higher-resolution runs of the same impact (runs 24 and 34) produce a debris disk. The number of particles dictates the coarseness of the gravity potential, especially in a drawn-out bar of material. When the scale of particles is larger than the wavelength of the physical instability in the continuum those particles represent, collapse might ensue, whereas a more finely resolved system would be drawn apart by tidal shearing. This artefact of low-resolution was noticed in models[19] for the tidal disruption of comet Shoemaker–Levy 9.

Figure 2 shows scaled data from Table 1. General trends with increasing impact parameter are seen across runs with different total masses and γ values; there is also a dependence of the maximum yield on γ. For $b > 0.85$, orbiting material contains more than 3% iron by mass, with this fraction increasing progressively until a disk with a similar iron-mantle composition to that of the impactor is achieved for grazing impacts[6].

Models of protolunar disk accretion[20–22] find that a large moon forms at a characteristic distance of about $1.2a_{\mathrm{Roche}}$, with a mass that is a function of the initial disk mass and angular momentum:

$$M \approx 1.9L_{\mathrm{D}}/\sqrt{GM_{\oplus}a_{\mathrm{Roche}}} - 1.1M_{\mathrm{D}} - 1.9M_{\mathrm{esc}} \qquad (2)$$

where M_{esc} is the amount of escaping material during accretion. Using $M_{\mathrm{esc}} = 0.05M_{\mathrm{D}}$ (refs 21, 22), we estimate the moon mass that would result in the final column of Table 1; simulations for which this value is at least a lunar mass with $M_{\mathrm{Fe}}/M_{\mathrm{D}} \leqslant 0.03$ are shown in boldface.

The successful impacts involve an impactor-to-target mass ratio $\gamma \approx 0.1$–0.11, or an impactor with a mass of $\sim$6–6.5 $\times$ 10^{26} g—the mass of Mars. This class of impacts represents the least restrictive impact scenario, requiring little or no dynamical modification of the Earth–Moon system after the moon-forming impact. Problems associated with a period of extended terrestrial growth subsequent to the event are avoided; in addition, the smaller impact we advocate is more likely than the 2–3 times more massive impactors postulated by recent works[4,5]; this is because in collisional populations small objects are more common than large ones (the number of objects, dN, in a mass range dm is typically proportional to $\mathrm{d}N \propto m^{-q}\mathrm{d}m$, with[23] $q \approx 1.5$ to 1.8). It is, in retrospect, interesting that the Mars-mass impactor that now appears to be the most promising Moon-forming candidate is essentially that originally proposed[1], a decade before the first Moon-forming impact simulations. □

Received 23 April; accepted 12 June 2001.

1. Cameron, A. G. W. & Ward, W. R. The origin of the Moon. *Lunar Sci.* **7,** 120–122 (1976).
2. Hartmann, W. K. & Davis, D. R. Satellite-sized planetesimals and lunar origin. *Icarus* **24,** 504–515 (1975).
3. Cameron, A. G. W. The origin of the Moon and the single impact hypothesis V. *Icarus* **126,** 126–137 (1997).
4. Cameron, A. G. W. in *Origin of the Earth and Moon* (eds Canup, R. M. & Righter, K.) 133–144 (Univ. Arizona Press, Tucson, 2000).
5. Cameron, A. G. W. From interstellar gas to the Earth–Moon system. *Meteor. Planet. Sci.* **36,** 9–22 (2001).
6. Canup, R. M., Ward, W. R. & Cameron, A. G. W. A scaling law for satellite-forming impacts. *Icarus* **150,** 288–296 (2001).
7. Hood, L. L. & Zuber, M. T. in *Origin of the Earth and Moon* (eds Canup, R. M. & Righter, K.) 397–409 (Univ. Arizona Press, Tucson, 2000).
8. Benz, W., Slattery, W. L. & Cameron, A. G. W. The origin of the Moon and the single impact hypothesis I. *Icarus* **66,** 515–535 (1986).
9. Benz, W., Slattery, W. L. & Cameron, A. G. W. The origin of the Moon and the single impact hypothesis II. *Icarus* **71,** 30–45 (1987).
10. Benz, W., Cameron, A. G. W. & Melosh, H. J. The origin of the Moon and the single impact hypothesis III. *Icarus* **81,** 113–131 (1989).
11. Cameron, A. G. W. & Benz, W. The origin of the Moon and the single impact hypothesis IV. *Icarus* **92,** 204–216 (1991).
12. Melosh, H. J. & Kipp, M. E. Giant impact theory of the Moon's origin: first 3-D hydrocode results. *Lunar Sci.* **20,** 685–686 (1989).
13. Lucy, L. B. A numerical approach to the testing of the fission hypothesis. *Astron. J.* **82,** 1013–1024 (1977).
14. Stewart, G. R. in *Origin of the Earth and Moon* (eds Canup, R. M. & Righter, K.) 217–223 (Univ. Arizona Press, Tucson, 2000).
15. Tillotson, J. H. Metallic equations of state for hypervelocity impact. Report No. GA-3216, July 18 (General Atomic, San Diego, California, 1962).
16. Canup, R. M. & Asphaug, E. Outcomes of planet-scale collisions. *Lunar Sci.* [CD-ROM] **32,** (2001).
17. Melosh, H. J. & Pierazzo, E. Impact vapor plume expansion with realistic geometry and equation of state. *Lunar Sci.* **28,** 935 (1997).
18. Melosh, H. J. A new and improved equation of state for impact computations. *Lunar Sci.* **31,** 1903 (2000).
19. Asphaug, E. & Benz, W. Size, density, and structure of comet Shoemaker-Levy 9 inferred from the physics of tidal breakup. *Icarus* **121,** 225–248 (1996).
20. Ida, S., Canup, R. M. & Stewart, G. Formation of the Moon from an impact-generated disk. *Nature* **389,** 353–357 (1997).
21. Kokubo, E., Canup, R. M. & Ida, S. in *Origin of the Earth and Moon* (eds Canup, R. M. & Righter, K.) 145–163 (Univ. Arizona Press, Tucson, 2000).
22. Kokubo, E., Makino, J. & Ida, S. Evolution of a circumterrestrial disk and formation of a single Moon. *Icarus* **148,** 419–436 (2001).
23. Greenberg, R. in *Origin and Evolution of Planetary and Satellite Atmospheres* (eds Atreya, S. K., Pollack, J. B. & Matthews, M. S.) 137–164 (Univ. Arizona Press, Tucson, 1989).
24. Nelson, A., Benz, W., Adams, F. & Arnett, D. Dynamics of circumstellar disks. *Astrophys. J.* **502,** 342–371 (1998).
25. Hernquist, L. & Katz, N. TREESPH—A unification of SPH with the hierarchical tree method. *Astrophys. J. Suppl.* **70,** 419–446 (1989).
26. Benz, W. in *Proc. NATO Adv. Res. Workshop on Numerical Modelling of Nonlinear Stellar Pulsations* (ed. Buchler, J. R.) 1–54 (Kluwer Academic, Boston, 1990).

Acknowledgements

We wish to thank Southwest Research Institute's Internal Research program for its support of development efforts for the methods utilized here, D. Terrell for securing a portion of the computational time, and P. Tamblyn for aid with some of the analysis software. A review by A. Halliday and comments provided by W. Ward, C. Agnor, D. Korycansky and R. Mihran helped to improve the paper. This research was supported by the National Science Foundation and NASA.

Correspondence and requests for materials should be addressed to R.M.C. (e-mail: robin@boulder.swri.edu).

Sub-Planck structure in phase space and its relevance for quantum decoherence

Wojciech Hubert Zurek

Theory Division, T-6, MS B288, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

Heisenberg's principle[1] states that the product of uncertainties of position and momentum should be no less than the limit set by Planck's constant, $\hbar/2$. This is usually taken to imply that phase space structures associated with sub-Planck scales ($\ll \hbar$) do not exist, or at least that they do not matter. Here I show that this common assumption is false: non-local quantum superpositions (or 'Schrödinger's cat' states) that are confined to a phase space volume characterized by the classical action A, much larger than $\hbar$, develop spotty structure on the sub-Planck scale, $a = \hbar^2/A$. Structure saturates on this scale particularly quickly in quantum versions of classically chaotic systems—such as gases that are modelled by chaotic scattering of molecules—because their exponential sensitivity to perturbations[2] causes them to be driven into non-local 'cat' states. Most importantly, these sub-Planck scales are physically significant: a determines the sensitivity of a quantum system or environment to perturbations. Therefore, this scale controls the effectiveness of decoherence and the selection of preferred pointer states by the environment[3–8]. It will also be relevant in setting limits on the sensitivity of quantum meters.

One of the characteristic features of classical chaos is the evolution of the small-scale structure in phase-space probability distributions. As a consequence of the exponential sensitivity to initial conditions, an initially regular 'patch' in phase space with a

characteristic size Δ will, after a time t, develop structure on the scale

$$z \approx \Delta \exp(-\Lambda t) \quad (1)$$

where Λ is, in effect, the Lyapunov exponent[9,10]. This is a consequence of the exponential stretching and of the conservation of phase-space volume in reversible hamiltonian evolutions.

What happens with small-scale structure in quantum versions of classically chaotic systems? We shall investigate this question using the Wigner function $W(x, p)$—the closest quantum analogue of the classical phase-space distribution[11]:

$$W(x,p) = \frac{1}{2\pi\hbar}\int \exp(ipy/\hbar)\rho\left(x - \frac{y}{2}, x + \frac{y}{2}\right)\mathrm{d}y \quad (2)$$

Above, ρ is the density operator, x is position and p is momentum. When the quantum state is a pure $|\psi\rangle$, the integrand becomes $\exp(ipy/\hbar)\psi^*(x + y/2)\psi(x - y/2)$.

In a quantum system the smallest spatial structures in the wavefunction $\psi(x)$ will be set by the highest phase-space frequencies available. Given a finite energy, $\psi(x)$ and, hence, $W(x, p)$ cannot reach scales as small as the exponential squeezing of equation (1) would eventually imply. Thus, in a quantum system, there must be a scale below which structure should not appear. That some limit must exist was apparent some time ago: Berry and Balazs (who, as we shall see below, have an enviable record of correctly anticipating aspects of quantum chaos) have conjectured[12,13] that structure saturates on scales given by the Planck constant. If this were the case, W should be smooth on scales small compared with $2\pi\hbar$.

I shall show that copious structure appears in the Wigner function W on much smaller sub-Planck scales associated with the action of the order of

$$a \approx \hbar \times \frac{\hbar}{A} \quad (3)$$

and explain its origin. Most importantly, I shall demonstrate that a has physical consequences: it determines the sensitivity of the system (or of the environment) to decoherence. Above, A is the classical action of the system, given approximately by the product of the range of effective support of its state (the phase space envelope within which Wigner function is significantly different from zero) in position $\Delta x = L$ and momentum $\Delta p = P$:

$$A \approx L \times P \quad (4)$$

The values of L and P are in turn set by the available energy E, and by the form of the potential $V(x)$. That is, $P \leq \sqrt{(2mE)}$, $E - V(L) \geq 0$, and so on, which determine the effective support of the probability distribution in phase space. We shall eventually see that many of the calculations can be carried out using the state vector of the system in the appropriate representation. Nevertheless, intuitive understanding of the significance of the sub-Planck scale a is easiest to attain starting with a more comprehensive view of the phase-space image of quantum states afforded by the Wigner representation.

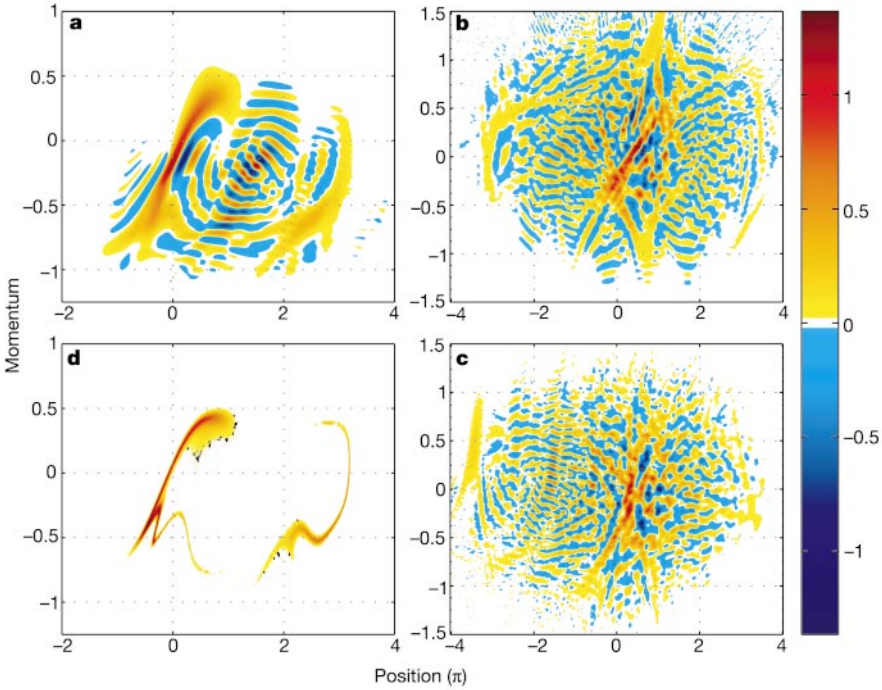

Figure 1 Snapshots of the quantum Wigner distribution and of the classical probability density in phase space of an evolving chaotic system. This system has the hamiltonian $H = (p^2/2m) - \kappa\cos(x - l\sin t) + ax^2/2$. For parameters $m = 1$, $\kappa = 0.36$, $l = 3$ and $a = 0.01$ the system exhibits chaos[24] with Lyapunov exponent $\Lambda \approx 0.2$. Initial probability density was given by the same gaussian in both the quantum cases (**a**–**c**; $\hbar = 0.16$) and the classical case (**d**). Structure on sub-Planck scales appears and saturates quickly. Note the contrast between the ever decreasing smallest dimensions of the probability density in the classical case (**d**) and of the Wigner distribution at the corresponding time (**a**). Exponential shrinking of the smallest scales makes it impossible to simulate accurately (for example, reversibly) classical evolution much beyond the time shown in **d**. By contrast, structures in quantum Wigner distribution saturate on scales $a \approx \hbar^2/A$ soon after $t \approx 20$ (which is of the order of the estimated t_h, equation (13a)). The scale of structure saturation of a state can be inferred from the volume of the domain containing it: smallest structures have dimensions $\delta_x = \hbar/P$ ($\delta_p = \hbar/L$), where P and L define the extent of the envelope of the effective support of the state ($P^2 \approx \langle p^2\rangle - \langle p\rangle^2$, $L^2 \approx \langle x^2\rangle - \langle x\rangle^2$). Action associated with the smallest structures will then be $a \approx \hbar^2/LP \approx \hbar^2/A$ in one spatial (two phase-space) dimensions. Generalization to the case of many dimensions is straightforward. One way to understand structure saturation is through the menu of wavevectors available in the system that have momenta restricted to $P \approx \sqrt{E}$, where E is the total energy. The complementary argument ($V(L) \approx E$, where $V(x)$ is the potential) can be made for the smallest scale of 'corrugation' of the Wigner distribution in p.

It is evident from Fig. 1 that, at least in chaotic systems, the structure on sub-Planck scales appears and saturates quickly. Let us now illustrate its origin. The smallest scales in phase space arise from interference. Consider, for instance, a familiar 'Schrödinger cat' coherent superposition of two minimum-uncertainty gaussians[6]:

$$W(x,p) = \frac{G(x+x_0,p)+G(x-x_0,p)}{2} + (\pi\hbar)^{-1}\exp\left(-\frac{p^2\xi^2}{\hbar^2}-\frac{x^2}{\xi^2}\right)\cos\left(p\frac{2x_0}{\hbar}\right) \quad (5)$$

where Wigner functions of the two gaussians 'east' and 'west' of the centre are

$$G(x\pm x_0,p-p_0) = (\pi\hbar)^{-1}\exp(-(x\pm x_0)^2/\xi^2-(p-p_0)^2\xi^2/\hbar^2) \quad (6)$$

The cosine term W_{WE} in equation (5) is a symptom of interference. Its ripples have a frequency proportional to the separation $L = 2x_0$ between the two peaks. When we define the frequency of the ripple pattern in momentum f_p through $\cos(Lp/\hbar) = \cos(f_p p)$, then $f_p = L/\hbar$. The ridges and valleys of such an interference pattern are always parallel to the line of sight between the two gaussian peaks. Thus, standing on top of one gaussian peak, one could still see the other peak through the valleys (and between the ridges) of the interference term, even though its envelope is a factor of two higher than either gaussian.

Coherent states form an overcomplete set. We can therefore express an arbitrary pure state as a superposition of coherent states. The smallest interference structures in such an expansion arise from pairs of coherent states separated by the whole range available to the system in phase space. Therefore, we expect smallest scales with

$$f_x = P/\hbar, \quad \text{or} \quad \delta_x = \hbar/P \quad (7)$$

$$f_p = L/\hbar, \quad \text{or} \quad \delta_p = \hbar/L \quad (8)$$

As an example, consider the compass state, a Schrödinger-cat-like superposition of four minimum-uncertainty gaussians, one pair located north and south of the common centre, the other east and west (see Fig. 2). The Wigner distribution is quadratic in the wavefunction. Therefore, W of any superposition can be reconstructed from the contributions corresponding to, at most, pairs of states. The structure of the Wigner distribution of a superposition of a pair of gaussians (equation (5); note that this is reflected in the patterns of the sides of the square in Fig. 2) can be then used to infer W of the compass state:

$$W_{NWSE} = G_N + G_W + G_S + G_E)/4 \quad (9a)$$

$$+(W_{NW} + W_{WS} + W_{SE} + W_{EN})/2 \quad (9b)$$

$$+(W_{NS} + W_{EW})/2 \quad (9c)$$

in an obvious 'geographical' notation. The last line of equation (9),

$$W_{NS} + W_{EW} = (\pi\hbar)^{-1}\exp\left(\frac{p^2\xi^2}{\hbar^2}-\frac{x^2}{\xi^2}\right)\left(\cos\frac{pL}{\hbar}+\cos\frac{xP}{\hbar}\right) \quad (10)$$

is of greatest interest. Above, I have assumed that the shapes of all the gaussians are identical, so that the exponential envelope in equation (9c) is common to both terms.

The resulting interference term (the centre of Fig. 2) has a chessboard pattern. The size of the single 'tile' can be obtained from zeros of the oscillatory factor of equation (10):

$$\cos(pL/\hbar)+\cos(xP/\hbar) = 2\cos\frac{Px+Lp}{2\hbar}\cos\frac{Px-Lp}{2\hbar} \quad (11)$$

Individual squares, four per tile, are defined by zeros that occur when $x = \pm\pi\hbar/2L$, $p = \pm\pi\hbar/2P$. The fundamental periodic tile has an area of

$$a = \frac{2\pi\hbar}{L}\times\frac{2\pi\hbar}{P} = (2\pi\hbar)^2/A \quad (12)$$

which, with the identification of $A = LP$, yields action associated with the smallest scales present in quantum phase space.

The above calculation shows by construction that a quantum state spread over a phase space of volume $A = LP$ can accommodate Wigner distribution structures as small as a. Such states arise naturally. Evolution will force almost any system (with a notable exception of a harmonic oscillator) into a 'Schrödinger cat' state[2]—a coherent non-local superposition that, after the time it takes the wavefunction to spread over the phase-space volume A, inevitably develops an interference pattern on the scale given by equations (7–12).

It is not necessary to invoke chaos: after sufficient time, even an integrable nonlinear system may spread coherently throughout the available phase space, and, consequently, saturate small-scale

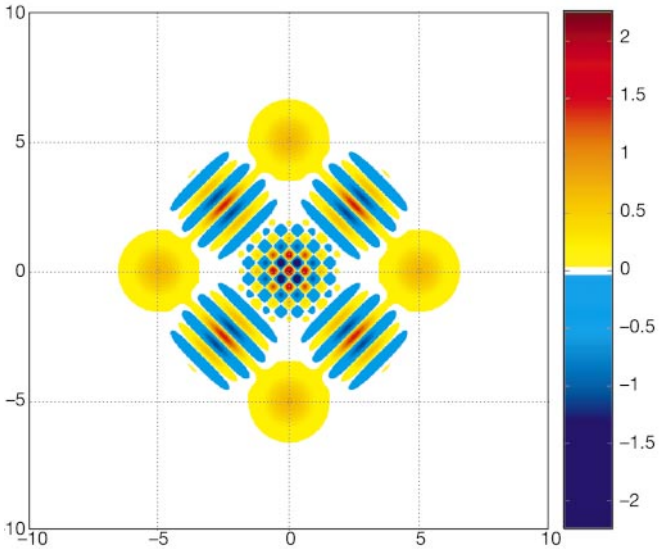

Figure 2 The compass state, equation (9). This is a superposition of minimum-uncertainty states ($|N\rangle$, $|W\rangle$, $|S\rangle$, $|E\rangle$) placed 'north',…, 'east', of the common centre. Units are selected so that the gaussians used are round and $\hbar = 1$. The form of the central interference pattern in W_{NWSE} can be inferred from the familiar structure of the superposition of two gaussians, equation (5), also apparent on the sides of the square above: the interference terms corresponding the north–south and east–west pairs superpose, creating a chessboard pattern. The area of individual tiles is set by the dimensions of the NWSE cross and corresponds to the classical action, A, of phase-space area of the effective support of its envelope, equation (12). The appearance of this interference pattern explains the origin of the structure saturation seen in Figs 1 and 3: a system that can be effectively confined to phase-space volume A cannot develop structure on scales smaller than a. Sensitivity of the compass state to perturbations is controlled by a. This is readily seen in the 'sparse limit'; that is, when $L \gg \xi$, $P \gg \hbar/\xi$. For simplicity, consider a shift $\boldsymbol{\delta} = \delta_x\hat{x} + \delta_p\hat{p}$ small compared with the sizes of the gaussians, $\delta_x \ll \xi$, $\delta_p \ll \hbar/\xi$. The square of the overlap of the original and displaced states $|\langle\psi_{NWSE}|\psi^{\delta}_{NWSE}\rangle|^2 = 2\pi\hbar\int W_{NWSE}(x,p)W_{NWSE}(x+\delta_x,p+\delta_p)\mathrm{d}x\mathrm{d}p$ is approximately $|\langle N|N^{\delta}\rangle + \ldots + \langle E|E^{\delta}\rangle|^2 \approx (\cos\delta_x P/2\hbar + \cos\delta_p L/2\hbar)^2/4$. To get this simple result we have ignored both the additive corrections (such as $\langle N|E^{\delta}\rangle$) that are small in the sparse limit, and a multiplicative correction $\sim|\langle E|E^{\delta}\rangle|^2$ that is very close to unity in the limit of small displacements $|\delta| \ll \sqrt{\hbar}$. The striking kinship of this form of the overlap with the interference term, equations (10) and (11), is no accident: the magnitudes of the shifts that produce orthogonality are $\delta_x = 2\pi\hbar/P$, $\delta_p = 2\pi\hbar/L$. Thus, displacing the states by the size of the square in the central interference pattern defined by equations (11) and (12) suffices to cause decoherence, if the environment starts in the NWSE state.

structure. It is just that when the evolution is chaotic, such small scales will be attained faster, on a timescale given by[14]:

$$t_\hbar = \Lambda^{-1} \ln \Delta p \chi/\hbar \tag{13a}$$

Above, Δp is the characteristic spread of the initial smooth probability distribution. χ characterizes the scale on which potential is significantly nonlinear. It is typically given by $\chi \approx \sqrt{(V'/V''')}$. Similar estimates are obtained from the formula

$$t_r = \Lambda^{-1} \ln A/\hbar \tag{13b}$$

deduced some time ago[12,15].

There is one more suggestive way to express saturation scale a: the number of distinct (orthogonal) states that can fit within phase space of volume A is $N = A/(2\pi\hbar)$. The structure we are discussing appears therefore on the scale $a \approx \hbar/N$. Here N is, in effect, the dimension of the available Hilbert space.

In accord with Heisenberg's principle, a quantum system cannot be localized to a sub-Planck volume in phase space. Hence, one might be tempted to dismiss sub-Planck scales as unphysical even if they appear in the Wigner distribution. In particular, if a state cannot be confined, by measurements, to a volume less than $\sim\hbar$, then one might expect that it will not be noticeably perturbed by displacements much smaller than $\sqrt{\hbar}$. I now show that this expectation is false, and that $a \ll \hbar$ plays a decisive role in determining the sensitivity of quantum systems (or of quantum environments) to perturbations: phase space displacements $\delta \approx \sqrt{a} \ll \sqrt{\hbar}$ shift a state with a dominant structure on scale a enough to make it orthogonal to—that is, distinguishable from—the unshifted original. Sensitivity to perturbations in turn sets the limit on the efficiency of decoherence.

There are two complementary aspects to this connection between a and decoherence. When a system with a sub-Planck scale in W is coupled to the environment, decoherence can be thought of as monitoring, by the environment, of some of its observables[3–8]. Its effect—suppression of quantum coherence—can be traced to Heisenberg's principle: the observable that is complementary to the one monitored by the environment becomes less determined, in effect smearing Wigner distribution along the corresponding phase-space direction[8]. When this smearing obliterates interference structures on scale a, coherence on the large scales corresponding to $A \approx \hbar^2/a$ will be suppressed, and the Schrödinger cat state will have lost its quantum nonlocality[2,14,16].

We turn now to a complementary aspect of the same story. It involves the situation when the state of the environment—the cause of decoherence—is a 'Schrödinger cat' spread over the phase-space region A. The environment entangles with the system, acting as a 'monitoring' apparatus[3–8]. The sensitivity of the environment to perturbations is therefore of the essence. We shall test the sensitivity of such a Schrödinger cat environment by allowing it to interact with a Schrödinger cat system, assumed to be initially in a superposition of two perfect pointer states $\{|+\rangle, |-\rangle\}$. Pointer states, taken one at a time, perturb the state of the environment but do not entangle with it. However, each pointer state perturbs the environment differently[3]. Therefore, a system prepared in a general superposition state will leave a pointer-state-dependent imprint on the environment, and hence entangle with it. We shall show that the displacement $\delta \approx \sqrt{a}$ sets the size of the smallest perturbations distinguished by the environment, which in turn controls its ability to decohere the system. Thus, when $|+\rangle$ and $|-\rangle$ shift the state of the environment differently, the off-diagonal terms of the density matrix of the system will be suppressed by a factor $|\langle\epsilon_+|\epsilon_-\rangle|$, where

$$|\langle\epsilon_+|\epsilon_-\rangle|^2 = 2\pi\hbar \int W_+ W_- \mathrm{d}x\mathrm{d}p \tag{14}$$

Above, kets $|\epsilon_\pm\rangle$ and their Wigner distributions ($W_\pm$) represent states that evolve from the original state of the environment $|\epsilon\rangle$ through the interaction (induced, say, by a system–environment interaction hamiltonian such as $\sim(|+\rangle\langle+| - |-\rangle\langle-|)(\hbar/i)\partial_x$ with the states $|\pm\rangle$ of the system, respectively.

Let us now demonstrate how the behaviour of the magnitude of the overlap

$$|\langle\epsilon_+|\epsilon_-\rangle| = \left| \int \epsilon^*(x) e^{i\delta_p x/\hbar} \epsilon(x+\delta_x) \mathrm{d}x \right| \tag{15}$$

is controlled by a. Above $\boldsymbol{\delta} = (\delta_x, \delta_p)$ is the net displacement

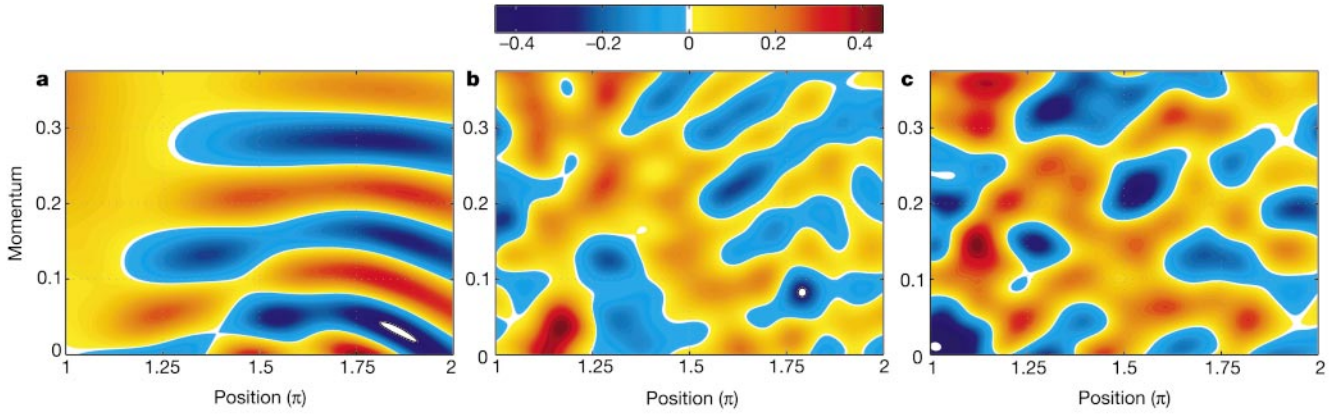

Figure 3 Snapshots of area $2\pi\hbar$ extracted from Fig. 1a–c. The smallest scales saturate on sub-Planck scale a, so that—as can be estimated from Fig. 1—the individual patches appear on the scale consistent with Figs 1 and 2 and equation (3). Such small phase-space substructure is physically significant. Displacement of the order of $\sqrt{a}$ (the size of a typical patch) suffices to make the perturbed state roughly orthogonal to its old (unperturbed) self. Thus, a sets the limit on the sensitivity of the state to perturbations, and is therefore relevant for decoherence. Note that the overlap of the original and displaced states will behave differently when the interference pattern is irregular (as it is here) rather than essentially periodic (as was the case in the compass state of Fig. 2). This can be understood by expressing the state as a superposition $|\psi\rangle = \Sigma_k^n \alpha_k |k\rangle$ of identical minimum-uncertainty gaussians $|k\rangle = G(x - x_k, p - p_k)$, each centred on x_k, p_k (see equation (6)). In the sparse limit (that is, when the gaussians in the superposition of $|\psi\rangle$ do not overlap significantly) the overlap of the original and displaced states is approximately $|\langle\psi|\psi^\delta\rangle| \approx |\Sigma_k^n |\alpha_k|^2 e^{i(\delta_p x_k + \delta_x p_k)/\hbar}| = |\Sigma_k^n w_k e^{i\phi_k(\delta)}|$, where $w_k = |\alpha_k|^2$, $\Sigma_k^n w_k = 1$ are the weights (probabilities) of finding the system in different states $|k\rangle$. It is obvious that, when many sparsely distributed gaussians participate in $|\psi\rangle$, so that $w_k \approx 1/n$, the size of the overlap is given by the distance covered by an eventually random walk in a complex plane where individual steps have magnitude w_k and directions determined by phases ϕ_k. Hence, as ϕ_k become random, the overlap will rapidly decrease from unity to about $1/\sqrt{n}$, and will probably remain small. (Such sums are a standard way of recovering a gaussian probability distribution, and have already been studied in the context of decoherence some time ago[4].)

between $|\epsilon_+\rangle$ and $|\epsilon_-\rangle$ corresponding to $W_+ = W(x+\delta_x^+, p+\delta_p^+)$ and $W_- = W(x+\delta_x^-, p+\delta_p^-)$, respectively. The displacement $\boldsymbol{\delta}$ is the difference between the shifts caused by $|+\rangle$ and $|-\rangle$, $\boldsymbol{\delta} = \boldsymbol{\delta}^+ - \boldsymbol{\delta}^- \neq 0$. The heuristic argument for the size of displacement that causes orthogonality $\langle\epsilon_+|\epsilon_-\rangle \approx 0$, and hence decoherence, is easiest to follow when phrased in terms of Wigner functions. Suppose W_+ and W_- in equation (14) have a small-scale structure with patches of alternating sign, as seen in Figs 1–3. The integral of their product can reach a maximum value of unity only when W_+ and W_- are not shifted with respect to each other. For shifts small compared with the typical size of the patches the integrand will still be positive almost everywhere, but as the magnitude of the shift increases, the right-hand side of equation (14) will become small compared with unity. If the interference pattern in the Wigner distribution is periodic (as is the case in Fig. 2) the oscillation will be also periodic with a period related to the size of the fundamental 'tile'. When, however, patches are random (as for the typical case illustrated in Fig. 3), the overlap—having decayed after a displacement $|\boldsymbol{\delta}| \approx \sqrt{a}$—will not significantly recur. A more formal version of this heuristic argument is put forward and backed up by numerical simulations elsewhere (Z. Karkuszewski *et al.*, manuscript in preparation). A simple back-of-the-envelope calculation valid for a generalization of compass states (that is, when the state of the environment can be approximated by a 'sparse' collection of identically shaped minimum-uncertainty gaussians) is given in Fig. 3.

This intuitive picture, already supported by the example of Fig. 2, can be further confirmed by a general yet straightforward calculation based on equation (15). To simplify notation we consider the case when $\boldsymbol{\delta} = (0, \delta_p)$, that is, when the net shift is aligned with one of the axes. (This assumption can be made with no loss of generality, as the axes in phase space can be rotated to align one of them with an arbitrary $\boldsymbol{\delta}$.) In that case:

$$\langle\epsilon_+|\epsilon_-\rangle = \int |\epsilon(x)|^2 e^{i\delta_p x/\hbar} \mathrm{d}x \qquad (16)$$

This is a simple, general, and compelling result: suppression of the off-diagonal terms in the density matrix of the system is given by a Fourier transform of the probability distribution in the environment along the direction perpendicular to the net relative shift induced by its coupling with the system. This leads back to sub-Planck scales: for small displacements the exponent in the integrand can be expanded. This yields

$$|\langle\epsilon_+|\epsilon_-\rangle|^2 \approx 1 - \delta_p^2(\langle x^2\rangle - \langle x\rangle^2)/\hbar^2 \qquad (17a)$$

The spatial extent of the wavefunction is naturally defined as $L = \sqrt{\langle x^2\rangle - \langle x\rangle^2}$, where the averages of the observable x are over the state of the environment. The estimate of the shift δ_p leading to orthogonality is then

$$\delta_p \approx \hbar/\sqrt{\langle x^2\rangle - \langle x\rangle^2} \qquad (17b)$$

in accord with the estimate of the dimensions of the sub-Planck structures, equation (8). The same simple formula holds when the environment is initially in a mixed state (see Methods).

Obvious generalizations of equations (16) and (17) are valid for arbitrary displacements, and for the case of many states of the system. This treatment can be also extended to interactions between the system and the environment that cannot be represented as simple shifts, although calculations become more complicated. The role played by the spread L of the state of the environment in the behaviour of the overlap (and hence in decoherence) is a direct consequence of the properties of the Fourier transform. Detailed analysis of other consequences of equation (16) is beyond the scope of this paper. Using elementary properties of the Fourier transform, readers can nevertheless confirm that the overlap decreases to near zero around δ_p (equations (8) and (17)), and that, for typical initial states of the environment, it remains much less than unity. 'Revivals' of the overlap we have seen in the case of the compass state (Fig. 2) are now easily understood and dismissed as an exception. Equation (16) shows that they can happen only when the initial probability distribution of the environment in the direction perpendicular to displacement is localized to a few peaks, so that $|\epsilon(x)|^2$, the spectrum of the displacement-dependent overlap, is essentially discrete.

To conclude, I have shown that Wigner functions can, and generally will, develop phase space structures on scales as small as, but not generally smaller than, $a \sim \hbar^2/A$, equation (3). This is surprising, as sub-Planck scales are often regarded as unphysical. I have shown that they have remarkable physical implications: displacements given by $\delta \approx \hbar/\sqrt{A}$ suffice to induce orthogonality. They are a factor $\sim\sqrt{N} = \sqrt{(A/2\pi\hbar)}$ smaller than the displacement $\delta \approx \sqrt{\hbar}$ dictated by the Planck constant and needed to move a typical minimum-uncertainty gaussian in a random direction enough to noticeably reduce its overlap with its (old) self.

Nonlinear, and, especially, chaotic dynamics lead to states that quickly spread over much of the available phase space[2,12–15]. Hence, one would expect that environments with unstable dynamics would be much more efficient decoherers, as they are constantly evolving into delocalized states. Yet standard models of decoherence[17–19] use harmonic oscillators. They make up for this inefficiency by using many (infinity) of them, so that each becomes slightly displaced by the interaction with the system.

On a more mundane level, these results allow one to anticipate the structure of the mesh required to simulate evolution of a quantum system in phase space. They are related to the quantization of discrete chaotic maps on a torus[20], which turn out to require a mesh similar to the one given by equations (7) and (8). This is in contrast to simulations of classical systems, which are doomed by equation (1), as it implies resolution exponentially increasing with time. Most importantly, a controls the sensitivity of quantum states to perturbations, with obvious implications for decoherence outlined above. Moreover, sensitivity to perturbations will set limits on the 'hypersensitivity' of quantum chaotic systems[21], and help one to understand the enhanced capacity of quantum chaotic systems for entanglement[22].

The sensitivity of quantum systems in highly delocalized states may not just be a cause of accelerated decoherence (and hence an impediment to truly quantum applications) but in certain settings it may be beneficial. This is not as far-fetched as it may seem at first. After all, a detector in a compass-like state of Fig. 2 would be sensitive to perturbations $\sim\hbar/\sqrt{A}$ that are minute compared with the 'standard quantum limit'[23]. And, when the weak force to be detected perturbs its state enough to make it orthogonal to the initial state, a record that is in principle distinguishable has been made. The sensitivity of quantum meters would be then limited by a, in accord with equations (3), (16) and (17). Thus, it seems possible that the Schrödinger cat may eventually follow the path of other paradoxical quantum *gedanken* experiments that have found—after dusting off the classical preconceptions and shedding the aura of the paradox—far more useful employment in potential applications of quantum physics. □

Methods

Decoherence happens to a quantum system $\mathcal{S}$ as a consequence of a measurement-like interaction with the environment $\mathcal{E}$, which entangles their states:

$$|s\rangle|\epsilon\rangle = (\alpha|+\rangle + \beta|-\rangle)|\epsilon\rangle \rightarrow \alpha|+\rangle|\epsilon_+\rangle + \beta|-\rangle)|\epsilon_-\rangle = |\Phi_{S\mathcal{E}}\rangle$$

Above, we have assumed a system with a two-dimensional Hilbert space spanned by the $\{|+\rangle, |-\rangle\}$ orthonormal basis. The two conditional states of the environment $|\epsilon_\pm\rangle = U_\pm|\epsilon\rangle$ evolve under the unitary transformations U_+ (U_-) induced by the system in the state $|+\rangle$ ($|-\rangle$) respectively. When the effect of the interaction is simply a displacement in phase space, then $U_\pm = D_\pm = \exp i(\delta_x^\pm p + \delta_p^\pm x)/\hbar$, where D is the displacement operator, and $\boldsymbol{\delta}^\pm = (\delta_x^\pm, \delta_p^\pm)$ are the resulting shifts. Such a displacement could be induced by a hamiltonian of interaction $H_{S\mathcal{E}} = g(|+\rangle\langle+| - |-\rangle\langle-|)(\hbar/i)\partial_x$, where g is a

coupling constant. More general conditional evolutions of the environment $\mathcal{E}$ can be of course considered.

Following entanglement, the state of the system alone is described by the reduced density matrix obtained from $|\Phi_{S\mathcal{E}}\rangle$ by a trace over the environment:

$$\rho_S = \mathrm{Tr}_{\mathcal{E}}|\Phi_{S\mathcal{E}}\rangle\langle\Phi_{S\mathcal{E}}| = |\alpha|^2|+\rangle\langle+| + z\alpha\beta^*|+\rangle\langle-| + z^*\alpha\beta^*|-\rangle\langle+| + |\beta|^2|-\rangle\langle-|$$

Disappearance of the off-diagonal terms signifies perfect decoherence. In the $\{|+\rangle, |-\rangle\}$ basis this is guaranteed when the overlap $z = \langle\epsilon_+|\epsilon_-\rangle = \mathrm{Tr}|\epsilon_-\rangle\langle\epsilon_+|$ disappears. It is therefore natural to measure effectiveness of decoherence by the magnitude of the overlap of the two conditional states of the environment, which in turn determines the degree of suppression of the off-diagonal terms in ρ_S.

This two-paragraph 'crash course' is no substitute for a more complete discussion of decoherence[6–8]. We have swept a number of issues under the rug. Foremost among them is einselection—the emergence, in the course of decoherence, of the preferred set of pointer states that habitually appear on the diagonal of ρ_S essentially independently of the initial states of either $\mathcal{E}$ or $\mathcal{S}$. Stability of these pointer states (rather than the diagonality of the density matrix in some basis) is the key to the role played by decoherence in the transition from quantum to classical[2–8,18].

Another important subject avoided here is the likely situation when the state of the environment is represented by a mixture $\rho_{\mathcal{E}}$. Detailed discussion of this case (treated extensively before[6–8,17–19], although not from the point of view of the sub-Planck structures) is beyond the scope of this paper, but the basic conclusion is easy to state: the estimated magnitude of the smallest displacement leading to orthogonality is still given by equation (17). That is, it is still related to the smallest scales compatible with the classical action A associated with $\rho_{\mathcal{E}}$. Off-diagonal terms of ρ_S are suppressed by $z = \mathrm{Tr}U_-\rho_{\mathcal{E}}U_+^\dagger$, which is the relevant generalization of equation (16). This expression can be expanded for simple displacements $U_\pm = D_\pm$ in the limit of small shifts to recover equation (17), with the only difference arising from the fact that now the mixture $\rho_{\mathcal{E}}$ must be used to obtain the averages ($\langle x^2\rangle = \mathrm{Tr}x^2\rho_{\mathcal{E}}$, etc.).

Note that the size of the structures in phase space depends on the representation of the quantum state. For instance, one does not expect scale a of equation (3) in the Husimi function, that is, in the Wigner function smoothed on a Planck scale. This is no surprise and does not affect our conclusions concerning decoherence, as only the Wigner function yields scalar product through the simple equation (14).

Note that the discussion throughout the paper is set in one spatial dimension. Generalization to d dimensions is as conceptually straightforward as it is notationally cumbersome. The structure saturates in volumes of a^d, etc. This has little effect on decoherence, as it depends on displacements that yield orthogonality, and these are still $\delta \approx \sqrt{a}$.

Received 25 September 2000; accepted 25 June 2001.

1. Heisenberg, W. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Z. Phys.* **43,** 172–198 (1927); The physical content of quantum kinematics and mechanics (Engl. Trans.) in *Quantum Theory and Measurement* (eds Wheeler, J. A. & Zurek, W. H.) (Princeton Univ. Press, Princeton, 1983).
2. Zurek, W. H. Decoherence, chaos, quantum-classical correspondence, and the algorithmic arrow of time. *Phys. Script.* **T76,** 186–198 (1998).
3. Zurek, W. H. Pointer basis of a quantum apparatus: Into what mixture does the wavepacket collapse? *Phys. Rev. D* **24,** 1516–1524 (1981).
4. Zurek, W. H. Environment-induced superselection rules. *Phys. Rev. D* **26,** 1862–1880 (1982).
5. Joos, E. & Zeh, H. D. The emergence of classical properties through the interaction with the environment. *Z. Phys. B* **59,** 229 (1985).
6. Zurek, W. H. Decoherence and the transition from quantum to classical. *Phys. Today* **44,** 36–46 (1991).
7. Giulini, D., Joos, E., Kiefer, C., Kupsch, J., Stamatescu, L.-O. & Zeh, H. D. *Decoherence and the Appearance of a Classical World in Quantum Theory* (Springer, Berlin, 1996).
8. Zurek, W. H. Decoherence, einselection, and the quantum origin of the classical. *Rev. Mod. Phys.* (in the press); also as preprint (quant-ph 010527) at ⟨http://xxx.lanl.gov⟩ (2001).
9. Haake, F. *Quantum Signatures of Chaos* (Springer, Berlin, 1991).
10. Casati, G. & Chrikov, B. *Quantum Chaos* (Cambridge Univ. Press, Cambridge, 1995).
11. Hillery, M., O'Connell, R. F., Scully, M. O. & Wigner, E. P. Distribution functions in physics: Fundamentals. *Phys. Rep.* **106,** 121–167 (1984).
12. Berry, M. V. & Balazs, N. L. Evolution of semiclassical quantum states in phase space. *J. Phys. A* **12,** 625–642 (1979).
13. Korsch, H. J. & Berry, M. V. Evolution of Wigner's phase-space density under a nonintegrable quantum map. *Physica D* **3,** 627–636 (1981).
14. Zurek, W. H. & Paz, J. P. Decoherence, chaos, and the Second Law. *Phys. Rev. Lett.* **72,** 2508–2511 (1994).
15. Berman, G. P. & Zaslavsky, G. M. Condition of stochasticity in quantum non-linear systems. *Physica* (Amsterdam) **91A,** 450 (1978).
16. Habib, S., Shizume, K. & Zurek, W. H. Decoherence, chaos, and the correspondence principle. *Phys. Rev. Lett.* **80,** 4361 (1998).
17. Caldeira, A. O. & Leggett, A. J. Path-integral approach to quantum Brownian motion. *Physica* **121A,** 587–616 (1983).
18. Paz, J. P. & Zurek, W. H. in *Les Houches Lectures Session LXXII* (eds Kaiser, R., Westbrook, C. and David, F.) 533–614 (Springer, Berlin, 2001).
19. Braun, D., Haake, F. & Strunz, W. A. Universality of decoherence. *Phys. Rev. Lett.* **86,** 2913–2917 (2001).
20. Hannay, J. H. & Berry, M. V. Quantization of linear maps on a torus—Fresnel diffraction by a periodic grating. *Physica* **1D,** 267–290 (1980).
21. Caves, C. in *Physical Origins of Time Asymmetry* (eds Halliwell, J. J., Pérez-Mercader, J. & Zurek, W. H.) 47–77 (Cambridge Univ. Press, Cambridge, 1993).
22. Miller, P. A. & Sarkar, S. Signatures of chaos in the entanglement of two coupled quantum kicked tops. *Phys. Rev. E* **60,** 1542 (1999).
23. Braginsky, V. B. & Khalili, F. Y. Quantum nondemolition measurements: the route from toys to tools. *Rev. Mod. Phys.* **95,** 703–711 (1996).
24. Karkuszewski, Z., Zakrzewski, J. & Zurek, W. H. Breakdown of correspondence in chaotic systems: Ehrenfest versus localization times. Preprint quant-ph/0010011 at ⟨http://xxx.lanl.gov⟩ (2000).

Acknowledgements

This research was supported in part by the National Security Agency. I thank A. Albrecht, N. Balazs, C. Jarzynski, Z. Karkuszewski and J. P. Paz for useful chaotic conversations.

Correspondence and requests for materials should be addressed to W.Z. (e-mail: whz@LANL.gov).

Crystalline ion beams

T. Schätz, U. Schramm & D. Habs

Sektion Physik, LMU München, D-85748 Garching, Germany

By freezing out the motion between particles in a high-energy storage ring, it should be possible[1–4] to create threads of ions, offering research opportunities beyond the realm of standard accelerator physics. The usual heating due to intra-beam collisions should completely vanish, giving rise to a state of unprecedented brilliance. Despite a continuous improvement of beam cooling techniques, such as electron cooling and laser cooling, the ultimate goal[5] of beam crystallization has not yet been reached in high-energy storage rings. Electron-cooled dilute beams of highly charged ions show liquid-like order[6,7] with unique applications[8]. An experiment[5] using laser cooling[9,10] suggested a reduction of intra-beam heating, although the results were ambiguous. Here we demonstrate the crystallization of laser-cooled Mg^+ beams circulating in the radiofrequency quadrupole storage ring PALLAS[11,12] at a velocity of 2,800 m s^{-1}, which corresponds to a beam energy of 1 eV. A sudden collapse of the transverse beam size and the low longitudinal velocity spread clearly indicate the phase transition. The continuous ring-shaped crystalline beam shows exceptional stability, surviving for more than 3,000 revolutions without cooling.

In ion storage rings[13], gaseous ion beams are heated through scattering within the beam. In combination with the varying focusing and bending elements, this mechanism couples part of the beam energy into the random ion motion. Cooling increases the phase-space density of the beam, and further amplifies the scattering rate[14]. In high-energy rings, this vicious circle can only be overcome by increasing the number of focusing sections[15], or, in practice, by choosing a sufficiently low density of stored ions[5]. Recently, space-charge-limited densities have been reached[10], but no beam crystallization has been observed so far, in striking contrast to the routine generation of elongated ion crystals at rest in linear[16,17] and ring traps[12,18]. As an illustration, in Fig. 1 we present images of ion crystals at rest[12], gained with our storage ring PALLAS (Paul laser cooling acceleration system), described below. Ions remain ordered because their mutual Coulomb repulsion overcomes their mean kinetic energy. The overall Coulomb repulsion is compensated by an external parabolic trapping potential ψ. The formation of the crystalline structure is well understood[4,17–20]. It develops from a linear chain of ions over a zig-zag band to three-dimensional helices when the linear ion density[19] λ is increased, either by adding more ions or by reducing the confining potential.

The difference between the behaviour in storage rings and traps might be caused by the predicted excitation of ion crystals passing through the periodic bending and focusing sections of a ring[20–22]. To bridge this gap and experimentally elucidate the conditions necessary for obtaining crystalline beams in high-energy rings, we