Logistic regression and Bayes-classifier study of classification of songs to genres based on timbre, pitch and rhythm of the music signal

Ville Virkkala
Aalto University, P.O. Box 11100, FI-00076 Aalto, Finland
(Dated: November 30, 2017)

A genre of a song can be estimated based on its music signal's characteristics . In this work we use two linear classifiers, Bayes Classifer and Logistic classifer, to classify songs into one of ten possible genres. The two classifers are trained against the training data and their performance is compared against each other. In this work we show that both classifiers perform much better compared to random guess. However their cabability to classify all songs is clearly limited the accuracy for both classifiers being around 60%.

I. INTRODUCTION

An automatic music transcription, *i.e.*, notating a piece of music to a speficic genre, *e.g.*, Blues, dates back to 1970s when first attempts towards automatic music transcription were made[1]. Since then interest in automatic transcription of music has grown rapidly and various approaches, statistical methods, modelling human auditory system, have been applied to music transcription problem. However even today an expert human musician often beats a state-of-the-art automatic transcription system in accuracy.

Characteristics of music signal that are useful in classification of a song are *timbre*, *rhytm*, *pitch*, *loudness* and *duration*[1] from which the three first one, descriped below are used in this work.

- The timbre of the music can be most easily described as the factor which separates two sources of music from each other. For example if the same song is played by violin or a quitar the timbre is called the characted which separates the violin from the quitar.
- The pitch is related to frequency scale of a song a can be defined as the frequency of the sine-wave fitted to target sound by human listener.
- The rhythm of the music can be described as arrangement of sounds as time flows.

In classification problem the object is classified into a certain class based on it's characteristics called features. A linear classifier does the classification by making a linear combination of the features and converting the resulting value into a class or a probability that the object belongs to given class. In logistic regression the feature vector of the object is transformed into a probality by taking a linear compination of features and mapping the result into interval $[0, 1]$ using a sigmoid function. The Bayes-classifier incontrast assumes that the feature vector is drawn from a multidimensional-gaussian distribution. The posterior probalility of the object belonging to a certain class is then obtained as a product of the prior of the class and the probability of to sample the given feature vector from it multidimensional gaussian distribution.

The paper is organized as follows. The used data-set and the computational methods are described in detail in Sec. II. In Sec. III the results for the both logistic regression- and Bayes-classifier are given. Sec. IV is a summary of the results and the differences between the two classfiers are discusssed.

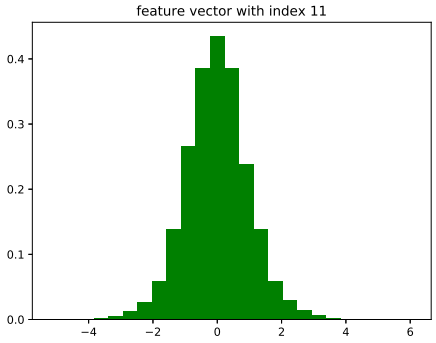
II. USED DATA-SET AND COMPUTATIONAL METHODS

A. Used data-set

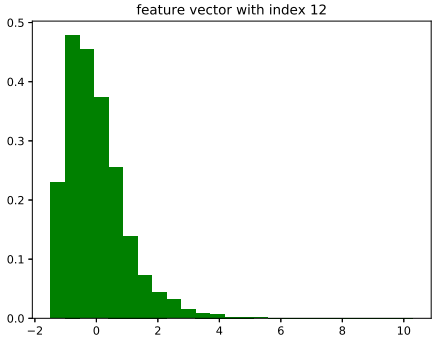
The data-set consisted of 4363 songs and was divided into training and test data sets including every third song to test set and rest of the songs to training set. Each song contained 264 features and the songs were labeled to 10 different categories. The gatecories were: 1 Pop Rock, 2 Electronic, 3 Rap, 4 jazz, 5 Latin, 6 RnB, 7 International, 8 Country, 9 Reggae and 10 Blues. The musical characteristics of the songs were packed to a feature vector of length 256. The first 48 elements in the feature vector can be associated to timbre, the next 48 elements to pitch and the final 168 features to rhythm. The distribution of the features resembled in most cases a gaussian distribution or a skew symmetric distribution. This is illustrated figures 1a and 1b.

B. Computational methods

In this work two different methods were used to classify the songs to different genres. First method is the gradient descent method in which the logistic-loss is minimized iteratively using the gradient descent method. The other method used is the Bayes-classifier which classifies the song to certain gategory that gives the maximum posterior probability with respect to label i. Both methods are described below in detail. In addition we studied the effect of feature extraction and for that purpose we used principal component analysis method to exlude features with little impact.



(a)



(b)

FIG. 1: Visualization of typical distributions of features, gaussian distribution (a) and skew-symmetric distribution (b).

1. *Gradient descent method*

In logistic regression for binary classifier problem the starting point is the minimization of the loss function

$$\mathcal{E}((\mathbf{x},y);\mathbf{w}) = \min_{\mathbf{w}\in\mathbf{R}^2} (1/N)\sum_{i=1}^{N} L((\mathbf{x}^{(i)},y^{(i)});\mathbf{w}),$$

where the logistic loss $L((\mathbf{x},y);\mathbf{w})$ is defined as $L((\mathbf{x},y);\mathbf{w}) = \ln\left(1+\exp\left(-y(\mathbf{w}^T\mathbf{x})\right)\right)$ and $\mathbf{x}$ is the feature vector of a music signal, $\mathbf{w}$ are the coefficients of the linear expansion and y is the label 1 or -1 whether the song belongs to certain category or not. The minimization problem can be further converted to

$$\begin{aligned}
\mathcal{E}((\mathbf{x},y);\mathbf{w}) &= \frac{1}{N}\max_{\mathbf{w}\in\mathbf{R}^2}\sum_{i=1}^{N}\ln\left(\frac{1}{1+\exp(-y^i(\mathbf{w}^T\mathbf{x}))}\right)\\
&= \frac{1}{N}\max_{\mathbf{w}\in\mathbf{R}^2}\Big(\sum_{y^i=1}\ln\left(\frac{1}{1+\exp(-\mathbf{w}^T\mathbf{x}^{(y^i)})}\right)\\
&\quad+\sum_{y^i=-1}\ln\left(1-\frac{1}{1+\exp(-\mathbf{w}^T\mathbf{x}^{(y^i)})}\right)\Big) \quad (1)\\
&= \max_{\mathbf{w}\in\mathbf{R}^2}\Big(\sum_{y^i=1}\ln(p_{y^i=1})\\
&\quad+\sum_{y^i=-1}\ln(1-p_{y^i=1})\Big), \quad (2)
\end{aligned}$$

where $p_{y^i=1}$ is the probability that the song i is labeled belonging to certain category. There is no closed form solution for equation (2) and for that reason some numberical iterative solver must be used to find the optimal $\mathbf{w}$. One of the most popular methods to find the optimal solution is gradient descent (GD) method. In GD the weights $\mathbf{w}$ are updated at each iteration $\mathrm{k}+1$ according to equation

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} - \alpha\nabla\mathcal{E}(\mathbf{w}^{(k)}). \quad (3)$$

To be able to use the GD method we need know the gradients $\nabla\mathcal{E}(\mathbf{w}^{(k)})$. By marking $t^i = \max(0,y^i)$ we can write the equation (2) as

$$\begin{aligned}
\mathcal{E}((\mathbf{x},y);\mathbf{w}) = \min_{\mathbf{w}\in\mathbf{R}^2} -\sum_{i}^{N}&(t^i\ln(p_{y^i})\\
&+(1-t^i)\ln(1-p_{y^i})). \quad (4)
\end{aligned}$$

The derivative of p_{y^i} with respect to w_j can be written as $\frac{\partial p_{y^i}}{\partial w_j} = \frac{\partial p_{y^i}}{\partial(\mathbf{wx})}\frac{\partial(\mathbf{wx})}{\partial w_j} = p_{y^i}(1-p_{y^i})x_j^i$. Thus taking derivative of equation (4) with respect to w_j we get for $\nabla\mathcal{E}(w_j^k)$.

$$\begin{aligned}
\nabla\mathcal{E}(w_j^k) &= -\sum_i\Big(t^i\frac{1}{p_{y^i}}(1-p_{y^i})p_{y^i}x_j^i\\
&\quad+(1-t^i)\frac{1}{1-p_{y^i}}(-1)p_{y^i}(1-p_{y^i})x_j^i\Big)\\
&= -\sum_i\left(t^i - t^ip_{y^i} - p_{y^i} + t^ip_{y^i}\right)x_j^i\\
&= -\sum_i(t^i-p_{y^i})x_j^i. \quad (5)
\end{aligned}$$

Now the $\nabla\mathcal{E}(w_j^k)$'s can be written in vector form as

$$\nabla\mathcal{E}(\mathbf{w}^k) = -\left(\hat{\mathbf{y}} - \frac{1}{1+\exp(-(\mathbf{w}^k)^TX^T)}\right)^T X, \quad (6)$$

where $\hat{\mathbf{y}} = \max(\mathbf{0},\mathbf{y})$ is a vector of length N for which the i-th element is zero if $y_i = -1$ and one if $y_i = 1$. The matrix X is a matrix which rows are the feature vectors $\mathbf{x}^i$. Thus we can solve the minimization problem (4) iteratively using the GD method (3) and the gradients (6).

2. Bayes classifer

For a Bayes classifier we assume that the distribution of the feature vector of a music signal with respect to label y_i is a Gaussian distribution

$$p(\mathbf{x}|y_i;\mathbf{m}_i,\mathbf{C}_i) = \frac{1}{\sqrt{\det\{2\pi\mathbf{C}_i\}}} e^{-(1/2)(\mathbf{x}-\mathbf{m}_i)^T\mathbf{C}_i^{-1}(\mathbf{x}-\mathbf{m}_i)}. \tag{7}$$

Using the Baye's theorem the posterior probability $p(y_i|\mathbf{x};\mathbf{m}_i,\mathbf{C}_i)$ can be written as

$$p(y_i|\mathbf{x};\mathbf{m}_i,\mathbf{C}_i) = \frac{p(y_i)p(\mathbf{x}|y_i;\mathbf{m}_i,\mathbf{C}_i)}{p(\mathbf{x})}, \tag{8}$$

where the $p(\mathbf{x})$ is a normalization constant and can be omitted. To be able to use the equation (8) we need to find optimal values for parameters $p(y_i)$, $\mathbf{m}_i$ and $\mathbf{C}_i$. The prior $p(y_i)$ can be simply estimated as the fraction of labels y_i among all labels. Because the samples $\mathbf{x}^{(t)}, y^{(t)}$ are independent the parameters $\mathbf{m}_i$ and $\mathbf{C}_i$ and can be obtained by maximizing the respective log-likelihood function with respect to the parameters. The log-likelihood can be written as

$$\begin{aligned}\mathcal{L}_{y=i} &= \sum_{y^t=i} \log(P(\mathbf{x}^t|y^t=i;\mathbf{m}_i,\mathbf{C}_i)) \\ &= \sum_{y^t=i}(-\frac{1}{2}\log(2\pi^n) - \frac{1}{2}\log(\det(\mathbf{C}_i)) \\ &\quad -\frac{1}{2}(\mathbf{x}^t-\mathbf{m}_i)^T\mathbf{C}_i^{-1}(\mathbf{x}^t-\mathbf{m}_i)).\end{aligned} \tag{9}$$

Now the optimal $\mathbf{m}_i$ is obtained by setting the derivative of $\mathcal{L}_{y=i}$ with respect to $\mathbf{m}_i$ to zero, *i.e.*,

$$\frac{\partial\mathcal{L}_{y=i}}{\partial\mathbf{m}_i} = \mathbf{C}_i^{-1}\sum_{y^t=i}(\mathbf{x}^t-\mathbf{m}_i). \tag{10}$$

Setting eq. (10) to zero and solving for $\mathbf{m}_i$ gives $\mathbf{m}_i = \frac{\sum_{y^t=i}\mathbf{x}^t}{N}$. Similarly the $\mathbf{C}_i$ is obtained by setting the derivative of $\mathcal{L}_{y=i}$ with respect to $\mathbf{C}_i$ to zero giving

$$\begin{aligned}\frac{\partial\mathcal{L}_{y=i}}{\partial\mathbf{C}_i} &= \sum_{y^t=1}(-\frac{1}{2}\mathbf{C}_i^{(-1)} \\ &\quad +\frac{1}{2}\mathbf{C}_i^{(-1)}(\mathbf{x}^t-\mathbf{m}_i)(\mathbf{x}^t-\mathbf{m}_i)^T\mathbf{C}_i^{-1}).\end{aligned} \tag{11}$$

Setting eq. (11) to zero and solving for $\mathbf{C}_i$ gives $\mathbf{C}_i = \frac{1}{N}\sum_{y^t=i}(\mathbf{x}^t-\mathbf{m}_i)(\mathbf{x}^t-\mathbf{m}_i)^T$. In equations (10) and (11) the identieties $\frac{\partial(\mathbf{x}-\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{s})}{\partial\mathbf{s}} = 2\mathbf{W}(\mathbf{x}-\mathbf{s})$, $\frac{\partial\ln|\det(\mathbf{X})|}{\partial\mathbf{X}} = (\mathbf{X}^T)^{-1}$ and $\frac{\partial\mathbf{a}^T\mathbf{X}^{-1}\mathbf{b}}{\partial\mathbf{X}} = -\mathbf{X}^{-T}\mathbf{a}\mathbf{b}^T\mathbf{X}^{-T}$ from matrix cook-book[2] were used. In classification the parameters $p(y_i)$, $\mathbf{m}_i$ and $\mathbf{C}_i$ are optimized for all ten classes using the training data. The song is then classified to certain category i that maximizes the posterior probability (8), *i.e.*,

$$i = \underset{i}{\operatorname{argmax}}\, p(y_i|\mathbf{x};\mathbf{m}_i,\mathbf{C}_i). \tag{12}$$

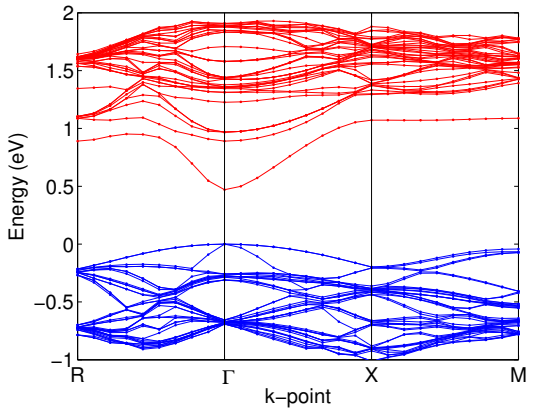


FIG. 2: (Color online) Band structure of the periodic $Ga_{256}As_{255}N$ system. The energy zero coincides with the top of the valence band. The vertical lines are guide to an eye and indicates the position of the high-symmetry points.

3. Principal component analysis

Let $\mathbf{X}$ be matrix which rows are the feature vectors, *i.e.*, the dimension of the matrix is $n \times p$ where n is the number of samples and p is the length of the feature vector. The sample covariance matrix $\mathbf{C}$ is then obtained as $\mathbf{C} = \mathbf{X}^T\mathbf{X}/(n-1)$ which can be diagonalized as

$$\mathbf{C} = \mathbf{V}\mathbf{L}\mathbf{V}^T, \tag{13}$$

where $\mathbf{V}$ are the eigenvectors of $\mathbf{C}$ and are called the principal axes. The matrix $\mathbf{X}$ can be decomposed using singular-value decomposition as $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, where $\mathbf{S}$ is a diagonal matrix containing the singular values of $\mathbf{X}$. Now the matrix $\mathbf{C}$ can be written as

$$\mathbf{C} = \mathbf{V}\frac{\mathbf{S}^2}{n-1}\mathbf{V}^T. \tag{14}$$

Comparing equation (14) to (13) shows that the principal axes are the same as the right singular vectors of $\mathbf{X}$. Now the principal components of $\mathbf{X}$ are obtained as $\mathbf{X}\mathbf{V} = \mathbf{U}\mathbf{S}$. Now we can include only feature vectors that have impact by exluding components that corresponds to singular values below some threshold. In this work we included singular values of which sum contained 90% of the total sum of the singular values and exluded rest.

III. RESULTS

SC-DFT calculations model alloys through band structures corresponding to the superlattice periodicity. Although this is an artificial approach for a random alloy it gives valuable first-principles information about the N-host and N-N interactions. Figure 2 shows the band structure of the periodic $Ga_{256}As_{255}N$ system in the small

first Brillouin zone corresponding to the large unit cell of 512 atoms. The lowest conduction band, *i.e.*, the CBE is flat outside the central region and bends down when approaching the Γ point. The flat region is due to the localized nitrogen-induced resonant states that hybridizes especially near the Γ point with GaAs bulk states. Anticipating our discussion below the strong dispersion near the Γ point can also be interpreted to originate from long-range nitrogen-induced resonant states interacting with each other.

The nitrogen-induced states result in characteristic features in the LDOS's. Fig. **??**a shows the LDOS's for $Ga_{256}As_{255}N$, $Ga_{108}As_{106}N$ and $Ga_{32}As_{31}N$ systems. LDOS's have a narrow peak and a low intensity tail toward lower energies corresponding to the flat and dispersive regions of the CBE, see Fig. 1, respectively. We find these features also in the case of $GaP_{1-x}N_x$ as is evident in Fig. **??**b for $Ga_{256}P_{255}N$, $Ga_{108}P_{107}N$ and $Ga_{32}P_{31}N$ systems. In comparison with GaAs host the LDOS peak related to nitrogen-induced states in GaP is at the same N concentrations lower in energy with respect to the CBM and the extent of the LDOS tail is clearly smaller. When the size of the SC is reduced from 512 to 216 and further to 64 atoms the peak corresponding to nitrogen-induced states broadens in both alloys but stays stationary with respect to the top of the valence band and the LDOS tail reaches deeper in to the band gap causing its reduction. This behavior is in a good qualitative agreement with the behavior of the A line in photoluminescence spectra for $GaP_{1-x}N_x$ alloys with increasing N concentration.[3]

The reduction of the band gap and the length of the LDOS tail below the peak corresponding to nitrogen-induced states for the GaAs and GaP SC's with a single N atom and of different sizes are given in Table I. For GaP the band gap reduction is in practice the same as the length of the LDOS tail reflecting the position of the nitrogen-induced states just at the bulk CBM at the Γ-point. In the case of GaAs a bias of about 0.5 - 0.6 eV should be subtracted from the length of the LDOS tail to obtain the band gap reduction. The bias is due to the location of the peak corresponding to nitrogen-induced resonant states within the bulk conduction band. The lengthening of the LDOS tail causing the band gap reduction is a common feature for the $GaAs_{1-x}N_x$ and $GaP_{1-x}N_x$ systems calling for a uniform picture for these materials.

Figure 3 shows the CBE partial charge density in GaAs around a N atom as a density isosurface. In GaP we find a similar behavior. The partial charge density is obtained by summing the densities from all the k-points used in the calculation. The SC used contains 512 atoms and the isosurface is viewed along a ⟨111⟩ direction so that a Ga atom is on top of the N atom in the center of the figure. The density agglomerates strongly around the N atom and stretches out toward the four nearest-neighbor

TABLE I: Band gap reductions and LDOS tail lengths (See Fig. 2) for different periodic $GaAs_{1-x}N_x$ and $GaP_{1-x}N_x$ systems. The band gap reduction is determined at the Γ point of the superlattice.

System SC	ΔE_g (eV)	Tail length (eV)
$Ga_{256}As_{255}N$	0.02	0.53
$Ga_{108}As_{107}N$	0.09	0.67
$Ga_{32}As_{31}N$	0.34	0.99
$Ga_{256}P_{255}N$	0.13	0.13
$Ga_{108}P_{107}N$	0.22	0.21
$Ga_{32}P_{31}N$	0.57	0.56

Ga atoms and the N-Ga back bonds. Thereafter, the agglomeration of the CBE partial charge density is directed toward the 12 zigzag directions (six of them are on a {111} plane perpendicular to the direction of view, three of them point upwards and three downwards from that {111} plane). We have checked that the CBE partial density from all the k-points, including those with dispersion and near the Γ point, have this anisotropic character. The interaction between the nitrogen-induced states takes place along the the zigzag chains with the CBE partial charge agglomeration. Thus, when the N concentration increases both the width of the narrow LDOS peak and the extent of the LDOS tail increase simultaneously with the strengthening of the CBE partial charge agglomeration on the connecting zigzag chains.

The tendency of the CBE charge to localize along the zigzag chains is strongly connected with ionic relaxations,

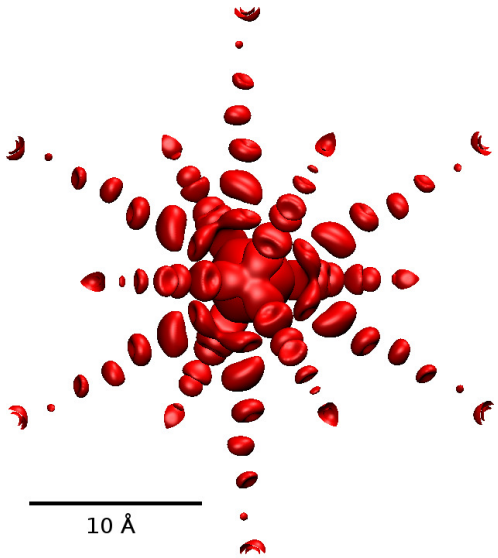


FIG. 3: (Color online). CBE partial charge density in the $Ga_{256}As_{255}N$ SC, viewed along the ⟨111⟩ direction. The isosurface shown corresponds to a density isovalue of 0.0012. The N atom is located at the center of the figure.

TABLE II: Band gap reduction due to a single N atom or two N atoms in the 512- and 216-atom GaAs SC's. Three different configurations are considered for two N atoms.

Configuration	ΔE_g (eV)
$Ga_{256}As_{255}N$	0.02
$Ga_{256}As_{254}N_2$ [(1,1,0)a]	0.09
$Ga_{256}As_{254}N_2$ [(1,1,1)a]	0.06
$Ga_{256}As_{254}N_2$ [(1,0,0)a]	0.04
$Ga_{108}As_{107}N$	0.09
$Ga_{108}As_{106}N_2$ [(1,1,0)a]	0.20
$Ga_{108}As_{106}N_2$ [(1,1,1)a]	0.17
$Ga_{108}As_{106}N_2$ [(1,0,0)a]	0.12
$Ga_{32}As_{31}N$	0.34

initiated by the strong and short Ga-N bonds also propagating along these chains. According to our SC-DFT calculations the Ga-N bond attains in GaAs and GaP nearly the same value as in bulk GaN resulting in an inward relaxation of the nearest neighbor Ga atoms by 13-16 % of the bond length of the GaAs and GaP lattices. A similar strong ionic relaxation along the zigzag directions has been previously observed in the case of the vacancy in silicon[4] and a similar anisotropy of localized N derived states in GaAs and GaP was predicted by Kent and Zunger.[5] We can see the agglomeration of the CBE partial charge density also when we omit the ionic relaxation from ideal lattice positions but then the effect is clearly weaker and shorter in range. This is in agreement with Kent and Zunger[5] who predicted a stronger band gap reduction when the ionic structures were relaxed in SC calculations.

We have studied the anisotropy and strength of the N-N interaction also by inserting two N atoms at different positions in GaAs SC's of 216 and 512 atoms. The studied representative configurations contain one N atom in the origin and another one in the (1,1,1)a, (1,1,0)a, or (1,0,0)a location, where a is the lattice parameter of the conventional unit cell. The results are given in Table II in terms of the reduction of the band gap with respect to the GaAs bulk band gap. The behavior of the band gap is qualitatively similar for the two SC sizes signalizing from the interaction between the N atoms inside the same SC. The strongest reduction in the band gap is observed for the (1,1,0)a configuration in accordance with the above discussion about the anisotropy of the nitrogen-induced resonant states. In the (1,0,0)a configuration the N atoms are not along the same zigzag chain and consistently the band gap reduction is modest compared to the case of a single N atom in the SC. In the (1,1,1)a configuration the band gap reduction is intermediate between those of the (1,1,0)a and (1,0,0)a configurations. This is due to the fact that the strong back bond at a Ga atom caused by one of the N atoms is next to the other N atom.

In order to study the effect of the interactions between the nitrogen-induced states on the band gap reduction in random structures in accordance with the experimental conditions we have developed on the basis of our *ab initio* results a TB model, in which only interactions between N atom sites connected through zigzag chains are included. In our model the non-diagonal matrix elements $h_{i,j}$, describing the interaction between the N atom sites i and j, are defined as $h_{i,j} = k/r_{i,j}^{\alpha}$ if the sites are connected through a zigzag chain and $h_{i,j} = 0$ otherwise. Here $r_{i,j}$ is the distance between the two N atoms. The power-law decay reflects the long-range tendency of the directional interaction. The diagonal terms $h_{i,i}$ are set to a constant value describing the energy level of the nitrogen-induced states. To determine the parameters k and α we used our LDOS results calculated within the SC-DFT scheme for four large structures of a single N atom in the SC's of 64-, 216-, 512-atoms (Fig. **??**) and of two N atoms at the nearest anion sites in the 216-atom SC (Fig. 4). The parameter values of $k = -0.67$ eVÅ$^{\alpha}$, $\alpha = 1.28$ reproduce the DFT LDOS peak and tail structures (See Fig. **??**) of N in GaAs and $k = -0.59$ eVÅ$^{\alpha}$, $\alpha = 1.43$ those of N in GaP. The implementation of the developed method is explained in more detail in Appendix.

Using our *ab initio* -based TB model we study changes in the CBE as a function of the N concentration. We randomly distribute from 384 up to 13,824 N atoms into a SC of 442,368 anion sites. The resulting Hamiltonian matrix is diagonalized and a broadening eigenvalue distribution around the original energy level of nitrogen-induced states is obtained. Between the extreme values

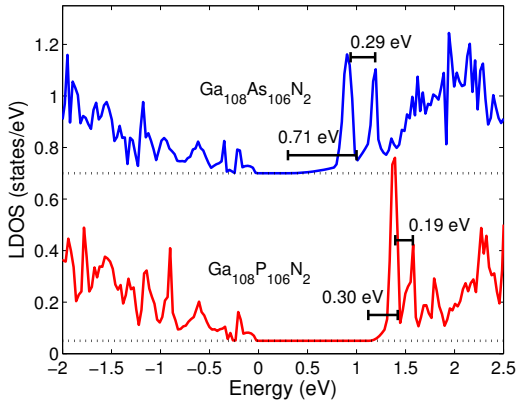


FIG. 4: (Color online) First-principles LDOS's corresponding to two N atoms at neighboring anion sites in $Ga_{108}As_{106}N_2$ and $Ga_{108}P_{106}N_2$ systems. The solid segments indicate the distance between the two peaks related to nitrogen-induced states and the distance between the center of mass of the two peaks and the minimum eigenvalue. The energy zero coincides with the top of the valence band.

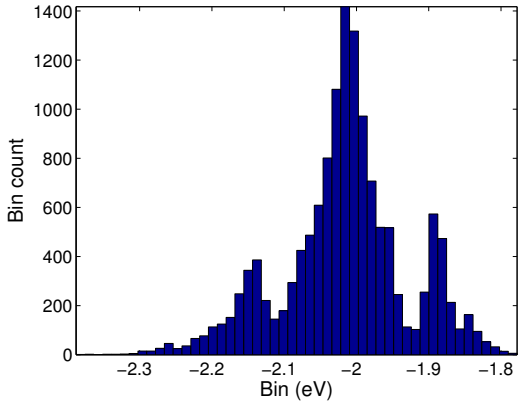
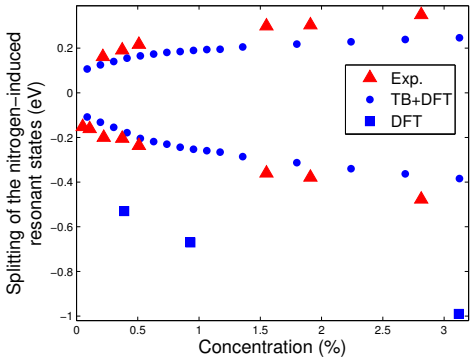


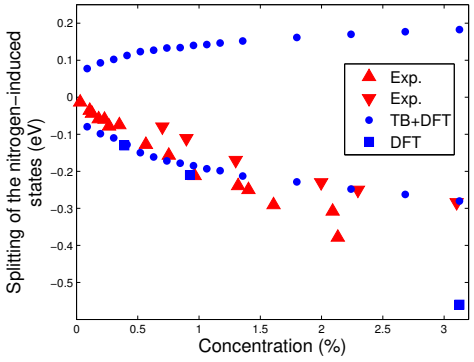
FIG. 5: (Color online) TB eigenvalue distribution corresponding to a random sample of 13,824 N atoms (3.1% concentration) in the $GaAs_{1-x}N_x$ alloy.

there exists a distribution of eigenvalues corresponding to the continuous broadening of the nitrogen-induced states seen in the LDOS's. Fig. 5 shows as an example the distribution for a large $GaAs_{1-x}N_x$ sample with a N concentration of 3.1%. The highest peak is due to the isolated N atoms and the two lower ones are related to N-N pairs. These features are in agreement with the measured scanning tunneling spectra .[6]

We plot, respectively, in Figs. 6a and 6b for $GaAs_{1-x}N_x$ and $GaP_{1-x}N_x$ for each N concentration the minimum and maximum values of the eigenvalue distribution. The values are averaged over 100 different random samples. The extreme values show a square root like behavior, which is, in the case $GaAs_{1-x}N_x$, in a good agreement with photomodulated reflectance measurements for the N-induced band gap reduction and a N-induced feature in the conduction band.[7] Similarly, for $GaP_{1-x}N_x$ our results are in good agreement with the band gap reduction measured by photoluminescence[3] and photomodulated transmission spectroscopy.[8] Our present SC-DFT results shown in Fig. 4 as well earlier SC-DFT calculations[5,9] give a linear dependence of the band gap reduction on the N concentration which is in a clear disagreement with experimental findings. The reason for the linear dependence is the fact that N atoms of the neighboring SC's are on the same zigzag chains resulting in a surplus of N-N interactions with relatively short distances in comparison with the random N atom distribution of the same concentration.



(a)



(b)

FIG. 6: (Color online) Broadening of the distribution of nitrogen-induced states as a function of N concentration in (a) $GaAs_{1-x}N_x$ and (b) $GaP_{1-x}N_x$. Blue circles and squares give our random-system TB and SC-DFT results, respectively. For $GaAs_{1-x}N_x$ red triangles are experimental data from Ref. 7 (measurement temperature 300 K), whereas for $GaP_{1-x}N_x$ red upright and downright triangles are experimental data from Refs. 3 (20 K) and 8 (room temperature), respectively. The energy zero is the energy level corresponding to isolated nitrogen-induced states (In calculations the position of the peak corresponding to nitrogen-induced states in LDOS). To align the experimental and calculated energy levels corresponding to nitrogen-induced states in $GaAs_{1-x}N_x$ the experimental data is shifted to locate symmetrically with respect to the energy zero.

IV. CONCLUSIONS

In this work we demonstrated using *ab initio* calculations how the nitrogen-induced states near the CBM propagate along zigzag chains in $GaAs_{1-x}N_x$ and $GaP_{1-x}N_x$ alloys. This results in coupling between states originating from different N atoms which becomes stronger with increasing N concentration leading to the broadening of the distribution of nitrogen-induced states. On the basis of our DFT results we constructed a TB model for the interaction of the nitrogen-induced states and applied it in large random systems of N atoms in GaAs and GaP. The model predicts a square-root-like broadening of the distributions of nitrogen-induced states as a function of the N concentration and a corresponding narrowing of the band gap in agreement with experi-

ments. The square-root-like behavior is due to the interplay between the directional and long-range characters of the interactions between the nitrogen-induced states. Thus, the band gap narrowing in dilute III-V nitrides can be qualitatively and quantitatively explained by *ab initio* calculations, and it is an inherent property of the interactions between nitrogen-induced states mediated by the host lattice, rather than nitrogen host material CBE interaction.

V. ACKNOWLEDGMENTS

We acknowledge the Suomen Kulttuurirahasto foundation for financial support. This work has been supported by the Academy of Finland through the Center of Excellence program. The computer time was provided by CSC – the Finnish IT Center for Science.

Appendix: Developed TB model

We have developed a TB model describing the interaction between nitrogen-induced states originating from N atoms substituting anions in a III-V compound semiconductors. The non-diagonal matrix elements $h_{i,j}$ corresponding to the N atoms i and j, are defined as $h_{i,j} = k/r_{i,j}^{\alpha}$ if atoms i and j, separated by the distance $r_{i,j}$, are connected trough a zigzag chain and $h_{i,j} = 0$ otherwise. The diagonal terms are set to an arbitrary chosen constant value E_{s^*} describing the energy level of the isolated nitrogen-induced states. The units are electron volts and angstroms. Using the supercell approximation with periodic boundary conditions and the Γ-point approximation the non-zero matrix elements $h_{i,j}$ become

$$h_{i,j} = \sum_{\mathbf{L}} \frac{k}{|\hat{\mathbf{r}}_{i,j} + \mathbf{L}|^{\alpha}}, \qquad \text{(A.1)}$$

where $\mathbf{L}$ is a vector of the superlattice. The restriction on the interactions to the zigzag chains and the use of simple-cubic supercells modifies Eq. A.1 to the form,

$$h_{i,j} = \sum_{\phi} \sum_{n=0}^{\infty} \frac{k}{\left(r_{i,j_\phi} + \sqrt{2}nL\right)^{\alpha}}, \qquad \text{(A.2)}$$

where ϕ runs over all directions where the N atoms i and j are connected through a zigzag chain and L is the side length of the cubic supercell. The diagonal terms become

$$h_{i,i} = E_{s^*} + 12 \sum_{n=1}^{\infty} \frac{k}{\left(\sqrt{2}nL\right)^{\alpha}}. \qquad \text{(A.3)}$$

The inner sum in Eq. A.2 is the Hurwitz zeta function and it can be evaluated efficiently using the Euler-Maclaurin summation formula[10].

To determine the free parameters k and α, we created large ordered structures corresponding to single N atom in 64-, 216-, 512-atoms supercells and two N atoms at the neighboring anion sites in the 216-atom supercell and fitted the parameters k and α so that the TB eigenvalue distribution reproduces the characteristic features in our first-principles LDOS results shown in Figs. **??**a-**??**b and 4. In the case of a single N atom the fitted feature is the tail length and in the case of two N atoms the fitted features are both the distance between the peaks corresponding to the bonding and antibonding wave functions and the distance between their center of mass and the conduction band minimum. The optimal k and α are found by searching for each value of α the optimal k value in the least squares sense. The obtained parameters are $k = -0.67$ eVÅ$^{\alpha}$, $\alpha = 1.28$ ($GaAs_{1-x}N_x$) and $k = -0.59$ eVÅ$^{\alpha}$, $\alpha = 1.43$ ($GaP_{1-x}N_x$).

In the case of GaAs we checked the possibility that the N-N interactions are not restricted to the zigzag chains and modelled them using the short-range exponential decay. However, this model does not reproduce the DFT results in Figs. **??**a-**??**b and 4 and the error in the fit becomes nearly four times larger than in the case where the interaction are restricted on the zig-zag chains.

To simulate real structures, we distributed randomly from 384 up to 13,824 N atoms into a supercell of 442,368 anion sites. The resulting Hamiltonian matrix is diagonalized and a broadening eigenvalue distribution around the original energy level of nitrogen-induced states is obtained. Fig. 5 shows the obtained distribution in the case of 13,824 N atoms. The side peaks around the nitrogen-induced peak results from the N pairs located at the neighboring anion sites.

[1] A. Klapuri, *Introduction to Music Transcription* (Springer, New York, 2006).
[2] K. B. Petersen and P. M. S., *The Matrix Cookbook* (2012).
[3] H. Yaguchi, S. Miyoshi, G. Biwa, M. Kibune, K. Onabe, Y. Shiraki, and R. Ito, Journal of Crystal Growth **170**, 353 (1997).
[4] C. Z. Wang, C. T. Chan, and K. M. Ho, Physical Review Letters **66**, 189 (1991).
[5] P. R. C. Kent and A. Zunger, Physical Review B **64**, 115208 (2001).
[6] L. Ivanova, H. Eisele, M. P. Vaughan, Ph. Ebert, A. Lenz, R. Timm, O. Schumann, L. Geelhaar, M. Dähne, S. Fahy,

H. Riechert, and E. P. O'Reilly, Physical Review B **82**, 161201 (2010).
[7] P. J. Klar, H. Grüning, W. Heimbrodt, J. Koch, F. Höhnsdorf, W. Stolz, P. M. A. Vicente, and J. Camassel, Applied Physics Letters **76**, 3439 (2000).
[8] W. Shan, W. Walukiewicz, K. M. Yu, J. Wu, J. W. Ager III, E. E. Haller, H. P. Xin, and C. W. Tu, Applied Physics Letters **76**, 3251 (2000).
[9] V. Virkkala, V. Havu, F. Tuomisto, and M. J. Puska, Physical Review B **85**, 085134 (2012).
[10] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: the art of scientific computing*, 3rd ed. (Cambridge University Press, New York, 2007).