Interstellar flyby scientific data downlink design

John I Davies

As earlier articles in Principium have illustrated, the interstellar downlink is a very substantial technical challenge. Most recently we reported on work by an i4is team, *i4is delivers Communications Study to Breakthrough Starshot*, in our last issue, P41 [1]. Here John Davies summarises a recent tutorial paper by three major contributors to the interstellar endeavour.

Introduction

In this new paper, *Interstellar flyby scientific data downlink design* (arxiv.org/abs/2306.13550) David Messerschmitt, Philip Lubin and Ian Morrison [2] provide a tutorial review of the interstellar downlink challenge [3].

Messerschmitt et al assume -

- acceleration by directed-energy beam
- the probe is ballistic (unpowered after initial acceleration)
- cruise at 10-20% c and flyby with no deceleration
- probe mass 1 to 1,000 grams
- optical communication using pulse-position modulation (PPM) with error-correction coding (ECC)
- data is downloaded during a period following encounter with the target star and any exoplanets
- very large receiver collection area on or near Earth is composed of individual incoherently-combined diffraction-limited apertures

They provide performance indices of interest to scientific investigators including -

- total launch-to-completion data latency
- total volume of data reliably recovered

And address issues including the interaction between the speed and mass of the probe and the duration of downlink transmission, transmit and receive pointing accuracy, beam size and receiver field of view.

The paper contains many parameters which requires the reader to hunt for their definition until memorised so this article has an Appendix with an index of them.

[1] i4is.org/wp-content/uploads/2023/05/News-Feature-i4is-delivers-Communications-Study-to-Breakthrough-Starshot-Principium41-23052291003-1.pdf

[2] David Messerschmitt is Professor Emeritus of Electrical Engineering and Computer Sciences at UC Berkeley, Philip Lubin is Professor of Physics at UC Santa Barbara, Ian Morrison is at the Curtin Institute for Radio Astronomy, Curtin University, Western Australia.

[3] See also *The Interstellar Downlink*, Principium 31 November 2020 (i4is.org/wp-content/uploads/2021/08/The-Interstellar-Downlink-Principium31-print-2011291231-opt.pdf) for an introduction to the subject and *The Icarus Firefly Downlink*, Principium 36, February 2022 (i4is.org/wp-content/uploads/2022/02/The-Icarus-Firefly-Downlink-Principium36-AW-2202191002opt.pdf) for the specific case of a large fusion powered probe.

Interstellar distances

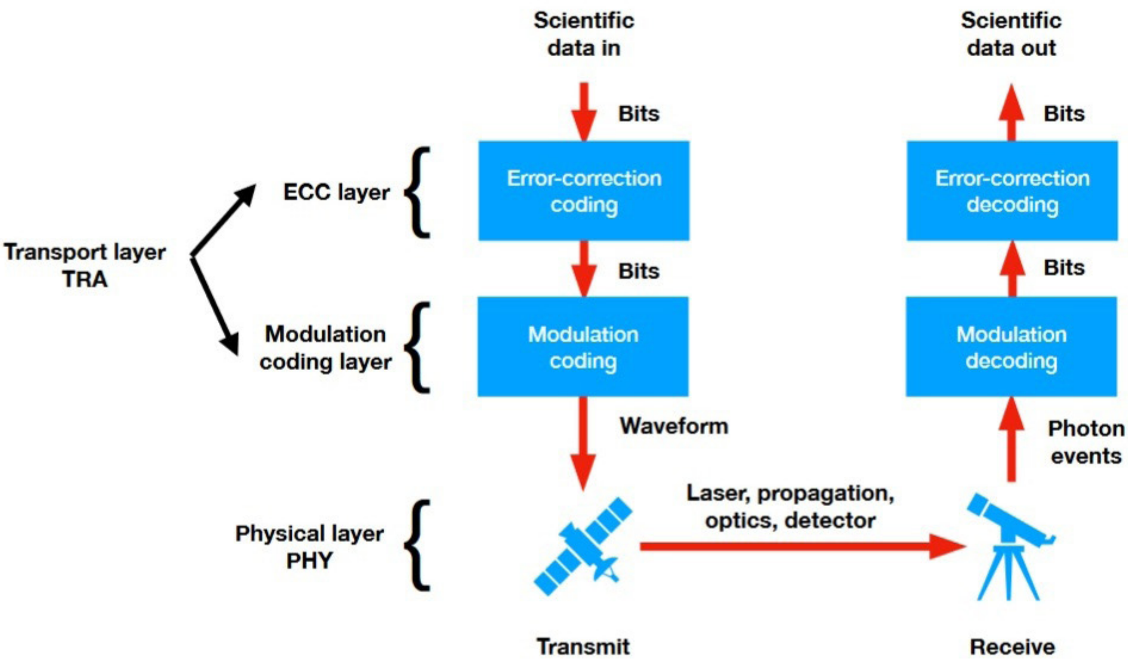
The numerical examples in the paper assume a mission to Proxima Centauri (the nearest star to our Sun) initially launched by directed-energy propulsion from the vicinity of Earth.
The paper assumes an idealised downlink equation [1] -

$$\frac{P_R}{P_T} = \frac{1}{\lambda^2 D_{star}^2} \cdot A_T A_C$$

- to give the relationship between transmit power and received power. To maximise beam collimation they assume the downlink will use the shortest wavelength that penetrates Earth atmosphere efficiently, 350—400 nm also taking into consideration interfering radiation from the target star Proxima Centauri [2].

Protocols

They present a protocol layering of the downlink [3].



Coordinated transmit-receive communications architecture.

Functionality is divided into three layers, each layer with a transmit and receive component. Logically each layer in the transmitter is coordinated with its counterpart in the receiver. The physical layer (PHY) includes everything from transmit laser to optical detection in the receiver. The transport layer (TRA), comprised of the modulation coding and ECC (error correction coding) layers, provides an intensity-vs-time waveform in the transmitter and recovers scientific data from the sequence of photon detection events in the receiver.
Credit (Image and caption): Messerschmitt et al, Figure 2 [4]

[1] Citing S Schelkunoff, H Friis, *Antennas: theory and practice*, Vol. 639, Wiley, 1952
[2] The received signal will be at a longer wavelength (redshifted) due to Doppler at 20%c. For example a transmit wavelength of 292 nm will produce a received signal at 350 nm. There will also be a relativistic effect at this significant fraction of the speed of light, see en.wikipedia.org/wiki/Relativistic_Doppler_effect and en.wikipedia.org/wiki/Relativistic_beaming.
[3] Protocol layering is fundamental to data communications design, see en.wikipedia.org/wiki/Communication_protocol#Protocol_layering
[4] Source: Fig 25 of D G Messerschmitt, P Lubin, I Morrison, *Challenges in scientific data communication from low-mass interstellar probes,* The Astrophysical Journal Supplement Series 249 (2) (2020) 36. iopscience.iop.org/article/10.3847/1538-4365/aba126/meta

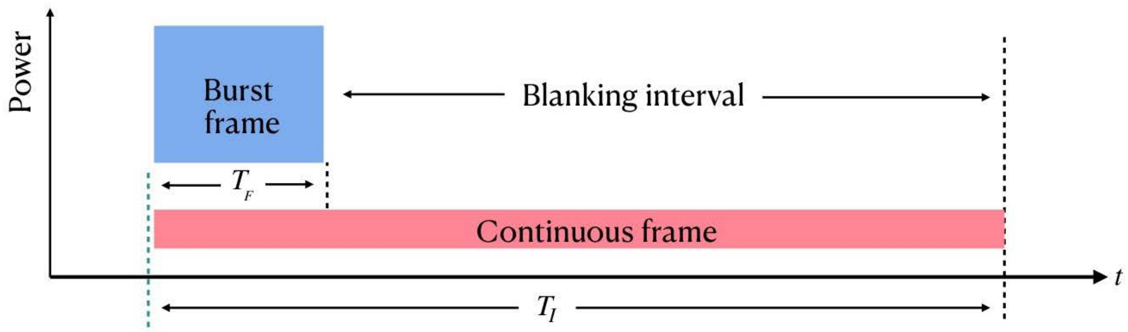
Background radiation

The paper categorises all unwanted data at the receiver as background radiation. It divides this into -

- Noise is radiation that accompanies, but cannot be separated from the data signal because of its overlap in time, wavelength, and direction with sources including cosmic background radiation, the deep star field, zodiacal radiation (due to solar-system dust emissions and scattering) and sunlight and moonlight scattered in the earth's atmosphere. Sunlight is an absolute block for a terrestrial optical receiver so only half of each day, on average, is available to the receiver.
- Dark counts are spurious signals originating in the receiver optics or in the optical detectors. Cooling optics and detectors helps here as does optical bandpass filtering (ie blocking unwanted parts of the spectrum). The remaining unwanted signals tend to appear for all detectors so this problem grows if there are many detectors for one signal.
- Interference can be distinguished from the wanted signal by its time, wavelength, or direction. The major source in this case is the target star itself. But a swarm of probes can produce mutual interference and multiplexing is required. Options are time division multiplexing (TDM probes wait their turn to transmit), frequency division multiplexing (FDM probes transmit at different wavelengths) and space division (SDM probe signals appear from differing directions and/or receivers are widely dispersed).

The paper considers the option of a space-based receiver, which would eliminate atmospheric interference, outages from weather and Earth rotation and wider choice of transmit wavelengths to avoid target star interference [1]. But it rapidly concludes that "given the likely necessity for a massive receive collector, this may not be affordable" [2].

The paper proposes burst-mode transmission, so that a higher power signal occupies just a fraction of possible transmit time, exchanging baseband bandwidth for improved signal to noise ratio.



An illustration of burst-mode transmission, in which a section of transmitted signal of duration T_I is compressed into a shorter duration $T_F < T_I$ at higher power, so that the total energy is not affected. This creates a blanking interval known to the receiver, during which there can be no signal photon detections and any background photons can be ignored. Credit (image and caption): Messerschmitt et al Figure 3

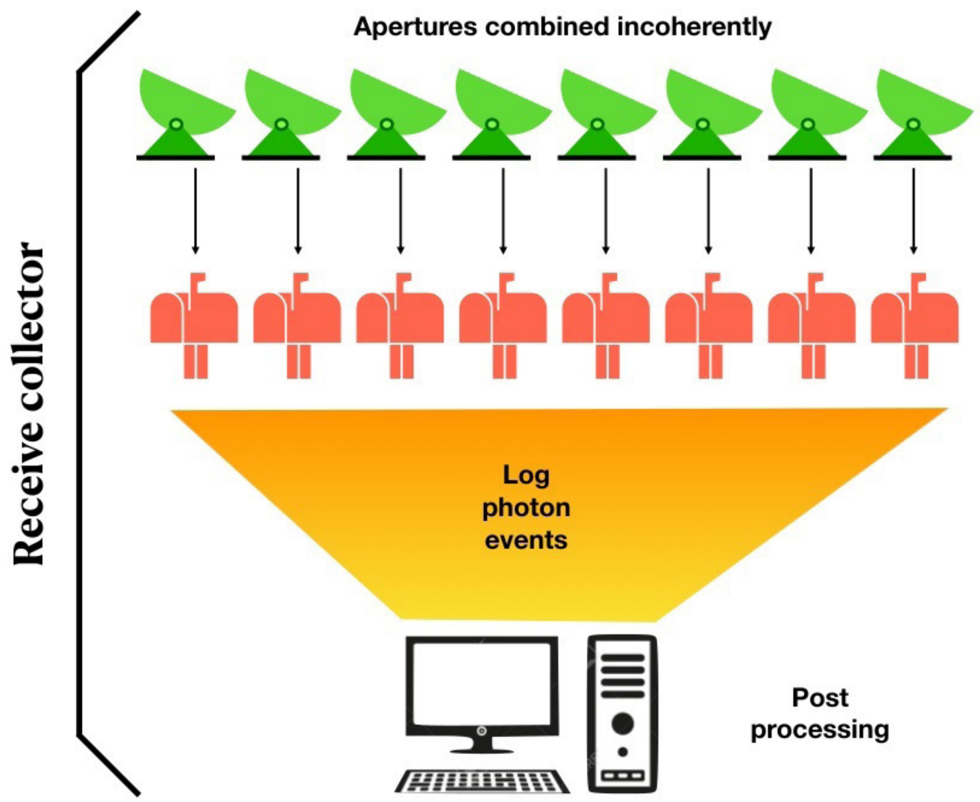
[1] It would also allow cooling of the receiver by radiation (as in the James Webb telescope) and, long term, allow scaling of the optical receivers unconstrained by Earth gravity.
[2] It would have been interesting to see a comparison with a space based downlink receiver which would allow a freer choice of wavelength and no outages either regular, due to rotation of the Earth, or intermittent, due to weather. Will the additional challenges of deploying in space have been overcome by the time probes are launched?

Transmitter and Receiver

The paper states "A single diffraction-limited aperture is the canonical building block of optical electromagnetic transmission and reception." and "Issues related to pointing accuracy are quite distinctive for space vs terrestrial platforms, and for transmit beam vs receive FOV [Field of view (of the receiver)]".

At the transmit end the downlink transmission is likely to be the only function following target encounter so transmit beam pointing is the only issue for spacecraft attitude control and the paper takes the view that the transmitter may be simply part of the spacecraft structure. This also eliminates the moving parts, always a reliability challenge, typical for the transmit antenna of deep space probes. At the receive end the challenge for either a ground-based or a space-based receiver has much less onerous requirements given that scaling to quite large apertures and more sophisticated processing are possible. The paper considers a range of aperture and pointing accuracy issues [1].

The target star and planet will be in motion ("proper motion") for the whole downlink period, assumed to be two years. The paper assumes that the probes will follow this motion but it is not clear to what extent this will happen given the very high velocity of the probes [2].



A receive collector composed of multiple apertures. This achieves simultaneously a FOV controlled by the aperture size together with a larger total collection area to achieve a signal photon counting rate. Λ_S commensurate with data rate $\mathcal{R}$ through the relation $\mathcal{R} = \Lambda_S \bullet$ BPP. For simplicity and without affecting the FOV, the apertures are combined incoherently by simply locating a photon-counting optical detector (shown as a mailbox icon) at each aperture and accumulating their photon counts. The resulting collector is not diffraction-limited, but this is fortunately unnecessary at the receiver.
Credit (image and caption): Messerschmitt et al Figure 5 (citing *Challenges in Scientific Data Communication from Low-mass Interstellar Probes*, Messerschmitt et al 2020, iopscience.iop.org/article/10.3847/1538-4365/aba126/meta Figure 3.)

[1] An additional issue is the fate of the downlink photons encountering the protons of the interstellar medium (ISM) constituting a "bow wind" for such long distance communications.

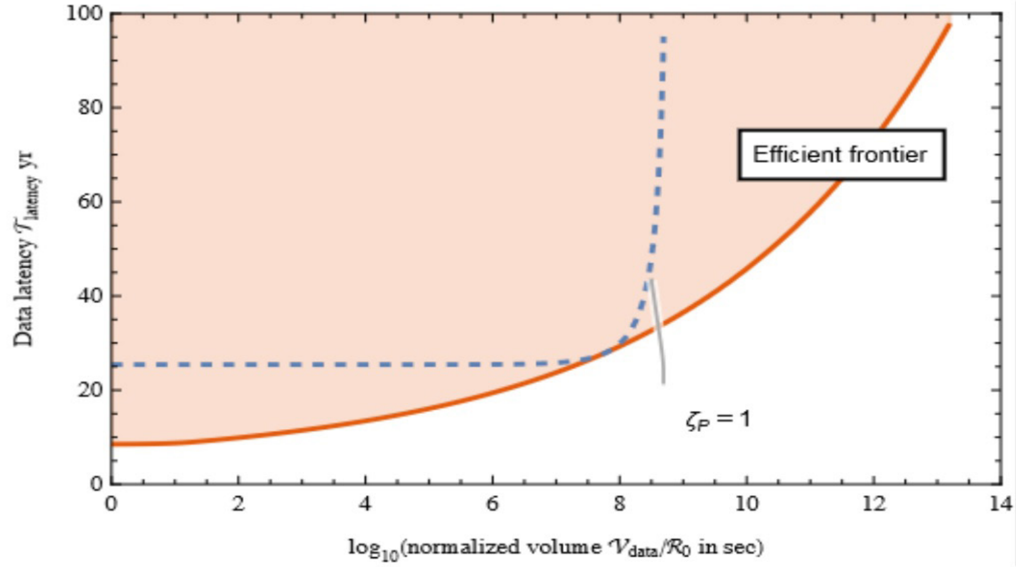
[2] See also the issue of space division multiplexing (SDM) under Background Radiation above.

The data rate $\mathcal{R}$ is proportional to frequency and thus the reciprocal of wavelength, λ, and the paper states that "Depending on area and the chosen technology, maintaining a transmit aperture close to the diffraction limit at shorter wavelengths will necessitate tighter fabrication tolerances proportional to the wavelength." and "With a large number of apertures and associated optical detectors, the average interval between signal photon detections is large. This is another indicator of the aggressive requirements on detector dark counts."
The choice of a terrestrial receiver imposes a number of outage challenges. These are similar to those of a terrestrial optical telescope including inability to receive during daylight hours, instability of light transmission through the atmosphere and cloud cover. More about this in Interstellar Distance and Background Radiation above - and in Error Correction below.

Downlink operation time

The paper assumes that the downlink will take place over several years and identifies a tradeoff between data volume and latency. It sees the consumers of the data, science stakeholders, as asking - how much data do we get back reliably, and how long do we have to wait for that data? This is the data volume V_{data} (total number of data bits returned), and the data latency $T_{latency}$ (elapsed time from probe launch to return of the data in its entirety). The mass ratio ζ_P, the ratio of the actual probe mass to some baseline value, is significant in that if the launch beam and power remain fixed across multiple probe launches, the speed of the probe is directly affected by the probe mass, and hence by ζ_P. In particular the speed decreases as $\zeta_P^{1/4}$ with increasing mass, and the total launch energy increases as $\zeta_P^{3/4}$. The mission design parameters- $\{\zeta_P , T_{down}\}$ can be varied to manipulate the mission performance metrics $\{V_{data} , T_{latency}\}$ to achieve the best $\{V_{data}, T_{latency}\}$ tradeoff. For example, if the launch beam remains fixed (except possibly for the time duration of probe acceleration), a larger ζ_P results in a lower cruise speed for the probe u_P and a longer $T_{latency}$. Overall travel-time increase is "deleterious" but all the other impacts are beneficial. The paper defines a normalised volume of data against data rate $V_{data} / \mathcal{R}_0$ as a performance metric to guide the choice of the mass ratio ζ_P and duration of the operation of the downlink data transmission T_{down} characterised as $\{\zeta_P , T_{down}\}$. $\mathcal{R}_0$ is the data rate (in bits per second) at the beginning of downlink transmission [1].
The paper states that choice of a mission operation point somewhere on the efficient frontier (lower boundary of feasible region of operation) provides flexibility in setting mission priorities. There are several compelling reasons to consciously select different operating points along the efficient frontier for different missions sharing a common launch infrastructure. Considerations include -

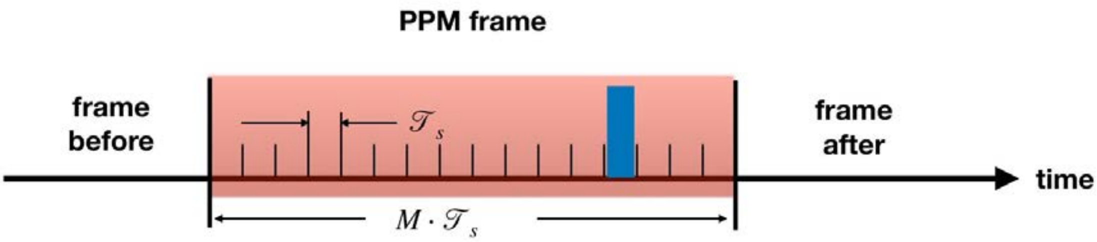
- Priority of large V_{data} versus small $\mathcal{R}_0 / T_{latency}$ (initial data link rate/elapsed time from probe launch to return of the data in its entirety).
- Different probes may carry different types of instrumentation entailing different mass and data volume requirements.
- Evolution of probe technology over time with technology validation first and greater scientific data return later.
- Missions to different targets at different distances.



Plots of data latency Tlatency (in years) vs the log of the normalized data volume Vdata/$\mathcal{R}_0$ (in seconds) where $\mathcal{R}_0$ is the data rate (in bits per second) at the beginning of downlink transmission (data rate declines from there as the square of propagation distance) for mass ratio $\zeta_P = 1$. V_{data} (in bits) is found by multiplying by the assumed value for $\mathcal{R}_0$. Any volume- latency mission operating point within the shaded region is feasible. The lower boundary of this region, called the efficient frontier, is an efficient operating point in the sense of maximizing the volume for a given latency, or minimizing latency for a given volume. The set of operation points obtained by fixing $\zeta_P = 1$ and varying downlink operation duration T_{down} are shown as a dashed curve.
Caption (image and caption): Messerschmitt et al Figure 6

Modulation

The paper chooses pulse-position modulation (PPM) which exploits the advantages of burst-mode in combating noise in the highly demanding environment of tiny probes transmitting over interstellar distances, as explained in section *Background radiation* above [1].
The paper states that PPM combined with an appropriate error correcting code (ECC) protocol layer "can achieve close to theoretical constraints on photon efficiency, subject to assumptions about peak and average power".



Pulse-position modulation divides the received signal power at the optical detector into frames, where each frame is further subdivided into M slots each of duration Ts. The convention is that the received signal power is all within a single slot within the frame, which communicates $\log_2 M$ bits of "raw" information (if the slots are equally likely to be so-energized). The average number of photons detected for each energized slot is K_s, and this is also the average number of photons detected for the whole frame. The peak-to-average power ratio is thus PAR = M.
Credit (image and caption): Messerschmitt et al Figure 7

[1] See also - Report of the paper *The Starshot Communication Downlink* in Principium 27, November 2019, page 28; News item *Challenges in Scientific Data Communication from Low Mass Interstellar Probe* in Principium 28, February 2020, page 17; Survey article *The Interstellar Downlink*, section 4.4.2 Burst pulse-position modulation (BPPM) in Principium 31, November 2020, page 38; *The Icarus Firefly Downlink*, section 5 Possible Laser downlink in Principium 36, February 2022, page 8.

The paper states that the peak power for a PPM transmitter on these tiny spacecraft is limited by the capabilities of semiconductor lasers [1] leading to the need for optical detectors at the receive end which exhibit the lowest possible dark counts and the use of pulse compression.

Error Correction

The example given for transmit error correction suggests an 83.4% overhead for error correction emphasising the considerable challenge of delivering error-free scientific data in this challenging environment [2]. The paper deals in detail with these challenges. One key factor is distribution of the ECC job over substantial amounts of the science data since the ECC performance tradeoff against overhead favours protecting longer blocks of data [3]. There is a big difference between encoding to include the ECC bits at the transmit end and decoding to perform the correction at the receive end. The former must be done with the minimum computing resources (storage and computation) while the latter has the enormous resources available on Earth (or in near Earth space).

The problem of outages, especially for a terrestrial receiver including prolonged weather events, means that even a single image must be transmitted over many times the longest expected outage. The paper gives the example of a one week maximum outage requiring that the image data must be spread over a whole year.

Conclusion

This article has only given a simplified overview of this valuable paper. As the paper remarks -

> *This tutorial has attempted to capture these dependencies as well as the local considerations coming into play within each subsystem. A required scope of core principles is large, spanning quantum mechanics, optics, and device physics on the one hand to information theory and finite field algebra [4] on the other. Such an undertaking is best conducted as collaboration among different types of expertise. In setting requirements and making concrete tradeoffs, that collaboration should include the ultimate stakeholders, which includes funding sources and domain scientists.*

The interstellar downlink is at least as big a challenge as propulsion for a near term interstellar mission. There is much to be done in both theoretical and practical work to achieve that first close-up image of an exoplanet and all the other data we will need to achieve a step forward in our knowledge as large as that between telescope observation of Pluto and the flyby by the New Horizons probe.

This article has benefited from comments kindly provided by Peter Milne and T Marshall Eubanks. Any remaining errors and ommissions are, of course, the responsiblity of the author.

[1] The limits appear to be at around 1 watt, see Song et al, *Processes of the Reliability and Degradation Mechanism of High-Power Semiconductor Lasers*, 2022, www.mdpi.com/2073-4352/12/6/765

[2] Contrast, for example, this with typical Forward Error Correction for mobile phones of 14-25% www.nokia.com/blog/what-the-fec/

[3] For example see en.wikipedia.org/wiki/Hamming_distance and note that to correct an error in one bit of information requires two additional bits, an overhead of 66%, while seven information bits require only three additional bits, an overhead of only 30%.
Error correction techniques have advanced enormously since Claude Shannon first defined the fundamentals of communications theory in the 1940s and the paper simply gives an example of Reed-Solomon coding (en.wikipedia.org/wiki/Reed%E2%80%93Solomon_error_correction) but clearly expects that substantial work is still to be done in this area of design. Note that in most terrestrial communications these techniques are described as Forward Error Correction (FEC) to distinguish from the more common situation in data communications where only error detection is required since corrupted data can be re-transmitted since latency is not critical. Re-transmission with 8 years round trip latency is clearly not feasible.

[4] See math.libretexts.org/Bookshelves/Abstract_and_Geometric_Algebra/Abstract_Algebra%3A_Theory_and_Applications_(Judson)/22%3A_Finite_Fields

APPENDIX - Parameter Index

In the interest of readability of this article and the source paper here are the parameters used, in the order of their appearance in the paper, with the defining page number in the paper.

Parameter	Measure	Page ref
c	velocity of light 300,000 km/sec	1
A_C	size of the collector at the receiver (the receiver aperture area)	3
A_T	transmit aperture area (size of the downlink transmitter antenna)	3
D_{star}	distance to the target star/exoplanet	3
P_R	downlink received power	3
P_T	downlink transmit power	3
λ	wavelength of downlink signal	3
ECC	error correcting code (additional downlink data which permits the receiver to correct some data errors)	5
PHY	physical layer of the downlink protocol stack - the lower layer	5
TRA	transport layer of the downlink protocol stack - the upper layer	5
BPP	bits per photon	7
$\mathcal{R}$	data rate (bits reliably recovered per unit time)	7
Λ_S	photon detections per unit time	7
FOV	field of view (of the receiver)	8
SBR	signal-to-background power ratio	8
PPM	Pulse position modulation	12
T_F	Duration of compressed signal	12
T_I	Duration of uncompressed signal	12
W	signal bandwidth	12
δ	duty cycle factor (for burst-mode)	12
$T_{latency}$	data latency - elapsed time from probe launch to return of the data in its entirety	23
V_{data}	total number of data bits returned	23
T_{down}	duration of downlink transmission	23
u_P	cruise speed of the probe	24
$\mathcal{R}_0$	data rate (in bits per second) at the beginning of downlink transmission [1]	24
Efficient frontier	lower boundary of feasible region of operation	25
ζ_P	mass ratio - the ratio of the actual probe mass to some baseline value.	26
K_s	average number of photons detected for each energized slot (in a PPM frame)	29
M	Each transmitted data frame is subdivided into M slots each of duration T_s.	29
T_s	Each transmitted data frame is subdivided into M slots each of duration T_s.	29