



دانشگاه شهید بهشتی

تشخیص بهینه مبتنی بر فلورسانس در فرایند سنتز تک مولکولی

نگارنده :

علی نیک اختر

استاد راهنما :

دکتر علی کتائفروش

بهمن 1400

## چکیده

- توالی یابی تک مولکولی برای تمام زمینه های تجزیه و تحلیل های ژنتیکی مانند ژنومیک، رونویسی ژنها، آزمایشهای بالینی، داروسازی، و غربالگری سرطان ضروری است.
- علاوه بر این برای توالی یابی تک مولکولها بخصوص در زمینه های توالی یابی DNA آنها در اولویت اول و قبل از استفاده از روشهای دیگر از روش توالی یابی مبتنی بر فلورسانس استفاده میشود.
- روش مدرن برجسب گذاری فلورسانس از یک دوربین با قابلیت عکس برداری فوری از تعدادی پیکسل از روی توالی های تک مولکولی بهره میگیرد.
- در این مقاله روش تشخیص برجسب در فلورسانس با استفاده از یک تک پیکسلی مورد بحث قرار گرفته است.
- ویژگیهای این روش از جمله برتری در میزان دقت و استفاده بهینه از منابع در شرایط فقدان سیگنال با کیفیت مناسب و همچنین توان عملیاتی این روش در برابر سایر روش ها نیز در این مقاله مورد بحث قرار گرفته است.
- در این مقاله ما سعی داریم تا فرایند سنتز تک مولکولی را با استفاده از یک توزیع دو جمله ای منفی مدلسازی کنیم.
- علاوه بر این تلفیق روش حداکثر درستنمایی (maximum likelihood) و الگوریتم ویتربی (Viterbi) در این مدلسازی باعث افزایش دقت در تشخیص سیگنال ها میشود.
- مدل مبتنی بر فلورسانس دارای ویژگی شبیه سازی نزدیک به واقعیت فرایند های فلورسانس و همچنین ابزار مناسبی جهت درک روابط بین انتشارات حاصل از فلورسانس و رویداد های دریافت سیگنال می باشد.
- این مدل با پیشنهاد های مناسبی که جهت انتخاب رنگ فلورسانس ارائه می دهد، نتایج آزمایش دقیق تری را به همراه دارد.

## 1-مقدمه

به لطف پیشرفت جمعی در فناوری، در زمینه ی علم بیولوژیک نیز یک رشد تصاعدی را خصوصاً در زمینه ی توالی یابی نوکلئوتیدها شاهد بودیم و این منجر به تولد استراتژی های تحلیلی و روش های محاسباتی مختلفی شده است. به لطف روش های جدید تشخیص اسید نوکلئیک ها، توالی یابی ژنوم پیشرفت بسزایی هم از نظر عملکرد و هم از نظر دقت داشته است.

در بین همه این روش ها روش سیگنال دهی و برچسبگذاری فلورسانس کارایی بسیار بالایی از خودش نشون داده است.

این روش دارای ویژگی هایی از جمله حساسیت بالاتر، قابلیت های مرکب و چندگانه و تشخیص صفات فلورسانس به صورت همزمان میباشد که باعث محبوبیت آن شده است.

تا به امروز روش های مبتنی بر (agent) زیادی برای افزایش دقت در زمینه شناسایی مولکول های موجود در طبیعت مانند پروتئین ها و نوکلئیک اسیدها مورد استفاده قرار گرفته اند و با این حال روش برچسب گذاری فلورسانس به دلیل داشتن قابلیت تشخیص با استفاده از سیگنالهای نوری منتشر شده حاصل از تحریک agent و مولکول هدف در طول توالی یابی در سطح خود بسیار مطرح است.

علاوه بر این هر ایجنت فلورسانس مربوط می شود به ۱ صفت فلورسانسی یا یک طول موج نوری مشخص. در نتیجه دستگاه های تشخیص قادر به انجام مشاهدات همزمان بر روی بیش از یک مولکول هستند که این به معنی صرفه جویی در مدت زمان لازم برای تشخیص میباشد.

یک مولکول هدف که تحت یک رویداد فلورسنت قرار می گیرد تحریک شده و نور را با طول موج متفاوتی از آنچه قبلاً در معرض آن قرار گرفته بود منتشر می کند .

پلتفرم با استفاده از دوربین تعبیه شده ای تصاویر فرآیند انتشار را دریافت کرده و توالی را از روی داده ها تشخیص می دهد، با این حال این روش نیازمند استخراج بهینه سیگنال ها جهت افزایش نرخ نسبت سیگنال به نویز یا SNR می باشد که در واقع این همان عامل اصلی محدودیت در مباحث جمع آوری داده در روش های مرسوم تشخیص مبتنی بر فلورسانس میباشد.

از آنجایی که ما از این روشها جهت تشخیص سنتز های تک مولکولی استفاده میکنیم معایبی از قبیل نرخ خطای بالا، هزینه در ازای هر بیس بالاتر، و توان زمانی پایین تر وجود دارند.

اگرچه روش توالی یابی اجماعی جهت جبران نرخ خطای بالا استفاده می شود، اما این روش نیازمند توان پردازشی بسیار بالا، تمپلیت های پیچیده و همچنین صرف زمان بیشتری جهت تکرار عملیات ضروری میباشد که در این صورت دوره سنتز ها و شدت انتشار فلورسانس عملکرد و دقت فرایند را تحت تاثیر قرار میدهد.

جهت بهبود کارای روش رمزگشایی و کاهش هزینه به ازای هر بیس در توالی یابی، ما یک مدل آماری برای شبیه سازی سنتزهای پلیمرز و فرآیند انتشار فلورسنت آن می سازیم و داده ها را با یک فوتودیود ۳ اتصال ثابت می کنیم.

در الگوریتم مورد مطالعه برای اینکه بتوانیم سیگنالها را شناسایی کرده و امکان رمزگشایی توالی ها را در شرایطی که نرخ SNR پایین تری در اختیار داریم فراهم سازیم، از روش حداکثر درستی ML (maximum likelihood) و الگوریتم ویتربی (viterbi) استفاده میکنیم.

قابلیت شناسایی رویدادهای فلورسنت بسته به طول موجی که از رنگ های فلورسنت ارائه می شود متفاوت است. در سناریوهای خاصی، انتخاب رنگ های فلورسنت در دقت تشخیص نقش بسیار مهمی دارد، با این حال کارایی آنها را نمی توان تا قبل از زمانی که توالی یابی کامل شده باشد اندازه گیری کرد و در نتیجه زمان و منابع زیادی مصرف می شود.

مدل های زیر به انتخاب مواد فلورسنت کمک می کنند و نشان می دهند که چگونه پارامترهای مختلف بر عملکرد این روش تأثیر می گذارند.

## 2. مواد و مدل

### 2.1- فرایند سنتز تک مولکولی

نور منتشر شده توسط واکنش سنتز پلیمر از نقش مهمی در روشهای توالی یابی تک مولکولی مبتنی بر فلورسانس ایفا میکند. (شکل 1)

از آنجایی که DNA پلیمر از نوکلئوتیدها را در یک توالی سنتز می کند ما پلتفرم توالی یابی long-read را به عنوان نمونه ای برای نمایش فرآیند سنتز تک مولکولی (SMSP) در نظر می گیریم.

SMSP یا همان فرایند سنتز تک مولکولی به عنوان یک فرآیند تصادفی زمان گسسته  $\{Z(t), t \geq 1\}$  در فضای حالت  $\{A, T, G, C, d\}$  مدل سازی می شود، که در آن حالت A (T، G، یا C) نشان می دهد که یک dATP (dGTP، یا dCTP) در حال ترکیب شدن توسط DNA پلیمر است و حالت d نشان می دهد که هیچ dNTP ای ترکیب نشده است.

$S_n$  و  $T_{n-1}$  را زمان شروع و زمان توقف n امین ادغام dNTP برای  $n \geq 1$  در نظر میگیریم. پس داریم:

$$0 = T_0 \leq S_1 < T_1 \leq S_2 < T_2 \leq \dots \leq S_n < T_n \leq \dots$$

$X_n \equiv Z(S_n)$  را n امین نوکلئوتیدی که ترکیب می شود در نظر بگیرید.

. طول n امین مدت زمان میان پالسی  $W_n = S_n - T_{n-1}$  و طول n امین مدت دوره ترکیب یا عرض پالس (در اینجا به عنوان عرض پالس نشان داده شده است)  $Y_n = T_n - S_n$  است.

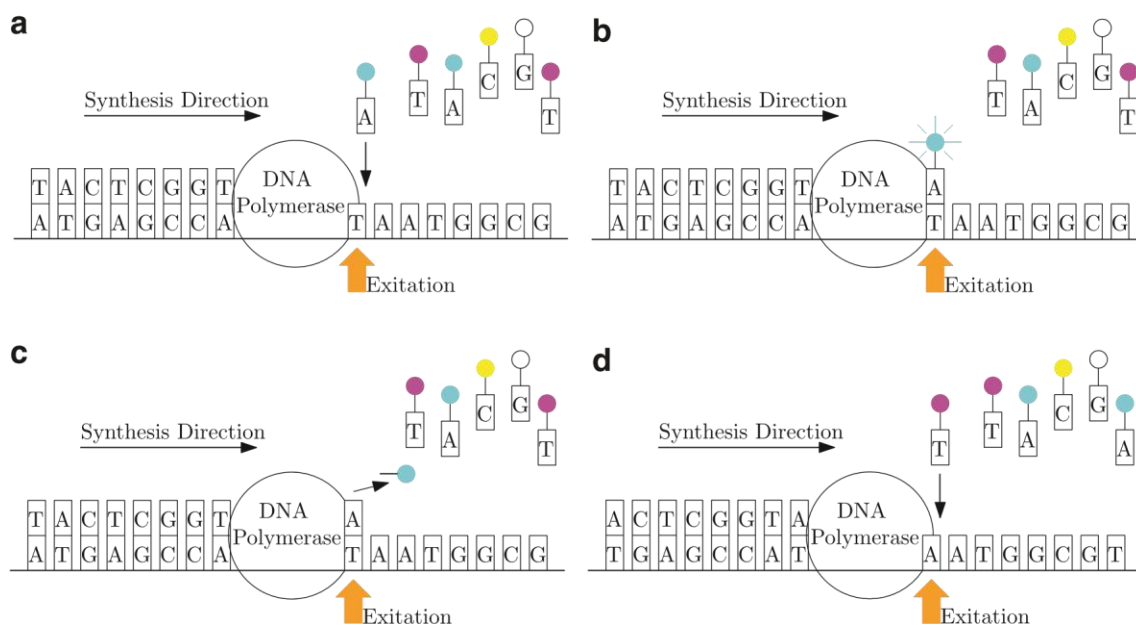
فرآیند  $\{(W_n, Y_n), n \geq 1\}$  یک فرایند متناوب در نظر گرفته می شود به طوری که  $W_n$  و  $Y_n$  از نظر آماری مستقل هستند.

به دلیل فقدان ساختار اطلاعاتی در یک قطعه کوتاه DNA، می توان فرض کرد که  $\{X_n, n \geq 1\}$  یک دنباله مستقل با توزیع یکسان از متغیرهای تصادفی با توزیع یکریخت روی  $\{A, T, G, C\}$  بوده که از فرایند متناوب  $\{(W_n, Y_n), n \geq 1\}$  نیز مستقل میباشد.

فرض کنید  $r$  کوچکترین دوره زمانی است که در طی آن فلورسانس نور ساطع می کند و  $l$  کوچکترین دوره زمانی است که در طی آن پلیمراز به مکان فعال بعدی منتقل می شود. طبق مطالعه قبلی، توزیع عرض پالس و مدت زمان بین پالس به طور یکنواخت توزیع نمی شوند و توزیع های  $W_n$  و  $Y_n$  را با توزیع های دوجمله ای منفی به ترتیب با پارامترهای  $(l, q)$  و  $(r, p)$  مدل سازی می کنیم.:

$$P(W_n = k) = C_{l-1}^{k-1} q^l (1-q)^{k-l}, k = l, l+1, l+2, \dots \quad (1)$$

$$P(Y_n = k) = C_{r-1}^{k-1} p^r (1-p)^{k-r}, k = r, r+1, r+2, \dots \quad (2)$$



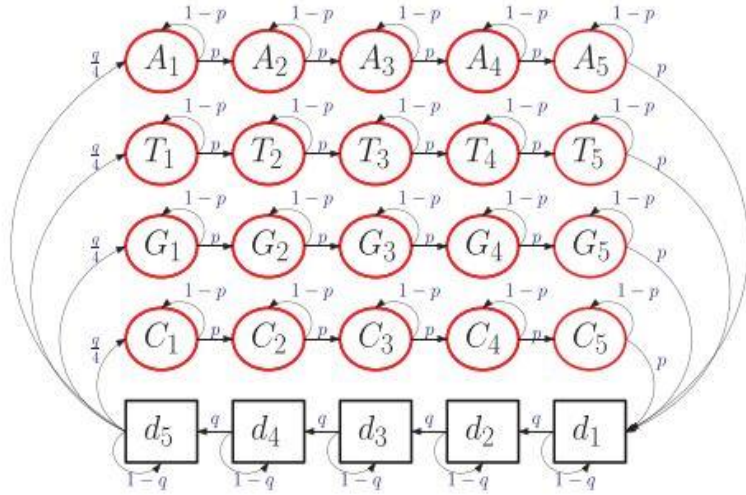
شکل 1. فرآیند سنتز پلیمراز. (a) نوکلئوتید مرتبط با فلورسانس به محل فعال پلیمراز متصل می شود. (b) رنگ فلورسانس هنگام برانگیختگی شروع به روشن شدن می کند. (c) رنگ فلورسانس از نوکلئوتید جدا می شود، بنابراین روشن نمی شود. (d) DNA پلیمراز به محل فعال بعدی منتقل می شود.

با مفروضات فوق،  $\text{SMSP} \{(W_n, Y_n, X_n), n \geq 1\}$  به یک زنجیره مارکوف زمان گسسته  $\{L(t), t \geq 1\}$  با فضای حالت

با ساین  $4r+l$  می باشد.  $S = \{d_1, d_2, \dots, d_l, A_1, A_2, \dots, A_r, T_1, T_2, \dots, T_r, G_1, G_2, \dots, G_r, C_1, C_2, \dots, C_r\}$

فرض کنید  $S_i = \{A_i, T_i, G_i, C_i\}$  نمودار انتقال حالت زنجیره مارکوف  $\{L(t), t \geq 1\}$  در شکل 2 است و احتمالات انتقال حالت عبارتند از

$$\pi_{d_i, s} = \begin{cases} 1-q, & \text{if } s = d_i, \\ q, & \text{if } s = d_{i+1}, \text{ for } 1 \leq i \leq l-1, \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$



شکل 2. نمودار انتقال حالت زنجیره مارکوف  $\{L(t), t \geq 1\}$ . در مثال نشان داده شده،  $r=5$  و  $l=5$ .

$$\pi_{d_i, s} = \begin{cases} 1-q, & \text{if } s = d_i, \\ \frac{q}{4}, & \text{if } s \in S_1, \\ 0, & \text{otherwise,} \end{cases} \quad (4)$$

$$\pi_{A_i, s} = \begin{cases} 1-p, & \text{if } s = A_i, \\ p, & \text{if } s = A_{i+1}, \text{ for } 1 \leq i \leq k-1, \\ 0, & \text{otherwise,} \end{cases} \quad (5)$$

$$\pi_{A_k, s} = \begin{cases} 1-p, & \text{if } s = A_k, \\ p, & \text{if } s = d_1, \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

با فرض معادله (2)، احتمالات انتقال،  $\pi_{C_i, s}$ ،  $\pi_{A_i, s}$ ،  $\pi_{T_i, s}$ ،  $\pi_{G_i, s}$ ، و  $\pi_{C_i, s}$  یکسان هستند.

در مرحله بعد، شباهت ویژگی های مدل SMSP خود را در معادلات (1) و (2) با مطالعه قبلی نشان می دهیم.

ما فرض می کنیم که کوچکترین عرض پالس و کمترین مدت زمان میان پالسی در SMSP هر دو 100 میلی ثانیه باشد که با واحد زمان موجود در ماژول مدار مجتمع 25 میلی ثانیه میشود. بنابراین یعنی در معادلات (1) و (2)،  $l=5$

و  $r=5$  و نمودار انتقال مربوطه در شکل 2 نشان داده شده است. علاوه بر این، بزرگترین عرض پالس و بیشترین مدت زمان بین پالس به ترتیب به 500 میلی ثانیه و 5 ثانیه محدود شده‌اند.

با محدود کردن احتمال  $p(Y_n \geq 21(500ms)) \leq 10^{-3}$  و  $p(W_n \geq 201(5s)) \leq 10^{-2}$ ، که به این معنی است که طول عرض پالس و مدت زمان بین پالس محدود می‌شود، توابع جرم احتمال تنظیم می‌شوند. (مانند شکل 3a به  $p=0.57$  و  $q=0.05$ ).

. توزیع نتایج شبیه سازی از نظر ویژگی های پالس و آمار ردیابی مشابه مطالعه قبلی (Eid et al., 2009) است که در شکل 3 نشان داده شده است.

## 2.2- فرایند انتشار

انتشار نور فلورسانس به معنای ترکیب شدن یک نوکلئید توسط DNA پلیمراز می‌باشد. ما از روشی استفاده می‌کنیم که از یک فتودیود سه اتصالی برای گرفتن تصویر در طول انتشار نور استفاده می‌کند.

در زمانی که هیچ نوکلئیدی در حال ترکیب شدن نیست، فقط نور محیط تشخیص داده می‌شود، در صورتیکه نور محیط در یک دوره ترکیب شدن نیز وجود دارد.

در نتیجه ردیاب نوری (آشکارساز) سیگنال فلورسانس به اضافه سیگنال شدت نور محیط در طول دوره ترکیب و سیگنال شدت نور محیط خارج از دوره زمانی عمل ترکیب را به عنوان خروجی می‌دهد.

فرض کنید  $\{E(t), t \geq 1\}$  سیگنال شدت خروجی آشکارساز نوری مرتبط با SMSP  $\{Z(t), t \geq 1\}$  باشد که فرآیند انتشار نامیده می‌شود. فرض کنید که بردار انتشار سه بعدی  $E(t)$  فقط به حالت  $Z(t)$  در زمان  $t$  بستگی دارد. توجه داشته باشید که  $\{Z(t), t \geq 1\}$  نشان دهنده SMSP  $\{(W_n, Y_n, X_n), n \geq 1\}$  و

$$Z(t) = \begin{cases} d & , \text{if } \sum_{i=1}^{n-1} (W_i + Y_i) < t \leq \sum_{i=1}^{n-1} (W_i + Y_i) + W_n, \\ X_n & , \text{if } \sum_{i=1}^{n-1} (W_i + Y_i) + W_n < t \leq \sum_{i=1}^n (W_i + Y_i). \end{cases}$$



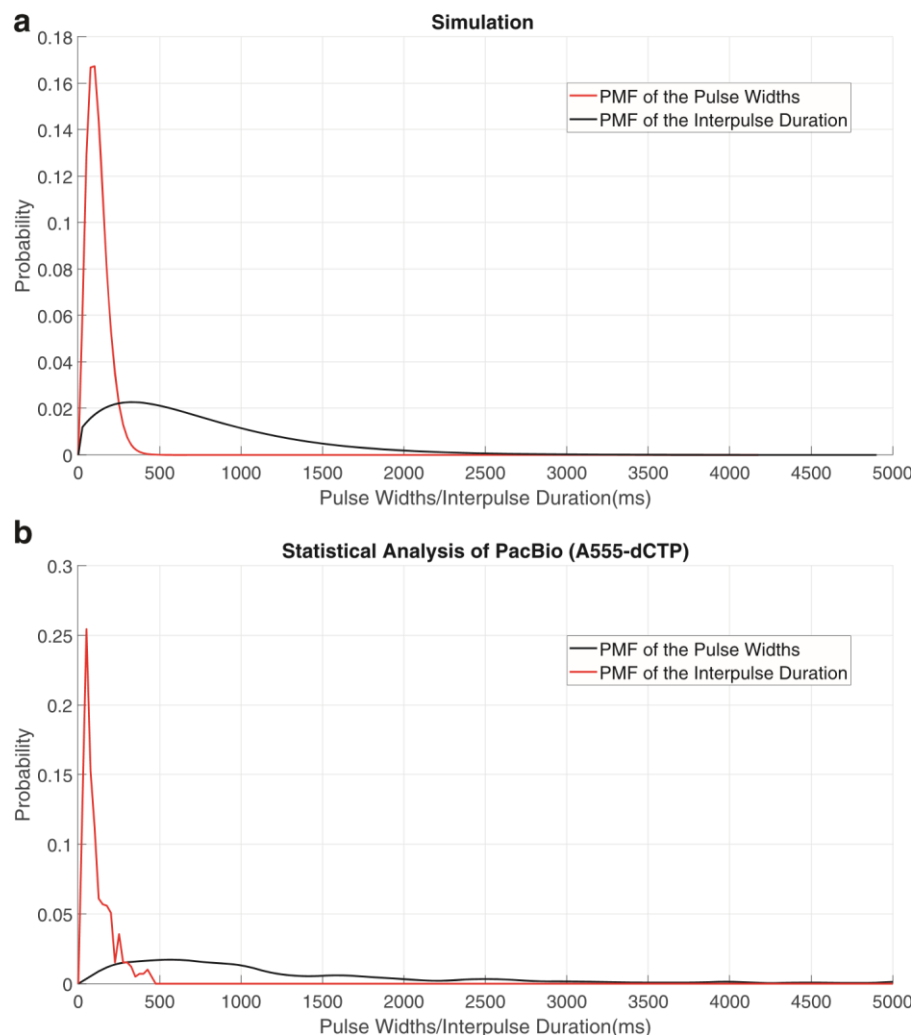
### 2.2.1- شدت سیگنال نور محیط

در شرایطی که منبع نور محیط از dNTP ها منشأ گرفته باشد یا در طول مدت interpulse (میان پالسی) در حالت ثابت باشد، سیگنال شدت نور محیط را می توان با یک بردار سیگنال ثابت  $az(t)$  مدل کرد، که مشخص نیست و برای ردیاب نور در قالب یک پیکسل تخمین زده می شود.

### 2.2.2- شدت سیگنال نور فلورسانس

سیگنال شدت نور فلورسانس در طول یک دوره ترکیب، به مولکول رنگ متصل به نوکلئوتید تحت سنتز و همچنین فاصله بین مولکول رنگ و فتودیود بستگی دارد. همچنین، واریانس در واکنش های بیوشیمیایی بر شدت نور تأثیر می گذارد. واکنش های بیوشیمیایی در یک فضای بسته انجام می شود که در آن نور فلورسانس می تواند به طور کامل توسط ردیاب نوری جذب شود. به چنین فضای بسته ای چاه سنتز می گویند.

فرض کنید که DNA پلیمراز در طول یک دوره ترکیب، یک موقعیت ثابت اسمی در چاه سنتز دارد، به طوری که dNTP گرفته شده به علاوه مولکول رنگ آنزیم یک فاصله اسمی ثابت از فتودیود داشته باشد. بنابراین، اگر  $Z(t)=x$  برای  $t \in [S_n, T_n-1]$ ، که در آن  $x \in \{A, T, G, C\}$  باشد، سیگنال شدت نور فلورسانس  $Y(t)s_x, t \geq 1$  خواهد بود که در آن  $s_x$  بردار سیگنال شناسایی شده فتودیود سه اتصالی به نور فلورسانس ساطع شده از مولکول رنگ است که با یک dxTP در فاصله اسمی از فتودیود متصل شده است.



شکل 3. (a) خط قرمز نشان دهنده PMF عرض پالس در رابطه (2) با  $p=0.57$ ، و خط سیاه نشان دهنده PMF مدت زمان interpulse (میان پالسی) در معادله (1) با  $q=0.05$  است. (b) نتایج PMF های آزمایش آقای Ed و همکارانش. (2009). PMF، تابع جرم احتمالی. (probability mass function)

ضریب محو شدن ناشی از تغییر فاصله واقعی مولکول رنگ تا فتودیود در زمان  $t$  دوره ادغام نسبت به  $Y(t)$  فاصله اسمی و  $Y(t)=0$  به صورت  $Z(t)=d$  است. از آنجایی که محدودیت‌هایی در موقعیت DNA پلیمراز و تغییرات واکنش‌های بیوشیمیایی وجود دارد، شدت انتشار  $Y(t)$  دارای حداقل و حداکثر مقادیر

به ترتیب هنگام سنتر ATP، TTP، GTP و CTP. مییاشد.  $(I_{T, \min}, I_{T, \max}), (I_{G, \min}, I_{G, \max}), (I_{C, \min}, I_{C, \max})$

هر بردار سیگنال مشخصه  $s_x, x \in \{A, T, G, C\}$  یک بردار پارامتر نامشخص است و می تواند تخمین زده شود.  $\{Y(t), t \geq 1\}$  فرآیند جوش و خروجی فلورسانس در فوتون/ms است.

### 2.2.3. انتشار در طول مدت زمان اینترپالس (میان پالسی)

سیگنال انتشاری در طول Nمین مدت زمان اینترپالس (میان پالسی)  $[T_{n-1}, S_n - 1]$  برابر است با :

$$E(t) = a_d, \quad t \in [T_{n-1}, S_n - 1].$$

### 2.2.4. انتشار در طول مدت زمان دوره های ترکیب

سیگنال انتشاری در طول Nمین دوره ترکیب  $[S_n, T_n - 1]$  برابر است با :

$$E(t) = Y(t)s_{Z(t)} + a_{Z(t)}, \quad t \in [S_n, T_n - 1].$$

### 2.3. فرایند دریافت شده

ردیاب نوری و همچنین مدار بازخوانی، نویز را به فرآیند انتشار وارد می کند. از اندازه گیری نویز ما از ردیاب نوری، همبستگی متقابل بین سه اتصال نزدیک به صفر است و بنابراین برای محاسبه ناچیز است. ما این نویز را به عنوان فرآیند بردار گاوسی سفید  $\{N(t), t \geq 1\}$  با ماتریس کوواریانس مدل می کنیم.

$$\Lambda = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix},$$

که در آن  $\sigma_i^2$  میانگین توان نویز کانال اتصال ام است و برای هر پیکسل تخمین زده می شود. تابع چگالی احتمال مشترک بردار نویز  $N(t)$  در زمان  $t$  است.

$$f_N(n) = \frac{1}{\sqrt{(2\pi)^3 |\Lambda|}} e^{-\frac{1}{2} n' \Lambda^{-1} n} = \prod_{i=1}^3 \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{n_i^2}{2\sigma_i^2}},$$

که در آن  $|\Lambda|$  تعیین کننده ماتریس کوواریانس  $\Lambda$  است. فرآیند نویز  $\{N(t), t \geq 1\}$  مستقل از  $\{Z(t), t \geq 1\}$  SMSP و فرآیند محو شدن  $\{Y(t), t \geq 1\}$  فرض می شود.

ما فرض می کنیم که فرآیند نویز  $\{N(t), t \geq 1\}$  به فرآیند انتشار  $\{E(t), t \geq 1\}$  افزودنی است به طوری که فرآیند سیگنال دریافتی  $\{R(t), t \geq 1\}$  است

$$R(t) = E(t) + N(t), \quad t \geq 1 \quad (7)$$

$$= \begin{cases} a_d + N(t), & \text{if } Z(t) = d, \\ Y(t)s_A + a_A + N(t), & \text{if } Z(t) = A, \\ Y(t)s_T + a_T + N(t), & \text{if } Z(t) = T, \\ Y(t)s_G + a_G + N(t), & \text{if } Z(t) = G, \\ Y(t)s_C + a_C + N(t), & \text{if } Z(t) = C. \end{cases} \quad (8)$$

### 3- روش ها و الگوریتمها

در این بخش، تمرکز بر روی الگوریتم های رمزگشایی قرار می گیرد. به عنوان اولین مرحله، پارامترهای پیکسل با مقادیر پیش فرض یا مقادیر تخمین زده شده از داده های آزمایش تهیه می شوند. مرحله بعد، کل دنباله توسط الگوریتم رمزگشایی و پارامترهای پیکسل تشخیص داده می شود.

1. فاز اولیه: پارامترهای پیکسل  $s_A, s_T, s_G, s_C, a_A, a_T, a_G, a_C, a_d$  و

با مقادیر پیش فرض داده می شوند یا از داده های آزمایش تخمین زده می شوند.  $\sigma = (\sigma_1, \sigma_2, \sigma_3)$

2. فاز رمزگشایی:  $\{I(t), t \geq 1\}$  و  $\{\alpha(t), t \geq 1\}$  از فرآیند متناوب  $\{L(t), t \geq 1\}$  و

فرآیند جوش و خروش فلورسانس  $\{Y(t), t \geq 1\}$  توسط نوکلئوتید آموزشی شناخته شده

$\{x_n, n \geq 1\}$  و پارامترهای پیکسل تخمین زده شده که داده شده اند

$$\sigma, s_A, s_T, s_G, s_C, a_A, a_T, a_G, a_C, a_d,$$

### 3.1- تخمین پارامترهای پیکسل

پارامترهای پیکسل  $\sigma, s_A, s_T, s_G, s_C, a_A, a_T, a_G, a_C, a_d,$  را می توان از داده های اندازه

گیریهای آزمایشگاهی تخمین زد. از آنجایی که چهار رنگ فلورسانس دارای طول موج های متمایز هستند، می توان از LED با طول موج های یکسان برای شبیه سازی فلورسانس و تخمین پارامترهای پیکسل استفاده کرد. در حالی که شدت نور LED با پیشرفت زمان افزایش می یابد، مقادیر دریافتی از فوتودیود سه اتصالی، بر حسب واحد

ولتاژ، از طریق تبدیل فوتوالکتریک (PC) اندازه گیری می شوند. سپس، پارامترهای پیکسل،  $\sigma, s_A, s_T, s_G, s_C, a_A, a_T, a_G, a_C, a_d,$  و

از طریق یک رگرسیون خطی ساده در معادله (8) تخمین زده می شوند.  $\sigma$  با اندازه گیری در  $x \in \{A, T, G, C\}$

مدت زمان interpulse محاسبه می شود. علاوه بر این،  $R^2$  R-squared برای تعیین اینکه آیا رگرسیون خطی ساده به خوبی برازش دارد یا خیر محاسبه می شود.

### 3.2- مرحله رمزگشایی

فرض کنید که پارامترهای پیکسل تخمینی  $\sigma, s_A, s_T, s_G, s_C, a_A, a_T, a_G, a_C, a_d,$  داده شده

است. تابع درستنمایی  $f_{R|Z, Y}(r|z, \alpha)$  سیگنال دریافتی  $\{R(t), t \geq 1\}$ ، باشد، با توجه به اینکه  $\{r(t), t \geq 1\}$

$\{Z(t), t \geq 1\}$ ، SMSP، باشد و فرآیند انتشار نور  $\{Y(t), t \geq 1\}$  اگر باشد  $\{x(t), t \geq 1\}$  برابر

است با:

$$f_{\mathbf{R}|\mathbf{Z}, \mathbf{Y}}(\mathbf{r}|\mathbf{z}, \boldsymbol{\alpha}) = \prod_{t \geq 1} f_{\mathbf{R}(t)|\mathbf{Z}(t), \mathbf{Y}(t)}(\mathbf{r}(t)|\mathbf{z}(t), \boldsymbol{\alpha}(t))$$

از آنجایی که فرآیند نویز  $\{N(t), t \geq 1\}$  یک فرآیند گاوسی سفید است، که در آن

$$f_{\mathbf{R}(t)|\mathbf{Z}(t), \mathbf{Y}(t)}(\mathbf{r}(t)|\mathbf{z}(t), \boldsymbol{\alpha}(t)) = \begin{cases} \prod_{j=1}^3 \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{1}{2}\left(\frac{r_j(t)-a_{d,j}}{\sigma_j}\right)^2}, & \text{if } z(t)=d, \\ \prod_{j=1}^3 \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{1}{2}\left(\frac{r_j(t)-a_{x,j}-\alpha(t)s_{x,j}}{\sigma_j}\right)^2}, & \text{if } z(t)=x \in \{A, T, G, C\} \end{cases}$$

توجه داشته باشید که  $\alpha(t)=0$  از آنجایی که  $z(t)=d$ .

اکنون با توجه به پارامترهای پیکسل تخمینی  $s_A, s_T, s_G, s_C, a_A, a_T, a_G, a_C, a_d$  و  $\sigma$ ، نسخه های

رمزگشایی شده  $\{\hat{z}(t), t \geq 1\}$  و  $\{\hat{x}(t), t \geq 1\}$  از  $\{Z(t), t \geq 1\}$  و فرآیند انتشار نور  $\{Y(t), t \geq 1\}$  را می توان با روش ML به دست آورد،

$$(\hat{z}, \hat{\alpha}) = \arg \max_{(z, \alpha)} \sum_{t \geq 1} \ln f_{\mathbf{R}(t)|\mathbf{Z}(t), \mathbf{Y}(t)}(\mathbf{r}(t)|\mathbf{z}(t), \boldsymbol{\alpha}(t)),$$

where  $\ln f_{\mathbf{R}(t)|\mathbf{Z}(t), \mathbf{Y}(t)}(\mathbf{r}(t)|\mathbf{z}(t), \boldsymbol{\alpha}(t))$

$$= \begin{cases} \sum_{j=1}^3 \ln \frac{1}{\sqrt{2\pi}\sigma_j} - \frac{1}{2} \sum_{j=1}^3 \left( \frac{r_j(t)-a_{d,j}}{\sigma_j} \right)^2, & \text{if } z(t)=d, \\ \sum_{j=1}^3 \ln \frac{1}{\sqrt{2\pi}\sigma_j} - \frac{1}{2} \sum_{j=1}^3 \left( \frac{r_j(t)-a_{x,j}-\alpha(t)s_{x,j}}{\sigma_j} \right)^2, & \text{if } z(t)=x \in \{A, T, G, C\}. \end{cases}$$

از آنجایی که عبارت  $\sum_{j=1}^3 \ln \frac{1}{\sqrt{2\pi}\sigma_j}$  به فرآیند بیشینه سازی بی ربط است، متریک  $m(z(t), \alpha(t))$  را به صورت زیر تعریف می کنیم:

$$m(z(t), \alpha(t)) = \begin{cases} \sum_{j=1}^3 \left( \frac{r_j(t)-a_{d,j}}{\sigma_j} \right)^2, & \text{if } z(t)=d, \\ \sum_{j=1}^3 \left( \frac{r_j(t)-a_{x,j}-\alpha(t)s_{x,j}}{\sigma_j} \right)^2, & \text{if } z(t)=x \in \{A, T, G, C\}. \end{cases}$$

سپس نسخه رمزگشایی شده ML از  $\{Z(t), t \geq 1\}$  و  $\{Y(t), t \geq 1\}$  برابر است با:

$$\begin{aligned}
(\hat{z}, \hat{\alpha}) &= \arg \min_{(z, \alpha)} \sum_{t \geq 1} m(z(t), \alpha(t)) \\
&= \arg \min_z \sum_{t \geq 1} \min_{\alpha(t)} m(z(t), \alpha(t)).
\end{aligned}$$

متعاقباً، با توجه به SMSP فرضی  $\{Z(t), t \geq 1\}$ ، به حداقل رساندن مجموع  $\sum_{t \geq 1} m(z(t), \alpha(t))$  متریک

بیش از شدت انتشار فرآیند  $m(z(t), \alpha(t))$  را می توان با به حداقل رساندن معیارهای  $\{\alpha(t), t \geq 1\}$ ،  
روی شدت  $\alpha(t)$  در هر زمان  $t$  با  $z(t)$  داده شده انجام داد.

$$z(t) \in \{A, T, G, C\}, \alpha^*(t|z(t)) = \arg \min_{\alpha(t)} m(z(t), \alpha(t)).$$

وقتی پس ما داریم

$$\alpha^*(t|z(t)) = Q(I_{x, \min}, \alpha^\#(t|z(t)), I_{x, \max}), \forall z(t) = x \in \{A, T, G, C\},$$

where for  $a < c$ ,

$$Q(a, b, c) = \begin{cases} a, & \text{if } b < a, \\ b, & \text{if } a \leq b \leq c, \\ c, & \text{if } b > c, \end{cases}$$

and

$$\alpha^\#(t|z(t)) = \frac{\sum_{j=1}^3 \left( \frac{r_j(t) - a_{s,j}}{\sigma_j} \right) (s_{x,j} / \sigma_j)}{\sum_{j=1}^3 (s_{x,j} / \sigma_j)^2}, \text{ if } z(t) = x \in \{A, T, G, C\}$$

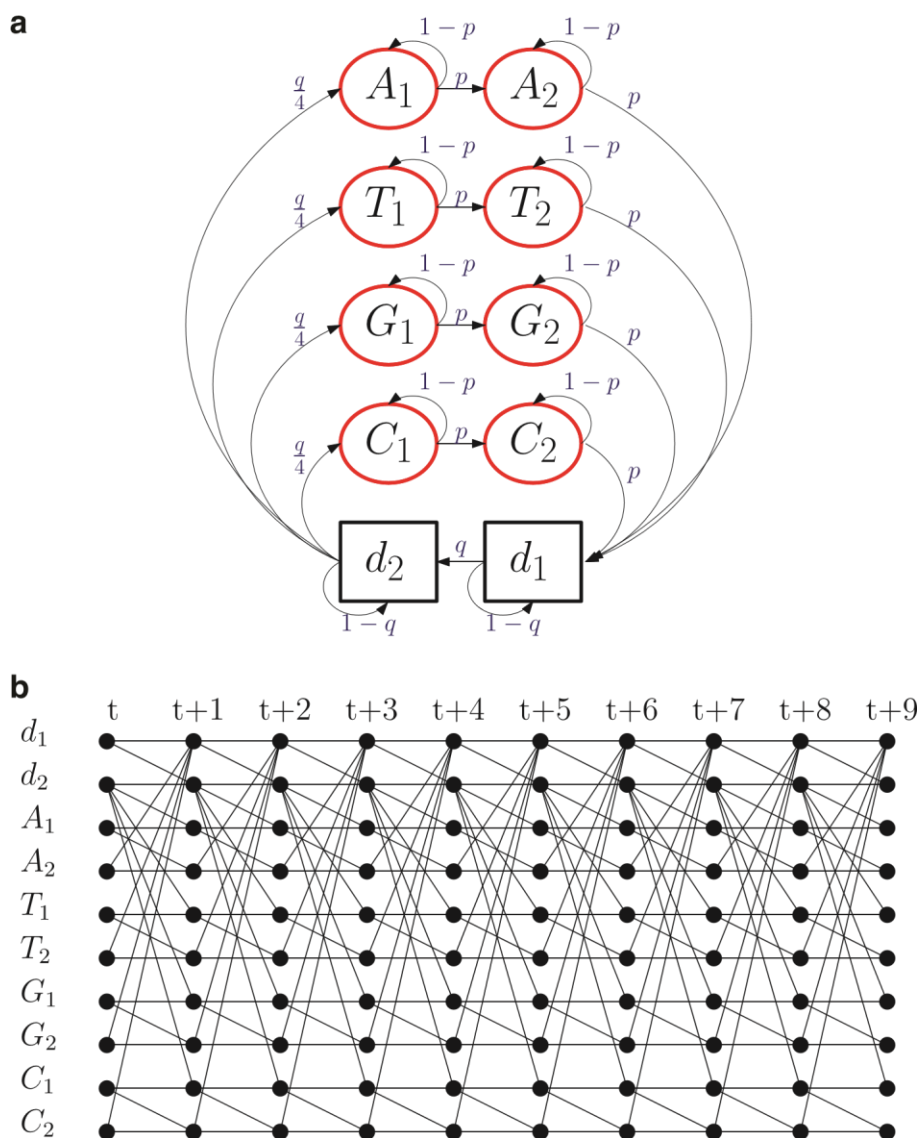
با انجام کمینه سازی متریک  $m(z(t), \alpha(t))$  روی تمام  $\alpha(t) \in \mathbb{R}$ .

$$m^*(z(t)) = \begin{cases} \sum_{j=1}^3 \left( \frac{r_j(t) - a_{d,j}}{\sigma_j} \right)^2, & \text{if } z(t) = d, \\ \sum_{j=1}^3 \left( \left( \frac{r_j(t) - a_{s,j}}{\sigma_j} \right) - \alpha^*(t|z(t)) \left( \frac{s_{x,j}}{\sigma_j} \right) \right)^2, & \text{if } z(t) = x \in \{A, T, G, C\}. \end{cases} \quad (9)$$

اکنون نسخه های رمزگشایی شده  $\{Z(t), t \geq 1\}$  و  $\{\alpha(t), t \geq 1\}$  از SMSP و فرآیند شدت فلورسانس  $\{\Upsilon(t), t \geq 1\}$  برابرند با:

$$(\hat{z}, \hat{\alpha}) = \arg \min_{(z, \alpha^*(t|z(t)))} \sum_{t \geq 1} m^*(z(t)). \quad (10)$$

ما الگوریتم Viterbi (Viterbi, 1967) را با نمودار انتقال حالت (شکل a4) و نمودار ترالیس مربوطه در شکل b4 بکار می‌بریم تا مساله کمینه‌سازی در معادله (10) را حل کنیم.



شکل 4. (a) نمودار انتقال حالت زنجیره مارکوف  $\{L(t), t \geq 1\}$  در مثال نشان داده شده،  $l=2$  و  $r=2$ . (b) نمونه

ای از نمودار ترالیس با نمودار انتقال حالت در (a) مطابقت دارد.

### 3.3- عملکرد اشتباه با وجود اطلاعات کامل



انتخاب رنگ یکی از مهم ترین عوامل در روش برچسب گذاری فلورسانس است. ترکیب های مختلف انتخاب رنگ بر عملکرد تشخیص سیگنال فلورسانس، به ویژه در فرآیند توالی یابی تأثیر می گذارد. با مدل ایجاد شده در این مطالعه، ما حد بالایی احتمال تشخیص اشتباه هر ترکیب متمایز فلورسانس را محاسبه می کنیم، که به انتخاب گروه بهینه رنگ ها کمک می کند. ویژگی های اصلی رنگ های فلورسانس حاکم بر تشخیص سیگنال از طریق مدل توضیح داده شده است.

در سناریوی ایده آل، فرض کنید منبع شدت نور محیط در حالت ثابت است، به طوری که سیگنال شدت نور محیط را می توان با یک  $a_d$  بردار سیگنال ثابت مدل کرد. همچنین ضریب محو شدن  $\bar{Y}(t) = Y_n, \forall t \in [\bar{s}_n, t_n - 1]$  را فرض کنید.

با اطلاعات کامل، یعنی

احتمال خطا برابر است با:  $\hat{a}_x = a_d, \hat{s}_x = s_x, \hat{\sigma}_i^2 = \sigma_i^2, \hat{\alpha}_n = \alpha_n, \text{ and } \hat{s}_n = s_n, \hat{t}_n = t_n \forall x \in \{A, T, G, C\},$

$$\begin{aligned}
P(X_n \neq \widehat{X}_n) &= \sum_{x \in \{A, T, G, C\}} P(X_n = x) P(\widehat{X}_n \neq x | X_n = x) \\
&= \frac{1}{4} \sum_{x \in \{A, T, G, C\}} P(\widehat{X}_n \neq x | X_n = x) \\
&= \frac{1}{4} \sum_{x \in \{A, T, G, C\}} P(\cup_{y \in \{A, T, G, C\}, y \neq x} F_{yx} | X_n = x) \\
&\leq \frac{1}{4} \sum_{x \in \{A, T, G, C\}} \sum_{y \in \{A, T, G, C\}, y \neq x} P(F_{yx} | X_n = x),
\end{aligned}$$

where  $F_{yx}$  is the event that

$$\sum_{k=s_n}^{t_n-1} (\mathbf{r}(k) - \alpha_n \mathbf{s}_y - \mathbf{a}_d)^t \Lambda^{-1} (\mathbf{r}(k) - \alpha_n \mathbf{s}_y - \mathbf{a}_d) \leq \sum_{k=s_n}^{t_n-1} (\mathbf{r}(k) - \alpha_n \mathbf{s}_x - \mathbf{a}_d)^t \Lambda^{-1} (\mathbf{r}(k) - \alpha_n \mathbf{s}_x - \mathbf{a}_d).$$

Given  $X_n = x$ , we have

$$\mathbf{r}(k) = \alpha_n \mathbf{s}_x + \mathbf{a}_d + \mathbf{n}(k), \quad k \in [s_n, t_n - 1],$$

so that

$$\mathbf{r}(k) - \alpha_n \mathbf{s}_y - \mathbf{a}_d = \alpha_n (\mathbf{s}_x - \mathbf{s}_y) + \mathbf{n}(k), \quad k \in [s_n, t_n - 1],$$

and

$$\mathbf{r}(k) - \alpha_n \mathbf{s}_x - \mathbf{a}_d = \mathbf{n}(k), \quad k \in [s_n, t_n - 1].$$

Thus, the event  $F_{yx}$  is

$$\begin{aligned}
\sum_{k=s_n}^{t_n-1} (\alpha_n (\mathbf{s}_x - \mathbf{s}_y) + \mathbf{n}(k))^t \Lambda^{-1} (\alpha_n (\mathbf{s}_x - \mathbf{s}_y) + \mathbf{n}(k)) &\leq \sum_{k=s_n}^{t_n-1} \mathbf{n}(k)^t \Lambda^{-1} \mathbf{n}(k) \\
&\Downarrow \\
\frac{\alpha_n}{2} (\mathbf{s}_y - \mathbf{s}_x)^t \Lambda^{-1} (\mathbf{s}_y - \mathbf{s}_x) &\leq (\mathbf{s}_y - \mathbf{s}_x)^t \Lambda^{-1} \frac{\sum_{k=s_n}^{t_n-1} \mathbf{n}(k)}{t_n - s_n} \\
&\Downarrow \\
\frac{\alpha_n}{2} \sum_{i=1}^3 \left( \frac{s_{y,i} - s_{x,i}}{\sigma_i} \right)^2 &\leq \sum_{i=1}^3 \left( \frac{s_{y,i} - s_{x,i}}{\sigma_i} \right) \left( \frac{1}{t_n - s_n} \sum_{k=s_n}^{t_n-1} \frac{n_i(k)}{\sigma_i} \right).
\end{aligned}$$

$$\frac{1}{t_n - s_n} \sum_{k=s_n}^{t_n-1} \frac{N_i(k)}{\sigma_i}, \quad i=1, 2, 3,$$

از آنجایی که نویزهای مقیاس شده متوسط

مستقل هم میباشند، نویز ترکیب شده برابر است با:

$$N = \sum_{i=1}^3 \left( \frac{s_{y,i} - s_{x,i}}{\sigma_i} \right) \left( \frac{1}{t_n - s_n} \sum_{k=s_n}^{t_n-1} \frac{N_i(k)}{\sigma_i} \right)$$

$$\sigma^2 = \frac{1}{t_n - s_n} \sum_{i=1}^3 \left( \frac{s_{y,i} - s_{x,i}}{\sigma_i} \right)^2 \quad N \left( 0; \frac{1}{t_n - s_n} \sum_{i=1}^3 \left( \frac{s_{y,i} - s_{x,i}}{\sigma_i} \right)^2 \right)$$

نرمال با میانگین و واریانس زیر که دارای توزیع \*

بنابراین احتمال شرطی  $P(F_{yx}|X_n=x)$  برابر است با :

$$\begin{aligned} P(F_{yx}|X_n=x) &= P\left(N \geq \frac{\alpha_n}{2} \sum_{i=1}^3 \left(\frac{s_{y,i} - s_{x,i}}{\sigma_i}\right)^2\right) \\ &= P\left(N \geq \frac{\alpha_n}{2} (t_n - s_n) \sigma^2\right) \\ &= P\left(\frac{N}{\sigma} \geq \frac{\alpha_n}{2} (t_n - s_n) \sigma\right) \\ &= Q\left(\frac{\alpha_n}{2} (t_n - s_n) \sigma\right), \end{aligned}$$

بطوریکه تابع  $Q$  به این شکل تعریف شده باشد :

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt, \quad \forall t > 0.$$

تابع  $Q$  دارای کران بالا و پایین نزدیک به همی دارد

$$\frac{1}{\sqrt{2\pi x}} \left(1 - \frac{1}{x^2}\right) e^{-\frac{x^2}{2}} < Q(x) < \frac{1}{\sqrt{2\pi x}} e^{-\frac{x^2}{2}}, \quad \forall x > 0$$

و یک کران بالای مناسب :

$$Q(x) < \frac{1}{2} e^{-\frac{x^2}{2}}, \quad \forall x > 0.$$

حالا یک کران بالا برای تابع احتمال  $P(F_{yx}|X_n=x)$  وجود دارد

$$P(F_{yx}|X_n=x) < \frac{1}{2} \exp\left\{-\frac{\alpha_n^2 (t_n - s_n)^2 \sigma^2}{8}\right\} = \frac{1}{2} \exp\left\{-\frac{\alpha_n^2 (t_n - s_n)}{8} \sum_{i=1}^3 \left(\frac{s_{y,i} - s_{x,i}}{\sigma_i}\right)^2\right\}$$

بنابراین کران بالای احتمال خطا نیز به این روش محاسبه میشود :

$$\begin{aligned} P(X_n \neq \widehat{X}_n) &\leq \frac{1}{4} \sum_{x, y \in \{A, T, G, C\}, y \neq x} P(F_{yx}|X_n=x) \\ &< \frac{3}{2} \exp\left\{-\frac{\alpha_n^2 (t_n - s_n)}{8} \min_{x \neq y} \sum_{i=1}^3 \left(\frac{s_{y,i} - s_{x,i}}{\sigma_i}\right)^2\right\} \end{aligned} \quad (11)$$

$$-\frac{\alpha_n^2(t_n - s_n)}{8} \min_{x \neq y} \sum_{i=1}^3 \left( \frac{s_{y,i} - s_{x,i}}{\sigma_i} \right)^2$$

با در نظر گرفتن عبارت، ، در رابطه (11)، چهار

عامل بر نتایج تشخیص سیگنال فلورسانس تأثیر می گذارد. an با توان دو بیان را تحت تأثیر قرار می دهد، در حالی که دوره زمانی n ادغام، tn - sn، با توان یک است. حداقل فاصله بین طول موج رنگهای متمایز که با

نشان داده شده است با توان یک است و توان نوین نیز با توان یک است. در

همان محیط و سیستم، مدت زمان ادغام و به طور خاص حداقل فاصله بین رنگ های فلورسانس بر عملکرد تأثیر می گذارد در حالی که دانشمندان مایلند ترکیبی از رنگ ها را انتخاب کنند.

جدول 1. بردارهای تبدیل فوتوالکتریک،  $s_x$  و  $a_x$ ، از اندازه گیری طول موج های متمایز انتشار نور LED محاسبه می شوند.

Wavelength (nm)	RGB PC vector $s_x$ (ms·mv/photon)	RGB PC vector $a_x$ (mv)
530	(0.153, 0.063, 0.025)	(1.390, 1.457, 1.105)
590	(0.098, 0.063, 0.059)	(1.915, 1.456, 0.924)
625	(0.078, 0.057, 0.076)	(1.856, 1.416, 0.891)
656	(0.069, 0.054, 0.087)	(1.595, 1.193, 0.527)

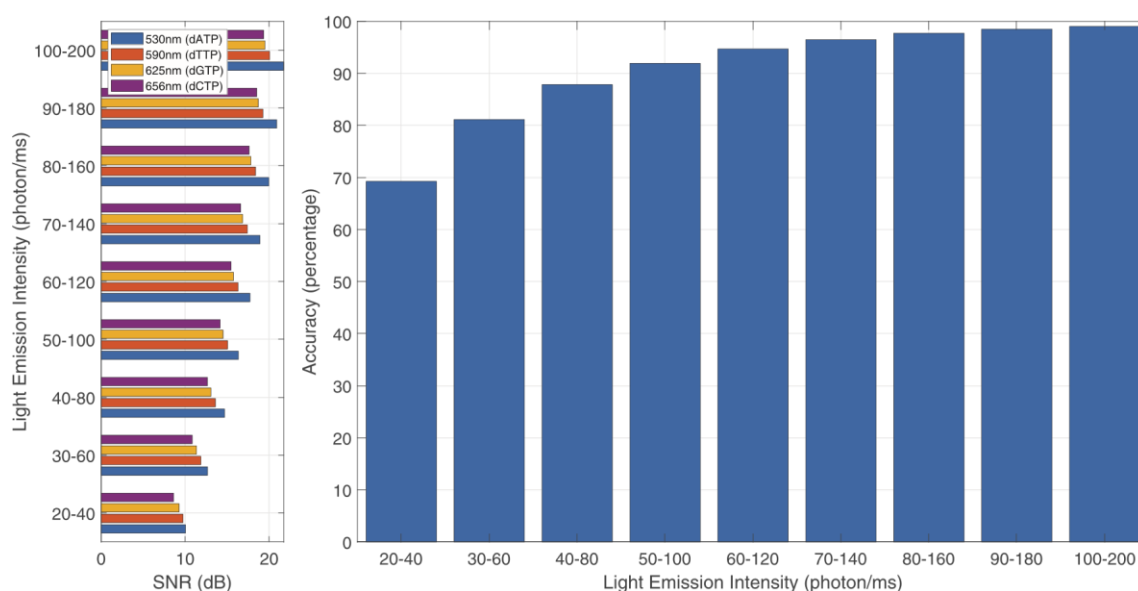
PC, photoelectric conversion; RGB.

#### 4- شبیه سازیها

داده های اندازه گیری شده برای استفاده از چهار LED تک طول موج (530، 590، 625، و 656 نانومتر) برای شبیه سازی اثر فلورسانس بر روی نوکلئوتیدها (dGTP، dTTP، dATP، و dCTP)، که در آن 530 نانومتر LED نشان دهنده انتشار pholinks برای phoATPphoATP است. و به همین ترتیب ... فرض کنید که میانگین عرض پالس و مدت زمان بین پالس به ترتیب 100 و 200 میلی ثانیه باشد و پارامترهای p و q به ترتیب از معادلات (1) و (2) تخمین زده شوند. از آنجایی که عرض پالس و مدت زمان میان پالس در مقایسه با مطالعه قبلی کوتاهتر شده است، الگوی ساطع نور نیز محدودتر خواهد بود، به این معنی که محیط های داخلی و خارجی آزمایش ها اجازه انعطاف پذیری بیشتری دارند. جمع آوری داده ها شامل روش های نمونه گیری است که در ادامه شرح داده شده است. برای مقایسه داده های شبیه سازی شده با مطالعه قبلی (Eid و همکاران، 2009)، ما داده هایی را با درجات

مختلف شدت نور برای دستیابی به SNR های مختلف ایجاد کردیم تا به عنوان معیاری برای مقایسه استفاده شود. شدت انتشار نور LED از مسیرهای تصادفی در 9 حوزه محدود پیروی می کند: [20,40]، [30,60]، [40,80]، [50,100]، [60,120]، [70,140]، [80,160]، [90,180] و [100,200] (photon/ms). ما 100 نمونه برای هر الگو و 1000 نوکلئوتید تصادفی برای هر نمونه تولید می کنیم. برای هر نوکلئوتید، فرآیند انتشار توسط مدل پیاده روی تصادفی تولید می شود و SMSP مطابق نمودار انتقال در شکل 2 تولید می شود.

داده های دریافتی از فرآیند انتشار LED به دست می آیند به اینصورت که فتودیود، نور LED را دریافت می کند که این نور شامل ولتاژهای تبدیل شده از انتشار نور و به دست آمده از اثر فوتوالکتریک است.



شکل 5. نمودار سمت چپ SNR مربوطه را نشان می دهد که با سطوح مختلف شدت چهار فلورسانس مجزا محاسبه شده است. نمودار اصلی دقت روش رمزگشایی ما را با شدت انتشار نور در نوسان متفاوت نشان می دهد. SNR، نسبت سیگنال به نویز.

جدول 2. مقایسه عملکرد بین پلتفرم PacBio و تشخیص ما

معماری (الگوریتم حداکثر احتمال-ویتری در یک فتودیود سه اتصالی)

<i>Platform: PacBio</i>		<i>SNR: 22–30 dB</i>		
	<i>A555-dATP</i>	<i>A568-dTTP</i>	<i>A647-dGTP</i>	<i>A660-dCTP</i>
Pulse width (ms)	133 ± 22	91 ± 13	117 ± 14	96 ± 10
Interpulse duration (ms)	770 ± 250	670 ± 220	960 ± 210	790 ± 230
	Correct	Mismatches	Insertions	Deletions
Performance (%)	82.9	4.4	5.1	7.6
<i>ML-Viterbi algorithm</i>		<i>SNR: 14.16–16.32 dB</i>		
	<i>530-dATP</i>	<i>590-dTTP</i>	<i>625-dGTP</i>	<i>656-dCTP</i>
Pulse width (ms)	100 ± 5.77	100 ± 5.77	100 ± 5.77	100 ± 5.77
Interpulse duration (ms)	200 ± 18.26	200 ± 18.26	200 ± 18.26	200 ± 18.26
	Correct	Mismatches	Insertions	Deletions
Performance (%)	91.94	8.06	0	0
<i>ML-Viterbi algorithm</i>		<i>SNR: 17.6–19.5 dB</i>		
	<i>530-dATP</i>	<i>590-dTTP</i>	<i>625-dGTP</i>	<i>656-dCTP</i>
Pulse width (ms)	100 ± 5.77	100 ± 5.77	100 ± 5.77	100 ± 5.77
Interpulse duration (ms)	200 ± 18.26	200 ± 18.26	200 ± 18.26	200 ± 18.26
	Correct	Mismatches	Insertions	Deletions
Performance (%)	97.72	2.28	0	0

در حالی که شدت نور LED با پیشرفت زمان افزایش می یابد، ما مقادیر RGB دریافتی را از طریق PC اندازه گیری می کنیم و این رابطه در جدول 1 نشان داده شده است. مقادیر PC متمایز از مطالعه قبلی ما است (چن و لو، 2019) زیرا در محیط های مختلف مانند رنگ های متمایز و دماهای متفاوت مقادیر مختلف PC را دریافت خواهیم کرد. میانگین SNR مربوطه از این سطوح مختلف شدت نور همانطور که در نمودار سمت چپ شکل 5 محاسبه شده است.

## 5-نتایج

نتیجه تخمین توالی های نوکلئوتیدی از روش تشخیص ما به نام الگوریتم ML-Viterbi بدست آمده است. ما الگوریتم اسمیت-واترمن را برای تطابق بین توالی های مورد انتظار و توالی های اندازه گیری شده اعمال می کنیم. تجزیه و تحلیل عمیق داده ها در شکل 5 و جدول 2 به تفصیل نشان داده شده است.

جدا از بحث مدل سازی در بالا، دقت توالی یابی حتی با خروجی شدت انتشار نور در نوسان بین 80 تا 160 فوتون بر میلی ثانیه و SNR کمتر از 19 دسی بل، فراتر از 97٪ است. علاوه بر این، حذف و درج، به جز عدم تطابق، توالی در سناریوی SNR بیش از 15 دسی بل رخ نخواهد داد. در مقایسه، شرح داده شده از مطالعه قبلی در Eid و

همکاران. (2009)، یعنی پلت فرم PacBio، نسبت سیگنال به نویز  $10 \pm 24 \sim$  است که از 22 تا 30 دسی بل متغیر است. دقت تقریباً  $>85\%$  بدون پشتیبانی از روش توالی یابی اجماعی با SNR بالاتر است. تفاوت‌ها این اهمیت را نشان می‌دهد که الگوریتم ML-Viterbi در تشخیص سیگنال حتی تحت یک محیط نامطلوب با SNR کم، درخشان عمل میکند.

## 6. بحث و نتیجه گیری

با فرض شرایط عینی یکسان برای دو روش تشخیص، از نظر سرعت پردازش تشخیص، روش تشخیص تک فتودیودی بر فناوری پردازش تصویر فعلی (ورودی‌های چند پیکسلی از دوربین CCD) به دلیل پردازش تک پیکسلی فائق می‌آید.

به طور خاص، روشی که در این مقاله به عنوان تکنیک برچسب گذاری فلورسانس با یک فتودیود منفرد اعمال می‌شود، عملکرد بهتری را، به ویژه در توالی‌یابی DNA، با دقت زیاد (نرخ خطا کمتر)، توان عملیاتی بالاتر و هزینه بالقوه کمتر ارائه می‌دهد. با ترکیب روش توالی یابی اجماع، میزان دقت داده‌ها را می‌توان به طور خوش بینانه تا تقریباً بدون خطا بهبود داد (هیات و همکاران، 2010). علاوه بر این، مدل سازی بر روی انتخاب رنگ فلورسنت و ترکیب بهینه، زمان و منابع مورد نیاز را در مقایسه با آزمایش واقعی کاهش می‌دهد. با این حال، سطح توان عملیاتی بسیار وابسته به توانایی محاسباتی پیشرفته در توالی یابی همزمان گروه‌های فتودیود منفرد است. علاوه بر این، دقت تشخیص ممکن است با افزایش تفاوت بین ویژگی‌های فلورسانس برای برچسب زدن چهار نوع dNTP افزایش یابد. به طور خلاصه، الگوریتم توضیح داده شده در این مقاله قطعاً کمک بسیار خوبی برای توسعه توالی یابی مبتنی بر فلورسانس است.