

The Bridge | Digital Talent Accelerator
Bootcamp Data Science
Part-Time



Estudio de estructura interna de nanopartículas de oro.

Análisis exploratorio de datos

Josefina Cresta

18 de mayo de 2022

Resumen

La nanociencia es el estudio de los procesos fundamentales que ocurren en estructuras de 1 a 100 nanómetros, escala en la cual las propiedades físicas y químicas de los sistemas difieren de las de los sistemas macroscópicos. El objetivo de la nanotecnología es innovar y desarrollar nuevos dispositivos de impacto en diferentes áreas, como por ejemplo, en la conversión de energía, almacenamiento de información, administración de fármacos y mejora de la calidad de los procesos médicos. Se denomina nanomateriales a aquellos materiales que miden entre 1 y 100 nm en al menos una de sus dimensiones. En particular, las nanopartículas son aquellos nanomateriales que poseen menos de 100 nm en todas sus dimensiones. El presente trabajo se focaliza en nanopartículas de metales nobles, especialmente las compuestas por átomos de oro. Se realiza un análisis detallado de una base de datos que cuenta con información de características estructurales y energéticas de este tipo de nanopartículas, con el objetivo de determinar si su estructura interna difiere de la del oro macroscópico que conocemos de nuestro día a día, nombrado a continuación como oro en bulk, o a granel.

Índice

1. Objetivos	2
2. Recursos:	2
3. Base de datos	2
4. Análisis de datos:	4
4.1. Longitud media de enlaces	4
4.2. Coordinación promedio	5
4.3. Energía	6
4.4. Energía y coordinación	7
5. Conclusiones:	7
A. Figuras:	8

1. Objetivos

En los últimos años, se ha extendido el uso de nanopartículas de metales nobles, como el oro, con diversos fines, desde la química analítica y la ingeniería hasta la medicina (Schmid, 2011). Estas nanopartículas aparecen en un gran número de aplicaciones comerciales, aprovechando las propiedades únicas que surgen de la naturaleza nanométrica del material, las cuales varían en función de sus características fisicoquímicas como la concentración, la distribución de tamaño, la composición, la forma, las propiedades de superficie, la aglomeración y la naturaleza del material (Rasmussen y col., 2018).

La intención del presente trabajo es realizar un análisis exploratorio de datos sobre la estructura atómica interna de nanopartículas de oro ya que es un punto crítico para la comprensión de sus propiedades físicas y químicas. Para esto se investiga detalladamente una base de datos descripta a continuación (Sección 3), con el objetivo de afirmar la hipótesis de que a pesar que las propiedades de las nanopartículas de oro son considerablemente diferentes a las propiedades del oro sólido, a partir de ahora oro en bulk, las características estructurales internas de estas dos configuraciones coinciden.

Específicamente se busca calcular estadísticamente los siguientes valores y se analizan las siguientes correlaciones:

- Longitud media de enlace entre átomos. Dependencia de la longitud de enlace según la cantidad de átomos que componen la nanopartícula (Sección 4.1)
- Número de coordinación de cada átomo interno y de cada átomo superficial (Sección 4.2).
- Correlación de la energía total con cantidad de átomos que componen la nanopartícula. Cálculo de E energía media por átomo (Sección 4.3).
- Correlación de energía con coordinación de átomos en bulk y superficiales (Sección 4.4).

En cada sección se describe con mayor detalle los motivos por los cuales se analizan estas propiedades.

2. Recursos:

La base de datos utilizada es obtenida del *Portal de acceso a datos de CSCIO* (link) el cual es una fuente de datos de investigación, software y otros activos digitales en una variedad de disciplinas.

Para el desarrollo de los programas para análisis de datos se utiliza *Python*, lenguaje de programación interpretado, multiplataforma y destacado por su sencillez y gran cantidad de librerías útiles en ámbitos científicos. Además se utilizan *Jupyter Notebooks* para el desarrollo de los análisis, ya que permiten contener códigos en vivo, con documentación, gráficos, diagramas y visualizaciones.

Más específicamente, se utiliza la librería *Pandas*, especializada en el manejo y análisis de estructuras de datos, *Numpy*, especializada en el cálculo numérico y se realizan análisis gráficos mediante las librerías *Matplotlib* y *Seaborn*.

Finalmente se utiliza para el análisis estadístico de las variables el módulo *Stats* de *Scipy*, el cual contiene una gran cantidad de distribuciones de probabilidad, estadísticas de resumen y frecuencia, funciones de correlación, pruebas estadísticas y estadísticas enmascaradas.

3. Base de datos

A partir de una exhaustiva búsqueda de bases de datos sobre nanopartículas de oro se encuentra con el buscador de bases de datos de Google (link) se encuentra una base de datos interesante (Barnard y Opletal, 2019), que está disponible para su descarga gratuita en el siguiente (link). En esta se encuentra un conjunto de datos de 4000 configuraciones finales de nanopartículas de oro las cuales han sido optimizadas (totalmente relajadas) usando dinámica molecular con un potencial interatómico de átomo incrustado EAM Project (Dr. Howard Sheng, [Link]), a varias temperaturas y tasas de crecimiento. Es decir que las entradas que se encuentran en dicha base de datos se refieren a una interesante variedad de configuraciones de nanopartículas de oro en sus configuración más estable y características calculadas con un tipo de simulación molecular computacional que permite analizar el comportamiento o evolución de un sistema (físico, químico o biológico) a través del tiempo, calculando las fuerzas entre los átomos que lo conforman mediante las ecuaciones del movimiento de Newton.

Los tamaños de las nanopartículas presentes en el set de datos van desde 236 átomos hasta 14277 átomos, con configuraciones y regiones tanto cristalinas como no cristalinas. Cada nanopartícula ha sido ampliamente caracterizada en base a descriptores atómicos, moleculares, topológicos y cristalográficos,

A continuación se describe la convención de nomenclatura y se presenta un breve detalle de las características que se encuentran a disposición.

1. **ID**: Identificador único, nombre de archivo ID.xyz (disponibles para descargar en el siguiente [link](#))
2. **T**: Temperatura, K
3. τ : Tasa de crecimiento, atoms/ns
4. **time**: Tiempo, ns
5. **N total**: Número total de átomos
6. **N bulk**: Número total de átomos a granel
7. **N surface**: Número total de átomos en la superficie
8. **Volume** : Volumen total de nanopartículas m^3
9. **R min**: radio mínimo, Å
10. **R max**: Radio máximo, Å
11. **R diff**: Radio de nanopartícula mínimo, Å
12. **R avg**: Promedio de radio de nanopartícula, Å
13. **R std**: Desviación estándar del radio, Å
14. **R skew**: Sesgo del radio de nanopartícula, Å
15. **R kurt**: Curtosis del radio de nanopartícula Å
16. **S 100**: Número de átomos ubicados en superficies (100)
17. **S 111**: Número de átomos ubicados en superficies (111)
18. **S 110**: Número de átomos ubicados en superficies (110)
19. **S 311**: Número de átomos ubicados en superficies (311)
- 20-37. **Curve $n - n + 10$** : Átomos con ángulo de curvatura superficial entre n y $n + 9$ grados. Se divide en 18 intervalos de 10 grados.
38. **Avg total**: Número de coordinación promedio de todos los átomos
39. **Avg bulk**: Número de coordinación promedio de todos los átomos a granel
40. **Avg surf**: Número de coordinación promedio de todos los átomos de la superficie
- 41-61. **TCN n**: Cantidad de átomos totales con número de coordinación n , (n entre 0 y 20).
- 62-82. **BCN n**: Cantidad de átomos internos con número de coordinación n , (n entre 0 y 20).
- 83-103. **SCN n**: Cantidad de átomos superficiales con número de coordinación n , (n entre 0 y 20).
104. **Avg bonds**: Longitud promedio de enlace entre átomos de la nanopartícula, Å
105. **Std bonds**: Desviación estándar de la longitud del enlace entre átomos de la nanopartícula, Å
106. **Max bonds**: Longitud máxima de enlace, Å
107. **Min bonds**: Longitud mínima de vinculación, Å
108. **N bonds**: Número total de enlaces
109. **angle avg**: Ángulo de unión promedio
110. **angle std**: Desviaciones estándar del ángulo de enlace
111. **FCC**: Número de átomos en arreglo de Bravais cúbica centrada en las caras.
112. **HCP**: Número de átomos en arreglo de Bravais hexagonal empaquetada cerrada.
113. **ICOS**: Número de átomos en arreglo de Bravais icosaédrica
114. **DECA**: Número de átomos en arreglo de Bravais decaédrica

Las demás columnas se refieren a la cantidad de átomos totales/interiores/superficiales con armónico esférico $q6.q6 > 0,7$ de n . (n entre 0 y 20 y mayor a 20).

Por ultimo las dos columnas finales son etiquetas energéticas:

184. **Total E**: Energía total de la nanopartícula de la simulación LAMMPS, eV
185. **Formation E**: Energía de formación de la nanopartícula (Total E- N total*Bulk E/atom), eV.

La energía de granel por átomo (Bulk E/ átomo) se proporciona en el sitio web del proyecto de potenciales EAM (Dr. Howard Sheng, [Link](#)).

Para realizar el análisis de algunas de estas tantas variables, se agrupan en 5 conjuntos diferentes que capturan diferentes tipos de información sobre las nanopartículas. Los 5 conjuntos de características se denominan Bulk (B), Superficiales (S), Totales (T) y Principal (P). El conjunto de características de bulk solo contiene

características que pertenecen a los átomos interiores, incluidos los números de coordinación, los tipos de red y los parámetros de orden. El segundo conjunto solo contiene características que pertenecen a los átomos exteriores, incluidos los números de coordinación, las orientaciones de las facetas de la superficie y los rangos de curvaturas de la superficie. El conjunto de características totales solo contiene características que involucran medidas globales (como el radio de nanopartículas), promedios y números de coordinación y parámetros de orden calculados sobre todos los átomos en la partícula.

Finalmente, el conjunto principal contiene una combinación de varias funciones de los conjuntos de bulk, de superficie y totales, que serán de principal interés en el presente trabajo en el momento de realizar el análisis de datos sobre las nanopartículas.

4. Análisis de datos:

A continuación se realiza trabajo estadístico sobre algunas de las variables. Se utiliza la muestra aleatoria de 4000 configuraciones y a partir de esta se realizan test de hipótesis con estimaciones de parámetros y análisis de correlaciones, los cuales permitan formular conclusiones sobre la estructura y energía interna de las nanopartículas.

4.1. Longitud media de enlaces

Esta variable es continua y mide la distancia promedio entre átomos dentro de cada nanopartícula. Se realiza contraste de hipótesis para validar o invalidar la afirmación de que las distintas longitudes de enlace sigue una distribución normal.

Uno de los métodos gráficos más empleados para el análisis de normalidad consiste en representar los datos mediante un histograma y superponer la curva de una distribución normal con la misma media y desviación estándar que los datos disponibles. Con el histograma (Fig. 1) se observa que la distribución tiende a la normalidad.

Los test de hipótesis teóricos más empleados para analizar la normalidad son los de Shapiro-Wilk y D'Agostino's K-squared (d'Agostino, 1971). En ambos, se considera como hipótesis nula que los datos proceden de una distribución normal. El p-value de estos test indica la probabilidad de obtener unos datos como los observados si realmente procediesen de una población con una distribución normal con la misma media y desviación que estos. Por lo tanto, si el p-value es menor que un determinado valor (típicamente 0.05), entonces se considera que hay evidencias suficientes para rechazar la normalidad.

Como el test de Shapiro-Wilk se desaconseja cuando se dispone de muchos datos (más de 50), se verifica la normalidad de la distribución de longitud de enlace entre átomos de nanopartículas de oro con el test k^2 de D'Agostino. Se halla un *P-valor* nulo lo que indica que hay que rechazar la hipótesis nula y no es posible asegurar que la variable longitud de enlace sigue una distribución normal. Con un gráfico de cuantiles teóricos (Gráficos Q-Q), el cual compara los cuantiles de la distribución observada con los cuantiles teóricos de una distribución normal con la misma media y desviación estándar que los datos, es posible observar (Fig. 2) que los datos se desvían de la recta que indica normalidad. Se procede a detectar valores atípicos (outliers) de la variable en estudio, pero en el caso de la longitud de enlace no se encuentran.

Finalmente como la muestra de 4000 configuraciones es suficientemente grande, es posible tomar muestras aleatorias, calcular sus medias y a partir de ellas hallar la media muestral, que va a ser el valor medio de la longitud de enlace aceptado finalmente en el presente análisis. Es decir, para calcular la media muestral de longitud de enlace se seleccionan, de los datos, aleatoriamente n muestras cada vez, un total de m veces. Luego, se toman los m grupos de muestras para encontrar el valor promedio respectivamente. Finalmente se verifica el teorema del límite central que asegura que la distribución de estos promedios se acerca a una distribución normal.

El teorema del límite central asegura que si se toma un número suficiente de muestras aleatorias de tamaño suficiente de cualquier tipo de distribución con alguna varianza, la distribución de las medias muestrales será una distribución normal. Esta nueva distribución se llama distribución de muestreo. La media de la distribución muestral debe ser aproximadamente igual a la media de la población, valor de interés en el presente análisis ya que será la longitud de enlace medio aceptado para las nanopartículas de oro de entre 236 y 14277 átomos. La desviación estándar de las medias muestrales se denomina más comúnmente error estándar (SE). Lo interesante

del teorema del límite central es que es posible calcular el error estándar simplemente usando la desviación estándar de la población y el tamaño de la muestra.

Como la muestra de 4000 configuraciones es suficientemente grande, es posible tomar muestras aleatorias de mas de 30 configuraciones cada una. Se toman 1000 muestras de 200 configuraciones cada una y se calcula la media muestral y su error estandar. Se obtiene la gráfica (Fig. 3) y longitud de enlace media con error estandar:

$$\bar{L} = (2,914 \pm 0,002)\text{\AA}$$

En la bibliografía (Bouju, 2017) se encuentra que las distancias interatómicas entre átomos de oro en bulk es de $2,8 \pm 0,1$, con lo cual el valor encontrado mediante el análisis de datos no difiere significativamente de este valor y es posible asegurar que la distancia entre átomos internos de las nanopartículas no tiene implicación en las diferencias que estas tienen con el oro metálico en bulk.

Además mediante técnicas experimentales (Legoas y col., 2002; Miller y col., 2006) se encuentra que a medida que disminuye el tamaño de las partículas metálicas, la longitud del enlace metal-metal tiende a disminuir. En general, existe un buen acuerdo utilizando varias técnicas, lo que indica que en películas delgadas y partículas metálicas sin soporte de Cr, Fe, Cu, Ag, Pd, Au y Pt menores de aproximadamente 30 Å, se produce un acortamiento de la distancia de enlace interatómico considerable. Por ejemplo, la distancia de enlace del Au evaporado en Mylar de una partícula de radio 30Å tiene una distancia de enlace de 2,84Å, mientras que una partícula de radio 8Å tenía una distancia de enlace de 2,72Å.

Esta información puede corroborarse en la base de datos analizada en el presente trabajo. Ya que se cuenta con nanopartículas de esos tamaños.

Además se realiza un análisis de correlación entre la cantidad de átomos que conforman la nanopartícula y la longitud de enlace y se observa (Fig 4) que efectivamente existe dicha relación. A menor cantidad de átomos se pueden encontrar distancias interatómicas más pequeñas. Como no se conoce la distribución y las posibles relaciones entre dos variables, pero si las dos muestras de datos son de la misma longitud, se utiliza el coeficiente de correlación de Spearman como herramienta para validar o rechazar la hipótesis nula de que la composición de la nanopartícula y la longitud de enlace son independientes.

Como el *P-valor* da cero se rechaza la hipótesis nula y se puede asegurar que existe correlación significativa entre el número de átomos que conforman la nanopartícula y la longitud de enlace entre dos de ellos.

4.2. Coordinación promedio

Otra característica estructural del oro en granel, que es conocida en física de materia condensada, es que los átomos adoptan una configuración estructural cúbica centrada en las caras o FCC (Face Centered Cubic). Es decir que cada celda unitaria consta de ocho átomos en las esquinas de un cubo y un átomo en el centro de cada una de las caras. Como cualquier sólido cristalino se construye a partir de la repetición de esta estructura elemental celda unitaria, en el caso del oro cristalino cada átomo tendrá a su alrededor 12 átomos vecinos. Dicho valor se conoce como número de coordinación.

En la presente sección se analiza si esto se cumple también en el caso de tener estructuras con cantidad de átomos pequeña. Referido a esto, en la base de datos se encontraban tres variables, el número de coordinación promedio de todos los átomos, de los átomos internos y de los de la superficie. Al sacar el promedio de coordinación de los átomos internos, sale que es 12,1 con una desviación estándar de 0,2.

Para no trabajar con promedios, es posible utilizar la subtabla de las características internas, osea el subdataframe llamado de bulk y realizar un análisis gráfico a partir de las variables de número de coordinación en bulk de 0 a 20 (BCN_0 a BCN_{20}).

Se emplea el diagrama de tallo y de hojas de *matplotlib*. Este gráfico se utiliza para mostrar la frecuencia absoluta en diferentes clases, es decir mostrar la relación entre una variable numérica y una categórica, en este caso la cantidad de átomos con un determinado número de coordinación y la etiqueta de ese número de coordinación. El gráfico de diagrama de tallo y hoja es principalmente adecuado para conjuntos de datos pequeños, por lo que primero se realiza para una muestra aleatoria de 40 configuraciones (Fig 5) y luego con todas las configuraciones (Fig 6).

y se observa que efectivamente para los átomos en bulk, su coordinación es 12 la mayoría de las veces. Si se calcula el porcentaje da que más del 98 % de las estructuras tienen sus átomos internos con 12 átomos vecinos a su alrededor.

Si se realiza el mismo análisis pero ahora utilizando la sub base de datos de características superficiales, se observa lo esperado de que la coordinación ahora será menor y tendrá más dispersión, alrededor del valor típico de 9 átomos vecinos, ver (Fig 7)

4.3. Energía

A partir de esta sección se comienza a analizar una de las características fundamentales de las nanopartículas que es su energía. Unos de los atributos de la base de datos es la energía total de cada configuración calculada mediante el simulador masivo paralelo atómico/molecular a gran escala (LAMMPS) que es un programa de dinámica molecular de los Laboratorios Nacionales Sandia.(, [Link](#)). LAMMPS hace uso de la interfaz de paso de mensajes (MPI) para la comunicación paralela y es un software gratuito y de código abierto (, [GitHub](#)), distribuido bajo los términos de la Licencia Pública General GNU.

Lo primero que se analiza es si existe correlación entre la energía total y la cantidad de átomos que forman la nanopartícula. Se formula el test de hipótesis, donde la hipótesis nula es que las variables son independientes entre ellas. Se utiliza un nivel de confianza del 95 %. Es decir que si se obtiene $P\text{ valor} < 0,05$, se rechaza la hipótesis nula y existe una correlación significativa entre las variables estudiadas. En caso contrario no es posible afirmar que la correlación difiera significativamente de 0, es decir no es posible afirmar que el número de átomos que forman la nanopartícula influya a la energía de la misma.

En primer lugar se representan las dos variables mediante un diagrama de dispersión (scatterplot de matplotlib) para intuir gráficamente si existe relación o no. Se observa en la (Fig 8) que los puntos claramente forman una recta con pendiente negativa la cual indica una fuerte relación lineal inversa entre las variables.

De todas formas se busca el coeficiente de correlación. Para poder elegir el adecuado, es decir el estadístico que cuantifica la asociación lineal entre dos variables numéricas, se tiene que analizar el tipo de variables y la distribución que presentan. Existen diferentes tipos, de entre los que destacan el Pearson, Rho de Spearman y Tau de Kendall.

En este caso, ambas variables son cuantitativas continuas y pueden ordenarse para convertirlas en un ranking, por lo que, a priori, los tres coeficientes de correlación podrían aplicarse. La elección se hará en función de la distribución que presenten las observaciones. Si se analizan los gráficos laterales de la (Fig 8) las distribuciones respectivas, de números de átomos y de energías muestran evidencias suficientes de que no se puede asumir normalidad en ninguna de las dos variables. Siendo estrictos, este hecho excluye la posibilidad de utilizar el coeficiente de Pearson, dejando como alternativas el de Spearman o Kendall. Por último se escoge el de Spearman debido a la gran cantidad de datos con los que se cuenta.

Finalmente se acepta una correlación lineal inversa alta, con coeficiente de correlación de Spearman prácticamente igual a -1 , y existe claras evidencias estadísticas de que la relación observada no se debe al azar $P\text{ valor}$ nulo. Lo cual está de acuerdo con la base teórica física de que cada átomo aportará al sistema un cierto grado de energía.

Para lograr calcular un estadístico que describa a la energía de las distintas configuraciones, se independiza de la cantidad de átomos, y se trabaja con energía por átomo.

Con el objetivo de hallar la energía por átomo media, se realiza el procedimiento de análisis de normalización, trabajo de valores atípicos y aplicación del teorema del límite central, análogo al realizado en la sección 4.1 con la variable longitud de enlace.

La diferencia ahora es que se encuentran outliers (Fig. ??, pero no genera gran inconveniente ya que solo son 49 de 4000 datos, por lo que se procede a eliminarlos y no tenerlos en cuenta en los próximos análisis de energía por átomo.

Luego de generar 1000 muestras de 200 configuraciones aleatorias, se tiene la distribución muestral normal (Fig ??) se encuentra la media muestral de la energía por átomo y su error estándar:

$$\overline{E} = (-3,609 \pm 0,006) \text{Å}$$

En la bibliografía (Lee y col., 2003) se encuentra que la energía por átomo para el oro a granel es de $-3,93$, también con cálculos realizados mediante simulaciones.

Existe una discrepancia significativa con el valor hallado. Por esta razón se realiza con el objetivo de ver si esto es una tendencia generada debido a la disminución del tamaño, una correlación entre la energía por átomo y el número de átomos que conforman la nanopartícula.

Se realiza la correlación siguiendo la metodología aplicada en el caso de la variable longitud de enlace. Planteando como hipótesis nula que las variables número de átomos y energía por átomo son independientes. Mediante el p valor de Spearman se llega a la conclusión de que se debe rechazar la hipótesis nula y aceptar que existe correlación significativa entre las variables. Gráficamente (Fig 11) se observa que efectivamente la energía las nanopartículas es más alta (menos negativa) a medida que el tamaño, es decir la cantidad de átomos se reduce.

4.4. Energía y coordinación

Por último se realiza test de hipótesis de si la energía por átomo esta correlacionada con la coordinación de los átomos internos y/o con la coordinación de los átomos externos.

Nuevamente se realiza análisis visual de la posible correlación, mediante un gráfico de dispersión entre las variables y numéricamente con el coeficiente y p valor de Spearman.

En el caso de la correlación entre energía por átomo con la coordinación que tengan los átomos internos, se obtiene un p valor de Spearman mayor a 0.05. Con lo cual no es posible afirmar correlación significativa entre las variables. También gráficamente (Fig. 12). Concluyendo así que la energía por átomo no tiene relación con el comportamiento de los átomos internos de las nanopartículas.

Por el contrario, al realizar el análisis de correlación entre la energía por átomo y la coordinación de los átomos superficiales, se obtiene p valor nulo. Por lo tanto las variables están correlacionadas. Lo cual también es posible observar gráficamente (Fig 13). Pudiendo concluir así que la energía por átomo se ve afectada por el comportamiento de los átomo superficiales de las nanopartículas.

5. Conclusiones:

En resumen, a lo largo de el presente análisis de datos sobre 4000 configuraciones, se llegaron a las siguientes afirmaciones sobre la población de nanopartículas de entre 236 a 14277 átomos:

- La media muestral de longitud de enlace es $\bar{L} = (2,914 \pm 0,002)\text{\AA}$, la cual no discrepa significativamente con la del oro en bulk.
- La longitud de enlace se modifica según la cantidad de átomos que conforman la nanopartícula.
- Es posible confirmar que la coordinación del 98 % de los átomos internos de las nanopartículas tiene coordinación 12. Lo cual coincide con la coordinación de átomos de oro en bulk.
- La energía total de la nanopartícula está correlacionada con la cantidad de átomos. Se observa una fuerte relación lineal inversa.
- La energía por átomo tiene una media de $\bar{E} = (-3,609 \pm 0,006)\text{\AA}$ la cual discrepa con la energía por átomo del oro en bulk.
- Se observa correlación entre dicha energía y la cantidad de átomos que conforman la nanopartícula.
- Se observa que no existe correlación entre la energía por átomo y la coordinación de los átomos internos. Pero si existe correlación entre la energía y los átomos externos.

Finalmente es posible concluir, a pesar de que se deberían comprobar y correlacionar más variables, que las propiedades de las nanopartículas que difieren de las propiedades del oro se deben a las características superficiales de las primeras, ya que sus estructuras internas coinciden.

Bibliografía

- . (Link). Laboratorio Nacional de Sandia [<https://www.sandia.gov/about/>, Last accessed on 2017-11-30].
- . (GitHub). LAMMPS software package. [<https://github.com/lammps/lammps>, Last accessed on 2017-11-30].
- Barnard, A. & Opletal, G. (2019). Gold Nanoparticle Data Set. <https://doi.org/10.25919/5D395EF9A4291>
- Bouju, X. (2017). Gold monolayer islands on a polar AlN (0001) surface.
- d'Agostino, R. B. (1971). An omnibus test of normality for moderate and large size samples. *Biometrika*, 58(2), 341-348.
- Dr. Howard Sheng. ([Link]). EAM potentials project. <https://sites.google.com/site/eampotentials/>
- Lee, B.-J., Shim, J.-H. & Baskes, M. (2003). Semiempirical atomic potentials for the fcc metals Cu, Ag, Au, Ni, Pd, Pt, Al, and Pb based on first and second nearest-neighbor modified embedded atom method. *Physical Review B*, 68(14), 144112.
- Legoas, S. B., Galvao, D. S., Rodrigues, V. & Ugarte, D. (2002). Origin of anomalously long interatomic distances in suspended gold chains. *Physical review letters*, 88(7), 076105.
- Miller, J., Kropf, A., Zha, Y., Regalbuto, J., Delannoy, L., Louis, C., Bus, E. & van Bokhoven, J. A. (2006). The effect of gold particle size on AuAu bond length and reactivity toward oxygen in supported catalysts. *Journal of Catalysis*, 240(2), 222-234.
- Rasmussen, K., Rauscher, H., Mech, A., Sintes, J. R., Gilliland, D., González, M., Kearns, P., Moss, K., Visser, M., Groenewold, M. y col. (2018). Physico-chemical properties of manufactured nanomaterials-Characterisation and relevant methods. An outlook based on the OECD Testing Programme. *Regulatory Toxicology and Pharmacology*, 92, 8-28.
- Schmid, G. (2011). *Nanoparticles: from theory to application*. John Wiley & Sons.

A. Figuras:

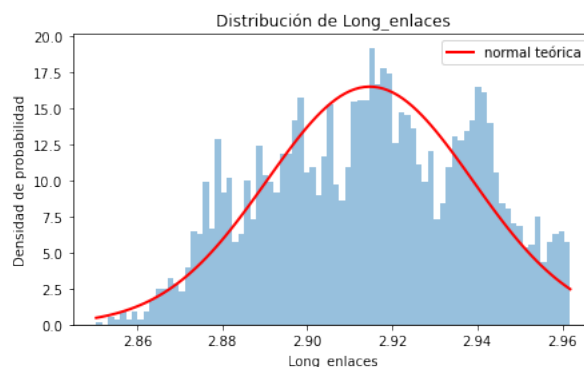


Figura 1: Análisis de normalidad para la longitud de enlace mediante histograma

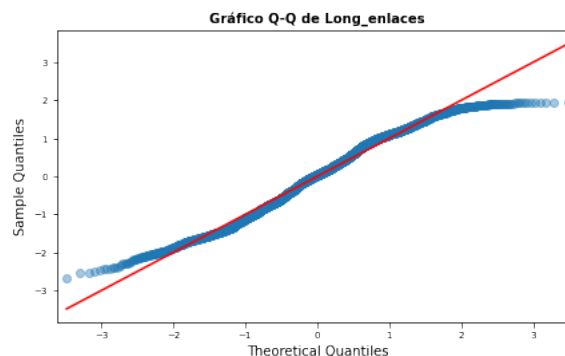


Figura 2: Análisis de normalidad para la longitud de enlace mediante gráfico Q-Q

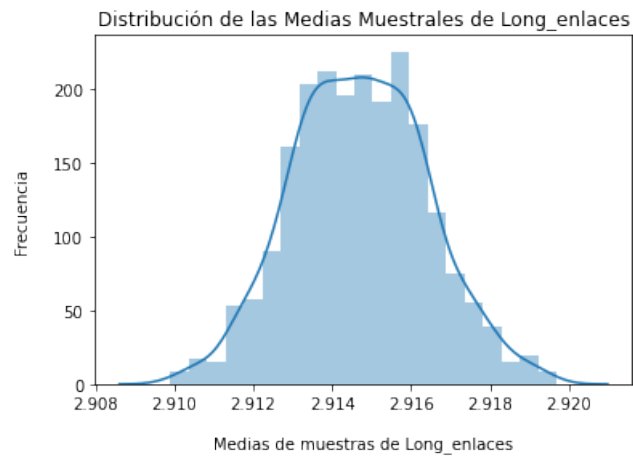


Figura 3

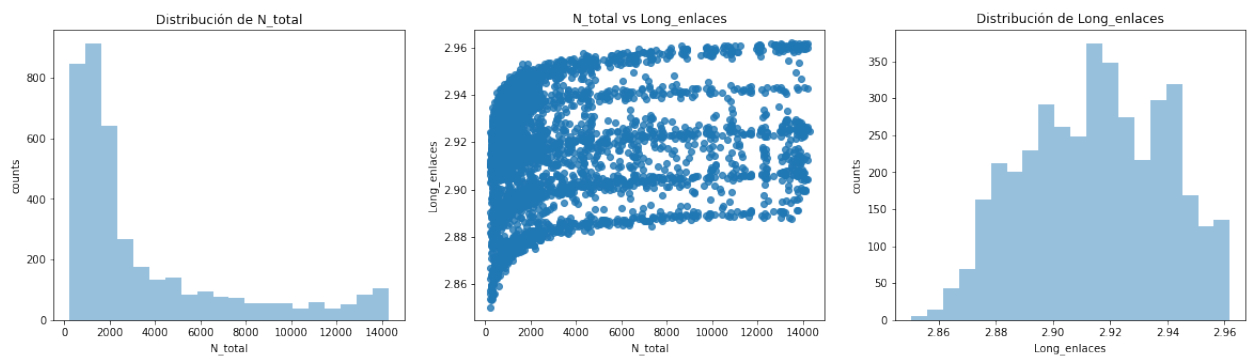


Figura 4: Correlación entre la longitud de enlace y la cantidad de átomos

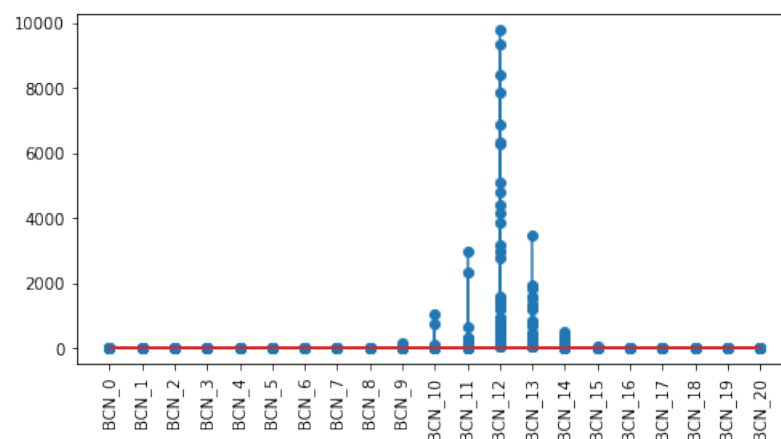


Figura 5: Distribución de coordinación entre átomos internos vecinos. Solo 40 configuraciones aleatorias

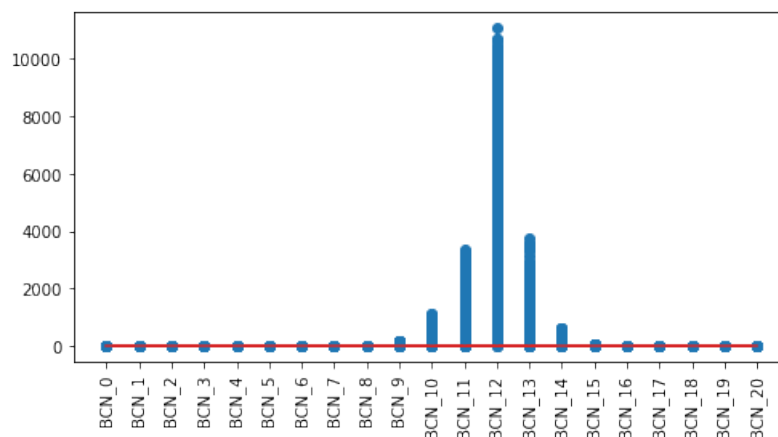


Figura 6: Distribución de coordinación entre átomos internos vecinos. Todas las configuraciones

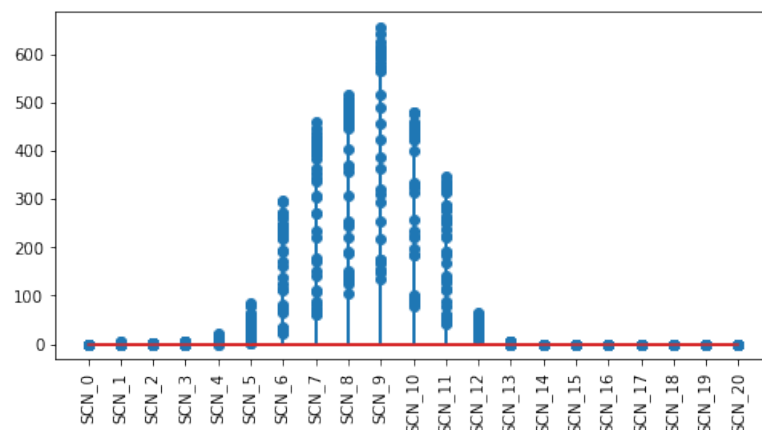


Figura 7: Distribución de coordinación entre átomos superficiales vecinos. Solo 40 configuraciones aleatorias

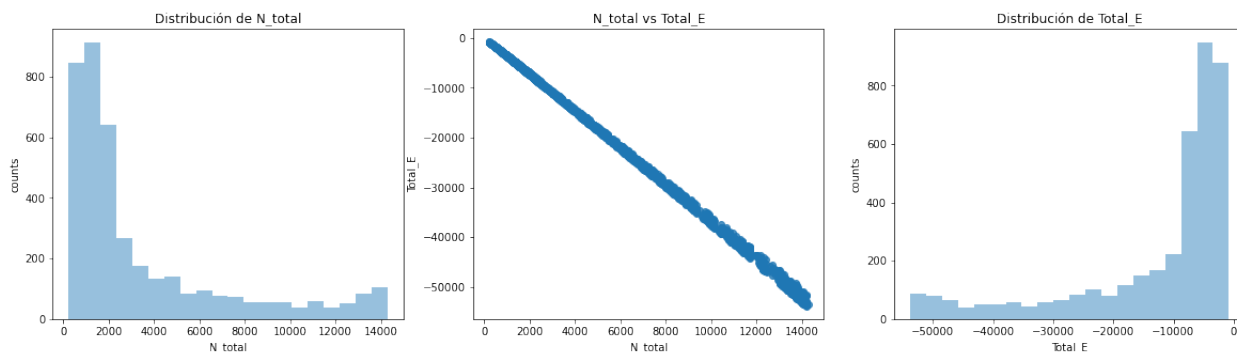


Figura 8: Distribución de coordinación entre átomos superficiales vecinos. Solo 40 configuraciones aleatorias

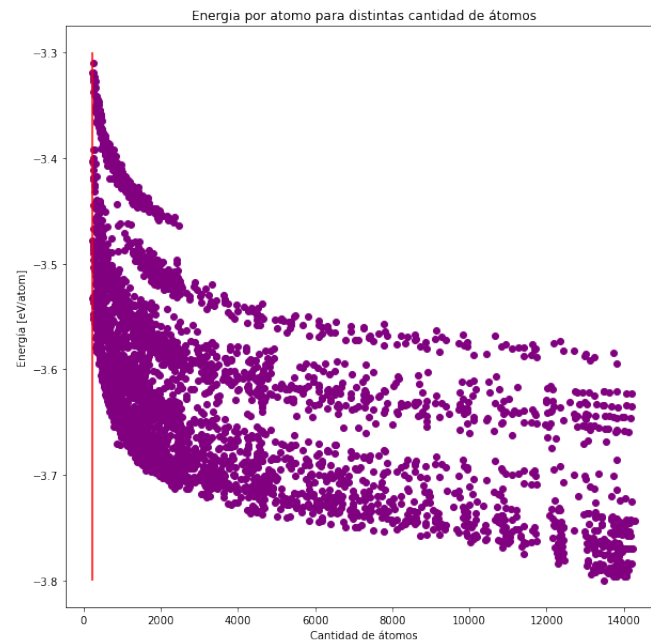


Figura 9: Correlación entre cantidad de átomos que conforman la estructura y su energía.

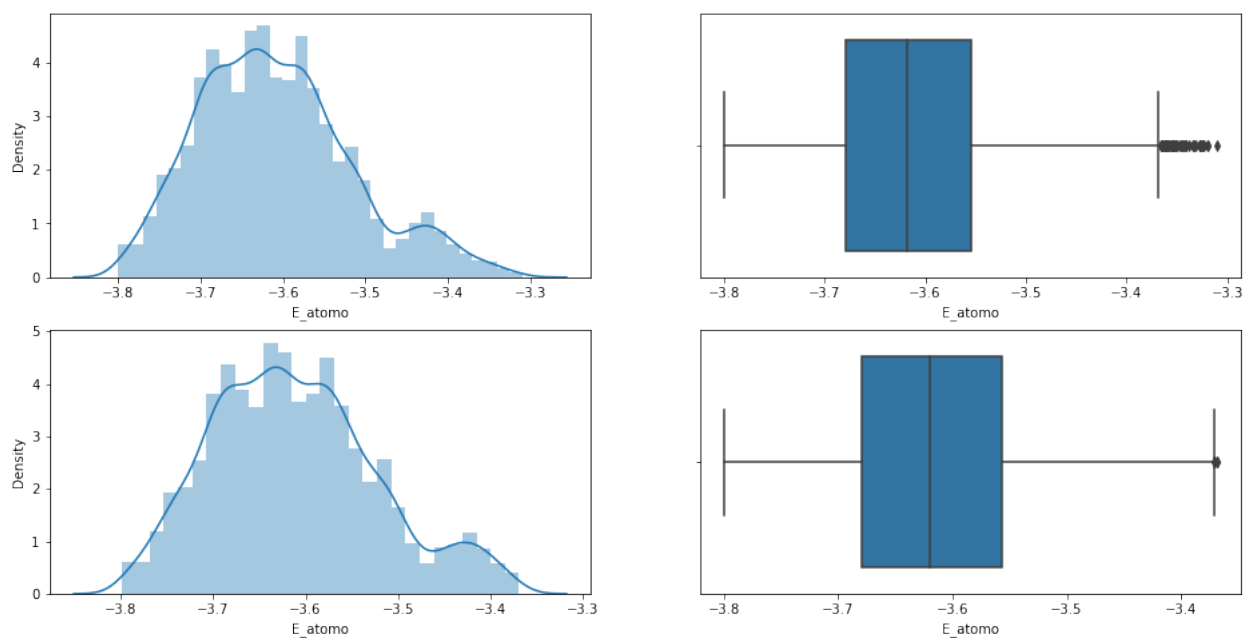


Figura 10: Gráfico de cajas que determinan outliers de la distribución de energías.

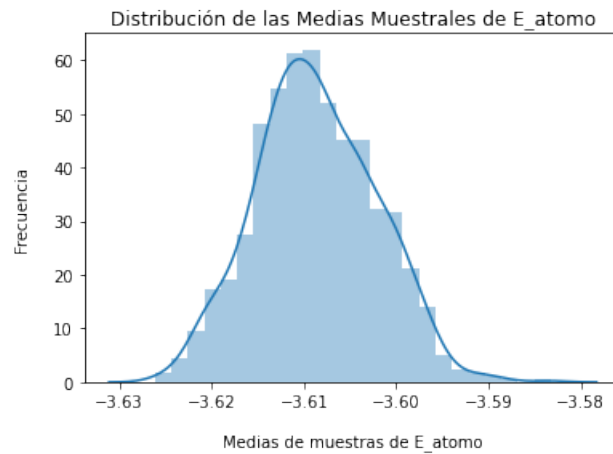


Figura 11

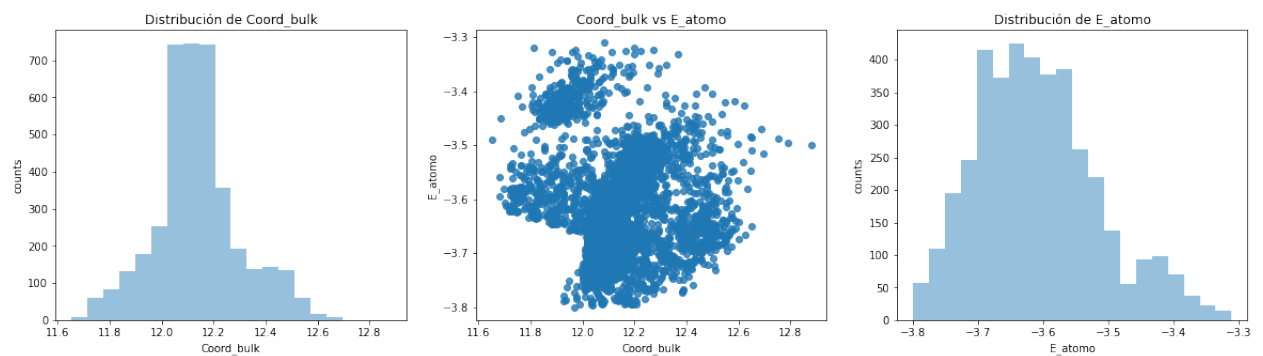


Figura 12: Correlación entre la coordinación de los átomos internos de la nanopartícula y la energía por átomo de la misma.

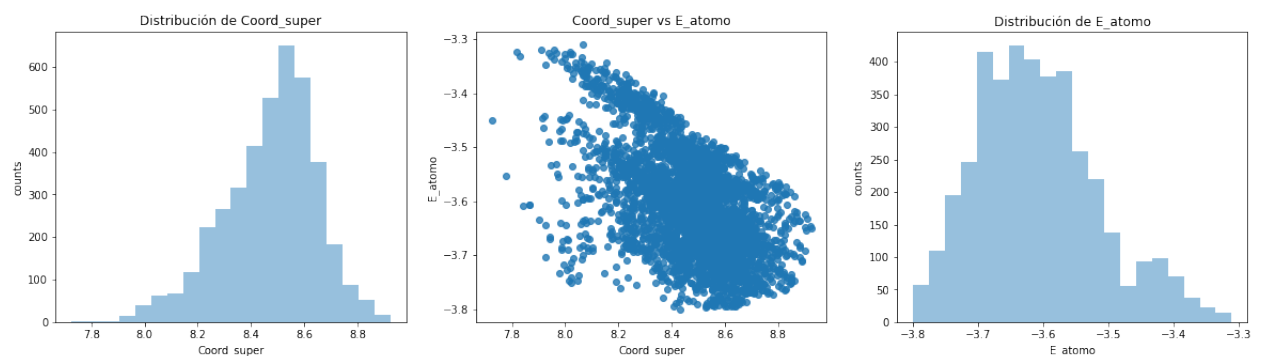


Figura 13: Correlación entre la coordinación de los átomos superficiales de la nanopartícula y la energía por átomo de la misma.