NUCLEIC ACID DETECTION USING BIOLUMINESCENCE REGENERATIVE CYCLE AND STATISTICAL SIGNAL PROCESSING

H. Vikalo, A. Hassibi‡, and B. Hassibi**

*California Institute of Technology, Pasadena, CA 91125
‡Stanford University, Stanford, CA 94309

ABSTRACT

An important emerging research area is the study and development of signal processing techniques for rapid real-time nucleic acid detection [1]. In this paper, we focus on the newly developed bioluminescence regenerative cycle (BRC) technique, and apply statistical signal processing to the data identification problem. This extended summary provides a description of the BRC platform and experiments, the statistical model employed for analysis, and some preliminary experimental results.

1. ON NUCLEIC ACID DETECTION

The identification and quantification of nucleic acid molecules is an eminent part of genomics, as well as other related fields in life sciences and medicine. Currently, an assortment of detection platforms for genetic analysis are available, with their applications varying from structural analysis systems such as DNA sequencing [2] or single-nucleotide-polymorphism (SNP) detection [3], to quantification by DNA microarrays [4]. A certain group of these methodologies takes advantage of the intrinsic capability of DNA to replicate itself. As a rule, in such systems, incorporation of various nucleotides (dNTPs) by polymerase enzyme (e.g. Klenow [5]) onto the strand, results in a measurable quantity associated with a particular characteristic of the molecule. The detection in these systems is typically carried out subsequent to the polymerization event, when the kinetics of nucleotide incorporation [6] is not apparent.

One polymerization-based system which carries out sequence by synthesis is pyrosequencing [7]. This technique detects the byproduct of DNA polymerization (inorganic pyrophosphate, PPi) by linking it to a bioluminescence assay which consist of ATP-sulfurylase, and Firefly luciferase. The enzymatic mix initially generates an ATP molecule per single PPi, and later generates a single photon from that ATP.

Bioluminescence is essentially the generation of electromagnetic radiation as light by the process of releasing energy from a biochemical reaction. While the generated photons can be emitted in a wide range of wavelengths from ultra violet to infrared, those that emit visible light are the most common in science. Bioluminescence and other various luminescence processes are often described as "cold light" since no external energy source is required and the chemical reaction generates photons spontaneously. There are two main types of bioluminescence assays: substrate detection and catalyst detection. Both generate little heat (loss), and have high photon generation efficiency (e.g., luciferase/luciferin process has quantum efficiency of 0.88 [8]).

The photon generation rate is a function of many variables: the molecules participating in the luminescence process, their reaction kinetics and the exact implementation of the luminescence entities in the assay. Exact calculation of the observed light intensity as a function of time, $I(t)$, requires knowledge of the initial concentrations of the substrate $[S_0]$ and the enzyme (catalyst) $[E_0]$, the forward and reverse reaction rates (k_f and k_r, respectively), the bioluminescence quantum efficiency α, the reaction volume V, and the optical path loss Φ.

For substrate detection it turns out that

$$I(t) = (\alpha VA)\Phi[S_0][E]k_f e^{-(k_f+k_r)[E_0]t}, \tag{1}$$

whereas for enzyme detection,

$$I(t) = (\alpha VA)\Phi[S_0][E_0](k_f - k_r t), \tag{2}$$

where A is the Avogadro number.

As can be seen from (1), in substrate detection systems (e.g., in Luciferase/Luciferin ATP detection assays for biocontamination), the light generation has an exponential decay. This stands in contrast to enzyme detection assays where, permitting neglect of the inhibition factor, the light is relatively steady. As far as photo-sensors are concerned, assays with steady-state light are much more favorable since time-averaging schemes become applicable. However, the light intensity peak of such assays are very low, since enzyme concentration is typically minuscule. To overcome these problems, *Bioluminescence Regenerative Cycle* (BRC) has been proposed which has the light intensity of a substrate detection system, while having the light stability of the steady-state bioluminescence process [9].

2. THE BRC TECHNIQUE

BRC is a novel bioluminometric method of quantifying nucleic acid (NA) molecules where the assay does not require any molecular modification or fluorescent labeling and merely counts the inorganic pyrophosphate (PPi) molecules released during the polymerization of the nucleic acid by a polymerase enzyme (e.g., Klenow). This technique implements a bioluminescence reaction activated by the generated PPi molecules, the amount of which is proportional to the number of target NA molecules. The regenerative cycle is comprised of an ATP-sulfurylase enzyme which converts PPi to adenosine triphosphate (ATP) by consuming adenosine phosphosulfate (APS) and firefly luciferase (bioluminescence enzyme). The luciferase, in turn, consumes ATP as an energy source to generate photons that are detected. This light generation process also generates PPi as a byproduct (see Figure 1). The detailed biochemical reactions involved in BRC are as follows:

1. Generating PPi from DNA:

$$(\text{DNA})_n + \text{dNTP} \overset{\text{Polymerization}}{\longrightarrow} (\text{DNA})_{n+1} + \text{PPi}$$

2. Generating ATP from PPi:

$$\text{PPi} + \text{APS} \overset{\text{ATP-Sulfurylase}}{\longrightarrow} \text{ATP} + \text{SO}_4^{-2}$$

3. Photon generation and PPi regeneration:

$$\text{ATP+luciferin+O}_2 \overset{\text{Luciferase}}{\longrightarrow}$$
$$\text{AMP+oxyluciferin+CO}_2 + 0.8h\nu + \text{PPi}$$

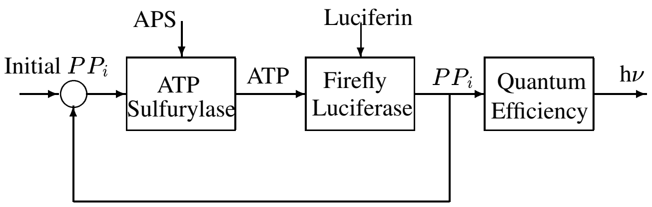

Fig. 1. The BRC block-diagram. The unity-gain feedback triggers the light generation process continually, by creating PPi in the photon generation process.

As implied by 1-3 above, the regeneration process can potentially create a unity-gain positive feedback for PPi, which in turn stabilizes the light intensity. In an ideal case where the concentration of the enzymes is unlimited, and no product inhibition is present, the expected light intensity ($\lambda = 565$nm) in the BRC assay is

$$I(t) = (\alpha VA)\Phi k_L[NA]L_p, \tag{3}$$

where $k_L$is the steady-state turn-over rate of ATP-sulfurylase and luciferase together, $[NA]$ is the nucleic acid concentration (typically between 10amol to 10pmol), and L_p is polymerization length (between 10 and 500).

Although it was demonstrated that this proportionality is in fact measurable over 5 orders of magnitude of NA concentration (i.e., assay dynamic range between 100amol to 10pmol) [9], in practice some light intensity decay is also observed. Moreover, it turns out that the reaction mix continually generates some amount of background light even without the presence of indigenous ATP and PPi or polymerized DNA.

3. MATHEMATICAL MODEL

In order to be able to deal with a realistic (non-ideal) BRC experiment one needs to develop a simple, yet relatively accurate model. This is also useful since in many applications the information of interest goes beyond the mere nucleic acid concentration, and may include the detection of certain reactions (SNP's, say) or even DNA sequences.

A simple model, which our experimental data confirms is sufficient for capturing the kinetics of the BRC reaction, is obtained by focusing on the pyrophosphate and APS (the first block in the loop of Figure 1) and by ignoring the luciferin (the second block) since it has a much higher reaction rate. With this simplification, everything of interest can be expressed in terms of the concentration of the APS, $e(t)$, and the concentration of the pyrophosphate, $p(t)$.

The relationship between $e(t)$ and $p(t)$ can be given by the kinetics of the simplified model, i.e.,

$$\frac{de(t)}{dt} = -K_1 e(t)p(t). \tag{4}$$

which states that the rate at which APS is burned is negatively proportional to the amount of APS and PPi available in the assay. Of course, (4) is not totally precise since it refers to the mean values of the concentrations and not necessarily their instantaneous ones. To construct a more precise model, let us begin by discretizing (4)

$$\frac{e(k+1) - e(k)}{\Delta} = -K_1 e(k)p(k),$$

and therefore

$$e(k+1) = e(k) - K_1\Delta e(k)p(k).$$

The constant $K_1\Delta p(k)$ can be interpreted as the probability that a particular APS molecule will be consumed during the time interval $[k, k+1]$ (and thus contribute to the decrease in the total APS concentration $e(k+1)$). We find that this probability can be efficiently modeled by

$$P_r = Pr(\text{an APS molecule is consumed}) = \frac{p(k)}{e(0) + p(k)},$$

where $e(0)$ denotes initial concentration of the APS.

Therefore the *average* number of APS molecules that are consumed during $[k, k+1]$ is given by $P_r e(k) = \frac{e(k)p(k)}{e(0)+p(k)}$. Assuming that the reactions are statistically independent, the *actual* number of APS molecules is given by

$$e(k+1) = \frac{e(0)e(k)}{e(0)+p(k)} + v(k), \tag{5}$$

where $v(k)$ is the uncertainty in the reaction process. $v(k)$ has clearly variance equal to $\sigma_v^2 = P_r(1-P_r)e(k) \approx P_r e(k) = \frac{e(k)p(k)}{e(0)+p(k)}$ and is well modeled as a Gaussian random process $\mathcal{N}(0, \sigma_v^2)$. (Note that the variance is proportional to the signal, reminiscent of the "shot-noise".)

The information that we seek is contained in the incoming PPi concentration. If we denote the information signal by $u(k)$, then we clearly have

$$p(k+1) = p(k) + u(k). \tag{6}$$

Finally, we need to describe the measurement process. We assume that we measure the light intensity (or number of photons). This is proportional to the number of PPi molecules consumed in the light generation process (which is the same as the number of APS molecules consumed). Thus, the measured signal is

$$y(k) = \frac{p(k)e(k)}{e(0)+p(k)} + v(k) + w(k), \tag{7}$$

where $w(k)$ is Gaussian measurement noise $\mathcal{N}(0, \sigma_w^2)$.

4. ML DETECTION

Our goal is to recover the information signal $u(\cdot)$ from the measurement signal $y(\cdot)$. Since (6) allows us to recover $u(\cdot)$ from $p(\cdot)$, we will focus on recovering $p(\cdot)$ from $y(\cdot)$. Since we have a full statistical description of the BRC process, we will employ the following maximum-likelihood (ML) criterion

$$\max_{e,p} f(e, p|y) \tag{8}$$

where the maximization in (8) is over time-sequences $e(\cdot)$ and $p(\cdot)$, and where $f(\cdot)$ denotes probability density function of its argument. Now, since

$$\arg\max_{e,p} f(e,p|y) = \arg\max_{e,p} f(y|e,p)f(e|p)f(p)\frac{1}{f(y)},$$

optimization problem (8) is equivalent to

$$\max_{e,p} f(y|e,p)f(e|p). \tag{9}$$

We note that

$$f(y|e,p) = \phi_y e^{-\sum_{t=1}^{T} \frac{1}{2\sigma_w^2}\left(y(k) - \frac{p(k)e(k)}{e(0)+p(k)} - e(t+1) + \frac{e(0)e(k)}{e(0)+p(k)}\right)^2},$$

where $\phi_y^{-1} = \prod_{i=1}^{T} \sqrt{2\pi\sigma_w^2}$. Furthermore,

$$f(e|p) = \phi_e e^{-\sum_{t=1}^{T} \frac{\left(e(t+1) - \frac{e(0)e(k)}{e(0)+p(k)}\right)^2}{2\frac{e(k)p(k)}{e(0)+p(k)}}},$$

where $\phi_e^{-1} = \prod_{t=1}^{T} \sqrt{2\pi \frac{e(k)p(k)}{e(0)+p(k)}}$.

We solve the optimization problem (9) numerically.

5. EXPERIMENTAL RESULTS

In this section we present preliminary experimental results. The BRC process is observed for various nucleic acid templates. In Figure 2, we consider the experiment where the PPi sequence $u(\cdot)$ is generated by a 10pmol, 20-base long TG template (see appendix for the detailed description of the DNA template). The plots in Figure 2, form top to bottom, show the measured signal $y(\cdot)$, the estimated PPi concentration $p(\cdot)$, and the estimated input sequence concentration $u(\cdot)$, respectively. The BRC process is observed over 180 seconds, as indicated on the horizontal axis. However, the nucleic acid polymerization lasts less than 10 seconds, as indicated by the shape of the recovered sequence $u(\cdot)$.

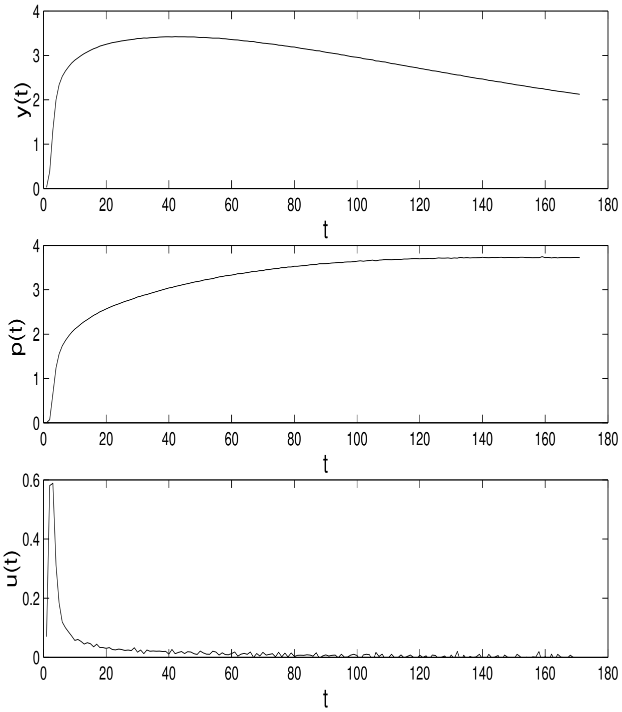

Fig. 2. The measurement, $y(\cdot)$, and the estimated $p(\cdot)$ and $u(\cdot)$ for the BRC process initiated by the polymerization of a $[NA] = $ 10pmol, 20-base long TG template.

The plots in Figure 3 have the same meaning, only for a 10pmol, 20-base long AG template. Comparing Figure 3 with Figure 2, we note that the input sequences $u(\cdot)$, generated by the two templates (TG and AG) that we used in the experiments and recovered by the proposed estimation technique, are clearly distinct.

It is of high interest to further examine the correlation between $u(\cdot)$ and the corresponding template that has generated it. This would be an important step in the direction of using the proposed technique to performing real-time DNA sequencing.

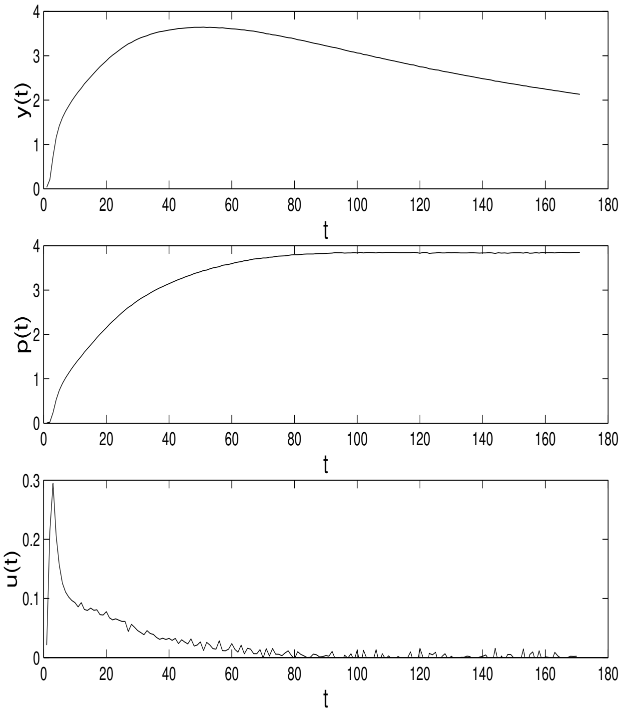

Fig. 3. The measurement, $y(\cdot)$, and the estimated $p(\cdot)$ and $u(\cdot)$ for the BRC process initiated by the polymerization of a $[NA] = 10$pmol, 20-base long AG template.

6. APPENDIX: EXPERIMENT SPECIFICATIONS

DNA Template:

The DNA oligo-loop structures (self-primed strands) were all HPLC purified after synthesis (Qiagen, USA) with the loop structure sequence of 5'-GCCGTCGTTTTACAACGG AACGTTGTAAAACGACGG C-3'. The overall sequence of the strands contained an additional 20 base extension template attached to the 5' end of the loop such that 5'-(X)20-loop-3' (X being A,C,G, or T) or 5'-(Y)10-loop-3' (Y being AC, AG, TC, or TG).

Assay Procedure:

The 45ul Bioluminescence assay mixture contained: 0.1M Tris-acetate (pH7.75), 0.5mM EDTA (Sigma,USA), 5 mM Mg-acetate (Sigma, USA), 0.1%(w/v) bovine serum albumin (Sigma, USA), 2.5 mM dithiothreitol (Sigma, USA),

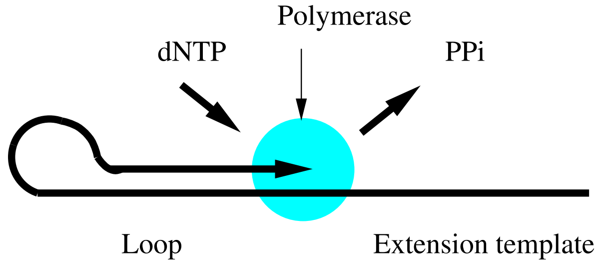

Fig. 4. DNA template.

2ul of exonuclease-deficient Klenow DNA polymerase (10 U/ul; Fermentas, NY, USA) 10uM adnosine 5'-phosphosulfate (APS) (Biolog, Germany), 0.4mg/ml polyvinylpyrrolidone (MW 36000), 100ug/ml D-luciferin (BioThema, Uppsala, Sweden), 3.0ug luciferase (Promega, USA) and 50mU recombinant ATP-sulfurylase (Sigma, USA) or recombinant thermo-stable ATP-sulfurylase (from Aquifex aeolicus, purified in Panoroma Research Inc., USA). 10pmol of oligo-loop diluted in 0.1M Tris-acetate (pH7.75) was added to the reaction mix, and subsequently to generate the bioluminescence light, 2ul of all 4 nucleotides (dCTP, dGTP, dTTP and dATPαS) with 1.4 mM final concentration was injected into the reaction mix. The emitted photons were measured using H5783-03 PMT (Hamamatsu, Japan).

7. REFERENCES

[1] D. Anastassiou, "Genomic signal processing," *IEEE Signal Processing Magazine,* vol. 18, no. 4, July 2001.

[2] L. Alphey, *DNA Sequencing: From Experimental Methods to Bioinformatics,* Bios Scientific Pub Ltd, 1997.

[3] Hajeer A., Worthington, J., John, S., *SNp and Microsatellite Genotyping: Markers for Genetic Analysis,* Eaton Pub Co, 2000.

[4] M. Schena, D. Shalon, R. Heller, A. Chai, P. O. Brown, and R. W. Davis, *Proc Natl Acad Sci USA*, vol. 93, pp. 10614-9, 1996.

[5] V. Derbyshire, P. S. Freemont, M. R. Sanderson, L. Beese, J. M. Friedman, C. M. Joyce, and T. A. Steitz, *Science*, vol. 240, pp. 199-201, 1988.

[6] M. E. Dahlberg and and S. J. Benkovic, *Biochemistry*, vol. 30, pp. 4835-43, 1991.

[7] M. Ronaghi, M. Uhlen, and P. Nyren, *Science*, vol. 281, pp. 363-365, 1998.

[8] T. O. Baldwin and V. A. Green, *Methods Enzymol*, vol. 305, pp. 180-188, 2000.

[9] A. Hassibi, S. Ghazuini, adn B. Hassibi, "Methods and Apparatus for Pathogen Detection, Identification, and/or Quantification," US Patent Application no. 10/627,332.