Reply to "Comment on 'Not all counterclockwise thermodynamic cycles are refrigerators'" [Am. J. Phys. 85, 861–863 (2017)]

R. H. Dickerson, and J. Mottmann

Citation: American Journal of Physics **85**, 864 (2017); doi: 10.1119/1.5005929
View online: http://dx.doi.org/10.1119/1.5005929
View Table of Contents: http://aapt.scitation.org/toc/ajp/85/11
Published by the American Association of Physics Teachers

Comment on "Not all counterclockwise thermodynamic cycles are refrigerators" [Am. J. Phys. 84, 413–418 (2016)]

João P. S. Bizarro[a)]
Instituto de Plasmas e Fusão Nuclear, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisboa, Portugal

(Received 17 June 2016; accepted 17 September 2017)

Contrary to what Dickerson and Mottmann [Am. J. Phys. **84**, 413–418 (2016)] state, the temperatures at which a refrigerator's working fluid absorbs heat need not always lie below those at which it expels heat; nor must a refrigerator's thermodynamic cycle have two adiabats. Moreover, what Dickerson and Mottmann call a "comparison Carnot cycle" cannot always be defined. These conclusions are illustrated here using a counter-clockwise Stirling cycle without regeneration. A refrigerator's cold reservoir can absorb some heat and its hot reservoir can expel some heat, so long as the *net* heat flow is still out of the cold reservoir and into the hot reservoir.

https://doi.org/10.1119/1.5005928

I. INTRODUCTION

In a recent article in this journal,[1] Dickerson and Mottmann (D&M) stressed the fact that not all counterclockwise (CCW) thermodynamic cycles represent refrigerators. However, some of the criteria that D&M suggested to be necessary for a working refrigerator are too stringent. Specifically, it is not true that the range of temperatures over which the working fluid absorbs heat must be entirely below the range of temperatures over which it expels heat; nor is it true that a refrigeration cycle must employ two adiabats. As explained below, a simple Stirling refrigerator (i.e., one employing no heat regeneration) provides a counter-example to both of these criteria.[2]

The flaw in D&M's analysis is the assumption that a refrigerator's working fluid can extract heat *only* from the cold reservoir and reject heat *only* to the hot reservoir. The correct statement, instead, is that over the course of a full cycle the cold reservoir must have given up a *net* amount of heat (and therefore, necessarily, the hot reservoir must have absorbed a *net* amount of heat). Much of the analysis in D&M's paper remains valid, and it is especially important to understand their point that in many refrigeration cycles the extreme temperatures are not the same as (or even close to) the reservoir temperatures.

II. THE CCW STIRLING CYCLE

Consider the quasistatic CCW Stirling cycle shown in Fig. 1. The working fluid is n moles of an ideal gas, with a constant molar specific heat C_V. Its volume varies from $V_{\min}$ to $V_{\max}$ and it is alternately in thermal contact with reservoirs at temperatures $\mathcal{T}_{\text{in}}$ and $\mathcal{T}_{\text{out}}$ (there is no regenerator). The cycle consists of four steps, with heat transfers that can be written in terms of the preceding variables and the gas constant R:

- $a \to b$, an isothermal expansion from $V_{\min}$ to $V_{\max}$ with the fluid in contact with the cold reservoir, from which heat $Q^*_{\text{in}} = nR\mathcal{T}_{\text{in}}\ln(V_{\max}/V_{\min})$ is extracted;
- $b \to c$, an isochoric heating from $\mathcal{T}_{\text{in}}$ to $\mathcal{T}_{\text{out}}$ with the fluid in contact with the hot reservoir, from which heat $Q^* = nC_V(\mathcal{T}_{\text{out}} - \mathcal{T}_{\text{in}})$ is extracted;
- $c \to d$, an isothermal compression from $V_{\max}$ to $V_{\min}$ with the fluid in contact with the hot reservoir, to which heat $Q^*_{\text{out}} = nR\mathcal{T}_{\text{out}}\ln(V_{\max}/V_{\min})$ is rejected; and
- $d \to a$, an isochoric cooling from $\mathcal{T}_{\text{out}}$ to $\mathcal{T}_{\text{in}}$ with the fluid in contact with the cold reservoir, to which heat $Q^* = nC_V(\mathcal{T}_{\text{out}} - \mathcal{T}_{\text{in}})$ is rejected.

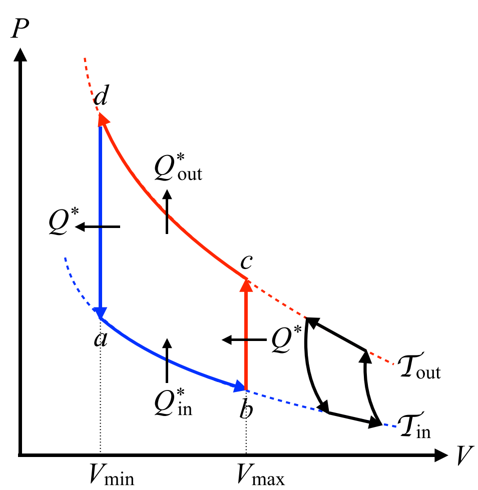


Fig. 1. A CCW Stirling cycle. The working fluid is in contact with the cold reservoir along $d \to a \to b$, and with the hot reservoir along $b \to c \to d$. The dashed lines are continuations of the two isotherms. Also sketched to the right is a CCW "comparison Carnot cycle" that utilizes the same two reservoirs.

Note that the reservoirs with which the fluid is in contact during the isochors $b \to c$ and $d \to a$ are such that heat spontaneously "flows downhill," consistently with the second law.[3]

The *net* amount of heat extracted from the cold reservoir is $Q_{\rm in} = Q^*_{\rm in} - Q^*$, while the *net* amount of heat rejected to the hot reservoir is $Q_{\rm out} = Q^*_{\rm out} - Q^*$, so the coefficient of performance (COP) for the CCW Stirling cycle is

$$K_{\rm Stirling} = \frac{Q_{\rm in}}{W} = \frac{Q_{\rm in}}{Q_{\rm out} - Q_{\rm in}} = \frac{Q^*_{\rm in} - Q^*}{Q^*_{\rm out} - Q^*_{\rm in}} = \left(1 - \frac{Q^*}{Q^*_{\rm in}}\right) K_{\rm Carnot}, \tag{1}$$

where W is the net work delivered to the fluid in a cycle and $K_{\rm Carnot} = \mathcal{T}_{\rm in}/(\mathcal{T}_{\rm out} - \mathcal{T}_{\rm in})$ is the COP for a CCW Carnot cycle operating between the same reservoir temperatures (also sketched in Fig. 1). Note also that $\mathcal{T}_{\rm out} > \mathcal{T}_{\rm in}$ entails $Q^*_{\rm out} > Q^*_{\rm in}$ and that, since $Q^*_{\rm out}$ and $Q^*_{\rm in}$ obey the Carnot-like relation $Q^*_{\rm out}/\mathcal{T}_{\rm out} = Q^*_{\rm in}/\mathcal{T}_{\rm in}$, the net entropy produced in a single cycle can be written

$$\Delta S_{\rm Stirling} = \left(\frac{1}{\mathcal{T}_{\rm in}} - \frac{1}{\mathcal{T}_{\rm out}}\right) Q^*. \tag{2}$$

This last result clearly shows that irreversibility in a quasi-static Stirling cycle arises from heat exchanges over finite temperature differences, and corresponds to a net transfer of heat from $\mathcal{T}_{\rm out}$ to $\mathcal{T}_{\rm in}$. This effect can be viewed as the equivalent of a heat leak between the two reservoirs and would vanish if the Stirling refrigerator used a regenerator (so that Q^* would not have to be moved from the hot to the cold reservoir).

As a refrigerator, the CCW Stirling cycle has to effectively extract heat from the cold reservoir $\mathcal{T}_{\rm in}$ (i.e., the "freezer") while rejecting heat to the hot reservoir $\mathcal{T}_{\rm out}$ (i.e., the "kitchen"). Hence $Q_{\rm in} > 0$, or $Q^*_{\rm in} > Q^*$, must apply, which implies $Q^*_{\rm out} > Q^*$, or $Q_{\rm out} > 0$, ensuring not only that $K_{\rm Stirling} > 0$ but also that $K_{\rm Stirling} < K_{\rm Carnot}$, in accordance with the second law. This CCW Stirling cycle represents a refrigerator where neither are the temperatures over which the fluid absorbs heat entirely below the temperatures over which it rejects heat, nor are there any adiabats, thus contradicting D&M.

If $Q^* > Q^*_{\rm out}$ instead of $Q^*_{\rm in} > Q^*$, so $Q^* > Q^*_{\rm in}$ also, then $Q_{\rm in} < 0$ and $Q_{\rm out} < 0$ both follow and the CCW Stirling cycle no longer represents a refrigerator but is instead what D&M refer to as a "cold pump," a device that uses work to help move heat from a hot to a cold reservoir.[4]

In the intermediate case $Q^*_{\rm out} \geq Q^* \geq Q^*_{\rm in}$, so $Q_{\rm in} \leq 0$ and $Q_{\rm out} \geq 0$, the CCW Stirling cycle describes neither a refrigerator nor a "cold pump", but a device that takes work and delivers heat to both reservoirs (or to a single reservoir if one of the equalities applies). This device, which converts work integrally into heat (i.e., a heater), can be named a "Joule pump" in memory of Joule's famous paddle-wheel experiment.[5,6] D&M's statement that all CCW cycles can be divided into two categories, refrigerators and "cold pumps", is thus contradicted, as has been reported previously.[2] Note that the CCW Stirling "Joule pump" does not admit what D&M call a "comparison Carnot cycle," as none of the reservoirs has heat extracted from it in this mode of operation (i.e., there is no freezer). The usefulness of a comparison Carnot cycle to distinguish between the different types of devices represented by CCW cycles is thus limited, inasmuch as the net heat flows to and from the reservoirs have to be worked out (which immediately identifies the device) before one can ascertain the direction of the "comparison Carnot cycle".

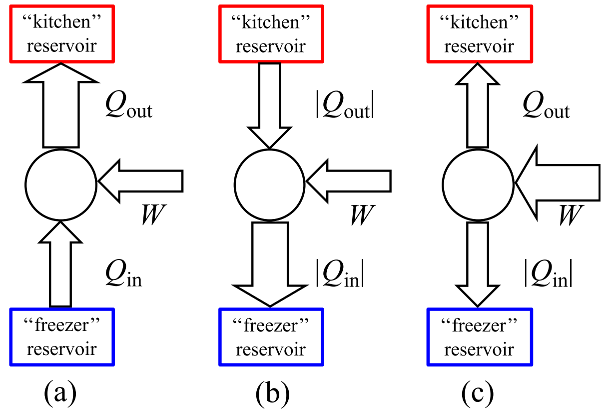


Fig. 2. Three possible modes of operation of a CCW Stirling cycle: (a) a refrigerator or heat pump, (b) a "cold pump", and (c) a "Joule pump".

Figure 2 summarizes the three possible modes of operation of a CCW Stirling cycle, i.e., $Q^*_{\rm in} > Q^*$ (refrigerator or heat pump), $Q^* > Q^*_{\rm out}$ ("cold pump"), and $Q^*_{\rm out} \geq Q^* \geq Q^*_{\rm in}$ ("Joule pump").[7] All three modes are consistent with the first and second laws, and all three can be achieved through different choices of the temperature ratio $\mathcal{T}_{\rm out}/\mathcal{T}_{\rm in}$ and the expansion ratio $V_{\rm max}/V_{\rm min}$. For example, if $\mathcal{T}_{\rm out} = 2\mathcal{T}_{\rm in}$ and $C_V = 5R/2$, the cycle represents a refrigerator if $\ln(V_{\rm max}/V_{\rm min}) > 5/2$, a "cold pump" if $\ln(V_{\rm max}/V_{\rm min}) < 5/4$, and a "Joule pump" at intermediate values of the expansion ratio. All three cases are represented by a diagram similar to that of Fig. 1, with varying isotherm locations and lengths. Note that, with $K_{\rm Stirling} \leq 0$ for "cold pumps" and "Joule pumps", the COP condition $0 < K_{\rm Stirling} < K_{\rm Carnot}$ provides a solid, quantitative criterion to discriminate between Stirling refrigerators and the other CCW Stirling cycles. Moreover, a similar analysis, with C_V replaced by $C_P = C_V + R$, can be applied to the CCW Ericsson cycle (formed by two isotherms and two isobars).

III. REFRIGERATION CYCLES IN GENERAL

CCW cycles with two adiabats or two isotherms can represent refrigerators, examples of the former being the CCW Otto, Diesel, Brayton, and Atkinson cycles, and of the latter the CCW Stirling and Ericsson cycles.[2] However, neither adiabats nor isotherms are a necessity for refrigeration. In fact, there is no need for the expansion process $a \to b$ in Fig. 1 to be isothermal, as long as the temperature of the working fluid does not go above $\mathcal{T}_{\rm in}$ and the amount of heat $đQ$ absorbed by the fluid at each point is never negative, thus ensuring heat does "flow downhill" and from the "freezer" into the fluid. For n moles of an ideal gas with constant specific-heat ratio $\gamma = C_P/C_V$, these two conditions imply $PV/nR \leq \mathcal{T}_{\rm in}$ and $dP/dV \geq -\gamma P/V$ for the corresponding curve $P(V)$. The last inequality means that the slope of such a $P(V)$ curve at each point never lies below that of the adiabat through the same point, and stems from imposing[8–11]

$$\text{đ}Q = nC_V\,dT + P\,dV = \frac{C_V}{R}(\gamma P\,dV + V\,dP) \geq 0. \qquad (3)$$

A similar discussion applies, *mutatis mutandi*, to the compression process $c \to d$, during which $PV/nR \geq \mathcal{T}_{\rm out}$ and $dP/dV \geq -\gamma P/V$ must simultaneously hold, if heat is indeed to be transferred from the fluid to the "kitchen".[12]

Each of the two curves replacing, slightly below or above, the isotherms in such a modified CCW Stirling cycle has their endpoints at the same temperature, so the respective change in internal energy is zero and the work delivered or received by the ideal gas during the process must equal the heat absorbed from, or expelled to the cold or hot reservoir, respectively. For the expansion curve lying below the $\mathcal{T}_{\rm in}$ isotherm, but joining it at the endpoints, the heat extracted from the "freezer" (the area under the $a \to b$ curve) is less than $Q^*_{\rm in}$ in the CCW Stirling cycle, while the total work input (the area enclosed by the cycle) is larger, thus leading to a COP for the modified cycle smaller than the original $K_{\rm Stirling}$. Although it is less efficient than the original CCW Stirling cycle, the modified CCW cycle just described is indeed a refrigerator, but has neither adiabats (once again contradicting D&M) nor isotherms. For instance, the isotherms in Fig. 1 can be replaced with stepwise, stair-like $P(V)$ curves formed by a succession of alternating short adiabats and isochors.

Refrigeration (or heat pumping) requires that, in a cyclic manner, a working fluid be put in thermal contact with a "freezer" and a "kitchen", from and to which, respectively, it must extract and give off heat. Suppose the fluid has just finished expelling heat to the "kitchen" and must come subsequently into contact with the "freezer"; there are many ways of doing this, as long as it is guaranteed that the fluid does absorb some heat $Q^*_{\rm in}$ from the "freezer" somewhere along the process. For the rest of the process the "freezer" may eventually draw heat from the fluid in some quantity $Q^*_{\rm cold}$ ($Q^*_{\rm cold} = Q^*$ in Fig. 1), provided $Q^*_{\rm cold} < Q^*_{\rm in}$ so the *net* heat rejected to the fluid is $Q_{\rm in} = Q^*_{\rm in} - Q^*_{\rm cold} > 0$. Similarly when, after extracting heat from the "freezer", the fluid comes into contact with the "kitchen": during this process the fluid must effectively expel some heat $Q^*_{\rm out}$ to the "kitchen", which may also give off some heat $Q^*_{\rm hot}$ ($Q^*_{\rm hot} = Q^*$ in Fig. 1), as long as the *net* amount of heat absorbed from the fluid obeys $Q_{\rm out} = Q^*_{\rm out} - Q^*_{\rm hot} > 0$. In addition, it must be ensured that, when $Q^*_{\rm in}$ and $Q^*_{\rm out}$ are exchanged between fluid and reservoirs, they are transferred in the appropriate direction, implying for the corresponding thermodynamic processes $dP/dV \geq -\gamma P/V$, in the case of an ideal gas.

The CCW Otto, Diesel, Brayton, and Atkinson cycles,[2] which operate with two adiabats (as too restrictively required by D&M), correspond to putting $Q^*_{\rm cold} = Q^*_{\rm hot} = 0$ and letting $Q^*_{\rm in}$ and $Q^*_{\rm out}$ be exchanged along isochors or isobars, whose slopes are, respectively, $dP/dV = \infty$ and $dP/dV = 0$, both larger than the ideal-gas adiabat slope $dP/dV = -\gamma P/V$. In particular, the CCW Otto cycle (formed by two adiabats and two isochors and studied in detail by D&M) corresponds to $P(V)$ curves whose slopes are precisely the extreme values allowed by the condition $dP/dV \geq -\gamma P/V$. If $Q^*_{\rm cold}$ and $Q^*_{\rm hot}$ do not vanish and are transferred isochorically or isobarically, while the fluid temperature is pushed to the maximum and minimum values that are still consistent with $Q^*_{\rm in}$ and $Q^*_{\rm out}$ "flowing downhill" (which implies isothermal heat transfers at $\mathcal{T}_{\rm in}$ and $\mathcal{T}_{\rm out}$), the CCW Stirling and Ericsson cycles follow.[2]

IV. SUMMARY AND CONCLUSIONS

The analysis by D&M[1] on the ability of CCW thermodynamic cycles to describe actual refrigerators has been extended here to include cycles that do not employ two adiabats and where the temperatures at which heat is expelled by the working fluid do not lie entirely above those at which heat is absorbed. The concept of a "comparison Carnot cycle" has also been shown to be of no great use in distinguishing between the different types of devices described by CCW cycles. Basically, all that is needed for a CCW cycle to represent a refrigerator, a device that keeps the "freezer" (i.e., the reservoir from which a *net* amount of heat is absorbed during the cycle) cooler than the "kitchen" (i.e., the reservoir to which a *net* amount of heat is rejected during the cycle), is that its COP be physically meaningful (i.e., positive and not greater than the COP of the Carnot cycle operating between the same two reservoirs).

ACKNOWLEDGMENTS

The author acknowledges R. H. Dickerson and J. Mottmann for the discussions that followed the submission of this Comment, and is indebted to the anonymous reviewers, whose criticism and suggestions have greatly improved the manuscript. J. S. Ferreira helped to prepare Figs. 1 and 2. This work received financial support from the Fundação para a Ciência e a Tecnologia (FCT, Lisboa) through project No. UID/FIS/50010/2013. The views and opinions expressed herein do not necessarily reflect those of FCT, of IST, or of their services.

[a)]Electronic mail: bizarro@ipfn.tecnico.ulisboa.pt

[1]R. H. Dickerson and J. Mottmann, "Not all counterclockwise thermodynamic cycles are refrigerators," Am. J. Phys. **84**, 413–418 (2016).

[2]The CCW Stirling cycle has also been analyzed by M. F. F. da Silva, "Some considerations about thermodynamic cycles," Eur. J. Phys. **33**, 13–42 (2012).

[3]Within the framework of classical thermodynamics, heat "flowing downhill" includes the marginal situation of heat exchange along isothermals in which the working fluid and reservoir are at the same temperature. The only situation strictly forbidden by the second law is that of heat "flowing uphill."

[4]A "cold pump" uses work to do something that nature does for free, so its COP $|Q_{\rm in}|/W$ or $|Q_{\rm out}|/W$ (whether it is used to heat the cold reservoir or cool the hot reservoir) is finite for $W > 0$ and becomes infinite for $W = 0$.

[5]F. Reif, *Fundamentals of Statistical and Thermal Physics* (McGraw-Hill, New York, 1981), pp. 184–186.

[6]A "Joule pump" merely converts work into heat, so its COP $(|Q_{\rm in}| + Q_{\rm out})/W$ never goes above unity.

[7]See also Fig. 4 of Ref. 2.

[8]R. H. Dickerson and J. Mottmann, "On the thermodynamic efficiencies of reversible cycles with sloping, straight-line processes," Am. J. Phys. **62**, 558–562 (1994).

[9]See Fig. 7 of Ref. 8.

[10]Note the distinction between slope and steepness, respectively, dP/dV and $|dP/dV|$: for instance, wherever the curve $P(V)$ replacing the isotherm $a \to b$ in Fig. 1 has a negative slope, the condition $dP/dV \geq -\gamma P/V$ does not allow it to be steeper than the local adiabat, but it is steeper than the latter for positive slopes larger than $\gamma P/V$.

[11]In points of the $P(V)$ curve where the fluid temperature is not increasing, its slope must also not go above that of the local isotherm, meaning $dP/dV \leq -P/V$.

[12]Note that during compression one has $dV < 0$, which is to be taken into account when working out the condition $\text{đ}Q \leq 0$.

Reply to "Comment on 'Not all counterclockwise thermodynamic cycles are refrigerators'" [Am. J. Phys. 85, 861–863 (2017)]

R. H. Dickerson[a)] and J. Mottmann
Department of Physics, California Polytechnic State University, San Luis Obispo, California 93407

(Received 25 July 2017; accepted 17 September 2017)
https://doi.org/10.1119/1.5005929

I. INTRODUCTION

Our paper[1] discussed the conditions for which a theoretical counterclockwise (ccw) and inherently irreversible cycle could be a refrigerator. We found three necessary and self-consistent criteria for refrigeration. One criterion required that all heat absorption occurred at temperatures below those at which heat was emitted. In addition we also *speculated* that refrigerators required two adiabatic segments, although by themselves adiabats were not a sufficient condition. Dr. Bizarro's Comment[2] presents a ccw Stirling cycle as a counterexample to our findings.

We maintain that the ccw Stirling cycle cannot function as a viable refrigerator since it violates one of our criteria mentioned earlier, and has no adiabats. The Stirling cycle has two isotherms and two isochors, as pictured in Fig. 1. As we explain below, it is Bizarro's handling of the isothermal segments that is the primary source of an error that invalidates his analysis and conclusions. In Sec. III, we discuss isothermal processes in some detail. There will be no need to do complex calculations. We never do more than apply the simplest form of the second law: heat flows from hot to cold.

II. BIZARRO'S ANALYSIS

In the discussion that follows we distinguish between T (the temperature of a cycle's working substance) and $\mathcal{T}$ (the temperature of a heat reservoir). Figure 1 is a PV diagram of a ccw Stirling cycle. The labels for the various cycle segments are those from Bizarro's Comment. The cycle is irreversible and operates with just two heat reservoirs, which

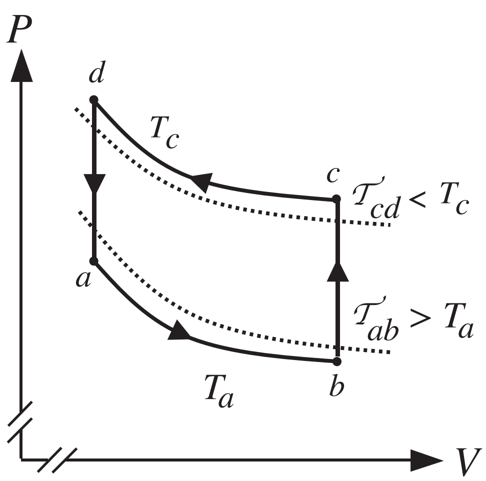


Fig. 1. A PV diagram of a ccw Stirling cycle. Temperatures T_a and T_c are those of the working substance while undergoing isothermal processes. The dotted isotherms at $\mathcal{T}_{ab}$ and $\mathcal{T}_{cd}$ represent the reservoir temperatures as envisioned by Bizarro.

Bizarro asserts are at temperatures T_c and T_a. Figure 1 includes as dotted lines the isotherms of the heat reservoirs at temperatures $\mathcal{T}_{ab} = T_a$ and $\mathcal{T}_{cd} = T_c$.

Bizarro claims that during the isothermal compression $c \to d$ the working substance at temperature T_c ejects heat into a reservoir at the exact same temperature T_c. During the isothermal expansion $a \to b$ the working substance at temperature T_a absorbs heat from a reservoir also at T_a.

A refrigerator's basic function is to move heat from a "cold place" into a "hot place." Bizarro's analysis achieves this since according to him the working substance absorbs heat at low temperature T_a and ejects heat at high temperature T_c. But in the effort to make the Stirling a viable refrigeration cycle Bizarro also requires that each of the reservoirs at T_a and T_c must both absorb and eject heat during various parts of the cycle. The absorption and ejection of heat by the same reservoir is a novel but necessary requirement of the Comment.

Dr. Bizarro's analysis is incorrect. He employs misconceptions about isothermal processes that unfortunately may also be accepted by others. We discuss this below.

III. ISOTHERMS

Consider first the isothermal expansion $a \to b$. It is vital for Dr. Bizarro's analysis that $\mathcal{T}_{ab}$ equals T_a. An endnote in the Comment attempts to justify $\mathcal{T}_{ab} = T_a$: "Within the framework of classical thermodynamics, heat 'flowing downhill' includes the marginal situation of heat exchange along isothermals in which the working fluid and reservoir are at the same temperature. The only situation strictly forbidden by the second law is that of heat 'flowing uphill.' " This statement is clearly wrong; neither 'downhill' nor 'uphill' exists if the two objects are at equal temperatures. The second law also prohibits heat exchange between objects in thermal equilibrium.

Compare the above quote to a statement by Marcella (emphasis added): "In an isothermal expansion the system at constant temperature T absorbs an amount of heat Q from its environment, *necessarily at a higher temperature*."[3] This unambiguous statement makes clear that heat flows from hot to cold and thus, during the isothermal process $a \to b$, heat must come from a reservoir warmer than T_a. The appropriate value for $\mathcal{T}_{ab}$ is then $\mathcal{T}_{ab} = T_a + dT > T_a$, for some positive difference dT, and "the transfer of heat approaches reversibility only if the temperature difference is infinitesimal."[3] A temperature difference, even if infinitesimal, is always mandatory if there is to be any heat exchange. Figure 2 shows an isotherm for a reservoir at temperature $\mathcal{T}_{ab} > T_a$ as required by the second law.

Dr. Bizarro also requires that the working substance and reservoir be at identical temperatures during the isothermal compression $c \to d$. With apologies to Marcella we adapt his quote given earlier: "In an isothermal compression the system at constant temperature T ejects an amount of heat Q

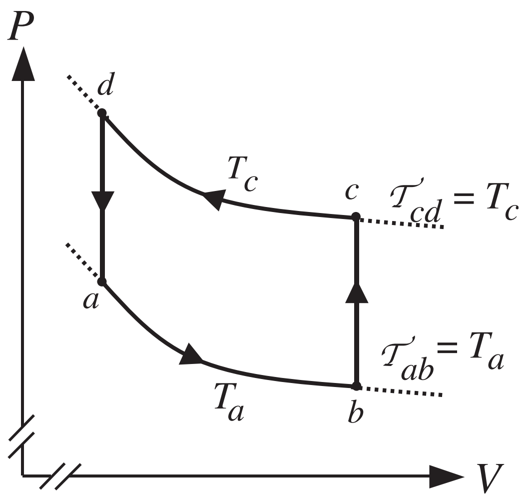


Fig. 2. The Stirling cycle of Fig. 1, with reservoir temperatures modified to make the isothermal processes physically possible. However, these reservoirs cannot account for the isochoric segments.

into its environment, necessarily at a lower temperature." The reservoir must be colder than the working substance if it absorbs heat: $\mathcal{T}_{cd} < T_c$.

IV. THE STIRLING CYCLE

Figure 2 shows the isotherms of the two reservoirs at $\mathcal{T}_{ab} > T_a$ and $\mathcal{T}_{cd} < T_c$ that are mandatory if the working substance is to absorb and eject heat as desired during the isothermal steps. The required value for $\mathcal{T}_{ab}$ is $T_{ab} = T_a + \mathrm{d}T$ while the required value for $\mathcal{T}_{cd}$ is $\mathcal{T}_{cd} = T_c - \mathrm{d}T$. But these temperature limitations bring about a severe complication, since the two reservoirs cannot account for how the working substance actually gets up to temperature T_c above $\mathcal{T}_{cd}$ or down to temperature T_a below $\mathcal{T}_{ab}$.

The isochoric path $b \to c$ cannot be accomplished by the reservoir at temperature $\mathcal{T}_{cd}$ since that reservoir is not hot enough. In order for the working substance to reach T_c isochorically, it would have to absorb heat from a reservoir at some temperature $\mathcal{T}_{bc} > T_c$. Similarly, the isochoric path $d \to a$ cannot be accomplished by the reservoir at temperature $\mathcal{T}_{ab}$, because it is not cold enough. For the working substance to reach T_a isochorically, there must be a reservoir at some temperature $\mathcal{T}_{da} < T_a$.

We have already seen that the isothermal processes plotted in Fig. 1 violate the second law. Yet if we modify the reservoir temperatures slightly, in order to fix this problem, then the isochoric processes violate the second law. There is no way to set the two reservoir temperatures to avoid both of these problems, and therefore Bizarro's analysis fails. A ccw Stirling cycle cannot function as a refrigerator.

Perhaps Dr. Bizarro might be interested in pursuing an idea that was suggested to us by a reviewer. From Fig. 2, we know that the isochoric path $b \to c$ cannot be fully accomplished by the reservoir at $\mathcal{T}_{cd}$ since it is not hot enough. However, in order to reach T_c one could insert an adiabatic process in place of the sharp corner at c. This can be accomplished by ending the isochor slightly below $\mathcal{T}_{cd}$ and continuing up to T_c via a short adiabat. The sharp corner at a can be similarly replaced by another short adiabat. These connecting adiabatic segments can be made arbitrarily short and thus the shape of the cycle closely mimics that of a Stirling. These connecting adiabats would eliminate the need for reservoirs above T_c and below T_a.

V. CONCLUSION

Complex calculations and complicated logic cannot overcome a basic property of the universe: heat flows from hot to cold. During an isothermal process the working substance has a constant temperature by definition, but the term *isothermal* does not imply that the working substance and its thermally connected reservoir are at identical temperatures. Bizarro's analysis requires that the working substance and connected reservoir have identical temperatures during parts of the cycle, thus violating the second law. Since this is impossible, it follows that much in the Comment is likewise incorrect, or at least suspect.

[a]Electronic mail: rdickers@calpoly.edu

[1]R. H. Dickerson and J. Mottmann, "Not all counterclockwise thermodynamic cycles are refrigerators," Am. J. Phys. **84**(6), 413–418 (2016).

[2]J. P. S. Bizarro, "Comment on 'Not all counterclockwise thermodynamic cycles are refrigerators,'" Am. J. Phys. **85**, 861–863 (2017).

[3]T. V. Marcella, "Entropy production and the second law of thermodynamics: An introduction to second law analysis," Am. J. Phys. **60**(10), 888–890 (1992).

Three new roads to the Planck scale

Valerio Faraoni[a)]

Physics Department, Bishop's University, 2600 College Street, Sherbrooke, Québec J1M 1Z7, Canada

(Received 18 October 2016; accepted 25 May 2017)

Three new heuristic derivations of the Planck scale are described. They are based on basic principles or phenomena of relativistic gravity and quantum physics. The Planck scale quantities thus obtained are within one order of magnitude of the "standard" ones. We contemplate the pair creation of causal bubbles so small that they can be treated as particles, the scattering of a matter wave off the background curvature of spacetime that it induces, and the Hawking evaporation of a black hole in a single burst at the Planck scale.

https://doi.org/10.1119/1.4994804

I. INTRODUCTION

General relativity and quantum mechanics are two great achievements of twentieth century physics. Gravity is completely classical in Einstein's theory of general relativity, and quantum mechanics (broadly defined to include quantum field theory and particle physics) incorporates special relativity but excludes gravity. It is believed that these two completely separate theories should merge at the Planck scale, at which general-relativistic effects become comparable to quantum ones. No definitive theory of quantum gravity is available, although much work has gone into string theories, loop quantum gravity, and other approaches (e.g., Refs. 1–4, see also Ref. 5, and see Ref. 6 for a popular exposition).

The Planck scale was introduced by Planck himself[7] in 1899, therefore predating the Planck law for blackbody radiation. The importance of the Planck units was realized by Eddington[8] and the idea that gravitation and quantum mechanics should be taken into account simultaneously at this scale was spread by Wheeler[9,10] and has bounced around ever since. The themes that a fundamental system of units exists in nature and that the values of these units can perhaps be derived in a super-theory have been the subject of a large literature (see Ref. 11 for an excellent introduction).

All derivations of the Planck scale more or less correspond to taking various combinations of the fundamental constants G (Newton's constant) associated with gravity, c (the speed of light *in vacuo*) characterizing relativity, and the Planck constant h [or the reduced Planck constant $\hbar \equiv h/(2\pi)$] which signals quantum mechanics. Usually the Planck scale is deduced, following Planck, on a purely dimensional basis[7] or it is derived using the concept of a black hole in conjunction with that of a matter wave. The simplest derivation of the Planck scale notes that by combining the three fundamental constants G, c, and $\hbar$ one obtains a unique quantity with the dimensions of a length, the Planck length

$$l_{\rm pl} = \sqrt{\frac{G\hbar}{c^3}} = 1.6 \times 10^{-35}\,{\rm m}\,. \quad (1)$$

By combining $l_{\rm pl}$ with G and c one then obtains the Planck mass

$$m_{\rm pl} = \frac{l_{\rm pl} c^2}{G} = \sqrt{\frac{\hbar c}{G}} = 2.2 \times 10^{-8}\,{\rm kg}\,, \quad (2)$$

the Planck energy

$$E_{\rm pl} = m_{\rm pl} c^2 = \sqrt{\frac{\hbar c^5}{G}} = 1.3 \times 10^{19}\,{\rm GeV}\,, \quad (3)$$

the Planck mass density

$$\rho_{\rm pl} = \frac{m_{\rm pl}}{l_{\rm pl}^3} = \frac{c^2}{l_{\rm pl}^2 G} = \frac{c^5}{\hbar G^2} = 5.2 \times 10^{96}\,{\rm kg\,m^{-3}}\,, \quad (4)$$

and the Planck temperature

$$T_{\rm pl} = \frac{E_{\rm pl}}{k_{\rm B}} = \frac{l_{\rm pl} c^4}{k_{\rm B} G} = \sqrt{\frac{\hbar c^5}{G k_{\rm B}^2}} = 1.4 \times 10^{32}\,{\rm K}\,, \quad (5)$$

where $k_{\rm B}$ is the Boltzmann constant. We denote with $x_{\rm pl}$ the Planck scale value of a quantity x determined by dimensional analysis as in the above. Two suggestive alternative derivations of the Planck scale appear in the literature and are reviewed in Secs. I A and I B. At least six more roads to the Planck scale, which are slightly more complicated, are known and have been discussed in Ref. 15. How many ways to obtain the Planck scale without a full quantum gravity theory are possible? The challenge of finding them can be fun and very creative. Other possibilities to heuristically derive the Planck scale certainly exist: in Secs. II–IV we propose three new ones based on pair creation of "particle-universes," the propagation of matter waves on a curved spacetime, or the Hawking radiation from black holes.

A. A Planck size black hole

In what is probably the most popular derivation of the Planck scale, one postulates that a particle of mass m and Compton wavelength $\lambda = h/(mc)$, which has Planck energy, collapses to a black hole of radius $R_{\rm S} = 2Gm/c^2$ (the Schwarzschild radius of a spherical static black hole of mass m (Refs. 12 and 13)). Like all orders of magnitude estimates, this procedure is not rigorous since it extrapolates the concepts of black hole and of Compton wavelength to a new regime in which both concepts would probably lose their accepted meanings and would, strictly speaking, cease being valid. However, this is how one gains intuition into a new physical regime.

Equating the Compton wavelength of this mass m to its black hole radius gives

$$m = \sqrt{\frac{hc}{2G}} = \sqrt{\pi}\, m_{\rm pl} \simeq 1.77\, m_{\rm pl}\,. \quad (6)$$

B. A universe of size comparable with its Compton wavelength

It is not compulsory to restrict to black holes in heuristic derivations of the Planck scale, although black holes certainly constitute some of the most characteristic phenomena predicted by relativistic gravity.[12,13] Why not use a relativistic universe instead of a black hole? This approach is followed in the following argument proposed in John Barrow's *Book of Universes*[16] (but it does not appear in the technical literature and it definitely deserves to be included in the pedagogical literature).

Cosmology can only be described in a fully consistent and general way by a relativistic theory of gravity and one can rightly regard a description of the universe as phenomenology of relativistic gravity on par with the prediction of black holes. Consider a spatially homogeneous and isotropic universe which, for simplicity, will be taken to be a spatially flat Friedmann-Lemaître-Robertson-Walker spacetime with line element

$$ds^2 = -dt^2 + a^2(t)(dx^2 + dy^2 + dz^2)\,, \quad (7)$$

and with scale factor $a(t)$ and Hubble parameter $H(t) \equiv \dot{a}/a$. An overdot denotes differentiation with respect to the comoving time t measured by observers who see the 3-space around them homogeneous and isotropic. The size of the observable universe is its Hubble radius cH^{-1} which is also, in order of magnitude, the radius of curvature (in the sense of four-dimensional curvature) of this space. Consider the mass m enclosed in a Hubble sphere, given by

$$mc^2 = \frac{4\pi}{3}\,\rho\left(H^{-1}c\right)^3 = \frac{H^{-1}c^5}{2G}\,, \tag{8}$$

where ρ is the cosmological energy density and in the last equality we used the Friedmann equation[12,13]

$$H^2 = \frac{8\pi G}{3c^2}\,\rho \tag{9}$$

(note that, following standard notation, $\rho_{\rm pl}$ and ρ denote a *mass density* and an *energy density*, respectively). The Planck scale is reached when the Compton wavelength of the mass m is comparable with the Hubble radius, i.e., when

$$\frac{c}{H} \sim \lambda = \frac{h}{mc}\,. \tag{10}$$

This procedure implies that quantum effects (Compton wavelength) are of the same order of gravitational effects (cosmology described by the Friedmann equation). Clearly, we extrapolate Eq. (9) to a new quantum gravity regime from the realm of validity of general relativity and we extrapolate the concept of Compton wavelength from the realm of ordinary quantum mechanics. This extrapolation is necessary in order to learn something about the Planck scale, although it is not rigorous.

The expression (8) of m then gives

$$H^2 = \frac{c^5}{2Gh}\,. \tag{11}$$

Using again Eq. (9) yields the energy density

$$\rho \sim \frac{3c^7}{16\pi G^2 h} = \frac{3c^2}{32\pi^2}\,\rho_{\rm pl} \simeq 10^{-2}c^2\rho_{\rm pl}\,, \tag{12}$$

from which the other Planck quantities (1)–(5) can be deduced by dimensional analysis. One obtains

$$l = \frac{c}{\sqrt{G\rho}} \simeq 10\, l_{\rm pl}\,, \tag{13}$$

$$m = \frac{lc^2}{G} \simeq 10\, m_{\rm pl}\,, \tag{14}$$

$$E = mc^2 \simeq 10\, E_{\rm pl}\,, \tag{15}$$

$$T = \frac{E}{k_{\rm B}} \simeq 10\, T_{\rm pl}\,. \tag{16}$$

At first sight, the argument of a universe with size comparable with its Compton wavelength is not dissimilar in spirit from the popular argument comparing the Schwarzschild radius of a black hole with its Compton wavelength. In fact, it is commonly remarked that the universe is a relativistic system by showing that the size of the observable universe is the same as the Schwarzschild radius of the mass m contained in it, for

$$R_{\rm S} = \frac{2Gm}{c^2} = \frac{2G}{c^2}\left(\frac{4\pi R^3}{3}\,\frac{\rho}{c^2}\right) = \frac{2G}{c^2}\,\frac{4\pi}{3}\,\frac{\rho}{c^2}\left(H^{-1}c\right)^3$$
$$= \frac{8\pi G}{3c}\,H^{-3}\rho\,. \tag{17}$$

Equation (9) then yields $R_{\rm S} \simeq cH^{-1}$, which is often reported in the popular science literature by saying that the universe is a giant black hole. This argument is definitely too naive because the Schwarzschild radius pertains to the Schwarzschild solution of the Einstein equations,[12,13] which is very different from the Friedmann-Lemaître-Robertson-Walker one. If one accepted this argument, then comparing the size of the visible universe cH^{-1} with the Compton wavelength of the mass contained in it would be numerically similar to comparing its Schwarzschild radius with this wavelength. However, the step describing the visible universe as a black hole (which is extremely questionable if not altogether incorrect) is logically not needed in the procedure expressed in Eq. (10).

Turning things around but in keeping with the spirit of the derivation above, it has also been noted that equating the Planck density to the density of a sphere containing the mass of the observable universe produces the size of the nucleus (or the pion Compton wavelength) as the radius of this sphere.[17]

II. PAIR CREATION OF PARTICLE-UNIVERSES

Another approach to the Planck scale is the following. The idea of a universe which is quantum-mechanical in nature has been present in the literature for a long time and the use of the uncertainty principle to argue something about the universe goes back to Tryon's 1973 proposal that the universe may have originated as a vacuum fluctuation.[18] This notion of creation features prominently also in recent popular literature.[19] Consider now universes so small that they are ruled by quantum mechanics and regard the mass-energies contained in them as elementary particles. At high energies, there could be production of pairs of such "particle-antiparticle universes." Again, one goes beyond known and explored regimes of general relativity and ordinary quantum mechanics by extrapolating facts well known in these regimes to the unknown Planck regime. The Heisenberg uncertainty principle $\Delta E \Delta t \geq \hbar/2$ can be used by assuming that ΔE is the energy contained in a Friedmann-Lemaître-Robertson-Walker causal bubble of radius $R \sim H^{-1}c$ containing the energy $\Delta E \simeq 4\pi\rho R^3/3$. Setting $\Delta t \sim H^{-1}$ (the age of this very young universe), $\Delta E \Delta t \simeq \hbar/2$ gives

$$\frac{4\pi}{3}\,\rho\left(H^{-1}c\right)^3 H^{-1} \simeq \frac{\hbar}{2} \tag{18}$$

which can be rewritten as

$$\frac{8\pi G}{3}\,\rho\,\frac{c^3}{GH^4} = \hbar\,. \tag{19}$$

Equation (9) then yields the mass density

$$\frac{\rho}{c^2} \simeq \frac{3c^5}{8\pi G^2\hbar} = \frac{3}{8\pi}\,\rho_{\rm pl}\,, \tag{20}$$

one order of magnitude smaller than the "standard" Planck mass density (4). The other Planckian quantities can then be derived from ρ and the fundamental constants G, c, and h.

III. SCATTERING OF A MATTER WAVE OFF THE BACKGROUND CURVATURE OF SPACETIME

The second alternative road to the Planck scale comes from the fact that, in general, waves propagating on a curved background spacetime scatter off it.[20–23] This phenomenon is well known and can be interpreted as if these waves had an effective mass induced by the spacetime curvature. It is experienced by waves with wavelength λ comparable with, or larger than, the radius of curvature L of spacetime. High frequency waves do not "feel" the larger scale inhomogeneities of the spacetime curvature and, as is intuitive, essentially propagate as if they were in flat spacetime.[12,21–23] The phenomenon is not dissimilar from the scattering experienced by a wave propagating through an inhomogeneous medium when its wavelength is comparable with the typical size of the inhomogeneities. Again, we extrapolate the backscattering of a test-field wave by the (fixed) background curvature of spacetime to a new regime in which this wave packet gravitates, bends spacetime and, at the Planck scale, impedes its own propagation. Clearly, this extrapolation is not rigorous, like all order of magnitude estimates. However, we can gain some confidence in this procedure *a posteriori* by noting that it produces a Planck scale of the same order of magnitude as that obtained by the other methods exposed here.

Consider now *a matter wave* associated with a particle of mass m and Compton wavelength $\lambda = h/(mc)$ scattering off the curvature of spacetime. The Planck scale can be pictured as that at which the spacetime curvature is caused by the mass m itself and the radius of curvature of spacetime due to this mass is comparable with the Compton wavelength. Essentially, high frequency waves do not backscatter but, at the Planck scale, there can be no waves shorter than the background curvature radius. Dimensionally, the length scale L associated with the mass m (the radius of curvature of spacetime) is given by $m = Lc^2/G$ and quantum and gravitational effects become comparable when $\lambda \sim L$, which gives

$$\frac{h}{(Lc^2/G)c} \sim L \tag{21}$$

or

$$L = \sqrt{\frac{Gh}{c^3}} = \sqrt{2\pi}\, l_{\rm pl} \simeq 2.51\, l_{\rm pl}\,. \tag{22}$$

In other words, if we pack enough energy into a matter wave so that it curves spacetime, the curvature induced by this wave will impede its own propagation when the Planck scale is reached. When the energy of this wave becomes too compact, the propagation of the matter wave is affected drastically.

IV. HAWKING EVAPORATION OF A BLACK HOLE IN A SINGLE BURST

Hawking's discovery that, quantum mechanically, black holes emit a thermal spectrum of radiation allowed for the development of black hole thermodynamics by assigning a non-zero temperature to black holes.[14] In the approximation of a fixed black hole background and of a test quantum field in this spacetime, a spherical static black hole of mass m emits a thermal spectrum at the Hawking temperature

$$T_{\rm H} = \frac{\hbar c^3}{8\pi G k_{\rm B} m}\,. \tag{23}$$

As is well known, the emitted radiation peaks at a wavelength $\lambda_{\max}$ larger than the horizon radius $R_{\rm S} = 2Gm/c^2$. In fact, using Wien's law of displacement for blackbodies

$$\lambda_{\max} T_{\rm H} = b = \frac{hc}{4.9651 k_{\rm B}} \simeq 2.8978 \times 10^{-3}\,{\rm m\,K} \tag{24}$$

and Eq. (23), one obtains

$$\lambda_{\max} = \frac{b}{T_{\rm H}} = \frac{8\pi^2}{4.9651}\,\frac{2Gm}{c^2} \simeq 15.90 R_{\rm S}\,. \tag{25}$$

Therefore, most of the thermal radiation is emitted at wavelengths comparable to, or larger than, the black hole horizon, giving a fuzzy image of the black hole.

Heuristically, one can extrapolate Hawking's prediction to a Planck regime in which the loss of energy is comparable with the black hole mass. Then the Planck scale is reached when the entire black hole mass m is radiated in a single burst of N particles of wavelength $\sim\lambda_{\max}$ and energy

$$E = \frac{hc}{\lambda_{\max}} \sim \frac{hc}{16 R_{\rm S}} = \frac{hc^3}{32 Gm}\,. \tag{26}$$

Although certainly not rigorous, this procedure provides a Planck scale of the same order of magnitude as the other procedures considered (which is all that one can expect from an order of magnitude estimate). Assuming N of order unity (say, $N=2$) and equating this energy with the black hole energy mc^2 yields

$$m \simeq \frac{NE}{c^2} \simeq \sqrt{\frac{hc}{16G}} = \sqrt{\frac{\pi}{8}}\, m_{\rm pl} \sim 0.627\, m_{\rm pl}\,. \tag{27}$$

V. DISCUSSION

Although black holes are a most striking prediction of Einstein's theory of gravity,[12,13] they do not constitute the entire phenomenology of general relativity and there is no need to limit oneself to the black hole concept in heuristic derivations of the Planck scale. One can consider cosmology as well, which is appropriate since cosmology can only be discussed in the context of relativistic gravity. This approach leads to Barrow's new heuristic derivation of the Planck scale[16] by considering, in a Friedmann-Lemaître-Robertson-Walker universe, a Hubble sphere with size comparable to the Compton wavelength of the mass it contains. Alternatively, one can consider the pair creation of causal bubbles so small that they can be treated as particles, or one can derive the Planck scale using the scattering of waves off the background curvature of spacetime which leads again, in order of magnitude, to the Planck scale when applied to matter waves. Alternatively, one can consider a black hole that evaporates completely in a single burst at the Planck scale. Of course, other approaches to the Planck scale are in principle conceivable. Although quantum gravity is certainly not a subject of undergraduate university courses, the exercise of imagining new heuristic avenues to the Planck scale can be fun and can stimulate the imagination of both undergraduate and graduate students, as well as being an exercise in dimensional analysis.

ACKNOWLEDGMENTS

The author is grateful to John Barrow for a discussion and for pointing out Ref. 8, and to two referees for helpful suggestions. This work was supported by Bishop's University and by the Natural Sciences and Engineering Research Council of Canada.

[a]Electronic mail: vfaraoni@ubishops.ca

[1]M. B. Green, G. H. Schwarz, and E. Witten, *Superstring Theory* (Cambridge U.P., Cambridge, UK, 1987).
[2]J. Polchinski, *String Theory* (Cambridge U.P., Cambridge, UK, 2005).
[3]C. Rovelli, *Quantum Gravity* (Cambridge U.P., Cambridge, UK, 2007).
[4]C. Kiefer, *Quantum Gravity* (Oxford U.P., Oxford, UK, 2004).
[5]D. Oriti, *Approaches to Quantum Gravity* (Cambridge U.P., Cambridge, UK, 2009).
[6]L. Smolin, *Three Roads to Quantum Gravity* (Weidenfeld & Nicolson, London, UK, 2000).
[7]M. Planck, "Ueber irreversible Strahlungsvorgänge," Sitzungsberichte der Koniglich Preussischen Akademie der Wissenschaften zu Berlin **5**, 440–480 (1899). Also as M. Planck, Ann. Phys. **11**, 69–122 (1900), translated in M. Planck, *The Theory of Heat Radiation*, translated by M. Masius (Dover, New York, 1959).
[8]A. S. Eddington, "Report on the relativity theory of gravitation," *Physical Society of London* (Fleetway Press, London, 1918).
[9]J. A. Wheeler, "Geons," Phys. Rev. **97**, 511–536 (1955).
[10]J. A. Wheeler, *Geometrodynamics* (Academic Press, New York and London, 1962).
[11]J. D. Barrow, *The Constants of Nature* (Pantheon Books, New York, 2002).
[12]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[13]R. M. Wald, *General Relativity* (Chicago U.P., Chicago, 1984).
[14]S. W. Hawking, "Particle creation by black holes," Comm. Math. Phys. **43**, 199–220 (1975); Erratum **46**, 206–206 (1976).
[15]R. J. Adler, "Six easy roads to the Planck scale," Am. J. Phys. **78**, 925–932 (2010).
[16]J. D. Barrow, *The Book of Universes* (W.W. Norton & C., New York, 2011), p. 185 and p. 260.
[17]N. A. Misnikova and B. N. Shvilkin, "A possible relation of the mass of the Universe with the characteristic sizes of elementary particles," preprint arXiv:1208.0824 (2012).
[18]E. P. Tryon, "Is the universe a vacuum fluctuation?," Nature **246**, 396–397 (1973).
[19]L. Krauss, *A Universe from Nothing* (Free Press, Simon & Schuster, New York, 2012).
[20]B. S. DeWitt and R. W. Brehme, "Radiation damping in a gravitational field," Ann. Phys. **9**, 220–259 (1960).
[21]J. Hadamard, *Lectures on Cauchy's Problem in Linear Partial Differential Equations* (Dover, New York, 1952).
[22]W. Kundt and E. T. Newman, "Hyperbolic differential equations in two dimensions," J. Math. Phys. **9**, 2193–2210 (1968).
[23]F. G. Friedlander, *The Wave Equation on a Curved Spacetime* (Cambridge U.P., Cambridge, UK, 1975).

Comment on "Magnetic field calculation for arbitrarily shaped planar wires" [Am. J. Phys. 68(3), 254–258 (2000)]

S. Gratkowski[a)]
Department of Electrical and Computer Engineering, West Pomeranian University of Technology, 70-313 Szczecin, Poland

(Received 17 August 2017; accepted 11 September 2017)
https://doi.org/10.1119/1.5005528

In the paper in the title of this comment, J. A. Miranda[1] (JAM) derives a simple and compact expression to compute the magnetic field caused by a current in a plane loop of wire at a point lying in the plane.

Starting from the Biot–Savart law

$$d\mathbf{B} = \frac{\mu_0 I}{4\pi}\frac{d\mathbf{s}\times\hat{\mathbf{r}}}{r^2}, \tag{1}$$

where μ_0 is the permeability of free space, $d\mathbf{s}$ is an element of length (pointing in the direction of current flow) of a wire which carries a current I, $\hat{\mathbf{r}}$ is the unit vector pointing from the element of length to the observation point $\mathcal{O}$, and r is the distance from the element of length to the observation point, he obtains the following simple expression for the magnitude of the total magnetic flux density at point $\mathcal{O}$:

$$B = \frac{\mu_0 I}{4\pi}\oint\frac{d\theta}{r}. \tag{2}$$

The expression is conveniently written in terms of the wire's shape $r = r(\theta)$ in polar coordinates, where r is the distance of the point on the wire from the origin at $\mathcal{O}$, and θ is the counterclockwise angle made by the line joining the point and the origin, and the reference x-line (usually horizontal). JAM illustrates the usefulness of formula (2), Eq. (4) in his paper, calculating the magnetic field at specific points in the wire's plane due to currents flowing in conic curves, spirals, and harmonically deformed circular circuits.

While formula (2) is certainly correct and generally valid (see also Ref. 2), the derivation has a weak point, as we discuss later. The following procedure is at once simpler and more general.

The element of length of a wire $d\mathbf{s}$ in polar coordinates can be decomposed into

$$d\mathbf{s} = \pm(-dr\hat{\mathbf{r}} + rd\theta\hat{\boldsymbol{\theta}}), \tag{3}$$

where $-\hat{\mathbf{r}}$ is the unit vector pointing in the direction of increasing r [compare with the definition of $\hat{\mathbf{r}}$ in Eq. (1)], $\hat{\boldsymbol{\theta}}$ is the unit vector tangent to the circle of radius r, pointing in the direction of increasing θ. The sign $\pm$ is connected with the direction of current flow. From this, we get

$$|d\mathbf{s}\times\hat{\mathbf{r}}| = |\pm(-dr\hat{\mathbf{r}} + rd\theta\hat{\boldsymbol{\theta}})\times\hat{\mathbf{r}}| = |\mp rd\theta\hat{\mathbf{z}}| = rd\theta, \tag{4}$$

where $\hat{\mathbf{z}} = -\hat{\mathbf{r}}\times\hat{\boldsymbol{\theta}}$ is the unit vector normal to both $-\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\theta}}$. By combining Eqs. (1) and (4), we immediately obtain the simple and compact expression (2).

JAM states that Eq. (4) can be obtained with the help of Fig. 1 in his paper. The figure is reproduced here as Fig. 1(a). JAM writes: "Denoting the angle between the vectors $d\mathbf{s}$ and $\hat{\mathbf{r}}$ by φ… we readily see that $\theta = \varphi - \pi/2$." (page 255). In point of fact, this statement is in disagreement with Fig. 1(a). It is obvious that if the relation is true, the figure should be modified. The modified version is shown in Fig. 1(b).

The statement given in a caption of Fig. 1 in the JAM paper

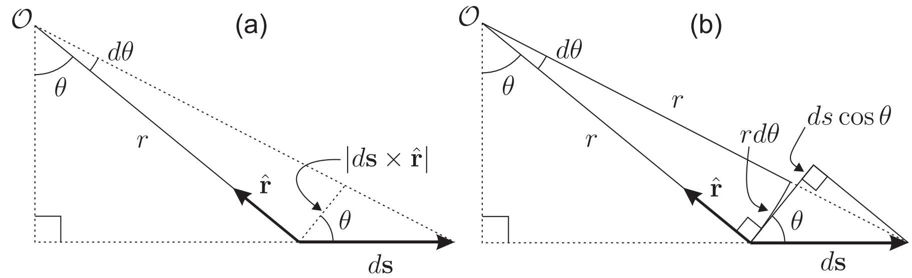


Fig. 1. (a) The plane diagram from the JAM paper (Ref. 1); (b) Modified version of the diagram.

$$r d\theta = ds \cos\theta. \tag{5}$$

means that the length of the arc $r d\theta$ is equal to the length of the segment $ds \cos\theta$ [see Fig. 1(b)]. Let us check when it is possible. The differential of arc length ds of a curve $r = r(\theta)$ is given by

$$ds = \sqrt{r^2 + (dr/d\theta)^2} d\theta. \tag{6}$$

Our task is to find a function $r = r(\theta)$ for an arbitrarily shaped wire which fulfills $r d\theta = ds \cos\theta$. By combining this with Eq. (5) we obtain the following ordinary differential equation:

$$r^2 = (r^2 + r'^2) \cos^2\theta, \tag{7}$$

where $r' = dr/d\theta$. After simple manipulations, we get

$$\left(\frac{r'}{r}\right)^2 = \tan^2\theta \quad \rightsquigarrow \quad \frac{r'}{r} = \pm \tan\theta. \tag{8}$$

Equation (7) has two solutions depending on the sign before the $\tan\theta$

$$r = r_0 \frac{\cos\theta_0}{\cos\theta}, \tag{9a}$$

and

$$r = r_0 \frac{\cos\theta}{\cos\theta_0}, \tag{9b}$$

which are equations of a straight line, the horizontal line in Fig. 1(a), and a circle whose diameter lies on the polar axis $\theta = 0$ with one end at the origin $\mathcal{O}$, respectively. The point (r_0, θ_0) represents the position of the element $d\mathbf{s}$. This can be easily seen if we express Eqs. (9a) and (9b) in Cartesian coordinates x–y. Substituting

$$x = r\cos\theta \quad \text{and} \quad r^2 = x^2 + y^2, \tag{10}$$

we get

$$x = r_0 \cos\theta_0 = \text{const.} \tag{11}$$

for Eq. (9a), and

$$\left(x - \frac{r_0}{2\cos\theta_0}\right)^2 + y^2 = \left(\frac{r_0}{2\cos\theta_0}\right)^2 \tag{12}$$

for Eq. (9b). The functions in Eq. (9) satisfy $r d\theta = ds \cos\theta$, while, for example, the function $r(\theta) = 2f/(1 - \cos\theta)$, considered in Ref. 1 for a parabolic wire, does not (f is the distance from the parabola's vertex to its focus located at the origin $\mathcal{O}$).

The derivation given by JAM is valid only for an infinitely long straight wire or a circular wire passing through the observation point $\mathcal{O}$ (or for fragments of such wires), not for arbitrarily shaped wires. The statement cited above that $\theta = \varphi - \pi/2$ is generally not true. The major limitation of JAM's derivation is simply the assumption that the angle between the dotted line and the element of length of a wire $d\mathbf{s}$ in Fig. 1(a) is equal to the polar angle θ. If JAM had used any other notation for that angle, his derivation would be general (Griffiths[2] and Zangwill[3]). Regardless of the fact that the main formula [Eq. (4) in Ref. 1] and all subsequent results presented in the JAM paper are correct, and that the author should be commended for a good paper, the simple use of Eq. (3) proposed here offers a more general derivation.

Warmest thanks to my friend Marek Ziolkowski and younger colleague Marcin Ziolkowski for stimulating discussions.

[a)]Electronic mail: Stanislaw.Gratkowski@zut.edu.pl

[1]J. A. Miranda, "Magnetic field calculation for arbitrarily shaped planar wires," Am. J. Phys. **68**(3), 254–258 (2000).
[2]D. J. Griffiths, *Introduction to Electrodynamics*, 4th ed. (Pearson Education, Inc., Glenview, IL, 2013), Prob. 5.51, p. 260.
[3]A. Zangwill, *Modern Electrodynamics* (Cambridge U. P., New York, 2012), Prob. 10.8(a), p. 331.

A demonstration of decoherence for beginners

L. Lerner[a)]
Department of Chemistry and Biochemistry, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53211

(Received 25 May 2017; accepted 5 September 2017)

We present a simplified analysis suitable for a beginners' course in quantum mechanics of a recently presented model of positional decoherence in a gas of scatterers. As such, no reference is made to

the density matrix formalism, many body theory, or even operator algebra. We only make use of the properties of quantum states and the position and momentum wavefunctions, which students typically encounter in a first quantum mechanics course.
https://doi.org/10.1119/1.5005526

Quantum mechanics describes nature in terms of state vectors $|\phi\rangle$ belonging to a Hilbert space, which in the Schrödinger picture evolve in time. This description has the apparent problem that many quantum mechanical states do not correspond to classical macroscopic objects. For example, a particle in state $|\phi_X\rangle$ localized at position X has a classical counterpart, while a particle in state $|\Phi\rangle = (|\phi_X\rangle + |\phi_Y\rangle)/\sqrt{2}$ appears to be located at both X and Y simultaneously and thus has no classical counterpart. Furthermore, macroscopic objects initially in classical-like states can evolve into states with no classical analogs. This is demonstrated in the famous Schrödinger-cat paradox, where a macroscopic cat, initially in a live state, evolves into states where it is both dead and alive.[1]

This problem can be addressed by noting that observable predictions of the quantum theory are expectation values of Hermitian operators A so that physically measurable consequences of the non-classical nature of states, such as $|\Phi\rangle$, are contained in the interference term $\langle\phi_Y|A|\phi_X\rangle$. The decoherence theory demonstrates that an interaction resulting in an entanglement with a many-particle system can produce rapid decay of such interference terms, leading to classical behavior at the macroscopic level. Here, we show how the action of a gas of scatterers on a macroscopic particle via the Schrödinger equation leads to the approximately exponential decay of $\langle\phi_Y|A|\phi_X\rangle$ using a simplified scattering model.[2] In physical terms, scattering effectively "measures" the position of the particle, forcing it into a position eigenstate.

Consider a particle with the center of mass at X, represented by state $|\phi_X\rangle$ with the corresponding position wavefunction $\phi_X(y) = \langle y|\phi_X\rangle$ and surrounded by a gas of scatterers, each in state $|\varphi\rangle$ and with the corresponding position wavefunction $\varphi(x) = \langle x|\varphi\rangle$. For simplicity, we assume that scattering is one dimensional and the scatterers are much lighter than the particle.[2] Scattering is then an elastic reflection from an infinite potential located at $x = X$, with the momentum of the scatterer changing the sign (see Fig. 1).

Because $p = m\,dx/dt$, the (scattering) transformation $p \xrightarrow{S} -p$ is effected by a change in the sign of the position, accompanied by a possible translation, $x \xrightarrow{S} a - x$. This transformation is accomplished by the scattered wavefunction $\varphi(a - x)$, which reverses the momentum distribution associated with $\varphi(x)$. Because the potential is infinite, the wavefunction must vanish at the wall;[1] thus, the constant a is determined[3] by the condition $\varphi(X) - \varphi(a - X) = 0$ so that $a = 2X$. The result is that scattering changes the product of particle and scatterer wavefunctions as

$$\phi_X(y)\varphi(x) \xrightarrow{S} -\phi_X(y)\varphi(2X - x). \tag{1}$$

Consider the wavefunction for a superposition of two orthonormal states corresponding to the particle centered at X and Y

$$\Phi(y, x) = \frac{1}{\sqrt{2}}[\phi_X(y) + \phi_Y(y)]\varphi(x). \tag{2}$$

Combining Eqs. (1) and (2), scattering changes the superposition of Eq. (2) as

$$[\phi_X(y) + \phi_Y(y)]\varphi(x) \xrightarrow{S} -\phi_X(y)\varphi(2X - x) - \phi_Y(y)\varphi(2Y - x). \tag{3}$$

Here, $\varphi(x)$ cannot be factored from the scattered state, which entangles the scatterer and particle states, and this is a key feature of decoherence.

Non-classical behavior is represented by interference of the two states in expressions for expectation values of Hermitian operators A

$$\langle\Phi|A|\Phi\rangle = \frac{1}{2}(\langle\phi_X|A|\phi_X\rangle + \langle\phi_Y|A|\phi_Y\rangle + r + r^*), \tag{4}$$

where we have defined the interference term $r = \langle\phi_Y|A|\phi_X\rangle$ and assumed that the expectation value of A does not depend on the state of the scatterer $|\varphi\rangle$. We now show that the contribution of interference terms to the expectation value tends to zero in the presence of scattering when the expectation value does not depend on $|\varphi\rangle$. The interference amplitude of the scattered state is evaluated as follows

$$r \xrightarrow{S} r\int_{-\infty}^{\infty} dx\, \varphi^*(2Y - x)\varphi(2X - x) \tag{5}$$

$$= r\int_{-\infty}^{\infty} dy\, \varphi^*(y)\varphi(2X - 2Y + y) \tag{6}$$

$$= r\int_{-\infty}^{\infty} dp\, |\tilde{\varphi}(p)|^2 e^{2ip(X-Y)/\hbar}, \tag{7}$$

where we have used the properties of the Fourier transform of a complex convolution in Eq. (7) and $\tilde{\varphi}(p)$ is the momentum wavefunction, that is, the Fourier transform of $\varphi(x)$.

As an aside, Eq. (7) is a special case of the standard positional decoherence expression for environmental scattering[4,5] when the scattering amounts to reflection, that is, the T-matrix satisfies $|T(p, q)|^2 = Rt\,\delta(p + q)$, where $\delta(p)$ is the delta function and R is the scattering rate.

The exponential in Eq. (7) can be expanded in a Taylor series as $e^{2i\alpha p} = 1 + 2i\alpha p - 2\alpha^2 p^2 + \cdots$, giving

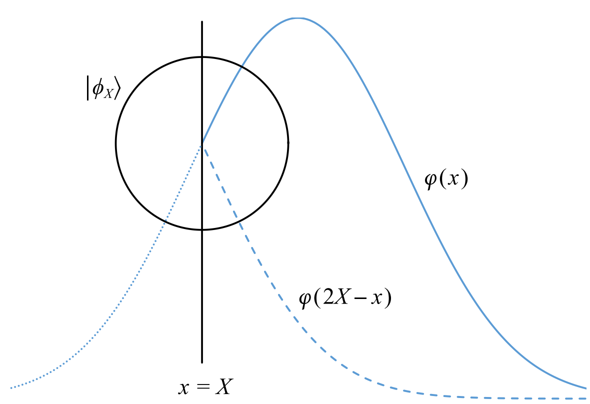


Fig. 1. One-dimensional scattering of an environmental particle with wavefunction $\varphi(x)$ from a very heavy particle in state $|\phi_X\rangle$ located at $x = X$. Due to the mass of the particle, scattering is equivalent to reflection from an infinite potential.

$$r \xrightarrow{S} r\left[1 + 2i(X-Y)\langle p\rangle/\hbar - 2(X-Y)^2\langle p^2\rangle/\hbar^2 + \cdots\right], \tag{8}$$

where $\langle p^n\rangle$ denotes the expectation value of p^n. Since scattering is assumed to be symmetric with respect to both sides of the particle, we have $\langle p\rangle = 0$. At a scattering rate R, a total of $N = Rt$ independent scatterers will collide with the particle in time t,[2] each with initial wavefunction $\varphi(x)$, so that r varies as

$$r(t) = r(0)\left[1 - 2(X-Y)^2\langle p^2\rangle/\hbar^2\right]^{Rt} \tag{9}$$

$$\approx r(0)e^{-2Rt(X-Y)^2\langle p^2\rangle/\hbar^2}, \tag{10}$$

provided that $(X-Y)^2\langle p^2\rangle/\hbar^2$ is small. Therefore, if scattering events are independent, which typically occur when the environment consists of a large number of scatterers, interference of particle states decays approximately exponentially, at a rate proportional to the square of the separation for small separations. The effect in Eq. (4) is that after a time inversely proportional to $(X-Y)^2$, only the first two terms contribute to the expectation value, and these are the expectation values of A with the particle at X or at Y but not both at the same time. This is what one would expect for a classical mixture of particles in states $|\phi_X\rangle$ and $|\phi_Y\rangle$, with equal weight 1/2 for both states.

Anticipating a more advanced treatment, we note that the expectation value in Eq. (4) can be written as the trace of a product of matrices

$$\langle\Phi|A|\Phi\rangle = \mathrm{tr}\left[\begin{pmatrix}\langle\phi_X|A|\phi_X\rangle & \langle\phi_X|A|\phi_Y\rangle \\ \langle\phi_Y|A|\phi_X\rangle & \langle\phi_Y|A|\phi_Y\rangle\end{pmatrix}\begin{pmatrix}1/2 & 1/2 \\ 1/2 & 1/2\end{pmatrix}\right], \tag{11}$$

where the second matrix is the density matrix representing the particle state in Eq. (2). It follows from the above paragraph that decoherence diminishes the off-diagonal elements of the density matrix so that with time it tends to a diagonal matrix. Such a density matrix cannot be obtained from a pure quantum state as in Eq. (2) but rather resembles a classical mixture of two quantum states of the particle.

In summary, we have shown that as a consequence of decoherence, the interference of a particle with itself can generally be neglected for macroscopic particles. Ultimately, this leads to a localization of the particle and the fact that classical particles have trajectories, that is, they appear to transition through a set of position eigenstates.[5]

[a]Electronic mail: lernerl@uwm.edu

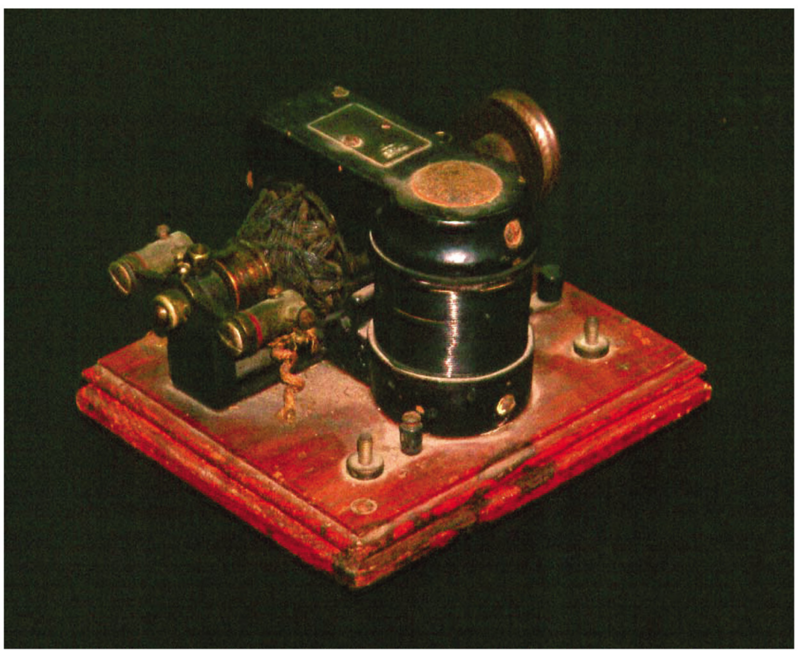
[1]D. J. Griffiths, *Introduction to Quantum Mechanics*, 2nd ed. (Prentice Hall, London, 2005).

[2]J. K. Gamble and J. F. Lindner, "Demystifying decoherence and the master equation of quantum Brownian motion," Am. J. Phys. **77**, 244–252 (2009).

[3]The wavefunction $\varphi(x+vt) - \varphi(2X - x + vt) = \psi(x,t)$ represents a non-dispersive wave-packet $\varphi(x)$ and its mirror image about $x = X$ traveling in opposite directions. This solution satisfies both the initial condition $\psi(x,t) = \varphi(x+vt)$ for large negative times and $x > X$, and the boundary condition $\psi(X,t) = 0$. The wave-packet $\varphi(x)$ is assumed to vanish at infinity.

[4]K. Hornberger and J. E. Sipe, "Collisional decoherence reexamined," Phys. Rev. A **68**, 012105-1–16 (2003).

[5]M. A. Schlosshauer, *Decoherence and the quantum to classical transition* (Springer, New York, 2007).



Electric Motor

This small bipolar electric motor and generator is at the apparatus collection of Franklin and Marshall College in Lancaster, Pennsylvania. An almost identical one is shown in the 1916 catalogue of the L. E. Knott Apparatus Co. of Boston at $7.75. As a motor it ran on 6 to 12 V and put out about 1/25 hp. When connected as a generator it put out 12 V at 2.5 A. (Picture and Notes by Thomas B. Greenslade, Jr., Kenyon College)