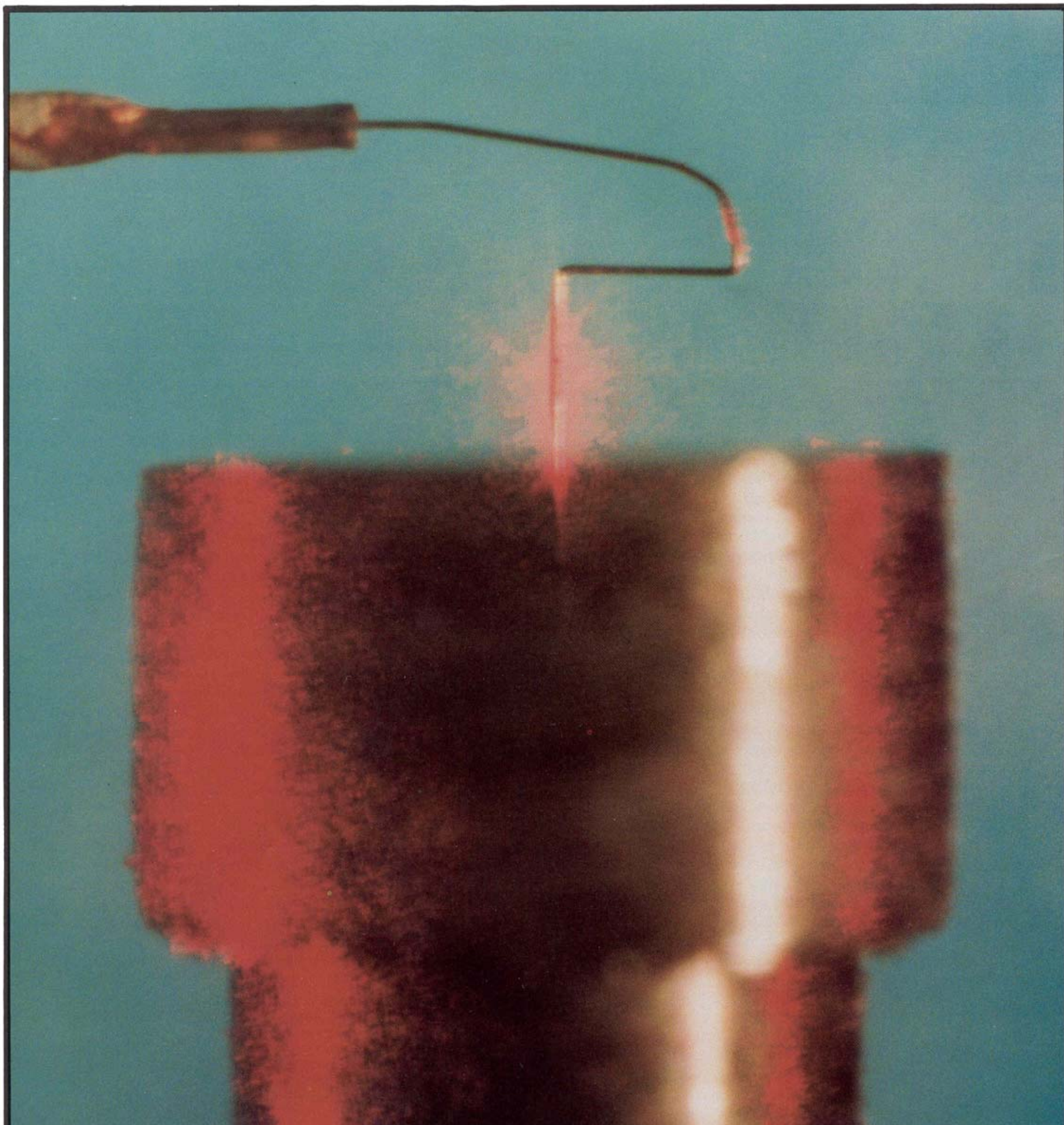
physics today

DECEMBER 1978

Frequency standards



# Time, frequency and physical measurement

**Perhaps all standards—including those of length, mass and temperature—can be redefined by linking them to the most accurate standard, that of time, which is known to an accuracy of $10^{-13}$.**

Helmut Hellwig, Kenneth M. Evenson and David J. Wineland

How many basic standards do we need? Standards are necessary to measurement, and for reasons of accuracy and convenience many measurements involve frequency. With atomic and molecular transitions serving as references, measurement precisions near $10^{-16}$ are possible. The duration of the second is determined by a resonance in the cesium atom, and international atomic time is the reference for all time and frequency measurements in the world. Furthermore, frequency measurements lead, *via* the speed of light, to the measurement of wavelengths and, *via* transducers, to the measurement of many other physical quantities such as temperature and pressure. As a result, time and frequency metrology is at the root of any thinking to revise or improve our system of basic standards of measurement.

Thus the question of how many basic standards are necessary arises, with the accompanying thought that it might be possible—and beneficial—to derive all measurements from just one basic standard, that of time and frequency.

These questions arise at a time when the commercial availability and demonstrated reliability of atomic clocks and frequency standards make possible the establishment of large national and international time-ordered navigation and communication networks while new ideas in the physics of sub-Doppler spectroscopy stimulate major, fundamental advances in measurements of extreme resolution. This article attempts to focus on several of these developments which, we feel, are likely to make a substantial impact on many areas of science, technology and measurement of physical quantities.

The authors are staff members of the National Bureau of Standards in Boulder, Colorado; Kenneth M. Evenson is also professor adjoint at the University of Colorado.

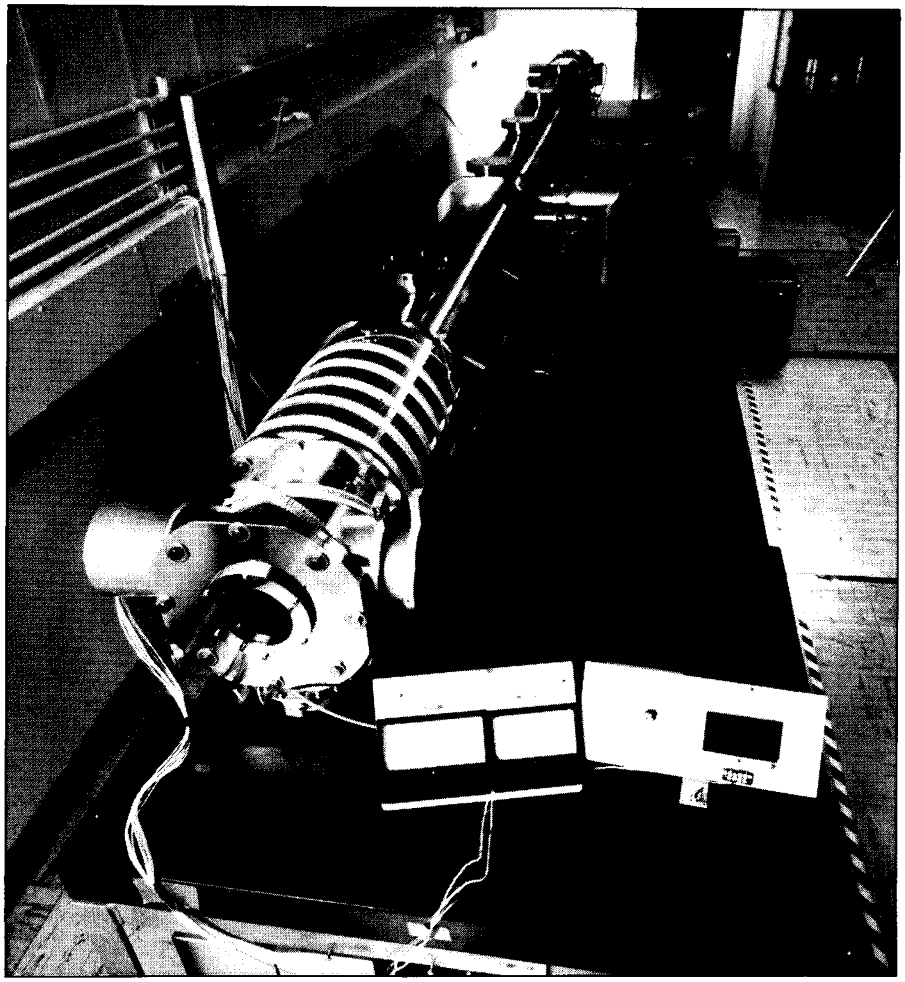
### Today's frequency standards

Atomic- and molecular-beam physics, beginning with the experiment of Otto Stern and Walther Gerlach in 1921, is the foundation of today's atomic time and frequency standards. I. I. Rabi's ideas of 1937 to use an oscillator-driven magnetic field to induce transitions[1] led to successful atomic-beam resonance experiments by Jerrold Zacharias, Polykarp Kusch, J. M. B. Kellogg and Norman Ramsey between 1938 and 1940.

Today's atomic cesium-beam frequency standard is strikingly similar to these first atomic-beam devices. Figure 1 is a photograph of a primary standard, the cesium atomic-beam device called "NBS-6" at the National Bureau of Standards in Boulder, Colorado, and figure 2 shows its basic configuration. Atoms are state selected by spatial deflection in the inhomogeneous magnetic field of magnet A. Similarly, these states are analyzed by a second inhomogeneous magnetic field (magnet B) with a detector placed in the proper position to intercept only one of the many possible energy states. We detect resonance transitions induced in the region between these two magnetic state selectors by the reduction of intensity that results from the depletion of the particular state being detected. A high reduction in Doppler line broadening yields very sharp resonances; fractional linewidths of $10^{-8}$ or better are possible. For a frequency standard the detector signal in a feedback loop controls the frequency of the interrogating[2] microwave radiation derived from a local oscillator, usually a quartz crystal. Cesium came to be used because of its convenient frequency (9.2 GHz), adequate vapor pressure near room temperature, and because of the efficient detectability of cesium by surface ionization on a hot wire.

Other atoms than cesium and other techniques than atomic beams have had an impact on the development of atomic time and frequency standards.[2] Optical pumping, developed by Alfred Kastler in 1950, led to the development of the rubidium gas-cell frequency standard shown in figure 3. The excitation of the microwave resonances changes the absorption of the pumping light, and the microwave transition can be detected by a reduction in the optical light intensity. This procedure leads to efficient detection and a good signal-to-noise ratio because, in principle, for every microwave photon absorbed one detects an optical photon. Rubidium 87, with its resonance at 6.8 GHz, is the preferred gas because the rubidium-85 isotope exhibits a light-resonance shift that allows the filtering of rubidium-87 light for selective optical pumping of one of the microwave energy levels. Unlike the nearly free cesium atoms, the rubidium atoms are stored in a vessel and collide with the walls of the vessel as well as with other atoms used as a buffer gas. This perturbation yields a lower accuracy than that found in the cesium-beam standard. However, rubidium standards are available commercially, and are useful when small size and low cost are important.

The first quantum electronic oscillator was the ammonia maser, built in 1954 by Eugene Gordon, Herbert Zeiger and Charles Townes. It was relegated to a place in history by the atomic hydrogen maser built and operated in 1960 by Mark Goldenberg, Daniel Kleppner and Ramsey. Figure 4 shows the principle. A vessel coated with a Teflon film is the hydrogen storage chamber. The wall collisions are essentially elastic and do not destroy phase coherence of the radiating



**The primary standard for the unit of time in the US** is this cesium atomic-beam device, NBS-6, at the National Bureau of Standards in Boulder, Colorado. See figure 2 for a schematic diagram of this device, which has an absolute accuracy of about $10^{-13}$ at its frequency of 9.2 GHz. Photograph by Geoffrey Wheeler, Boulder. Figure 1

atoms; it takes more than 10 000 collisions to induce relaxation of the atomic energy state. This property permits observation (that is, storage) times of more than one second, with fractional linewidths of $10^{-9}$ or better (the hydrogen resonance is at 1.4 GHz). Hydrogen masers of the Smithsonian Astrophysical Observatory have shown one of the best frequency stabilities of any known frequency source, reaching $7 \times 10^{-16}$ for averaging times of the order of hours. The present accuracy is limited to $10^{-12}$ in hydrogen maser standards by the residual phase shifts that the atoms suffer during the wall collisions.

Advances in quantum electronics have brought about the extension of radiofrequency techniques into the infrared and visible radiation regions. In the traditional atomic frequency standard, signals derived from crystal oscillators interrogate microwave hyperfine resonances in atoms. In the infrared and visible radiation regions, signals derived from laser oscillators interrogate atomic and molecular resonances ranging from electronic transitions in atoms to molecular rotations and vibrations. This development has led to highly stable devices such as the iodine stabilized helium–neon laser, and the methane-stabilized helium–neon laser.[3] Here, saturated absorption provides the needed reduction in first-order Doppler shifts and line broadening.

To represent the present state of development, we plot the stabilities and accuracies of some of the more common atomic frequency standards in figure 5.

### Physics of frequency standards

The guidelines for making a good frequency standard are fairly simple:

- it must be reproducible, and
- it must be "reasonably usable," in the sense that it should have an output frequency easily used in measurements.

The first requirement implies that bulk devices (such as quartz crystal resonators or superconducting cavity oscillators) are undesirable, because the frequency depends on parameters, such as size, that are difficult to control. This shortcoming does not, however, rule out the use of these devices as calibrated "flywheels." Atomic (or molecular) resonances provide the necessary reproducibility; one derives a "standard" frequency $\nu_0$ in terms of the energy difference between two states of the atom with energies $E_1$ and $E_2$ by the relation $h\nu_0 = E_2 - E_1$ where $h$ is Planck's constant. To ensure reproducibility between different observations, the measured frequency is usually referred to the value that would be obtained in free space; consequently, various corrections are necessary to take account of environmental factors, such as magnetic fields. The problem then reduces to measuring and correcting for the various perturbations to the measured frequency. Note that a frequency standard or clock based on atomic transitions is conceptually identical to the operation of a pendulum clock; in both cases a standard unit of time can be defined in terms of the time it takes for the completion of a certain number of periodic oscillations or vibrations.

Satisfying the second requirement depends on technological limitations and may rule out some interesting frequency-standard possibilities. In most cases the uncertainties in the perturbations to the measured resonance frequency scale as $Q^{-1}$ where $Q = \nu_0/\Delta\nu$, and $\Delta\nu$ is the width of the transition measured at the half power points; therefore a high $Q$ is desirable. Also, all measurements are limited in precision by various sources of noise. The fractional frequency stability[4] $\sigma(\tau)$ relates to $Q$ and signal-to-noise by

$$\sigma(\tau) \simeq \left[ Q \frac{S}{N}(\tau) \right]^{-1}$$

where $S/N(\tau)$ is the signal-to-noise ratio as a function of averaging time $\tau$. For example, if we detect the number of atoms that have made a transition in an atomic-beam apparatus, then we are often shot-noise limited by the random arrival times of the atoms at the detector. We can find the signal and noise terms by noting that $S$ is proportional to $I\tau$ and $N$ is proportional to $(I\tau)^{1/2}$ where $I$ is the mean number of atoms arriving at the detector per second; therefore $S/N(\tau) = (I\tau)^{1/2}$ and the frequency stability improves with averaging time as $\tau^{-1/2}$. Often, there is a trade-off between signal-to-noise and $Q$. Extremely high $Q$ does not guarantee a good frequency standard because, if the signal-to-noise ratio is small, it may take an impractical length of time to realize the accuracy.

Finally, the output of the device (or the operating frequency) must be convenient for general application. This measurement problem is being solved for laser frequency standards and it may soon be possible to intercompare frequency standards from very low frequencies to optical frequencies with the accuracy and resolution provided by the standards themselves. However, for the foreseeable future, certain transitions that can be realized at frequencies beyond the visible, such as those in nuclei by the Mössbauer effect, are ruled out for frequency standards purposes. Although the $Q$ realized in these transitions may be as high as $10^{15}$, they are not generally usable because neither frequency nor wavelengths can be accurately measured in the gamma-ray region.

To make a frequency *measurement* on an atom or molecule, the general requirements are the same as those faced by all spectroscopists; but to make a frequency *standard,* one looks for the special cases where the $Q$ and signal-to-noise are particularly high, and where the perturbations on the resonance are very small.

### Narrow spectral lines

The resolution is fundamentally limited by the Heisenberg uncertainty relation on time and energy:

$$\Delta E \Delta t \gtrsim \hbar/2$$

Thus, for a single atom, if we have a time $\Delta t$ to measure the energy-level difference $E_2 - E_1$, the measurement will be uncertain by at least an amount $\Delta E$. Specifically, $\Delta t$ may arise from the time of flight (transit time) of an atom through the apparatus, or from the lifetime of the atom in one or both of its energy states. The uncertainty relation yields a fractional uncertainty in energy of $\Delta E/(E_2 - E_1) = Q^{-1}$. We can, of course, make $\Delta E$ small by making $\Delta t$ large; this may be accomplished by slowing down the atoms as much as possible or by confining them.

Doppler effects are related to the particular method of confinement. They represent perhaps the most important problem limiting the accuracy and resolution of existing or proposed frequency standards. In the usual way we can say that if an absorber of radiation moves relative to the source, the observed resonance frequency is shifted to the value

$$\omega_{\text{abs}} \simeq \omega_0 + \mathbf{k} \cdot \mathbf{v} - \frac{1}{2}\omega_0 \left(\frac{v}{c}\right)^2 + \frac{\hbar \omega_0{}^2}{2Mc^2}$$

where the velocity **v** and wave vector **k** are measured relative to the source. The first-order Doppler shift ($\mathbf{k}\cdot\mathbf{v}$), the "second-order" Doppler shift, ($(v/c)^2$) and the recoil shift (the last term) can be understood in terms of conservation of energy and momentum in the emission or absorption process. Specifically, the so-called "second-order" Doppler shift is merely the relativistic time-dilation factor resulting from the movement of the atom relative to the apparatus. Its effect is small but important; this shift is especially small in the cesium-beam frequency standard, where it is approximately $10^{-13}$. We can describe the first-order shift in terms of the time dependence of a plane electromagnetic wave as seen by an atom. We have for the electric field

$$\mathbf{E}(x,t) = \mathbf{E}_0 \sin(\mathbf{k} \cdot \mathbf{x} - \omega t + \phi)$$

where **x** is the atom position, **k** is the field wave vector ($\mathbf{k}\cdot\mathbf{E}_0 = 0$), and $\phi$ is an arbitrary phase factor. If $\mathbf{x} = \mathbf{v}_x t$ then

$$\mathbf{E}(t) = \mathbf{E}_0 \sin[(\mathbf{k} \cdot \mathbf{v}_x - \omega)t + \phi]$$

and the particle sees a sinusoidally varying field of frequency $\omega' = \omega - \mathbf{k}\cdot\mathbf{v}_x$. The result is the familiar Doppler broadening because, typically, the atoms in a sample have a Maxwellian velocity distribution leading to a distribution of $\omega'$ values. If one observes the full Doppler width, then the $Q$ of the transition is limited to about $10^6$ for room-temperature samples.

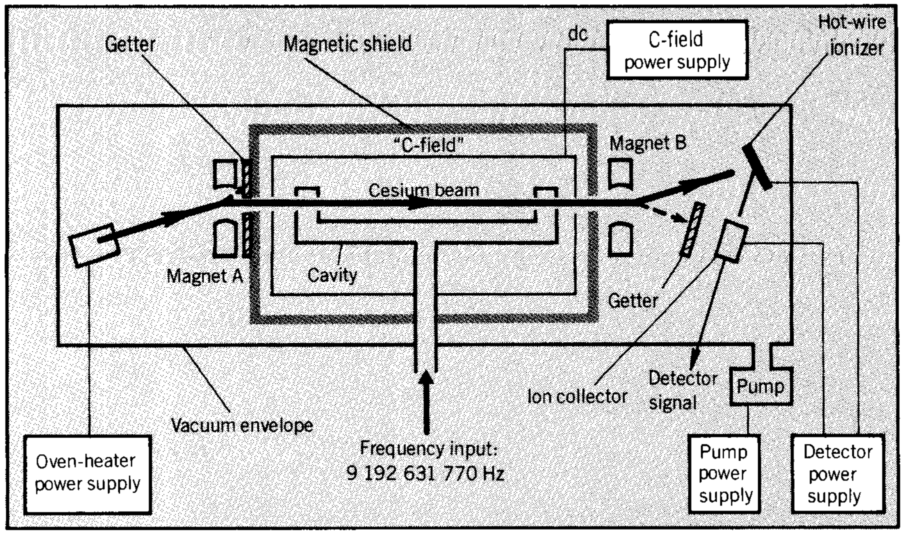
If the particle is confined within dimensions $|\mathbf{x}| < k^{-1}$ (the so-called "Dicke regime") the resonance spectrum has a sharp central feature with natural width (figure 6). This effect, first noted by Willis Lamb in 1939, was discussed in an important paper by Robert Dicke[5] in 1953. This condition is more easily realized in the microwave region and the initial "sub-Doppler" experiments were done there. We note however that the Dicke regime has also been realized in the gamma-ray region of the spectrum and is the basis for the Mössbauer effect, where, as in the microwave case, the photon recoil momentum is transferred to the whole lattice and does not appear as a frequency shift.

When we cannot meet the condition $|\mathbf{x}| \ll k^{-1}$, there is still the possibility of obtaining the same effect if we can satisfy the more general condition $\mathbf{k}\cdot\mathbf{v}_x \Delta t \lesssim 1$, where $\Delta t$ is the transit time of the atom through the apparatus. This is the general condition that must be met in a molecular beam apparatus. It allows for saturated absorption ("Lamb-dip") spectroscopy where atoms satisfying this condition are preferentially detected. Qualitatively, in this case, the detected atoms traverse the apparatus in a direction nearly normal to the traveling-wave propagation direction, and, therefore, the spatial phase change of the field experienced by the atoms is less than one radian.

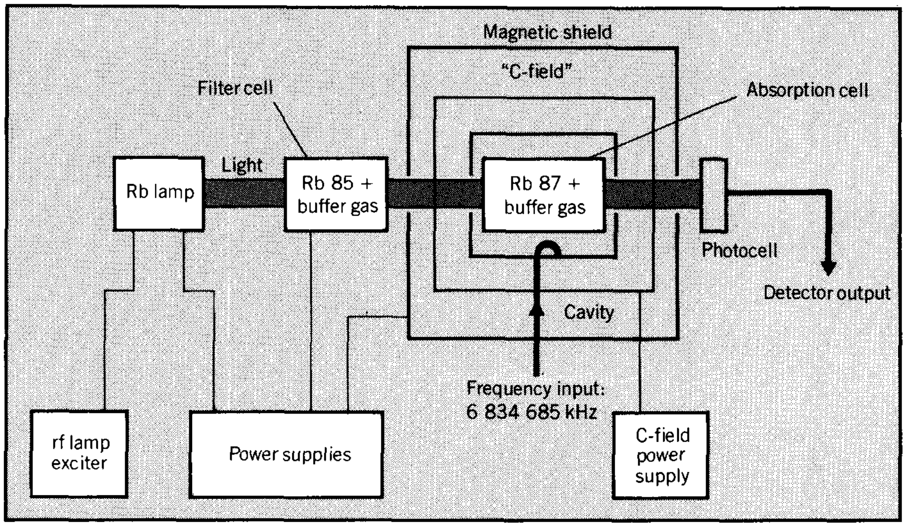
Even when the atoms interact with a non-plane wave, such as a standing wave in a microwave cavity, we can relax the condition on confinement if the atoms interact with the radiation in two, phase-coherent, spatially separated interaction regions. In each interaction region, the condition $\mathbf{k}\cdot\mathbf{v}_x \Delta t \lesssim 1$ is satisfied—where $\Delta t$ is now the transit time through one of the interaction regions. However, we must now calculate the $Q$ from the much larger transit time between interaction regions. This principle is the basis of Ramsey's separated oscillatory field technique,[6] which is used in all commercial and laboratory cesium clocks. This technique and its counterpart, the use of temporally separated oscillatory fields, were first used in the radiofrequency and microwave regions, and have now been realized in the optical region as well.[9]

Finally, the confinement problem has a rather unique solution, in the form of Doppler-free, two-photon spectroscopy. Here the atom interacts with counter-propagating plane waves of frequency $\nu_0/2$. The atom can resonantly absorb two photons simultaneously, one of frequency $\frac{1}{2}\nu_0(1 + \mathbf{v}\cdot\mathbf{k}/c)$ from one of the running waves and one with frequency $\frac{1}{2}\nu_0(1 - \mathbf{v}\cdot\mathbf{k}/c)$ from the counterrunning wave. The total energy from the two photons is $h\nu_0$, independent of the atomic velocity. Important applications exist in the optical region,[9] but the technique may be limited in accuracy by dynamic Stark shifts resulting from the required intense light field.

In the past, most frequency standards—such as the hydrogen-maser and the rubidium standard—have made use of Dicke narrowing or of the Ramsey technique as in the cesium beam. Although the first-order Doppler shift is nearly absent in all of these devices, the hydrogen maser has a large wall collision shift. The rubidium device, in addition,




**Cesium atomic-beam resonator,** which operates on principles similar to the Stern–Gerlach experiment. The frequency input is derived from a quartz-crystal oscillator (typically at 5 MHz) with a frequency multiplier and synthesizer to generate the atomic resonance frequency. A feedback servo from the detector output then controls this oscillator. Figure 2




**Rubidium atomic resonator,** which yields a lower accuracy than the cesium-beam device but is smaller and less costly. Rubidium and cesium standards are available commercially. Figure 3

suffers from large environmental shifts (light and buffer-gas shifts). The cesium device is free of the confining shifts but suffers from residual first-order Doppler shifts due to the presence of running-wave components in the interaction cavities. This inability to obtain pure standing waves generally affects all of the sub-Doppler techniques except Dicke narrowing. Thus, we have a tradeoff between the confinement techniques, which "eliminate" first-order Doppler shifts but introduce perturbations due to the confinement, and the "free"-atom techniques, which have no confinement perturbation but introduce Doppler shifts. Moreover, all of the techniques, including those employing Dicke narrowing, suffer from the "second-order" Doppler shift because the atoms have a non-zero motion. In some devices, such as hydrogen and rubidium, this is not a serious limitation, because the atoms are in thermal equilibrium with the container, whose temperature can be measured.

Improvements are still possible with the "sub-Doppler" techniques. In the separated oscillatory field technique, we can use different frequencies in each of the interaction regions and employ state-selected beams whose velocity distribution is trajectory independent (for example, by optical state selection and detection) to eliminate the residual first-order Doppler shifts.[7] In the microwave region there is also the possibility of using superconducting cavities, which nearly eliminate residual running waves.

In the optical region substantial improvements are possible; for example, in saturated absorption, operation with **k** parallel to **v** gives a precise knowledge of **v** (and therefore, precise knowledge of the second-order shift) and makes the method less sensitive to misalignments between saturating and probe beams.[8] We emphasize the importance and desirability of higher frequency operation because, if the interrogation time is held constant, the $Q$ is proportional to the frequency.

Both the first- and second-order Doppler shifts could be reduced, in a fundamental way, if the atoms could be slowed down or cooled. This cooling was attempted by Zacharias in 1953 in his "fountain" experiment[10] in which only the very slow atoms from an effusive source were selected by gravity. Unfortunately, the number of slow atoms available is too small to be useful. But there is the possibility of using radiation pressure to cool, both for neutral atoms and for ions stored in electromagnetic traps (see figure 7). This cooling technique has been incorporated into schemes for optical trapping[11] and has already been realized with stored ions.[12] Hans Dehmelt has long pointed out the advantages of the stored-ion technique, where the confining forces are relatively benign and where the Dicke criterion can be realized without substantial perturbations. The more recent availability of tunable, narrow band, optical sources will make detection of the small numbers of ions (typically $10^5$) much easier than in the past. Cooling of electromagnetically confined atoms or ions to less than a millikelvin should be possible, enabling one to satisfy the Dicke criterion in the optical region and thereby reduce the first- and second-order Doppler shifts by several orders of magnitude.

## Generation of time

Frequency standards have found an important use in the keeping and generation of time. First in the field were the quartz-crystal oscillators of the 1920's, which soon found extensive service in timekeeping and in interpolation between astronomical time fixes; however, the rotating Earth remained the primary frequency and time reference. Harold Lyons,[13] in 1949, was the first to demonstrate the atomic-clock principle; he used the ammonia inversion transition. Louis Essen and J. V. L. Parry of the National Physical Laboratory in the UK operated the first practical laboratory atomic cesium-beam clock in 1955. Their clock then saw extensive use as an actual frequency and time standard. In 1956, the first commercial atomic-beam frequency standard appeared on the market; the National Company produced it under the name "Atomichron." Also, several more accurate laboratory standards were constructed throughout the world. These developments demonstrated that cesium standards were capable of keeping time in a reliable way with substantially more accuracy than was possible by astronomical observations. Several international agreements and practices followed. Among them is the redefinition of the second in terms of the cesium resonance, which was concluded by the 13th General Conference on Weights and Measures in 1967:

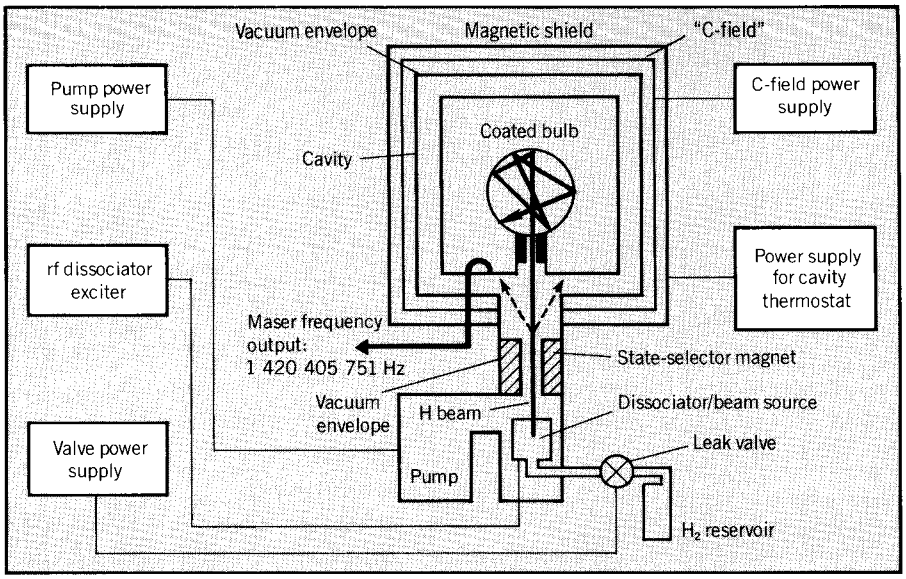
> "the second is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium-133 atom."

Atomic time scales have been kept since the mid-1950's by means of reliable and continuous integration of "cesium seconds." The Bureau International de l'Heure in Paris maintains International Atomic Time (abbreviated TAI) from the data of about a hundred atomic clocks scattered around the world. The available precision is about one microsecond per year. The source of most official and legal time of the world, however, is Coordinated Universal Time (abbreviated UTC), which is derived from TAI and approximates the changing rotation rate of the Earth by insertion or deletion, when necessary, of whole seconds (leap-seconds) with respect to TAI. For the last several years, only one leap second each year has been inserted.

## Clocks in science and technology

Two of the many uses of precise time and frequency standards in science are in radioastronomy and relativity. In a typical Very Long Baseline Interferometry experiment,[14] two or more radio antennas receive signals from one or more stellar radio sources. The greater the separation of the radio antennas, the higher the spatial resolution. Each antenna site requires a precise frequency standard to provide coherence. Interference fringes are produced by correlation of the two or more recordings. Since the observations are done in the microwave wavelength region and observation times of many hours are common, the experiments demand clock stabilities of $10^{-14}$ or even better; hydrogen maser standards have been of particular use.

Certain tests of theories of relativity involve experiments that detect changes in frequency and in accumulated time of clocks moving relative to each other and in different gravitational potentials. Such tests include spacecraft experiments and clocks flying in airplanes. Particularly interesting are the experiments that called for clock ensembles to be flown in commercial airliners around the world, demonstrating experimentally about a 400-nanosecond time difference between clocks flown *with* as compared to clocks flown *against* the rotation of the Earth.[15] Also, Robert Vessot and his co-workers have used a hydrogen maser space probe in a four-hour flight, 10 000 kilometers into space, returning in a ballistic trajectory back to Earth. This latter experiment yielded an accuracy of $10^{-4}$ in the measurement of the gravitational redshift.[16]




**Hydrogen maser used as a frequency standard.** In this device, unlike the cesium and rubidium resonators, the microwave input signal is phase-compared with the maser output for stabilizing the crystal oscillator. Accuracy is presently $10^{-12}$, although frequency stability up to $7 \times 10^{-16}$ has been achieved for periods up to several hours. Figure 4

Clocks and frequency standards are the references in almost every frequency and time-based measurement. Precise time is used for the operation of national power networks and permits precise measurement of the phase that allows the sharing of power by different networks. Frequency standards provide a high degree of phase stability in television broadcasts, which ensures color purity and permits fast switching and merging between different programs.

Time standards have been a part of navigation since John Harrison invented his chronometers in the 18th century. Modern navigators use either stellar observations or radio signals for position finding. In a more general sense, position finding amounts to obtaining a four-dimensional solution for the three spatial coordinates and the one time coordinate. A classical method of position finding on the surface of the Earth involves the observation of star positions and the use of a table giving the evolution of their positions in time, referenced to some fixed location on Earth. If such an observer has a sufficiently precise clock he can infer his own position.

A major space-based navigation system, currently in its first testing phase, makes use of atomic clocks in orbiting satellites. In the present system, rubidium and cesium standards operate on board the space vehicles, and improved rubidium, cesium and hydrogen standards are envisioned for the final system. For position finding one obtains range data for which, in principle, time measurements are converted to length measurements via the speed of light. Since a radio signal propagates at the speed of light, approximately $3 \times 10^8$ meters per second, a measurement precision of one nanosecond in time is equivalent to a distance or position accuracy of 0.3 meters. Therefore, if the clocks on board the satellite are to provide a clock-limited position accuracy of better than 0.3 meters, one needs a clock performance of about $10^{-14}$ over time periods of one day or a better performance if longer time periods are involved.

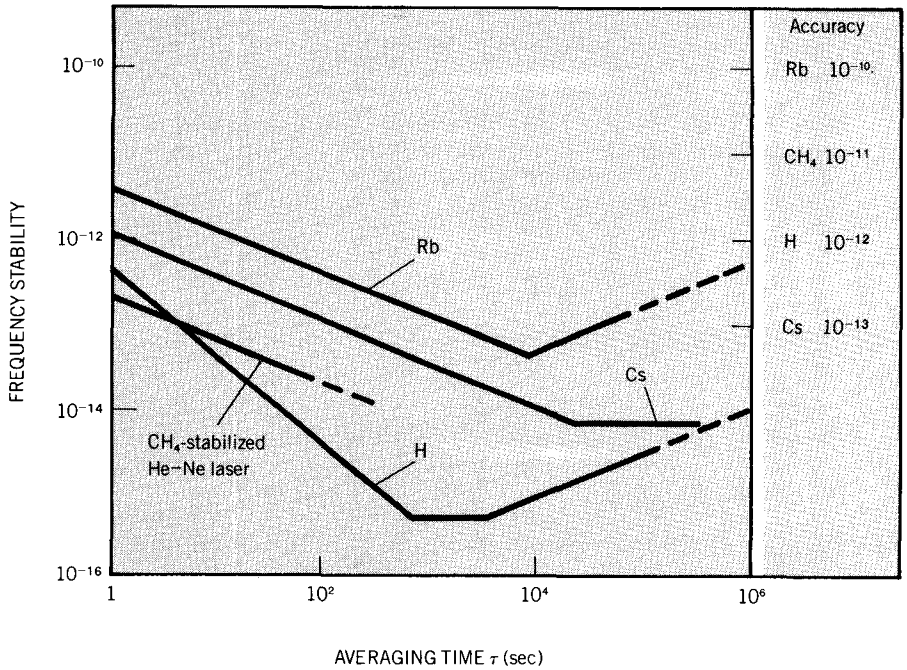
## Length and time

The standard of length, the meter, is defined in terms of the vacuum wavelength of the 0.6058-micron ($4.95 \times 10^5$ GHz) radiation from the krypton atom. Because the second is defined in terms of a resonance frequency (9.2 GHz) of the cesium atom, both the meter and the second are defined similarly in terms of quantum transitions and electromagnetic radiation.

Precise measurements of the absolute spectral characteristics in these two regions of the spectrum differ both in technique and in the precision attained. In the visible region (where the length standard is defined) wavelength techniques are generally appropriate and these are limited in accuracy by diffraction and optical imperfections to about $10^{-10}$. In the radiofrequency and microwave region, where the second is defined, direct frequency measurement techniques accurate to about $10^{-13}$ are used.

The measurement of dimensions of physical objects in the universe is limited mainly by irregularities in their surfaces with respect to the wavelength of visible light; hence accuracies of about one tenth of a visible wavelength are possible. In comparisons of wavelength, however, much greater accuracies are possible, especially when laser light, with its greater coherence, allows the observation of fringes over much greater distances. Any comparison of wavelength is done in interference devices—gratings, spectrometers and interferometers. Grating devices are limited in resolution by the number of lines, and by optical imperfections in the surfaces of the gratings, to a few parts in $10^7$. Interferometers are considerably more accurate; however, phase shifts at the mirror surfaces, uncertainties in mirror curvature, mode matching, and diffraction effects limit the comparative wavelength measurement to parts in $10^{11}$. The intercomparison also depends upon the coherence of the source and upon any asymmetries in the line shape. At present these effects limit the realization of the meter by the krypton lamp to about $4 \times 10^{-9}$. Lasers, in comparison, have much greater coherence, and measurements with lasers are not generally affected by lack of coherence. In fact, with frequency difference measurement techniques, stabilities approaching $10^{-14}$ are possible with the 3.39-micron He–Ne laser. Frequency measurements do not suffer from instrumental limitations as do wavelength measurements; rather the uncertainties are imposed by the sources themselves.

Laser oscillators generally do not provide very stable frequencies; the frequency can vary within the Doppler- and pressure-broadened gain curve of the laser medium. This gain curve might have a width from less than a hundred to several thousand megahertz, depending on the laser. Therefore, even though the instantaneous frequency of the laser can be measured with high precision, there is no reference point other than the broad gain curve whose center cannot be located very precisely. A solution to this problem was the development of the technique of saturated absorption,[8,9] which allows sub-Doppler observation of atomic and molecular resonances. The methane stabilized 3.39-micron He–Ne laser, for example, currently has an uncertainty in




**Fractional frequency stabilities** as a function of averaging time for cesium, rubidium and hydrogen frequency standards, and for a methane-stabilized helium–neon laser. The initial improvement of stability with time is due to increasingly precise averaging of fundamental noise processes; deterioration at longer averaging times is due to perturbations in the physical parameters affecting the frequency. The absolute accuracy of these devices appears on the right. Figure 5

its frequency reproducibility[8] of about $10^{-13}$. Thus, the radiation must have a similar wavelength reproducibility, that is, $\Delta\lambda/\lambda \simeq 10^{-13}$ which is more than $10^4$ times better than the current krypton length standard. With additional development in laser stabilization and infrared frequency synthesis some of these devices may be considered as contenders for the primary frequency standard role[3,8,9]

### Frequency measurements in the ir

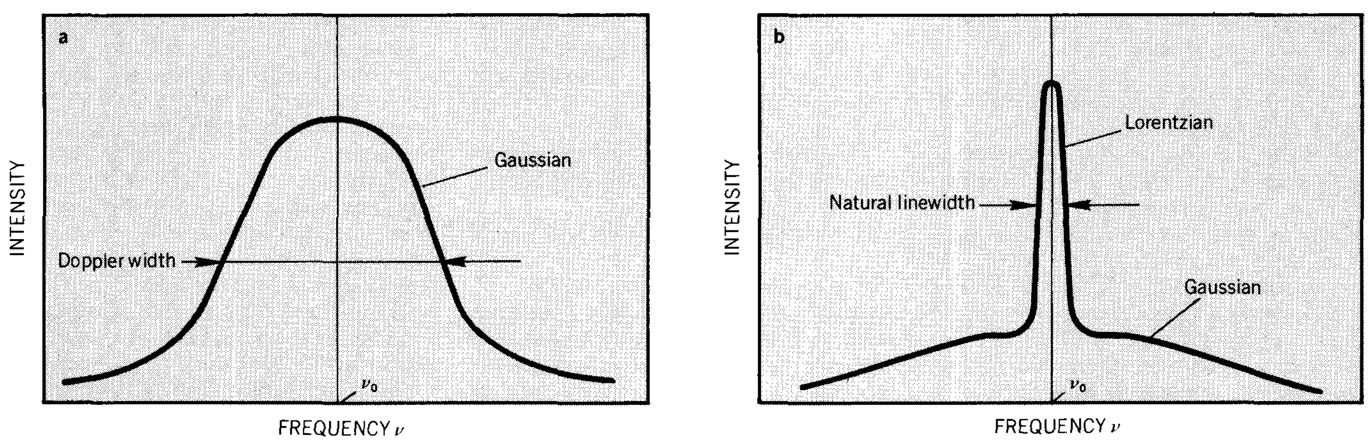
An essential part of frequency measurements in the infrared and visible radiation region is the interaction of electromagnetic radiation with some high-speed nonlinear device so that the sinusoidal wave is distorted, and hence possesses harmonics of the initial radiation. Thus, even though lasers provide coherent frequency sources in the infrared and visible, their frequencies cannot immediately be measured because no device exists capable of generating frequencies of 100 THz or more from multiple harmonics of microwave sources. This problem is currently overcome with a chain of stabilized lasers whose frequencies are connected by harmonic generation in the metal-oxide–metal point contact tunnel diode (see the cover of this issue of PHYSICS TODAY). This device[17] has now been shown to be operable to a frequency of 200 THz (or a wavelength of 1.5 microns). It is capable of generating twelve harmonics of far-infrared lasers and of adding and subtracting fundamentals and harmonics from as many as four lasers and a microwave source.[18]

This metal-oxide-metal diode is quite similar in operation to the cat's-whisker detector diode used in early crystal radios as well as in present microwave receivers. It consists of a polished nickel base, two or three millimeters in diameter, which is precisely positioned against a tungsten whisker. The whisker is sharpened to a tip of radius 0.06 microns. This junction is adjusted to a dc resistance of a few hundred ohms, and the resulting nonlinearity, although small, has an extremely fast response. Since it has permitted the generation of harmonics at 150 THz of a CO laser, the response time of this amazing detector may be inferred to be less than $10^{-14}$ seconds. In the far infrared, point-contact Josephson junctions also have been used successfully for the generation of high harmonics (several hundred) from microwave sources at the National Physical Laboratory and at the National Bureau of Standards.

At present, direct frequency measurements have been extended from the cesium standard all the way to the 1.15-micron radiation (260 THz) of the He–Ne laser. Thus, in the last decade, the portion of the electromagnetic spectrum accessible to direct frequency measurements has increased from HCN laser frequency at 0.9 THz (337-micron) made by Ali Javan and his co-workers to the present "world record" (achieved at the National Bureau of Standards) of 260 THz. This 300-fold increase of the region of direct frequency measurements represent most of the region from the microwave to the visible, and highly accurate, direct frequency measurements may soon be possible in the visible itself.

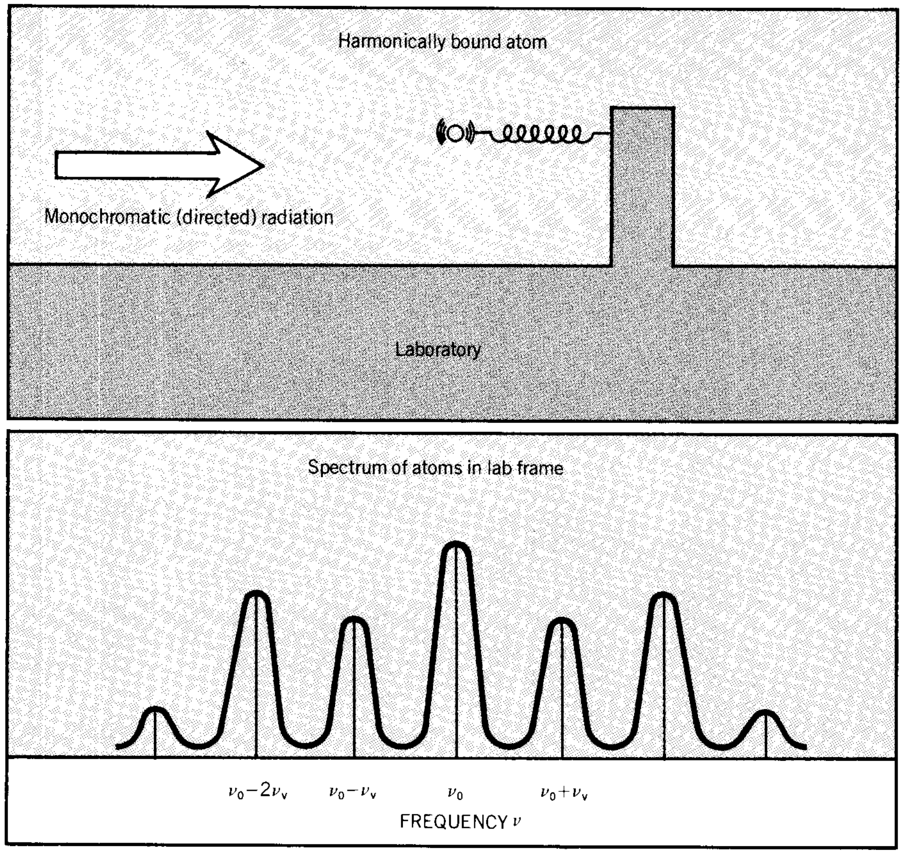
### The speed of light

The process of extending direct frequency measurements to the present world record included a measurement of




**Spectrum of an atomic transition.** On the left, part **a** shows the situation when the atoms are unbound and the resonance feature has the full Doppler width $\Delta\nu_D \approx (v/c)\nu_0$. When the atom is confined to dimensions less than the wavelength, the Doppler profile is suppressed and the central feature has the natural width $\Delta\nu$. This condition is most easily realized in the microwave region of the spectrum. Figure 6 the methane stabilized laser frequency, 3.39 microns. At this short wavelength, diffraction and mode uncertainties become smaller than uncertainties due to the incoherence in the krypton standard of length, and a highly accurate wavelength of this radiation is also possible. It was thus obvious that a highly precise value of the speed of light would result from the product of the frequency and wavelength of this radiation. In fact, the accuracy of $c$ improved by nearly a hundred-fold to a value uncertain only by the inaccuracy of the krypton realization of the meter.[19] Thus this value of $c$ can only be improved through a redefinition of the meter itself.

We feel it is increasingly important to have coherent, highly stable, visible sources of radiation whose frequency can be directly related to either the primary standard of length or of frequency. It is possible to use lasers as frequency references for certain kinds of problems such as high resolution optical heterodyne spectroscopy. The extension of absolute frequency measurements, linking the cesium standard (accurate to about $10^{-13}$) to these lasers, provides accuracy as well as resolution for the absolute frequencies involved. At the same time, the wavelength aspect of the radiation applies, for example, in precision long-path interferometry. Indeed, the increasing resolution these new spectroscopic techniques provide may very well usher in an era of precision and accuracy in frequency and length measurements undreamed of a few years ago.




**Radiation pressure cooling** for a bound atom. The spectrum in the lab frame is given by $I(\nu_0 \pm m\nu_v) = J_m^2(kx_0)$, where $\nu_v$ is the frequency of vibration. If the incident radiation is tuned to a lower "sideband," absorption occurs at frequency $\nu_0 - m\nu_v$. However, re-emission on the average occurs at nearly $\nu_0$; the energy deficit must come from the kinetic energy of the atom. Figure 7

**Our physical measurement system**

The basic standard for time interval, as realized by the cesium atomic-beam apparatus, is not only the most accurate frequency and time standard, but is also the most accurate of all basic standards[20] by a considerable margin, as is illustrated in figure 8. One of the four standards of figure 8, the standard for mass—the kilogram—is based on an artifact: It is a single macroscopic piece of matter made of material that is believed to vary as little as possible with time and environmental conditions. Length is defined in terms of the wavelength of a transition in krypton; temperature, in terms of the triple point of water. However, the realizations of the latter two standards suffer from a series of added experimental constraints, such as the prescription of certain pressures and sizes.

In comparison to these, the unit of time is realized by a laboratory measurement of the resonance frequency of the cesium atom. Corrections are made for all effects due to the cesium beam apparatus, the atomic motion and perturbations by fields, thus approaching the condition of an atom at rest in free space. The International System of Units features three additional base standards, those for electrical current (ampere), amount of substance (mole) and luminous intensity (candela). However, these standards depend in their definition[20] on one or more of the standards shown in figure 8.

We now examine the need for a certain set of base standards in the world of physical measurements. It is often thought that physical measurements require at least three fundamental and independent standards—those of time, length and mass. An example of this, familiar to physicists, is the cgs system of measurements. However, it is very important to realize that there is no fundamental physical reason for such an assumption.

Physical processes are expressed by equations connecting the variables. In some cases these equations contain physical constants. We now consider three such equations: The first one, Newton's equation

$$F = m \times a$$

relates the force $F$, to the acceleration, $a$, of an object of mass $m$. On the right side of the equation the quantity "mass" is directly based on an independent standard, and acceleration is directly derived from the fundamental standards of time and length. On the left side a derived variable, force, is defined by this equation. Therefore, no (fundamental) constant appears in this equation.

To illustrate our point further, consider the equation of electromagnetic wave propagation,

$$c = \lambda \times \nu$$

In this equation the physical variables, $\lambda$ and $\nu$, each relate to two independent physical standards in terms of our present system of measurement: $\lambda$ relates to the standard for length, and $\nu$ relates to the standard for time. Therefore, in this description of a fundamental physical phenomenon, the speed of light, $c$, becomes a fundamental physical constant linking two independent standards of physical measurement.

A third equation is that of the Josephson effect,

$$V = \frac{h}{2e}\nu$$

which relates voltage, $V$, to frequency via two fundamental physical constants, Planck's constant $h$ and the electron charge $e$. These quantities $h$ and $e$ are fundamental physical constants because $\nu$ relates to the standard of time and $V$ relates to the standards of mass, length and time.

If we had an independent standard for force (which, of course, is difficult to imagine), a fundamental constant would appear in the first equation. Conversely, if we remove the independent standard of length in our second equation, in the metrologist's view, $c$ would lose the quality of a fundamental physical constant and become a scale factor. In the third equation, for the Josephson effect, we have less freedom. We cannot simply make $h/2e$ a scale factor and thus create a new volt standard. The reason for this lies in the fact that the volt is related to several base standards, and the use of $h/2e$ as a scale factor is only possible if the standards of mass and length are simultaneously eliminated.

We must keep in mind the primary importance of providing physical measurements with references that are as precise and practical as possible and with an ability to compare results of the past with those of the present. Therefore, one must be very cautious in considering revisions of the system of fundamental standards and constants. Also, we often find practical realizations of derived units that are far more "accurate" than more complex realizations from the base standards. An example is the Josephson effect which allows a realization of the volt to better than $10^{-7}$.

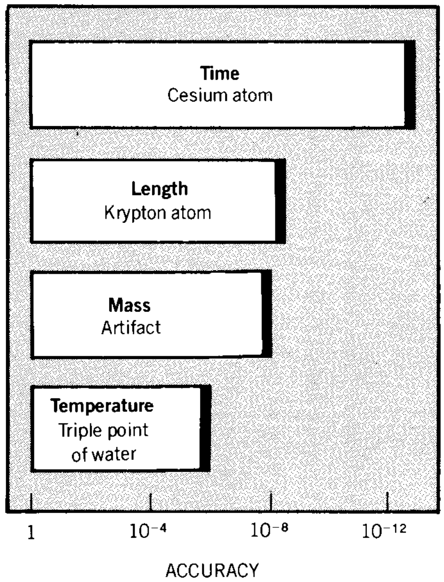
### Concept of the unified standard

What should now be clear is that it is more a matter of convenience and utility than a fundamental physical necessity to define a certain set of base standards and, with it, a unique set of fundamental physical constants. Therefore, if we change the number of independent base standards, we will also change the number of independent physical constants. In fact, reducing the number of independent base standards causes a conversion of fundamental physical constants into mere scale factors. These scale factors would be similar to quantities such as the dielectric constant or the magnetic susceptibility. Conversely, increasing the number of independent base standards leads to a corresponding increase in the number of fundamental physical constants. This may be a too simplistic view of physical realities. One may eventually have to allow for the possibility that these scale factors or fundamental constants are time dependent (as some theories of cosmology suggest) or are functions of other variables such as frequency, direction in space, and so on. However, as such effects have not yet been observed, the problem is not addressed here.

From a metrologist's viewpoint we could consider the possibility of just *one* base standard with all other reference quantities and physical measurements derived via scale factors from this base standard.

The concept of a unified base standard is not new. For example, Townes advocated the unified standard for frequency, time and length in his Nobel lecture in 1965.

The second equation above (for wave propagation) points out, as we discussed




**Absolute accuracies** available from four of the seven base standards included in the International System of Units. Figure 8

already, how to eliminate an independent standard of length. We need only define the speed of light and thus derive the length standard *via* a wavelength measurement of a known frequency. This procedure is possible because these two physical variables directly relate to their respective, present base standards, in a simple way. An essential requirement is that the corresponding physical measurement can be executed with ease and, what is more important, with precision and accuracy: We can only take the step of defining the speed of light in order to derive the length standard if we have an experimental possibility of measuring frequency and wavelength of an electromagnetic signal with a precision satisfactorily exceeding the accuracy of the inferior standard. The standard of length can be realized only to parts in $10^9$ (see figure 8), so we place the following two requirements on a possible redefinition of the length standard via a definition of the speed of light:

- the measurement must be done with about $10^{-10}$ precision (we need to exceed the present accuracy of the independent length standard by a significant amount), and
- this measurement must occur in a region of wavelength where practical length metrology is carried out, that is, in the visible region of the electromagnetic spectrum.

The first requirement is, as we pointed out previously, already fulfilled.[19] Not so easy is the realization of the frequency measurement in the visible region; however, this may well be accomplished in the near future. There is, of course, the additional concern of practicality. To be widely accepted by the metrology community, the realization of a unit of measurement must be simple enough to be duplicated in many standards laboratories.

As we have discussed, many wavelength measurements are already based on laser frequency references, and the length standard may formally be redefined in the not too distant future. The Josephson effect also has become the *de facto* primary reference for the volt *via* a frequency measurement. Magnetic fields are now measured in terms of frequencies *via* nuclear magnetic resonances. It is, therefore, useful and interesting to ask for possibilities to replace other base standards, as well as the length standard, such as the standards of mass and temperature, by defining physical constants and linking them to the standard of time. This is an exciting challenge facing metrologists and basic researchers.

### References


1. I. I. Rabi, Phys. Rev. **51**, 652 (1937).
2. H. Hellwig, Proc. IEEE **63**, 212 (1975), and N. F. Ramsey, Proc. IEEE Inst. and Meas. **21**, 90 (1972).
3. H. Walther, ed., *Laser Spectroscopy of Atoms and Molecules,* Springer, Berlin, (1976), chapter 6.
4. J. A. Barnes et al., Proc. IEEE **IM-20**, 105 (1971).
5. R. M. Dicke, Phys. Rev. **89**, 472 (1953).
6. N. F. Ramsey, *Molecular Beams,* Clarendon, Oxford (1956).
7. S. Jarvis Jr, D. J. Wineland, H. Hellwig. J. Appl. Phys. **48**, 5336 (1977).
8. J. L. Hall, Opt. Commun. **18**, 62 (1976).
9. T. W. Hänsch, PHYSICS TODAY, May 1977, page 34.
10. J. R. Zacharias, Phys. Rev. **94**, 751 (1954).
11. A. Ashkin, Phys. Rev. Lett. **40**, 729 (1978); also V. S. Letokhov, V. G. Minogin, B. D. Pavlik, Sov. Phys.-JETP **45**, 698 (1977).
12. D. J. Wineland, R. E. Drullinger, F. L. Walls, Phys. Rev. Lett. **40**, 1639 (1978); also W. Neuhauser, M. Hohenstatt, P. Toschek, H. Dehmelt, Phys. Rev. Lett. **41**, 233 (1978).
13. H. Lyons, Am. N. Y. Acad. Sci. **55**, 831 (1952).
14. W. K. Klemperer, Proc. IEEE **60**, 602 (1972).
15. J. C. Haefele, R. E. Keating, Science **177**, 166 (1972).
16. R. F. C. Vessot, M. W. Levine, in the proceedings of a conference on experimental gravitation published as: Accademia Nazionale della Linzei, Rome, **34**, 371 (1977).
17. A. Sanchez, C. F. Davis, K. C. Liu, A. Javan, J. Appl. Phys. **49**, 10 (1978).
18. E. Sakuma, K. Evenson, Proc. IEEE **QE-10**, 599 (1974).
19. K. Evenson, J. S. Wells, F. R. Petersen, B. L. Davidson, G. W. Day, R. L. Barger, L. Hall, Phys. Rev. Lett. **29**, 1346 (1972).
20. NBS Special Publication 330: "The International System of Units (SI)," 1977 edition. The official medium for publication of news by the International Bureau of Weights and Measures is the journal *Metrologia.* □