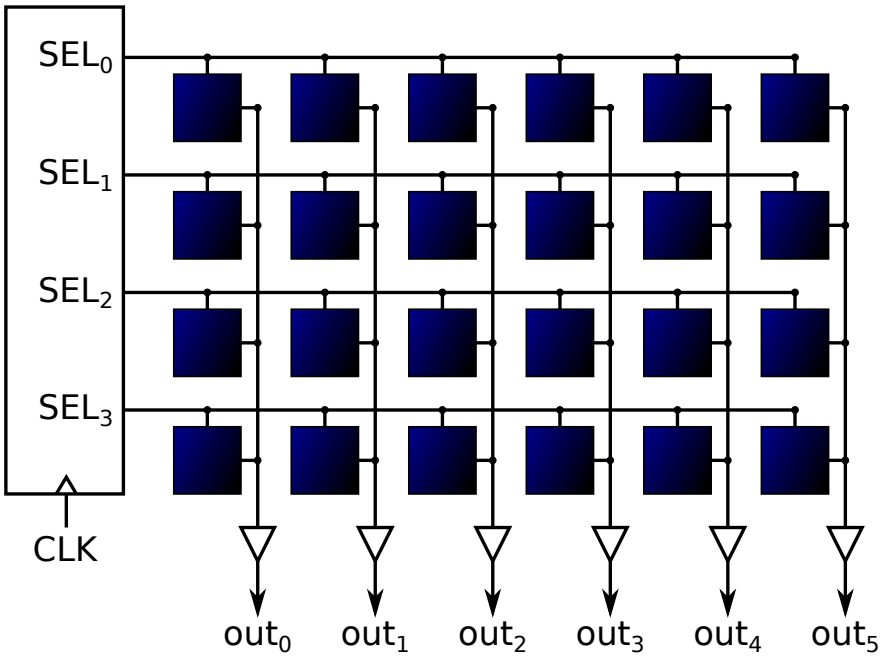


Figure 4: Row scanning image sensors only activate one SEL row-select line at a time to read out pixels.

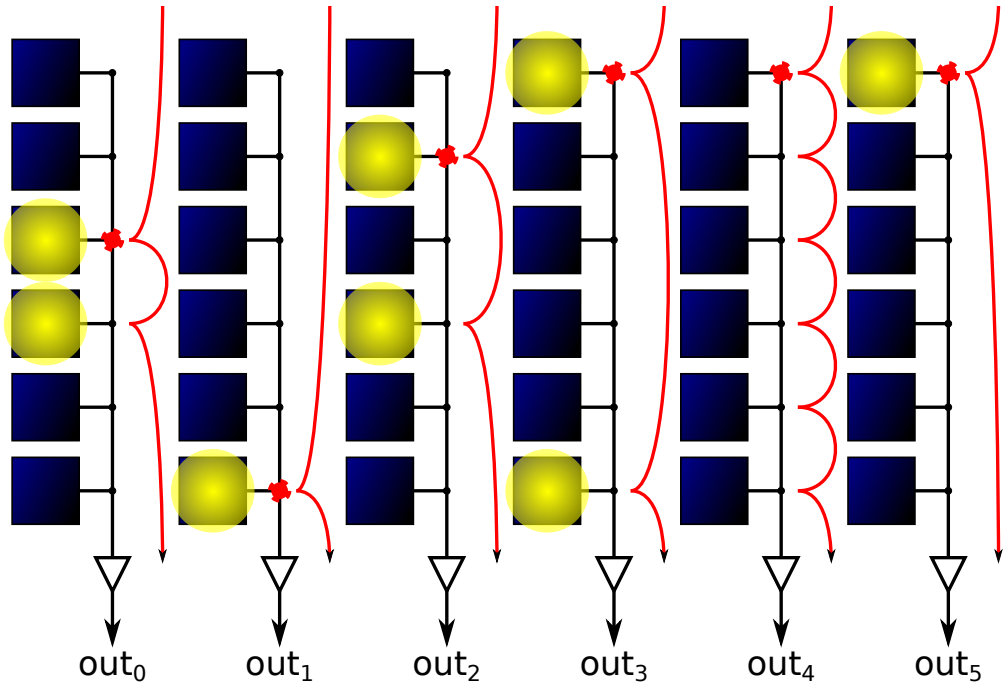


Figure 5: The path of tokens (red) in a 6x6 TBPS image sensor is illustrated with randomly distributed transmitter images (yellow).

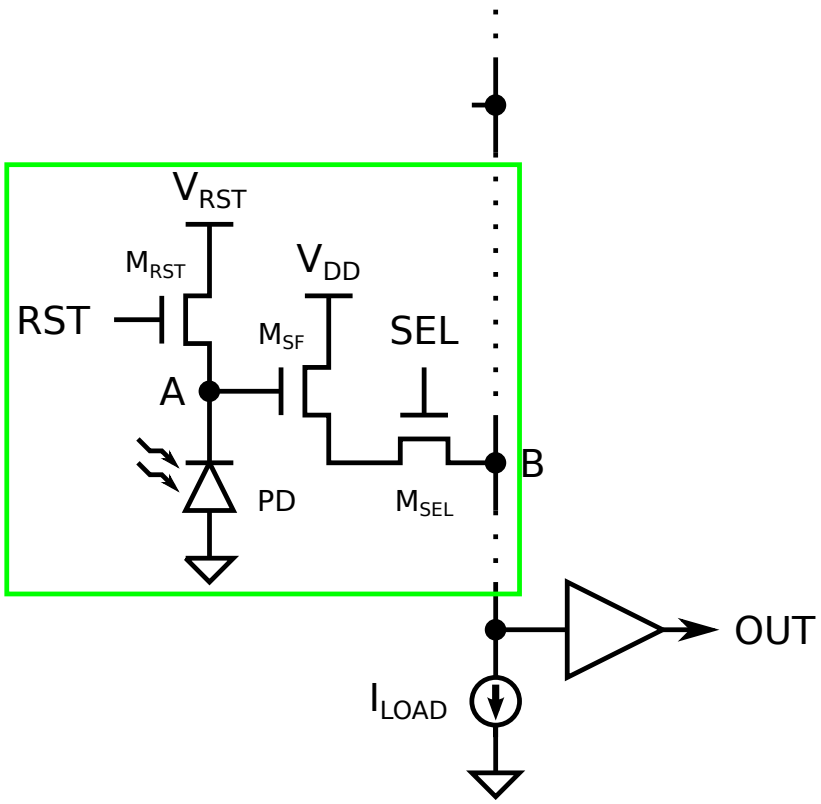


Figure 6: A basic CMOS active pixel connects to the shared output when its SEL input is raised high.

simultaneously read pixels in multiple rows.

4.1 Implementation Details

In TBPS, a pixel selector (PS) (shown in Figure 7) in each pixel replaces the row-select circuitry of the row-scanning image sensor to coordinate access to each column's shared output. The PS described in this paper is compatible with CMOS active pixels, such as the one illustrated in Figure 6.

4.1.1 Pixel Phases

CMOS active pixels acquire and outputs its samples in three phases. First, the pixel is reset to clear previously accumulated charges. Then, as photons hit the pixel, charges are accumulated through the pixel's photodiode; this accumulated charge represents the number of photons detected. Finally, after the exposure time passes, the pixel is connected to an output that reads the pixel to produce a discrete-time sample of the incoming optical signal.

Since the pixel's frequency response depends on its exposure time, the exposure time should be fixed to maintain the frequency response at a known value.

Certain other timing constraints must also be enforced for the image sensor to operate properly. Due to the shared outputs, no more than one pixel in each column may be in the read phase at any

given time. Furthermore, the presented pixel selector design can only reset at most one pixel at a time in each column. However, even in the same column, pixels can be concurrently exposed and different pixels can simultaneously be in different phases without conflicts.

4.1.2 Token-Passing Logic

To fix the exposure time and to pipeline the three phases, two tokens are used in each column: a reset (RST) token indicates which pixel is reset next and a select/read (SEL) token indicates which pixel is selected to output its sample.

In each column, the PSs pass a single SEL token using the logic shown in Figure 7b. Only the pixel with this SEL token may access the output. When a PS receives the SEL token (represented by a logic high at SIN, the SEL token input), it raises SEL on its CMOS active pixel to connect the pixel to the output. After the pixel is read, the PS passes the token onto the next pixel to be read.

A RST token is similarly passed from pixel to pixel in each column by the logic shown in Figure 7a. After initialization, the RST token moves in lockstep with the SEL token to ensure the proper exposure time for each sample.

Each PS is configured by an enable (EN) bit, which is high only if the connected pixel should be read. The EN bit can be implemented as addressable memory for easy configuration and may change at runtime to capture new VLC signals or to track moving transmitters. When EN is disabled for a pixel, the PS becomes transparent to tokens, allowing tokens to skip the disabled pixel.

The PSs in a column are connected together as shown in Figure 8.

4.1.3 Procedure for Operation

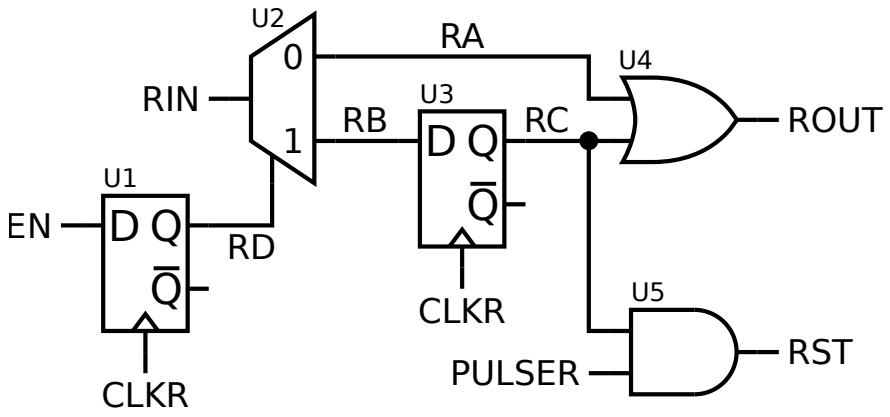
On initialization, with FLUSHS low, signal NEWS should be raised and CLKS should be cycled once for every pixel in the chain. This clears RS latch U6, illustrated in Figure 7b, for every pixel so that the RS latch can be used to detect whether the pixel has been reset. Once U6 has been cleared for each pixel, NEWS should be lowered.

The EN bit for each pixel should be initialized to a known state. Each column must always have at least one enabled pixel to avoid potentially losing the tokens.

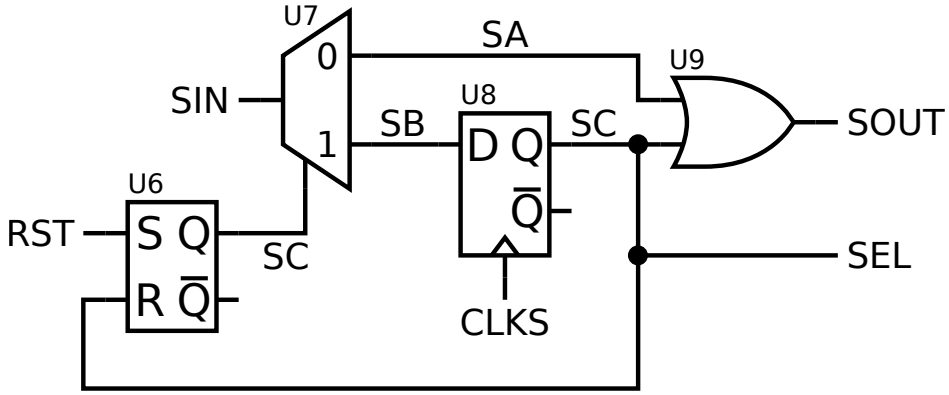
Next, both FLUSHR and FLUSHS should be raised. Cycle CLKR and CLKS once for each pixel in the chain to ensure that no tokens remain. When all tokens are flushed, return FLUSHR and FLUSHS to logic low. Finally, raise NEWR for one CLKR cycle and NEWS for one CLKS cycle to insert exactly one of each token into each chain. This initialization procedure ensures that each chain has only one RST token and one SEL token, and that the imaging receiver always knows where each token is (and thus, which pixel is being read).

Once the TBPS image sensor is initialized, each rising edge of clocks CLKR and CLKS advances the tokens to the next enabled PS. After a token passes the last pixel in the chain, it is returned to the first pixel so that the column's pixels are scanned in a round-robin fashion.

This TBPS architecture allows the image sensor's limited number of readout circuits to read the desired pixel while ignoring all other pixels. Furthermore, the PS logic has been designed so that after initialization, a pixel can be read on every clock cycle, which helps ensure that the PS does not slow down the image sensor.



(a) RST PS logic



(b) SEL PS logic

Figure 7: The logic for the pixel selector (PS) is shown. U1, U3, and U8 are D flip-flops; U2 and U7 are demultiplexers; and U6 is a SR latch. EN, a bit of addressable memory, is clocked through U1 to avoid glitches. U2 and U7 route incoming tokens, to either bypass this pixel or hold the token for one clock cycle. Tokens enter through RIN and SIN; they leave through ROUT and SOUT. The PS produces RST and SEL to control the connected pixel.

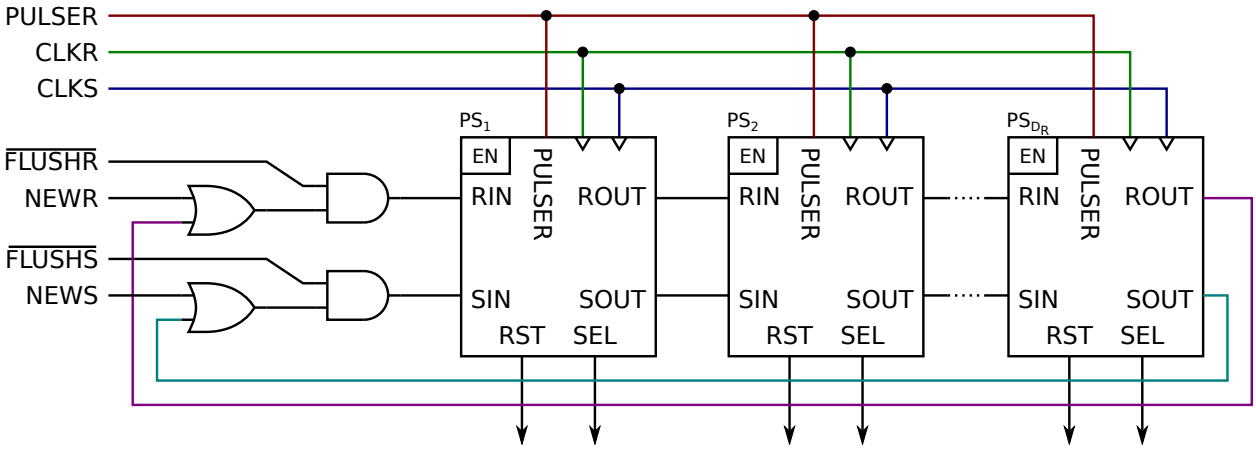


Figure 8: The connections for a column of pixels that share the same output is shown. FLUSH and NEW are used in initialization to flush tokens and to insert a token respectively. Each PS's SEL output is connected to its corresponding pixel.

5 Performance

To evaluate the performance of the token-based pixel selection (TBPS) architecture, we model the receiver's image sensor as a rectangular array of pixels with D_R rows and D_C columns, where each column has a separate output. Onto this array, we uniformly distribute M'_T transmitter images and calculate R_{PO}: the aggregate rate at which pixels with transmitter images are sampled and read out. Along with the signal-to-noise ratios, R_{PO} determines the maximum capacity of the imaging VLC receiver. We compare R_{PO} for both TBPS and windowing VLC receivers.

Although each transmitter image can land on multiple adjacent pixels, we assume for simplicity that each of the M'_T transmitter images lands on only one receiver pixel. This assumption is reasonable if transmitters are small or sufficiently far away, forming transmitter images that are small with respect to the receiver pixel size. This assumption may also be practical if the image sensor performs binning (as diversity combining) for transmitter images that land on multiple pixels before the pixels are read out.

5.1 Assumed Readout Timing

Modeling the reset mechanism as a switch to discharge the photodiode's capacitance (estimated to be $7\ \mathrm{pF}$) and assuming a switch resistance of less than $100\ \Omega$, we determine that

$$T_{\text{RST}} = 4\ \mathrm{ns}$$

as more than five time constants, is amply sufficient to completely reset the pixel.

The appropriate exposure time for the imaging VLC receiver is heavily scenario-dependent. For this analysis, we assume an exposure time of

$$T_{\text{int}} = 100\ \mathrm{ns}$$

Assuming $1\ \mu\mathrm{W}$ of incident light per pixel, $400\ \mathrm{mA/W}$ responsivity, and $7\ \mathrm{pF}$ capacitance, $100\ \mathrm{ns}$ is sufficient to generate a $5.7\ \mathrm{mV}$ voltage swing.

Finally, we use

$$T_{\text{read}} = 16ns$$

for the duration of the read phase. This duration is estimated by surveying commercial CMOS image sensors, which can read pixels at $62\ \mathrm{MHz}$ [6].

5.2 Analysis

Of the M'_T transmitter images incident on the image sensor, M'_{Tj} of those transmitter images land on column j. In column j, these M'_{Tj} transmitter images land on M'_{Rj} pixels that need to be read. Since multiple transmitter images can overlap, $M'_{Rj} \leq M'_{Tj}$.

Define D_{RW} to be the total number of rows needed to cover all M'_{Rj} pixels in all columns; this is the minimum number of rows that needs to be scanned in windowing image sensors to read every pixel with transmitter images.

$$M'_{Rj} \leq D_{RW} \forall j \in \{1, 2, ..., D_C\} \tag{4}$$

Given this setup, define $U_j(n)$ to be the total time needed to read each of the M'_{Rj} pixels (in column j) n times, including any initial setup overhead time. Let

$$T_{\text{cycle}j} \triangleq \lim_{n\to\infty} \frac{U_j(n)}{n} \tag{5}$$

Essentially, $T_{\text{cycle}j}$ is the additional time needed to scan every pixel to be read in column j one more time.

Using

$$T_{\text{inc}} = \max(T_{\text{RST}}, T_{\text{read}}) \tag{6}$$

for TBPS,

$$T_{\text{cycle}j} = \max(T_{\text{RST}} + T_{\text{int}} + T_{\text{read}}, M'_{Rj}T_{\text{inc}}) \tag{7}$$

and for regular windowing,

$$T_{\text{cycle}j} = \max(T_{\text{RST}} + T_{\text{int}} + T_{\text{read}}, D_{RW}T_{\text{inc}}) \tag{8}$$

Note that due to pipelining (e.g., other pixels in a column can be read while the first pixel is reset and exposed again), an output can take as long to scan one pixel as it does to scan multiple pixels.

Our desired metric, R_{PO}, is the aggregate rate at which pixels with transmitter images are read out:

$$R_{PO} = \sum_{j=1}^{D_C} \frac{M'_{Rj}}{T_{\text{cycle}j}} \tag{9}$$

where $M'_{Rj}/T_{\text{cycle}j}$ is the rate at which pixels with transmitter images are read out in column j.

5.3 Simulation

With this setup, we performed Monte Carlo simulations of TBPS and windowing image sensors at different image sensor resolutions over a range of values for M'_T. As shown in Figures 9 and 10, the simulations confirmed that (given the same values for T_{RST}, T_{int}, and T_{read}) TBPS always performs as well as or better than regular windowing.

Figure 9 shows that the relative advantage of TBPS increases with image sensor resolution: a trend that continues beyond the simulated 64x64 resolution. This advantage can be attributed to the increased likelihood for higher-resolution image sensors that transmitter images are scattered over more rows of the image sensor, reducing the number of desired pixels (those with transmitter images) on each row. Higher resolution image sensors also have more columns, further reducing the concentration of desired pixels per row.

Since windowing image sensors only read one row at a time, reducing the concentration of desired pixels per row increases the average fraction of time that each output reads "blank" pixels without VLC signals and reduces the average rate at which each windowing image sensor output read desired pixels. In contrast, TBPS image sensors can better utilize each available output because they can read from multiple rows at once, skipping over blank pixels to only read desired pixels.

The plot also shows that when the number of transmitter images incident on the image sensor (M'_T) is low, both TBPS and windowing have the same performance. In this regime, R_{PO} is

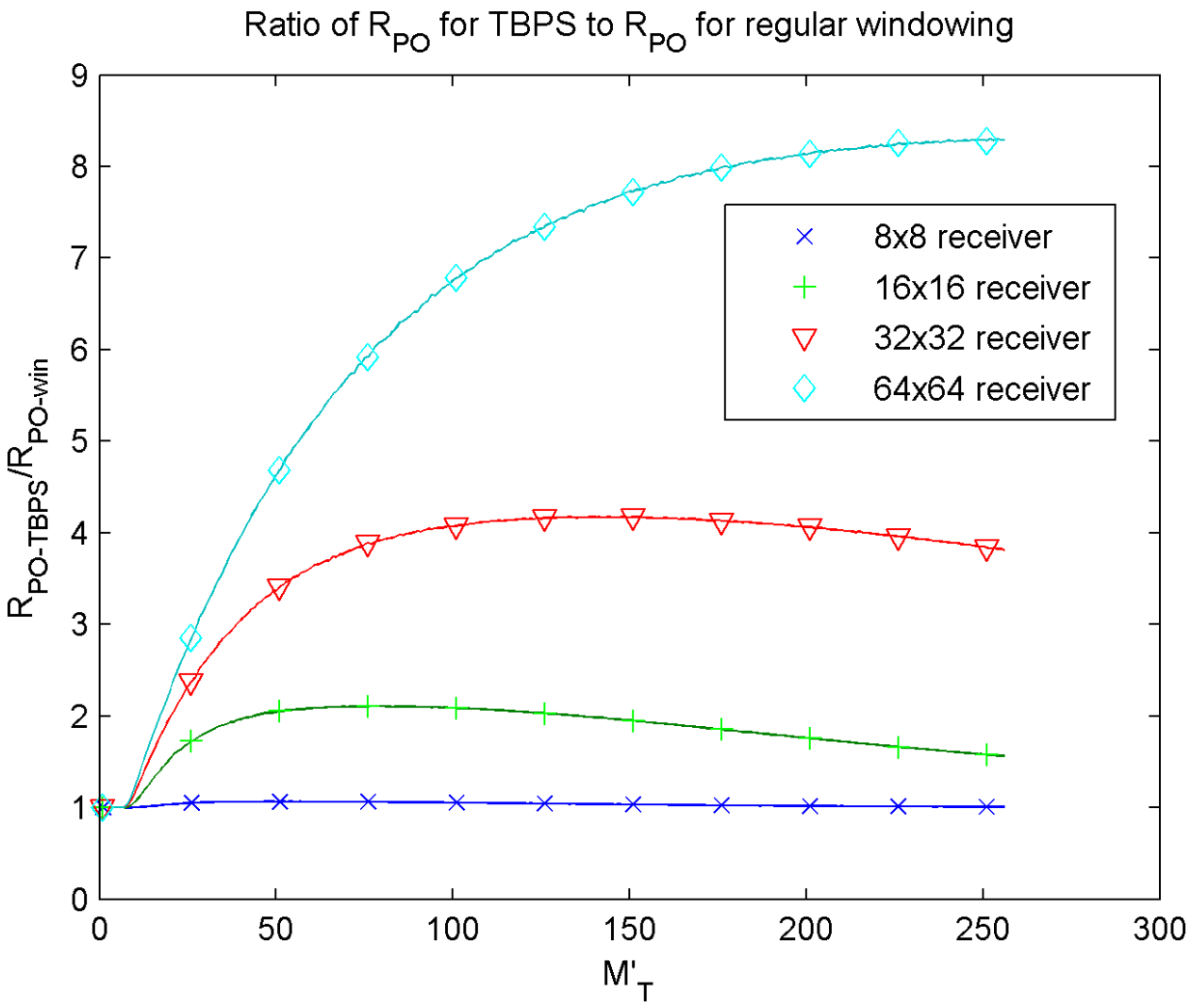


Figure 9: The relative performance gain of TBPS is shown.

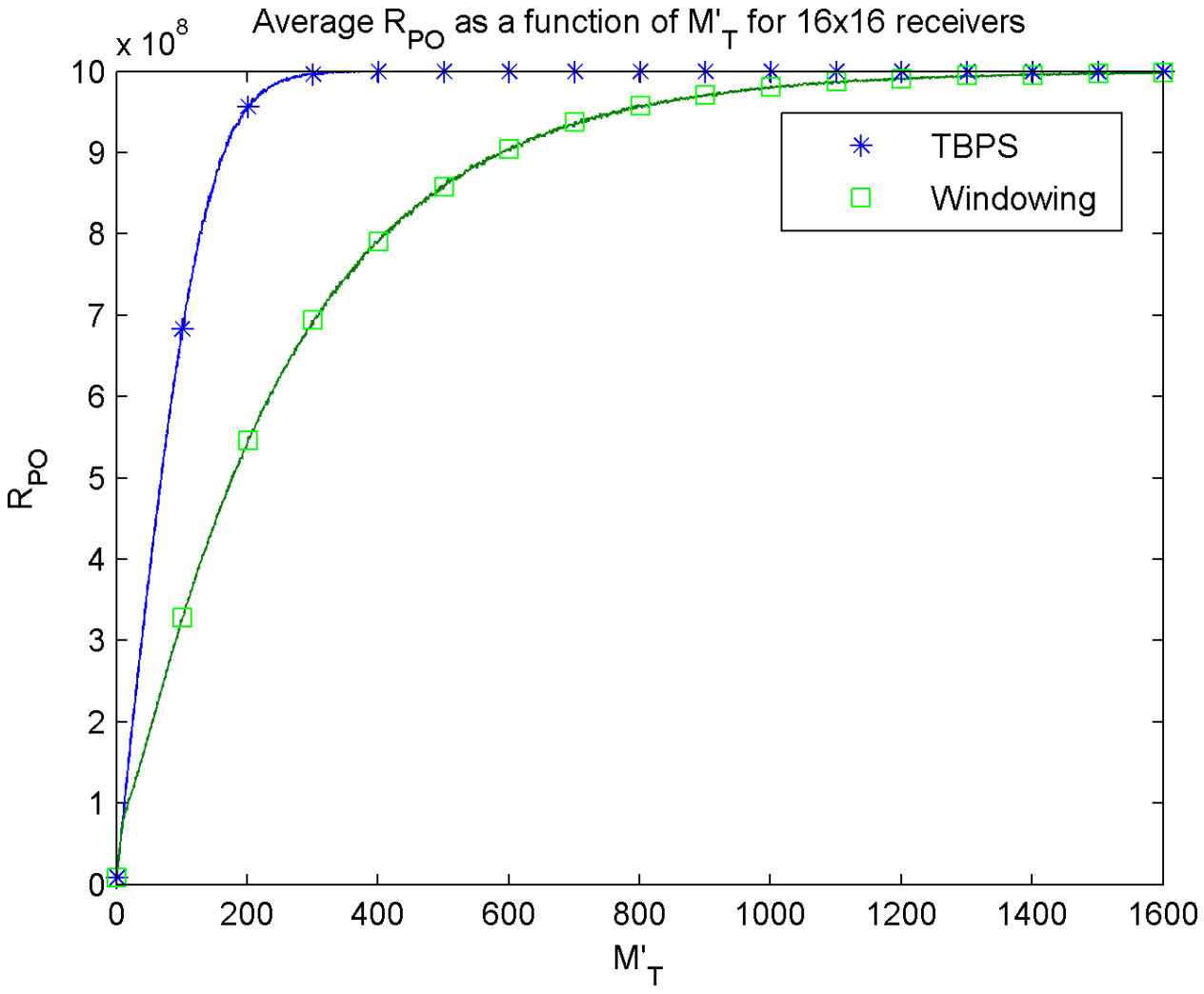


Figure 10: For any M'_T, since $M'_{Rj} \leq D_{RW}$, the TBPS performs better (or as well as) regular windowing. Extremely high values of M'_T are plotted to show that both TBPS and windowing approach the same maximum R_{PO}.

limited by the relatively long exposure time T_{int}; pipelining allows all of the reads to fit into the time needed to reset, expose, and read just one pixel. In these cases, $T_{\text{cycle}j} = T_{\text{RST}} + T_{\text{int}} + T_{\text{read}}$ for both windowing and TBPS, yielding the same R_{PO}.

As M'_T grows, the ratio of $R_{PO\text{-TBPS}}$ to $R_{PO\text{-win}}$ reaches a maximum and then decreases back towards unity. Although this maximum may appear to suggest that that an optimal M'_T exists for each resolution, R_{PO} continues to improve for both architectures as M'_T increases.

This trend is better explained with Figure 10, which shows that as M'_T grows, the performance of both the TBPS and windowing image sensors asymptotically approach the same maximum $R_{PO} = 10^9$ pixels/s, which occurs when every column reads desired pixels at the maximum rate. Since the TBPS image sensor approaches this maximum R_{PO} faster as M'_T grows, its R_{PO} plateaus first, allowing the R_{PO} of the windowing image sensor to catch up. As the R_{PO} of the windowing image sensor catches up, the ratio of $R_{PO\text{-TBPS}}$ to $R_{PO\text{-win}}$ decreases back towards one.

Although windowing performs as well as TBPS at extreme values of M'_T (with either a huge quantity of VLC transmitters or very few), TBPS yields much better performance than windowing in typical cases.

6 Conclusion

Dual-use lighting can provide ubiquitous wireless network connectivity by embedding VLC in the the lighting infrastructure. But the need for diffuse high-quality light can limit the effective data rates delivered by each lighting unit. To compensate for this limitation, we investigate spatial multiplexing as a technique to increase received data rates by aggregating multiple lighting units as independent transmitters. Specifically, we investigate the challenge of receiving these multiple data streams through the use of imaging (camera) receiver architectures.

This approach requires high-resolution image sensors that can quickly sample signals from many VLC transmitters simultaneously. Unfortunately, existing image sensor architectures do not scale well to meet this requirement. Instead, we propose a new imager architecture that is designed specifically for receiving spatially-multiplexed VLC transmissions and is compatible with well-established and low-cost CMOS active pixel sensor designs.

The proposed token-based pixel selection (TBPS) architecture allows the image sensor to selectively scan pixels of interest while ignoring all other pixels. Although TBPS image sensors require more control logic per pixel than simpler windowing image sensors, the additional design complexity is justified by the increase in sampling speed for each VLC stream and thus the overall reception capacity enabled by the device.

In addition to spatial multiplexing, imaging VLC receivers can also guarantee interference-free communications in multi-user scenarios, without requiring coordination between transmitters, through *space-division multiple access*, in which the spatial separation of transmitter images prevents interference between transmissions. Such guarantees may be useful in time-sensitive and safety-critical applications, such as in vehicular networks [2].

In the future, we plan to refine our simulation model for the placement of transmitter images, relax the simplifying assumption that each transmitter image only lands on one pixel, investigate VLC signal detection and tracking, explore diversity combining and MIMO decoding for CMOS image sensors, investigate the characteristics of CMOS active pixels for communications, and explore other applications of TBPS image sensors.

References

[1] A. Azhar, T.-A. Tran, and D. O'Brien. Demonstration of high-speed data transmission using MIMO-OFDM visible light communications. In *GLOBECOM Workshops (GC Wkshps), 2010 IEEE*, pages 1052–1056, 2010.

[2] J. Chau. Secure cooperative accident avoidance for vehicles. Master's project, Boston University, May 2011.

[3] J. Gancarz, H. Elgala, and T. Little. Impact of lighting requirements on VLC systems. *Communications Magazine, IEEE*, 51(12):34–41, December 2013.

[4] J. Grubor, O. Jamett, J. Walewski, S. Randel, and K.-D. Langer. High-speed wireless indoor communication via visible light. *ITG-Fachbericht-Breitbandversorgung in Deutschland-Vielfalt für alle?*, 2007.

[5] B. Nakhkoob, S. Ray, and M. Hella. CMOS integrated high speed light sensors for optical wireless communication applications. In *Sensors, 2012 IEEE*, pages 1–4, 2012.

[6] ON Semiconductor. VITA 1300 1.3 megapixel 150 FPS global shutter CMOS image sensor, June 2013.