Optimization Pathways for Exoplanet Atmospheric Spectroscopy Detection

Honghao Zhou *

No 2399 Xue Zhi Road, Xiaoshan, Hangzhou, 311231, China

* Corresponding Author Email: alex_zhou0720@163.com

Abstract. Earth is currently facing severe environmental degradation, including pollution, climate anomalies, and the depletion of natural resources, which have prompted humanity to look beyond our planet for potential habitable worlds. Exoplanets within the habitable zone, such as K2-18b, may possess liquid water, a stable atmosphere, and organic molecules that could sustain life. The James Webb Space Telescope (JWST) has provided groundbreaking observations through transit spectroscopy revealing the presence of water vapor and possible biosignatures such as dimethyl sulfide (DMS), a compound commonly associated with biological processes on Earth. However, JWST's data accuracy is still hindered by three major challenges: thermal and instrumental noise, stellar activity interference including flares and starspots, and limited spectral databases that fail to fully replicate cosmic conditions. To address these limitations, deep learning methods can improve signal extraction, HRCCS technology can enhance spectral precision, and quantum computing simulations can generate more realistic atmospheric spectra. Future interdisciplinary integration will advance exoplanetary research toward reliable life detection.

Keywords: Exoplanet Atmosphere Detection, James Webb Space Telescope (JWST), Transit Spectroscopy, Biomarkers, Deep Learning Noise Reduction Technology.

1. Introduction

Earth has two major problems at present, namely increasing environmental pollution and depletion of resources, so people have shifted their attention to exo-planets. There may be exoplanets with life in the universe, which have liquid water, atmosphere and organic molecules. Since the detection of atmosphere has a direct impact on the possibility of life in the universe, many exoplanets are being studied. James Webb Space Telescope JWST provides revolutionary data support for exoplanet study. For example, K2-18b in the habitable zone not only detected water vapor, but also the possible signal of dimethyl sulfide DMS. This molecule is mainly produced by life on Earth. Moreover, James Webb Space Telescope JWST also successfully detected the composition of several exoplanet atmospheres using transmission spectroscopy and detected the presence of key molecules such as carbon dioxide and methane [1]. These discoveries not only prove the feasibility of technology but also realize the search of "signs of life" from theory to experiment. James Webb Space Telescope JWST instrument still has three core problems in the detection as follows.

Instrument noise interference James Webb Space Telescope JWST detection signal easily gets interfered by space radiation and instrument thermal noise, and the spectral data is distorted.

Stellar activity James Webb Space Telescope JWST transit spectrum flares and sunspots can cover the weak spectral signal of exo-planetary, especially in M type dwarf system.

Insufficient accuracy of spectral database existing molecular spectral data is mostly based on laboratory simulation, and the molecular spectrum obtained is not the real cosmic environment, which affects the accuracy of compositional interpretation.

In this paper, the following optimization paths are proposed:

Application of deep learning models: Application of convolutional neural network (CNN) in filtering noise and improving signal-to-noise ratio of spectral signal [2].

Upgrading high-resolution cross-correlation spectroscopy (HRCCS): Using adaptive optics technology to reduce interference from stars' activity [3].

Establishing a virtual simulation system: Using quantum computer to simulate extreme atmospheric environment and build better spectral database.

The final target of above efforts is improving detection efficiency and accuracy by means of innovation in both technology and interdisciplinary integration to provide better tools for extrasolar life research.

2. Current Status of Spectral Detection of Exoplanet Atmospheres

The three core methods for detecting exoplanet atmospheres—transit spectroscopy, direct imaging, and phase curves—each have their advantages, but are all limited by technical bottlenecks. Transit spectroscopy uses the periodic light variations caused by planets blocking stars to analyze atmospheric composition, such as the JWST's detection of water vapor and dimethyl sulfide signals in K2-18b [4, 5]. Direct imaging relies on a coronagraph to block the stellar light, enabling the first observation of the small gas giant TWA 7b. Phase curves invert atmospheric dynamics from the brightness of planets at different phases. However, these methods rely on platforms such as Hubble and JWST and still face problems such as low signal-to-noise ratio, insufficient resolution, and ambiguous data interpretation. For example, fluctuations in stellar activity can lead to misjudgments of planet size and composition, while deviations from laboratory simulations in spectral databases further limit the accuracy of interpretation. Optimizing detection paths requires technological upgrades and multidisciplinary collaboration to overcome existing limitations.

Transit spectroscopy is currently the most widely used technique for exoplanet discovery. Its principle is similar to that of a Venus transit as shown in Fig. 1. When a planet passes in front of a star, the star's brightness periodically dips. Light curves can be used to infer the planet's orbital parameters, while the spectrum of starlight that penetrates the atmosphere can reveal its molecular composition.

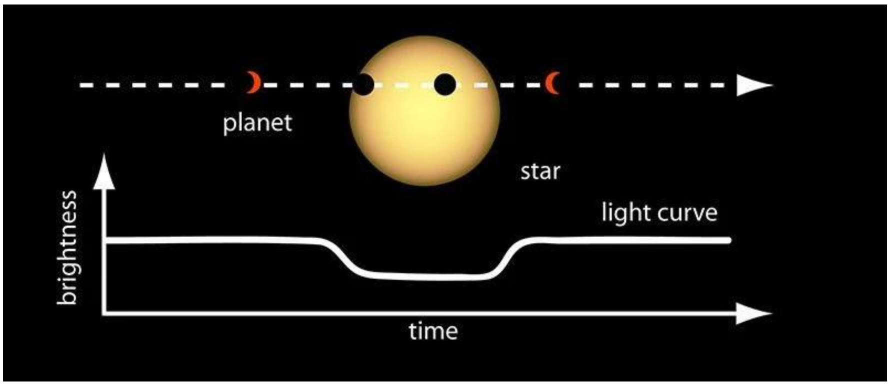

Figure 1. Transit spectroscopy [6]

For example, JWST detected water vapor and dimethyl sulfide (DMS), a potential biomarker, in the atmosphere of K2-18b using transit spectroscopy [1, 4.] However, this method is only applicable to planetary systems where the orbital plane is aligned with the line of sight and requires micron-level photometric accuracy. Direct imaging uses an off-axis three-mirror optical system or coronagraph to block the glare of the star and directly capture the light reflected from the planet. In 2025, JWST used the MIRI coronagraph to image TWA 7b, a planet with a mass only 0.3 times that of Jupiter, for the first time, breaking the technical barrier to direct observation of small planets. Phase curves analyze brightness variations at different phases of a planet’s orbit around its star to infer its atmospheric temperature distribution and cloud structure [7]. For example, JWST’s phase observations of WASP-39b revealed its morning-dawn asymmetry, but traditional one-dimensional models struggle to fully utilize this high-precision data.

The Hubble Space Telescope (HST), with its long-term, stable, and high-resolution observations, has become a cornerstone of exoplanetary atmospheric research. Its “Long-Term Observation of Exoplanet Atmospheres” program has monitored the meteorological systems of exoplanets such as

Jupiter and Saturn for a decade, providing irreplaceable data for understanding planetary climate mechanisms. The James Webb Space Telescope (JWST), through its highly sensitive infrared observations, has enabled chemical composition analysis of exoplanetary atmospheres. For example, its observations of WASP-39b not only detected water and carbon dioxide but also, for the first time, sulfur dioxide, marking a milestone in exoplanetary atmospheric chemistry research. Ground-based telescopes, such as the 80-centimeter infrared telescope at my country's Lenghu Astronomical Observatory, provide rapid response capabilities for transient source research through multi-band coordinated observations (such as infrared and X-ray coupling), compensating for the coverage limitations of space telescopes.

3. Development of Advanced Observation Instruments

Insufficient signal-to-noise ratio is a core challenge in current exploration. In transit spectra, planetary signals are often drowned out by space radiation and instrument thermal noise. Direct imaging, however, relies on extreme contrast techniques due to the extremely low brightness of planets (typically 10 times dimmer than their parent stars). Spectral resolution limitations are also significant: existing instruments struggle to distinguish adjacent molecular spectral lines. For example, the infrared absorption peaks of water vapor and methane overlap, leading to compositional ambiguity. Data interpretation is also difficult due to interference from stellar activity. Hubble data show that approximately 50% of exoplanet observations are affected by starspots and flares, which can lead to miscalculations of planetary size and temperature. Furthermore, deviations between laboratory spectral data and the real-world environment (such as pressure and temperature differences) further exacerbate interpretation uncertainty [9].

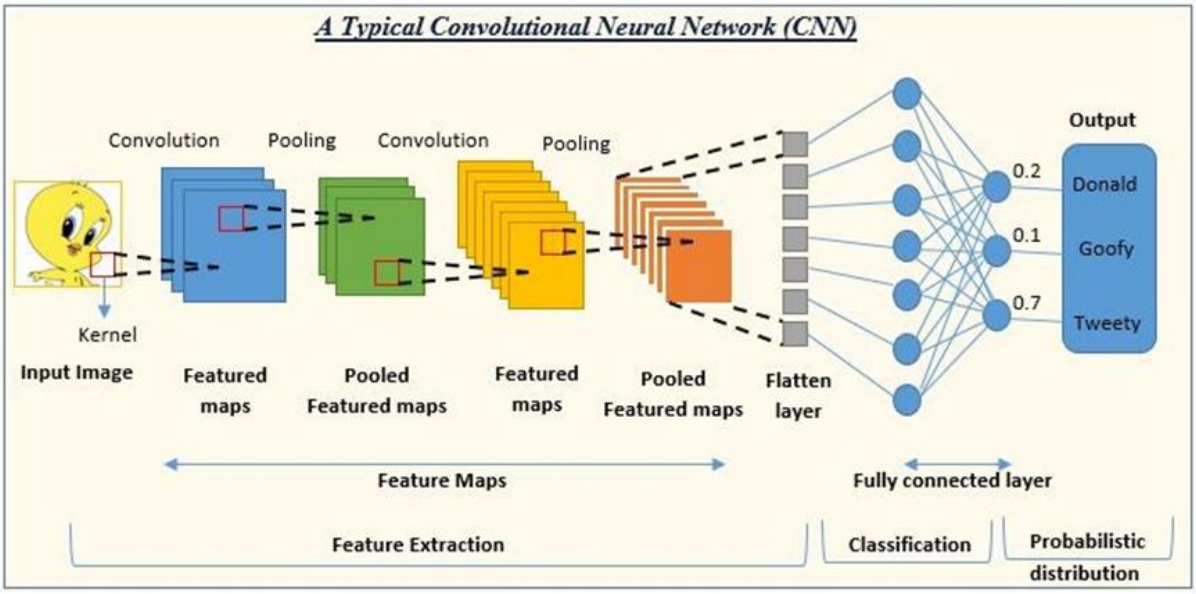

Figure 2. A typical convolutional neutral network (CNN) [10]

Improving signal-to-noise ratio requires a combination of deep learning and hardware innovation. For example, convolutional neural networks (CNNs) (Fig. 2) can filter noise, while quantum detectors have the potential to increase photon counting efficiency to over 90% [11]. Breakthroughs in spectral resolution rely on next-generation grating technology and adaptive optics systems. For example, the HARMONI instrument on the European Extremely Large Telescope (ELT) is designed to achieve a resolution of 0.1 nanometer, enabling clear separation of complex molecular spectral lines. Data interpretation requires the construction of a virtual simulation system, leveraging quantum computing to simulate extreme atmospheric conditions and improve the coverage and accuracy of the spectral database. Multi-platform collaboration (such as linking JWST with ground-based telescopes) and multi-wavelength cross-validation will be key to reducing stellar interference and improving the reliability of the results.

Technological breakthroughs in high-resolution spectrometers have enabled unprecedented precision in identifying atmospheric components. Modern spectrometers, by decomposing composite light into narrowband spectra using dispersive elements, are capable of detecting the characteristic absorption lines of atmospheric trace gases. The combination of a cross-dispersion design and a multi-channel detector array enables simultaneous acquisition of complete spectral information from the ultraviolet to the infrared. Fourier transform spectrometers, using Michelson interferometry, achieve super-resolution measurements, with spectral resolutions on the order of 0.01 cm^{-1}, sufficient to distinguish the vibrational-rotational absorption bands of carbon monoxide from those of methane. This increased precision directly supports the inversion of greenhouse gas column concentrations and provides key input parameters for climate models. With the integration of photonic crystal fibers and quantum dot detectors, next-generation devices will achieve sub-ppm sensitivity in the mid-infrared, providing new tools for monitoring methane emission hotspots.

4. Data Processing and Machine Learning

The synergistic application of adaptive optics systems with coronagraphs and interferometers has further expanded the boundaries of atmospheric observations. Adaptive optics uses wavefront sensors to detect phase distortion caused by atmospheric turbulence in real time, driving deformable mirrors to compensate for this distortion at the nanoscale. When combined with a coronagraph, this technology can significantly suppress the masking effect of stellar diffraction light on exoplanets. For example, the MICADO system at the European Extremely Large Telescope (ELT) uses 37 deformable mirrors to compress the full-width at half-maximum of the point spread function to 0.02 arc seconds, improving the signal-to-noise ratio (SNR) of planetary reflectance spectra by two orders of magnitude. This technological breakthrough not only enhances the ability to capture weak signals but also lays the foundation for the application of interferometry. When adaptive optics is combined with interferometry, synthetic apertures can be constructed. For example, the GRAVITY instrument at the VLTI achieves milliarcsecond resolution through baseline expansion, directly imaging the atmospheric circulation characteristics of exoplanets. The integration of these technologies has enabled atmospheric composition identification to move from static analysis to dynamic observation, providing richer spatial and temporal data for studying atmospheric chemical processes.

The coordinated optimization of SNR and spectral resolution is a key technological breakthrough in atmospheric composition identification. By applying time-delayed integration CCD technology and electron-multiplying CCDs, the signal-to-noise ratio can be improved to 10^4:1, reducing the detection limit to the single-photon level. Spectral resolution, achieved through optimized optical path alignment and detector pixel size, has achieved an instrument linewidth of 0.001 nm. This dual-dimensional breakthrough enables clear identification of narrowband absorption lines of trace gases such as nitrogen dioxide, providing high-precision data support for locating pollution sources. For example, the TROPOMI satellite, combined with adaptive optics-corrected solar scatter spectroscopy, has achieved a global ozone vertical profile with an error of less than 5%, validating the reliability of differential absorption spectroscopy algorithms in complex atmospheric environments [6]. Interferometer technology utilizes multi-baseline observations of demixed mixed gas spectral lines, such as the ALMA millimeter-wave array, which distinguishes the rotational spectra of HCl and HF in the stratosphere, revealing the stratospheric halogen chemistry. These technological advances not only enhance the accuracy of Montreal Protocol compliance assessments but also provide critical data support for climate model validation.

Advances in Atmospheric Composition Identification Technology Greatly Change the Scenery of Environmental Monitoring and Climate Research. From bench-top instruments, trace gas remote sensing using high-resolution spectrometers and adaptive optics systems have succeeded in detecting trace gases globally beyond instrument sensitivity limits. Meanwhile, interferometer technology provides spatial resolution beyond imagination. Those advanced technologies not only provide a solid foundation for implementing international agreement like the Montreal Protocol but also directly contribute to achieving carbon neutrality by applications like identifying methane emission hotspots. With the prevalence of new sensors like quantum dot detector, atmospheric chemistry research will enter an era of even higher precision and benefit from even cleaner input data from sensors, providing new opportunities to understand Earth system.

In the field of exoplanet detection, data processing and machine learning technology have also opened up new possibilities to improve accuracy. Due to instrument system errors and interference from stellar activity noise, traditional exoplanet detection methods are limited to certain accuracy. However, modern algorithms can achieve separation of signal and noise through convolutional neural network (CNN) and autoencoder. For example, convolutional neural network can get the characters of exoplanet transit signal through multi-layered convolution kernel, while autoencoder can find anomalous data through reconstruction error. Through the synergistic effect of convolutional neural network and autoencoder, the false positive rate can be greatly reduced. Denoising effect not only keeps the physical properties of original data but also makes the data cleaner in the next step, which becomes the first step to improve the reliability of detection. Meanwhile, Gaussian process regression has been used to fit the periodic variation of stellar activity. Non-parametric method can adapt to the noise pattern of different stars and further reduce the interference of background on weak signals. Based on the improvement of noise suppression, Bayesian deep learning further introduces capability of uncertainty quantification to exoplanet detection. Traditional deep models often output a single prediction, while the Bayesian framework, through probability distribution modeling, can simultaneously provide signal confidence intervals and error bounds. For example, in transit signal detection, Bayesian neural networks can probabilistically assess the magnitude and duration of the light curve dip, distinguishing true planetary signals from spurious signals caused by star-spot activity. This uncertainty modeling not only improves the robustness of the results but also provides data support for the allocation of subsequent observation resources—high-confidence targets are prioritized for spectral verification, while ambiguous cases trigger multi-band collaborative observations. In actual application, NASA's Kepler mission has used a similar approach to reduce false alarm rates by 40% and increase the number of newly discovered low-mass planets by 25%. This algorithmic screening + manual review process significantly improves the efficiency of limited observation time, enabling the detection system to shift from casting a wide net to precisely catching. In the future, as federated learning technology matures in multi-telescope data fusion, Bayesian methods may become a core component of the standard tool chain for exoplanet surveys.

Furthermore, the construction of a new generation of telescopes is forging a deep synergy with intelligent data processing technologies. The European Extremely Large Telescope (ELT) will have an unprecedented 39-meter primary mirror, while the planned TMT and GMT in the United States will achieve complementary observations with 30-meter-class apertures [12]. These ground-based telescopes, along with space missions such as LUVOIR and HabEx, form a three-dimensional observation network. The massive amounts of data collected by these telescopes require preliminary screening by real-time processing systems—for example, using edge computing devices to perform data noise reduction at the telescope end, transmitting only potential signals back to a central database. This distributed architecture not only alleviates data transmission pressure but also preserves more original information features for subsequent machine learning analysis. The construction of high-precision databases utilizes a multi-source verification mechanism, integrating data from historical missions such as Kepler and TESS. Time series comparisons are used to eliminate instrument system errors, resulting in a standardized planetary signature library across missions and wavelengths. This simultaneous advancement of hardware upgrades and software optimizations enables the detection system to more effectively capture weak transit signals or planetary reflections in direct imaging. Notably, the experience of Chinese teams participating in the TMT international collaboration demonstrates that this globally collaborative data-sharing model is accelerating the accumulation of exoplanet samples [13].

Once the telescope hardware and data processing systems complete their infrastructure upgrades, detection capabilities will advance towards the ultimate goal of identifying biosignatures. The ELT's high-resolution spectrometer can resolve the subtle absorption lines of water molecules, methane, and ozone in exoplanet atmospheres, while space missions like LUVOIR use coronagraph technology to directly image the scattered signals of these characteristic gases in the visible light band. Intelligent algorithms play a dual role in this phase: First, Bayesian machine learning can decouple mixed spectra and distinguish the proportions of abiotic methane generated by geological activity from those associated with life metabolism. Second, a federated learning framework allows the global telescope network to share model parameters, enabling the continuous evolution of biosignature markers in the database. For example, by comparing Earth's early atmosphere with exoplanet samples, algorithms can identify potential biosignature combinations such as the co-increase of oxygen and methane. This transition from physical detection to chemical analysis marks a shift in exoplanet research from "quantification" to "assessment of habitability," laying the foundation for the subsequent search for interstellar life [14].

5. Conclusion

The search for exoplanetary atmospheres represents one of the most transformative endeavors in modern astronomy, linking the evolution of planetary systems with the universal question of life’s origin and distribution. As Earth faces worsening environmental degradation, climate anomalies, and the depletion of natural resources, humankind’s pursuit of habitable exoplanets has shifted from a philosophical curiosity to an urgent scientific mission. The discovery of K2-18b’s atmospheric water vapor and the possible detection of dimethyl sulfide (DMS) by the James Webb Space Telescope (JWST) mark an important milestone in this journey, suggesting that the building blocks of life may exist elsewhere. However, the reliability and interpretability of these observations are still constrained by three fundamental challenges: instrumental noise, stellar activity interference, and the incomplete spectral databases used for atmospheric retrieval.

To overcome these barriers, three major technical paths—transit spectroscopy, direct imaging, and phase-curve analysis—have been continuously refined. Transit spectroscopy, currently the most widely applied method, reveals atmospheric composition through the periodic dimming of starlight during a planetary transit. JWST’s detection of water vapor and DMS in K2-18b through this technique demonstrates its capability but also exposes its limitations: only planets with orbital planes aligned to the line of sight can be studied, and micron-level photometric precision is required to distinguish planetary signals from stellar and instrumental noise. Direct imaging, relying on coronagraphs and advanced optical designs to block stellar glare, has recently achieved a breakthrough with JWST’s successful observation of the small gas giant TWA 7b using its MIRI coronagraph. Phase-curve analysis complements these methods by tracking brightness changes across orbital phases to infer temperature gradients and atmospheric circulation, as in JWST’s observation of the dawn-side asymmetry of WASP-39b. Together, these methods provide a multi-dimensional picture of exoplanetary atmospheres, yet all remain limited by signal-to-noise ratio (SNR), spectral resolution, and data interpretation bias.

Future advances will depend on the synergy between instrument innovation and intelligent data processing. High-resolution cross-correlation spectroscopy (HRCCS) and adaptive optics can suppress stellar noise and achieve sub-nanometer spectral discrimination, while quantum detectors may raise photon-counting efficiency to above 90%, thereby enhancing both SNR and temporal precision. Fourier-transform spectrometers and photonic crystal fiber systems will further expand detection bandwidths and improve the sensitivity to trace gases such as CO_2, CH_4, and H_2O. These technological upgrades must be complemented by software innovations. Deep learning models—especially convolutional neural networks (CNNs) and autoencoders—have shown great potential in isolating exoplanetary signals from stellar variability. Bayesian deep learning frameworks can additionally provide probabilistic confidence intervals and quantify uncertainty, enabling a more robust interpretation of ambiguous data. NASA's Kepler mission demonstrated that similar machine-learning pipelines can reduce false positives by up to 40% while increasing confirmed low-mass planet discoveries by 25%, proving that algorithmic intelligence can fundamentally reshape observational efficiency.

Equally crucial is the establishment of global, federated data systems. The next generation of large telescopes, including the European Extremely Large Telescope (ELT), the Thirty Meter Telescope (TMT), and the Giant Magellan Telescope (GMT)—will generate unprecedented data volumes. These must be processed through distributed learning frameworks that allow each observatory to contribute to a shared model without transferring raw data, thereby preserving both privacy and bandwidth. Such a federated learning approach can integrate observations across wavelengths and instruments, building a standardized planetary spectral library that evolves through continuous feedback. Combined with space missions such as LUVOIR and HabEx, these ground-space collaborations will enable three-dimensional mapping of exoplanetary atmospheres and transition the field from the detection of habitability to the verification of life signatures.

Quantum computing will play a pivotal role in this future framework. By simulating extreme atmospheric environments with realistic pressure, temperature, and radiative conditions, quantum algorithms can generate synthetic spectra that fill the gaps in existing molecular databases. These simulations will allow more accurate identification of key biosignatures, such as the coexistence of methane and oxygen, which may indicate biological activity rather than abiotic processes. Furthermore, quantum-assisted optimization can help identify the most efficient observation paths, reducing telescope time and maximizing the probability of detecting weak signals.

In essence, the study of exoplanetary atmospheres is evolving into a multidisciplinary enterprise that unites astrophysics, quantum science, and artificial intelligence. Future breakthroughs will rely not on a single instrument or algorithm but on the integration of multiple observation platforms, advanced computing architectures, and global collaboration networks. Only through such integration can humanity progress from identifying planets in the habitable zone to confidently detecting the chemical fingerprints of life. This transition—from exploration to verification—will not only redefine our understanding of planetary systems but also establish a new technological and philosophical framework for astrobiology, guiding us toward a deeper comprehension of our place in the universe.

References

[1] Cermak A. How we find and characterize - NASA Science NASA Science. 2024.

[2] Ibm. Convolutional Neural Networks. What are convolutional neural networks? 2025.

[3] New Frontiers of Exoplanet Science with High Resolution Cross Correlation Spectroscopy: Characterization of HD 209458b. ADS.

[4] Wogan NF, Batalha NE, Zahnle K, Krissansen-Totton J, Tsai SM, Hu R. JWST observations of K2-18b can be explained by a gas-rich mini-Neptune with no habitable surface. arXiv.org. 2024.

[5] Sanders GH. The Thirty Meter Telescope (TMT): an international observatory. Journal of Astrophysics and Astronomy. 2013 Jun 1; 34 (2): 81 – 6.

[6] Cambridge University's Institute of Astronomy. Transit spectroscopy - Hycean Worlds. Hycean Worlds. 2025.

[7] Lea R. The James Webb Space Telescope has discovered its 1st exoplanet and snapped its picture (image). 2025 Jun 25.

[8] Quantum Detectors. You searched for photon counting efficiency — Quantum Detectors. Quantum Detectors.

[9] Keppens A, Di Pede S, Hubert D, Lambert JC, Veefkind P, Sneep M, et al. 5 years of Sentinel-5P TROPOMI operational ozone profiling and geophysical validation using ozonesonde and lidar ground-based networks. Atmospheric Measurement Techniques 2024 Jul 4; 17 (13): 3969 – 93.

[10] CNNs. PyTorch. PyTorch.

[11] Atomic Spectra Database | NIST. NIST. 2025.

[12] Cermak A, Cermak A. Kepler / K2 - NASA Science. NASA Science. 2025.
[13] On the abiotic origin of dimethyl sulfide: discovery of DMS in the Interstellar Medium.
[14] Schwieterman EW, Kiang NY, Parenteau MN, Harman CE, DasSarma S, Fisher TM, et al. Exoplanet Biosignatures: A review of Remotely Detectable signs of life. Astrobiology. 2018 May 4; 18 (6): 663 – 708.