



yenwen / iStock

THERMODYNAMIK

Schaltenergien am Limit

Eine thermodynamische Betrachtung der Informationsverarbeitung

Jan Klärs

Der Betrieb von Datenzentren verursacht mehr als zwei Prozent des weltweiten Verbrauchs an Elektrizität. Dabei entsteht geschätzt ebenso viel CO₂ wie durch die weltweite Luftfahrt. Daher sind neue Ansätze nötig, um den steigenden Ressourcenbedarf unserer Informationsgesellschaft einzudämmen. Dafür erweist es sich auch als erforderlich, über fundamentale physikalische Grenzen nachzudenken.

Die Digitalisierung unserer Gesellschaft hat eine allgegenwärtige Informationsverarbeitung entstehen lassen. Mobiltelefone, Laptops, Arbeitsplatzrechner, smarte Haushaltsgeräte sind der sichtbare Teil dieser gigantischen Informationsmaschine. Weniger sichtbar, wengleich für den reibungslosen Betrieb unerlässlich, ist die Infrastruktur, die für die Vernetzung all dieser Geräte verantwortlich ist. Mobilfunk und glasfaserbasierte Netze

ermöglichen heutzutage einen noch nicht dagewesenen Informationsaustausch. Dass wir enorme Energiemengen für den Unterhalt dieser Technologien bereitstellen müssen, ist eine der Kehrseiten dieser Entwicklung. Der Internetkonzern Google hat berechnet, dass die Verarbeitung einer Suchanfrage genauso viel Energie verbraucht wie das Leuchten einer 60-Watt-Glühlampe für 17 Sekunden. Dabei sollte man sich vor Augen führen, dass Google im Jahr 2016 rund 3,3 Billionen Suchanfragen verarbeitet hat, sodass diese insgesamt rund 900 Millionen Kilowattstunden Strom verbrauchen. Das entspricht dem Stromverbrauch von 300 000 Haushalten mit zwei Personen.

Was aber passiert mit der Energie, die eine große Serverfarm für das Verarbeiten von Suchanfragen verbraucht? Das Beispiel der Glühlampe gibt auch hier Aufschluss. Letztlich gelangt diese Energie als Wärme in die Umgebung. Daher gilt es, Serverfarmen aktiv zu kühlen, etwa durch kaltes



Abb. 1 Ein Ansatz, um den Energieverbrauch zur Kühlung von Datenzentren zu reduzieren, ist, diese im Meer zu versenken.

Meerwasser, wie es im Projekt Natick von Microsoft der Fall ist (**Abb. 1**) [1]. Eine bessere Lösung wäre es natürlich, das Entstehen von Abwärme so weit wie möglich zu verhindern. Hier gab es durch die fortschreitende Miniaturisierung von integrierten Halbleiterschaltkreisen in den letzten Jahrzehnten enorme technologische Fortschritte. Generell teilt sich die Energiedissipation in CPUs in zwei ungefähr gleich wichtige Beiträge: eine statische Verlustleistung verursacht durch Leckströme und dynamische Verluste durch Schaltprozesse. Die dynamischen Verluste entstehen dann, wenn Logikgatter ihren Zustand verändern.

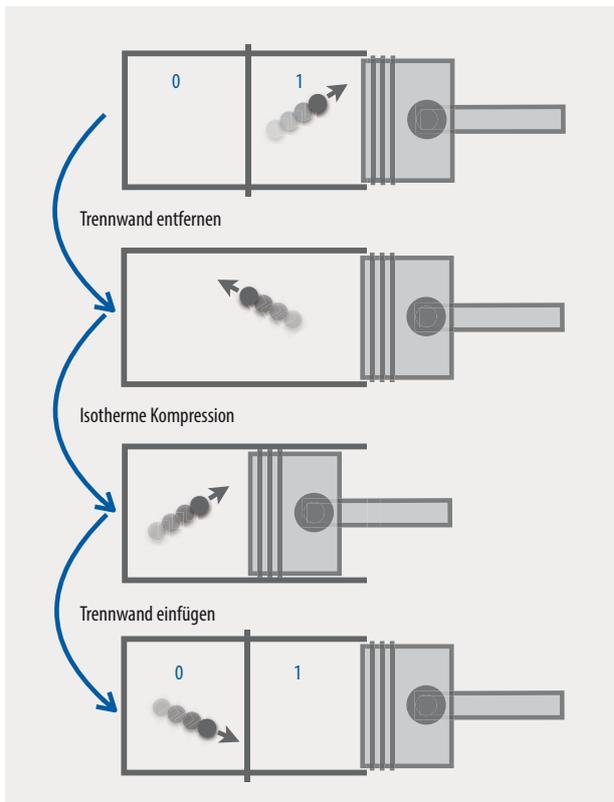


Abb. 2 Landauers Modell für einen 1-Bit-Speicher besteht aus einem Behälter mit einer Trennwand in der Mitte, in dem sich ein einzelnes Teilchen aufhält. Befindet sich das Teilchen auf der linken bzw. rechten Seite der Trennwand, so ordnet man dem Speicher den logischen Zustand 0 bzw. 1 zu. Das Initialisieren bzw. das Löschen des Speichers lässt sich durch das Entfernen der Trennwand (freie Expansion), eine reversible isotherme Kompression sowie das Wiedereinfügen der Trennwand realisieren.

Physikalisch geschieht dies durch das Auf- oder Entladen von Kapazitäten innerhalb des Gatters, was mit Energieverlusten verbunden ist. Die für einen Schaltprozess notwendige Energie hat sich in den letzten Jahrzehnten um viele Größenordnungen verringert. Während der erste Intel-Pentium-Prozessor von 1994 noch eine Schaltenergie von etwa 130 Femtojoule aufwies, unterschritten Prozessoren ab etwa 2005 die Marke von einem Femtojoule. Mittlerweile liegen die Schaltenergien um die 10 Attojoule.

„Information ist physikalisch“

Die Frage, inwieweit sich die Schaltenergie weiter reduzieren lassen könnte, ist keine rein technologische Frage. Sie berührt auch fundamentale physikalische Prinzipien. Schon weit vor dem Siegeszug der Computer hat sich der deutsch-amerikanische Physiker Rolf Landauer eingehend mit dieser Thematik beschäftigt. Er entwickelte in den frühen 1960er-Jahren ein genial einfaches Modell, mit dem er grundlegende Fragen zur Physik der Informationsverarbeitung untersuchen konnte [2]. Sein Modell beantwortet konkret die Frage nach der Schaltenergie in Computerkernen.

Landauers Ausgangspunkt ist die Überlegung, dass jede Information physikalisch codiert ist, d. h. jeder logische Zustand entspricht einem physikalischen Zustand. Landauers Gedankenexperiment basiert auf einem Teilchen, das sich in einem quaderförmigen Behälter bewegt. Durch wiederholte Stöße mit den Wänden nimmt dieses 1-Teilchen-Gas die Temperatur der Umgebung an. Der Behälter lässt sich zudem mithilfe einer Trennwand in zwei Bereiche unterteilen (**Abb. 2**). Falls sich das Teilchen auf der linken Seite befindet, sagen wir, dass das System im logischen Zustand „0“ sei. Der Fall, dass sich das Teilchen im rechten Teil des Kastens befindet, entspricht dem logischen Zustand „1“. Damit hatte Landauer ein minimalistisches mechanisches Modell für einen 1-Bit-Speicher kreiert, das mithilfe einfacher physikalischer Gesetzmäßigkeiten zu untersuchen war.

Allgemeiner betrachtet betrifft das Landauer-Limit auch das Verhältnis von Information und Thermodynamik, wie es etwa durch das Dilemma des Maxwellschen Dämons veranschaulicht wird. Doch dazu später mehr.

Das Landauer-Limit

Wichtig für die Untersuchung der Schaltenergie sind das Initialisieren bzw. Löschen des Speichers. Das Initialisieren versetzt den Speicher von einem unbekanntem Ausgangszustand in den Zustand „0“ (**Abb. 2**). Landauer hatte sich dafür folgende Prozedur überlegt:

- Die Trennwand wird entfernt. Unabhängig von seiner ursprünglichen Lage nimmt das Teilchen nun also das gesamte Volumen des Behälters ein (freie Expansion).
- Ein Kolben wird von der rechten Seite in den Behälter eingeführt, der das 1-Teilchen-Gas komprimiert und so dafür sorgt, dass sich das Teilchen auf der linken Seite des Behälters wiederfindet (isotherme Kompression). Für die Kompression ist eine bestimmte Arbeit aufzuwenden.

- Schließlich setzt man die Trennwand wieder in die Mitte des Behälters ein und entfernt den Kolben. Der Speicher ist damit im Zustand „0“.

In physikalischen Systemen mit wenigen Freiheitsgraden, wie es ein Gas aus einem einzelnen Teilchen darstellt, werden die thermodynamischen Größen, z. B. der Druck, zu stochastischen Variablen. Die Größen fluktuieren als Funktion der Zeit, erfüllen aber die bekannten Gesetze der Thermodynamik wie das ideale Gasgesetz bei hinreichend langer Mittelung. Die thermodynamische Analyse zeigt nun, dass sich beim Löschen des 1-Bit-Speichers die Entropie des Universums um $\Delta S = k_B \ln 2$ erhöht, was mit einem Energiebedarf von $W = T\Delta S = k_B T \ln 2$ einhergeht.

Die Tragweite dieses Resultats mag auf den ersten Blick nicht ganz ersichtlich sein. Was hat dieses Gedankenexperiment mit den Schaltprozessen in einem Computer zu tun? In einer abstrakteren Formulierung des Experiments lässt sich der bewegliche Kolben durch ein zeitabhängiges Potential ersetzen, das die Position des Teilchens kontrollieren kann. Solche zeitabhängigen Potentiale treten auch in elektronischen Schaltkreisen auf. Auch hier steuert das Anlegen von Spannungen zeitlich den Elektronenfluss, was zu einer Abfolge von bestimmten logischen Zuständen führt. Das Landauer-Limit lässt sich in dieser Weise recht allgemein auf logisch irreversible Schaltprozesse in der Elektronik anwenden. Das betrifft sowohl den Speicher-Initialisierungsprozess als auch viele Logikgatter (AND, OR, ...). Nicht betroffen sind dagegen gänzlich andersgeartete Rechnerarchitekturen, die ausschließlich auf thermodynamisch und logisch reversiblen Berechnungsprozessen basieren. Das trifft üblicherweise auf Quantencomputer zu, ist aber nicht auf die Quantenwelt beschränkt. Edward Fredkin und Tommaso Toffoli schlugen beispielsweise einen hypothetischen „Billardball-Computer“ vor, der einen Booleschen Schaltkreis ohne Energiedissipation simulieren kann [3]. Gegenstand theoretischer Untersuchungen waren darüber hinaus auch reversibel arbeitende zelluläre Automaten. Eine Verallgemeinerung des Landauer-Limits auf beliebige logische Prozesse unabhängig von ihrer physikalischen Realisierung haben Forscher der ETH Zürich 2014 vorgeschlagen [4].

Energiesparen durch die Hintertür

Die Schaltenergien in heutigen Transistoren liegen etwa drei bis vier Größenordnungen oberhalb des Landauer-Limits, sind also im Moment noch rein technisch bedingt. In bestimmten, gut kontrollierten Modellsystemen gelang es bereits vor einigen Jahren, die physikalische Signifikanz des Landauer-Limits experimentell nachzuweisen [5]. Die Entwicklung der optischen Pinzette erlaubt es, das ursprüngliche Gedankenexperiment von Landauer weitgehend unverändert in die Realität umzusetzen. Ein mikrometergroßes Glasteilchen wird dazu im Fokus eines Laserstrahls gefangen (Abb. 3). Durch eine schnelle Modulation des Laserstrahls im Ort und in der Intensität kann effektiv ein Doppelmuldenpotential für das Glasteilchen entstehen. Die Position des Teilchens codiert wiederum den logischen Zustand des Systems. Durch Variation des Laserstrahls lässt

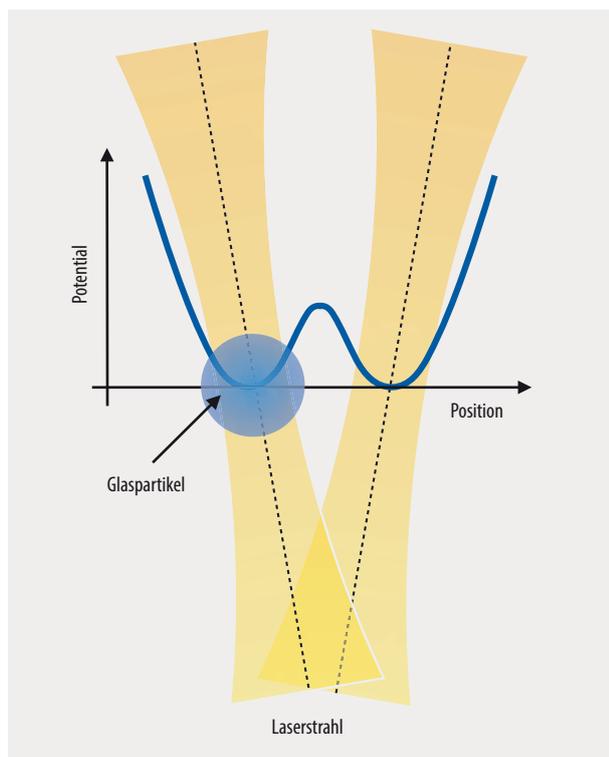


Abb. 3 Mithilfe einer optischen Pinzette ist es möglich, das Gedankenexperiment von Landauer in die Realität umzusetzen. Ein mikrometergroßes Glasteilchen wird im Fokus eines Laserstrahls gefangen. Eine schnelle Modulation des Laserstrahls erzeugt ein Doppelmuldenpotential. Die Position des Teilchens in diesem Potential codiert den logischen Zustand des Systems.

sich das von Landauer vorgeschlagene Speicher-Lösch-Protokoll durchführen. Eine Kamera verfolgt dabei die Bewegung des Teilchens. Dessen Trajektorie erlaubt dann Rückschlüsse auf den Wärmetransfer zwischen dem Teilchen und seiner Umgebung. Dies führte letztlich zur ersten experimentellen Bestätigung des Landauer-Limits.

Extrapoliert man die rasante technologische Entwicklung integrierter Schaltkreise in den letzten Jahrzehnten, sollte das Landauer-Limit etwa in ein bis zwei Jahrzehnten auch in elektronischen Systemen erreicht sein. Da diese Grenze eine direkte Konsequenz thermodynamischer Gesetze ist, besteht eigentlich keine allzu große Hoffnung auf weitergehende Verbesserungen. Sowohl hinsichtlich des Energiebedarfs als auch hinsichtlich der realisierbaren Prozessorgeschwindigkeiten von Computern wäre das von Nachteil. Neue Forschungsergebnisse zeigen aber, dass es eine „Hintertür“ gibt, die sich in den Annahmen von Landauers Gedankenexperiment versteckt, und die es uns prinzipiell erlaubt, Schaltenergien unterhalb des Landauer-Limits zu realisieren [6, 7]. Seine Herleitung basiert auf der Annahme, dass sich das Teilchen im thermischen Gleichgewicht mit seiner Umgebung befindet. Dieses kommt dadurch zustande, dass während der Kollisionen mit den Wänden zufällige Kräfte auf das Teilchen wirken, die durch die thermischen Fluktuationen der Wände verursacht werden. Auf diese Weise übertragen sich die thermischen Eigenschaften der Umgebung auf den Bewegungszustand des Teilchens. Ist die Umgebung im thermischen Gleichgewicht, wie bei der Herleitung des Landauer-Limits an-

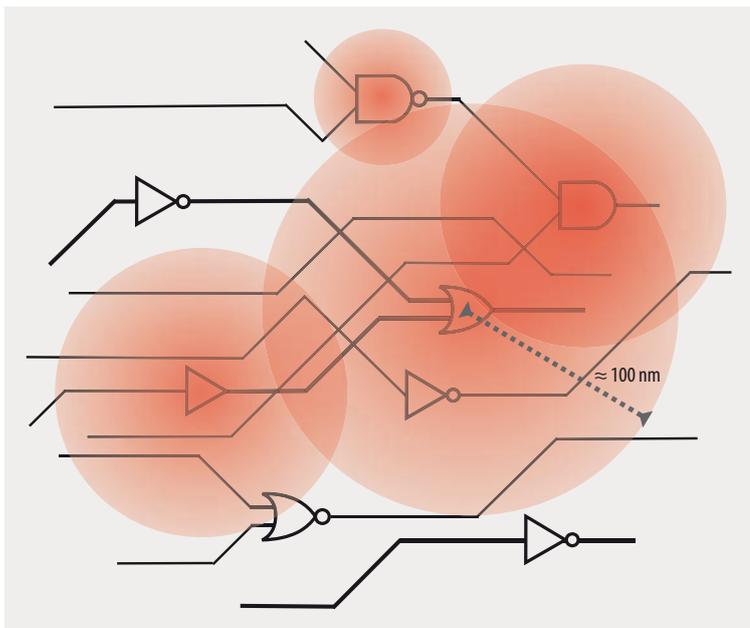


Abb. 4 In einem Computer emittieren logische Gatter durch periodische Schaltprozesse Wärme in ihre Umgebung und erzeugen Bereiche gequetscht thermischer Fluktuationen (symbolisiert durch die roten Bereiche). Dadurch beeinflussen sie den Schaltprozess benachbarter Logikgatter.

genommen, sind die Kräfte, die auf das Teilchen wirken, rein zufällig; also frei von statistischen Korrelationen. Die Frage ist nun: Ist diese Annahme korrekt bzw. lässt sie sich noch präzisieren?

Auf das Timing kommt es an

Die Schaltprozesse in einer CPU finden keineswegs zufällig statt, sondern sind durch den Takt der CPU-Uhr synchronisiert. Das bedeutet nichts anderes, als dass auch die Energieverluste durch Schaltvorgänge zeitlich synchronisiert sind. Etwas vereinfachend kann man sich ein Logikgatter in einer CPU als eine Wärmequelle vorstellen, die periodisch mit dem Takt Energie in ihre Umgebung dissipiert (**Abb. 4**). Solch eine periodische Wärmeabgabe bewirkt, dass die Umgebungstemperatur in der näheren Umgebung der Wärmequelle mit der Frequenz des CPU-Takts zwischen einem minimalen Wert T_k und einem maximalen Wert T_h hin und her oszilliert. In der statistischen Physik heißt solch ein Zustand „gequetscht thermisch“ (engl. „squeezed thermal“). Im Allgemeinen ist er durch eine mittlere Temperatur T und einen „Quetschparameter“ r charakterisiert, wobei $T_{h,k} = T \exp(\pm 2r)$ gilt (**Abb. 5a**). Die räumliche Ausdehnung des gequetscht thermischen Bereiches lässt sich auf etwa einige hundert Nanometer abschätzen.

Wie lässt sich die periodisch oszillierende Temperatur nutzen, um Energie bei Schaltprozessen einzusparen? Hier kommt es auf das richtige Timing an. In Landauers Gedankenexperiment entsteht die Schaltenergie von $W = k_B T \ln 2$ durch die isotherme Kompression des 1-Teilchen-Gases. In einer gequetscht thermischen Umgebung besteht die Möglichkeit, dass wir den Kolben nur zu jenen Zeitpunkten bewegen, in denen die effektive Temperatur des 1-Teilchen-Gases geringer ist als im zeitlichen Mittel. Diese Zeitpunkte

sind aufgrund des periodischen Verhaltens von gequetscht thermischen Reservoiren vorhersagbar. Moduliert man also die Geschwindigkeit des Kolbens periodisch mit der Frequenz des Reservoirs und achtet dabei auf die korrekte Phasenlage, so führt man die Kompression gegen den Druck eines effektiv kälteren Gases durch (**Abb. 5b**). Auf diese Weise sinkt die Schaltenergie auf $W = k_B T \ln 2 \exp(-2r)$, was einer exponentiellen Verringerung des Landauer-Limits mit dem Quetschparameter r entspricht [6]. In gewisser Weise basiert diese Energieeinsparung auf einem Recycling-Prozess: Die Abwärme früherer Schaltprozesse erzeugt einen thermischen Nichtgleichgewichtszustand, dessen Eigenschaften nützlich sind, um die Schaltenergie zu reduzieren. In einer CPU würde dies konkret bedeuten, dass die Logikgatter asynchron zu schalten sind, um die Zeitpunkte verringerter Wärmefluktuationen auszunutzen.

Der gezähmte Dämon

Neben dem technologischen Aspekt hat das Gedankenexperiment Landauers auch viel zum fundamentalen Verständnis von Thermodynamik und Information beigetragen. So lässt es sich beispielsweise nutzen, um das Dilemma des Maxwellschen Dämons aufzulösen. In diesem Gedankenexperiment wird ein Gasbehälter durch eine Trennwand in zwei Teile geteilt, die aber eine kleine verschließbare Öffnung aufweist. Dabei soll das Öffnen und Schließen der Klappe keine Energie kosten. Zu Beginn des Experiments enthalten beide Teile des Behälters Luftmoleküle bei Raumtemperatur. Somit gehorchen die Moleküle einer bestimmten Geschwindigkeitsverteilung. Ein intelligentes Wesen, der „Dämon“, betrachtet nun die Moleküle in der Nähe der Öffnung. Er öffnet und schließt die Klappe so, dass die schnellen Moleküle nur von links nach rechts und die langsamen Moleküle nur von rechts nach links durch die Öffnung gelangen können. Auf diese Weise baut sich „aus dem Nichts“ eine Temperaturdifferenz zwischen den beiden Teilen des Behälters auf. Das verletzt aber offenbar den zweiten Hauptsatz der Thermodynamik.

Zur Auflösung dieses Dilemmas gilt es, die Rolle des Dämons kritisch zu untersuchen. Der amerikanische Physiker und Informationstheoretiker Charles Bennett argumentierte etwa, dass die Geschwindigkeitsinformationen, auf deren Basis der Dämon entscheidet, ob er die Klappe öffnet oder schließt, nur durch das Speichern von Positionsdaten einzelner Moleküle zu gewinnen sind [8]. Wenn aber das Gedächtnis bzw. der Speicher des Dämons komplett mit Positionsdaten belegt ist, müssen diese gelöscht bzw. überschrieben werden. Das Landauer-Limit schreibt hierfür eine minimale Entropieproduktion bzw. einen minimalen Energieaufwand vor. Diese stellen letztlich wieder die Konsistenz zum zweiten Hauptsatz her und lösen so das Dilemma des Maxwellschen Dämons auf.

Jenseits der Informationsverarbeitung

Auch jenseits der Thermodynamik von Computerspeichern haben sich gequetscht thermische Reservoire als fruchtbares Forschungsobjekt erwiesen. So gelang es, zu-

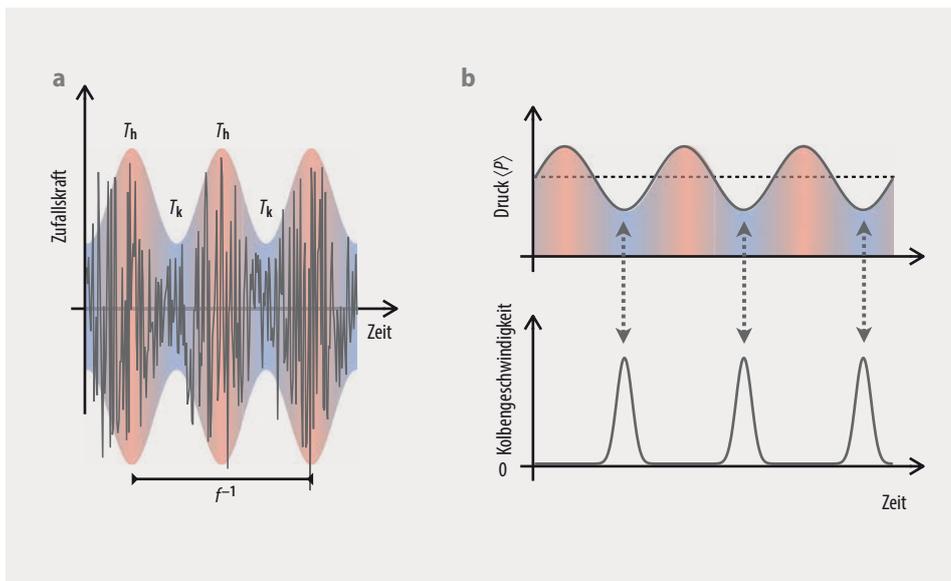


Abb. 5 Ein thermisches Reservoir kann als eine Zufallskraft auf ein gegebenes physikalisches System modelliert werden (a): Je höher die Temperatur des Reservoirs, desto größer ist die Amplitude der Zufallskraft. Ein gequetscht thermisches Reservoir erzeugt eine Zufallskraft, die durch schnelle Amplitudenmodulationen gekennzeichnet ist. Das lässt sich als schnelle periodische Temperaturvariation zwischen einem Minimal- und einem Maximalwert (T_k bzw. T_h) mit Frequenz f betrachten. Wegen der gequetscht thermischen Umgebung ist der Druck P , den ein 1-Teilchen-Gas auf einen Kolben ausübt, periodisch moduliert (b). Um die für die Kompression des Gases notwendige Arbeit zu minimieren, wird die Geschwindigkeit des Kolbens ebenfalls periodisch moduliert: Der Kolben wird nur zu den Zeitpunkten bewegt, bei denen der erwartete Druck geringer ist als im zeitlichen Mittel.

nächst theoretisch [9] und später experimentell [10] zu zeigen, dass solche Reservoirs ein ausgesprochen kraftvoller Treibstoff für Wärmekraftmaschinen sind. Diese nehmen üblicherweise Energie aus einem heißen Reservoir auf, konvertieren diese teilweise in Arbeit und geben Restwärme an ein kaltes Reservoir ab. Das Vorhandensein zweier thermischer Reservoirs mit unterschiedlichen Temperaturen ist eine notwendige Voraussetzung für ihren Betrieb. Das kann man auch am berühmten Carnot-Koeffizienten ablesen, der die Effizienz der Umwandlung von Wärme in Arbeit beschreibt und gegen Null geht, falls die Temperaturdifferenz der Reservoirs gegen Null geht. Neuere Experimente haben gezeigt, dass sich die Arbeitsleistung von Wärmekraftmaschinen steigern lässt, wenn man das heiße Reservoir durch ein gequetscht thermisches Reservoir ersetzt. Die erreichte Konversionseffizienz kann dann sogar oberhalb des Carnot-Limits liegen.

Bemerkenswerterweise sind Kreisprozesse möglich, die Arbeit aus einem einzelnen gequetscht thermischen Reservoir gewinnen. Diese Möglichkeit rührt letztlich daher, dass ein gequetscht thermisches Reservoir mehrere Temperaturen in sich vereint (**Abb. 4a**) und daher sowohl die Rolle des heißen als auch des kalten Reservoirs übernehmen kann. Die in den Experimenten verwendete Wärmekraftmaschine erfüllt noch keinen praktischen Anwendungszweck. Ein längerfristiges Ziel dieser Forschungsaktivitäten besteht aber darin, miniaturisierte Maschinen herzustellen, die eine bestimmte Aufgabe erledigen können, wie den Transport von Teilchen in einer Flüssigkeit oder die Manipulation von biologischer Materie. Thermische Nichtgleichgewichts-Zustände in der Umgebung könnten dafür als Treibstoff dienen.

Fazit

Periodisch getriebene, dissipative Systeme, beispielsweise Computerprozessoren, sind natürliche Kandidaten für das Auftreten von gequetscht thermischen Reservoirs. Das thermische Verhalten von Prozessoren auf langsamen

Zeitskalen ist gut verstanden und lässt sich von jedem Computerbesitzer durch interne Temperatursensoren leicht nachvollziehen. Doch die statistischen Eigenschaften thermischer Fluktuationen auf der Zeitskala des CPU-Takts sind bisher noch weitgehend unbekannt. Wie das Beispiel des Landauer-Limits zeigt, könnte man aus solchen Untersuchungen wichtige Schlussfolgerungen für das Wärme- und Arbeitsmanagement und letztlich auch für die Energieeffizienz und Arbeitsgeschwindigkeit von Computerprozessoren ziehen.

Literatur

- [1] natick.research.microsoft.com
- [2] R. Landauer, IBM J. Res. Dev. **5**, 183 (1961)
- [3] E. Fredkin und T. Toffoli, Int. J. Theor. Phys. **21**, 219 (1982)
- [4] P. Faist, F. Dupuis, J. Oppenheim und R. Renner, Nat. Commun. **6**, 7669 (2015)
- [5] A. Bérut et al., Nature **483**, 187 (2012)
- [6] J. Klaers, Phys. Rev. Lett. **122**, 040602 (2019)
- [7] M. Konopik, A. Friedenberger, N. Kiesel und E. Lutz, arXiv:1806.01034
- [8] C. H. Bennett, Int. J. Theor. Phys. **21**, 905 (1982)
- [9] J. Roßnagel et al., Phys. Rev. Lett. **112**, 03602 (2014)
- [10] J. Klaers, S. Faelt, A. Imamoglu und E. Togan, Phys. Rev. X **7**, 031044 (2017)

Der Autor



Dr. Jan Klärs studierte an der RWTH Aachen und promovierte 2011 an der Universität Bonn im Institut für Angewandte Physik. Darauf folgten Postdoc-Aufenthalte in Bonn und am Institut für Quantenelektronik an der ETH Zürich. Seit 2017 ist er Assistent Professor für Nanophotonik am

Lehrstuhl für Komplexe Photonische Systeme an der Universität Twente in den Niederlanden. Er forscht zur Thermodynamik von nanoskaligen optischen und optomechanischen Systemen und zu optischen Spin-Glas-Simulatoren.

Dr. Jan Klärs, Complex Photonic Systems, MESA+ Institute for Nanotechnology, University Twente