Microscopic Origins of Macroscopic Behavior

by JOEL LEBOWITZ (Rutgers)

This article is mostly based on a talk I gave at the March 2021 meeting (virtual) of the American Physical Society on the occasion of receiving the Dannie Heineman prize for Mathematical Physics from the American Institute of Physics and the American Physical Society. I am greatly indebted to many colleagues for the results leading to this award. To name them all would take up all the space allotted to this article. (I have had more than 200 collaborators so far), I will therefore mention just a few: Michael Aizenman, Bernard Derrida, Shelly Goldstein, Elliott Lieb, Oliver Penrose, Errico Presutti, Gene Speer and Herbert Spohn. I am grateful to all of my collaborators, listed and unlisted.

I would also like to acknowledge here long time support form the AFOSR and the NSF.

Introduction

Let me begin by quoting Freeman Dyson, an earlier recipient of this prize, about his definition of mathematical physics:

“Mathematical physics is the discipline of people who try to reach a deep understanding of physical phenomena by following the rigorous style and method of mathematics.”

— Freeman Dyson, From Eros to Gaia, p 164-165

Freeman was a good friend and this talk is dedicated to his memory, as well as to the memory of my teachers Melba Phillips and Peter Bergmann.

My own area of mathematical physics is statistical mechanics, which is concerned with the microscopic origin of macroscopic behavior. Since our mathematical abilities for dealing with strongly interacting many particle systems are quite limited it is fortunate that many striking features of macroscopic systems can be obtained from simplified microscopic models.

We therefore often take as our lowest level starting point an idealized description of atoms. As put by Feynman [6]:

“If in some cataclysm all of scientific knowledge was to be destroyed, . . ., what statement would contain the most information in the fewest words? I believe it is . . .that all things are made of atoms — little particles that move around in perpetual motion, attracting each other when they are a little distance apart, but repelling upon being squeezed into one another.”
Microscopic Origins of Macroscopic Behavior

Unfortunately even such idealized systems are too difficult to deal with in any detail. In fact many details would just be confusing. I will therefore focus on describing, in a qualitative way, the microscopic origin of those behaviors which are (almost) always observed in isolated macroscopic systems both in equilibrium and out of it. I will relate this to the fact that this behavior is typical for systems represented by the usual Gibbs measures or those derived from them. These take small phase space volume to indicate small probability. I will not try to justify this here.

This means that for equilibrium macroscopic systems these behaviors occur for an overwhelming majority of the microstates in the micro-canonical ensemble, i.e. they are typical. In fact, the fraction of systems with noticeable macroscopic deviations from the average behavior, computed in such an ensemble, is exponentially small in the number of degrees of freedom of the system: the functions on the phase space which correspond to such typical behavior will be described later.

An analogous statement holds for the (exponentially small) subsets of the micro-canonical ensemble which describe systems in nonequilibrium macrostates (to be defined below). It thus includes the time asymmetric approach to equilibrium, encoded in the second law and observed in individual macroscopic systems. Once one accepts the applicability of these measures to physical systems the observed behavior does not require explanations based on ergodicity, time averaging, or subjective information theory.

This property of typicality of behavior predicted by the measures used to represent macroscopic systems is true both classically and quantum mechanically. It explains why these ensembles can be used to predict the observed behavior of individual macroscopic systems and not just some average behavior.

I will begin with classical systems where the situation is easier to visualize.

Classical Systems

In classical mechanics, the microstate of a system of $N$ particles confined to a region $V$ in $\mathbb{R}^d$ is a point $X$ in the $2dN$-dimensional phase space,

$$X = (\vec{r}_1, \vec{p}_1, \ldots, \vec{r}_N, \vec{p}_N), \quad \vec{r}_i \in V \subset \mathbb{R}^d, \quad \vec{p}_i \in \mathbb{R}^d$$ (1)

Its time evolution is given by a Hamiltonian $H(X)$ which conserves energy, so $X(t)$ will be confined to $\Gamma_E$, a thin shell surrounding the energy surface $H(X) = E$. A macroscopic system is one with “very large” $N$, say $N \gtrsim 10^{20}$.

Macrostates

To describe the macroscopic state of such a system of $N$ particles in a box $V$, we make use of a much cruder description than that provided by the microstate $X$. We shall denote by $M$ such a macroscopic description: $M(X)$ is the macrostate of the system in the microstate $X$. As an example we may divide $V$ into $K$ cells, where $K$ is large but still $K \ll N$, and specify the
number of particles, the momentum and the amount of energy in each cell, with some tolerance. Clearly there are many $X$’s (in fact a continuum) which correspond to the same $M$. Let $\Gamma_M$ be the region in $\Gamma_E$ consisting of all microstates $X$ corresponding to a given macrostate $M$ and denote by $|\Gamma_M|$ its Liouville volume.

It can be proven [14] that, generally, that in every $\Gamma_E$ of a macroscopic system there is one region $\Gamma_M$ which has most of the volume of $\Gamma_E$. This is called the equilibrium macrostate $M_{\text{eq}}$,

$$\frac{|\Gamma_{M_{\text{eq}}}|}{|\Gamma_E|} = 1 - \varepsilon \tag{2}$$

with $\varepsilon \ll 1$. When $M(X)$ specifies a nonequilibrium state, $|\Gamma_M|$ is much smaller. Thus for a gas consisting of $N$ particles in a volume $V$ the ratio of $|\Gamma_M|$, the volume of a macrostate $M$ in which all the particles are in the left half of the box, and $|\Gamma_{M_{\text{eq}}}|$, the volume of the macrostate $M_{\text{eq}}$ in which there are $(\frac{1}{2} \pm 10^{-10})N$ particles in the left half of the box, is of order $2^{-N}$; see Figure 1. The second picture is slightly more faithful. Neither shows the topology or differences in relative sizes of the different $\Gamma_M$’s. In general, the closer $M$ is to $M_{\text{eq}}$ the larger $\Gamma_M$.

A system is then in macroscopic thermal equilibrium iff $X \in \Gamma_{M_{\text{eq}}}$.

Relevant properties of macroscopic systems depend only on sums over the entire system of functions which depend only on the coordinates and momenta of a few particles. The values of the sums, suitably scaled with $N$, are approximately the same for almost all $X \in \Gamma_{M_{\text{eq}}}$, hence they are typical and observed for (almost) all systems in equilibrium. In fact since $|\Gamma_{M_{\text{eq}}}| \sim |\Gamma_E|$ they are also typical of $X \in \Gamma_E$.

This justifies the use of the microcanonical ensemble to compute relevant properties of an equilibrium system: independent of whether or not the dynamics is ergodic in a mathematical sense.

Figure 1: Schematic Picture of the decomposition of $\Gamma_E$. 

IAMP News Bulletin, April 2021 5
Microscopic Origins of Macroscopic Behavior

Nonequilibrium States

Thus, most microstates in $\Gamma_E$ of a macroscopic system correspond to the system being in equilibrium. A similar statement is true for most wave functions, in fact also for most energy eigenstates, in $\mathcal{H}_E$, the energy shell of the Hilbert space: see later. Fortunately there are also microstates which correspond to macroscopic systems which are out of equilibrium (or we would not be here).

Given an $X(t_0)$ in such a $\Gamma_M, M \neq M_{eq}$, at an “initial” time $t_0$, we would like to know how the properties of a macroscopic system, isolated for $t > t_0$, change with time.

Approach to Equilibrium

Boltzmann (also Maxwell, Kelvin, . . .) argued that given the disparity in the sizes of the $\Gamma_M$ corresponding to the various macrostates, the evolution of a “typical” microstate $X$, specified to be at $t = t_0$ in the phase space region $\Gamma_M$, will be such that $|\Gamma_M(X(t))|$ will not decrease (on a macroscopic scale) for $t > t_0$. In fact for any $\Gamma_M$ the relative volume of the set of microstates $X \in \Gamma_M$ for which this is false over some time period $\tau$ during which the macrostate undergoes a macroscopically noticeable change, but not longer than the age of the universe, goes to zero exponentially in the number of atoms in the system. This explains and describes the evolution towards equilibrium of macroscopic systems which start in the macrostate $\Gamma_M, M \neq M_{eq}$, and are kept (effectively) isolated afterwards.

Boltzmann’s Entropy

To make a connection with the Second Law, Boltzmann defined the (Boltzmann) entropy of a macroscopic system in a microstate $X$ as

$$S_B(X) = \log |\Gamma_M(X)| = S_B(M).$$

Boltzmann then showed that the Clausius thermodynamic entropy of a gas in equilibrium is equal to $\log |\Gamma_E| \sim \log |\Gamma_M_{eq}|$.

The above heuristic argument, based on relative phase space volume, is the correct explanation for the behavior typically observed in actual macroscopic systems. It is, however, very far from a mathematical theorem and contains no quantitative information about time scales. A desirable result would be the rigorous derivation from the microscopic dynamics of the kinetic and hydrodynamic equations commonly used to describe the time asymmetric, entropy increasing, observed behavior of macroscopic systems out of equilibrium. This has been achieved so far only for the Boltzmann equation for dilute gases. This was done rigorously (in appropriate limits) by Oscar Lanford in 1975. (I will not discuss derivations which include some external randomness in the dynamics by the Varadhan school. I will also not discuss the derivation of a diffusion equation for non-interacting particles moving among Sinai billiards. Those cases show what we could do if only our mathematics was better.)
The Boltzmann Equation for Dilute Gases

Following Boltzmann, we refine the description of a macrostate $M$ by noting that the microstate $X = \{r_i, v_i\}, i = 1, \ldots, N$, can be considered as a set of $N$ points in six dimensional one particle space. We then divide up this one particle space into $J$ cells $\Delta_\alpha$, centered on $(r_\alpha, v_\alpha)$, of volume $|\Delta_\alpha|$. A macrostate $M_f(X)$ can then be specified by a distribution $f(x, v)$ in the one-particle space such that the fraction of particles in each $\Delta_\alpha$, is given by

$$N_\alpha/N \cong \int_{\Delta_\alpha} dx dv \; f(x, v).$$

(4)

Boltzmann then used his deep physical intuition (and hints from Maxwell) to derive his eponymous equation for the time evolution of the macrostate $M_f(X(t))$ given by $f(x, v, t)$. I will not reproduce this equation here as this can be found in all textbooks on kinetic theory, c.f. [4]. The reasoning behind it is carefully explained in Lanford’s beautiful non-technical article [15], one of the best mathematical-physics articles I have ever read. I strongly recommend it.

I will now give a bird’s eye view of what I think is the essence of that article:

Let $f_0(x, v), x \in V \subset \mathbb{R}^3, v \in \mathbb{R}^3$, be a smooth function of integral one. Then consider a gas consisting of $N$ spheres of diameter $d$ in $V$. Keeping $V$ fixed consider now a sequence of states with different particle numbers, $N \to \infty$, $d \to 0$, such that $Nd^2 \to b > 0$, while $Nd^3 \to 0$. This is called the Boltzmann-Grad (BG) limit. Consider now all the phase points $X_N$ of this gas such that the fraction of particles in $\Delta_\alpha$, satisfies $N_\alpha/N \cong \int_{\Delta_\alpha} f_0(x, v)dx dv$ with,

$$\lim_{BG} N_\alpha/N = \int_{\Delta_\alpha} f_0(x, v)dx dv$$

(5)

The system with $N$ particles evolves according to Hamiltonian dynamics for elastic collisions going from $X_N$ to $X_N(t), t > 0$. Lanford’s theorem then says:

There exists a $\tau > 0$ such that for $t < \tau$ the $N_\alpha(t)$ corresponding to asymptotically almost all such $X_N(t)$ satisfy

$$\lim_{BG} \frac{N_\alpha(t)}{N} = \int_{\Delta_\alpha} f(x, v, t)dx dv$$

(6)

where $f(x, v, t)$ evolves according to the Boltzmann equation with initial condition $f_0$. Here again almost all is with respect to the relative phase space volume. (6) holds for all reasonable sets of $\Delta_\alpha$’s.

The time $\tau$ for which Lanford’s theorem holds is about one fifth of the mean free time between collisions, but that is a purely technical problem. This time is long enough for the Boltzmann entropy per particle of the macrostate $M_f$ to increase by a finite amount.

The Boltzmann entropy of the macrostate $M_f$, associated with the distribution $f$ is defined as in (3),

$$S_B(f) = S_B(M_f) = \log |\Gamma_{M_f}|$$

(7)
where $|\Gamma_{M_f}|$ is the phase space volume corresponding to $M_f$. $S_B(f)$ was actually computed by Boltzmann. He showed that, up to constants, this is given for a dilute gas, by

$$\frac{1}{N} S_B(f) = - \int_x \int_{\mathbb{R}^3} f(x, v) \log f(x, v)$$

This agrees with the Gibbs-Shannon entropy per particle for a system in a product measure, with each particle having distribution $f(x, v)$, but is conceptually not the same at all (see below).

The maximum of $S_B(f)$ over all $f$ with a given energy, which is here just the kinetic energy, is given by the Maxwell distribution

$$f_{eq} = \frac{N}{|V|} (2\pi kT/m)^{-3/2} \exp\left[-m v^2/2kT\right]$$

where $kT = 2/3(E/N)$.

In this case

$$\frac{1}{N} S_B(f_{eq}) = \frac{3}{2} \log T - \log \frac{N}{|V|} + \text{Const.}$$

the same as the Clausius entropy for a dilute gas.

When $f \neq f_{eq}$ then $f$ and consequently $S_B(f)$ will change in time.

The second law, now says that for typical $X$, $f_X(x, v) = f(x, v, t) = f_t$ has to be such that $S_B(f_t) \geq S_B(f_{t'})$, for $t' \geq t'$.

This is exactly what happens for a dilute gas described by the Boltzmann equation.

$$\frac{d}{dt} S_B(f_t) \geq 0, \quad \text{Boltzmann’s } \mathcal{H}-\text{theorem}$$

As put by Boltzmann [3]:

“In one respect we have even generalized the entropy principle here, in that we have been able to define the entropy in a gas that is not in a stationary state.”

**More General Hydrodynamic Equations**

Suppose, more generally, that the time evolution of the macrostate $M$, given by $M(X(t)) = M_t$, effectively satisfies an autonomous deterministic time asymmetric equation, such as the Navier-Stokes equation or the heat equation or the Boltzmann equation just discussed.

Such an equation means that if $M_{t_1} \rightarrow M_{t_2}$, for $t_2 > t_1$, and $M_{t_2} \rightarrow M_{t_3}$, for $t_3 > t_2$, then the microscopic dynamics $T_t$ carries $\Gamma_{M_{t_1}}$ inside $\Gamma_{M_{t_2}}$, i.e. $T_{t_2-t_1} \Gamma_{M_{t_1}} \subset \Gamma_{M_{t_2}}$ and $T_{t_3-t_2} \Gamma_{M_{t_2}} \subset \Gamma_{M_{t_3}}$, with negligible error. Put otherwise a typical phase point in $\Gamma_{M_{t_1}} = \Gamma_{M_1}$ will go to $\Gamma_{M_{t_2}}$ and then to $\Gamma_{M_3}$, i.e. $T_{t_3-t_1} \Gamma_{M_1} \subset \Gamma_{M_3}$.

The fact that phase space volume is conserved by the Hamiltonian time evolution implies that $|\Gamma_{M_1}| \leq |\Gamma_{M_2}| \leq |\Gamma_{M_3}|$, and thus that $S_B(M_3) \geq S_B(M_2) \geq S_B(M_1)$. A deterministic
Figure 2: Time evolution of a macrostate

Macroscopic equations, for increasing time, then has to satisfy the inequality \( \frac{d}{dt} S_B(M_t) \geq 0 \), [10, 18].

A crucial point here is that the phase points in the region in \( \Gamma_2 \) coming from \( \Gamma_1 \) behave, forward in time, as microstates typical of \( \Gamma_2 \). They are, however, very atypical backwards in time: if we reverse all the velocities in \( \Gamma_2 \), then at a later time, \( t' = t_2 + (t_2 - t_1) \) all of the points initially in \( \Gamma_{M_1} \) will again be in \( \Gamma_{M_1} \) (with their velocities reversed) a smaller region than \( \Gamma_{M_2} \). (The same is true for any sequence of positive times.)

The reason for this asymmetry in typical behaviors is due to initial conditions. That is, when nature or the experimentalist who is part of nature, starts out with a nonequilibrium system in an initial state \( X \in \Gamma_M \) we can assume that \( X \) is typical of \( \Gamma_M \), and continues to be so in the forward time direction.

But how did all this get started? In the Lanford derivation reversing the velocities at some \( t < \tau \) violates the assumptions on the initial conditions required for the derivation of the Boltzmann equation. But what about real life situations? Somewhat surprisingly, if one thinks about it, one has to go back to the very beginning of the world we live in. This was already fully understood by Boltzmann and others as the quotes below show.

**Initial Conditions**

“From the fact that the differential equations of mechanics are left unchanged by reversing the sign of time without changing anything else, Herr Ostwald concludes that the mechanical view of the world cannot explain why natural processes always run preferentially in a definite direction. **But such a view appears to me to overlook that mechanical events are determined not only by differential equations, but also by initial conditions.** In direct contrast to Herr Ostwald I have called it one of the most brilliant confirmations of the mechanical view of Nature that it provides an extraordinarily good picture of the dissipation of energy, as long as one assumes that the world began in an initial state satisfying certain conditions. I have called this state an improbable state.”

— L. Boltzmann [2]
“It is necessary to add to the physical laws the hypothesis that, in the past the universe was more ordered in the technical sense, [i.e. low $S_B$] than it is today ... to make an understanding of irreversibility.”

— R.P. Feynman [7]

Figure 3: “Creation of the universe: a fanciful description! The Creator’s pin has to find a tiny box, just 1 part in $10^{10^{123}}$ of the entire phase-space volume, in order to create a universe with as special a Big Bang as we actually find.” from R. Penrose, The Emperor’s New Mind [19]

The “tiny box” in Fig. 3 is a macrostate with low $S_B$. N.B. It is not necessary to select a particular microstate. Almost all microstates in a low-entropy macrostate will behave in a similar way.

It may be relevant to mention here a question I was asked during a talk I gave on the subject:

Q: What does the initial state of the universe have to do with the fact that when I put my sugar cube in my tea it dissolves irreversibly? A: Nothing directly but the fact that you, the sugar cube and the tea are all here is a consequence of the initial low entropy state of the universe.

**Boltzmann vs. Gibbs Entropies**

Given an ensemble (probability) density $\mu(X)$, the Gibbs-Shannon entropy is given by

$$S_G \equiv -k \int_{\Gamma} \mu \log \mu \ dX. \quad (12)$$

Clearly if $\mu = \tilde{\mu}_M$, where

$$\tilde{\mu}_M = \begin{cases} |\Gamma_M|^{-1}, & \text{if } X \in \Gamma_M; \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

then

$$S_G(\tilde{\mu}_M) = k \log |\Gamma_M| = S_B(M). \quad (14)$$
This is essentially the case for the microcanonical ensemble and thus the Gibbs and Boltzmann entropies are equal for equilibrium systems.

However $S_G(\mu)$ does not change in time for isolated systems and therefore is “useless” for such systems not in equilibrium, while $S_B(M(X))$ captures the essence of typical macroscopic behavior.

**Quantum Systems**

Let me turn now to quantum systems.

I will take the microstate of a system to be its wave function $\psi \in \mathcal{H}_E$, where $\mathcal{H}_E$ is a shell in Hilbert space of thickness $\Delta E$, i.e. $\psi$ is a linear combination of energy eigenfunctions in the range $(E, E + \Delta E)$, $\Delta E \ll E$ but very large compared to the spacing between levels.

This is not perfect (problems with Schrödinger’s Cat) but it will have to do for the present (see below).

The question then is which $\psi$ correspond to the system being in macroscopic thermal equilibrium (MATE), i.e. what is the analog of a classical microstate $X$ lying in $\Gamma_{M\text{eq}}$.

Following von Neumann (Chapter 5 of [17]), we take the macro-observables $M$ to commute with each other. We assume that this can be achieved by suitably “rounding off”, i.e. coarse-graining, the operators representing the macro-observables.

The coarse-grained energy operator commutes with the other coarse-grained macro-observables. Thus all $M$’s can be regarded as operators on $\mathcal{H}_E$. Their joint spectral decomposition defines an orthogonal decomposition

$$\mathcal{H}_E = \bigoplus \mathcal{H}_\nu, \quad (15)$$

The subspaces $\mathcal{H}_\nu$ (“macro spaces”), the joint eigenspaces of the macro-observables, correspond to the different macro states. This corresponds to the division of the classical energy shell $\Gamma_E$ into disjoint regions $\Gamma_M$.

A system is in a macrostate $M_\nu$ if its wave function $\psi$ is “close” to $\mathcal{H}_\nu$, i.e. $\langle \psi | P_\nu | \psi \rangle \geq 1 - \delta$, $\delta \ll 1$, with $P_\nu$ being the projection to $\mathcal{H}_\nu$. As noted earlier due to the Schrödinger’s Cat problem there will be $\psi$ which are a superposition of $\psi$’s in different macrostates. To remedy this one has to go beyond the Copenhagen interpretation of the wave function being a complete description, c.f. [1, 8]. For the present let me say that I would interpret such $\psi$’s as giving probabilities of being in different macrostates.

The “volume” of each macro space $\mathcal{H}_\nu$ is its dimension $d_\nu$.

As in the classical case, it is generally true that one of the $\mathcal{H}_\nu$, denoted $\mathcal{H}_{eq}$, has most of the dimensions of $\mathcal{H}_E$, i.e.,

$$\frac{\dim \mathcal{H}_{eq}}{\dim \mathcal{H}_E} = 1 - \varepsilon \quad (16)$$

with $\varepsilon \ll 1$.

A macroscopic system is in MATE if

$$\langle \psi | P_{eq} | \psi \rangle \geq 1 - \delta. \quad (17)$$
Figure 4: Boltzmann’s grave in Zentralfriedhof, Vienna, with bust and entropy formula. Credit: Daderot at English Wikipedia, CC BY-SA 3.0, via Wikimedia Commons
The Boltzmann entropy $S_B(\psi)$ of a system in a macrostate $M$ is then given by the log of the dimension of the macro space $H_\nu$; $S_B(M_\nu) = \log d_\nu$: $d_\nu$ is the quantum analogue of $|\Gamma_{M_\nu}|$.

Most everything we said about isolated classical macroscopic systems then also holds for quantum systems. In particular for $\varepsilon \ll \delta$, most pure states in $H_E$ are in MATE.

When I say “for most $\psi$”, I mean that $\psi$ is typical with respect to a uniform measure on the unit sphere in $H_E$. This measure on “wave functions” was considered already by Schrödinger and particularly Felix Bloch [11]. It yields the microcanonical measure $\rho^{mc}$ but goes beyond it.

The Boltzmann argument for increase in entropy of isolated macroscopic systems out of equilibrium is then similar to that in the classical case [13]. Unlike classical systems however, where any subsystem of a system in a pure state is also in a pure state, a subsystem $S$ of a quantum system with a wave function $\psi$ will be described by a density matrix $\rho_S^\psi$.

We can now define [9] a system with wave function $\psi$ to be in microscopic thermal equilibrium (MITE) if, for any not-too-large subsystem $S$, say subsystems with linear dimension $\ell < \ell_0$, the reduced density matrix of $S$ is close to the thermal equilibrium density matrix of $S$

$$\rho_S^\psi \approx \rho_S^{mc}$$

where

$$\rho_S^\psi = Tr_{S^c} |\psi\rangle \langle \psi|$$

is the reduced density matrix of $S$ obtained by tracing out the complement $S^c$ of $S$, and

$$\rho_S^{mc} = Tr_{S^c} \rho^{mc}$$

$\rho^{mc}$ is the microcanonical density matrix corresponding to a uniform distribution over energy eigenstates in $H_E$. For macroscopic systems $\rho_S^{mc}$ can be replaced by $\rho_S^{ca}$, where $\rho^{ca}$ is the canonical density matrix.

The distinction between MITE and MATE is particularly relevant for systems with many-body localization (MBL) for which the energy eigenfunctions fail to be in MITE while necessarily most of them, but not all, are in MATE.

The argument for most energy eigenfunctions being in MATE is based on the fact that, calling $D$ the dimension of $H_E$, we have for energy eigenfunctions $|n\rangle$

$$\frac{1}{D} \sum_{n=1}^{D} \langle n | P_{\nu eq} | n \rangle = \frac{1}{D} Tr(P_{\nu eq}) = 1 - \varepsilon$$

Noting that $\langle n | P_{\nu eq} | n \rangle \leq 1$, the average being close to 1 means that most eigenstates are close to $H_{\nu eq}$. This is consistent with the Eigenfunction Thermalization Hypothesis (ETH).

In fact for generic macroscopic systems, including those with MBL, most wave functions in an energy shell are in both MATE and MITE.

This follows from the following result.
**Canonical Typicality**

Consider an isolated system consisting of two parts. Call them system 1 and 2 or system and reservoir. Then [11] we have the following result, see also [16].

Let $H$ be the Hamiltonian of the whole system and let the number of particles in system 1 and 2 be $N_1 \ll N_2$. Let $\mathcal{H}_E \subset \mathcal{H}_1 \otimes \mathcal{H}_2$ be an energy shell. Then for most $\psi \in \mathcal{H}_E$ with $||\psi|| = 1$,

$$\text{tr}_2 |\psi\rangle \langle \psi| \approx \text{tr}_2 \rho_{\text{mc}},$$

where $\rho_{\text{mc}}$ is the microcanonical density matrix of the whole system at energy $E$, i.e. equal weight to all energy eigenstates in $\mathcal{H}_E$.

The theorem says that most wave functions in the energy shell $\mathcal{H}_E$ are both in MATE and in MITE. In fact for macroscopic systems one can show that MITE implies MATE. The opposite is however not true. This is particularly relevant when one considers energy eigenfunctions $|n\rangle$. While most energy eigenstates, including those for systems with MBL must, as shown, generally be in MATE, most energy eigenfunctions for systems with MBL are not in MITE.

There is no analog to MITE for a classical system where any subsystem of a composite system in state $X^{(1,2)}$ is also in a unique state $X^{(2)}$.

When the interaction between systems 1 and 2 is weak, $H \approx H_1 \otimes I_2 + I_1 \otimes H_2$, then, as is well known,

$$\text{tr}_2 \rho_{\text{mc}} \approx \frac{1}{Z} e^{-\beta H_1},$$

for $\beta = \beta(E) = dS_{\text{eq}}(E) / dE$.

If MATE-ETH holds strictly, i.e., if all energy eigenstates in $\mathcal{H}_E$ are in MATE, then every state $\psi \in \mathcal{H}_E$ will sooner or later reach MATE and spend most of the time in MATE in the long run. That is because, writing $\overline{f(t)} = \lim_{T \to \infty} \frac{1}{T} \int_0^T f(t)dt$ for time averages, $|n\rangle$ for the energy eigenstate with eigenvalue $E_n$ and $\psi_t = e^{-iHt}\psi$,

$$\langle \psi_t | P_{\text{eq}} | \psi_t \rangle = \sum_{n,n'} \langle \psi | n \rangle \overline{e^{iE_nt}} \langle n | P_{\text{eq}} | n' \rangle e^{-iE_{n'}t} \langle n' | \psi \rangle$$

$$= \sum_n |\langle \psi | n \rangle|^2 \langle n | P_{\text{eq}} | n \rangle \geq \sum_n |\langle \psi | n \rangle|^2 (1 - \delta)$$

$$= 1 - \delta,$$

provided $H$ is non-degenerate, i.e., $E_n \neq E_{n'}$ for $n = n'$ (using $e^{iEt} = 1$ if $E = 0$ and $= 0$ otherwise).

A similar statement is true when there is degeneracy.
Microscopic Origins of Macroscopic Behavior

Summary of Boltzmann’s Ideas (also Maxwell, Kelvin, Feynman)

Time-asymmetric behavior as embodied in the second law of thermodynamics is observed in individual macroscopic systems. It can be understood as arising naturally from time-symmetric microscopic laws when account is taken of a) the great disparity between microscopic and macroscopic sizes, b) initial conditions, and c) that what we observe is “typical” behaviors — not all imaginable ones. Common alternate explanations, such as those based on equating irreversible macroscopic behavior with ergodic or mixing properties of ensembles (probability distributions) already present for chaotic dynamical systems having only a few degrees of freedom or on the impossibility of having a truly isolated system, are either unnecessary, misguided or misleading.

Let me end this article by quoting Einstein’s tribute to Boltzmann.

“On the basis of kinetic theory of gases Boltzmann had discovered that, aside from a constant factor, entropy is equivalent to the logarithm of the “probability” of the state under consideration. Through this insight he recognized the nature of course of events which, in the sense of thermodynamics, are “irreversible”. Seen from the molecular-mechanical point of view, however all courses of events are reversible. If one calls a molecular-theoretically defined state a microscopically described one, or, more briefly, micro-state, then an immensely large number (Z) of states belong to a macroscopic condition. Z is then a measure of the probability of a chosen macro-state. This idea appears to be of outstanding importance also because of the fact that its usefulness is not limited to microscopic description on the basis of mechanics.”

— A. Einstein, Autobiographical notes

Acknowledgements

I thank S. Goldstein and E. Speer for many very useful comments. I also thank A.J. Krueger for very helpful technical support in putting this paper together.

References


Microscopic Origins of Macroscopic Behavior


The Ionization Problem

by PHAN THÀNH NAM (Munich)

The question: “How many electrons can a nucleus bind?” is as old as quantum mechanics, but its rigorous answer based on the many-body Schrödinger equation remains a difficult challenge to mathematicians. Nevertheless, there has been remarkable progress in this problem in the past four decades. We will review the current understanding of the Schrödinger equation and then turn to simplified models where the problem has been solved satisfactorily. We will also discuss the connection to the liquid drop model, which is somewhat more classical, but no less interesting.

1 Atomic Schrödinger equation

For us, an atom is a system of $N$ quantum electrons of charge $-1$ moving around a heavy classical nucleus of charge $Z \in \mathbb{N}$ and interacting via Coulomb force (we use atomic units). The wave functions of $N$ electrons are normalized functions in $L^2(\mathbb{R}^{3N})$ satisfying the antisymmetry

$$\Psi(x_1, \ldots, x_i, \ldots, x_j, \ldots, x_N) = -\Psi(x_1, \ldots, x_j, \ldots, x_i, \ldots, x_N), \quad \forall i \neq j,$$

where $x_i \in \mathbb{R}^3$ stands for the position of the $i$-th electron (we will ignore the spin for simplicity). The Hamiltonian of the system is

$$H_N = \sum_{i=1}^{N} \left(-\frac{1}{2}\Delta x_i - \frac{Z}{|x_i|}\right) + \sum_{1 \leq i < j \leq N} \frac{1}{|x_i - x_j|}.$$ 

The self-adjointness of $H_N$ follows from a famous theorem of Kato. We are interested in the ground state problem

$$E_N = \inf_{||\Psi||_{L^2}=1} \langle \Psi, H_N \Psi \rangle.$$ 

By a standard variational method, we know that the minimizers, if they exist, are solutions to the Schrödinger equation

$$H_N \Psi = E_N \Psi.$$ 

The existence/nonexistence issue is related to the stability of the system, namely whether all electrons will be bound, or some of them may escape to infinity. Obviously, $H_N$ and $E_N$ also depend on $Z$, but let us not include this dependence in the notation.

It is natural to guess that there is a sharp transition when $N$ crosses the value $Z + 1$. Heuristically, if $N < Z + 1$, then the outermost electron sees the rest of the system as a large nucleus of the effective charge $Z - (N - 1) > 0$. Hence, this electron will “prefer to stay” by the
Coulomb attraction. On the other hand, if \( N > Z + 1 \), then the outermost electron will “prefer to go away” by the Coulomb repulsion.

Part of the above heuristic guess was justified by Zhislin in 1960.

**Theorem 1.** If \( N < Z + 1 \), then \( E_N \) has a minimizer.

More precisely, he proved that if \( N < Z + 1 \), then \( E_N < E_{N-1} \). This strict binding inequality prevents any electron from escaping to infinity, thus ensuring the compactness of minimizing sequences for \( E_N \). On a more abstract level, if \( N < Z + 1 \), then \( E_N \) is strictly below the essential spectrum of \( H_N \). In fact, the essential spectrum of \( H_N \) is \([E_{N-1}, \infty)\) due to the celebrated Hunziker–van Winter–Zhislin (HVZ) theorem.

Thus, Zhislin’s theorem ensures the existence of all positive ions and neutral atoms. On the other hand, the nonexistence of highly negative ions is much more difficult, and often referred to as the “ionization conjecture”; see e.g. [23, Problem 9] and [15, Chapter 12].

**Conjecture 2.** There exists a universal constant \( C > 0 \) (possibly \( C = 1 \)) such that if \( N > Z + C \), then \( E_N \) has no minimizer.

Note that the above heuristic argument is purely classical and it is too rough to understand the delicate quantum problem at hand. In 1983, Benguria and Lieb [1] proved that if the anti-symmetry condition of the wave functions is ignored, then the atoms with “bosonic electrons” always exist as soon as \( N \leq t_c Z \) with a universal constant \( t_c > 1 \) (numerically \( t_c \approx 1.21 \), computed by Baumgartner). Thus the ionization problem requires a deep insight, as the particle statistics, more precisely Pauli’s exclusion principle, play an essential role.

## 2 Known results

A rigorous upper bound to the question “How many electrons can a nucleus bind?” was first derived by Ruskai [19] and Sigal [21] independently in 1982. They proved that there exists a critical value \( N_c(Z) < \infty \) such that if \( N > N_c(Z) \), then \( E_N \) has no minimizer. In these works, they applied certain inequalities on classical point particles to the quantum problem via the geometric localization method. In particular, Sigal realized that for every collection \( \{x_i\}_{i=1}^N \subset \mathbb{R}^3 \) with \( N > 2Z + 1 \), the energy contributed by the farthest electron, \( x_N \), says, is always positive because of the triangle inequality

\[
-\frac{Z}{|x_N|} + \sum_{i=1}^{N-1} \frac{1}{|x_i - x_N|} \geq -\frac{Z}{|x_N|} + \frac{N-1}{2|x_N|} > 0.
\]

This leads to the upper bound \( \lim_{Z \to \infty} \sup_{N_c(Z)/Z \leq 2} \frac{N_c(Z)}{Z} \leq 2 \) in [22].

Later, Lieb, Sigal, Simon and Thirring [13] found the following improvement: for every \( \{x_i\}_{i=1}^N \subset \mathbb{R}^3 \) with \( N \) large, one has

\[
\max_{1 \leq i < j \leq N} \left\{ \sum_{1 \leq i \neq j} \frac{1}{|x_i - x_j|} - \frac{N + o(N)}{|x_j|} \right\} \geq 0. \tag{2.1}
\]
The Ionization Problem

Consequently, they obtained the asymptotic neutrality

$$\lim_{Z \to \infty} \frac{N_c(Z)}{Z} = 1.$$  

It is unclear whether one can improve the quantity $N + o(N)$ in (2.1) to $N + O(N^\alpha)$ with some constant $0 \leq \alpha < 1$.

In 1990, Fefferman and Seco [4], and Seco, Sigal and Solovej [20], proved

**Theorem 3.** When $Z \to \infty$, we have $N_c(Z) \leq Z + O(Z^{5/7})$.

This bound was obtained by comparing it with the Thomas–Fermi theory (that we will revisit below) and taking into account quantitative estimates for Scott’s correction (studied by Hughes, and by Siedentop and Weikard). There has been no further improvement in the past three decades!

Instead of the asymptotics as $Z \to \infty$, one may also be interested in explicit bounds for all $Z$ (in fact, $1 \leq Z \leq 118$ for realistic atoms in the current periodic table). The best known result in this direction is

**Theorem 4 ([12, 17]).** For all $Z \geq 1$, $N_c(Z) < \min(2Z + 1, 1.22Z + 3Z^{1/3})$.

Let us quickly explain Lieb’s proof of the bound $2Z + 1$ in [12], since it is short and important. The starting point is the following identity, which follows from the Schrödinger equation

$$\langle |x_N| \Psi_N, (H_N - E_N) \Psi_N \rangle = 0.$$  

The idea of “multiplying the equation by $|x|$” was also used by Benguria on a simplified model. Then we decompose

$$H_N = H_{N-1} - \Delta_N + \sum_{i=1}^{N-1} \frac{1}{|x_i - x_N|}.$$  

For the first $(N - 1)$ electrons, we use the obvious inequality

$$H_{N-1} \geq E_{N-1} \geq E_N.$$  

For the $N$-th electron, we use the operator inequality

$$(-\Delta)|x| + |x|(-\Delta) \geq 0$$  

on $L^2(\mathbb{R}^3)$ (which is equivalent to Hardy’s inequality). Consequently,

$$Z > \sum_{i=1}^{N} \left\langle |x_N| \frac{\Psi}{|x_i - x_N|}, \Psi \right\rangle = \frac{1}{2} \sum_{i=1}^{N} \left\langle |x_N| + |x_i| \frac{\Psi}{|x_i - x_N|}, \Psi \right\rangle > \frac{N - 1}{2}.$$  

Thus $N < 2Z + 1$. Here, we have used the symmetry of $|\Psi|^2$ and the triangle inequality.

To get the bound in [17], we multiply Schrödinger’s equation with $|x_N|^2$ instead of $|x_N|$ and proceed similarly. In this case, the operator $(-\Delta)|x|^2 + |x|^2(-\Delta)$ on $L^2(\mathbb{R}^3)$ is not positive,
but its negative part can be controlled using a special property of the ground state. The key point is, instead of using the triangle inequality, we now have
\[ Z \geq \inf_{\{x_i\}_{i=1}^N \subset \mathbb{R}^3} \left\{ \sum_{1 \leq i < j \leq N} \frac{|x_i|^2 + |x_j|^2}{|x_i - x_j|^2} \right\} + O(N^{2/3}) = \beta N + O(N^{2/3}) \]
with the statistical value
\[ \beta := \inf_{\rho \text{ probability measure in } \mathbb{R}^3} \left\{ \frac{\iint_{\mathbb{R}^3 \times \mathbb{R}^3} x^2 + y^2 \rho(x) \rho(y) \, dx \, dy}{\iint_{\mathbb{R}^3} |x| \rho(x) \, dx} \right\} . \]

It is nontrivial to compute \( \beta \), but we can estimate it using the inequality
\[ \iint_{\mathbb{R}^3 \times \mathbb{R}^3} \frac{x^2 + y^2}{|x - y|} \rho(x) \rho(y) \, dx \, dy \geq \iint_{\mathbb{R}^3 \times \mathbb{R}^3} \left( \max(|x|, |y|) + \frac{\min(|x|, |y|)}{|x - y|} \right) \rho(x) \rho(y) \, dx \, dy , \]
which is a consequence of (2.1). This gives \( \beta \geq 0.82 \), leading to the bound \( 1.22 Z + 3Z^{1/3} \) (as \( \beta^{-1} \leq 1.22 \)).

3 Thomas-Fermi theory

Since the Schrödinger equation is too complicated, for practical computations one often relies on approximate models which are nonlinear but dependent on less variables. In density functional theory, a popular method in computational physics and chemistry, one replaces the \( N \)-body wave function \( \Psi \) with its one-body density
\[ \rho_{\Psi}(x) = N \int_{\mathbb{R}^3} |\Psi(x, x_2, ..., x_N)|^2 \, dx_2 ... dx_N . \]
Clearly, \( \rho_{\Psi} : \mathbb{R}^3 \to [0, \infty) \) and \( \int_{\mathbb{R}^3} \rho_{\Psi} = N \).

The oldest density functional theory was proposed by Thomas and Fermi in 1927. In the Thomas–Fermi (TF) theory, the ground state energy \( E_N \) is replaced by its semiclassical approximation
\[ E_{\text{TF}}(N) = \inf_{\rho = N} \left\{ C_{\text{TF}} \int_{\mathbb{R}^3} \left( \rho^{5/3}(x) - \frac{Z}{|x|} \rho(x) \right) \, dx + \frac{1}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{\rho(x) \rho(y)}{|x - y|} \, dx \, dy \right\} \]
with a constant \( C_{\text{TF}} > 0 \). The existence and properties of the TF minimizers was studied by Lieb and Simon in [14]. In particular, they proved

**Theorem 5.** \( E_{\text{TF}}(N) \) has a minimizer if and only if \( N \leq Z \).
The Ionization Problem

By standard techniques, we find that the TF functional is convex and rotation invariant. Therefore, if a minimizer exists, it is unique and radial. Moreover, it satisfies the TF equation
\[
\frac{5}{3} C_{TF} \rho(x)^{2/3} = [Z|x|^{-1} - \rho \ast |x|^{-1} - \mu]_+
\]
for some chemical potential \( \mu \leq 0 \).

The existence of the TF minimizer is rather similar to Zhislin’s theorem for the Schrödinger equation. The nonexistence is more challenging. The original proof of Lieb and Simon is based on a clever use of the maximum principle. Here we offer another proof, using a variant of the Benguria–Lieb argument.

Proof of \( N \leq Z \) \([18]\). Assume that the TF equation has a radial solution \( \rho \). Multiplying the equation with \( |x|^k \rho(x) \), \( k \geq 1 \), we have the pointwise inequality
\[
(Z|x|^{-1} - \rho \ast |x|^{-1} - \mu) \rho(x)|x|^k = \frac{5}{3} C_{TF} \rho(x)^{5/3}|x|^k \geq 0.
\]
Then we integrate over \( \{|x| \leq R\} \). Note that \( \mu \leq 0 \). Moreover, since \( \rho \) is radial, by Newton’s theorem we have
\[
\rho \ast |x|^{-1} = \int_{\mathbb{R}^3} \frac{\rho(y)}{\max(|x|, |y|)} dy.
\]
Consequently,
\[
Z \int_{|x| \leq R} |x|^{k-1} \rho(x) \geq \int_{|x| \leq R} |x|^k \rho(x)(\rho \ast |x|^{-1}) dx
\]
\[
\geq \frac{1}{2} \iint_{|x|,|y| \leq R} \frac{(|x|^k + |y|^k) \rho(x) \rho(y)}{\max(|x|, |y|)} dxdy.
\]
On the other hand, by the AM-GM inequality,
\[
\frac{|x|^k + |y|^k}{\max(|x|, |y|)} \geq \left(1 - \frac{1}{k}\right) (|x|^{k-1} + |y|^{k-1}).
\]
Thus
\[
Z \int_{|x| \leq R} |x|^{k-1} \rho(x)dx \geq \left(1 - \frac{1}{k}\right) \left( \int_{|x| \leq R} |x|^{k-1} \rho(x) dx \right) \left( \int_{|y| \leq R} \rho(y) dy \right).
\]
Taking \( R \to \infty \) and \( k \to \infty \), we conclude that \( \int_{\mathbb{R}^3} \rho \leq Z \) \( \square \).

When \( N = Z \), the TF minimizer has the perfect scaling property:
\[
\rho_{TF}^Z(x) = Z^2 \rho_{TF}^1 (Z^{1/3} x), \quad \forall x \in \mathbb{R}^3
\]
where the function \( \rho_{TF}^1 \) is independent of \( Z \). Moreover, it satisfies the TF equation with chemical potential 0. Thus if we denote the TF potential
\[
\varphi_{TF}^Z(x) = Z|x|^{-1} - \rho_{TF}^Z \ast |x|^{-1},
\]

then the TF equation can be written as the nonlinear Schrödinger equation
\[ \Delta \psi_{Z}^{TF}(x) = 4\pi \frac{5}{3} C_{TF}^{-3/2} \psi_{Z}^{TF}(x)^{3/2}. \]

This leads to the following Sommerfeld estimate [24, Theorem 4.6].

**Theorem 6.** Denote
\[ A_{TF} = (5 C_{TF})^{3} (3\pi^{2})^{-1} \]
and \( \zeta = (\sqrt{73} - 7)/2. \) Then
\[ A_{TF} \geq \psi_{Z}^{TF}(x)|x|^{4} \geq A_{TF} - C \left( Z^{1/3} |x| \right)^{-\zeta}, \quad \forall x \neq 0 \]
(3.1)

In particular, when \( |x| \gg Z^{-1/3}, \) then the TF potential \( \psi_{Z}^{TF} \) is more or less independent of \( Z. \) This universality makes the TF approximation much more useful than what can normally be explained by its semiclassical nature. More precisely, the standard semiclassical analysis ensures that the TF theory gives a good approximation for the electron density in the distance \( |x| \sim Z^{-1/3}. \) However, we may expect that the TF theory gives a good approximation for larger distances, possibly up to \( |x| \sim 1. \) We refer to [25] for a detailed discussion.

### 4 Hartree-Fock theory

Invented shortly after the discovery of the Schrödinger equation, the Hartree–Fock (HF) theory has been a very useful computational method to describe electronic orbitals. In this theory, one restricts \( N \)-body wave functions to Slater determinants, or equivalently to their one-body density matrices which are trace class operators on \( L^{2}(\mathbb{R}^{3}) \) satisfying
\[ 0 \leq \gamma \leq 1, \quad \gamma = \gamma^{2}, \quad \text{Tr} \gamma = N. \]

The HF ground state energy is
\[ E_{HF}(N) = \inf_{\text{Tr} \gamma = N} \left( \text{Tr}((-\Delta - Z|x|^{-1})\gamma) ight. \\
\left. + \frac{1}{2} \iint_{\mathbb{R}^{3} \times \mathbb{R}^{3}} \frac{\rho_{\gamma}(x)\rho_{\gamma}(y) - |\gamma(x; y)|^{2}}{|x - y|} dxdy \right) \]
where \( \rho_{\gamma}(x) = \gamma(x; x) \) (the kernel of \( \gamma \) is defined properly via the spectral decomposition).

The existence of Hartree–Fock minimizers when \( N < Z + 1 \) was proved by Lieb and Simon in 1977. The nonexistence was proved later by Solovej in 2003 [24].

**Theorem 7.** There exists a universal constant \( C > 0 \) such that if \( N > Z + C, \) then \( E_{HF}(N) \) has no minimizer.

To explain the proof, let us go back to the heuristic argument discussed before. Assume that we have an efficient method to separate \( m \) outermost electrons. Then these particles see the rest of the system as a big nucleus with the effective nuclear charge \( Z' = Z - (N - m). \) Thus by the Benguria–Lieb method, we may hope to get a bound like \( m < 2Z' + 1. \) Since \( Z' \)
The Ionization Problem

is smaller than $Z$, the loss of the factor 2 becomes less serious. If the procedure can be iterated to bring $Z'$ down to order 1, then we can conclude that $N - Z$ is of order 1.

In [24], this approach is carried out rigorously by studying the screened nuclear potential

$$
\Phi^\text{HF}_Z(x) = \frac{Z}{|x|} - \int_{|y| < |x|} \frac{\rho^\text{HF}(y)}{|x - y|} dy.
$$

This function will be compared with the corresponding TF version

$$
\Phi^\text{TF}_Z(x) = \frac{Z}{|x|} - \int_{|y| < |x|} \frac{\rho^\text{TF}(y)}{|x - y|} dy.
$$

Similar to the TF potential $\varphi^\text{TF}_Z(x)$, $\Phi^\text{TF}_Z(x)$ behaves as $|x|^{-4}$ for $|x| \gg Z^{-1/3}$. It turns out that this property holds true for the HF screened potential as well. The key ingredient of the analysis in [24] is

**Theorem 8.** There exist constants $C > 0$, $\epsilon > 0$ such that for all $x \neq 0$,

$$
|\Phi^\text{HF}_Z(x) - \Phi^\text{TF}_Z(x)| \leq C(1 + |x|^{-4+\epsilon}).
$$

This estimate can be proved by induction in $|x|$. First, for $|x| \leq Z^{-1/3+\epsilon}$, it follows by the semiclassical approximation. For longer distances, one repeatedly uses the Sommerfeld estimate (3.1) to get refined information for “inner electrons”, and then controls the “outer electrons” in terms of the screened potential. At the end of the day, the universality of the TF potential makes a miracle happen!

Let us explain why Theorem 8 implies the ionization bound. First, Theorem 8 implies that for $|x| = r \sim 1$,

$$
\int_{|y| < r} \frac{\rho^\text{HF}(y) - \rho^\text{TF}(y)}{|x - y|} dy \leq C r.
$$

We can replace $x$ by $\nu x$ with $\nu \in S^2$, then average over $\nu$ and use Newton’s theorem. This gives

$$
Z' := \int_{|y| \leq r} (\rho^\text{HF}(y) - \rho^\text{TF}(y)) dy \leq C r.
$$

The number of outermost electrons, namely $\int_{|y| > r} \rho^\text{HF}$, can be controlled by a constant time $Z'$, leading to the final bound $N - Z \leq C$.

Clearly, this proof strategy requires an efficient way of splitting the problem from the inside and the problem from the outside. This can be done for the Hartree–Fock theory, because the energy functional has been greatly simplified to a one-body functional. For the $N$-body Schrödinger equation, such a splitting would require a novel many-body localization technique which is not available at the moment.
5 Liquid drop model

Now let us turn to a related problem in the liquid drop model which is somewhat more classical than the ionization conjecture. This model was proposed by Gamow in 1928 and further developed by Heisenberg, von Weizsäcker and Bohr in the 1930s. Recently, it has gained renewed interest from many mathematicians [3].

In modern language, a nucleus is described in this theory by an open set $\Omega \subset \mathbb{R}^3$ which solves the minimization problem

$$E^G(m) = \inf_{|\Omega|=m} \left\{ \text{Per}(\Omega) + \frac{1}{2} \int_{\Omega} \int_{\Omega} \frac{1}{|x-y|} \, dx \, dy \right\}.$$ 

Here $m$ stands for the number of nucleons (protons and neutrons) and $\text{Per}(\Omega)$ is the perimeter in the sense of De Giorgi (which is the surface area of $\Omega$ when the boundary is smooth). The Coulomb term captures the electrostatic energy of protons.

It is generally assumed in physics literature that if a minimizer exists, then it is a ball. Consequently, by comparing the energy of a ball of volume $m$ with the energy of a union of two balls of volume $m/2$, one expects the nonexistence of minimizers if $m > m_*$ with

$$m_* = \frac{2 - 2^{2/3}}{2^{2/3} - 1} \approx 3.518.$$ 

Conjecture 9 ([2]). If $m \leq m_*$, then $E^G(m)$ is minimized by a ball. If $m > m_*$, then $E^G(m)$ has no minimizer.

In particular, the nonexistence of minimizers for large $m$ is consistent with nuclear fission of heavy nuclei, which was discovered experimentally by Hahn and Strassmann in 1938.


Mathematically, it is nontrivial to analyse $E^G(m)$ due to the energy competition: among all measurable sets of a given volume, a ball minimizes the perimeter (by the isoperimetric inequality) but maximizes the Coulomb self-interaction energy (by the Riesz rearrangement inequality).

In 2014, Knüpfer and Muratov [9] proved the following

IAMP News Bulletin, April 2021 25
The Ionization Problem

**Theorem 10.** There exist constants $0 < m_1 < m_2$ such that:

(i) If $m < m_1$, then $E^G(m)$ has a unique minimizer which is a ball;
(ii) If $m > m_2$, then $E^G(m)$ has no minimizer.

The proof in [9] uses deep techniques in geometric measure theory, including a quantitative isoperimetric inequality proved by Fusco, Maggi and Pratelli in 2008. Independently, the existence of small $m$ was proved by Julin [10] and the nonexistence of large $m$ was proved by Lu and Otto [16]. In 2016, with Rupert Frank and Rowan Killip, we offered a new proof of the nonexistence which also provides the quantitative bound $m_2 \leq 8$. Let us explain the short proofs in [10] and [5].

**Proof. Existence for $m$ small [10].** Consider

$$D(\Omega) := \text{Per}(\Omega) + \frac{1}{2} \iint_{\Omega} \frac{1}{|x-y|} dxdy - \text{Per}(\Omega^*) - \frac{1}{2} \iint_{\Omega^*} \frac{1}{|x-y|} dxdy$$

where $\Omega^*$ is the ball centered at 0 with volume $|\Omega^*| = |\Omega| = m$. We need to prove that if $m$ is small, then $D(\Omega) > 0$ unless $\Omega$ is a ball. Denote

$$f = \chi_{\Omega^*} - \chi_\Omega, \quad V = f * |x|^{-1}.$$  

By a quantitative isoperimetric inequality in [8], there exists a universal constant $\epsilon_0 > 0$ such that after an appropriate translation of $\Omega$, we have

$$\text{Per}(\Omega) - \text{Per}(\Omega^*) \geq \epsilon_0 \int_{\mathbb{R}^3} \frac{f(x)}{|x|} dx = \epsilon_0 V(0).$$

Note that by Hardy–Littlewood rearrangement inequality, $V(0) > 0$ unless $\Omega$ is a ball. For the Coulomb terms, we can write

$$\frac{1}{2} \iint_{\Omega} \frac{1}{|x-y|} dxdy - \frac{1}{2} \iint_{\Omega^*} \frac{1}{|x-y|} dxdy = \frac{1}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{f(x)f(y)}{|x-y|} dxdy + \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \chi_{\Omega^*}(x)f(y) dxdy$$

$$= \frac{1}{8\pi} \int_{\mathbb{R}^3} |\nabla V(x)|^2 dx + \int_{\Omega^*} V(x) dx.$$

In the last equality we used $-\Delta V = 4\pi f$. This Poisson equation also shows that $V$ is superharmonic in $\Omega^*$ (as $f \geq 0$ in $\Omega^*$), and hence

$$\int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{\chi_{\Omega^*}(x)f(y)}{|x-y|} dxdy = \int_{\Omega^*} V(x) dx \leq |\Omega^*| V(0) = mV(0).$$

Thus in summary, if $m < \epsilon_0$ and $\Omega$ is not a ball, then

$$D(\Omega) \geq (\epsilon_0 - m)V(0) > 0.$$
The Ionization Problem

**Nonexistence if** \( m > 8 \) [5]. Assume that \( E^G(m) \) has a minimizer \( \Omega \). We split \( \Omega \) into two parts, \( \Omega = \Omega^+ \cup \Omega^- \), by a hyperplane \( H \) and then move \( \Omega^- \) to infinity by translations. Since \( \Omega \) is a minimizer, we obtain

\[
\text{Per}(\Omega) + \int_{\Omega} \int_{\Omega} \frac{1}{|x - y|} dxdy \leq \text{Per}(\Omega^+) + \int_{\Omega^+} \int_{\Omega^+} \frac{1}{|x - y|} dxdy \\
+ \text{Per}(\Omega^-) + \int_{\Omega^-} \int_{\Omega^-} \frac{1}{|x - y|} dxdy
\]

which is equivalent to

\[
2\mathcal{H}^2(\Omega \cap H) \geq \int_{\Omega^+} \int_{\Omega^-} \frac{1}{|x - y|} dxdy.
\]

Here \( \mathcal{H}^2 \) is the two-dimensional Hausdorff measure. Next, we parameterize:

\[
H = H_{\nu, \ell} = \{ x \in \mathbb{R}^3 : x \cdot \nu = \ell \}
\]

with \( \nu \in S^2 \), \( \ell \in \mathbb{R} \). The above inequality becomes

\[
2\mathcal{H}^2(\Omega \cap H_{\nu, \ell}) \geq \int_{\Omega} \int_{\Omega} \frac{\chi(\nu \cdot x > \ell > \nu \cdot y)}{|x - y|} dxdy.
\]

Integrating over \( \ell \in \mathbb{R} \) and using Fubini’s theorem we get

\[
2|\Omega| \geq \int_{\Omega} \int_{\Omega} \frac{[\nu \cdot (x - y)]^+}{|x - y|} dxdy.
\]

Finally, averaging over \( \nu \in S^2 \) and using

\[
\int [\nu \cdot z]_+ \frac{d\nu}{4\pi} = \frac{|z|}{2} \int_0^{\pi/2} \cos \theta \sin \theta d\theta = \frac{|z|}{4}
\]

with \( z = (x - y) \), we conclude that \( 2|\Omega| \geq \frac{1}{4} |\Omega|^2 \), namely \( |\Omega| \leq 8 \).

With Rupert Frank and Hanne Van Den Bosch, we used the cutting argument in the liquid drop model to study the ionization problem in the Thomas–Fermi–Dirac-von Weisäcker theory in [6], and in the Müller density matrix functional theory in [7]. In these theories, the standard Benguria–Lieb method does not apply, but we can replace it by an appropriate modification of the minimizers, leading to an efficient control of the number of particles “outside” in terms of particles “inside”. This enables us to employ Solovej’s bootstrap argument to establish the uniform bound \( N - Z \leq C \).

### 6 Related problems

The ionization problem is an example of a question that is easy to find in physics textbooks, but difficult to answer mathematically. Below we list some related open problems for the Schrödinger operator \( H_N \).
The main concept in the ionization problem is that in a large atom, although most of electrons stay in the domain $|x| \sim Z^{-1/3}$, the binding property only depends on a few outermost electrons in the region $|x| \sim 1$. In fact, only this outer region is relevant to chemical reactions in everyday life. Therefore, an important quantity of an atom is its radius. To fix the notation, we define the radius $R_\Psi$ of a wave function $\Psi$ by requiring
\[
\int_{|x| \geq R_\Psi} \rho_\Psi(x) dx = 1.
\]

**Conjecture 11** ([24]). There exist two universal constants $0 < R_1 < R_2$ such that if $N \geq Z$ and $E_N$ has a minimizer $\Psi$, then $R_1 \leq R_\Psi \leq R_2$.

Another important quantity is the ionization energy $I_N = E_{N-1} - E_N$.

**Conjecture 12** ([23, 15]). There exists a universal constant $C > 0$ such that if $N \geq Z$, then $I_N \leq C$.

**Conjecture 13** ([15]). The function $N \mapsto I_N$ is non-increasing (equivalently $N \mapsto E_N$ is convex).

See [20] for partial results on Conjectures 11 and 12. A consequence of Conjecture 13 is that if $E_{N-1} > E_N$ (namely the nucleus can bind $N$ electrons), then $E_{N-2} > E_{N-1}$ (the nucleus can bind $N - 1$ electrons). This “obvious fact” is still not proved mathematically!

So far we have only focused on the ground state problem for $H_N$. Recall from the HVZ theorem that the essential spectrum of $H_N$ is $[E_{N-1}, \infty)$. Conjecture 2 mainly concerns the existence of eigenvalues below $E_{N-1}$. Since the existence of embedded eigenvalues is generally not expected, we have the following stronger version of Conjecture 2.

**Conjecture 14.** There exists a universal constant $C > 0$ such that if $N > Z + C$, then $H_N$ has no eigenvalue.

The last issue has been studied by Lenzmann and Lewin in [11], who proved that $H_N$ has no eigenvalue if $N > 4Z + 1$. This question is related to the scattering theory of dispersive PDEs with long-range interaction potentials, which is interesting in its own right.

**Short CV:** Phan Thành Nam is a mathematical physicist working on analytical problems from quantum physics. He is particularly interested in the spectral theory of many-body quantum systems, including atoms, molecules and Bose gases. He is currently a professor of mathematics at LMU Munich.

This article is reprinted with permission from the EMS Newsletter, where it first appeared in the December, 2020, issue, pp. 22–27.

**References**


**Time’s Arrow**

This new column in the *News Bulletin* will list scientific anniversaries, commemorative conferences, obituaries, and other commemorative items. We welcome your input for future issues, which can be sent to bulletin@iamp.org.

**Scientific anniversaries**

1821. Augustin-Louis Cauchy published his *Cours d’Analyse*, introducing the notion of the limit and setting the path for all future textbooks on infinitesimal calculus.

1921. The *Niels Bohr Institute*, then known simply as the Institute for Theoretical Physics, was founded on 3 March. (Celebrations have been deferred to 2022.)


1971. The first network e-mail message was sent, by Ray Tomlinson of BBN.

**Recent celebratory conferences**


Barry Simon’s 75th Birthday Conference, A mini-conference in celebration of the 75th birthday of Barry Simon, April 18, 2021

**Lost luminaries**

Detlef Dürr, 3 January, 2021.
Georgi Raikov, 10 March 2021.

In 2020, in addition to those whose passing has already been acknowledged in these pages, we remember:

Freeman Dyson, 28 February, 2020.
A memorial tribute is planned for a future issue of the *News Bulletin*, jointly published with the *Notices of the AMS*.

A memorial tribute is planned for a future issue of the *News Bulletin*.

Vaughn Frederick Randal Jones, 6 September, 2020.
A memorial tribute to Vaughn Jones appeared in the December *Newsletter of the EMS*.


A memorial tribute to Misha Shubin appeared in the January *Notices of the AMS*.
News from the IAMP Executive Committee

The IAMP Executive Committee wishes to express its deep appreciation to Jan Dereziński and Daniel Ueltschi for creating the One World Mathematical Physics Seminar and for so expertly running it for the past year.

For the continuation of the seminar they are now joined by Executive Committee members Kasia Rejzner and Marcello Porta.

Flora Koukiou and Hal Tasaki organized a brief tribute to the past and current organizers of the seminar, which took place after the talk by Jan Wehr on March 30, 2021. A recording of the tribute, including short speeches by Robert Seiringer, Bruno Nachtergaele, and Elliott Lieb can be found here: http://youtu.be/uFpegN8ALYw.

The program for the seminar and links to recordings of past talks are available on the IAMP website:


New individual members

IAMP welcomes the following new members

1. DR. ALLENSANDRO ZAMPINI, Università degli Studi Di Napoli Federico II, Napoli, Italy
2. DR. SIMONE RADEMACHER, IST, Klosterneuburg, Austria
3. MR. THÉOTIME GIRARDOT, LPMMC, Grenoble, France
4. DR. TYLER HELMUTH, University of Durham, United Kingdom
5. PROF. STEPHEN P.A.A. DIAS BARRETO, Padre Conceicao College of Engineering, Verna, India
6. DR. ALVIN MOON, University of Copenhagen, Denmark
7. DR. PER MOOSAVI, ETH Zurich, Zurich, Switzerland
8. DR. PETER MADSEN, Université Paris-Dauphine, Paris, France
9. MR. SEVERIN SCHRAVEN, University of Zurich, Zurich, Switzerland
10. PROF. ELIO CONTE, School of Advanced International Studies on Applied Theoretical and non Linear methodologies of Physics, Bari, Italy
11. MR. GOGOI DHRUBA JYOTI, Dibrugarh University, India
12. DR. TOBIAS RIED, Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany
13. DR. ISHMAEL TAKYI, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Upcoming conferences

**Random Matrix Theory and Statistics**
April 30, 2021, Royal Statistical Society, online meeting.

**XX International Congress on Mathematical Physics**
**Geneva, August 2-7, 2021.**

The International Congress on Mathematical Physics (ICMP) is the most important conference of the International Association of Mathematical Physics (IAMP). In 2021, the Congress will take place in Geneva, August 2-7, at the International Conference Centre (CICG).

The organization of the Congress is under way. The list of plenary and session speakers, together with other useful information, can already be found at the Congress webpage [https://www.icmp2021.com/](https://www.icmp2021.com/). As usual, the Congress will be preceded by the Young Research Symposium (YRS) that will be held at the University of Geneva, July 29-31, 2021.

Registration for ICMP and YRS, application for financial help for junior participants and abstract submission will open in the next weeks. Because of the uncertainty related to the COVID 19 pandemic, we are planning to stream ICMP talks. For this reason, we will offer two separate registration options, for full on-site participation and for remote participation (with a reduced fee). At any time, participants will be able to upgrade to on-site participation or to downgrade to remote participation, at essentially no additional costs. We very much hope that by next summer, the situation will improve and that we will be able to welcome many of you in Geneva!

On behalf of the local organizing committee, we would like to invite all IAMP members to attend ICMP 2021.

For an updated list of academic job announcements in mathematical physics and related fields visit


Michael Loss (IAMP Secretary)
Contact coordinates for this issue

**JOEL LEBOWITZ**  
Center for Mathematical Sciences Research  
Rutgers, The State University of New Jersey  
110 Frelinghuysen Road  
Piscataway, NJ 08854-8019, USA  
lebowitz@math.rutgers.edu

**PHAN THÀNH NAM**  
Mathematics Institute  
Ludwig-Maximilians-Universität  
Theresienstrasse 39  
D-80333 Munich, Germany  
nam@math.lmu.de

**MICHAEL LOSS**  
School of Mathematics  
Georgia Institute of Technology  
Atlanta, GA 30332-0160, USA  
secretary@iamp.org

**EVANS HARRELL**  
School of Mathematics  
Georgia Institute of Technology  
Atlanta, GA 30332-0160, USA  
bulletin@iamp.org