# Supercolor Coding Methods for Large-Scale Multiplexing of Biochemical Assays 

Aditya Rajagopal,,$^{*}{ }^{\dagger}$ Axel Scherer, ${ }^{\dagger}$ Andrew Homyk, ${ }^{\dagger}$ and Emil Kartalov ${ }^{\dagger}{ }^{\dagger}{ }^{\dagger}$<br>${ }^{\dagger}$ California Institute of Technology, Department of Electrical Engineering, MC 200-36, 1200 E. California Blvd., Pasadena, California 91125, United States<br>${ }^{\ddagger}$ Pathology Department, Keck School of Medicine, University of Southern California, 2011 Zonal Ave, HMR301A, Los Angeles, California 90089-9092, United States

## (S) Supporting Information


#### Abstract

We present a novel method for the encoding and decoding of multiplexed biochemical assays. The method enables a theoretically unlimited number of independent targets to be detected and uniquely identified in any combination in the same sample. For example, the method offers easy access to 12-plex and larger PCR assays, as contrasted to the current 4-plex assays. This advancement would allow for large panels of tests to be run simultaneously in the same sample, saving reagents, time, consumables, and manual labor, while also avoiding the traditional loss of sensitivity due to sample aliquoting. Thus, the presented method is a major technological breakthrough with farreaching impact on biotechnology, biomedical science, and clinical diagnostics. Herein, we present the mathematical theory behind the method as well as its experimental proof of principle using Taqman PCR on sequences specific to infectious diseases.



fluorophores, this has meant that only 4 to 6 questions are typically asked of the same sample at the same time. ${ }^{5,6}$ Wherever higher factors of multiplexing are required, the solution has been to combine fluorescence with aliquoting, spatial arraying, or sequential processing. ${ }^{7}$ However, those lead to labor-intensive workflow and/or significantly more expensive and bulkier systems. Thus, it would be highly beneficial if the multiplexing factor were increased through spectral coding only.

Early attempts at "supercolor" multiplexing $(M>N)$ have been only partially successful. "Chromosome painting" ${ }^{8}$ allowed for FISH-based identification of 24 chromosomes, while two PCR Taqman-based schemes could identify eight foodborne pathogens ${ }^{9}$ or fifteen types of HPV. ${ }^{10}$ However, the first scheme made use of the application-specific condition that each chromosome was also spatially resolved from the others, while the other two schemes assumed that no more than one of the targets is present at a time. If these restrictions are relaxed, all three schemes produce "degeneracy", i.e., the same test result decodes into more-than-one possible outcomes. A general multiplexing scheme (e.g., coding for panels of infectious

[^0]diseases, drug-resistant bacterial strains, foodborne pathogens, or genetic markers) cannot allow ambiguity of outcome and cannot assume that only one target is present at a time.
Herein, we report on a novel coding method that is the first to solve the degeneracy problem and thus the first to ensure the unambiguous identification of any combination of present targets. Also, its multiplexing factor $M$ is mathematically unlimited and is practically much larger than the number of available colors $(M \gg N)$. Thus, our method is a major breakthrough in the field.
As the method uses standard fluorophores and oligos, it is easily implemented with the existing oligo-synthesis techniques and measurement infrastructure. Once applied in practice, the method would produce a major increase of the productivity of the existing fleet of machines (e.g., for PCR, qPCR, digital PCR) as no changes in the hardware would be needed to accommodate the new multiplexing capabilities.
In clinical diagnostics, splitting a profuse sample into aliquots and testing each aliquot for a different target is a valid but expensive and labor-intensive multiplexing strategy. ${ }^{11-19}$ However, the amount of sample is rather limited in point-ofcare diagnostics. ${ }^{20}$ As a result, any splitting is undesirable, because the resulting amount of analyte in each aliquot may fall below the detection limit of the assay. In contrast, our method would pose multiplexed questions to the whole sample, thus avoiding aliquots and the loss in sensitivity. Hence, our method is particularly significant to point-of-care PCR assays, where it would also allow for cheap disposables to test for large panels of infectious diseases and multiple strains of drug-resistant bacteria. The low cost of its implementation makes it even more important to diagnostics in low-resource settings.

## - RESULTS

The crux of the presented method is the combined use of "colors", the multiplicity of signal intensity, and mathematical strategies to circumvent degeneracy and ensure an infinite number of unique codes that can be unambiguously decoded in any combination of occurrence. While the method is not limited to fluorescence, PCR, and Taqman probes, ${ }^{21}$ they are the natural choice as a system for proof of principle for the method.
Taqman probes are short pieces of DNA that have a fluorophore on one end and a quencher on the other end. If the target sequence is present, the probe and primer hybridize to it. As the polymerase extends the primer, its $5^{\prime}$ exonuclease activity breaks up the probe sequence. The released fluorophore and quencher separate by diffusion, and so the fluorophore emits fluorescence signal. Conversely, if the target sequence is not present, the probe remains intact, so the quencher prevents fluorescence emission. The end result is a strong fluorescence signal when the target is present and a weak or no signal when the target is not present.

Standard fluorophores have wide emission spectra, so to avoid overlap and false positives, only a few colors (typically four) are used simultaneously in multiplexed assays. Each color is assigned to a different Taqman probe. Then, the presence or absence of a sequence is judged by the fluorescence signal in the respective color. For example, if the color set is named \{blue, green, yellow, red\} by excitation, a traditional experimental result of 1100 means the sequences coded by blue- and green-excited fluorophores are present, while the other two are absent.

Unlike the traditional approach, the presented method generally uses more than one color per sequence. Table 1

Table 1. Combinatorial Coding of Targets

| Sequence | $\mathbf{B}$ | $\mathbf{G}$ | $\mathbf{Y}$ | $\mathbf{R}$ |
| :--- | :--- | :--- | :--- | :--- |
| Control Z | 1 | 0 | 0 | 0 |
| Sequence A | 1 | 0 | 0 | 1 |
| Sequence B | 1 | 0 | 1 | 0 |
| Sequence C | 1 | 1 | 0 | 0 |
| Sequence D | 1 | 0 | 1 | 1 |
| Sequence E | 1 | 1 | 0 | 1 |
| Sequence F | 1 | 1 | 1 | 0 |
| Sequence G | 1 | 1 | 1 | 1 |

shows one such coding scheme in four colors. Code 1000 is assigned to a control sequence Z that will always be present and should always amplify. ${ }^{22}$ All sequences have a probe in the control color. This design ensures that the multiplicity of the signal intensity in the "control" color reports the number of unique sequences that have successfully extended and are thus present in the sample.

Mathematical symmetry dictates that any color can be the control color. However, as the highest multiplicity would likely be observed in the control color, it makes sense to choose it to be the color that is best detected in the particular system. The Roche Lightcycler 480 used for the experimental proof has only a blue excitation, while the other colors count on wide excitation tails. Hence, blue was the logical choice here.

The possible measurement results and their decoded meaning in terms of sequences present are shown in Table 2. Three conditions have been stipulated: First, the positive control always produces a positive outcome, i.e., it works appropriately. Second, in each color, the signal is additive and scales in the same way with probe concentration, regardless of which probe it comes from. Third, in each color, each probe produces the same unit of signal. Essentially, this means the signals are additive and digital. Ensuring the first condition is just a matter of proper preparation of the assay. The second condition is attainable under fluorescence. In practice, the third condition need only be approximately right, as we show further below.

A measurement outcome is denoted as a sequence of signal multiplicities in the respective "colors". Each multiplicity is calculated within its own color. For example, outcome 4321 means the sample's measured intensities are $4 \times$ in blue, $3 \times$ in green, $2 \times$ in yellow, and $1 \times$ in red. This outcome is valid, because it can be achieved by adding sequence codes from Table 1, so it is found in the decoding Table 2. It produces the answer ZCFG, which means only sequences Z, C, F, and G are present. Conversely, the outcome 4000 is invalid, because it cannot be achieved by adding sequence codes from Table 1 in a digital fashion and thus is not listed in Table 2. An invalid outcome means that the assay malfunctioned in the particular experiment. This logic filter provides a powerful tool for judging the validity of experimental results.

Table 2 shows all the valid answers, and thus, it is the exhaustive set for the encoding Table 1. In Table 2, "Rank" is defined as the number of present sequences, which under the encoding Table 1 , is equal to the multiplicity of the blue bin. The lowest rank is 1 , with a single valid outcome of 1000 and a test result of Z. This means only the control sequence was present. At rank 2, the control sequence and one other are present, so there are 7 such cases. At rank 3, the control sequence and two other sequences are present, so the number

Table 2. Exhaustive Combinations of Code

| B | G | Y | R | Read |
| :---: | :---: | :---: | :---: | :---: |
| 1 | 0 | 0 | 0 | Z |
| 2 | 0 | 0 | 1 | ZA |
| 2 | 0 | 1 | 0 | ZB |
| 2 | 1 | 0 | 0 | ZC |
| 2 | 0 | 1 | 1 | ZD |
| 2 | 1 | 0 | 1 | ZE |
| 2 | 1 | 1 | 0 | ZF |
| 2 | 1 | 1 | 1 | ZG |
| 3 | 0 | 1 | 1 | ZAB |
| 3 | 1 | 0 | 1 | ZAC |
| 3 | 0 | 1 | 2 | ZAD |
| 3 | 1 | 0 | 2 | ZAE |
| 3 | 1 | 1 | 1 | ZAF, ZCD, ZBE |
| 3 | 1 | 1 | 2 | ZAG, ZDE |
| 3 | 1 | 1 | 0 | ZBC |
| 3 | 0 | 2 | 1 | ZBD |
| 3 | 1 | 2 | 0 | ZBF |
| 3 | 1 | 2 | 1 | ZBG, ZDF |
| 3 | 2 | 0 | 1 | ZCE |
| 3 | 2 | 1 | 0 | ZCF |
| 3 | 2 | 1 | 1 | ZCG, ZEF |
| 3 | 1 | 2 | 2 | ZDG |
| 3 | 2 | 1 | 2 | ZEG |
| 3 | 2 | 2 | 1 | ZFG |
| 4 | 1 | 1 | 1 | ZABC |
| 4 | 0 | 2 | 2 | ZABD |
| 4 | 2 | 2 | 0 | ZBCF |
| 4 | 2 | 0 | 2 | ZACE |
| 4 | 1 | 1 | 2 | ZABE, ZACD |
| 4 | 1 | 2 | 1 | ZABF, ZBCD |
| 4 | 2 | 1 | 1 | ZACF, ZBCE |
| 4 | 1 | 2 | 2 | ZADF, ZABG, ZBDE |
| 4 | 2 | 1 | 2 | ZACG, ZCDE, ZAEF |
| 4 | 2 | 2 | 1 | ZBCG, ZBEF, ZCDF |
| 4 | 2 | 2 | 2 | $\begin{aligned} & \text { ZCDG, ZAFG, ZBEG, } \\ & \text { ZDEF } \end{aligned}$ |
| 4 | 1 | 1 | 3 | ZADE |
| 4 | 1 | 3 | 1 | ZBDF |
| 4 | 3 | 1 | 1 | ZCEF |
| 4 | 1 | 2 | 3 | ZADG |
| 4 | 3 | 2 | 1 | ZCFG |
| 4 | 2 | 1 | 3 | ZAEG |
| 4 | 2 | 3 | 1 | ZBFG |
| 4 | 1 | 3 | 2 | ZBDG |
| 4 | 3 | 1 | 2 | ZCEG |
| 4 | 2 | 2 | 3 | ZDEG |
| 4 | 3 | 2 | 2 | ZEFG |
|  |  |  |  |  |

of distinct results is a combination of 7 choose 2 , or $7!/(5!*$ $2!)=21$. At ranks $4,5,6,7$, and 8 , the number of distinct results is $35,35,21,7$, and 1 , respectively. Table 2 shows the same numbers of results in each rank, so the table is exhaustive.

In "supercolor" multiplexing, degeneracy is the phenomenon wherein the same valid outcome corresponds to multiple distinct combinations of present sequences. ${ }^{9,10}$ For example, outcome 5233 in Table 2 is degenerate, because it can be decoded as either ZADFG or ZBDEG. As a counterexample, outcome 3110 is not degenerate because it is decoded to a single result (ZBC).
It turns out degeneracy can be eliminated. For example, outcome 4112 is valid and can be decoded as either ZABE or ZACD. So, if sequence $D$ is dropped from the set, the test result ZACD is no longer possible, so the only remaining test result associated with outcome 4112 is ZABE. Similarly, it turns out that dropping any two of $\{\mathrm{D}, \mathrm{E}, \mathrm{F}\}$ completely eliminates degeneracy, while the coding remains "supercolor" $(M>N)$ as 6 sequences are unambiguously identifiable in any combination while using just 4 colors.

| $\mathbf{B}$ | $\mathbf{G}$ | $\mathbf{Y}$ | $\mathbf{R}$ | Read |
| :--- | :--- | :--- | :--- | :--- |
| 5 | 1 | 2 | 2 | ZABCD |
| 5 | 2 | 1 | 2 | ZABCE |
| 5 | 2 | 2 | 1 | ZABCF |
| 5 | 2 | 2 | 2 | ZABCG, ZABEF, ZBCDE, ZACDF |
| 5 | 1 | 2 | 3 | ZABDE |
| 5 | 2 | 1 | 3 | ZACDE |
| 5 | 1 | 3 | 2 | ZABDF |
| 5 | 2 | 3 | 1 | ZBCDF |
| 5 | 3 | 1 | 2 | ZACEF |
| 5 | 3 | 2 | 1 | ZBCEF |
| 5 | 2 | 2 | 3 | ZABEG, ZACDG, ZADEF |
| 5 | 2 | 3 | 2 | ZABFG, ZBCDG, ZBDEF |
| 5 | 3 | 2 | 2 | ZACFG, ZBCEG, ZCDEF |
| 5 | 1 | 3 | 3 | ZABDG |
| 5 | 3 | 1 | 3 | ZACEG |
| 5 | 3 | 3 | 1 | ZBCFG |
| 5 | 2 | 3 | 3 | ZADFG, ZBDEG |
| 5 | 3 | 3 | 2 | ZBEFG, ZDEFG |
| 5 | 3 | 2 | 3 | ZCDEG, ZAEFG |
| 5 | 2 | 2 | 4 | ZADEG |
| 5 | 2 | 4 | 2 | ZBDFG |
| 5 | 4 | 2 | 2 | ZCEFG |
| 5 | 3 | 3 | 3 | ZDEFG |
| 6 | 2 | 2 | 3 | ZABCDE |
| 6 | 2 | 3 | 2 | ZABCDF |
| 6 | 3 | 2 | 2 | ZABCEF |
| 6 | 2 | 3 | 3 | ZABCDG, ZABDEF |
| 6 | 3 | 2 | 3 | ZABCEG, ZACDEF |
| 6 | 3 | 3 | 2 | ZABCFG, ZBCDEF |
| 6 | 4 | 3 | 2 | ZBCEFG |
| 6 | 4 | 2 | 3 | ZACEFG |
| 6 | 3 | 3 | 3 | ZABEFG, ZACDFG, ZBCDEG |
| 6 | 3 | 4 | 2 | ZBCDFG |
| 6 | 2 | 4 | 3 | ZABDFG |
| 6 | 3 | 2 | 4 | ZACDEG |
| 6 | 3 | 2 | 4 |  |
| 6 | 2 | 3 | 4 | ZABDEG |
| 6 | 4 | 3 | 3 | ZCDEFG |
| 6 | 3 | 4 | 3 | ZBDEFG |
| 6 | 3 | 3 | 4 | ZADEFG |
| 7 | 4 | 4 | 3 | ZBCDEFG |
| 7 | 4 | 3 | 4 | ZACDEFG |
| 7 | 3 | 4 | 4 | ZABDEFG |
| 7 | 4 | 3 | 3 | ZABCEFG |
| 7 | 3 | 4 | 3 | ZABCDFG |
| 7 | 3 | 3 | 4 | ZABCDEG |
| 7 | 3 | 3 | 3 | ZABCDEF |
| 8 | 4 | 4 | 4 | ZABCDEFG |

To demonstrate the scheme experimentally, representative sequences were chosen from important diseases and designed respective primers and probes labeled with FAM, Cy3, ROX, and Cy5. The chosen coding was 1000 (HIV PolyProt), 1100 (HIV P17), 1010 (malaria), 1001 (herpes), 1101 (tuberculosis), and 1111 (dengue). Separate positive controls were performed for each sequence, as well as the full panel for several combinations of sequences present. The experimental details are in the Supporting Information.

The signals of the positive controls were added to produce the expected cumulative signal for every possible combination of present sequences, in each color. Each expected cumulative signal was plotted as its own level in a "chromatogram". Expected cumulative signals corresponding to combinations of the same rank in the same color were organized into their own "band". Doing this for all colors produced a level and band structure, against which the experimental results of each combination could be judged. Figure 1 shows the chromatograms of experimental combinations 4112, 3121, and 3102. The


Figure 1. Proof of principle of supercolor fluorescence intensity coding in PCR. (Left) Fluorescence intensity measurements of a multiplexed PCR reaction including sequences from TB, malaria, HSV-2, and HIV TPP in four frequency channels. The actual measured intensity measurements (black dots) are compared with all possible fluorescence codes (represented by color bands). The code 4112 uniquely maps to the combination of TB, malaria, HSV-2, and HIV TPP. (Center) Fluorescence intensity measurements of a multiplexed PCR reaction including sequences from dengue, malaria, and HIV TPP. (Right) Fluorescence intensity measurements of a multiplexed PCR reaction including sequences from TB, HSV-2, and HIV TPP.
experimental results for each combination in each color are shown by black dots
"Calling" the results to determine multiplicity in each color relies on three criteria: First, the analog result of fluorescence signal must be converted into a digital result of the multiplicity. Thus, if the dot falls inside a band, then the multiplicity of that band is the answer. If a dot is in between two bands, the result can be either multiplicity but not both. Second, illegal results can be excluded or considered a failed experiment. Third, if the dot is between two legal results, the "called" result is the one whose band is closer to the dot.

Using these criteria, the first experiment is "called" 4112, which is the correct answer. The second experiment should be "called" 2121 or 3121 , but 2121 is illegal, so the result must be 3121 , which is the correct answer. The third experiment is 3102 or 3101, both of which are legal, but the dot is clearly closer to 3102, so that should be the call, and that is the correct answer. These results provide an experimental demonstration and proof-of-principle for our coding method.
The results show a trend of slightly lower combined signal than what would be expected from a simple summation of positive-control signals. This leads us to believe that the combination is not perfectly linear, so there are additional effects, for which the chromatogram construction does not account at present. We conjecture that this particular effect has to do with the absolute concentration of quenchers. The higher the absolute concentration of quenchers, the larger the percentage of reaction volume they "black out", so the same percentage of released fluorophores fails to contribute to the cumulative signal. One solution is to use low concentrations of probes, so that the percentile loss is negligible, regardless of how much of the quencher is released. This issue is a subject of ongoing research.
Figure 1 shows that each rank containing multiple combinations has a relatively wide band. The width of the band is ultimately the difference between the highest and
lowest cumulative signals. If all positive controls in the particular color produced exactly the same fluorescence signal, then the width of each band would be zero. Since instead those signals are somewhat different, the resulting bands are relatively wide. Fortunately, there is a simple means to tighten the bands. Instead of loading all probes at the same concentration as is done here, each probe concentration can be adjusted so that the end-point fluorescence intensity of each positive control, in every color, is the same. This approach would significantly tighten the bands, making calling easier.

The coding method can be further improved by switching to a single hybridization probe per target sequence. To maintain the coding in Table 1, each such probe would be present with corresponding color labels at the concentrations prescribed by the band-tightening optimization above. Making one sequence, aliquoting it and labeling each aliquot with a corresponding color, is far less expensive than making multiple sequences each with its own color. As a result, the overall coding scheme is far cheaper to deploy in this second modality. Furthermore, it is far easier to fit just one hybridization probe to a target sequence, than two, three, or four.

Under the coding scheme of Table 1, the maximalmultiplicity combination is 6323, which means all coded sequences are present. On the other hand, Table 1 has the tacit restriction that each sequence can have only $1 \times$ signal in any color. If that restriction is relaxed, a sequence $1 X 00$ can be added, which preserves the sequence counting property in the blue color. What is the minimal $X$, such that the new addition preserves the general structure of Table 2, keeps degeneracy broken, and offers a new tier of multiplexing? The answer is $X_{\min }=4$. Basically, $X$ has to be the maximal previous cumulative multiplicity plus 1 . Then, any answer of $0-3$ in the green digit is interpreted as before and means sequence 1400 is not present, while any answer of $4-7$ in green means 1400 is present and the full answer must have 1400 subtracted from it to obtain the rest of the present sequences using Table 2.


Figure 2. Preliminary proof of principle of binary fluorescence intensity coding. Three sequences from dengue, HIV P17, and HIV TPP were coded as $1 \times, 2 \times$, and $4 \times$, respectively, all in blue excitation. The integer multiples of the $1 \times$ positive control were used to build a ladder of possible cumulative outcomes, which we call a "chromatogram". Then, the result of each possible non-null combination of the occurrence of the three templates was plotted as a black dot on the same fluorescence scale. The results show excellent agreement between expectation and measurement, proving the binary scheme for 3 sequences in 1 color.

Obviously, this system can be analogously expanded with additional sequences 1030 and 1004, for yellow and red, respectively. An analogy with Table 1 coding then suggests additional sequences 1404 and 1434. Iterating the same idea to the third tier will lead to further added sequences of 1800 , 1060, 1008, 1808, and 1868. Thus, if just 3 tiers are used, the total number of sequences that are unambiguously identifiable in any combination, is $5 \times 3=15$, while the maximalmultiplicity result would be $16-39-14-39$, when all sequences are present.
This expansion of the Table 1 coding offers a theoretically infinite number of additional sequences that can be unambiguously identified in any combination. However, there are practical limitations on the assay, e.g., the ability to tell the difference between bands $B$ and $B+1$ in the multiplicity call. Thus, practically, there is a limited multiplicity bandwidth to be utilized by any coding scheme. Hence, the most efficient scheme is the one that uses that bandwidth most sparingly, so that $M$ is maximized for a fixed $B_{\max }$.
These ideas lead to the following third coding scheme. Each sequence has one probe with only one color. The sequences of each color are coded by the multiplicities of their probes. The first few sequences in blue are 1000, 2000, 4000, and 8000. Similarly, the first few sequences in green are 0100, 0200, 0400, and 0800 . This scheme is theoretically infinite as before, but it allows for more sequences per unit multiplicity bandwidth. For example, if we take the first 4 tiers, the number of unambiguously identifiable sequences in any combination is 4 $\times 4=16$, while the "all-present" outcome is $15-15-15-15$. This third scheme is better than the second scheme in terms of bandwidth density and practicability, but it does not possess the self-correcting qualities of the first and second schemes. Furthermore, it makes the most efficient possible use of the bandwidth, since all rungs of the digital signal are used and decode to unique combinations of present sequences. We call
this third scheme "binary" as the coded intensities correspond to single digits of different orders in binary counting.

To test the binary scheme experimentally, three sequences were chosen, dengue Virus, HIV TPP, and HIV P17, in just one color (blue). The fluorescence signals from their probes in positive-control end-point PCR reactions were measured and used to calculate the respective probe concentrations that would produce a $1 \times, 2 \times$, and $4 \times$ signal intensity, respectively. Then, all 7 non-null combinations of target occurrence were processed in a batch experiment using the same Masterplex mixture of primers and probes. The fluorescence signal of each case was plotted in a chromatogram (Figure 2). The multiplicity levels were assembled as the expected integer multiples of the $1 \times$ signal intensity, while their widths were calculated as the propagated uncertainty of the $1 \times$ measurement. That uncertainty was equated to the standard deviation of the fluorescence signals of the last five PCR cycles in saturation.

The results in Figure 2 show a virtually perfect agreement between expectations and experiment. In every case, the result correctly decodes to the exact list of present sequences. Furthermore, the experimental set is exhaustive, as all theoretical combinations are accounted for and measured for the case of 1 color and 3 sequences. These results show proof of principle for the binary coding scheme.

## DISCUSSION

The so-far described schemes all utilize standard fluorescence. However, Forster resonance energy transfer (FRET) can be used in conjunction with standard fluorescence to enlarge the set further. FRET-based Taqman probes ${ }^{23}$ function in essentially the same way as standard Taqman probes but have an orthogonal spectral output, because they emit one step further into the larger wavelengths. Thus, if fluorescence interrogation of the sample is performed one color at a time,
the signal from the standard probe will be shifted by one step but the signal from the FRET probe will be shifted by two steps. Then, detection in each spectral band separately will produce independent signals. The same coding as described above can then be independently and simultaneously applied to the FRET probes. This should double the number of sequences that are unambiguously identifiable in any combination. Thus, for the third scheme used to the fourth tier, this means $4 \times 4 \times$ $2=32$ sequences in the set. The trade-off is that the analyzer systems would have to be FRET-capable. That limits usability with systems already embedded in practice, but the large gain in multiplexing would more than justify the feature in emerging systems.
The presented method of coding is a universal solution for systems limited by the requirement to fit multiple reporters within a certain bandwidth while avoiding overlap, crosstalk, and false positives. Thus, the particular coding shown in Table 1 for four basis "colors" is just one of the possibilities under the general method. If the basis "color" set is expanded, e.g., by future fluorophores with tighter emission spectra, the described coding method is still fully applicable and will offer larger multiplexing factors. Thus, the fundamental utility of the coding method will grow further with advances in probes.
Also, while standard fluorophores are discussed here as the reporter molecules, the same coding idea can be used with other systems, e.g., chemiluminescence, or in fact any quantitative reporting where the output signal scales with the constituent signals of the same color, frequency, physical size, absorption band, etc. The cumulative signal does not have to be digital or scale linearly with number and intensity of constituent signals. For example, if the physical principle of measurement is absorption, the cumulative attenuation is a product of constituent attenuations while the constituent concentrations are additive, due to the exponential nature of Beer-Lambert's law. The logarithm of the cumulative attenuation will then scale linearly with constituent concentrations in each absorption band (the equivalent of color), so the same multiplexing method is still applicable.

While the presented description focused on multiplexed endpoint PCR for qualitative panels, the same method can be applied to certain types of qPCR . Digital PCR systems (e.g., BioRad's digital droplets and Fluidigm's dynamic arrays) produce highly sensitive quantification of copy numbers, and so are qPCR systems by output, but their actual mode of operation is end-point PCR ${ }^{24,25}$ This means that our presented method can be immediately integrated into those systems and would drastically expand their productivity.
Furthermore, we can confidently prophesy that the presented multiplexing method would work within traditional real-time PCR. First, the real-time data would be fully recorded as the PCR runs to completion. The end-point values would be decoded into the list of present sequences. Second, individual $C_{\mathrm{T}} \mathrm{S}$ can be gleaned in each color by detecting the maxima of the second derivative of the real-time curve in that color. Third, as a particular sequence gets amplified, fluorophores are released from its Taqman probes in the same ratios as the color multiplicities in the coding of that sequence. That is so because the amplification breaks up the same probe and thus has no bias among the colors for that sequence. Hence, e.g., under the second coding scheme, 1100 will have the same $C_{T}$ in blue and green, while 1400 will have its green $C_{T}$ precede its blue $C_{T}$ by 2 cycles. This set of clues should be rich enough to
decipher the identity and starting quantity of each present sequence.

Finally, the applicability of the method is not limited to PCR or even biochemical assays. Any qualitative test, producing a fixed amount of signal that is independent of the starting amount of the test's target, is subject to the same mechanic and thus can benefit from this multiplexing method. If there is a "spectral" dimension to the physical basis of the test, the concept of "color" would apply and, thus, the presented mathematical apparatus can be engaged to full force. If there is no "spectral" dimension, single-color multiplexing is still applicable, e.g., using the binary scheme presented here.

## ■ MATERIALS AND METHODS

General Reagents. TE buffer, pH 7 (Life Technologies, Carlsbad, CA), UltraPure RNase-free Water (Life Technologies, Carlsbad, CA), and Taq $5 \times$ Master Mix (Fisher Scientific Company, Tustin, CA) were used.

DNA Sequences, Primers, and Probes. Five nucleic acid pathogens of clinical relevance were chosen for this study: human immunodeficiency virus 1 (HIV-1), Plasmodium falciparum (malaria), herpes simplex virus-2 (HSV-2), Mycobacterium tuberculosis (TB), and dengue virus type 3 (dengue Fever).

Two targets of diagnostic relevance on the HIV-1 genome, ${ }^{26-29} \mathrm{p} 17$ and polyprotease, were selected from the Los Alamos National Laboratory HIV-1 reference sequence. ${ }^{30}$ A diagnostic sequence from the malaria falciparum ChR7 gene ${ }^{31,32}$ was obtained from the UCSC Plasmodium falciparum Genome Browser. ${ }^{33}$ A sequence for herpes simplex virus- $2^{34}$ was synthesized from the sequence obtained from the European Molecular Biology Library. ${ }^{35}$ Similarly, a diagnostic sequence for rpoB gene in Mycobacterium tuberculosis ${ }^{36-38}$ was synthesized from a sequence obtained from the European Molecular Biology Library. ${ }^{39}$ A PCR diagnostic sequence for dengue Virus Type $3^{40}$ was obtained from the National Institute of Health genetic sequence database. ${ }^{41}$

All oligonucleotides were synthesized by Integrated DNA Technologies (Coralville, Iowa). Diagnostic sequences were input into IDT's OligoAnalyzer 3.1 tool. Probes and primer pairs for each target were chosen, from the OligoAnalyzer's set of generated sequences, to minimize homology with unintended targets, probes, and primers. TaqMan sense probes for all targets were synthesized with a fluorophore at the $5^{\prime}$ end and a quencher at the $3^{\prime}$ end. Sequence information is tabulated in Tables S1-S6, Supporting Information. Nucleic acid products were synthesized and lyophilized by IDT. These products were reconstituted with TE buffer and aliquoted for experimental use.

Polymerase Chain Reactions. All PCR reactions were performed on a Roche 480 LightCycler instrument (Roche Applied Science, Penzberg, Germany). The PCR cycling reaction was run for 45 cycles, with a 60 s hot-start at $95{ }^{\circ} \mathrm{C}$. The cycling conditions were: denaturation for 45 s at $95^{\circ} \mathrm{C}$, annealing for 50 s at $65^{\circ} \mathrm{C}$, and extension for 60 s at $70^{\circ} \mathrm{C}$. Each experiment was run in quintuplicate, with a reaction volume of $15 \mu \mathrm{~L}$. Fluoresence measurements in $483-533 \mathrm{~nm}$ (FAM), 523-568 nm (Cy3), $558-610 \mathrm{~nm}$ (ROX), and 615670 nm (Cy5) were first taken after the hot-start and again taken at the end of 45 thermal cycles. The change in fluorescence intensity between these two measurements, for each instance of an experiment, determined the fluorescent signal.

Experiment Conditions. Positive control experiments were performed to determine baseline fluorescence levels for each target and set of probes. Only targets with their associated probes were cycled in experiments $1,3,5,7$, and 9 , tabulated in Tables S7, S9, S11, S13, S15, and S17, Supporting Information. These were positive-control experiments that provided baseline fluorescence intensity for each oligo target. The change in fluorescence intensity, in each color, was used to assemble the expected cumulative signal levels in the chromatograms in Figure 1. The combinatorial superposition of fluorescence levels, in each independent wavelength channel, determined the actual bands for each count. Experiments 2, 4, 6, 8, and 10 measured the extent of cross-talk.

Experiments 11, 12, 13, and 14, tabulated in Tables S9-S12, Supporting Information, were multiplex coding experiments. Their changes of fluorescence intensity in each color were presented by the black circles in the chromatograms of Figure 1.

Experiments 23-27, tabulated in Tables S21-S27, Supporting Information, were binary-coded multiplex experiments in the FAM channel. The change in fluorescence intensity at the end of the saturated PCR experiment was presented by the black circles in the chromatograms of Figure 2. In each case, the multiplicity count corresponded to a unique combination of present targets. The baseline concentration for a $1 \times$ count was determined by a set of fluorophore titration experiments. A $200 \mathrm{nM} 1 \times$ concentration of probes in the FAM channel allowed for a sufficiently strong signal that minimized crosschannel bleed through.
In any particular experiment, the uncertainty in the cycling data was determined by the spread of values in the last five cycles of the particular amplification reaction. This uncertainty did not scale with the value of the total signal, which implied that the source of uncertainty was instrumental rather than experimental. A $1 \times$ fluorphore $(200 \mathrm{nM})$ baseline was determined by statistical analysis of a set of data on 200 nM concentration. The expected multiplicative signal levels were determined by multiplying this baseline by the multiplicity.

## - ASSOCIATED CONTENT

## (S) Supporting Information

Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

## AUTHOR INFORMATION

## Corresponding Author

*E-mail: arajagop@caltech.edu.

## Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The authors thank the Jacobs Institute at Caltech for access to a Roche 480 LightCycler instrument.

## - REFERENCES

(1) Liew, M.; Groll, M. C.; Thompson, J. E.; Call, S. L.; Moser, J. E.; Hoopes, J. D.; Voelkerding, K.; et al. BioTechniques 2007, 42 (3), 327333.
(2) Wiese, R.; Belosludtsev, Y.; Powdrill, T.; Thompson, P.; Hogan, M. Clin. Chem. 2001, 47 (8), 1451-1457.
(3) Sanger, F.; Nicklen, S.; Coulson, A. R. Proc. Natl. Acad. Sci. U.S.A. 1977, 74 (12), 5463-5467.
(4) Tyagi, S.; Marras, S.; Kramer, F. R. Nat. Biotechnol. 2000, 18 (11), 1191-1196.
(5) Lee, L. G.; Livak, K. J.; Mullah, B.; Graham, R. J.; Vinayak, R. S.; Woudenberg, T. M. BioTechniques 1999, 27 (2), 342-349.
(6) Vet, J. A. M.; Majithia, A. R.; Marras, S. A. E.; Tyagi, S.; Dube, S.; Poiesz, B. J.; Kramer, F. R. Proc. Natl. Acad. Sci. U.S.A. 1999, 96 (May), 6394-6399.
(7) Fodor, S. P. A.; Rava, R. P.; Huang, X. C.; Pease, A. C.; Holmes, C. P.; Adams, C. L. Nature 1993, 364, 555-556.
(8) Speicher, M. R.; Ballard, S. G.; Ward, D. C. Nat. Genet. 1996, 12, 368-376.
(9) Huang, Q.; Hu, Q.; Li, Q. Clin. Chem. 2007, 53 (10), 1741-1748.
(10) Huang, Q.; Zheng, L.; Zhu, Y.; Zhang, J.; Wen, H.; Huang, J.; Niu, J.; et al. PloS One 2011, 6 (1), No. el6033.
(11) Henegariu, O.; Heerema, N. A.; Dlouhy, S. R.; Vance, G. H.; Vogt, P. H. BioTechniques 1997, 511 (September), 504-511.
(12) Chamberlain, J. S.; Gibbs, R. A.; Ranierl, J. E.; Nguyen, P. N.; Thomas, C. Nucleic Acids Res. 1988, 16 (23), 11141-11156.
(13) Waters, L. C.; Jacobson, S. C.; Kroutchinina, N.; Khandurina, J.; Foote, R. S.; Ramsey, J. M. Anal. Chem. 1998, 70 (1), 158-162.
(14) Oliveira, D. C.; Lencastre, H. D. Antimicrob. Agents Chemother. 2002, 46 (7), 2155-2161.
(15) Lao, K.; Xu, N. L.; Yeung, V.; Chen, C.; Livak, K. J.; Straus, N. A. Biochem. Biophys. Res. Commun. 2006, 343 (1), 85-89.
(16) Zhang, K.; McClure, J.-A.; Elsayed, S.; Conly, J. M.; Louie, T. J. Clin. Microbiol. 2005, 43 (10), 5026-5033.
(17) El-hajj, H. H.; Marras, S. A. E.; Tyagi, S.; Kramer, R.; Alland, D. J. Clin. Microbiol. 2001, 39 (11), 4131-4137.
(18) Chong, S. S.; Boehm, C. D.; Higgs, D. R.; Cutting, G. R. Blood 2000, 95 (1), 360-362.
(19) Paton, A. W.; Paton, J. C. J. Clin. Microbiol. 1998, 36 (2), 598602.
(20) Urdea, M.; Penny, L. A.; Olmsted, S. S.; Giovanni, M. Y.; Kaspar, P.; Shepherd, A.; Wilson, P.; et al. Nature 2006, 73-79.
(21) Livak, K. J.; Flood, S. J. A.; Marmaro, J.; Giusti, W.; Deetz, K. Genome Res. 1995, 4, 357-362.
(22) Hartman, L. J.; Coyne, S. R.; Norwood, D. A. Mol. Cell. Probes 2005, 19 (1), 51-59.
(23) Jothikumar, P.; Hill, V.; Narayanan, J. BioTechniques 2009, 46 (7), 519-524.
(24) Chun, J.-Y.; Kim, K.-J.; Hwang, I.-T.; Kim, Y.-J.; Lee, D.-H.; Lee, I.-K.; Kim, J.-K. Nucleic Acids Res. 2007, 35 (6), No. e40.
(25) Lee, H. R.; Kim, S. Y.; Chang, H. E.; Song, S. H.; Lee, H. S.; Park, K. U.; Song, J.; et al. J. Clin. Microbiol. 2010, 48 (12), 46124614.
(26) Patterson, B. K.; Till, M.; Otto, P.; Furtado, M. R.; McBride, L. J.; Wolinsky, S. M. Science 1993, 260 (5110), 976-979.
(27) Roth, W. K.; Weber, M.; Seifried, E. Lancet 1999, 353 (9150), 359-363.
(28) Ptak, R. G.; Gallay, P. A.; Jochmans, D.; Halestrap, A. P.; Ruegg, U. T.; Pallansch, L. A.; Bobardt, M. D.; et al. Antimicrob. Agents Chemother. 2008, 52 (4), 1302-1317.
(29) Ou, C.-Y.; Kwok, S.; Mitchell, S. W.; Mack, D. H.; Sninsky, J. J.; Krebs, J. W.; Feorino, P.; et al. Science 1988, 239 (4837), 295-229.
(30) http://www.hiv.lanl.gov/content/index. Accessed 02/24/2012.
(31) Heidari, M.; Assmar, M.; Daloii, M. R. N. J. Sci, Islamic Repub. Iran 2005, 16 (1), 21-24.
(32) Tirasophon, W.; Ponglikitmongkol, M.; Wilairat, P.; Boonsaeng, V.; Panyim, S. Biochem. Biophys. Res. Commun. 1991, 175 (1), 179184.
(33) http://microbes.ucsc.edu/cgi-bin/hgGateway?hgsid= $612764 \&$ clade $=$ eukaryota-protista\&org $=0 \& d b=0$. Accessed 04/02/ 2012.
(34) Weidmann, M.; Meyer-König, U.; Hufert, F. T. J. Clin. Microbiol. 2003, 41 (4), 1565-1568.
(35) http://www.ebi.ac.uk/Tools/dbfetch/emblfetch?db=embl\&id= AJ303204\&format=default\&style=default\&Retrieve=Retrieve. Accessed 04/02/2012.
(36) Noordhoek, G. T.; Kolk, A. H. J.; Bjune, G.; Catiy, D.; Dale, J.
W.; Fine, P. E. M.; Godfrey-Faussett, P.; et al. J. Clin. Microbiol. 1994, 32 (2), 277-284.
(37) Rosenstraus, M.; Wang, Z.; Chang, S.-Y.; DeBonville, D.; Spadoro, J. P. J. Clin. Microbiol. 1998, 36 (1), 191-197.
(38) Beige, J.; Lokies, J.; Schaberg, T.; Finckh, U.; Fischer, M.; Mauch, H.; Lode, H.; et al. J. Clin. Microbiol. 1995, 33 (1), 90-95.
(39) http://www.ebi.ac.uk/Tools/dbfetch/emblfetch?db=embl\&id= GQ395623\&format=default\&style=default\&Retrieve=Retrieve. Accessed 04/02/2012.
(40) dos Santos, H. W. G.; Poloni, T. R. R. S.; Souza, K. P.; Muller, V. D. M.; Tremeschin, F.; Nali, L. C.; Fantinatti, L. R.; et al. J. Med. Virol. 2008, 80, 1426-1433.
(41) http://www.ncbi.nlm.nih.gov/nuccore/M93130. Accessed 04/ 02/2012.


[^0]:    Received: April 30, 2013
    Accepted: June 13, 2013

