Review of Fundamental Characteristics of Coherent and Direct Detection Doppler Receivers and Implications to Wind Lidar System Design

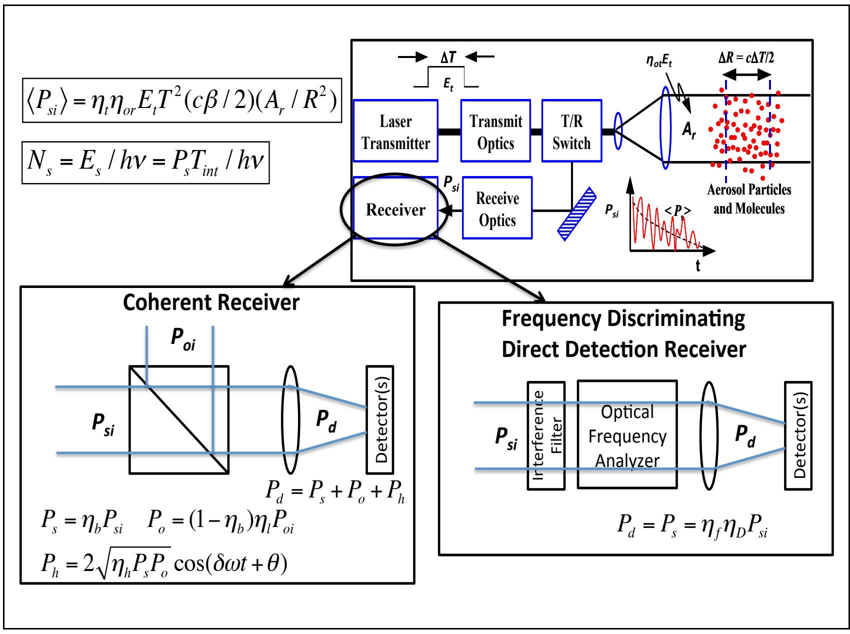
Sammy W. Henderson
Beyond Photonics, 1650 Coal Creek Drive, Ste B, Lafayette, CO 80026 sammy@beyondphotonics.com

Coherent and direct detection lidar systems for wind measurement applications have many similarities but differ fundamentally in photon detection and frequency estimation sensitivity. In this paper we review the key characteristics of each receiver type and implications of these fundamental characteristics on the design of Doppler measurement lidar systems. Although signals from aerosol and molecular scattering are included in the discussion, the emphasis in this paper is on systems utilizing narrowband aerosol backscatter for the wind measurement.

Simplified system designs and approximate equations are used to better illustrate the fundamental characteristics of each receiver type without being bogged down by fine design details of any given system. Of course more detail must be added, beyond what is presented here, in order to describe the performance of any specific real lidar system to a high level of fidelity.

1. Brief Review of Doppler Lidar Systems

Doppler lidar systems of many varieties have been described in detail in numerous past publications with most major system types well summarized in a book chapter dedicated to wind lidar systems.[1] A simplified diagram of a generic Doppler lidar system is shown in *Figure 1*. Most coherent and direct detection Doppler lidar systems utilize single-frequency pulsed transmitters and a beam expanding telescope (or two telescopes in bistatic designs) to transmit and receive the signal. Coherent systems mix the signal with a cw local oscillator signal prior to detection. The resulting heterodyne signal contains the signal amplitude and frequency information. Direct detection systems utilize an optical frequency analyzer, typically an interferometer, and direct detection of the resulting signal intensity from the frequency analyzer to gain information regarding the frequency of the return signal.



P_{si} – Signal power incident on receiver
P_s – Signal power incident on detector(s)
P_o – LO power inc. on coherent det. (CD)
P_h – Heterodyne beat power inc. on CD
E_t – Pulse energy from transmitter
ΔT – Transmit pulse FWHM duration
T_{int} – Total integration time of signal
E_s – Signal energy inc. on detector in T_{int}
N_s – Signal photons inc. on detec. in T_{int}
A_r – Area of receiver aperture
R – Range from lidar to target
T – One way transmission to/from target
β – Target volume backscatter coefficient
η_t – Transmit efficiency (opt. + trunc.)
η_{or} – Receiver optics efficiency
η_b – Coherent receiver beam splitter eff.
η_h – Coherent receiver heterodyne eff.
η_f – Direct receiver interference filter eff.
η_D – Direct receiver freq. analyzer eff.
ν – Signal frequency
h – Planck's constant
c – Speed of light

Figure 1. Simplified Doppler lidar system diagram illustrating key similarities of many of the subsystems and differences of coherent and direct detection receivers. Equations describe collected signal power incident on the receiver and detector(s) for both coherent and direct detection.

A small fraction of the pulse energy transmitted by the lidar is backscattered by the target, which consists of aerosol particles or molecules in the case of wind measurements. The signal power collected by the receiver telescope and incident upon the receiver input (coherent or direct) is given by P_{si} and the number of photons incident on the detector (or detectors) in a given signal integration time, T_{int}, is N_s. The relevant equations are given in Figure 1, and are further described in Reference 1. For fixed aperture size and transmitted energy the number of signal photons presented to the receiver in a given integration time is independent of whether the receiver is coherent or direct.

The contrast between the lidar types is driven primarily by the differences in the manner the detected signal photons are processed by the receiver. The overall receiver efficiency is $\eta_{rec} = \eta_f \eta_D \eta_d \eta_q$ for the direct detection receiver and $\eta_{rec} = \eta_b \eta_h \eta_q$ for the coherent detection receiver, where η_q, η_d, and η_h are the detector quantum efficiency, the detector truncation efficiency (direct detection), and the heterodyne efficiency (coherent detection) respectively. For a coherent system the heterodyne efficiency contains all losses due to wavefront errors – e.g., misalignment, optical aberrations, turbulence – and truncation. The total number of detected signal photons is $N_{sd} = \eta_q \eta_d N_s = \eta_{rec} N_{si}$ for the direct detection receiver and $N_{sd} = \eta_q \eta_h N_s = \eta_{rec} N_{si}$ for the coherent detection system. For the direct detection receiver the detected signal photons are simply the generated signal photoelectrons, but due to effective amplification of the signal when it is mixed with the LO, [1] this in not the case for the coherent receiver. We chose to use detected signal photons instead of photoelectrons to describe the detected signal throughout this paper.

As noted above, the other efficiencies of the lidar system are similar in both coherent and direct detection systems resulting the same number of signal photons being incident on the receiver no matter its type. It is primarily the efficiency, noise character, and frequency sensitivity of the receiver itself that discriminates the different lidar types. The fundamental measurement performance of different receiver types when these attributes are considered is described in the following section.

2. Signal Detection Sensitivity and Velocity Measurement Precision

Coherent Detection Receiver

In Reference 1 a composite model is developed for the variance of the radial velocity measurement precision that a coherent lidar can achieve assuming a fully speckle modulated incident signal. This model originates from the sum of the Cramer Rao lower bound variance estimate without speckle, as derived by Van Trees[2], and the variance estimate due to speckle saturation, as derived by Doviak and Zrnic.[3] The approximate composite model estimating the variance for velocity measurements is given by

$$var(V_{rm}) \sim var(V_{rm})_{CNR} + var(V_{rm})_{sat}\,, \qquad \text{Eqn. 1}$$

where

$$var(V_{rm})_{CNR} = \frac{var(V_s)}{M_e}\left[\frac{2}{CNR_n^2} + \frac{2}{CNR_n}\right] \quad \text{and} \quad var(V_{rm})_{sat} = \frac{var(V_s)}{2M_e}\,.$$

In this expression

- $var(V_s) = \left(\frac{\lambda}{2}\right)^2 var(\nu_s)$ is the velocity variance of the signal incident on the coherent receiver and $var(\nu_s)$ is the frequency variance of the incident signal,
- CNR_n is the narrowband carrier to noise ratio of the signal (the height of the spectral peak above the shot noise floor in the spectral domain[4]), and
- M_e is the total effective diversity of the measurement, including independent pulses, independent coherence times within a range gate, independent range gates, independent polarizations, independent spatial samples (detectors), etc. To increase the diversity, it is important that the samples are independent, uncorrelated, samples.

Equation 1 assumes that the mean narrowband signal to noise ratio, CNR_n has little variation during the total measurement time over which the different independent samples (M_e of them) are collected.

The CNR dependent variance term can be represented in a more fundamental form by recognizing that total number of photons detected is $N_{sd} = M_e CNR_n$, and also recognizing that for a coherent detection system that the total effective number of noise photons detected during a measurement is equivalent to the total diversity during the measurement, i.e., $N_{nd} = M_e$. Although it is ascribed to the local oscillator shot noise in the semi-classical description, the noise floor of one photon per diversity mode is fundamentally due to the coherent system detecting the zero-point energy fluctuations of the vacuum.

Using these substitutions yields

$$var(V_{rm})_{CNR} = var(V_s)\frac{2}{SNR_{ns}^2}\,, \quad \text{where} \quad SNR_{ns} = \frac{N_{sd}}{\sqrt{N_{sd} + N_{nd}}}, \qquad \text{Eqn. 2}$$

SNR_{ns} is the ratio of the signal to the shot noise fluctuations of the signal plus noise without speckle fluctuations.

The saturation term in the sum variance model (Equation 1) is due to *actual frequency variations (random phase variation with time) of the speckle-modulated signal, which are present even at infinite CNR.* This is present even when there is no velocity spread in the target volume. This limiting measurement precision is only improved by averaging over multiple independent samples – hence the $1/M_e$ dependence.

Using Equation 2 in Equation 1 results in

$$var(V_{rm}) \sim var(V_s)\left[\frac{2N_{nd}}{N_{sd}^2} + \frac{2}{N_{sd}} + \frac{1}{2M_e}\right] \qquad \text{Eqn. 3}$$

$$= \frac{var(V_s)}{2M_e}\left[\frac{(N_{sd} + 2M_e)^2}{N_{sd}^2}\right]$$

It should be noted that the simple sum variance model compares well with measurements made from both hard and distributed targets (with varying levels of speckle) and with Monte Carlo simulations. High CNR performance can continue to improve, not fully saturating, if range gates defined by soft-edged functions are used in the processing (see pages 564-567 in Reference 1). At extremely high CNR, a processing window, like a Gaussian window, will have high CNR (useful signals) very far from the window peak, picking up independent speckle phase samples, and effectively increasing the diversity.

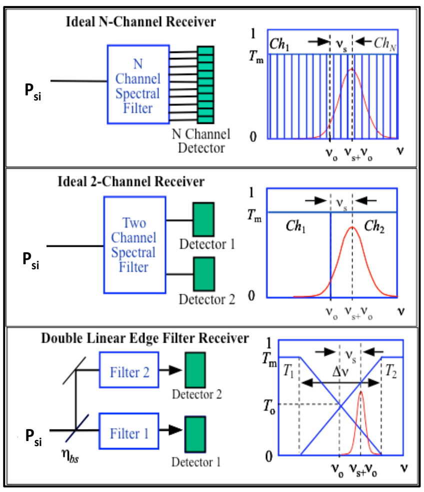
Direct Detection Receiver

Analysis of multiple direct detection system types shows that the variance of the radial velocity measurement can also be approximated using a variance sum model with a SNR dependent term like Equation 2,[1] but with the factor of 2 being replaced with a factor, κ, that is appropriate for the sensitivity of the given receiver design (higher κ for lower sensitivity receivers). For the saturation term it is assumed that the *real phase fluctuations of the speckle-modulated signal* will limit the high SNR performance just as it does in a coherent system. The difference being that for a direct detection system the diversity M_e is typically very large, resulting in a very high level of speckle averaging, so that the saturation level is usually not practically reached. A low diversity direct detection receiver with a narrowband speckle-modulated signal input will experience velocity precision saturation. The general form of Equation 3 approximating the performance for both coherent and direct detection receivers is given by.

$$var(V_{rm}) \sim var(V_s)\left[\frac{\kappa N_{nd}}{N_{sd}^2} + \frac{\kappa}{N_{sd}} + \frac{1}{2M_e}\right] \qquad \text{Eqn. 4}$$

Table 1 summarizes the sensitivity levels for several common receiver types. The receiver types are described in detail in Reference 1. It should be noted that most of the wind lidar community refers to the Two Channel receiver as a Double Edge receiver since the two channels work on the edges of the spectrum. The ratio of κ/η_{rec} is a figure of merit for the receiver, with smaller κ values and higher η_{rec} values improving velocity measurement performance. Note that unlike the other listed receivers, the Double Edge receiver sensitivity is proportional to the width of the edge $\Delta\upsilon$, (determined by the velocity search interval), this can significantly impact the relative performance if the velocity search bandwidth is large compared the spectral width of the signal, $\delta\upsilon$.

Table 1. Sensitivity factors and receiver efficiencies for common receiver types.



Receiver	$\kappa(\nu)$	$\kappa(\nu_o)$	η_{rec}	$\kappa(\nu_o)/\eta_{rec}$
SN-limited Coherent	2	2	$\eta_b\eta_h\eta_q$	$2/\eta_b\eta_h\eta_q$
Ideal Multi-Channel	F_d	F_d	$T_m\eta_f\eta_d\eta_q$	$F_d/T_m\eta_f\eta_d\eta_q$
Ideal Two-Channel	$F_d\frac{\pi}{2}\left[1-\text{erf}\left(\frac{\nu_s}{\sqrt{2}\sigma_S}\right)^2\right]\exp\left(\frac{\nu_s}{\sigma_S}\right)^2$	$F_d\frac{\pi}{2}$	$T_m\eta_f\eta_d\eta_q$	$\pi F_d/2T_m\eta_f\eta_d\eta_q$
Ideal Double-Edge	$F_d\frac{1}{8}\left[\frac{\Delta\nu}{\sigma_S}\right]^2\left[1-4(\nu_s/\Delta\nu)^2\right]$	$F_d\frac{1}{8}\left[\frac{\Delta\nu}{\sigma_S}\right]^2$	$T_m\eta_f\eta_d\eta_q$	$\frac{F_d}{8T_m\eta_f\eta_d\eta_q}\left[\frac{\Delta\nu}{\sigma_S}\right]^2$

Although κ/η_{rec}, the receiver figure of merit, is important to performance, just as important, if not more so, is the noise characteristic of the receivers. The velocity measurement variance estimate given in Equation 4 assumes that the measurement is being made near the actual signal peak, i.e., it does not account for anomalies where random noise peaks significantly removed from the signal peak are mistaken for the signal location.

The three terms of Equation 3 (upper form) represent the contribution from the other noise (left), the shot noise of the signal itself (middle), and the speckle fluctuations of the signal (right). The different regimes of operation are:

Weak Signal Regime: $N_{sd} \ll N_{nd}$, mean signal dominated by mean noise, or $CNR_n \ll 1$ for the coherent. As noted above, for a coherent system the mean noise count per unit bandwidth is equivalent to the diversity, $N_{nd} = M_e$. The signal to noise ratio in this weak signal regime is $SNR_{ns} \sim N_{sd}/\sqrt{N_{nd}}$, which is $\sqrt{M_e}CNR_n$ for a coherent receiver. The proper detection of the signal location in the surrounding noise floor will only be reliable if $SNR_{ns} \geq 1$, otherwise anomalies will result (random noise peaks mistaken for the signal location). Equation 4 for the variance of the measured velocity is only valid if measurements are near the signal location (no anomalies). If the measurement is at the signal peak and not anomalous then the velocity variance in this weak signal regime is given by $var(V_{rm}) \geq var(V_s)(\kappa N_{nd}/N_{sd}^2)$. Therefore, minimizing N_{nd} is very beneficial.

Moderate Signal Regime: $N_{sd} \sim N_{nd}$, or $CNR_n \sim 1$ for the coherent receiver. In this regime the shot noise of the signal itself becomes important. For the coherent receiver, when the signal photon number is approximately equal to the noise, the measurement variance scales as $1/N_{sd}$ ($= 1/M_e\,CNR_n$ for coherent).

Strong Signal Regime: $N_{sd} \gg N_{nd}$, or $CNR_n \gg 1$ for the coherent receiver. When the signal photon number is large compared to the noise count then the variance clamps at $var(V_{rm}) \geq var(V_s)/2M_e$. As noted above, *this is due to actual frequency variations (random phase variation in time) of the speckle-modulated signal even at infinite CNR.*

Figures 2 and 3 provide examples of the performance for coherent and a sample direct detection receiver.

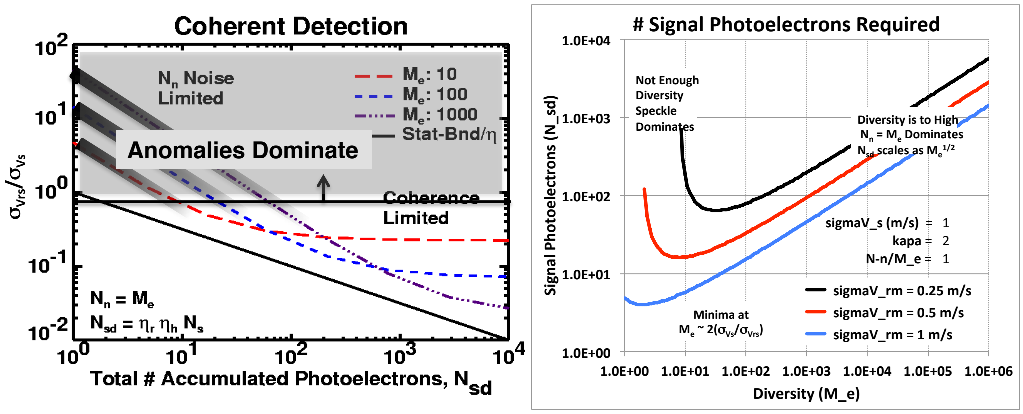


Figure 2. Coherent receiver performance predicted by Equation 4 with $N_{nd} = M_e$ and κ =2. Right plot uses model parameters given in the figure.

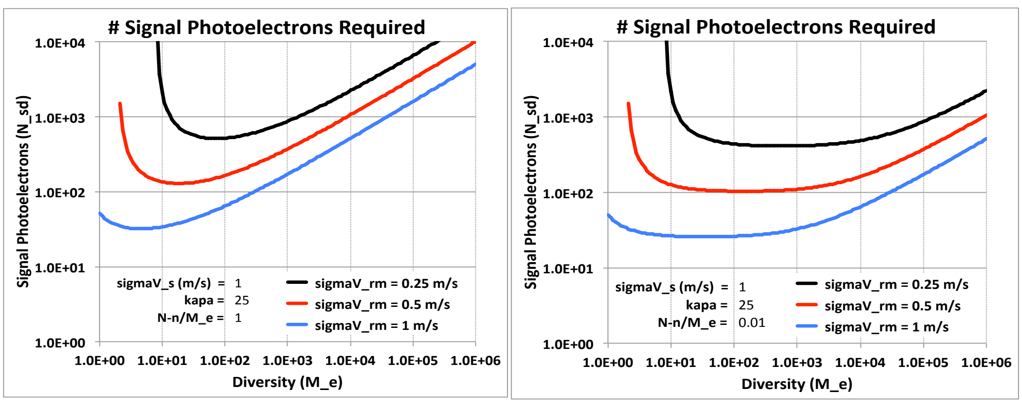


Figure 3. Detected signal photons required for direct detection receiver with κ = 25 (appropriate for a edge based interferometer with moderate velocity search window). Left: total noise counts equivalent to diversity. Right: lower noise receiver with total noise counts being 1% of the diversity. Model parameters are given in figures.

As shown in Figure 2 (right) and Figure 3, sufficient diversity is required to average down the speckle-induced frequency fluctuations. For this approximate performance model, a given level of precision is impossible to achieve unless the diversity meets the condition $M_e \geq var(V_s)/\left(2var(V_{rm})\right)$.

A fundamental difference between coherent and direct detection systems is the noise character of the receivers.

A coherent system has both the advantage, and simultaneously the disadvantage, that the noise level is "only" one-detected-noise-photon per diversity mode. Due to this inherent noise, a coherent detection system design is most efficient when order of one signal photon is detected from the desired target volume per diversity mode (e.g., per resolution element per pulse). The minimum number of photons occurs when $M_e = 2\, var(V_s)/var(V_{rm})$ and at this diversity only 2 photons per diversity mode are required. The inherent coherent detection noise also results in the number of required detected signal photons increasing as $\sqrt{M_e}$ when the diversity is significantly above the optimal value – or equivalently stated, when the signal is significantly below one photon per diversity mode. The advantage of a coherent receiver is that it only requires a few photons per diversity mode near the optimal diversity point, even at relatively high detector noise or background light levels. The bad news is that this optimal diversity region is fairly narrow. In practice what this means is that the coherent system design is optimized by ensuring that $CNR_n \sim 1$, and that sufficient pulses (and/or range gates, if allowed by range resolution requirements) are averaged so that the required velocity measurement precision is realized. If the lidar designer attempts to use pulse energies that are too low, so that significantly less than one signal photon is detected per diversity mode (i.e., $CNR_n \ll 1$), then the desired performance can still be achieved at increased diversity (more pulses and ranges gates averaged) but at the cost of increased average transmitter power. For a fixed range resolution, $E_t\sqrt{PRF}$ must be held approximately constant, which increases the average power transmitted by $\sqrt{PRF}$, where E_t and PRF are the pulse energy and pulse repetition rate of the transmitter respectively. As a specific example, if a 1 J 10 Hz (10 W) transmitter allows optimal performance, then a transmitter operating at a 1 kHz PRF would require 100 mJ pulses (100 W average power).

Direct detection systems do not have this fundamental noise constraint – since they do not sense the vacuum fluctuations. With state-of-the-art detectors, electronics, and interference filters (to reduce background noise counts), the noise in direct detection systems can be significantly less than one-detected-photon per diversity mode. Direct detection systems therefore have the potential to *achieve efficient operation by averaging, even when the signal is substantially less than one-detected-photon per resolution element per pulse*. Example performance curves are shown for moderate- and low-noise direct detection receivers in Figure 3. Transmitter energy and PRF can be traded over a wider range for the low noise direct detection receiver. Following the coherent example above, if the noise of the direct detection receiver is sufficiently low, then a 10 W 1kHz transmitter will work as well as a 10 W 10 Hz transmitter. For a real system with a finite noise rate, the PRF can be increased up to the point where the noise is comparable to the signal level – for example, if the noise rate is at 0.01 noise detections per resolution element per pulse then it is best to keep the signal level above that rate. If the signal detection rate drops significantly below the noise detection rate, you wind up having the same $\sqrt{PRF}$ increase in transmitter power as the coherent system.

Receiver efficiency is also very important so that the photons incident on the receiver have the opportunity to be detected. Receiver efficiencies, η_{rec}, of about 10% are possible for both well-designed coherent and direct detection lidar systems utilizing aerosol backscatter. If the coherent transmitter can be operated reliably and efficiently at the optimal point indicated in Figure 2 (right) then it will be hard for a direct detection system to perform as well due to the lower frequency sensitivity of the receiver (higher κ values) when working with the narrowband aerosol signal. On the other hand, if technology and efficiency considerations drive the coherent system significantly away from the optimal operating point then a coherent receiver will be very competitive due to its wider operating region.

These concepts and implications on the design of wind lidar systems will be discussed further in the presentation.

Acknowledgments: I have had many useful discussions with Phil Gatt on these topics.

[1] S.W. Henderson, P. Gatt, D. Rees, and R.M. Huffaker, "Wind Lidar," book chapter in Laser Remote Sensing, Eds. Fujii and Fukuchi, CRC Press, Taylor and Francis Group, Boca Raton, FL, p 469-722, (2005)

[2] H. L. Van Trees, Detection, Estimation, and Modulation Theory, Part III: Radar-Sonar Signal Processing and Gaussian Signals in Noise, John Wiley & Sons, New York, 1971, Eq. 10.95

[3] R.J. Doviak and D.S. Zrnic, Doppler Radar and Weather Observations , Equation 6.22b (Academic Press, New York, 1984).

[4] D Jacob and P Gatt, Defining CNR, Diversity, Window Efficiency and Resolution of Coherent Lidar for Optimizing System Performance, 16th Coherent Laser Radar Conference, Long Beach, CA 2009