An Introduction to Phase-Stable Optical Sources.

JOHN L. HALL and MIAO ZHU

Joint Institute for Laboratory Astrophysics
National Institute of Standards and Technology
University of Colorado - Boulder, Colorado 80309-0440

1. – Introduction, and ultimate motivation for the study.

This lecture deals with one of the impressive and exciting topics of our time—the development and application of superstable lasers. Their relevance to this School arises in large measure from the incipient marriage of such stable laser sources with the remarkable samples of free-floating quantum absorbers which can be assembled, manipulated and cooled using light forces on atoms—the latter forming the main course of our study here. In addition, such gas samples and stable laser sources represent now in many cases the tools of choice in ultra-precise experiments designed to test our cherished beliefs about the fundamental physical laws. Indeed, a large number of important and impressive measurements have already been carried out using stable-laser techniques. Still, many of us in the field believe that the best is just ahead, becoming freshly enabled in a significant number of examples by progress in the "atom manipulation business". A Smorgasbord selection of interesting physical applications of phase-stable lasers could include the following:

A) Atomic-clock research—narrow transitions in Mg, Ca, Sr, Ba, Ag, Hg⁺, Sr⁺, Yb⁺ ....

B) Ultra-narrow-resonance spectroscopy—Ramsey fringes, atomic fountains ....

C) Investigations of "squeezed" light effects.

D) Special-relativity tests, e.g. Kennedy-Thorndike experiment—look for frequency shifts at 1 cycle/sidereal day.

E) Interferometric gravity wave detectors, ground- or space-based.

F) Passive laser gyro—G.R. "frame-dragging" effect, ....

G) Special atoms: hydrogen, positronium, muonium—QED tests ....
H) "Correlated-emission" laser devices—sub-Schawlow-Townes relative linewidths.

I) Search for long-term drift of the physical constants.

J) Influence of parity nonconservation—spectroscopy of special atoms and chiral molecules.

K) Clean source for "stochastic-excitation" laser spectroscopy.

L) Useful for atom interferometers—atom state preparation, phase readout....

M) Test for atom charge—equality of the fundamental charges.

N) Search for atom electric dipole—test of T-reversal symmetry.

O) Demonstration and measurement of the "geometric phase" in optical spectroscopy.

2. Why insist on phase-stable lasers?

Momentarily we will be discussing laser stabilities at levels beyond those even dreamed possible just a few years ago. For example, we will soon show measured frequency stability at the $10^{-16}$ level. Surely this is good enough? The answer must be only "maybe". It depends on the measurement time we are discussing. The important criterion will be whether the phase excursions induced by the frequency noise are greater than 1 radian. If the unplanned phase deviation stays below 1 radian for the entire duration of our measurement, it is appropriate to speak of a "phase-stable" source. If not, we are effectively dealing with band-filtered optical noise. The central frequency may be well defined on the average, but excessive phase excursions prevent us from following the elapsed phase exactly. The difference between these two cases may seem small and only quantitative, but it has dramatic consequences. Consider how the precision of our measurements scales with measurement time for the two cases.

2'1. Measurement precision with frequency measurement techniques increases as $\tau$. Imagine an experiment to compare a phase-coherent source against a precise frequency standard. The two frequencies are in general incommensurable and, in addition, there will be some amount of amplitude and phase noise. We will characterize the measurements by putting the zero-crossings of the two waves into correspondence. In the measurement time $\tau$ there will be $N = f\tau$ elapsed cycles. At the ends of the measurement interval there can be an uncertainty of up to $\pm 1$ count, so with naive frequency measurements we can expect a measurement imprecision of $1/f\tau = 1/N$. Is this the best we can do? Not at all!
22. Measurement precision with a phase-stable source increases as $3^{3/2}$. For a phase-stable source, a reasonable strategy would be to divide the total measurement time into six parts $N/6 = f/6$. In each of the two end regions we will count $N/6 = f/6$ cycles in an optimized attempt to find the fine fractional-cycle phase relationship between the two waves. The phase measurement of our signal relative to the reference will be limited by such things as noise and digitizing precision, so ideally we may expect in the first end region to find the relative phase to a precision $\sim 1/\sqrt{f/6}$. (This corresponds to the usual $1/\sqrt{N}$ uncertainty rule.) We can measure equally well at the far end in the process of determining the vernier phase offset between the unknown and reference waves at our chosen fiducial point there. Of course, in between the ends we counted off $2N/3 = 2f/3$ actual full cycles. The fractional uncertainty of the total measurement can now be seen to be $(\sqrt{3}/2)/N^{3/2}$. If we can invent some scheme to take better advantage of the high $S/N$ to improve the phase measurement at the ends, so much the better. That will change the numerical constants, while retaining our wonderful $N^{3/2}$ scaling law: with $N \sim 10^{15}$ cycles corresponding to measuring an optical frequency for some seconds, it is important to try to win the $N^{1/2}$ "extra" factor due to averaging the phase in the end regions.

This measurement strategy may have been first pointed out by Snyder [1] in connection with LambdaMeter wavelength measurements. In view of this powerful advancement of potential measurement accuracy, it is attractive to study the factors that control the phase coherence of a feedback oscillator. First, however, in this lecture we need to develop the tools for discussing the performance of such sources.

3. – Thinking about frequency stability—a minitutorial.

3'1. Basic ideas: AM vs. FM. – We can begin our discussion usefully by emphasizing the conceptual distinction between amplitude modulation and frequency modulation. While such modulations on radiofrequency waves may be useful for purposes such as selling soap flakes or breakfast cereals, for us both of these modulations are troublesome: we prefer to have just the sine wave carrier, free of modulation! Such a pure electromagnetic field could be written in the form

$$
\varepsilon = \hat{E}E_0 e^{-i\omega t - i\xi_0} + c.c.,
$$

where the polarization, amplitude, frequency and initial phase are all fixed quantities. Of course, a real field as generated and measured in the laboratory will have a much more general form, as all these quantities will be in general time dependent. A useful level of generality for us will be to allow the ampli-

43 - Rendiconti S.I.F. - CXVIII
tude and phase to be slow functions of time:

\[ \epsilon = \hat{\epsilon} E(t) \exp \{-i\omega_0 t - i\xi(t)\} + \text{c.c.} \]

We will ignore the vector character of the field in the following discussions, focussing instead on the slowly varying amplitude and phase functions, \( E(t) \) and \( \xi(t) \).

It is useful to consider whether we can ignore as well the amplitude noise effects. For example, we will be discussing lasers well above threshold, emitting a substantial coherent power level, say > 1 \( \mu \)W. This corresponds to \( \sim 3 \times 10^{12} \) photons/s. Suppose we take the electronic measurement time interval to be 1 ns, during which we would have collected some 3000 photo-electrons in our detector. The amplitude uncertainty would be \((3000)^{1/2}\), and the fractional amplitude uncertainty \( \sim 1/50 \). At this level it would be exceedingly unlikely for us to mistake one cycle, but the measured frequency would have to be restricted to \( f < 500 \) MHz because of the 1 ns sampling bins.

In order that the measured frequency should be so low—only 1 p.p.m. of the optical frequency—it will be appropriate to use an optical heterodyne setup to compare this laser with a stable reference laser. When the two laser fields are combined into the same spatial mode by a beamsplitter, the detected power will show a beat. (We could change the offset between the two lasers to control this difference frequency. See later discussion.) Suppose we can see that the laser beat is as stable as one would expect, based on the signal/noise obtained. Then, if the beat frequency were reduced, longer measurement bins could be used, and this would bring improved signal/noise performance. So one can see that—for a stable enough source—the natural time scale for the measurement is no longer the minimum convenient electronic measurement time. Rather, we focus instead on the time scale fixed by the source itself, i.e. the inverse of its spectral linewidth.

As an example, a decent HeNe laser may have a 10 kHz linewidth. That means it has a phase slip of \( \sim 1 \) radian only after 100 \( \mu \)s. Remembering the amplitude calculation, then, we have the leisure to collect photodetected charges for this much longer time scale, yielding \( 3 \times 10^8 \) electrons. The fractional noise in this longer averaging time is now seen to be \( 1/17000 \), i.e. only some 60 parts per million. The quantum amplitude effects are rapidly disappearing!

The bottom line of this argument is the following: The quantum noise is not very important for our purposes. Furthermore, lasers generally operate under conditions under which the internal fields rapidly increase under the influence of the multipass gain until the medium becomes saturated, giving just one more photon to the field in response to each new atomic-excitation event. For some lasers (notably CO\(_2\) and He-Ne), it can turn out that the technical noise in the pumping process can be reduced to near or below such a fluctuation level as estimated above. (One is helped in this noise reduction direction by the fact that the laser pumping rate passes through a maximum as a function of the dis-
charge current.) So the fundamental amplitude noise is very small and basically not interesting in the well-above-threshold domain of interest to us. Therefore, in what follows we will only focus on the *phase noise*, *i.e.* we take $E(t) \Rightarrow E_0$.

### 3.2. FM theory for the laser field without amplitude noise.

In eq. (2) we have represented all frequency noise aspects of the laser by a slowly varying phase, $\dot{\varphi}(t)$. Eventually we will discuss this as being a stochastic variable. For a beginning, we can study the simplest case, that of a single sinusoidal *phase* modulation. We take the laser phase to be

\begin{equation}
\dot{\varphi}(t) = \beta \sin \omega_m t,
\end{equation}

where the parameter $\beta$, called the modulation index, is expressed in radians and represents the peak phase excursion induced by the modulation. The modulation rate is $\omega_m$. The laser field now appears as

\begin{equation}
\varepsilon = \hat{e} E_0 \exp[-i\omega_0 t - i\dot{\varphi}(t)] = \hat{e} E_0 \exp[-i\omega_0 t - i\beta \sin (\omega_m t)].
\end{equation}

In eq. (4), we may identify the instantaneous optical frequency with the instantaneous rate of change of the total phase, *i.e.*

\begin{equation}
\text{instantaneous frequency} = \omega_0 + d\varphi(t)/dt.
\end{equation}

(As an aside, we can see that the spectral character of the laser field will be controlled only by the time-varying part of the phase. It does not matter if the central frequency was $\omega_0$, or $2\omega_0$ or $\omega_0/2$. The fact that the central frequency can be changed by optical heterodyne techniques will be useful to us in the laboratory, allowing effective use of highly refined radiofrequency techniques in the optical domain.)

We want to consider the frequency changes due to the phase term

\begin{equation}
\omega^* = (d/dt) \dot{\varphi}(t) = \beta \omega_m \cos \omega_m t,
\end{equation}

\begin{equation}
\omega^* = \Delta \omega \cos \omega_m t.
\end{equation}

One can see that there is an effective frequency excursion $\Delta \omega \equiv \beta \omega_m$. To discern the presence of such a modulation, we will be making sequential frequency measurements over many small time intervals, $\tau \ll 1/\omega_m$.

Suppose that, instead of considering the process in *phase* modulation terms, we saw the laser frequency noise on our instruments, and were led to consider it to be a *frequency* modulation process. We could have then chosen the laser field
to be in the form

\begin{equation}
\varepsilon' = \hat{\varepsilon} E_0 \exp \left[ -i(\omega_0 + \Delta\omega \cos(\omega_m t)) t \right].
\end{equation}

Comparing eqs. (6) and (7), we can see an alternative «definition» of \( \beta \), viz.

\begin{equation}
\beta = \Delta\omega / \omega_m.
\end{equation}

Note that, in FM, \( \beta \equiv \text{frequency excursion/frequency modulation rate} \). If there is but a single modulating frequency, the phase vs. frequency modulation distinction is not very meaningful.

3.3. Representation of phase modulation by a spectrum of sidebands. – An important mathematical transformation allows us to rewrite eqs. (4) and/or (7) in terms of Bessel functions of the type denoted by \( J_n(\beta) \). We will work with eq. (4):

\begin{equation}
\varepsilon = \hat{\varepsilon} E_0 \exp \left[ -i\omega_0 t - i\beta \sin(\omega_m t) \right] =
\end{equation}

\begin{align*}
&= \hat{\varepsilon} E_0 \sum_{n=-\infty}^{\infty} J_n(\beta) \exp \left[ -i(\omega_0 + n\omega_m) t \right] = \\
&= \hat{\varepsilon} E_0 \left[ J_0(\beta) \exp \left[ -i\omega_0 t \right] + \sum_{n=1}^{\infty} J_n(\beta) \left[ \exp \left[ -i(\omega_0 + n\omega_m) t \right] + \right. \\
&\left. \quad + (-1)^n \exp \left[ -i(\omega_0 - n\omega_m) t \right] \right] \right].
\end{align*}

This remarkable transformation (especially in its somewhat daunting representation on the two lower lines) actually shows us clearly what has become of our optical sine wave when we impressed a sinusoidal phase modulation upon it: it has developed a «comblike» structure in frequency space. We speak of the \( J_0(\beta) \) term at the original frequency \( \omega_0 \) as the optical «carrier» frequency. The \( \omega_0 + n\omega_m \) set of terms in the summation appear at successively higher frequencies, their common separation being the modulation frequency \( \omega_m \). The amplitudes are given by the higher-order Bessel functions, \( J_n(\beta) \), which rapidly become small for \( n > \beta \). In the second set of terms in the sum, the \( \omega_0 - n\omega_m \) frequency components are all below the carrier. An interesting and important oddity appears in this second term as well, namely the sign alternations. Odd terms—we say odd-order sidebands—below the carrier have negative amplitudes.

We can calculate the energy in the wave, by calculating \( |\varepsilon \cdot \varepsilon^*| \), to appreciate the physical meaning of these sign reversals. Certainly we expect heavy cancellations to occur. Direct calculation gives

\begin{equation}
\varepsilon \cdot \varepsilon^* = E_0^2 \sum_{n=-\infty}^{\infty} J_n(\beta)^2 = E_0^2.
\end{equation}
This shows the apparently miraculous result that, in spite of the complicated, multifrequency time dependence contained in this wave, its power is time independent! Following the work of engineers and physicists for more than half a century, we will be able to make good use of the details of this phase-modulated wave when we try to build laser frequency discriminators and optical-frequency servo systems.

It is worth discussing the number of sidebands which have significant amplitudes. As noted before, the Bessel function amplitudes $J_n(\beta)$ decay rapidly for $n > \beta$. The large-$\beta$ case could represent a laser with FM mode locking between its successions of cavity orders. However, of interest in connection with highly stable lasers will be the case of small $\beta$, ideally $\beta \ll 1$. In this domain, the spectrum consists mainly of the carrier and the two first-order sidebands. The case of best stability will be when even these sidebands are very weak. From the mathematical character of $J_1(\beta)$ for $\beta \ll 1$, we have relative sideband power $\sim J_1(\beta)^2 = \beta^2/4$. Looking back at eq. (8), our second relation «defining» $\beta$, we can see the conditions under which a significant value may be expected for $\beta$. Suppose we have a certain frequency deviation $\Delta \nu$, driven, for example, by white noise in the frequency discriminator (more about this below!). From eq. (8) it is clear that low modulation frequencies, $\omega_m$, will be the dangerous ones, since they offer us the «best» chance for $\beta \geq 1$. In such a case we will see pronounced modulation effects, manifested as many modulation sidebands, mainly at the very low frequencies. Of course, low-level modulation effects may exist at higher modulation frequencies, but, as is clear from this discussion, one will expect only small $\beta$ and so only weak sidebands at these higher modulation frequencies. We will return several times to the underlying physical idea: when the modulation process causes the laser frequency to deviate from a constant value, we will accumulate a larger phase error in proportion to the period of the modulating wave form. So even a small frequency deviation at low frequency can lead to $\beta \geq 1$, which in turn leads to prominent sideband structures. (Note that in such a case the spectrum of the laser field contains components at harmonics of the basic perturbation frequency.)

So far, we have considered a process with but a single modulation frequency. It is clear that the language of frequency modulation and that of phase modulation are absolutely equivalent for such sinusoidal perturbations. We will need to generalize our representation of the undesired laser frequency noise, and develop language suitable for describing a continuous distribution of perturbing frequencies. But at this point it may be helpful first to compare optical and r.f. sources.

34. Comparing r.f. and optical sources. – In the r.f. domain, technology is rather well developed, and low-noise amplifiers are ubiquitous. Similarly, r.f. resonators are well developed and stable. So an r.f. source may have little extraneous broad-band noise contained within its oscillation loop, beyond the min-
imum associated with thermal noise, etc. On the other hand, periodic modulation by mains frequency harmonics seems rather likely, for example due to inadequate filtering of the supply voltages or, more likely, direct magnetic pickup in some r.f. component whose phase is just the slightest bit sensitive to the ambient magnetic field. Considering the relatively low carrier frequency, just a little FM will appear as a rather large fractional frequency modulation. Thus for r.f. domain devices it is rather typical that the most conspicuous FM sidebands are linelike, sharp spectral lines carrying the (Fourier) frequency of their source(s). Sometimes additional radiofrequencies are present within the one apparatus, as used for frequency synthesis purposes, for example. It is quite expected that the output will then also contain these frequencies or nonlinear mixing products arising from their presence. One speaks of «spurious» outputs, or «spurs».

On the other hand, many optical laser sources are based on special discharges run under conditions appropriate to produce population inversion and gain. Often these are not the same conditions which would produce an ideally quiescent plasma discharge. In some cases the plasma instabilities are nearly essential to the laser action! In a low-power laser medium such as helium-neon, as one increases the current from its lowest value, it is often possible to trace through the onset of sinusoidal plasma oscillations. Period doubling, limit cycle oscillations and a variety of other routes very often lead the plasma discharge down the road to chaos before one has reached the traditional laser operating point. However, well-designed HeNe laser tubes may have only a single sinusoidal current oscillation peak, typically at 1 MHz or somewhat lower. Understandably, the temporally changing plasma refractive index leads to frequency changes of the laser. Generally there is a great tendency for laser frequency sources, even without servo control, to exhibit frequency noise with a continuous distribution. In addition there may be a few relatively sharp spectral peaks due to acoustic resonances in the plasma tube design, etc. But the native laser phase noise is even worse: with a frequency \( \sim 10^{18} \) Hz, even a minuscule fractional change will offer us an extra radian in a very short time!

So, finally, to reach the domain of optical-frequency stability of present interest, \textit{i.e.} the sub-Hertz domain, it is basically certain that the laser will be operating under servo control of its output frequency. The control information will be obtained by detecting part of the laser emission with a photodetector, measuring either the sharp resonance of a stable cavity or quantum absorber, or some kind of an optical heterodyne signal. In either case, it is easy to see that shot noise will be present in the detector's output, and that the servo mechanism must feed back this continuous noise onto the laser's frequency control actuator. Thus it is useful for us to become more precise about the specification of frequency noise, particularly that deriving from a continuous process.

In summary of these qualitative ideas, we may say that different sources have different spectral distributions of frequency noise. In addition we may dis-
tunguish spectral «lines» vs. a distribution of broad-band noise. We may find frequency noise components which are fixed, or slowly drifting in Fourier frequency, or which may be of a modulated frequency.

3.5. Spectral representation of noise: phase spectral density $S_{\phi}(f)$. — We now begin an introduction to a description of our noise modulation processes when they form a continuous distribution, rather than the monochromatic modulation process discussed up to now. It is useful to deal with power in one of the laser sidebands generated by a modulation process. Using eq. (9) in the limit of small modulation index, we have

\begin{equation}
\frac{P_{\text{SSB}}(f)}{P_C} = \left( \frac{E_{\text{SSB}}(f)}{E_C} \right)^2 = (J_1(\Delta\phi_{pk}))^2 \propto \frac{1}{4} (\Delta\phi_{pk}(f))^2 = \frac{1}{2} (\Delta\phi_{\text{r.m.s.}}(f))^2 = \frac{1}{2} \Delta\phi_{\text{r.m.s.}}^2(f),
\end{equation}

where $P_{\text{SSB}}$ and $P_C$ mean single sideband and carrier power, respectively, and the modulation process causing the phase modulation $\Delta\phi$ is occurring at the Fourier frequency $f$. Moving towards a continuum description of the noise modulation source, we will view these noise sidebands as arising from a spectral density of frequency excursions—all contributing to form a resultant average $\Delta\nu$ and $\Delta\phi$. The noise around the Fourier center frequency $f$ will be associated with a bandwidth $b$ Hz, leading to

\begin{equation}
\left. \frac{P_{\text{SSB}}(f)}{P_C} \right|_{b^{-1} \text{Hz}} = \frac{1}{2} \frac{\Delta\phi_{\text{r.m.s.}}^2(f)}{b} \left[ \frac{\text{rad}^2}{\text{Hz}} \right].
\end{equation}

This fractional-power-in-one-sideband measure, denoted by $L(f)$, is sometimes used in the frequency standards community. But in general it is regarded [2] as an inferior specification tool, because it depends on the small-angle approximation used to obtain eq. (11). In the following, we generally will drop the designation «r.m.s.» on $\Delta\phi$ as it leads to no ambiguity.

A robust and useful quantity is the phase spectral density, denoted by $S_{\phi}(f)$. It measures the phase variance, per unit bandwidth, at a frequency $f$ from the optical carrier frequency, $v_0$, i.e.

\begin{equation}
S_{\phi}(f) = \frac{\Delta\phi_{\text{r.m.s.}}^2(f)}{b} \left[ \frac{\text{rad}^2}{\text{Hz}} \right].
\end{equation}

Evidently we have the relation $L(f) = 1/2 S_{\phi}(f)$, for weak sidebands.

3.6. Power spectral density: $P_\phi(\nu)$. — In the above, while we have taken the power in one optical sideband in defining $L(f)$, the frequency variable is measured relative to the optical carrier. A more transparent and universal
kind of optical-power density can be defined using the optical frequency as the argument.

We make the definition that power spectral density, \( P_E(\nu) \), is the optical (noise) power within a 1 Hz bandwidth centered on the optical frequency \( \nu \). Power spectral density can be defined on a per-Hz basis, as is done here, or on a per-radian/s basis. We use the symbol \( P_E(\nu) \) to indicate optical-power spectral density (rather than \( S_p \)) to emphasize that it is a local measurement in the optical domain, without any necessary reference to a nearby laser line. Some authors call this the optical-power spectrum, but this language is tricky when both bright spectral lines and noise are covered in the same measurement.

3.7. Phasor picture of phase noise. – A valuable tool for understanding a.c. circuit issues is the «phasor picture» in which the main frequency-dependent phase evolution has been suppressed by transformation to a suitable rotating frame of reference. In this frame currents and voltage differences can be expressed vectorially, with the phasor angles representing the relative phases. Mapping this construct over to our optical case, the phasors will be called upon to represent optical electric fields. As we noted, phase noise will be relatively conspicuous in our optical problems.

\[
\begin{align*}
\Delta \phi_{\text{r.m.s.}} &= \sqrt{\frac{E_{\text{n, r.m.s.}}}{E_{\text{r, r.m.s.}}}} \\
E_{\text{peak}} &= \sqrt{2} E_{\text{r.m.s.}}
\end{align*}
\]

for small \( \Delta \phi_{\text{r.m.s.}} \):

\[
\Delta \phi_{\nu}^2 = \left( \frac{E_{\text{n, r.m.s.}} \sin \theta}{E_{\text{r, r.m.s.}}} \right)^2.
\]

A similar phase jitter is due to noise in the lower sideband, \( \Delta \phi_{\text{L}} \):

\[
\Delta \phi^2 = \Delta \phi_0^2 + \Delta \phi_{\text{L}}^2.
\]

We consider that this phase noise is produced by dense spectral components within a bandwidth \( b \):

\[
S_{\phi}(f) = \frac{\Delta \phi^2(f)}{b \text{ rad}^2/\text{Hz}}.
\]

\( S_{\phi}(f) \) is the basic quantity: the spectral density of phase fluctuations.

Fig 1. – The physical interpretation of \( S_{\phi}(f) \).

Suppose we know something about the signal and noise fields. A useful way of dealing with this case is indicated in fig. 1. We show the average r.m.s. elec-
electric field as a fixed phasor. While its phase is really evolving at the optical frequency \(\omega_0 = 2\pi\nu_0\), we work in a suitable frame in which this vector is fixed. To represent the fluctuating amplitudes and phases of the full actual instantaneous electric field, we can invoke an appropriate «noise field» \(E_{N,r.m.s.}\). This could be taken as the fluctuating part of our actual field, or it could be some other fluctuating field in the problem. In the limit, it would be the vacuum field of the spatial-temporal mode into which we also have put real photons. For the systems of interest in this discussion, we will believe that this fluctuation or noise field has zero mean and a well-defined mean-squared value. This effective noise phasor is measured from the tip of the average field (we work with r.m.s. measures of noise and signal fields). Since the noise field has a random phase, its vector tip will describe a circle as indicated in fig. 1. For our well-stabilized laser case, we can hope the phase noise \(\Delta \phi_{r.m.s.}\) is small. While the random variation of direction of \(E_{N,r.m.s.}\) affects both amplitude and phase of the total field, we are interested only in the transverse projection that gives phase changes. Recalling that \(E_N\) has zero mean, a very simple approximate formula can be given:

\[
\Delta \phi^2 = \left( \frac{E_{N,r.m.s.}^U \sin \theta}{E_{S,r.m.s.}} \right)^2 = \frac{1}{2} \left( \frac{E_{N,r.m.s.}^U}{E_{S,r.m.s.}} \right)^2 \frac{P^U}{P_S} (\nu_0 + f). \tag{14}
\]

Here the average of \(\sin^2 \theta = 1/2\) has been used, since there is no \textit{a priori} phase defined at the modulation frequency \(f\). This can be consistent with small \(\Delta \phi_{r.m.s.}\) only when \(E_N \ll E_{S,r.m.s.}\). See further caveats below.

Here we have estimated the phase noise associated with noise-based electric-field components rotating at frequency \(\omega_u = 2\pi f\) faster than the optical-field carrier frequency. In the «sideband» language introduced above, we would call this the «upper noise sideband»—this accounts for the «U» designation. Of course, an equal contribution to the total squared phase fluctuation of the resultant optical field arises from noise processes of frequency symmetrically below the carrier. Summing these upper and lower sideband contributions to the total phase variance is to be done quadratically.

To progress further, it is useful to remember that we are associating these phase fluctuations with noise processes within the bandwidth \(b\). So we again are led to the quantity \(S_\phi(f)\)

\[
S_\phi(f) = \frac{\Delta \phi^2(f)}{b} = \frac{\Delta \phi^2_U(f) + \Delta \phi^2_L(f)}{b}. \tag{15}
\]

\(S_\phi(f)\) has dimensions of \(\text{rad}^2/\text{Hz}\) and is called the \textit{spectral density of phase fluctuations}. It is the basic quantity of interest in describing the noise performance of high-stability frequency sources. From the phasor picture discussed above,
we see that

\[ S_\phi(f) = \frac{P_N^0 + P_N^1}{bP_{\text{sig}}}, \]

where \( P_{\text{sig}} \) represents the carrier power at \( \nu_0 \), and \( P_N^1 \) represents the noise power per 1 Hz bandwidth in a sideband at the frequency \( f \) above \( \nu_0 \). A similar definition can be made for the lower noise sideband.

These arguments of this subsection suffice to indicate the strong relationship between sideband power and \( \Delta \phi^2(f) \). However, in eq. (14) we have been a bit cavalier in numerically comparing a stochastic field with a sine wave field. So, while this intuitive vector picture gives us an excellent way to understand the phase fluctuation effects, this treatment gives only approximate numerical results (compare the estimate of eq. (16) with eq. (11), which is rigorous within the small-modulation limit).

\section*{38. Measuring phase noise spectral density: \( S_\phi(f) \) [rad^2/Hz].} For simplicity, we here discuss measuring the spectral density of phase fluctuations only for the case \( \Delta \phi^2 < 1 \). The clearest case is to heterodyne the unknown source with a stable reference to produce a difference frequency signal for measurement purposes. In addition to the two d.c. photocurrents, there will be a cross-term between the two sources. The \( \sin(\alpha) \cdot \sin(\beta) = (1/2)[\cos(\alpha - \beta) - \cos(\alpha + \beta)] \) identity shows us that the full phase information of the high-frequency unknown is preserved in the difference frequency term. This phase conservation under heterodyne frequency offset is an absolutely essential point which bears emphasis. For example let us say \( \alpha = \omega_0 t + \phi(t) \). The difference frequency term, \( \cos(\alpha - \beta) \), will contain \( \phi(t) \) explicitly, without any proportional reduction or scaling. But note that the full phase noise \( \dot{\phi}_n(t) \) of our reference oscillator \( \beta = \omega_r t + \dot{\phi}_n(t) \) will be present in the output as well.

This approach is patterned after conventional measurement practice in the r.f. domain, where we can use a balanced mixer to perform the multiplication. By using a slow phase-locking loop to control the reference oscillator \( \beta \) to be in phase quadrature relative to the average phase of the unknown \( \alpha \), we see from the expansion of \( \cos(\alpha - \beta - \pi/2) \) that we can obtain a direct display of \( \phi(t) \). To adapt this practice to study the phase noise of optical fields, we can apply this heterodyne concept twice, transforming the unknown optical signal down to the r.f. domain using another optical source (of suitably low phase noise), followed by the r.f. phase comparison with a balanced mixer driven by a phase-lockable crystal oscillator. Lacking the perfect optical reference source, a reasonable procedure is to heterodyne together two equivalent optical systems. One then estimates the conversion to the single-source-with-phase-noise basis by taking 1/2 the measured \( S_\phi(f) \).

Another common approach is to use a delayed version of the laser itself for the reference function, as a way to avoid the necessity for an auxiliary source of
superior phase stability. In this self-referenced heterodyne method, the delayed sample and the direct beam are recombined and detected for analysis at radiofrequencies. The recombined beams will show the appropriate phase quadrature condition at a number of r.f. measurement frequencies, depending on the length of the fiber-optic delay line. By having a long fiber, one gets a sufficient number of samples of the phase noise within the envelope of its spectral width. Evidently this measurement system becomes less sensitive at low frequencies, well below the fiber's inverse delay.

A third technique for obtaining the required phase-quadrature reference so that we can measure \( S_c(f) \) can be based on the Pound-Drever optical-frequency discriminator [3]. When the laser is locked to this high-finesse optical cavity, a strong ringing field builds up inside. Its leakage back toward the source can be used as a kind of self-tracking phase reference for the desired phase noise measurement. Modulation sideband techniques provide frequency-offset «test» fields which lie outside the resonator passband. They simply reflect from the input mirror, and carry information about the present phase of the laser. Taken with the ringing leakage field, they allow one to detect an optical beat in the cavity-reflected signal. Then in the subsequent r.f. mixing process, we provide the appropriate quadrature phase reference drive to the r.f. mixer in order to extract the phase variations of the optical signal. Since the cavity ringing time is limited to perhaps 1 ms with even contemporary mirrors, it can be easily seen that this measurement system will not easily measure phase noise at low frequencies. The system has lost its phase memory and has reverted to a frequency measurement algorithm. This brings us to the next topic.

3'9. Frequency noise spectral density. – For some purposes it is convenient to have a measure of the spectral density of noise, expressed as frequency modulation noise processes. Analogous to eq. (13), using eq. (7) we may introduce the spectral density of frequency noise \( S_c(f) \) by the relation

\[
S_c(f) = \frac{\Delta v_{r.m.s.}(f)}{b} \text{ [Hz}^2/\text{Hz]}. \tag{17}
\]

The allocation of frequency noise to lower and upper sidebands proceeds as before. This representation is useful in considering a servo-controlled frequency source being driven in part by noise associated with the error signal. After processing by the servo loop filter, the detected noise reaches the laser which operates on this signal as a voltage-to-frequency converter, producing stochastically varying frequency excursions. The discriminator senses these changes, adds a little noise, and sends them around the loop again.... The ultimate performance limits of such a servo system have been discussed [4]. Basically the signal/noise achieved at the discriminator controls the best possible performance. Sufficient
Fig. 2A. – Power spectral distribution of laser source after an optical-transmission filter cavity. In this optical example, a coherent source ($\omega = c$) is passing through a slightly mistuned resonant cavity ($d$) which is, however, effective in reducing the amount of broadband spontaneous-emission noise present in transmission. Eventually, at frequencies far from the cavity resonance ($a, f$), we can see the broad-band spontaneous-emission background as a nonresonant leakage through the cavity.

Fig. 2B. – Spectral distribution of residual FM noise. Here, the optically filtered laser source of fig. 2A has been applied to an optical-frequency discriminator. So now we are looking at the frequency distribution of the FM on the laser field. Usually there will be prominent FM at the power mains frequency and its harmonics ($b, c, d$). The unsymmetrical optical noise sidebands (represented as the detuned-filtered spontaneous emission in fig. 2A) lead to FM noise in frequency region $e$. At sufficiently high frequencies ($f$) the additive noise of the photodetection process may mask the "real" signal from the optical-frequency discriminator. At the lowest frequencies ($f < a$) it is difficult to avoid some level of contributions due to "d.c. drift" and "$1/f$" noise from the amplifiers.
servo gain must be provided to reduce the laser’s intrinsic noise to a negligible level, after feedback. One is then mapping the discriminator noise into frequency noise.

In both frequency and phase modulation cases, for small enough modulation a sideband occurs only at the perturbing frequency \(\omega_m\). For heavy phase modulation, a number of harmonic sidebands may appear as well. This possibility of a nonunique mapping of frequency noise spectral density between perturbation and modulation result needs to be remembered when we try to diagnose a system.

Figure 2 gives an example in which a laser source is being sent through a band-pass filtering resonator prior to measurement. Although the resonator is slightly mis-tuned, as evidenced by offset of the center of the background fluorescence, the optical carrier is visible clearly in transmission in fig. 2A. Because of the cavity mis-tuning, however, the optical noise sideband structure is not symmetrical around the laser frequency. When we look at the associated spectral distribution (in fig. 2B), we will be trying to represent the full intensity distribution of fig. 2A as an FM spectrum written onto the coherent source. Thus the laser’s spontaneous emission, which began as an additive white process, has been converted into an apparent FM because of the filter mis-tuning (region e). In practice it is hard to avoid modulation of the laser frequency at the power line frequency and its harmonics (b, c, d). At low frequencies the optical-frequency discriminator’s resonant frequency may not precisely match the input frequency, changing, for example, due to temperature fluctuations and perhaps vibrations. By waiting longer (lower Fourier frequencies, \(f < a\)) it is more likely that a big excursion will occur, leading to a stronger negative spectral slope at the very low frequencies. These spurious processes operate above and beyond the «real FM» at low frequencies which would also be displayed, at least in principle. At high frequencies the discriminator slope may be reduced because of bandwidth limitations, and the laser FM noise is probably small also. But photodetection noise is inevitable, and the following amplifier has a finite noise level as well, so the observable discriminator output has a freshly added broad-band noise floor contributed by the measurement system (region f).

3'10. Measuring \(S_n(f)\) [Hz²/Hz]. – In subsect. 3'8, we indicated that, at low frequencies, one may measure the spectral density of frequency variations using a Pound-Drever discriminator below the cavity ringing-time frequency. An earlier and simpler idea is to build an optical-frequency discriminator using an optical resonant cavity in transmission. One can stabilize on the cavity fringe side, for example, to recover the frequency variation information. Unfortunately this signal will be contaminated by omnipresent laser amplitude noise. The photodetector is measuring the desired highly sloped cavity transmission function, but multiplied by the laser intensity. Noise in the latter can be substantially suppressed by subtracting a photocurrent representing an appropriate
fraction of the laser intensity [5]. A clever circuit for automatically establishing the precisely correct scaling factor for this subtraction was recently published by Hobbs [6]. Of course, one really should prefer to divide the measured transmission fringe signal by the separately determined laser intensity. In implementation, however, one finds some problems with differential time delays, denominator feedthrough of the noise itself setting a suppression limit, and various other electronic limitations. It is difficult to obtain deep noise suppression with a ratio system above \( \sim 10 \) MHz. The best bet is to do a fast subtraction first, and then stabilize the scale factor by dividing by the intensity in a lower bandwidth loop. (Actually one multiplies in a fast multiplier by the low-pass-filtered inverse of the intensity.)

As noted earlier, another approach to measuring the spectral density of frequency noise is based on heterodyning the unknown down into the r.f. domain for FM analysis. Evidently this approach also requires a second reference laser for the heterodyne reference. A useful instrument is a low-noise converter of radiofrequency to voltage. Its output will contain a d.c. voltage representing the average frequency, and voltage changes which mirror frequency changes in the source. Measurement of the noise spectral density of the output voltage with an audio spectrum analyzer—perhaps an FFT-based instrument—allows extraction of the desired laser frequency noise information, after calibration of the scale factor. One frequency/voltage converter design, based on a digitally operated «charge pump», could operate in the range 1 MHz \( < f < 36 \) MHz and displayed a low-noise floor corresponding to \( < 150 \) Hz r.m.s. frequency noise equivalent in a 10 kHz bandwidth [7]. One generally finds that frequency measurement is useful for relatively high Fourier frequencies, but that phase measurement methods excel below a few hundred hertz.

3’1. Relationship of \( S_\phi(f) \) and \( S_\nu(f) \). — To make connection between these two closely related measures, we may make use of the basic relationship between phase and frequency. We denote the frequency by \( \Delta \nu(t) \) to indicate that we are considering frequency changes relative to a constant value. Correspondingly, \( \Delta \phi(t) \) refers to changes beyond the constant linear accumulation of phase with time, that is to phase noise:

\[
\Delta \nu(t) = (2\pi)^{-1} (d/dt) \Delta \phi(t).
\]

We may write this same relationship in Fourier space, where \( d/dt \to 2\pi f \):

\[
\Delta \nu(f) = (2\pi)^{-1} 2\pi f \Delta \phi(f).
\]

This leads us to see that

\[
\Delta \nu_{r.m.s.}^2 = f^2 \Delta \phi_{r.m.s.}^2.
\]
Fig. 3. – Calculating residual FM due to phase noise. Fourier components between \( f = a \) and \( f = b \) contribute to laser spectral width weighting of \( f^2 \), i.e.

\[
\Delta \nu = \sqrt{\int_a^b f^2 S_\nu(f) \, df}.
\]

Recalling our definition that \( S_\nu(f) = \Delta \nu_{\text{r.m.s.}}(f)/b \), we obtain the desired relationship

\[
S_\nu(f) = f^3 S_\nu(f) \, [\text{Hz}^2 / \text{Hz}].
\]

As an application example of this useful relationship, consider that we know the phase noise density \( S_\nu(f) \) between two frequencies, \( a \) and \( b \) (as indicated in fig. 3). The question is: what is the FM noise associated with these excursions? Applying eq. (21) and the definition of spectral density, we have

\[
\Delta \nu = \text{residual FM due to phase noise between frequencies } a, b = \sqrt{\int_a^b f^2 S_\nu(f) \, df}.
\]

3'12. Summary of model of laser output. AM vs. FM revisited. – While an ideal laser produces an unchanging sine wave, a real laser (at least a reasonably stable one!) may be usefully modelled as having a slowly varying amplitude and phase. We have shown how to discuss the instantaneous frequency of the laser: an analogous treatment of instantaneous intensity could be easily made. In the case of time-varying intensity, we would produce amplitude modulation (AM) sidebands, in addition to those sidebands due to the phase/frequency modulation processes (FM). In general the optical-power spectral density measured at any frequency will have contributions from both sources, but, as we have indi-
inated, for stable laser systems the amplitude sidebands tend to be much less prominent.

Another interesting distinction here is that the intensity modulation sidebands have even symmetry around the optical carrier. In calculating the a.c. photocurrent produced by the detected optical power, $|e \cdot e^*|$, we will have cross-terms between the carrier and the sidebands. In the AM case, the even symmetry of the sidebands will lead to a resultant variation in the photocurrent, revealing at frequency $f$ the fluctuating noise power that began as sidebands above and below the carrier by the interval $f$. By contrast, the antisymmetry of the odd FM spectral sidebands leads to cancellation of the cross-term of carrier × lower sideband by the current due to the carrier × upper sideband.

313. FM lineshapes for $\Delta\nu_{r.m.s.} \gg b$ or $\Delta\nu_{r.m.s.} \ll b$. - Interesting analytic results can be obtained for the laser lineshape under rather general assumptions. For brevity, we restrict ourselves to negligibly small AM, and focus on FM noise processes. Middleton[8] considers that the continuous phase/frequency noise of an oscillator can be discussed as a Gaussian, ergodic process. The question for us is: what is the spectral lineshape of a laser, given that we know the r.m.s. frequency modulation excursion and the bandwidth of the noise? Stewart[9] considered both frequency modulation and phase modulation of an oscillator by a stochastic voltage. For simplicity of discussion, he assumed the frequency noise density to be $S_{\nu}(f) = \Delta\nu_{r.m.s.}^2 / b$, flat from $f = 0$ up to an upper sharp limit of $b$ hertz. This brings an echo of our earlier identification of the sine wave phase modulation index $\beta$ as the ratio frequency deviation/frequency rate of the deviation. For sinusoidal modulation, small $\beta$ led to a 1:1 map, while large $\beta$ leads to many orders of sidebands. The case of frequency modulation by a stochastic wave is richer. As shown by Stewart[9], the wide-bandwidth, small-excursion case leads to a Lorentzian profile. The lineshapes, changing between the two limiting regimes of interest, were later studied in more detail by Elliott, Roy and Smith[10].

313.1. $b \ll \Delta\nu_{r.m.s.}$. This is the limit of slow, rather large excursions, and leads to a large modulation index and a Gaussian lineshape. With these assumptions, the power spectral density (optical lineshape) is

$$P_{\nu}(\omega) = \frac{E_0^2}{\Delta\nu_{r.m.s.}} \exp \left[ -\frac{(\omega - \omega_0)^2}{8\pi^2 \Delta\nu_{r.m.s.}^2} \right]$$

(23)

and the FWHM linewidth is

$$\Delta\nu_{\text{FWHM}} = 2.355 \sqrt{\Delta\nu_{r.m.s.}^2}.$$
The noise bandwidth does not appear explicitly in this result—it is the
frequency excursion (rather than the rate) which is important.

3'132. \( b \gg \Delta \nu_{r.m.s.} \). This is the limit of fast, small excursions, leading in the
line wings to weak sidebands of only first order. Under these conditions, the
optical-power spectral density takes the form of a Lorentzian profile:

\[
P_E(\omega) = \frac{E_0^2}{\pi} \frac{\frac{\pi^2 \Delta \nu_{r.m.s.}^2}{b}}{(\omega - \omega_0)^2 + \left( \frac{\pi^2 \Delta \nu_{r.m.s.}}{b} \right)^2}
\]

and the linewidth is

\[
\Delta \nu_{\text{FWHM}} = \frac{\pi \Delta \nu_{r.m.s.}}{b} = \pi S_r(f).
\]

(As an aside, the reader should note the use here of the power spectral-den-
sity function \( P_E(\omega) \) to represent the optical power within a 1 radian/s band-
width at optical frequency \( \omega_0 \). One distinction is that there we have elected to
use radians/s rather than hertz. However, the important distinction is that this
spectral-density function would be called «single-sided» to contrast it with our
power spectral-density function \( S_r(f) \) discussed earlier in which the frequency
axis for \( f \) begins at zero, i.e. \( f = |\nu - \omega_0| \). The latter type of spectral density
may be called «baseband» to indicate it is a spectrum of the modulation pro-
cesses. It would also be called «double-sided» to indicate that it contains contribu-
tions from processes both above and below the carrier.)

Elliott et al. [10] give a useful discussion of the lineshape behavior be-
tween these limiting cases, and of experimental tricks to prepare fields with
known statistical properties. But in the general laboratory case, the frequency
noise is given by some complicated spectrum, and usually shows regions in
which different physical processes dominate the noise. So it is useful to develop
a less restrictive framework for the discussion. One approach is based on know-
ing the autocorrelation function of the optical electric field, and makes use of a
remarkable theorem relating this function to the spectrum.

3'14. Calculating the laser spectrum from the field autocorrelation func-
tion [11]. – In a substantial part of our discussion up to now, attention has been
focussed on the optical phase as representing the information most significant
to us as we learn how to characterize our laboratory laser sources. The full time
dependence of the phase, our function \( \hat{\phi}(t) \), represents the total information
available, so all properties can be derived from it. However, the insight which
this knowledge produces may be limited by our ability to analyze and digest it.
 Somehow, in a big picture, every detail is not so interesting: We fundamentally
want to catch the main effects. Features which are prominent in the data be-
cause of their strength or their longevity attract our attention. One of the most important things we would like to know is whether the carrier is still present when the laser has been modulated in some specified way [12].

A very useful way to reduce the information content of our record is to consider the autocorrelation function of the electric field. Experimentally, this will be done most conveniently by heterodyning the field with a stable reference laser, thus producing a quasi-sinusoidal voltage wave which can be captured in the time domain. Forming the autocorrelation function from these heterodyne data will reduce the output information from that of \( \xi(t) \) in three ways. First, we will certainly include phase noise from the reference laser. Second, the autocorrelation function does not catch the two-quadrant information needed to fully reconstruct the wave form. Finally, in the end we will probably agree to make some approximations (that the correlation has basically gone to zero after some time interval), so that the actual number of numbers we retain may be drastically reduced.

The electric-field autocorrelation function \( R_E(\tau) \) may be defined as

\[
R_E(\tau) = \langle E(t) E^*(t + \tau) \rangle,
\]

Here the asterisk on \( E \) is meant to represent the complex conjugate operation, and the average is on time, taken over the particular data example of interest. As the delay variable \( \tau \) is scanned, we can expect to pick up a significant autocorrelation only to the extent that the field is similar to itself after a delay. Note that the basic fast time dependence is suppressed in the form of eq. (27), so we will be effectively studying the phase noise function \( \phi_N(t) \). (Again we have here neglected amplitude noise for simplicity [11].) The remarkable Wiener-Khintchine theorem [8] gives the power spectrum of the optical electric field, \( P_E(\omega) \), as the Fourier transform of the electric-field autocorrelation function \( R_E(\tau) \), i.e.

\[
P_E(\omega) = \frac{1}{2\pi} \int \exp[-i\omega \tau] R_E(\tau) \, d\tau,
\]

\[
P_E(\omega) = \frac{1}{2\pi} \int \exp[-i\omega \tau] \langle E(t) E^*(t + \tau) \rangle \, d\tau.
\]

Using eq. (4), the autocorrelation function in eq. (28) takes the form

\[
R_E(\tau) = E_0^2 \exp[i\omega_0 \tau] \langle \exp[i\phi_N(t + \tau) - i\phi_N(t)] \rangle.
\]

The Gaussian moment theorem allows us to evaluate eq. (30) to obtain

\[
\langle \exp[i\phi_N(t + \tau) - i\phi_N(\tau)] \rangle = \exp[R_x(\tau) - R_x(0)],
\]

where

\[
R_x(\tau) = \langle \phi_N(t) \phi_N(t + \tau) \rangle
\]
is the autocorrelation function of the noise source. (Some parts of this calculation are worked out in detail by ELLIOTT et al. [10].)

Since we have already in mind a model representing the phase behavior of $E(t)$, it is possible to make this discussion much more concrete. In eq. (4) we supposed a single modulation frequency to be the agent of phase modulation. Then we generalized this idea to describe a distribution of modulation frequencies which, in the limit, was be taken to be a continuous Gaussian distribution. (There are some interesting subtleties here involving stochastic modulation [8, 11, 13] and concerning time-dependent spectral measurements [14].) The essential result is that phase noise modulation can be represented in eq. (4) by a time-dependent stochastic phase $\phi_N(t)$ with its magnitude summarized by an r.m.s. phase noise variance, $\Delta_{\phi_{r.m.s.}}^2 = \langle \phi_N^2(t) \rangle$. In terms of the noise source autocorrelation function, $R_\phi(t)$, we have

$$R_\phi(0) = \Delta_{\phi_{r.m.s.}}^2,$$

and

$$R_\phi(t) \to 0 \quad \text{as} \quad t \to \infty.$$

Thus in a very natural way we can obtain information about the power remain-

<table>
<thead>
<tr>
<th>#</th>
<th>$\Delta_{\phi_{r.m.s.}}$ (rad)</th>
<th>$\Delta_{\phi_{r.m.s.}}^2$ (rad$^2$)</th>
<th>Carrier remaining (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1.8</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>37</td>
</tr>
<tr>
<td>3</td>
<td>0.83</td>
<td>0.69</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>0.5</td>
<td>0.25</td>
<td>78</td>
</tr>
<tr>
<td>5</td>
<td>0.32</td>
<td>0.11</td>
<td>90</td>
</tr>
<tr>
<td>6</td>
<td>0.23</td>
<td>0.05</td>
<td>95</td>
</tr>
<tr>
<td>7</td>
<td>0.18</td>
<td>0.03</td>
<td>97</td>
</tr>
<tr>
<td>8</td>
<td>0.10</td>
<td>0.01</td>
<td>99</td>
</tr>
<tr>
<td>9</td>
<td>0.03</td>
<td>0.001</td>
<td>99.9</td>
</tr>
<tr>
<td>10</td>
<td>0.01</td>
<td>0.0001</td>
<td>99.99</td>
</tr>
</tbody>
</table>
ing in the laser carrier after all the phase noise modulation sidebands have deducted their power[12]. The result is

\begin{equation}
R_E(\tau) = E_0^2 \exp[i\omega_0 \tau] \exp[-\Delta \varphi^2_{r.m.s.}] + E_0^2 \exp[i\omega_0 \tau] \exp[R_s(\tau)](1 - \exp[-\Delta \varphi^2_{r.m.s.}]) \tag{carrier}
\end{equation}

+ E_0^2 \exp[i\omega_0 \tau] \exp[R_s(\tau)](1 - \exp[-\Delta \varphi^2_{r.m.s.}]) \tag{sidebands}

This equation shows explicitly the importance of the phase variance, \(\Delta \varphi^2\), in controlling the distribution of optical power between the carrier and the sidebands. Specifically, one can see that the fractional power remaining in the carrier is \(\exp[-\Delta \varphi^2_{r.m.s.}]\).

In the laboratory, one will have a servo working to reduce the unplanned phase excursions, basically by the closed-loop gain factor. The question often comes up: How much gain do we really need to phase-lock a laser? How much more carrier power do we earn by improving the servo performance by some factor? Table I helps answer these kinds of questions, qualitatively.

In practice, one can remember that, if the phase variance is \(1 \text{ rad}^2\), the carrier has about 1/3 of the total power. If the phase noise is reduced to 1/3 radian, then most of the power (90%) is in the optical carrier. If one happens to be in the region of phase noise around 1 rad, it is clear that a little extra electronics work pays off handsomely!

3.15. Laser linewidth in terms of spectral-density measurements. From the above indication of the carrier power remaining under a certain phase modulation condition, it is possible to write an approximate relationship which will give the laser linewidth in terms of measured spectral-density information. We saw that the total power would be about equally distributed between carrier and sidebands when the phase variance is \(\approx 1 \text{ rad}^2\). Using the phase spectral density we can write an implicit equation for the 3 dB linewidth \(f_{3dB}\) (HWHM basis):

\begin{equation}
1 \text{ rad}^2 = \int_{f_{min}}^{f_{max}} S_{\varphi}(f) \, df. \tag{36}
\end{equation}

The equivalent approximate relation, based on the spectral density of frequency noise, would be

\begin{equation}
1 \text{ rad}^2 = \int_{f_{min}}^{f_{max}} \frac{S_{\omega}(f)}{f^2} \, df. \tag{37}
\end{equation}

While lineshape effects are rather dependent upon the details of the noise modulation, the rapid variation of the carrier power with phase variance \(\Delta \varphi^2_{r.m.s.}\) leads one to believe that these estimates should be fairly robust.
4. Representing frequency noise in the time domain: Allan’s variance.

4.1. Motivation and overview. It is a matter of common experience that measurements made over a restricted time interval often cluster more tightly than when one considers the full set of measurements made over an extended period of time. Unconsciously we seem to choose conditions in the laboratory so as to avoid the fast variations which become apparent when we signal-average over too short an interval. Many repeated measurements made within the hour give basically the same answer. But come back tomorrow, or even after a coffee, and the values appear to be changed. Often to values «impossibly» different, if we try to take the previous small scatter as indicative of the Gaussian uncertainty. What is going on?

Statistical experts are likely to «explain» this phenomenon by discussing the «1/f problem», or by invoking even more aggressively divergent low-frequency noise processes which appear to prevent us from making our study satisfactorily. Such elaborate modelling discussions may have utility, but tend to cloud the basic physical reality: the longer we wait, the more likely it becomes that a giant number of unlikely fluctuations will have all been overlapping somewhat in time. Measuring longer does not help, because the system is unlikely to have all of these difficult barriers re-crossed for our convenience. So averaging becomes less help: Depending on the nature of the noise, the most recent measurement may be the best estimator of the quantity or process of interest!

In the frequency standards community these problems are often encountered. Probably the most effective and useful suggestion was made by Allan[15], who noticed that simply comparing adjacent measurements to each other would introduce a profound attenuation of any long-term noise process since the latter does not lead to rapid sample-sample changes. In this way one is able to separate and isolate processes based on their time scales. For example, random noise should decrease (on a relative basis) with increasing measurement time as \( \tau^{-1/2} \). So Allan’s first-difference technique allows us to focus on the random-noise part of the problem, and to identify the presence of strong low-frequency divergent processes[2, 15].

Imagine we are counting or otherwise measuring a frequency \( \nu(t) = \omega(t)/2\pi \). We count the input pulses arriving during a time interval \( \tau \), record the result as \( \nu_i(t_i) \), and repeat the measurement to form the data set \( \{ \nu_i(t_i) \} \). Rather than calculating means and standard deviations of this data set, Allan’s suggestion is to deal with the frequency differences \( \{ d_i \} \equiv \{ |f_{i+1} - f_i| \} \). Because the truly random noise will be uncorrelated in the two measurements, we can later correct the scale of \( d_i \) by \( 1/\sqrt{2} \). Another important concept is that we need not seek a total description capable of recreating the data from our representation. Rather, we seek to identify and characterize the underlying physical processes in the observed noise.
The usual quantity of interest will be the normalized frequency variations, \( \bar{y}_i = v_i/v_0 \). (The average is indicated on \( y \) because the frequency is, in fact, averaged during the measurement time \( \tau \). See ref. [2, 15].) In these terms, the definition of Allan's variance is

\[
\sigma_y^2(\tau) = \frac{1}{2(N-1)} \sum_{i=1}^{N-1} |\bar{y}_{i+1} - \bar{y}_i|^2.
\]

Basically, this is a two-point operator parameterized by the counting time \( \tau \). Experience confirms that a very useful measure of our frequency noise can be obtained by plotting \( \sigma_y^2(\tau) \) vs. the common frequency measurement time interval \( \tau \). (More exactly, one usually makes a double logarithmic plot of the square root of Allan's variance. Another small grumble about common usage is the usual tendency to speak of a frequency stability of \( 10^{-13} \), or whatever. Logic would suggest that this measure is of frequency in-stability.) To acquire data for the plot, rather than repeating the measurements with increasing \( \tau \), a long measurement series can be recorded using the shortest appropriate counting gate time. If the dead time is small, the result which would have been obtained over a longer counting interval can be just calculated by summing an appropriate number of adjacent samples before calculating Allan's variance. Other gate times can be supposed, and so the full plot can be calculated from a small number of data sets. This real-time efficiency [2] also helps us avoid the long-time divergent effects. An integral transform is known [2, 15] which gives \( \sigma_y^2(\tau) \) in terms of the phase spectral density \( S_p(f) \). Reference [2] gives a useful discussion and examples of this relationship between frequency domain and time domain measures of frequency stability.

4.2. Behavior of Allan's variance: synthetic data. – Before presenting an Allan variance plot based on actual physical measurements of a stable laser system, it is instructive to consider fig. 4 which shows the results obtained from eq. (38) when the input is synthetic data. We chose the (r.f. beat) frequency of 500 kHz, adding a random frequency noise of Gaussian standard deviation of 5 kHz. Several useful pieces of information can be read off the Allan variance plot rather directly.

4.2.1. Looking at the short-time domain, we can see a \( -1/2 \) slope on Allan's logarithmic plot. The level of this frequency noise is fixed by the voltage noise in the frequency discriminator output, in comparison with the slope of the frequency discriminator function. The servo system has no way to distinguish noise outputs caused by real frequency variations from noise outputs of a shot noise or a purely electronic origin: its behavior is always to reduce such fluctuations by the appropriate gain factor [4]. So shot noise is mapped by the servo system into an equivalent frequency excursion. Such shot noise processes have
a white-noise spectrum, that is $S_v(f) = k(\text{Hz}^2/\text{Hz})$, independent of frequency $f$. This will show up as a $\tau^{-1/2}$ line on an Allan variance plot. To understand this $\tau^{-1/2}$ dependence, consider a time interval of $s$ units of the basic measurement interval. In each we will have accumulated $s$-fold more contributions to the average value. In this same interval the input noise to the servo—which will appear as a random walk of the controlled frequency—will be increased by the factor $s^{1/2}$, leading to a $s^{-1/2}$ scaling of the relative frequency variation to be plotted as the (square root of) Allan’s variance.

4.2.2. The long-time drift/divergence problem discussed above leads in the Allan variance plot to a domain of $\tau^0$ behavior, i.e. a variance which is measurement time independent. This is the so-called “flicker of frequency” domain, in which the noise reductions won by longer averaging are just annulled by changes in the mean value. In the experiments these changes are usually associated with changes of systematic offsets. As the technical solutions improve, an early system with flicker level $S_1$ may be improved to level $S_2$. An indication of
what can be achieved is that—after 40 years dedicated work on cesium atomic-beam standards—the laboratory standard Cs clocks provide line splitting in an accuracy sense of $\sim 10^{-6}$ linewidths. Internal-cell optical systems tend to show worse stability ($\sim 10^{-4}$ linewidths) and another 10-fold worse accuracy. See discussion below for improved approaches.

4'2.3. Often it will happen that highly stable sources will still exhibit a frequency drift in time. For the simplest case, we could hope that over our measurement interval the drift would be approximately linear, say at the rate $\delta$ Hz/s. Then it is easy to see that the (square root of) Allan variance will be increased by $\delta^2 \cdot \tau$, i.e. that we have developed a domain of $\tau^{-1}$ at the longest times. If the frequency drift is quadratic or unpredictable, a memory of the $\tau^{-1}$ behavior will appear even after removal of the linear drift term before calculating Allan’s variance. Of course, it will become dominant again only at longer measurement times $\tau$. Reference [2] considers other cases.

4'2.4. Sometimes a stabilized laser to be measured has a significant sinusoidal FM, for example as used for generating the locking signals with phase-sensitive detection. This FM gives rise to a localized rise in the Allan variance for averaging times near $\tau = 1/2f$.

4'3. *Distinguishing between spectrally filtered noise and a drifting monochromatic source.* – In the foregoing discussion, we have concentrated on the aspect that the most interesting frequency sources are phase-coherent. Basically this property means that the phase noise remains below 1 rad and identifies a source which will appear monochromatic. Perhaps its frequency will evolve in time, but it moves slowly enough that we may speak of a slowly drifting monochromatic source. Alternatively, for another source the phase noise may be somewhat larger than 1 rad. Then the noise sidebands ($= J_0^2(\beta)$) will have become well developed and the character of the source is fundamentally changed; it now looks like narrow-band-filtered noise. As soon as the modulation index is increased even a little more, the carrier may no longer be observable. It is useful to re-state this Fourier space distinction in the measurement time domain, using Allan variance language.

Consider fig. 5, in which we show the Allan variance data for several high-performance optical-frequency sources, plotted on double logarithmic coordinate axes. The figure also shows a major diagonal line, running with slope of $-1$, from $\sigma_\tau = 3 \cdot 10^{-12}$ at $10^{-4}$ s to $3 \cdot 10^{-16}$ at 1 s. If the Allan variance for a source lies well below and to the left of this line, we would call this a *phase-coherent, monochromatic* source. In this domain, the maximum phase modulation would be below 1 rad even for a source frequency of 500 THz (which corresponds to the red wavelength of 600 nm). Also plotted on this graph are a number of other diagonal lines of slope $-1/2$, corresponding to the time rate of
Fig. 5. – Distinguishing a spectral line from filtered noise. In the lower-left sector the fast phase noise is below one radian, so that the source looks like a spectral line which is (slowly) drifting in frequency. In the upper-right triangle, the phase noise exceeds one radian, and so no appreciable carrier remains. Thus the field has only a short-time coherence and looks like white noise which has been spectrally filtered, in our case by the resonance filter represented by the frequency servo loop. Note that the present (e)xperimental cavity-locked results correspond to a line spectrum only out to a few seconds, while the (t)heoretical S/N-determined performance would give a phase coherence time of $\sim 300 \text{ s}$ if it were not for environmental perturbation of the reference cavity. In general, the iodine resonances are too broad and/or weak to give a sharp line/good short-term stability.

growth of signal/noise ratio for frequency control systems limited only by white frequency noise. The theoretical performance for our cavity-locked system[16] (marked (t)) is shown for short measurement times to be strongly into the monochromatic, phase-coherent domain. Considering that the signal/noise improves with $\tau^{1/2}$—while the number of cycles elapsed increases as $\tau^1$—it is sure that a frequency-discriminator-locked laser must eventually run out of signal/noise ratio adequate to define the optical phase to within 1 rad. In this experimental example, this cross-over point between a sharp spectral line and filtered noise, that is to say the phase coherence time, is about 200 s. The corresponding width would appear to be $1/\pi \tau = 1.6 \text{ mHz}$! However, as we have discussed in the section about spectral-noise density, there is an additional factor of $\pi$ which comes in when we try to deduce the Lorentzian
linewidth from frequency noise excursions. Even so, 4.4 mHz is not a bad linewidth for an optical-frequency source!

It is useful to explore the relation between linewidth and Allan variance data. In the $\tau^{-1/2}$ domain, a line on the Allan variance plot represents the FM noise density which the laser carries, which we know decreases for longer measurement times as the signal averaging process evolves. We showed before how the fast parts of these frequency excursions are relatively harmless: they reverse their direction before there has been enough time to integrate up substantial phase. From the definition $\sigma_\nu(\tau) = \delta(\tau)/\nu$, and the scaling $\sigma_\nu(\tau) = \sigma_1/\sqrt{\tau}$, we introduce $\sigma_1$, the Allan variance scaled to 1 s. This is a measure of the white-noise spectral density from the frequency discriminator which tries to «correct» the oscillator for false frequency variation information. An exercise for the reader is to show that the Lorentz linewidth of an ideal frequency-servo-controlled laser is $\nu_{FWHM} = 2\pi^2(\sigma_1)^2$. Barnes et al. [15] give formulae for more general cases.

This theoretical, ideal white-noise performance can be confronted by experiment [16, 17]. To minimize extraneous noise factors, we locked two lasers to successive orders of a stable cavity. In fig. 5, for short times the experimental curve (marked «cavity (e)») approaches the theoretical Allan variance line («cavity lock (l)») gracefully enough, although it can be seen that, by even 1 s, the experimental result no longer is improving with integration time. Instead, it has reached a «flicker floor», with a $\tau^0$ dependence of $\sigma_\nu$ at about 1·10^{-15} in this case. (As a small aside, to the author’s knowledge, this represents the best optical stability yet reported [17].) Plots of the measured frequency vs. time reveal the presence of «jumps» at irregular times, where the frequency may change rather abruptly by a few tenths of a hertz or more, up to several hertz. These jumps are probably caused by changes in the optical contacts of the cavity mirrors. A 9000 s record yields perhaps a dozen domains during which the frequency is constant to about one radian equivalent for times of (50 ÷ 200) s, in rough correspondence with the phase coherence time estimated from the signal/noise performance of the device.

More generally, for the longer times in fig. 5 we believe that time-dependent changes of systematic offsets from the line center are the origin of this fresh noise which is revealed when the random-amplitude noise out of the cavity frequency discriminator has been reduced enough by averaging. This flicker, i.e. $\tau^0$ behavior, is characteristic of the medium to long-term averaging-time dependence of most high-performance frequency sources. At longer times we can anticipate more severe degradation as we can no longer suppose the continuity of our operating conditions. Finally at still longer times, we approach the resettability limit for the controlled oscillator. If the locking criterion can be referred to some atomic transition which could in principle be reproduced in another laboratory, we may speak of the reproducibility of a particular design of apparatus in providing the same locked frequency, in contrast to the resettability of a sin-
gle device. Standards laboratories also find it useful to distinguish accuracy capability, which refers to a synthetic frequency which would be obtained when measured shifts and their uncertainties are taken into account. One can soon see the advantage of good short-term performance which allows completion of a protocol of precise measurements of all known offsets within a time short enough that the offsets have not yet changed appreciably.

44. Performance of iodine-stabilized lasers. – In fig. 5, in the upper right domain where nonmonochromatic sources are located, we find some experimental and theoretical results for HeNe lasers stabilized to molecular iodine[18]. The curve labelled 〈633 (e)〉 shows the performance of a standards laboratory quality of HeNe stabilized laser operating at 633 nm, using 127I2 for the reference[17]. With serious effort, the flicker corner can be pushed to beyond 5000 s. Problems for this system include a rather small thermal population in the V' = 5 initial level, predissociation which leads to linewidth increase and contrast decrease, and asymmetric distortion of the absorption profile by gas lens effects, and perhaps also by effects of an uncontrolled geometric phase. The dashed lines R1 and R2 show the range of reproducibilities encountered in international intercomparisons[18].

The usual stabilized-laser instruments are designed with a rather low modulation frequency, e.g. 5 kHz, which lies in the range where laboratory vibrations still have an appreciable amplitude. The resulting laser FM has the disastrous consequence of converting the laser's intrinsic frequency noise into noise in the signal. If a higher modulation frequency were used, even for an intracavity device, one should be able to obtain a 10-fold performance improvement. Furthermore, a better transition in I2 can be accessed with the HeNe laser running at 612 nm.

A powerful method[19] called modulation transfer can be utilized with an external cell of iodine to yield improved results, in both short-term S/N and long-term accuracy. In fig. 5 we show the curve labelled 〈mod tranf, e(xperimental) S/N〉. This system employed a cavity-stabilized HeNe orange laser with the modulation transfer signal being obtained from a 1 m iodine cell located external to the laser[20]. A factor of about 1000 improvement was obtained relative to a standard metrological red HeNe/I2-stabilized laser. A factor about 10 reduction in the molecular linewidth was welcome, provided in part by reduced predissociation on the orange transition, and in part by lower operating pressure and light intensity. The change in measurement topology has advantages with respect to S/N performance, but the most significant advantages will accrue when one attempts to lock to the ever-elusive center of the absorption line.

The flicker level projected in fig. 5 for the 612 nm system is estimated on the basis of a stability of 10^-5 linewidths. We noted earlier that the laboratory standard Cs clocks provide line splitting in an accuracy sense of ~ 10^-5 linewidths,
so this stability level should be a reasonable expectation. (Remember this system uses an external location for its absorption cell.) In pursuit of ultra-stable lasers, sharper resonances are obviously helpful, so we close this discussion with a few dreams about the coming new epoch of spectroscopy with cold atoms.

5. The future of stabilized lasers: cold atoms and diode lasers?

From the above discussion it is clear that there are adequate challenges and rewards for further work on stabilized lasers. After the S/N ratio is a large number, say well above 10^4 in the measurement time considered, it is clear that progress should be sought also by reducing the resonance linewidth. In early methane studies, this was done by expanding the beam in order to reduce the transit broadening[21]. With iodine, some reduction is possible by going into the blue-green region, nearer to the dissociation limit. One good source for this work is provided by the doubled output of a Nd:YAG laser at 1.06 \mu m. The laser-diode-pumped monolithic version of this system (the so-called «NPRO») is particularly nice with its good intrinsic frequency stability and \approx (10 \div 100) \text{ mW} output power. With 100 \text{ mW} of i.r. available, one can expect to obtain \approx 30 \text{ mW} of green output by doubling using LiNbO_3, KTP, or KNbO_3 crystals in an external ring cavity. The iodine linewidth near 534 nm is \approx 50 \text{ kHz}, and several lines with high absorption can be reached with the doubled Nd output[22].

But to really make serious qualitative advances, it is clear that drastic linewidth reductions are necessary. Consider, for example, Ca which has a red intercombination line at 657 nm with an intrinsic linewidth near 400 Hz! However, to realize such a linewidth in a measurement of a thermal beam, one needs an interaction length of \approx 1/3 m. This would be possible using Ramsey’s separated-oscillating-fields method. But a thermal beam has such a large velocity that the quadratic Doppler effect (considered equivalently as the relativistic time dilation) causes profound lineshape distortion. These effects have been studied before in Ramsey fringe spectroscopy based on saturated absorption[23] in Ca and in two-photon absorption[24] in Bi. Now, more than a decade later we finally have the atom manipulation tools to offer a gentle interaction of long duration with atoms of small and well-controlled velocity. One can be particularly optimistic about the prospects for the Ca system which can be saturated by red diode laser output as shown by HOLLBERG[25], and cooled and trapped as shown by SHIMIZU[26]. An interesting beneficial accident is that the cooling light at 423 nm can be efficiently generated by frequency-doubling a powerful Ti:sapphire laser. Indeed, POLZIK et al.[27] have shown that > 650 mW can be obtained rather directly in this way! Later one will use a laser diode at 846 nm as the i.r. source. Finally one can foresee a possible stabilized laser

...
optics, in Quantum Optics IV, edited by J. D. Harvey and D. F. Walls (Springer-Verlag, Heidelberg, 1986), p. 273. Also see ref.[16].


[19] Typically several such measurements appear each year in this journal.


system of sub-hertz accuracy potential, based on Ca atoms interrogated and cooled by battery-powered diode lasers. Fabrication of such a system as transportable (within the size limits of airline cabin baggage) leads American scientists to look forward eagerly to future trans-Atlantic trips to intercompare such frequency standards with corresponding systems developed by friends and colleagues in European laboratories!

***

The author would like to express his appreciation to L. Hall for her cheerful friendship and patience, and her long-standing tolerance of unreasonable work habits. He also is grateful for the kind and generous hospitality provided by the scientific and administrative organizers of the Fermi School. He has profited enormously from collaboration in this work with a number of former students, including M. Zhu, M. P. Winters, L. Hollberg, J. C. Bergquist, and with colleagues and visitors, including D. Hils, L-S. Ma, J. Hough, H. G. Robinson, J. M. Chartier and H. J. Kimble. J.L.H. would like to thank Dr. M. Zhu in particular for his contributions to both theory and experiment in phase-locking lasers. The ultra-stable-cavity results exist only due to the patient and dedicated work of D. Hils. Special thanks are offered to colleagues in the frequency standards group at NIST, particularly F. L. Walls and D. W. Allan, for generously sharing their understanding and research results over the years. Thanks are also due to the California Institute of Technology and its Sherman Fairchild Distinguished Scholars Program under which auspices much of this manuscript was prepared. The work at JILA is supported in part by the Air Force Office of Scientific Research, the Office of Naval Research, the National Science Foundation and in part by the National Institute for Standards and Technology under its program of basic research on topics of possible application to fundamental standards and measurements.

REFERENCES


[4] An earlier article makes explicit the role of detection noise in controlling the laser linewidth. See J. L. Hall: Stabilizing lasers, for applications in quantum