An overview of optical communication for deep space: issues, challenges, solutions

by

Baris I. Erkmen, Bruce E. Moision, and Samuel J. Dolinar

Jet Propulsion Laboratory, California Institute of Technology

@ 2012 California Institute of Technology. Government sponsorship acknowledged.
Deep-space optical communication

• Why optical communications to, from, in space?

• Optical communication systems
  – system diagram, link equations, fundamental capacity bounds

• Impairments to the optical channel
  – atmosphere-induced losses
  – pointing-induced losses
  – receiver-induced losses

• Conclusions
Why optical communication?

- Ever-growing demand for data rate and data volume
  - increase in science return from interplanetary missions
  - increasing desire for connectivity via high-bandwidth links

- Optical frequencies much higher in the EM spectrum than radio frequencies
  - higher carrier frequency: smaller diffraction of transmitted beams
  - unallocated spectrum: significantly higher modulation bandwidth
Radio frequency (RF) versus optical links

- **I.** diffraction loss is significantly smaller than RF
- **II.** nominal aperture sizes are smaller
- **III.** photon energy is significantly higher
- **IV.** nominal transmitter and receiver efficiencies are lower

Representative RF and optical links:

<table>
<thead>
<tr>
<th>(r)</th>
<th>Ka-band Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f$</td>
<td>carrier frequency</td>
</tr>
<tr>
<td>$D_t$</td>
<td>transmit diameter</td>
</tr>
<tr>
<td>$D_r$</td>
<td>receiver diameter</td>
</tr>
<tr>
<td>$\eta$</td>
<td>system efficiency</td>
</tr>
<tr>
<td>$N_0$</td>
<td>noise spectral density</td>
</tr>
<tr>
<td>$W$</td>
<td>bandwidth</td>
</tr>
<tr>
<td>$P_t$</td>
<td>transmit power</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(o)</th>
<th>Near-Infrared Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
<td>wavelength</td>
</tr>
<tr>
<td>$D_t$</td>
<td>transmit diameter</td>
</tr>
<tr>
<td>$D_r$</td>
<td>receiver diameter</td>
</tr>
<tr>
<td>$\eta$</td>
<td>system efficiency</td>
</tr>
<tr>
<td>$\alpha_b$</td>
<td>noise spatial density</td>
</tr>
<tr>
<td>$T_s$</td>
<td>slot width</td>
</tr>
<tr>
<td>$P_t$</td>
<td>transmit power</td>
</tr>
</tbody>
</table>

\[ C_{\text{OPT}} = C_{\text{RF}} \times \left( \frac{\lambda^{(r)}}{\lambda^{(o)}} \right)^2 \left( \frac{D_{t}^{(o)} D_{r}^{(o)}}{D_{t}^{(r)} D_{r}^{(r)}} \right)^2 \left( \frac{N_0^{(r)}}{(hc/\lambda^{(o)})/\log_2 M} \right) \left( \frac{\eta^{(o)}}{\eta^{(r)}} \right) \]

- **I.** +76 dB
- **II.** -33 dB
- **III.** -12 dB
- **IV.** -16 dB

**net = +15 dB**
The link design must overcome channel impairments

Atmosphere-induced impairments:
- absorption: atmospheric extinction
- turbulence: scintillation, beam spread and angle-of-arrival spread

Pointing-induced impairments:
- time-dependent fluctuations in power delivered to receiver

Receiver-induced impairments:
- shot noise
- background radiation noise
- photodetector impairments: blocking, timing jitter, nonlinear responsivity
- post-detection electronic noise
The overall system performance depends on
- modulation method and detection method
- pointing, acquisition and tracking performance (both flight and ground)
- transmitter resources (e.g., photon flux)
- channel code performance
Link performance evaluation

\[
P_{rx} = P_t G_t G_r L_s L_a \eta_{pt} \eta_t \eta_r
\]

- Power incident on the detection subsystem
  - \( P_t \) : transmitted power
  - \( G_t, G_r \) : transmit & receive aperture gains
  - \( L_s \) : space loss
  - \( L_a \) : atmospheric loss
  - \( \eta_{pt} \) : pointing loss
  - \( \eta_t, \eta_r \) : transmit & receive efficiencies

\[
P_{rqd} = P_i / (L_b L_j L_f L_t \eta_{det} \eta_{imp} \eta_{code} \eta_{int})
\]

- Required signal power to support target data rate
  - \( P_i \) : minimum (ideal receiver) required power
  - \( L_b, L_j, \eta_{det} \) : detector blocking, jitter & efficiency losses
  - \( L_f \) : fading loss
  - \( L_t \) : truncation loss (angle-of-arrival spread)
  - \( \eta_{imp} \) : implementation efficiency
  - \( \eta_{code}, \eta_{int} \) : code & interleaver efficiencies

Link margin [dB] = \( 10 \log_{10}(P_{rx}/P_{rqd}) \)
Conventional detection methods

- **Direct detection:**
  - measures photon-flux of incident field
  - ideal limit yields Poisson statistics with rate equal to incident photon flux

- **Heterodyne detection:**
  - measures real and imaginary quadratures of field
  - ideal limit yields Gaussian statistics with mean equal to incident field amplitude and variance $\frac{1}{2}$ per dimension.

- **Homodyne detection:**
  - measures Real quadrature of field
  - ideal limit yields Gaussian statistics with mean equal to quadrature of incident field and variance $\frac{1}{4}$. 
Coherent versus incoherent detection

- Resource efficiency curves obtained via information theory
  - Photon efficiency \( (c_p) \): capacity [bits/s] / average photon flux [photons/s]
  - Bandwidth efficiency \( (c_d) \): capacity [bits/s] / modulation bandwidth [1/s]
- Direct detection receivers + binary intensity modulations asymptotically optimal in photon efficiency
- Heterodyne and homodyne receivers + coherent-state modulations encounter brick-wall asymptotes in photon efficiency
To achieve high photon efficiency, average photon number per channel use must be low
  – binary modulation alphabet is near-optimal: often transmit nothing, send a pulse sparsely
  – results in low duty cycle $\Rightarrow$ high peak-to-average photon flux

**Binary intensity modulation and photon-counting**

**On-Off-Keying (OOK):** 1 bit is represented by a slot, which may either be occupied by a pulse or not

**Pulse-Position-Modulation (PPM):** $\log_2 M$ bits are represented by a single pulse out of $M$ slots (here $M=16$).
Photon-counting statistics

- Ideal photon-counting yields photon arrivals as a Poisson point process with rate function proportional to incident photon flux.

- After synchronization an equivalent discrete channel can be defined.

\[ p(1) = \frac{1}{M} \]

\[ p(y|a = 0) = \frac{e^{-n_b}(n_b)^y}{y!} \]

\[ p(y|a = 1) = \frac{e^{-(n_s+n_b)}(n_s+n_b)^y}{y!} \]

\[ n_s = \text{mean signal photons per pulsed slot} \]

\[ n_b = \text{mean background photons per slot} \]
A simple example for reliable information transfer

Fix data
Add parity
Transmit symbols
Detect signal
Recover clock
Form statistics
Estimate data

1 0 1 1 0 1 0
1 0 1 1 0 1 0 1 1 1 1

0: 1:

Channel coding
Modulation
Slotwidth $T_s$
Detection
Synchronization
Receiving/Decoding

Fix data 1 0 1 1 0 1 0
0: 1:

Form statistics
1 0 1 1 0 1 0

Receiving/Decoding
Optimal processing of photon counts

- Deciding on each bit independently (based on photon counts in the slot) is suboptimal

\[
\begin{align*}
M=8 \\
\text{slot} & \quad 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
\text{count} & \quad 0 & 1 & 0 & 3 & 0 & 1 & 1 & 0 \\
\end{align*}
\]

- Photon-number gap is several dB between hard and soft decisions

\[
\hat{x} = 011 \\
p(x=000|y) = 0.05 \\
p(x=001|y) = 0.2 \\
p(x=010|y) = 0.05 \\
p(x=011|y) = 0.3 \\
\ldots
\]

- Typical region of operation is between 0.01 and 1.0 photons/slot

Average power required to achieve C = 1/8 bits/slot for M=16 slots
Channel impairments:
I. Atmospheric effects
Optical communication through the atmosphere

- Atmospheric layer mostly concentrated in 0-20km above ground, although it extends up to 100km
- Atmospheric effects on performance:
  - bad weather (e.g., snow, fog, rain), and particulates cause absorption and scattering
  - in clear weather *turbulence* causes fading, beam and angular spread

- Impact in Earth-space links are *asymmetric*
Absorption and scattering from aerosols (dust, etc.) and molecules (water vapor, etc.) attenuate the signal.

In bad weather (rain, snow, fog), attenuation can be severe, causing dropouts.

Even in clear-sky conditions must budget for attenuation.
  - Drives selection of bands with good clear-sky transmissivity.

Typical attenuation for Earth-space link in near-infrared at zenith 0.1 -- 0.3 dB.

FIGURE 8.12  Atmospheric transmittance in an Earth-to-space path at zenith. A rural aerosol composition with a surface visual range of 23 km is considered. The data refers to the case of an observer located at two elevations: sea level (lower transmittance) and 2 km above sea level.

Piazolla, Near-Earth Laser Communications 2009
Clear-air turbulence effects

- Random spatio-temporal mixing of air with different temperatures causes refractive-index variations
  - scintillation (constructive/destructive interference)
  - angle-of-arrival variations
  - beam spreading
  - beam wander

Space-to-Earth:
- Angle-of-arrival spread (spatial distortion)
- Scintillation (fading)

Earth-to-Space:
- Beam spread (attenuation)
- Beam wander and scintillation (fading)

atmosphere is mostly concentrated in 0-20 km
Impact on downlink

- Extended Huygens-Fresnel principle models paraxial quasimonochromatic propagation through turbulence

$$E_f(\rho) = \int d\rho' E_r(\rho') e^{i\phi(\rho')} h_f(\rho, \rho')$$

- Field incident on the photodetector plane has speckle
  - short-exposure brightest spot will wander on the photodetector surface
  - long-exposure average will be broader than diffraction limit

- To accommodate angle-of-arrival fluctuations, field-of-view of detector larger than vacuum \(\Rightarrow\) increase in background
• Extended Huygens-Fresnel principle yields

\[ E_r(\rho) = \int d\rho' E_t(\rho') e^{i\phi(\rho')} h_{fs}(\rho - \rho'; L) \]

vacuum-propagation Green's function

\[ \langle e^{\chi(\rho_1) - i\phi(\rho_1)} e^{\chi(\rho_2) + i\phi(\rho_2)} \rangle = e^{-D(\rho_2 - \rho_1)/2} \]

\[ D(\rho) = |\rho|^{5/3} / r_0^{5/3} \]

transmitter-plane-coherence length

• Taylor-series expanding phase term shows various effects
  – \( \phi_2(\rho) \): short-term beam spread (relative to vacuum propagation)
  – \( k_0 \): beam wander (randomly-varying tilt at transmitter plane)
  – \( \chi(\rho) \): scintillation

\[ \phi(\rho) = \phi_0 + \rho \cdot k_0 + \phi_2(\rho) \]

Time-dependent power fluctuations

- Turbulence causes the incident photon flux to fluctuate

- Fluctuations in weak turbulence are often well-modeled as log-normally distributed

\[ P_r(t) = P_0 V(t) \]
\[ \langle V(t) \rangle = 1 \]

Marginal distribution:
\[ f_V(v) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \frac{1}{v} \exp \left( -\frac{(\log v + \sigma_i^2/2)^2}{2\sigma_i^2} \right) \]

Biswas & Wright, Measured fluctuations over a 45-km mountain-top to mountain-top link, 2002
Block fading model for photon-flux fluctuations

- \( P_r(t) \) exhibits a coherence time on the order of \( \sim \) msec

- We define coherence time as inverse of 90\% bandwidth

\[
W(\xi) \equiv \left\{ B : \frac{\int_{-B}^{B} dfS_x(f)}{\int_{-\infty}^{\infty} dfS_x(f)} = \xi \right\}
\]

\[
T_{coh} \equiv \frac{1}{W(0.9)}
\]

- We simplify fading to a two-parameter model \( \{\sigma_l^2, T_{coh}\} \)

Model fades as drawn independently from a log-normal distribution every \( T_{coh} \) seconds, and constant over those intervals.
Mitigating outages with interleaving

• Codewords are significantly shorter than the rate of photon-flux fluctuations
  – with no mitigation there is always a finite probability of a fade deep enough to corrupt a codeword \( \Rightarrow \) unconditional reliability not possible

• To assure reliability, each codeword should see a ‘well-mixed’ channel
  – interleaving achieves this goal, as \( N \to \infty \) rate of reliable transmission converges to ergodic capacity
    \[
    C_E(\sigma_I^2, T_{coh}) \equiv \langle C(P_0, V) \rangle_V
    \]

Codeword duration = 0.06 msec

\[ t \text{ (msec)} \]

\[ t \text{ (msec)} \]
Capacity loss due to fading

- We have $C_E(P_0, \sigma^2_I, T_{coh}) \leq C_E(P_0, 0, \infty)$
  - fading dynamics cause *unrecoverable* loss at equal mean photon flux

- Finite-depth interleaving ($N < \infty$) results in a nonzero outage probability and an increase in required power

$$\eta_{int} \approx 16 \sqrt{\frac{\sigma^2_I}{N_f}} \text{ dB}$$

$$L_f \approx 2.5\sigma^2_I \text{ dB}$$
Channel impairments
II. Pointing effects
Transmitter-induced pointing errors

- Transmitter often performs active tracking on beacon (or received communication beam in bidirectional links) to minimize pointing errors.

- Assuming a circularly-symmetric Gaussian beam, and a Gaussian random process for $\theta(t)$

  The marginal probability distribution (averaged over a uniform $\phi$):

  $$p(\theta) = \frac{\theta}{\sigma^2} e^{-\left(\theta^2_{\text{m}} + \theta^2\right) / (2\sigma^2)} I_0\left(\frac{\theta_{\text{m}} \theta}{\sigma^2}\right)$$

- Impact on link is similar to that of turbulence-induced fades:

  $$P_r(t) = P_0 B_G(\theta(t))$$

  where $B_G(\theta(t))$ is the on-axis irradiance and $P_0$ is the normalized beam profile.
Receiver-induced pointing errors

- The telescope can be thought of as a *big* aperture and an effective lens focusing the incoming light on a photodetector.
- If telescope is off-pointed by $\theta$ relative to optical axis, the beam will focus off-center.

\[
\text{translation of spot is approximately } \frac{f \theta}{f}.
\]

- The long-exposure image integrates over many jitter-induced translations, resulting in broadened spot.
  - The average pulse approaches Gaussian as exposure time increases and has *approximate width* $f \sqrt{\theta_{\text{rms}}^2 + \left(\frac{\alpha \lambda}{D}\right)^2}$ (\(\alpha\) is some constant).

- Impact is similar to that of angle-of-arrival spread.
Channel impairments

III. Receiver effects
• The receiver collects background light along with signal
  – scattered sunlight (daytime)
  – light from point sources in field-of-view (both daytime and nighttime)

• Background light is treated as uniform incoherent illumination on the photodetector

\[ P_n = B_\lambda \times A_R \times \Omega \times \Delta \lambda \times \eta_R \]

Sky Radiance
Aperture area
Field-of-View
Bandwidth
Efficiency

Piazzolla, Near-Earth Laser Communications 2009
Efficiency trade-off with finite background

- **$K$** background noise modes (*iid*, complex-Gaussian amplitude distribution), $N$ counts/mode

$$p_1(k; K) = \frac{N^k}{(1+N)^{k+K}} L_k^{(K-1)} \left(\frac{-n_s}{N(1+N)}\right) e^{-n_s/(1+N)}$$

$$p_1(n; K) \xrightarrow[K \to \infty]{} \frac{(n_b+n_s)^n e^{-(n_b+n_s)}}{n!}$$

- Thermal noise bounds photon efficiency

$$c_p^{\text{Hol}}(n_b) \leq \log_2(1 + 1/n_b) \quad (n_b \equiv KN)$$

- Poisson approximation gives unbounded photon efficiency

  - Poisson approximation to multimode thermal noise must become inaccurate at large $c_p$ for any number of noise modes
**Dark noise** is spurious photoelectrons that are generated even with no incident light

- dark current is a Poisson-distributed *signal-independent background noise* $n_b$.

$$n_d = \ell_d A T_s$$  
**dark e/slot**

Surface dark rate  
Slot width  
Active area

Noise levels with $n_b > 10^{-5}$ incur large losses beyond 10 bits/photon.

<table>
<thead>
<tr>
<th>Device</th>
<th>$\ell_d$ (e/s/mm²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si GM-APD</td>
<td>$10^6$</td>
</tr>
<tr>
<td>InGaAsP GM-APD</td>
<td>$10^8$</td>
</tr>
<tr>
<td>NbN SNSPD</td>
<td>$10^2$</td>
</tr>
</tbody>
</table>
Efficiency trade-off with finite dark noise

- The photon-efficiency of OOK + photon-counting is effectively bounded because spectral efficiency drops off doubly-exponentially: \( c_d < \beta c_p 2^{-\beta c_p} \) \[ \beta = \max(1, en_b 2^{c_p}) \]
  - this approximate bound crosses the noiseless OOK and Holevo curves at
    \[ c_p \approx \log_2 \left( \frac{1}{en_b} \right) \]
  - The actual \( c_d \) breaks away sharply from the noiseless OOK curve when \( Mn_b \approx 1/e^4 = 0.018 \) noise counts/PPM symbol
• Non-ideal lasers ‘leak’ some power in the off state:
  – leaked power is proportional to power in the on state; the proportionality constant is the extinction ratio $\alpha$.

$P_0 = \frac{P_1}{\alpha} = $ power transmitted in the off state

$P_1 = $ power transmitted in the on state

Effective signal photons

‘Signal’ photons appearing as noise

• generates a Poisson-distributed background noise proportional to the signal $n_b = \frac{n_s}{\alpha}$

• With finite extinction ratio $\alpha$, the photon efficiency of OOK + photon counting is strictly bounded:
  $$c_p \leq \log_2(\alpha) - 1/\ln(2) \text{ (bits/effective signal photon)}$$
Each curve in these plots is the capacity efficiency tradeoff for a given PPM order $M$, and is generated by varying the average number of signal photons.
Timing jitter at the photodetector

- Jitter is the *random delay* from the time a photon is incident on a photo-detector to the time a photo-electron is detected.

- Jitter losses are a function of the *normalized jitter standard deviation*: \( \frac{\sigma_j}{T_s} \)

- Losses grow rapidly when \( \frac{\sigma_j}{T_s} > 0.1 \)

### Device Specifications

<table>
<thead>
<tr>
<th>Device</th>
<th>( \sigma / 1\text{ns} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>InGaAs(P) PMT</td>
<td>0.9</td>
</tr>
<tr>
<td>InGaAs(P) GM-APD</td>
<td>0.3</td>
</tr>
<tr>
<td>Si GM-APD</td>
<td>0.24</td>
</tr>
<tr>
<td>NbN SNSPD</td>
<td>0.03</td>
</tr>
</tbody>
</table>

### Graphical Representation

- **Incident signal intensity**
- **Equivalent “spread” signal intensity**
- **Knee at \( \sigma/T_s = 0.1 \)**
Photodetector blocking

- Photon-counting detectors are rendered inoperative (blocked) for some time $\tau$ (dead time) after each detection
  - 10-50 ns for SiGM-APD
  - 1-10 $\mu$s for InGaAs GM-APD
  - 3-20 ns, for NbN SNSPD

- Blocking can be modeled as a Markov chain

![Diagram showing incident light intensity, ideal output, observed output, ideal detector, blocking, and dark events.](image)

Ideal output: sequence of impulses at event times

Observed output

$\tau = \text{blocking duration}$

Blocking may be modeled via tracking of detector state with a Markov chain
Mitigating blocking

- Blocking may be mitigated by decreasing the peak incident photon rate (per detector)
- **Temporal**: Increase slot-width and reduce photon rate, while preserving photons/slot
  - lowers date rate, and integrates more noise

```
\begin{align*}
&\text{Ts} \\
&\rightarrow \\
&2\text{Ts}
\end{align*}
```

- **Spatial**: Increase focal length, and illuminate photodetector array instead of single photodetector
  - integrates more noise from multiple pixels

```
\begin{align*}
\text{F/D=8} & \quad & \text{F/D=16}
\end{align*}
```

- Make signal more diffuse in time
- Make signal more diffuse in space
Quantifying blocking loss

• Model blocking by splitting time into small time bins, and modeling the blockage as a Markov chain

\[ \mu \approx \frac{1}{1 + \tau} \]

• Approximate the photodetector output as a Poisson process

• Signal Power Loss: increase in power to achieve fixed capacity

\[ C_b(l_s') = C_u(l_s) \]

\[ \frac{l_s'}{l_s} \approx \begin{cases} \mu & \text{high SNR} \\ \sqrt{\mu} & \text{low SNR} \end{cases} \]

• Capacity Loss: decrease in capacity at fixed signal power

\[ \frac{C_b}{C_u} = \mu \]
Putting it back together

Photon-flux fluctuations caused by atmospheric turbulence and pointing errors are mitigated by interleaving the data.

Angle-of-arrival spread, receiver pointing jitter, and detector blocking are mitigated by increasing the field-of-view of receiver, but dark noise and background radiance are traded to find optimum.

Extinction ratio loss is mitigated with better engineered sources.

Detector temporal jitter is mitigated by utilizing slots that are wide enough.

\[
P_{\text{rx}} = P_t G_t G_r L_s L_a \eta_{\text{pt}} \eta_{\text{t}} \eta_{\text{r}}
\]

\[
P_{\text{rqd}} = \frac{P_i}{(L_b L_j L_f L_t \eta_{\text{det}} \eta_{\text{imp}} \eta_{\text{code}} \eta_{\text{int}})}
\]
Conclusions

• Free-space optical communication systems potentially gain many dBs over RF systems

• Deep-space optical communication links are *photon-starved*
  – presently binary intensity modulation + photon-counting is best known technique to achieve high photon efficiency

• Reliable information transmission requires both physical and link layer engineering:
  – active tracking systems are employed to maintain as stable and robust a line-of-sight as possible
  – Error-correction coding is employed to ensure reliability in the presence of errors

• Receiver noises, detector losses, pointing errors, and atmospheric effects must all be accounted for:
  – Theoretical models are used to analyze performance degradations
  – Mitigation strategies derived from this analysis are applied to minimize these degradations