Shot noise: a 100-year history, with applications to lithography

Chris A. Mack
Shot noise: a 100-year history, with applications to lithography

Chris A. Mack*
Fractilia, Austin, Texas, United States

Abstract. The term “shot effect” (schroteffekt) was coined in 1918 when Walter Schottky studied electrical noise in vacuum tubes. Earlier still, the foundations of shot noise theory go back to Einstein, who in 1905 explained the photoelectric effect as caused by discrete “particles” of light and Brownian motion as caused by discrete particles of matter. When the numbers of particles that affect observable outcomes are large, shot noise effects (variability in number as a fraction of the mean number) become small, and the continuum approximation (energy and matter are continuous) becomes accurate. For most of the history of semiconductor lithography, the continuum approximation has served well. But at small dimensional scales, where the number of discrete particles or events is small, the counting statistics of shot noise can dominate. The 100-year history of shot noise in science and engineering is today playing a role in our understanding of shot noise in lithography. © 2018 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMM.17.4.041002]

Keywords: shot noise; Poisson statistics; line-edge roughness; linewidth roughness; stochastic-induced roughness.

Paper 18056V received May 8, 2018; accepted for publication Jun. 7, 2018; published online Jul. 11, 2018.

“And for this reason it is the more right for you to give heed to these bodies, which you see jostling in the sun’s rays, because such jostlings hint that there are movements of the matter too beneath them, secret and unseen. For you will see many particles there stirred by unseen blows change their course and turn back, driven backwards on their path, now this way, now that, in every direction everywhere.”

Lucretius, On the Nature of Things, first-century BCE

1 Introduction

One hundred years ago, a new form of noise was discovered when Walter Schottky studied the output of vacuum tubes run at low currents. He called this noise the “shot effect” (schroteffekt). Earlier still, the foundations of shot noise theory go back to Einstein, who in 1905 gave convincing arguments that both light and matter are fundamentally discrete in nature. Einstein’s photon theory of light was inspired by Max Planck’s quantization of energy, and his argument for the atomic theory of matter provided an explanation for Brownian motion. Today, we recognize that shot noise plays a role in any system where the number of discrete events that are important is sufficiently small so that the counting statistics of those events must be considered.

In lithography, events such as photon absorption within a photoresist, chemical conversion of a light-sensitive component in that resist, and the chemical changes that make a molecule soluble in developer are all discrete events that experience shot noise. Consider a volume of photoresist and ask, “How many events occurred within that volume to change the solubility of that portion of resist?” If the volume of resist is sufficiently large that the number of events is large, the variability of that number relative to its mean (the shot noise) will be small, or even negligible. In this regime, it is safe to make the continuum approximation, where the discrete nature of light and matter is ignored. But if the volume of resist of interest is small enough, shot noise will dominate its behavior. As lithographic dimensions shrank from 25 μm in the early days of semiconductor manufacturing to 25 nm and below today, we have transitioned from a regime where shot noise can be completely ignored to one where it dominates errors in patterning.

This paper will review the history of shot noise, beginning almost 200 years ago with Brownian motion. We will then look at how the ideas of shot noise have been applied in lithography for semiconductor manufacturing over the last 40 years.

2 Brownian Motion

The term Brownian motion is named for Robert Brown (1773 to 1858), a well-known botanist of his day. In 1827, while studying pollen suspended in water under a microscope, he observed a random, quick motion of the pollen particles that at first he attributed to the “vital force,” an ill-defined concept of life that was in vogue at the time. Brown was not the first to notice such movements, but he was the first to make an extensive study of it. He quickly realized that inorganic particles (and in fact, any properly sized particles) would also exhibit the random motion, going so far as to test ground particles from the Egyptian Sphinx (courtesy of the British Museum where he worked). He systematically ruled out causes for the motion, such as living creatures, gradients and convection currents, evaporation, or vibrations. In the end, he left its cause a mystery. In 1827, he self-published a pamphlet describing his observations (Fig. 1), the contents of which were republished in a scientific journal the next year.

The mystery of the cause of Brownian motion simmered for over 75 years. Although hand-waving arguments were quickly proposed, they were slowly dispelled in turn through careful experiments, including the possibilities of chemical

*Address all correspondence to: Chris A. Mack, E-mail: chris@Lithoguru.com

J. Micro/Nanolith. MEMS MOEMS 041002-1 Oct–Dec 2018 • Vol. 17(4)
and electrical attractions. What was left was the theory of
molecular motion: energetic water molecules constantly
colliding with the suspended particle randomly push it in
various directions. However, the atomic/molecular theory
of matter was still controversial at the time, and many
were unconvinced by this proposed cause of Brownian
motion.

In 1859, James Clerk Maxwell described the distribution
of velocities of idealized gas particles statistically, the first
statistical law of physics. Ludwig Boltzmann expanded on
this result in 1871 into what we now call the Maxwell–
Boltzmann distribution of particle velocities. The success of
this theory in predicting and explaining the properties of
gases prompted many to apply the same concepts to explain
Brownian motion. If the water molecules could be described
by the same distribution of velocities, perhaps the motion of
the suspended particle could be explained and predicted as
well. However, attempts to both predict and measure the
velocity of the Brownian particles proved a failure. By the
close of the 19th century, many saw Brownian motion as
a possible proof for the existence of atoms, but that proof
remained lacking.

It is hard to imagine today that in 1900 the atomic/
molecular theory of matter was very much in debate. Fueled
by the positivist school of philosophy, which eschewed
explanations that could not be directly observed with the
senses, many scientists considered atoms a convenient fic-
tion, but not real. The statistics of ideal gases seemed ideally
suited to explain Brownian motion, but the connection
remained elusive. In 1905, Albert Einstein finished his

Fig. 1 The front page of Robert Brown’s self-published pamphlet describing his observations and
experiments on what came to be called Brownian motion.
In the dissertation and a second publication that year, Einstein applied Maxwell–Boltzmann statistics to the problem of Brownian motion. But instead of trying to mimic the Maxwell–Boltzmann theory and predict the velocity of the Brownian particle (an exercise that will later be proven to be impossible), Einstein instead focused on the mean position of the particle over time. In doing so, he provided the first theoretical derivation of the diffusion equation and showed that a Brownian particle’s position should follow Gaussian statistics.

Einstein’s theoretical achievement was quickly appreciated as providing the proof needed for the molecular theory of matter, but only if the predictions of his theory could be experimentally verified. The experimental verification came a few years later by the careful and detailed work of the French scientist Jean Baptiste Perrin. Perrin devised several tests of the molecular theory of Brownian motion and began by creating particles of uniform diameter and density, which he measured. Since the particles were denser than water, a suspension of these particles in water could test the molecular theory of Brownian motion by looking at how they settled. In the absence of molecular collisions, the particles should all settle to the bottom under the force of gravity. But molecular collisions would force some particles to stay suspended, creating an exponential concentration gradient in the vertical direction in the same way as barometric pressure varies with altitude. By measuring the concentration gradient (counting particles using a Zeiss microscope in water immersion mode, see Fig. 3), he was able to verify the predicted exponential distribution of particle counts. Further, by separately measuring quantities such as particle size and density, Perrin was able to calculate the one unknown quantity in his theoretical equation: Avogadro’s number. His result of $7.05 \times 10^{23}$ was only off by 17% from the value we know today and was the most accurate measurement of Avogadro’s number to date. Perrin also directly verified Einstein’s predictions about Brownian motion by measuring the particle position over time (Fig. 3).

Perrin’s work earned him the Nobel Prize in physics in 1926 and firmly established the atomic theory of matter: matter is not a continuum but instead is made up of discrete particles. But the study of Brownian motion was not complete. In 1908, Paul Langevin developed a stochastic differential equation for Brownian motion, adding an important mathematical tool to its study. In 1930, Uhlenbeck and Ornstein added friction to the motion of the particle, creating a more accurate model for Brownian motion (a model which remains useful today in describing photoresist stochastic behavior). The most complete model for Brownian motion was developed by the mathematician Norbert Wiener. Known as the Wiener process, it is a continuous-time stochastic process such that the distance a particle travels in time interval $\Delta t$ is Gaussian with mean 0 and variance proportional to $\Delta t$, and with independent, nonoverlapping time intervals. The use of this model required the development of stochastic random variables and their mathematical treatment. For example, the position of a particle in a Wiener process is a stochastic random variable that has no defined derivative (that is, undefined velocity). Wiener developed time-series analysis techniques for such variables and a stochastic version of the Fourier transform. He also showed that the power spectral density of a stochastic random variable is the Fourier transform of its autocorrelation function. The study of Brownian motion took another interesting turn when Benoit Mandelbrot applied the new concept of fractals to its description. Even today, Brownian motion remains an active area of research, with many interesting applications.

### 3 Discrete Light and Discrete Current: The Photon and the Electron

The year 1905 was special not just for Einstein’s theoretical proof of the existence of atoms, but for the concept of the photon: light is not a continuum of energy but instead is made up of photons, small units of energy that interact with matter discretely. Einstein examined the photoelectric effect, where light shining on a metal plate in a vacuum can release electrons. A surprising experimental result by Philipp Lenard in 1902 was that while the number of electrons emitted was a function of the intensity of light, the energy of the emitted electrons was independent of the intensity of light. And if the frequency of the light was below a certain threshold, no electrons would be emitted regardless of the light’s intensity. This behavior was completely unexplained by classical electromagnetic theory.
anomalous under the paradigm of continuous light energy. Einstein was able to make sense of it by applying Max Planck’s concept of the energy quantum: light is made up of discrete “particles,” each with an energy fixed by the frequency of the light. The intensity of the light is then determined by the number of photons.

The twin ideas of discrete matter and discrete energy events produced a remarkable paradigm shift from the continuum ideas prevalent in the 19th century. Even the atom was shown to be made of more elementary particles. In 1896, Joseph J. Thomson showed that cathode rays were made of negatively charged particles (later named electrons) and measured their mass and charge. In a series of experiments in 1908 to 1911, Ernst Rutherford, Hans Geiger, and Ernest Marsden showed that atoms are composed of a central positively charged nucleus surrounded by electrons.

The theoretical and experimental ideas of matter and energy as made up of discrete particles inspired mathematicians to delve into appropriate mathematical tools for their description, especially random variables with a Poisson distribution. For example, in 1909, Campbell derived the moments (mean and variance) of a sum of Poisson processes, motivated by the latest developments in physics: “The trend of modern theory is everywhere to replace by discontinuity the continuity which was the basis of science in the last century.”

4 Walter Schottky and the “Shot Effect”

It took some time for the full implications to become apparent of energy and matter (and their interaction) as composed of discrete, stochastic particles or events. In 1918, Walter Schottky was studying the behavior of vacuum tubes, such as a vacuum diode, under low-current conditions (Fig. 4). The measured current was noisy, a not unexpected result. But unlike other types of noise, this noise had a frequency spectrum that was constant over all frequencies and was not temperature dependent. Schottky called this noise the “Schroteffekt,” the shot effect. In coining this phrase, Schottky said, “The expression ‘shot’ points, as it does in common language use, to the occurrence of a large number of homogeneous elementary particles.”

At the time, other forms of electrical noise had well-known frequency and temperature dependences that differed greatly from Schottky’s new shot effect. Johnson noise, thermal noise of current flowing through a conductor or resistor, varied greatly with temperature. Flicker noise had a $1/f^\alpha$ frequency dependence. But the shot effect was white noise, with a flat frequency response over most of the frequency range. As Schottky said, “Because of the atomic structure of electricity the electrical transition is represented not as a continuously flowing current but as a hail of charge quanta which would cause current fluctuations even for a very regular temporal distribution.”

According to Partridge, noted lexicographer of slang, the word “noise” derives from nausea, from “the noise made by an ancient shipful of passengers groaning and vomiting in bad weather.” The Oxford English Dictionary defines noise as “random fluctuations that obscure or do not contain meaningful data or other information.” Over time, the phrase “shot effect” became “shot noise.” It is used to describe any random uncertainty in a physical quantity caused by the counting statistics of the discrete events that underlie the phenomenon. The most common statistical distribution appropriate for shot noise is the Poisson distribution, characterized by a variance of the distribution equal to its mean.

As an example of Poisson counting statistics, consider the concept of concentration, the average number of molecules per unit volume. Let $C$ be the average number of molecules...
per unit volume, and $dV$ a volume small enough so that at most one molecule may be found in it (thus requiring that the concentration be fairly dilute, so that the position of one molecule is independent of the position of other molecules). The probability of finding a molecule in that volume is just $CdV$. For some larger volume $V$, the probability of finding exactly $n$ molecules in that volume, $P(n)$, will be given by a binomial distribution. But for any reasonably large volume ($CV > 1$), this binomial distribution will also be well approximated by a Poisson distribution

$$P(n) = \frac{(CV)^n}{n!} e^{-CV}. \quad (1)$$

The average number of molecules in the volume will be $CV$, and the variance will also be $CV$. The relative uncertainty in the number of molecules in a certain volume will be

$$\frac{\sigma_n}{\langle n \rangle} = \frac{1}{\sqrt{\langle n \rangle}} = \frac{1}{\sqrt{CV}}. \quad (2)$$

Thus, if the mean number of particles in the process is large, the relative uncertainty in that number is small and the impact of shot noise is also small. But as the mean number of particles becomes small (for example, by examining a very small volume of material so that the expected number of particles in that volume is small), the relative uncertainty in that number becomes great.

For a given concentration, the relative importance of shot noise is determined by the size of the volume of interest. As will be described next, the scaling of lithography to small dimensions has meant that the volume of interest shrinks and the importance of shot noise in lithography processes continues to grow.

5 Shot Noise in Lithography

Today, shot noise is widely recognized as a serious problem when printing near the resolution limit of extreme ultraviolet (EUV) lithography.\textsuperscript{21,22} But the concept of shot noise as a limiter to lithographic performance is not new to EUV. In the 1970s, x-ray lithography was considered a potential successor to optical lithography. Like EUV today, x-ray light sources were not as bright as desired, so researchers wondered how low the exposure dose could be made before print quality becomes limited by shot noise. By 1976, Spiller and Feder, and coworkers at IBM were exploring the trade-off between resolution and resist sensitivity (what we now think of as a part of the resolution–linewidth roughness–sensitivity trade-off).\textsuperscript{23,24} Describing photoresist as a detector,

\textquote{Theoretically the final limit for the sensitivity of any detector is determined by the shot noise of the absorbed photons... resists with lower resolution can have higher sensitivities than resists with high resolution.}\textsuperscript{25}

They went on to show that the shot-noise limited resist sensitivity must scale as $1/R^3$, where $R$ is the resolution.\textsuperscript{26} They showed that the minimum incident dose $E_{\text{inc}}$ required to print a feature of dimension $\delta$ using photons of energy $h\nu$ in a resist with absorption coefficient $\alpha$ will be

$$E_{\text{inc}} = \frac{\pi h \nu}{\alpha \delta}. \quad (3)$$

where $\pi$ is the mean number of photons absorbed in the volume that are required to avoid detrimental shot noise effects. The implications of this scaling are unpleasant if one is limited by available light: if the feature size is reduced by a factor of 0.7, the dose required to print it must rise by a factor of 3. A similar scaling law for electron beam lithography was also derived in 1976.\textsuperscript{27}

Further insights into the stochastic limits of x-ray lithography were provided by a number of researchers who tried to turn the Spiller and Feder scaling relationship into hard numbers as to what dose would be required for which feature size. Smith\textsuperscript{28} derived a simple relationship between lithographic feature edge error ($\Delta x$), the image intensity slope at the edge, the Poisson statistics photon shot noise $\sigma_{\text{photon}}$, and what he called the \textquote{development uncertainty band} $\delta N$ related to photoresist contrast

$$\Delta x = \frac{\delta N + \sigma_{\text{photon}}}{\text{image slope}}. \quad (4)$$

Neureuther and Willson,\textsuperscript{29} in a remarkable 1988 paper, explored the impact of the photoresist in far more detail. They recognized that stochastics must be considered in two ways:

First the occurrence of a large random defect at anyone of some $10^{11}$ sites on chip could produce a catastrophic failure and second, intra- and interfeature variations in linewidths at any of several hundred sites could cause unacceptable performance degradation or reliability failure.

Neureuther and Willson described two feature size-scaling regimes. In addition to the $1/R^3$ dose scaling of Spiller and Feder, they described a \textquote{fixed resist aspect ratio} scaling where absorption is made to increase as $1/R$ so that required dose scales as $1/R^2$. They derived a model for the dose required to avoid various types of stochastic defects (insoluble bits of resist at different locations on a device), though their dose limit was \textquote{based on the occurrence of a defect so infrequent that constructing an experiment to observe this effect would be extremely difficult.}\textsuperscript{30}

The need to consider shot noise effects in EUV lithography was appreciated as early as 1994.\textsuperscript{31,32} Most of this early work sought to understand the effects of shot noise and other stochastic factors on the roughness of photoresist edges. Schechter et al.\textsuperscript{33} extended the stochastic defect model of Neureuther and Willson to chemically amplified resists and predicted, for a given resist, that a 54-mJ/cm$^2$ dose would be required to print 70-nm features. In 2001, O’Brien and Mason\textsuperscript{34} explored a different approach to predicting stochastic defects in contacts. By counting the total number of photons absorbed inside a contact hole, they considered shot noise as simply an exposure dose error. For a contact with a given exposure latitude, the probability that shot noise will produce simply a device has billions (or hundreds of billions) of contacts, even low probability contact failures can produce a high probability of device failure. In their model, the required dose is inversely proportional to the exposure latitude squared and proportional to the \textquote{number of sigmas} in the probability distribution that matters squared. For example, a five-sigma failure rate is about 1 failure per 3.5 million contacts, a six-sigma...
failure rate is about 1 failure per billion contacts, and a seven-sigma failure rate is about 1 failure per trillion contacts. Thus, the required dose not only scales with feature size, but also with the number of features found on a device.

More recently, the reduction in exposure latitude for undersized contacts has been added to O’Brien and Mason’s model, predicting a much higher probability of contact hole failure.\(^{21}\) These models, while calibrated with experimental data, must still be extrapolated very far along the tail of a probability distribution.\(^{22}\) Even if the critical dimension distribution of a million contacts was measured, it is still very hard to accurately predict the expected number of failures for 100 billion contacts.

One consequence of the statistics of stochastic effects will be the need to use higher than desired exposure doses in manufacturing. Early targets for EUV dose to size were 5 mJ/cm\(^2\). The throughput specifications for ASML’s NXE:3100 assumed a 10-mJ/cm\(^2\) dose and the NXE:3300 assumed a 15-mJ/cm\(^2\) dose.\(^{23}\) Today, a 20-mJ/cm\(^2\) dose is widely assumed when making throughput calculations, even though a workable resist with such a dose to size is more aspirational than realistic. The realities of high-yield manufacturing with EUV will require much higher doses, and these doses must increase every time the feature size is reduced.

The exact stochastic limits of EUV lithography remain uncertain. From a theoretical perspective, the volume \(V\) that should be used in a Poisson calculation of shot noise has not been exactly pinned to a measurable quantity, through a recent proposal to relate this volume to the total resist blur as manifest in a roughness correlation length has some potential.\(^{35}\) From a stochastic defect perspective, work is required to either measure or predict killer stochastic defects in one out of 10\(^11\) features. Still, progress in these areas could provide answers in the very near future.

6 Conclusions

One hundred years ago a new term, and a new concept, entered the lexicon of science: shot noise. This idea was a natural consequence of the changing view of the microscopic world from the continuous to the discrete. For large-scale phenomenon, where the number of particles or events in a volume of interest is large, the relative variation about the mean is small so that a continuum view of the world is quite accurate. But as the scale of interest shrinks, the importance of shot noise grows quickly.

Shot noise has been an understood part of lithographic phenomenon for over 40 years. But models that can predict the lithographic impact of stochastic phenomenon with sufficient accuracy are still lacking. Each new generation of lithographic process, with its smaller dimensions, suffers from increased stochastic variations. So for each new generation of lithographic process, there is a fear that stochastic variation will lead to unacceptable yield loss and a hard limit to progress, and a hope that this limit will occur at least one more generation in the future.

References


Chris A. Mack developed the lithography simulator PROLITH and founded and ran FINLE Technologies for ten years. He then served as a vice president of Lithography Technology for KLA-Tencor for five years, until 2005. He is a fellow of SPIE and IEEE and is an adjunct faculty member at the University of Texas at Austin. In 2012, he became editor-in-chief of the Journal of Micro/Nanolithography, MEMS, and MOEMS. In 2017, he cofounded Fractilia, where he is chief technical officer.