Origin of the Moon's orbital inclination from resonant disk interactions

William R. Ward & Robin M. Canup

Southwest Research Institute, 1050 Walnut Street, Suite 426, Boulder, Colorado 80302, USA

The Moon is generally believed to have formed from the debris disk created by a large body colliding with the early Earth[1,2]. Recent models of this process predict that the orbit of the newly formed Moon should be in, or very near, the Earth's equatorial plane[3,4]. This prediction, however, is at odds with the known history of the lunar orbit: the orbit is currently expanding, but can be traced back in time to reveal that, when the Moon formed, its orbital inclination relative to the Earth's equator was $I \approx 10°$ (refs 5, 6). The cause of this initial inclination has been a mystery for over 30 years, as most dynamical processes (such as those that act to flatten Saturn's rings) will tend to decrease orbital inclinations. Here we show that the Moon's substantial orbital inclination is probably a natural result of its formation from an impact-generated disk. The mechanism involves a gravitational resonance between the Moon and accretion-disk material, which can increase orbital inclinations up to ~15°.

The Moon-forming impact is believed to have produced a disk of debris[7–9] that was roughly centred at the Earth's Roche limit, $a_{\rm Roche}$, where $a_{\rm Roche} \sim 2.9R_\oplus$ and $R_\oplus$ is the radius of the Earth (6,378 km). The Roche limit is the distance interior to which a fluid body with no internal strength will be disrupted by a planet's gravity. Inside $a_{\rm Roche}$, collisional aggregation is inhibited by the Earth's tidal forces and orbiting material remains dispersed in a disk; consequently, it is the portions of the disk exterior to $a_{\rm Roche}$ that rapidly accumulate to become the Moon[3,4]. Accretion studies[4] find that producing an object the size of the Moon with an orbital radius of $a \approx 1.2a_{\rm Roche} \approx 3.5R_\oplus$ requires an initial disk containing at least two lunar masses, or $M_{\rm disk} \approx 2M_{\rm Moon}$. Such studies typically yield a moon with $I \approx 1°$. To date, only two explanations for the cause of the initial ~ 10° lunar inclination appear viable, and these require either another large impact improbably soon after the Moon-forming event, or multiple passes of the Moon through two orbital resonances involving the Sun under a rather specialized set of conditions[10]. The resonant interaction we identify here would precede either of these potential events, and is effective for a wide range of plausible conditions.

Estimates of the lifetime of an impact-generated protolunar disk have varied. Recent simulations[4] are computationally limited to describing the disk as several thousand "moonlets", each initially hundreds of kilometres in size. Such a disk persists for only about a year until objects are accreted by the growing Moon or scattered onto the Earth. However, moonlet-sized objects could not form within $a_{\rm Roche}$ to begin with, and so this short time estimate is probably applicable only to that portion of the disk exterior to the Roche limit. An inner disk composed of small debris would be more collisionally dominated and resistant to rapid depletion via scattering by the Moon than accretion simulations[4] suggest. Such a disk would respond to gravitational perturbations from a satellite in a collective fashion (by, for example, the generation of spiral waves), rather than through discrete scattering events. In addition, the rate of disk spreading (proportional to the disk's viscosity) predicted by the purely dynamical estimates[11,12] is so large that the corresponding energy release rate greatly exceeds the disk's radiative cooling rate[13]. A more appropriate limit would be a viscosity regulated by the disk cooling time, $\sim 10^2$ years.

These results suggest that the Moon would have co-existed for some time with an inner disk containing a good fraction of a lunar mass. The inclination-generating mechanism we propose involves a resonance between this inner remnant disk and the Moon. Disk–satellite resonances are well-studied processes associated with, for example, spiral waves observed in Saturn's rings[14] and the shepherding of the uranian ringlets[15].

Periodic perturbations by a satellite excite waves in a companion disk at locations where the ratio of the orbital period in the disk to that of the satellite is approximately equal to a ratio of two integers. Mean motion resonances that involve in-plane perturbations are known as Lindblad resonances; vertical resonances arise from perturbations perpendicular to the plane of the disk. The strongest resonances are zero-order, independent of either e (eccentricity) or I (inclination); the next strongest are first-order resonances which have a linear dependence on e or I. In an interior disk, the first-order $m = 2$ inner Lindblad resonance (ILR) occurs at the location in the disk where the orbital period is about one-third that of the satellite;

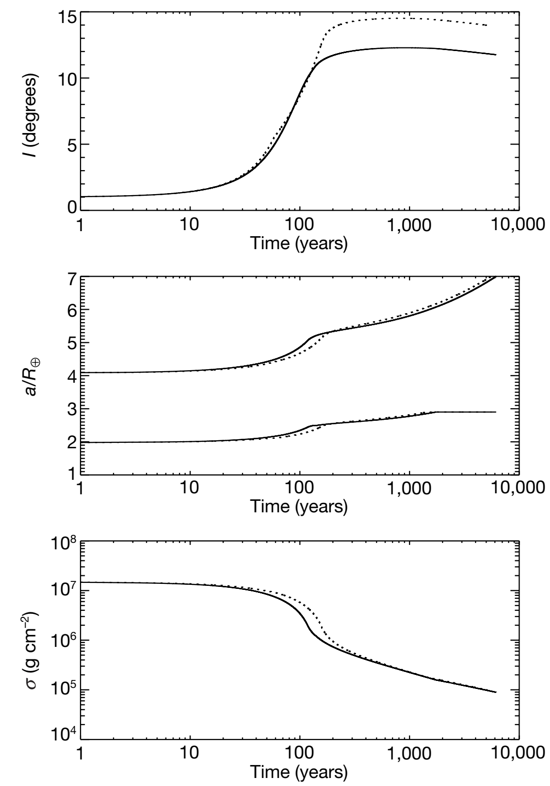

Figure 1 The evolution of the Moon's orbit as it resonantly interacts with an inner impact-generated disk. Values of inclination (I; top panel), radius ($a/R_\oplus$; middle panel), and disk surface density (σ; bottom panel) are shown as the Moon recoils away from a disk with $M_{\rm disk}(0) = 0.75M_{\rm Moon}$. Data are shown for $t_{\rm spread}(0) = 37$ years (solid lines) and 50 years (dotted lines). We assume that the Moon forms outside the Roche distance, and for some time co-exists with a disk that initially extended from $R_\oplus$ to $r_{\rm disk}(0) = a_{\rm Roche}$. The initial angular momentum of the system is defined by $M_{\rm disk}(0)$, and the assumption that material which comprises the Moon had the same starting angular momentum as a lunar mass orbiting at $\sim 1.2\ a_{\rm Roche} \sim 3.5\ R_\oplus$ (based on results to date of accretion simulations[3,4], although $I_{\rm max}$ is not overly sensitive to this choice). The disk edge (lower curves in middle panel) contracts to $r_{\rm disk} \approx 2R_\oplus$ by the time the Moon's orbit (upper curves in middle panel) expands to $\sim 4.1R_\oplus$ (due to zero-order and first order ILR torques) and the $m = 2$ IVR becomes the only first-order mean-motion resonance left in the disk. (We note that higher-order resonances also in the disk at this time will be rendered ineffectual by rapid eccentricity-damping by $m = 1$ secular ILR[20,21].) From $4.1R_\oplus$, the Moon's orbit is evolved due to the $m = 2$ ILR (if it can still be felt by the disk), the $m = 2$ IVR, and terrestrial and lunar tides (we use dissipation functions $Q_\oplus = Q_{Moon} = 50$, and tidal Love numbers $k_{2\oplus} = 10k_{\rm 2Moon} = 0.25$). The discontinuity at ~120 years is due to the switch from a radiation-limited viscosity[13] (equation (5)) to an instability-driven viscosity[11] (equation (6)).

here m corresponds to the number of spiral arms in the resulting density wave pattern. For an inclined satellite, spiral bending waves (which corrugate the disk) are launched at inner vertical resonances (IVRs)[14]. For resonances with $m > 1$, the interaction of a satellite with waves it generates in an interior disk causes a transfer of energy and angular momentum from the disk (which contracts) to the satellite (whose orbit expands). For a non-circular and/or inclined satellite orbit, disk torques also alter e and/or I (refs 16–18).The torque values applicable for a given resonance have been shown to be quite insensitive to the particular physical state of the disk material, such as its viscosity, pressure and self-gravity[19].

As a moon (or moons) first accumulates outside $a_{\rm Roche}$, there would be many resonances within a disk extending from the Earth's surface to the Roche limit. Torques from the strongest resonances would cause satellite-sized bodies to migrate outward at rates comparable to their month-to-year accumulation time[4], $t_{\rm acc}$. Satellite migration would be accompanied by a migration of the mean motion resonances (which each occur at a fixed fraction of a satellite's position, a); these resonances would pass outward through the disk, and finally leave it. As the strongest, zero-order resonances migrate out of the disk, the expansion rate of the moon(s) slows. In particular, when the orbital radius reaches $a \approx 3^{2/3} r_{\rm disk} \approx 2.08 r_{\rm disk}$, where $r_{\rm disk}$ is the radius of the disk's outer edge, the first-order $m = 2$ ILR leaves the disk, and the migration time for a lunar-sized body slows to a value that is much longer than $t_{\rm acc}$. At this point, the $m = 2$ IVR—which is located slightly interior[14] to the $m = 2$ ILR because of dissimilar precession rates of the orbital apse and node—becomes the only first-order mean-motion resonance remaining in the disk, and acts to increase the lunar inclination. Any eccentricity generated during the preceding recoil phase is quickly damped by the $m = 1$ secular resonance which resides in the disk interior to the $m = 2$ IVR (refs 20, 21).

The $m = 2$ IVR increases a and I, at the rates[14,16,17]:

$$\frac{\mathrm{d}a}{\mathrm{d}t} = \frac{3T_{\rm IVR}}{Ma\Omega} \quad \text{and} \quad \frac{\mathrm{d}I}{\mathrm{d}t} = \frac{T_{\rm IVR}}{Ma^2\Omega \sin I}\left(\frac{3}{2}\cos I - 1\right) \qquad (1)$$

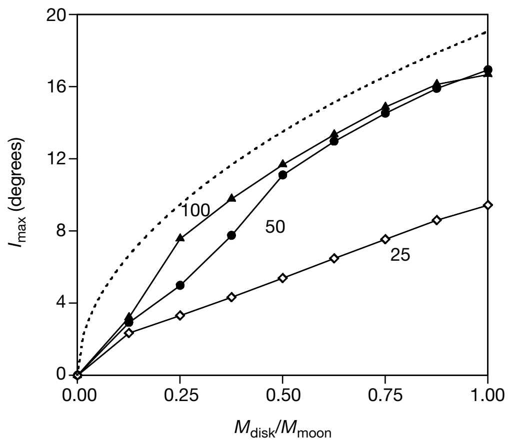

Figure 2 The Moon's initial orbital inclination. Maximum lunar orbital inclinations, $I_{\rm max}$, as a function of $M_{\rm disk}(0)$, are given for disk spreading times ($t_{\rm spread}$) of 25, 50 and 100 years and assuming that $a_o = 3.5R_\oplus$. A lower-mass disk yields a lower $I_{\rm max}$ for a given value of $t_{\rm spread}$ due to the dependence of the IVR torque on disk surface density. For a given $M_{\rm disk}$, shortening the disk spreading time also decreases $I_{\rm max}$. In this case, the time for $r_{\rm disk}$ to reach $a_{\rm Roche}$ or for the disk to be depleted (and thus the time over which the Moon experiences the IVR torque) is reduced. The dashed line is the analytical upper limit for $I_{\rm max}$, which neglects solid body tides and assumes that the IVR remains in the disk until the disk loses all of its mass to the Earth. In this case, the final angular momentum of the Moon equals the initial angular momentum in the disk and the Moon, minus that absorbed by the Earth as it accretes the disk, that is, $L_{\rm absorb} = M_{\rm disk}(GM_\oplus R_\oplus)^{1/2}$.

where M is the mass of the moon, Ω is its orbital frequency, and $T_{\rm IVR}$ is the torque on the moon due to the $m = 2$ IVR. While the IVR is in the disk, $T_{\rm IVR}$ is given by

$$T_{\rm IVR} = \sin^2 I \frac{2\pi^2\sigma}{3\Omega^2}\left(\frac{GM}{2a}\right)^2 [\alpha_{\rm IVR}^{5/2} b_{3/2}^2(\alpha_{\rm IVR})]^2 \qquad (2)$$

where σ is the disk surface mass density, G is the universal gravitational constant, $a_{\rm IVR}$ is the orbital radius of the IVR, $\alpha_{\rm IVR}$ is the ratio $(a_{\rm IVR}/a) \approx 3^{-2/3}$, and $b_{3/2}^2(\alpha_{\rm IVR})$ is a Laplace coefficient[22]. Assuming that the $m = 2$ IVR is the sole source of lunar migration and that $\cos I_0 \approx 1$, the rates in equation (1) can be divided and integrated to find

$$\cos I_{\rm max} = \frac{2}{3} + \frac{1}{3}\sqrt{\frac{a_1}{a_2}} \qquad (3)$$

where $I_{\rm max}$ is the maximum inclination produced by the IVR, and a_1 and a_2 are the lunar orbital radii when the first-order $m = 2$ ILR and the $m = 2$ IVR leave the disk, respectively. An estimate of a_1 is found by requiring conservation of mass, and that the sum of the angular momenta of the disk and the moon is equal to the starting angular momentum of the system. This is valid (independent of the details of earlier moon(s)–disk interactions) if the initial recoil is much faster than the disk can viscously spread, or than the moon can respond to solid body tides. The largest value of a_2 (and the largest lunar inclination) is achieved if the IVR remains in the disk while it spreads until the disk has lost all of its mass to the Earth.

A simple simulation illustrates the Moon's orbital evolution as it recoils from its precursor disk. We do not model the accretion phase, but assume that the Moon assembled quickly at an orbital radius $a_{\rm o}$ and has migrated outward while the disk contracts until $a(t) = 2.08 r_{\rm disk}(t)$. Recent impact simulations[9] yield debris disks with total masses $\sim 2M_{\rm Moon}$ and total angular momenta $\sim 0.3J_{\oplus\text{-M}}$, where $J_{\oplus-\rm M}$ is the current angular momentum of the Earth–Moon system. Such a disk yields a Moon on a circular orbit with $a_{\rm o} = 3.4R_\oplus$ (ref. 4); somewhat larger $a_{\rm o}$ values are possible given non-zero eccentricities for the Moon and inner disk debris. Accretion simulations[4] typically find ($a_{\rm Roche} \leqslant a_{\rm o} \leqslant 1.5a_{\rm Roche}$); for all values in this range, the $m = 2$ IVR is within the disk when the Moon forms.

We track the evolution of the lunar orbit as it evolves due to (1) the $m = 2$ ILR and IVR, (2) interaction with tides it raises on the Earth[5,6], and (3) energy dissipation due to lunar tides raised by the Earth[10]. The disk evolves due to both loss of angular momentum to the Moon via the $m = 2$ ILR and IVR, and viscous diffusive spreading. For the latter, we use a time-dependent disk viscosity, $\nu(t)$, equivalent to that predicted by dynamical simulations[12]and analytical estimates[11], unless this value implies an energy release rate that exceeds the disk's cooling rate. In the latter case, we assume a balance between these rates, so that $\sigma_{\rm S-B}T^4 = \frac{9}{8}\nu\sigma\Omega_{\rm disk}^2$, where $\sigma_{\rm S-B}$ is the Stefan–Boltzmann constant, T is the disk photospheric temperature, and $\Omega_{\rm disk}$ is the orbital frequency at $r_{\rm disk}$ (ref. 13). This balance yields a radiation-limited minimum disk spreading time in years, $t_{\rm spread}$, given by (ref. 13):

$$t_{\rm spread} = \frac{r_{\rm disk}^2}{\nu} \approx 50\left(\frac{r_{\rm disk}}{a_{\rm Roche}}\right)^{-3}\left(\frac{T}{2{,}000\,\rm K}\right)^{-4}\left(\frac{M_{\rm disk}}{M_{\rm Moon}}\right) \qquad (4)$$

The initial values of $t_{\rm spread}(0)$ and $M_{\rm disk}(0)$ (that is, those when $r_{\rm disk} = a_{\rm Roche}$) are input parameters. At each time step, $\nu(t)$ is updated to be the smaller of two values:

$$\nu(t) = \nu(0)\left(\frac{M_{\rm disk}(0)}{M_{\rm disk}(t)}\right)\left(\frac{r_{\rm disk}(t)}{a_{\rm Roche}}\right)^5 \qquad (5)$$

which is the radiation-limit for $\nu(t)$ from equation (4), or

$$\nu(t) = \left(\frac{M_{\rm disk}(t)}{M_\oplus}\right)^2 \Omega_{\rm disk} r_{\rm disk}^2(t) \qquad (6)$$

which is the viscosity obtained by a disk stability estimate[11]. The disk mass is decreased as material spreads onto the Earth or past a_{Roche}.

Figure 1 shows the evolution of I, a and σ as the Moon recoils from a disk of mass $M_{disk} = 0.75\ M_{Moon}$, assuming that $a_o = 3.5R_\oplus$ and $a_1 = 4.1R_\oplus$. The resulting maximum inclinations are 12.3° and 14.5° for $t_{spread} = 37$ and 50 years, respectively. The timescale, t_{IVR}, for growth in a due to the IVR is regulated by the location of the disk edge to be comparable to t_{spread}. As the IVR nears the edge of the disk, the back torque from the Moon retards the disk's tendency to spread, and the IVR begins to migrate out of the disk. But as this occurs, the back torque on the disk is decreased, allowing increased spreading of disk material outward, until once again the IVR is in the disk. In this manner, the location of $r_{disk}(t)$ and the location of the IVR move in lock step. Once $r_{disk} = a_{Roche}$, the edge of the continuous disk does not advance because material spreading beyond a_{Roche} can accrete into discrete objects which are not effective at sustaining wave action. The continued outward migration of the Moon then causes the IVR to leave the disk, preventing further resonant growth of I; the subsequent system evolution is determined by tidal interaction with the Earth and Sun[5,6,10]. Figure 2 shows I_{max} resulting from a variety of choices of $M_{disk}(0)$ and $t_{spread}(0)$.

Until now, an apparent deficiency of the impact hypothesis has been its inability to account for the inclination of the lunar orbit[23]. Here we identify a resonance that would naturally result as the Moon recoiled from an interior disk; this resonance could, for a range of plausible conditions, yield the requisite lunar inclination. The resulting inclination depends mainly on two parameters: $M_{disk}(0)$ and $t_{spread}(0)$. Simulations of potential Moon-forming impacts by Cameron[9] yield $M_{disk} \approx 1.5–2\ M_{Moon}$, with somewhat more than half of the mass exterior to a_{Roche}. Forming the Moon at the outer edge of such a disk would leave an inner remnant disk of $\sim$0.5–1 M_{Moon}. An inner disk whose spreading rate is regulated by its radiative cooling time with a nominal photospheric temperature of 2,000 K will have $t_{spread}(0) \approx 50(M_{disk}/M_{Moon})$ years. These values coincide well with those needed to yield $I_{max} \approx 10°$. The inclination of the Moon's orbit may thus represent an important clue to the early state of the post-impact system; we expect that this will motivate further modelling of the physics of an impact-generated protolunar disk. □

Received 8 September 1999; accepted 4 January 2000.

1. Cameron, A. G. W. & Ward, W. R. The origin of the Moon. *Lunar Sci.* VII, 120–122 (1976).
2. Hartmann, W. K. & Davis, D. R. Satellite-sized planetesimals and lunar origin. *Icarus* **24,** 504–515 (1975).
3. Canup, R. M. & Esposito, L. W. Accretion of the Moon from an impact-generated disk. *Icarus* **119,** 427–446 (1996).
4. Ida, S., Canup, R. M. & Stewart, G. Formation of the Moon from an impact-generated disk. *Nature* **389,** 353–357 (1997).
5. Goldreich, P. History of the lunar orbit. *Rev. Geophys.* **4,** 411–439 (1966).
6. Touma, J. & Wisdom, J. Evolution of the Earth-Moon System. *Astron. J.* **108,** 1943–1961 (1994).
7. Cameron, A. G. W. & Benz, W. The origin of the Moon and the single impact hypothesis IV. *Icarus* **92,** 204–216 (1991).
8. Cameron, A. G. W. The origin of the Moon and the single impact hypothesis V. *Icarus* **126,** 126–137 (1997).
9. Cameron, A. G. W. & Canup, R. M. The giant impact occurred during Earth accretion. *Lunar Planet. Sci.* **XXIX,** 1062–1063 (1998).
10. Touma, J. & Wisdom, J. Resonances in the early evolution of the Earth-Moon system. *Astron. J.* **115,** 1653–1663 (1998).
11. Ward, W. R. & Cameron, A. G. W. Disc evolution within the Roche limit. *Lunar Planet Sci.* **IX,** 1205–1207 (1978).
12. Kokubo, E., Ida, S. & Makino, J. Evolution of a circumterrestrial disk and formation of a single moon. *Icarus* (submitted).
13. Thompson, C. & Stevenson, D. Gravitational instability in two-phase disks and the origin of the Moon. *Astrophys. J.* **333,** 542–481 (1983).
14. Shu, F. in *Planetary Rings* (eds. Greenberg, R. & Brahic, A.) 513–561 (Univ. Arizona Press, Tucson, 1984).
15. Elliot, J. L. & Nicholson, P. D. in *Planetary Rings* (eds Greenberg, R. & Brahic, A.) 25–72 (Univ. Arizona Press, Tucson, 1984).
16. Goldreich, P. & Tremaine, S. Disk-satellite interactions. *Astrophys. J.* **241,** 425–441 (1980).
17. Borderies, N., Goldreich, P. & Tremaine, S. Excitation of inclinations in ring-satellite systems. *Astrophys. J.* **284,** 429–434 (1984).
18. Ward, W. R. & Hahn, J. M. Damping of orbital inclinations by bending waves. *Icarus* **110,** 95–108 (1994).
19. Meyer-Vernet, N. & Sicardy, B. On the physics of resonant disk-satellite interaction. *Icarus* **69,** 157–175 (1987).
20. Ward, W. R. & Hahn, J. M. The eccentricity of Neptune and the nature of the Kuiper belt. *Science* **280,** 2104–2106 (1998).
21. Tremaine, S. Resonant relaxation in protoplanetary disks. *Astron. J.* **116,** 2015–2022 (1998).
22. Brouwer, D. & Clemence, G. M. *Methods of Celestial Mechanics* (Academic, New York, 1961).
23. Boss, A. P. & Peale, S. J. in *Origin of the Moon* (eds Hartmann, W. K., Phillips, R. J. & Taylor, G. J. (Lunar and Planetary Institute Press, Houston, 1986).

Acknowledgements

We thank L. Dones, D. Stevenson, G. Stewart and J. Wisdom for comments, and A. G. W. Cameron for data from his SPH impact simulations. This work was supported by NASA's Planetary Geology and Geophysics, and Origins of Solar Systems, Programs.

Correspondence and requests for materials should be addressed to W.R.W. (e-mail ward@boulder.swri.edu).

Preparing pure photon number states of the radiation field

B. T. H. Varcoe*, S. Brattke*†, M. Weidinger* & H. Walther*†

* *Max Planck Institute for Quantum Optics, Hans-Kopfermann Strasse 1, D-85748 Garching, Germany*
† *Sektion Physik der Universität München, 85748 Garching, Germany*

The quantum mechanical description of a radiation field is based on states that are characterized by the number of photons in a particular mode; the most basic quantum states are those with fixed photon number, usually referred to as number (or Fock) states. Although Fock states of vibrational motion can be observed readily in ion traps[1], number states of the radiation field are very fragile and difficult to produce and maintain. Single photons in multi-mode fields have been generated using the technique of photon pairs[2,3]. But in order to generate these states in a cavity, the mode in question must have minimal losses; moreover, additional sources of photon number fluctuations, such as the thermal field, must be eliminated. Here we observe the build-up of number states in a high-*Q* cavity, by investigating the interaction dynamics of a probe atom with the field. We employ a dynamical method of number state preparation that involves state reduction of highly excited atoms in a cavity, with a photon lifetime as high as 0.2 seconds. (This set-up is usually known as the one-atom maser or 'micromaser'.) Pure states containing up to two photons are measured unambiguously.

Number states are also generated as trapping states in a micromaser[4]; however, it is not easy to study the purity of the generated states in that case. There are many experiments where single-photon exchange between subsequent atoms has been investigated (see, for example, refs 5 and 6) leading to an entanglement between subsequent atoms and possibly the photons. In those experiments, no state reduction of the photon field in the cavity is performed, and so the field itself is in a more complicated state. During the review process of this Letter, work[7] describing quantum non-demolition measurement of states containing an average of one or zero photons in the mode related to a parity measurement of the photon state was published (see ref. 8 for comparison).

In the micromaser, highly excited atoms (usually called Rydberg atoms) interact with a single mode of a superconducting cavity with a quality factor as high as 3×10^{10}, leading to the above-mentioned photon lifetime. The steady-state field generated in the micromaser has been the subject of many detailed studies, such as the observation of sub-poissonian photon statistics[9]. The dynamics of the atom–field photon exchange has also been investigated in the