Some Mathematical Reasoning on the Artificial Force Induced Reaction Method

Wolfgang Quapp*, Josep Maria Bofill†

November 7, 2019

Abstract

There are works of the Maeda-Morokuma group which propose the artificial force induced reaction (AFIR) method [J. Computat. Chem 2014, 35, 166, and 2018, 39, 233]. We study this important method from a theoretical point of view. The understanding of the proposers does not use the barrier breakdown point of the AFIR parameter which usually is half of the reaction path between the minimum and the transition state which is searched for. Based on a comparison with the theory of Newton trajectories, we could better understand the method. It allows us to follow along some reaction pathways from minimum to saddle point, or vice versa. We discuss some well known two-dimensional test surfaces where we calculate full AFIR pathways. If one has special AFIR curves at hand, one can also study the behavior of the ansatz.

Keywords: AFIR, Effective potential energy surface, Saddle point, Barrier breakdown point, Mechanochemistry, Newton trajectory.

*Leipzig University, Mathematisches Institut, Universität Leipzig, PF 100920, D-04009 Leipzig, Germany, quapp@uni-leipzig.de
†Universitat de Barcelona, Departament de Química Inorgànica i Orgànica, Universitat de Barcelona, and Institut de Química Teòrica i Computacional, Universitat de Barcelona, (IQTCUB), Martí i Franquès, 1, 08028 Barcelona, Spain, jmbofill@ub.edu
AFIR points on a 2D surface are blue. In black color are drawn minimums and saddles of index one. The AFIR points explore quite good the reaction pathways from a reactant minimum over an intermediate to a product minimum. Only near the intermediate minimum, the connection has a gap. Red points depict maximums of the AFIR-parameter.

INTRODUCTION

Considerable interest is attached to the search of reaction pathways in chemistry, especially the points which govern these ways: minimums and saddle points of index one (SP1) on the potential energy surface (PES) of a reaction system. The reaction pathway is defined as a one-dimensional description of a chemical reaction through a sequence of molecular geometries in an M-dimensional configuration space.

The AFIR method is an ansatz which disturbs the given PES by an external force.\(^1\,^2\) It is a generalized case of the treatment in mechanochemistry.\(^3\,^5\) It has some similarity with the SEGO method (standard and enforced geometry optimization).\(^6\) By the disturbance one moves the stationary points of the former PES to new locations. By following the successive force-displaced stationary points one gets a curve which can, in good cases, connect a minimum and a SP\(_1\) by a kind of reaction path. The AFIR path has analogous properties.

This paper has the following Sections: next we refer to the AFIR method, and we calculate a reaction path by pieces of a curve by consecutive AFIR points. A more theoretical tool is obtained by a variational formula for full AFIR curves. Further special properties like dependence of the AFIR curve on the coordinates, and avoided crossing (AC), are discussed.
separately by examples. At the end we add a Discussion and some Conclusions.

THE AFIR METHOD

The proposal of the Maeda-Morokuma group is to use an effective PES $^{1,2,7-10}$ with internal coordinates $\mathbf{r}$

$$V_{\text{eff}}(\mathbf{r}) = V(\mathbf{r}) + \alpha \sum_{i<j} \left( \frac{R_i + R_j}{r_{ij}} \right)^p r_{ij} \sum_{k<l} \left( \frac{R_k + R_l}{r_{kl}} \right)^p .$$

Here $V(\mathbf{r})$ is the original PES, $\alpha$ is a factor which plays the role of a numerical parameter which drives the calculation, $R_i$ and $R_j$ are covalent radii of atoms $i$ and $j$. The vector $\mathbf{r}$ with the components $r_{ij}$ contains the distances between the corresponding atoms $i$ and $j$ of the studied chemical system. The dimension of all $r_{ij}$ is maximally $M = \frac{N(N-1)}{2}$ for a molecule with $N$ atoms. It is possible to include a lower number of distances only.$^{1,11}$ Of course, all $r_{ij} > 0$. To imagine the external force, $\mathbf{f}$, directly, we write the components with numbers $i, j$

$$f_{ij} = \frac{\left( \frac{R_i + R_j}{r_{ij}} \right)^p}{\sum_{k<l} \left( \frac{R_k + R_l}{r_{kl}} \right)^p} ,$$

and the effective PES is

$$V_{\text{eff}}(\mathbf{r}) = V(\mathbf{r}) + \alpha \mathbf{f}(\mathbf{r})^T \cdot \mathbf{r} ,$$

thus the force, $\mathbf{f}$, acts on the current point, $\mathbf{r}$. If the force is zero, $\alpha=0$, then we have the original PES. Interesting are the cases with increasing amount of $\alpha$.

Eq. (2) controls the relation between different bonds. It means that the larger distances nearly disappear in the extra force for $p > 2$ but only the smallest distances make a contribution to the resulting direction of the force. So, eventually, the small distances of H-atom bonds, which do not react in the system of question, should not be used in $\mathbf{f}$.\textsuperscript{7}

Of course, if the extra force moves all stationary points of the PES out of their former places then a minimum and an SP can coalesce, and a former barrier can disappear. Such a situation occurs in a point labeled barrier breakdown point (BBP) with $\alpha=\alpha_{\text{max}}$, and it is instructive to compare it for Newton trajectories (NTs).\textsuperscript{12,13} So a new valley opens for
a contact between former distant minimums. Thus, one can use the ansatz to detect reaction valleys.\textsuperscript{1,2,9,11,14–21} To the purpose, one has to chose $\alpha \geq \alpha_{\text{max}}$. The important relation is not discussed in the AFIR papers. If it holds then the former initial minimum disappears and a new minimum exists on the corresponding effective PES which may be near a searched minimum of the original PES. Many examples are drawn recently for NTs.\textsuperscript{4,5,13,22–25}

Here, in contrast, we propose to calculate the full 'reaction path' between the initial minimum and the next SP of index 1 by AFIR. The aim will allow us to better understand the behavior of the AFIR method and its possible improvement. One starts at a known minimum with $\alpha = 0$. Like for NTs,\textsuperscript{4,5,26} a continuous increase of the strength of the force, $\alpha$, will move the stationary points of the effective new PES. In the first papers to AFIR, Meada et al. only searched for an increase of the force.\textsuperscript{1,10,11,14} In the last great review\textsuperscript{2} they propose to use only one fixed value of $\alpha$. We first remark that in the case the value should be larger than $\alpha_{\text{max}}$ because then the optimization on the effective PES does not go back near to the original minimum.

Here we propose to improve the method by two alternating pieces of the curve of new stationary AFIR points. We propose to use an increase of the parameter, $\alpha$, up to the BBP at $\alpha_{\text{max}}$ of the AFIR curve, and a decrease of the parameter, $\alpha$, after the BBP. Then the obtained curve points could fully describe the curve between two original stationary points over a BBP, like in the case of NTs.\textsuperscript{4} The maximal $\alpha$ determines the BBP. Note that the BBP is not an approximation of the original SP of the PES. The BBP is usually anywhere between the initial minimum and the next SP, see some instructive discussions for NTs.\textsuperscript{4} At the next stationary point the parameter $\alpha$ has again to converge to zero.

Since in the AFIR method only one test-$\alpha$ is used\textsuperscript{2} this has to be greater than $\alpha_{\text{max}}$. Because then the optimization can jump along a new valley to a minimum near to the searched one. If it is test-$\alpha < \alpha_{\text{max}}$ then the optimization of the effective PES will get a minimum before the BBP, near to the original initial minimum.

Figure 1 shows the result of calculations for changing values of $\alpha$ for the Rhee-Pande test surface.\textsuperscript{4,22,27} We use $(x, y)$ for only two abstract coordinates, thus dimension $M=2$, and
the exponent $p=6$ in Eq. (1) and put formally $R_1 = 1/2$ and $R_2 = 1/2$. Thus the equation becomes

$$V_{\text{eff}}(x, y) = V(x, y) + \alpha \frac{x^6 + y^6}{1 - \frac{1}{x^6} - \frac{1}{y^6}}.$$ 

The calculation goes step by step. We put a value of the parameter $\alpha$ near zero, and search the next stationary point of the effective PES, $V_{\text{eff}}$. This point is used for the next $\alpha$-step as the initial value for the optimization. We proceed in the same way after the convergence is reached.

![Figure 1: Test surface of Rhee and Pande with AFIR points (blue). The minimums and saddle points of index one are indicated in black. The BBPs are shown in pink color. The level lines start from zero level at $\mathbf{R}$. They increase in 1-steps up to a value of 15, and after that in 5-steps up to a value of 60.]

The AFIR points start at the right minimum, $\mathbf{R}$ at (4.03, 0.97) at zero energy, and they go with negative values of $\alpha$ up to $\alpha_{\text{max}} = -9.98$ at the BBP (pink) at point (4.7, 2.35). Then the parameter is again decreased (in absolute values). With the decreasing amounts of the parameter after the BBP, the curve at least correctly finds the rightmost SP at (4.25,
2.96). The same can continued in direction of the intermediate minimum, however, at the corresponding next BBP, the curve ends. There is a small gap to the intermediate where no AFIR points exist. An explanation of the fact comes below in the next Section.

The other half of the reaction path from the product minimum, $P$ at point (1.0, 4.0) and energy 3.64, is analogous. Up to a values of $\alpha = -6.5$ at the leftmost BBP (pink), at (2.8, 5.15), we increase the parameter, but the way to the SP before the intermediate, at (3.0, 4.56), is again get by a decrease. And the piece between the leftmost SP of index 1 and the intermediate minimum at (3.76, 4.03) at energy 5.96 is a next ordinary part of an AFIR curve.

The procedure to go up by small steps for $\alpha$ and optimize the corresponding stationary point, up to a BBP, and then go back to zero for $\alpha$ to find an SP$_1$, this procedure will work in every dimension $M$.

**AFIR curves by a variational formula**

We use a first variational structure$^{28}$ of the AFIR model

$$g(r) - \alpha \varphi(r) = 0$$

where $g(r)$ is the gradient of the PES, $\alpha$ is the Lagrange multiplier and $\varphi(r)$ is the derivation of the extra term, $\nabla_r(f(r)^T \cdot r)$. If we assume that $\varphi(r) \neq 0$ we can write the variation ansatz in another form

$$\left( U - \frac{\varphi(r)\varphi(r)^T}{\varphi(r)^T \varphi(r)} \right)g(r) = 0$$

where $U$ is the unit matrix. Here the task would be to derive the implicit tangent from the given term. However, the problem in the model ansatz Eq. (1), is the quite complicated expression of $\varphi(r)$. To make an attempt to achieve an expression, we execute the following derivation. First we rewrite Eq. (1) in the form $g^T(r) - \alpha(r)\varphi^T(r) = 0^T$, where $\alpha(r) = \varphi^T(r)g^T(r)[\varphi^T(r)\varphi(r)]^{-1}$. Now we define $H(r) := \nabla_r g^T(r)$ and $G(r) := \nabla_r \varphi^T(r)$. After some mathematical manipulations we obtain

$$\left[ H(r) - \alpha(r)G(r) - \varphi(r)\nabla^T_r \alpha(r) \right] \frac{dr}{dt} = 0$$
where $\nabla_r \alpha(r)$ has the form

$$\nabla^T_r \alpha(r) = \frac{g^T(r)}{\varphi^T(r)\varphi(r)} \left[ U - 2 \frac{\varphi(r)\varphi^T(r)}{\varphi^T(r)\varphi(r)} \right] G(r) + \frac{\varphi^T(r)}{\varphi^T(r)\varphi(r)} H(r).$$

(7)

We note that if $\varphi(r)$ is independent of $r$ then Eq.(6) reduces to the tangent of the reduced gradient following (RGF) model$^{26,29}$ which is another version of the NT model. From this point of view we can say that the AFIR method is a generalization of the NT-model.

Figure 2: Test surface of Rhee and Pande$^{27}$ with AFIR curves (blue). The minimums and saddle points of index one are shown in black, the BBPs in red.

In the case of $M = 2$, we can use Eq.(5) for a numeric search of a solution of the AFIR curves. We employ a Mathematica contour plot in Figure 2 for the zero contour of the square of the norm of the left hand side of Eq. (5). The point by point optimized AFIR points of Figure 1 fit well in the resulting curves. The gap between the rightmost SP of index 1 and the intermediate is an avoided crossing of two AFIR curves.

A next problematic property is, at the other side of the minimums, for positive $\alpha$ values, that the AFIR curves escape into the mountains. They do not converge to the upper SP$_1$ at (1.59, 1.45). This SP is very higher in energy, it is 24 units, than the pathways through the
intermediate where the left SP is at 12.84 energy units. Thus a reaction will proceed over the lower reaction path, and not over the upper one, however, from a theoretical point of view, one would like to know also the higher energy pathway. But the AFIR points from the two global minimums do not converge to this SP. An AFIR curve can start in this SP, but it connects the SP to the summit of the surface, an SP of index two. In contrast, steepest descent from the SP at (1.59, 1.45) will find the two global minimums.

Note that here the origin (0, 0) of the plane of the coordinates plays an exceptional role which is an artifact of this 2D test surface. (The origin was also excluded for the application of Eq. (5).) It seems that not all AFIR curves through the zero point have a geometric sense. In the AFIR ansatz of Eq. (1) are used only atomic distances of the chemical system which cannot be zero. Thus in real chemical calculations the zero of the distance coordinates does not appear.

Coordinate dependence of AFIR curves

Figure 3: (a) KAM test surface\textsuperscript{12} with AFIR curves (blue). Green lines are the $\text{Det}(H_{eff})=0$ points of the surface. The crossing with an AFIR curve is a BBP of this curve. (b) The same surface but all coordinates moved by (2, 2). By this transformation we avoid the point (0, 0) in Eq.(1). The new set of AFIR curves is different with respect to the previous set of AFIR curves.

Because of the nonlinearity of the AFIR ansatz, Eq. (1), the resulting curves do not only
depend on the PES, however, they also depend on the used coordinates. We demonstrate it with a very simple test surface with one minimum and two SP of index 1, the Konda-Avdoshenko-Makarov (KAM) surface.\textsuperscript{12,13} It is a surface with two reaction pathways between a reactant and the exit. Across to the exit it has a very flat ridge. Unfortunately, the minimum is the zero of the coordinates system. Nevertheless, an AFIR curve in Fig. 3(a) connects the minimum and the lower SP region, but near the lower SP it suffers from a small avoided crossing. Another AFIR curve follows nicely the flat ridge. Near point \((0.9, 0.8)\) this part of an AFIR curve has a turning point (TP). In this example, all AFIR curve directions through the minimum are different from the eigenvalues of the Hessian. It seems that not all AFIR curves through the zero point have a geometric sense.

In Fig. 3(b) we use the same surface but moved symmetrically by a coordinate transformation by a linearly moved origin with a distance \((2, 2)\). The AFIR curves have a quite other form! But they reflect well the ridge of the surface. A former direct pathway from the minimum to the lower SP of index 1 is missing. The AFIR curve now connects indirectly both SP with the minimum over turning points.

The green lines in Fig. 3 (and in the following Figures) are the \(\text{Det}(\mathbf{H}_{\text{eff}}) = 0\) points of the surface where \(\mathbf{H}_{\text{eff}}\) is the Hessian matrix of the second derivatives of the effective PES with respect of the coordinates. There the gradient norm of the effective PES has a maximum or a minimum if one goes along the AFIR curve. The crossing of a green line with an AFIR curve is a BBP of this curve.

The next example has another set of scaled coordinates. Fig. 4(a) shows the AFIR curves for a modified three-minimums surface\textsuperscript{13,30} obtained by the variational formula, Eq. (5). The modified surface is defined by

\[
V(x, y) = \frac{1}{3}(x^3 - 3xy^2) + \frac{1}{40}((x + \frac{7}{4})^4 + y^4) + 250 \exp[-0.15 ((x + 3)^2 + y^2)] .
\]

The three minimums may mean one reactant minimum, Min\(_R\), and two different product minimums, Min\(_{P1}\) and Min\(_{P2}\). The corresponding saddles are also so depicted. The example is chosen because the range of the coordinates is extended by a factor of 10. Nevertheless, here all AFIR curves suffer from avoided crossings (AC). No two stationary points are truly
connected. Especially the Min$_R$ is far away from the origin, but it has a large AC to the SP$_2$. AC means that the AFIR method can fail because the stationary points cannot coalesce on such separated AFIR curves.

Figure 4: (a) Three minimums BQC test surface with blue AFIR curves. Green lines are the BBP curves with Det($H_{eff}$)=0. (b) The same surface moved by the coordinate pair (20,20). By this transformation we avoid the point (0,0) in Eq.(1). Again, AFIR curves have strongly changed.

Fig. 4(b) shows the AFIR curves for the same surface but changed coordinates by the symmetric distance (20,20). Now the picture again changes totally. The movement of the coordinates origin out of the global bowl improves the situation. AFIR curves connect one each minimum with one next SP, but not with the corresponding other SP. The Min$_R$ is connected to SP$_1$, the Min$_{P1}$ is connected to SP$_P$, and the Min$_{P2}$ is connected to SP$_2$. Between SP$_P$ and SP$_2$ emerges an AFIR arc over the maximum. Again some avoided crossings exist. No two minimums are truly connected.

Further Examples

Fig. 5(a) shows the AFIR curves for the Eckhardt test surface$^{31}$ obtained by the variational formula. The example is used because here the 'forbidden' origin of the coordinates is the maximum of the surface reflecting to a certain degree the case of distance coordinates, r of
Figure 5: (a) Eckhardt test surface with blue AFIR curves. Green lines are the BBP curves \(\text{Det}(\mathbf{H}_{\text{eff}})=0\). (b) The same surface where the origin is moved by \((3, 3)\). By this transformation we avoid the point \((0, 0)\) in Eq.(1). The AFIR curves qualitatively change.

Eq.(1), where the case \(r=0\) is a really forbidden singularity. The Eckhardt surface is again a surface with two different reaction pathways between reactant and product minimums like the Rhee-Pande case. Here again all AFIR curves suffer from avoided crossings. No two stationary points are truly connected. But again the upper \(SP_u\) seems to be more isolated than the lower one, \(SP_l\).

Fig. 5(b) shows the AFIR curves for a moved by \((3, 3)\) Eckhardt surface. This situation is not better than under panel (a): the product minimum has no connection to other stationary points by an AFIR curve, the reactant minimum is connected to the lower \(SP_l\) by a strange AFIR curve, but the upper \(SP_u\) again is isolated.

Many AFIR curves show an AC. We could not assign any useful property of the PES to such ACs. It is in contrast to NTs. There the ACs indicate the neighborhood of a valley-ridge inflection (VRI) point which is crossed by a bifurcating, a singular NT. Singular NTs divide the 'regions of influence' of the different stationary points. However here, so to say, 'singular' AFIR curves with a bifurcation are very seldom because these curves do not form a dense family of curves. They are unique curves. One cannot try to change the 'search-direction' of the AFIR curve to get a nearby 'singular' AFIR curve like a singular NT. The bifurcation of
Figure 6: MB test surface\textsuperscript{32} with blue AFIR curves. Green lines are the curves $\text{Det}(\mathbf{H}_{\text{eff}})=0$. The crossing point between a green line and an AFIR curve is the BBP of this curve.

NTs is quite easier to calculate\textsuperscript{33} and depends directly on the Hessian of the PES. Because of the nonlinearity of the ansatz of Eq. (1) the connection to the effective Hessian will be quite more complicated.

Fig. 6 shows the AFIR curves for the well-known Müller-Braun surface\textsuperscript{32} obtained by the variational formula, Eq. (5). Here only the main SP\textsubscript{1} and the intermediate minimum are connected by an AFIR curve. All other AFIR curves suffer from avoided crossings. It is in contrast to the case of NTs.\textsuperscript{23} At every stationary point we detect one AFIR curve which coarsely follows an eigenvector direction of the Hessian. At two minimums they follow the smaller eigenvalue direction, thus the 'reaction valley', but at the third minimum the corresponding AFIR curve follows the larger eigenvalue direction. The rule for this pattern is not clear. At the main SP\textsubscript{1} near point (-0.8, 0.65) the AFIR curve here crosses along the ridge, not along the reaction path direction.

**DISCUSSION**

The examples demonstrate that the AFIR method can follow a valley from a minimum to an SP\textsubscript{1}, or vice versa, at least in good cases.
There are some specialties:

(i) There are often gaps by an avoided crossing of the AFIR curves. The hypothetical bifurcation points inside an avoided crossing seem to have no geometrical meaning. In contrast, regular NTs connect the minimums with the SP$_1$ of the PES. Bifurcation points of NTs are valley-ridge inflection points. Additionally, an AC can destroy the planned action of the AFIR method.

(ii) The AFIR curve can have a turning point. This means that the curve touches a level line. Such behavior is also known from NTs. If a turning point emerges then the corresponding curve should not serve for a model of a reaction path because the TP has usually a higher energy than the next SP$_1$.

(iii) A problematic property of the AFIR method, at least in the example of Fig.2, as bad as in others, is that here an unsatisfactory behavior emerges into the inverse directions of the two global minimums. The corresponding AFIR curves for positive values of $\alpha$ escape into the left and right mountains, however, they do not find the SP$_1$ at point (1.59, 1.45). Thus not every SP$_1$ which is connected to a minimum by a steepest descent is also connected with this minimum by an AFIR curve.

(iv) Usually one AFIR curve leads through the stationary points, correspondingly, to positive or negative values of the parameter, $\alpha$. It is again like for NTs, but there we can choose any direction which then is the leading direction of the NT. The NTs have a quite greater variability because around a stationary point all search directions are possible. The NTs form a dense net of curves on the PES. And the NTs are a linear ansatz, thus very easier to handle than the AFIR method.

(v) A search for optimal BBPs$^{34}$ is not possible with AFIR curves because they have their fixed direction at every point. To determine an optimal direction, the search direction must be continuously changeable to determine the optimal NT.

**CONCLUSIONS**

In former applications the AFIR method is handled as a 'black box'. It is not discussed that the $\alpha_{\text{max}}$ of the BBP plays the decisive role for the planned action. The use of only a
fixed value of the parameter $\alpha$ for a test calculation\textsuperscript{2} where then it is hoped to find a next minimum more or less accidentally, gives away possibilities of this ansatz. In good cases, a consecutive use of small $\alpha$-steps can follow a reaction path up to the searched SP\textsubscript{1} directly. But one has to be careful: $\alpha$ has to increase up to $\alpha_{\text{max}}$ at a ‘barrier breakdown point’ BBP and then to decrease back to zero at the next stationary point.

However, the emergence of ‘avoided crossings’ of AFIR curves can destroy their exploitability for a full reaction pathway.

The dependence on the coordinates of the external force makes the method somewhat tricky!

The last conclusion is that one should better use the simpler Newton trajectories, thus a fixed, constant force, $f$, in Eq. (3). NTs are better adapted to the task of the AFIR method.

**ACKNOWLEDGMENTS**

Financial support from the Spanish Ministerio de Economía y Competitividad, Project CTQ2016-76423-P, Spanish Structures of Excellence María de Maeztu program through grant MDM-2017-0767 and from the Generalitat de Catalunya, Departament d’Empresa i Coneixement, Project 2017 SGR 348 is fully acknowledged.

**References**


