

# DNA-Based Fixed Gain Amplifiers and Linear Classifier Circuits

David Yu Zhang<sup>1</sup> and Georg Seelig<sup>2</sup>

<sup>1</sup> California Institute of Technology, Pasadena, CA, USA

<sup>2</sup> University of Washington, Seattle, WA, USA

dzhang@dna.caltech.edu, gseelig@u.washington.edu

**Abstract.** DNA catalysts have been developed as methods of amplifying single-stranded nucleic acid signals. The maximum turnover (gain) of these systems, however, often varies based on strand and complex purities, and has so far not been well-controlled. Here we introduce methods for controlling the asymptotic turnover of strand displacement-based DNA catalysts and show how these could be used to construct linear classifier systems.

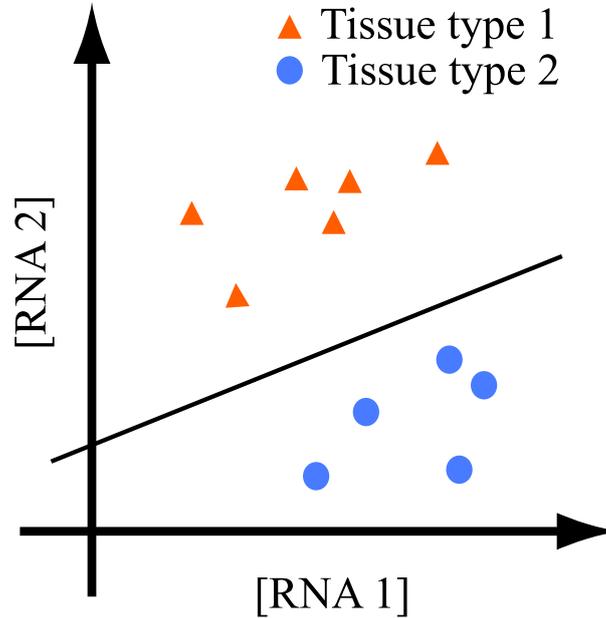
## 1 Introduction

DNA nanotechnology has utilized the specific binding properties [1] and the well-understood thermodynamics [2] and kinetics [3] [4] of nucleic acids to construct dynamic cascaded reactions, such as logic gates and circuits [5] [6] [7] [8], motors [9] [10], and amplification mechanisms [11] [12] [13] [14] [15] [16] [17].

DNA devices can operate in complex biochemical environments and can be programmed to specifically interact with biological nucleic acids such as messenger RNA (mRNA) or microRNA (miRNA). DNA circuits could be used to develop novel point-of-care diagnostic devices that integrate detection with analysis and do not require complex laboratory equipment. It has even been suggested to use DNA devices as “smart therapeutics” that operate inside living cells and integrate detection of specific disease markers with the activation of a therapeutic response based on the RNA interference pathway [18] [19], on antisense oligonucleotides [20] or ribozymes.

Such applications require nucleic acid circuitry that can reliably identify a specific disease state. Characteristic RNA markers that could serve as inputs to a DNA analytic circuit have been identified for many diseases. However, it is often not sufficient to simply detect the presence or absence of a set of RNA markers. Instead the classifiers that distinguish a disease tissue from healthy tissue (or other disease tissues) are often complex functions of the concentrations of multiple RNA markers (see Refs. [21] [22] for examples of microRNA-expression based classifiers of varying complexity).

Here we propose a molecular implementation for a specific class of classifiers, namely linear classifiers. The classifier circuit computes a linear combination



**Fig. 1.** Sketch of a hypothetical two-gene classifier. Samples from two different tissue can be clearly distinguished based on the expression profiles of two RNA molecules.

with arbitrary (positive or negative) weights on a set of inputs (e.g. RNA molecules) and compares the result to a threshold value. Fig. 1 shows a highly simplified sketch of a linear two-gene classifier: the line separating the two different tissue types is given by an equation of the form  $\alpha_1[RNA_1] + \alpha_2[RNA_2] = K$ . Given a sample of unknown origin, we can now classify it as tissue type 1 or 2 based on a measurement of two RNAs. Unlike in the more conventional case where the expression of each RNA is individually measured and the linear classification analysis is performed *in silico*, here both detection and analysis are done on the molecular level, allowing *in situ* and *in vivo* applications.

Previous DNA logic gates and circuits were mostly designed for a situation where inputs can be represented as Boolean variables and are either present at a high concentration or completely absent [8] [5]. This does not necessarily require the original inputs to be at a specific level; DNA-based signal restoration units consisting of a threshold gate and an amplifier can be used to restore an input with an arbitrary concentration to the expected logical TRUE or FALSE values. Still, the digital nature of such circuits is inherently incompatible with classification problems, in which the relative amounts of inputs determines the value of the final output. The fixed gain amplification methods presented here allow reliable tuning of analog signals encoded in the concentrations of nucleic acids.

## 2 Fixed Gain Amplifiers: Lowering Catalytic Turnover

One key component of the proposed linear classifier is a DNA-based catalytic amplifier, that allows one signal-stranded nucleic acid to specifically produce or release many single-stranded nucleic acid molecules of independent sequence. Importantly, this amplifier needs to have a finite and controllable gain  $\alpha$  such that each input on average releases  $\alpha$  copies of the output. Such a finite gain amplifier would be useful not only in a linear classifier, where each detected RNA species is assigned a different weight, but could also be used for pre-amplification of a set of low-concentration inputs while maintaining their relative concentrations.

Existing DNA amplifiers have an intrinsically finite turnover; strand displacement-based nucleic acid catalysts typically convert on the order of 10-100 substrates before being inactivated [14] [15]. Inactivation is most likely due to defective substrate complexes or fuels [14] [15] [17]. The details of the inactivation process depend on the specifics of the amplifier design, but it seems likely that imperfectly synthesized DNA strands are a major culprit. In practice, the maximal turnover obtained seems to depend strongly on sequence, purification procedures, strand orientation and similar experimental and design details. Therefore, while the gain is finite, it can be characterized for any particular system.

The question then becomes if, starting from an arbitrary but high turnover, we can lower the turnover controllably to a fixed value. Given the intrinsic turnover of a catalytic system, intuitively it seems clear that we can lower the turnover further either by increasing the fraction of imperfect substrate or through addition of an alternative competitive inhibitor that irreversibly binds to the catalyst. However, it may be less intuitive how to best adjust the turnover to any specific desired value.

To address the question of how to control turnover we first consider a simple model for a catalytic reaction with competitive inhibition. Afterwards, we simulate a specific DNA implementation using measured reaction parameters. A catalytic reaction in the presence of an impurity can be modeled as:

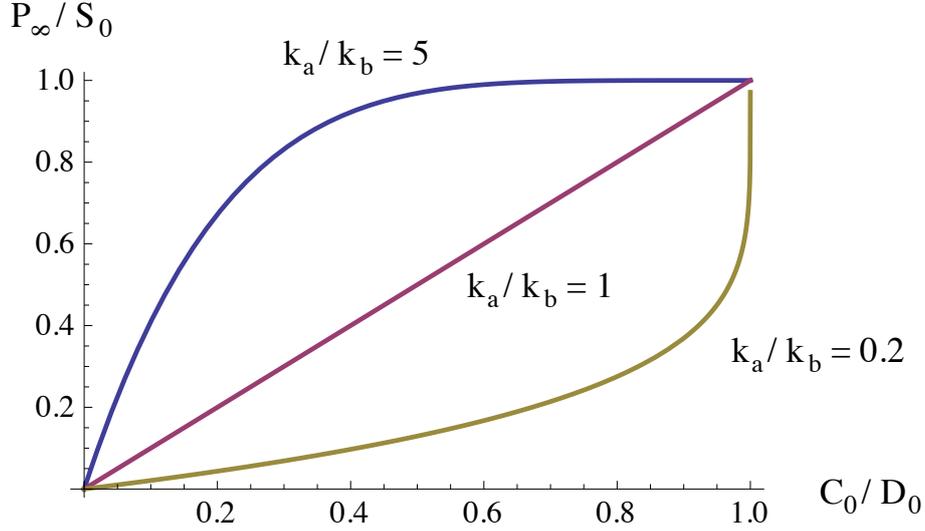


In the first reaction a catalyst  $C$  transforms a substrate  $S$  into a product  $P$ . The rate constant for this reaction is  $k_a$ . The catalyst can also participate in a second, unproductive reaction with an inhibitor (or damper)  $D$ . This reaction proceeds at a rate constant  $k_b$ .

The differential equations resulting from this model can be integrated with initial conditions  $C(0) = C_0$ ,  $S(0) = S_0$ ,  $D(0) = D_0$  and  $P(0) = 0$ . Solving for the product  $P(t)$  we get

$$P(t) = S_0 - S_0 \left( \frac{1 - \rho}{1 - \rho e^{k_b \Delta t}} \right)^{k_a/k_b} \quad (3)$$

where we introduced the ratio  $\rho = C_0/D_0$  and the difference  $\Delta = C_0 - D_0$  of the initial amounts of catalyst and inhibitor.



**Fig. 2.** Output produced in a catalytic reaction with competitive inhibition as a function of the catalyst concentration. Final product  $P_\infty$  is scaled by initial concentration of substrate  $S_0$ . Initial catalyst concentration  $C_0$  is measured in units of inhibitor concentration  $I_0$ . We obtain different I/O characteristics depending on the ratio between the rate constants  $k_a$  and  $k_b$  for the catalytic and the competitive reaction.

In an ideal system without competitive inhibition the final product concentration is always equal to the initial concentration of substrate. Given enough time the catalyst will convert all substrate into product. In a system with competitive inhibition this is not necessarily true. The final amount of product produced in that case can be computed by taking the limit  $t \rightarrow \infty$  in Eq. 3:

$$\lim_{t \rightarrow \infty} P(t) = P_\infty = \begin{cases} S_0, & C_0 \geq D_0 \\ S_0 - S_0 (1 - \rho)^{k_a/k_b}, & C_0 < D_0 \end{cases} \quad (4)$$

Not surprisingly, if we start out with more catalyst than inhibitor, the reaction will eventually go to completion. The opposite limit is more interesting.

First, consider the case where the rate for the catalytic reaction is much faster than the inhibition reaction,  $k_a > k_b$  (blue trace in Fig. 2). In this case the inhibitor has a relatively minor effect that is most pronounced at low concentrations of catalyst compared to the inhibitor.

In the limit where the catalytic reaction occurs at exactly the same rate as the inhibitory reaction, i.e.  $k_a = k_b$  (red trace in Fig. 2) Eq. 4 predicts that the final amount of product is linear in the initial amount of catalyst, i.e.  $P_\infty = \alpha C_0$  where  $\alpha = S_0/D_0$ . That is, by adjusting the relative concentration of substrate to inhibitor we can get any finite gain we need.

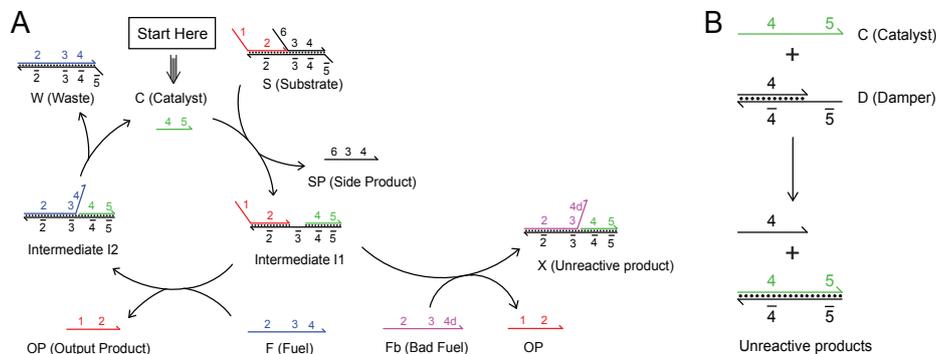
$S + C \xrightleftharpoons[k_2]{k_1} I1 + SP$	$k_0 = 5 \text{ M}^{-1} \text{ s}^{-1}$
$I1 + F \xrightarrow{k_2} I2 + OP$	$k_1 = 2.7 \cdot 10^5 \text{ M}^{-1} \text{ s}^{-1}$
$I2 \xrightleftharpoons[k_1]{k_3} W + C$	$k_2 = 1.1 \cdot 10^6 \text{ M}^{-1} \text{ s}^{-1}$
$S + F \xrightarrow{k_0} OP + SP + W$	$k_3 = 1.1 \cdot 10^{-2} \text{ s}^{-1}$
$I1 + Fb \xrightarrow{k_2} X + OP$	$k_{rep} = 4 \cdot 10^5 \text{ M}^{-1} \text{ s}^{-1}$
$S + Fb \xrightarrow{k_0} OP + SP + W2$	
$C + W2 \xrightarrow{k_1} X$	
$C + D \xrightarrow{k_1} X$	
$A + C \xrightleftharpoons[k_2]{k_1} IA1 + SP$	
$IA1 + F \xrightarrow{k_2} I2 + C$	
$I2 \xrightleftharpoons[k_1]{k_3} W + C$	
$A + F \xrightarrow{k_0} C + SP + W$	
$IA1 + Fb \xrightarrow{k_2} X + C$	
$A + Fb \xrightarrow{k_0} C + SP + W2$	

Table 9-1: Reactions simulated in Fig. 9-4.

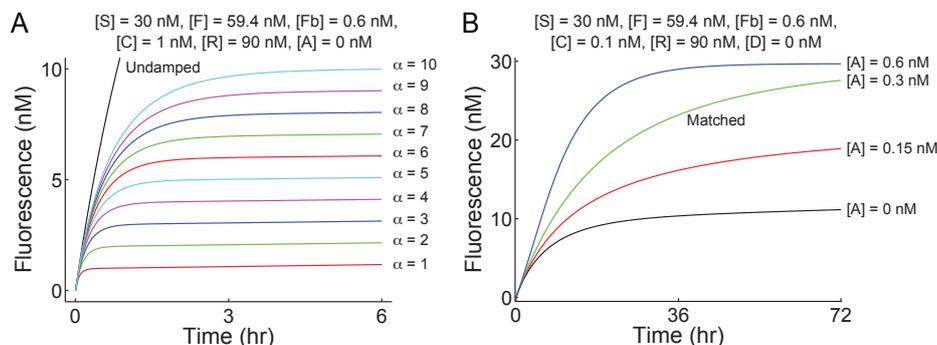
The situation where the rate for the inhibitor reaction is faster than the rate for the catalytic reaction is also interesting. In that case, the amount or product is sub-linear in the initial amount of catalyst for  $C_0 < D_0$  but reaches a fixed value  $S_0$  in the opposite regime. The concentration of the competitive inhibitor  $I$  therefore acts as a threshold for the catalytic reaction. Such a threshold element is useful for reliable signal propagation for example in the context of chemical digital circuits.

We now turn to a specific DNA implementation of such a system. Our implementation is based on the entropy-driven catalytic amplifier of Ref. [15] which was further characterized in Ref. [17]. Turnover for this amplifier was measured to be about 100. The reaction mechanism for this system including the side reactions leading to intrinsically finite turnover is shown in Fig. 3A. A reaction between catalyst strand and substrate relies on toehold mediated strand displacement. As a competitive inhibitor we here propose to use a damper DNA gate that irreversibly binds the catalytic input (Fig. 3B). In order to match the reaction rate constants of the catalyst with this inhibitor to that of the catalyst with the active substrate we simply choose the toeholds for both reactions to be identical.

In order to verify the predictions from our simple model Eq. 1 we simulated the full catalytic system of Ref. [15] with a parallel inhibitory reaction using the measured rate constants and reaction intermediates. The model is given in Table 9-1 and resulting data is shown in Fig 4A. As expected from our model, the final fluorescence depends linearly on the concentration of the damper gate.



**Fig. 3.** Methods for tuning catalytic turnover. **(A)** DNA amplification via catalysis, adapted from Zhang et al. [17]. Catalyst strand *C* reacts with *S* to form side product *SP* and intermediate *I1*, the latter of which subsequently reacts with *F* to release output product *OP*, waste *W*, and catalyst *C*. However, a small fraction of bad fuel with deletions and/or degradation near the 3' end, denoted as *Fb*, will bind to intermediate *I1* to form an unreactive product *X*, thus permanently trapping catalyst *C* and reducing the observed catalytic turnover of the reaction. The ratio  $\frac{[Fb]}{[F]+[Fb]}$  was estimated to be 0.01 for HPLC-purified fuel strands [17]. **(B)** The catalytic turnover of the reaction can be tuned to be lower via the addition of the damping complexes *D*. Because *C* binds by the toehold to *D* as to *S*, it is assumed that this rate constant is identical in value to that of  $k_1$ . **(C)** Schematic of a generalized catalytic reaction with arbitrary control over turnover. In the original work on entropy-driven DNA catalysts [15], it was shown that the catalytic reaction can be made autocatalytic by using an alternative substrate *A*, which releases as product an molecule identical to the catalyst. Turnover can be increased above the limit set by fuel purity with autocatalytic substrate.



**Fig. 4.** Modulating turnover. **(A)** Simulations of the entropy-driven catalyst system with damper. [17]. Various amounts of *D* were present to achieve the fixed turnover  $\eta$  shown, with  $[D] = \frac{30}{\alpha} - 0.3$  nM. See Table 9-1 for the full set of simulated reactions. **(B)** Simulations of the entropy-driven catalyst system with autocatalytic substrate. At  $[A] = 0.3$  nM, the increase of catalyst due to the autocatalytic substrate nearly matches the decrease of the catalyst due to bad fuel. With lower concentrations of *A*, asymptotic turnover is limited. With higher concentrations of *A*, the reaction adopts autocatalytic characteristics, and becomes less sensitive to the initial concentration of the catalyst.

### 3 Fixed Gain Amplifiers: Increasing Catalytic Turnover

The turnover of a catalytic reaction can be increased above the intrinsic limit set by defective oligonucleotides. It seems clear that it should be possible to compensate for the loss of catalyst in an unproductive reaction through the production of an extra catalyst in a parallel autocatalytic reaction that proceeds at the same rate. A simple model motivated by this intuition is



Here  $A$  is the substrate for the autocatalytic reaction which is present initially at a concentration  $A(0) = A_0$ . With the same initial conditions as above we can solve the resulting differential equation. The final product as a function of time then is

$$P(t) = S_0 - S_0 \left( \frac{1 - \sigma}{1 - \sigma e^{k_a \Gamma t}} \right)^{k_a/k_b}, \quad (8)$$

where  $\Gamma = C_0 + A_0 - D_0$  and  $\sigma = C_0/(D_0 - A_0)$ . The result is therefore of exactly the same form as Eq. 3 if we make the substitution  $D_0 \rightarrow D_0 - A_0$ . In the special case where the initial concentrations of the inhibitor  $D$  and the substrate  $A$  for the autocatalytic reactions are the same, i.e.  $A_0 = D_0$ , these reactions cancel each other out and  $P(t) = S_0 \exp(-k_a C_0 t)$  as expected for an ideal catalytic reaction. If  $A_0 > D_0$  the overall kinetics of the reaction is that of an autocatalytic reaction. In fact, for  $k_a = k_b$  Eq. 8 looks very similar to the logistic equation we obtain when solving a simple autocatalytic reaction. The different limiting cases for the amount of product  $P_\infty$  for  $t \rightarrow \infty$  follow from the discussion above if we make the substitution  $D_0 \rightarrow D_0 - A_0$ .

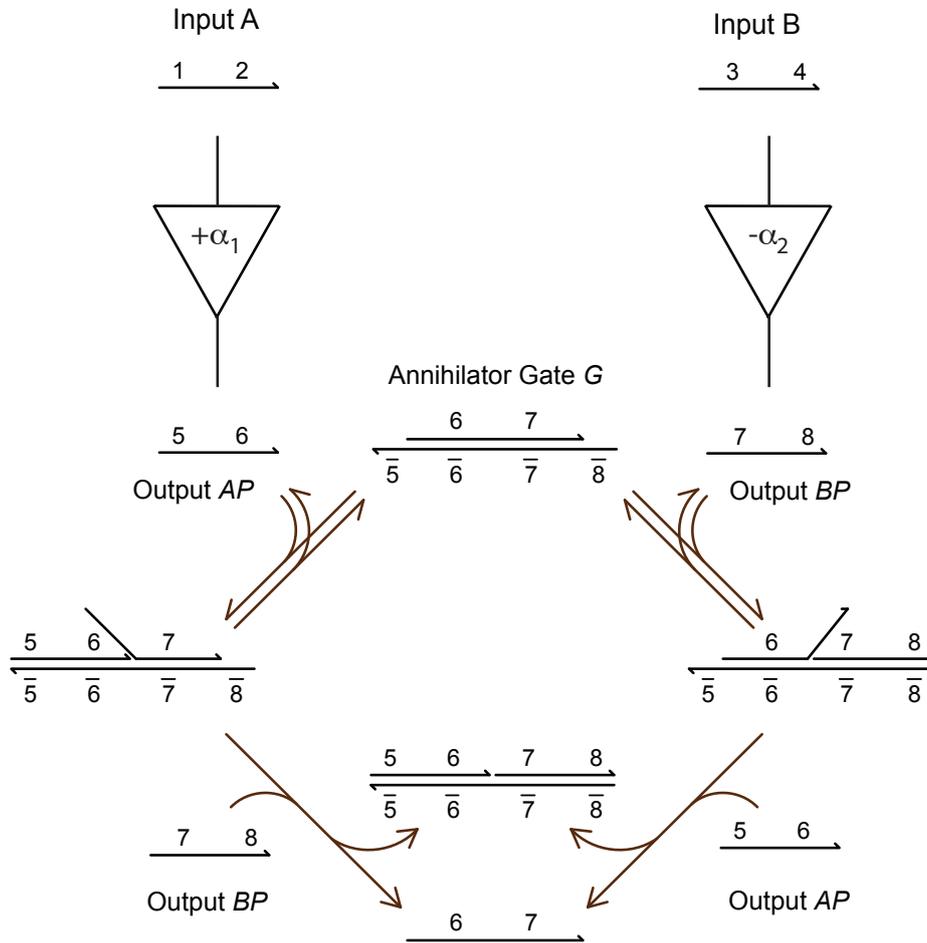
### 4 A Linear Classifier

Based on the fixed gain amplifier systems explained above we can now build a linear classifier that implements a function

$$\sum_i \alpha_i [m_i] = T. \quad (9)$$

Here  $\alpha_i$  are the weights,  $[m_i]$  the concentrations of the molecular species  $m_i$  and  $K$  is the threshold. A molecular implementation of this function thus requires that an initial concentration of  $m_i$  results in a concentration  $\alpha_i [m_i]$  of some signal molecules that can be compared to each other and to the concentration  $K$  of a threshold molecule.

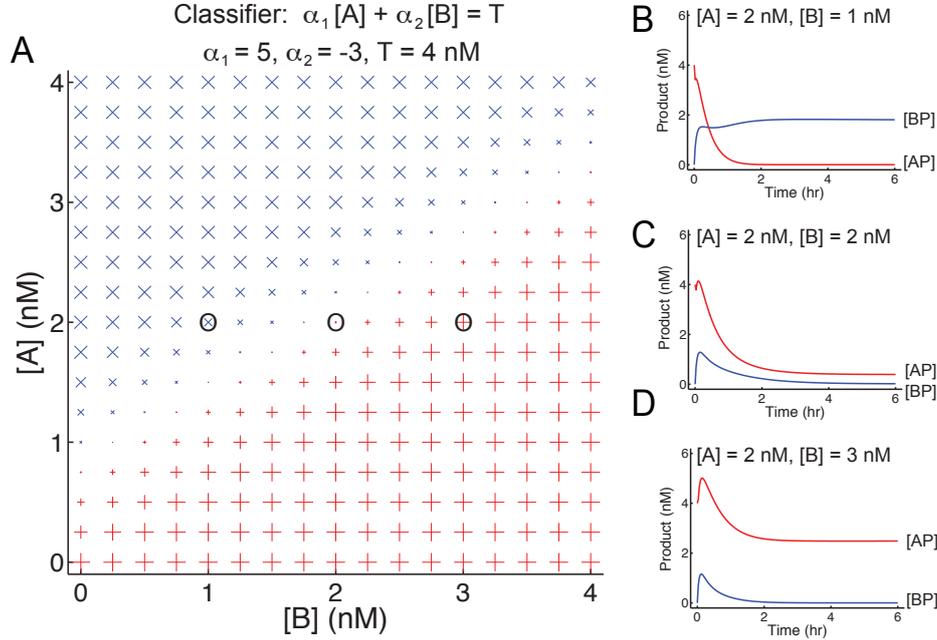
An element of the sum with a positive weight  $\alpha_i$  is implemented as a catalytic reaction with a fixed gain  $\alpha_i$ . An input  $m_i$  at initial concentration  $[m_i]_0$  results



**Fig. 5.** Implementing negative gain. We implement negative gain by having all inputs with positive gains catalytically produce one product  $AP$ , and all inputs with negative gains catalytically produce another product of independent sequence,  $BP$ . The products  $AP$  and  $BP$  stoichiometrically neutralize one another via the annihilator gate  $AG$  [24]. Excess  $AP$  at the end of the reaction denotes that the density classification expression evaluated to positive, while excess  $BP$  denotes the expression evaluated to negative.

in a final concentration  $\alpha_i[m_i]_0$  of an output strand  $AP$  of unrelated sequence. Importantly, the output strand is the same for all reactions with a positive  $\alpha_i$ . Similarly, every reaction with a negative  $\alpha_i$  is implemented as a catalytic reaction with a (positive) gain  $|\alpha_i|$  but a different output strand  $BP$ .

In principle, we could use reporters with two different colors to independently read out the the positive and negative output strands  $AP$  and  $BP$ . Using fluorescence calibration curves, we could then compute the respective concentrations



**Fig. 6.** DNA classifier. **(A)** Summary plot of the concentrations of *AP* and *BP* at the end of 6 hours of simulated reaction for various initial concentrations of *A* and *B*. Size of crosses denote the final concentration of *AP*; size of pluses denote final concentration of *BP*. **(B)**, **(C)**, **(D)** Sample concentration traces for *AP* and *BP*.

as well as the difference between them and compare the result to the threshold value  $T$ . However, such an approach would still require considerable intervention from an experimentalist meaning that only part of the computation is actually implemented as molecular computation.

To embed the comparison of the concentrations of *AP* and *BP* in the DNA molecules themselves, we use the annihilator gate design presented in Ref [24] (see also Fig. 5). In this design, each of *AP* and *BP* bind to annihilator gate *G* reversibly, but the combination of the two irreversibly binds to *G*, removing both from solution (Fig. 9-5). In an excess of annihilator gate *G*, only one of *AP* and *BP* will be present in solution at significant concentration. *G* is present in solution from the beginning of the beginning of the reaction, and serves to dynamically reduce the concentrations of both *AP* and *BP*. Note that a similar mutual annihilation reaction could also be implemented using the mechanism for implementing arbitrary bimolecular reactions explained in Ref. [23].

So far we have shown how to implement arbitrary positive and negative gains and how to perform molecular-level comparison of the concentrations of the resulting reporter strands *AP* or *BP*. This would be sufficient to implement a classifier with  $T = 0$ . To implement a non-zero value for the threshold  $T$  we simply add  $T$  units of *AP* or *BP* depending on the sign of  $T$ . In this way we

can implement a molecular classifier with arbitrary values for  $\alpha_i$  and  $T$  on the molecular level.

Fig. 6 shows an example of a simulation of a simple two-input linear classifier. The simulations use a realistic model for the underlying DNA reactions. Fig. 6A shows the expected final signal (i.e. the excess amount of  $AP$  or  $BP$ ) for a variety of “samples.” Each sample is characterized by a pair  $(A, B)$  of the two molecules of interest. Note that without further amplification of the final output (either  $AP$  or  $BP$ ) the signal linearly increases with the distance from the threshold line.

## 5 Conclusions

Here we have proposed a DNA implementation of a fixed gain amplifier and of linear classifier circuits. The fixed gain amplifier combines a DNA catalytic amplifier with a threshold element or an autocatalytic reaction in order to obtain arbitrary gain that can be lower or higher than the intrinsic gain of the DNA catalyist. Classifier circuits similar to the one proposed here can potentially be used for the embedded analysis of RNA expression levels in complex mixtures. Such classification circuits could find applications in point-of-care diagnostics or could even be used to analyze gene expression in living cells.

To apply the presented linear classifier circuit to actual cell state classification, however, the classifier must be able to deal with RNA input concentrations that are often low and can vary by orders of magnitude. While in theory the methods presented should be able to allow indefinitely high values of  $\alpha$ , the precise control of large values of  $\alpha$  will be difficult in practice, because the intrinsic turnover set by strand purities will not be known to great accuracy. Additionally, achieving high turnover will be slow, because each turnover requires a fixed amount of time for reaction.

Multi-stage fixed turnover amplifiers can be used to combat the aforementioned difficulties. That is, the products  $AP$  and  $BP$  can be themselves amplified by another fixed gain amplifier, and the gains of the two systems will be multiplied. Achieving high turnovers with a 2-stage system will also be quadratically faster. For extremely high turnovers, even more stages of fixed amplification can be cascaded.

There are a variety of alternatives to the specific implementation proposed here. In particular, the chemical reaction systems networks of Ref. [23] can be used to implement the reactions described here. However, the catalytic system of Ref. [15] is currently the best characterized and also fastest catalytic amplifier available which is why we chose to use this system for our design.

The reactions and mechanisms used to construct the linear classifier have either been demonstrated or are similar enough to well-understood reactions that they are expected to experimentally function as designed. All simulation results shown include modeling of relevant intermediate species and side reactions; similar modeling has been able to quantitatively predict the kinetics of similar DNA constructions [4] [17]. Thus, we are optimistic that we can experimentally demonstrate the density classifier circuit *in vitro* in the near future.

**Acknowledgments.** DYZ is supported by the Fannie and John Hertz Foundation. GS is supported by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund and an NSF CAREER award.

## References

1. Bloomfield, V.A., Crothers, D.M., Tinoco Jr., I.: *Nucleic Acids: Structures, Properties, and Functions*. University Science Books, Sausalito (2000)
2. SantaLucia, J., Hicks, D.: *Annu. Rev. Biophys. Biomol. Struct.* 33, 415 (2004)
3. Yurke, B., Mills, A.P.: *Genet. Prog. Evol. Mach.* 4, 111 (2003)
4. Zhang, D.Y., Winfree, E.: *J. Am. Chem. Soc.* 131, 17303 (2009)
5. Seelig, G., Soloveichik, D., Zhang, D.Y., Winfree, E.: *Science* 314, 1585 (2006)
6. Hagiya, M., Yaegashi, S., Takahashi, K.: *Nanotechnology: Science and Computation*, pp. 293–308 (2006)
7. Frezza, B.M., Cockroft, S.L., Ghadiri, M.R.: *J. Am. Chem. Soc.* 129, 14875 (2007)
8. Qian, L., Winfree, E.: A Simple DNA Gate Motif for Synthesizing Large-Scale Circuits. In: Goel, A., Simmel, F.C., Sosik, P. (eds.) *DNA 14. LNCS*, vol. 5347, pp. 70–89. Springer, Heidelberg (2009)
9. Yurke, B., Turberfield, A.J., Mills, A.P., Simmel, F.C., Neumann, J.L.: *Nature* 406, 605 (2000)
10. Bath, J., Turberfield, A.J.: *Nat. Nanotech.* 2, 275 (2007)
11. Turberfield, A.J., Mitchell, J.C., Yurke, B., Mills, A.P., Blakey, M.I., Simmel, F.C.: *Phys. Rev. Lett.* 90, 118102 (2003)
12. Bois, J.S., Venkataraman, S., Choi, H.M.T., Spakowitz, A.J., Wang, Z.G., Pierce, N.A.: *Nuc. Acid Res.* 33, 4090 (2005)
13. Green, S.J., Lubrich, D., Turberfield, A.J.: *Biophysical Journal* 91, 2966 (2006)
14. Seelig, G., Yurke, B., Winfree, E.: *J. Am. Chem. Soc.* 128, 12211 (2006)
15. Zhang, D.Y., Turberfield, A.J., Yurke, B., Winfree, E.: *Science* 318, 1121 (2007)
16. Yin, P., Choi, H.M.T., Calvert, C.R., Pierce, N.A.: *Nature* 451, 318 (2008)
17. Zhang, D.Y., Winfree, E.: *Nuc. Acid Res.* (2010, pre-published online doi:10.1093/nar/gkq088)
18. Masu, H., Narita, A., Tokunaga, T., Ohashi, M., Aoyana, Y., Sando, S.: *Angew. Chemie Int. Ed.* 48, 9481 (2009)
19. Xie, Z., Liu, S.J., Bleris, L., Benenson, Y.: *Nuc. Acids Res.* (2010, doi:10.1093/nar/gkq117)
20. Benenson, Y., Gil, B., Ben-Dor, U., Adar, R., Shapiro, E.: *Nature* 429, 423 (2004)
21. Lu, J., et al.: *Nature* 435, 834 (2005)
22. Rosenfeld, N., et al.: *Nat. Biotech.* 26, 462 (2008)
23. Soloveichik, D., Seelig, G., Winfree, E.: *Proc. Nat. Acad. Sci.* (2010, pre-published online doi:10.1073/pnas.0909380107)
24. Zhang, D.Y.: Cooperative DNA strand displacement for DNA quantitation, detection, and logic (submitted, 2010)