Lecture 3. Arrays of Obstacles and how polymers move in them.

First, some motivation and some history.

Long ago, before any of you were born, in the 1980's the Department of Energy in the USA decided that a great project would be to sequence the human genome.

At the time this was viewed as an enormous problem because the sequencing technology used (gel electrophoresis) was very slow and could not possibly scale to the human genome.
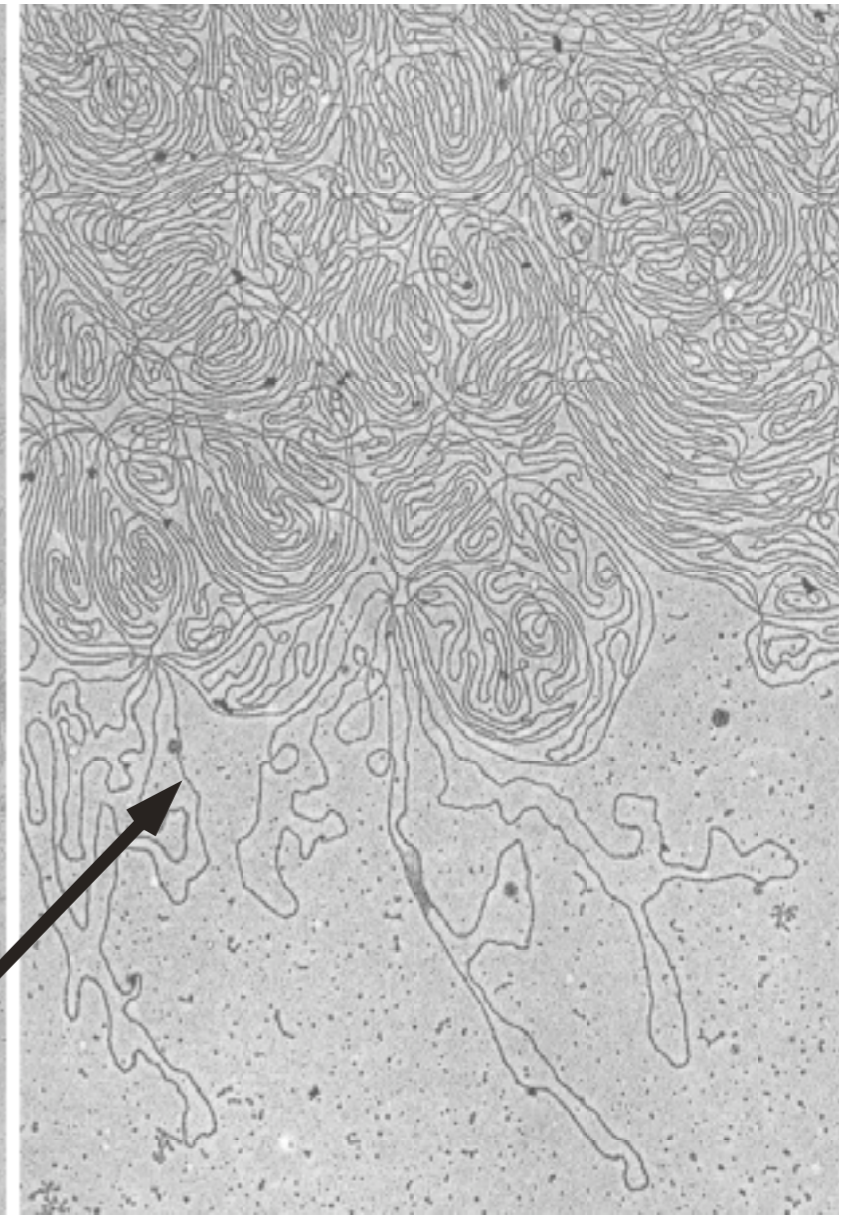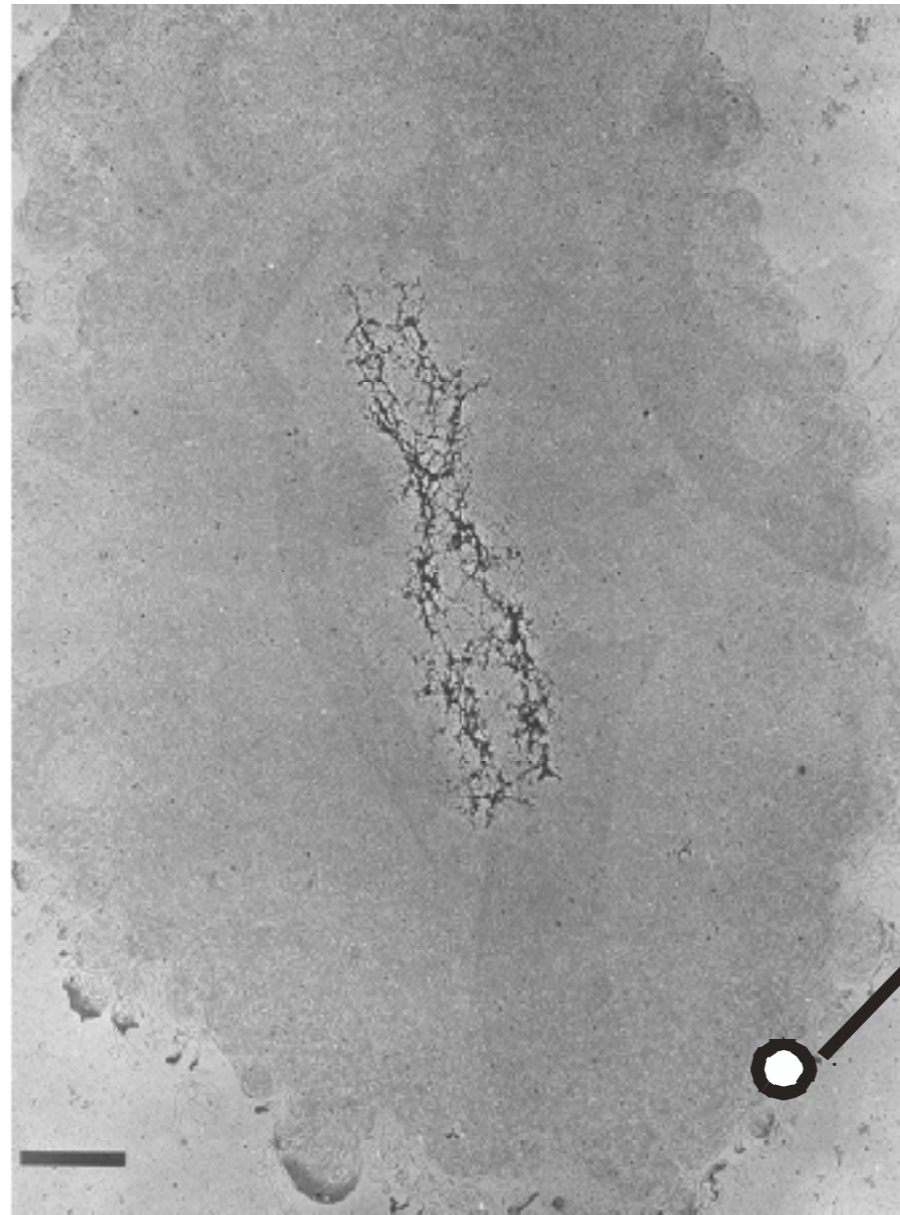
So, a call went out to the physicists to come up with new ideas for DNA sequencing.

I. A little biology about the importance of both gene sequence and transcription factors.
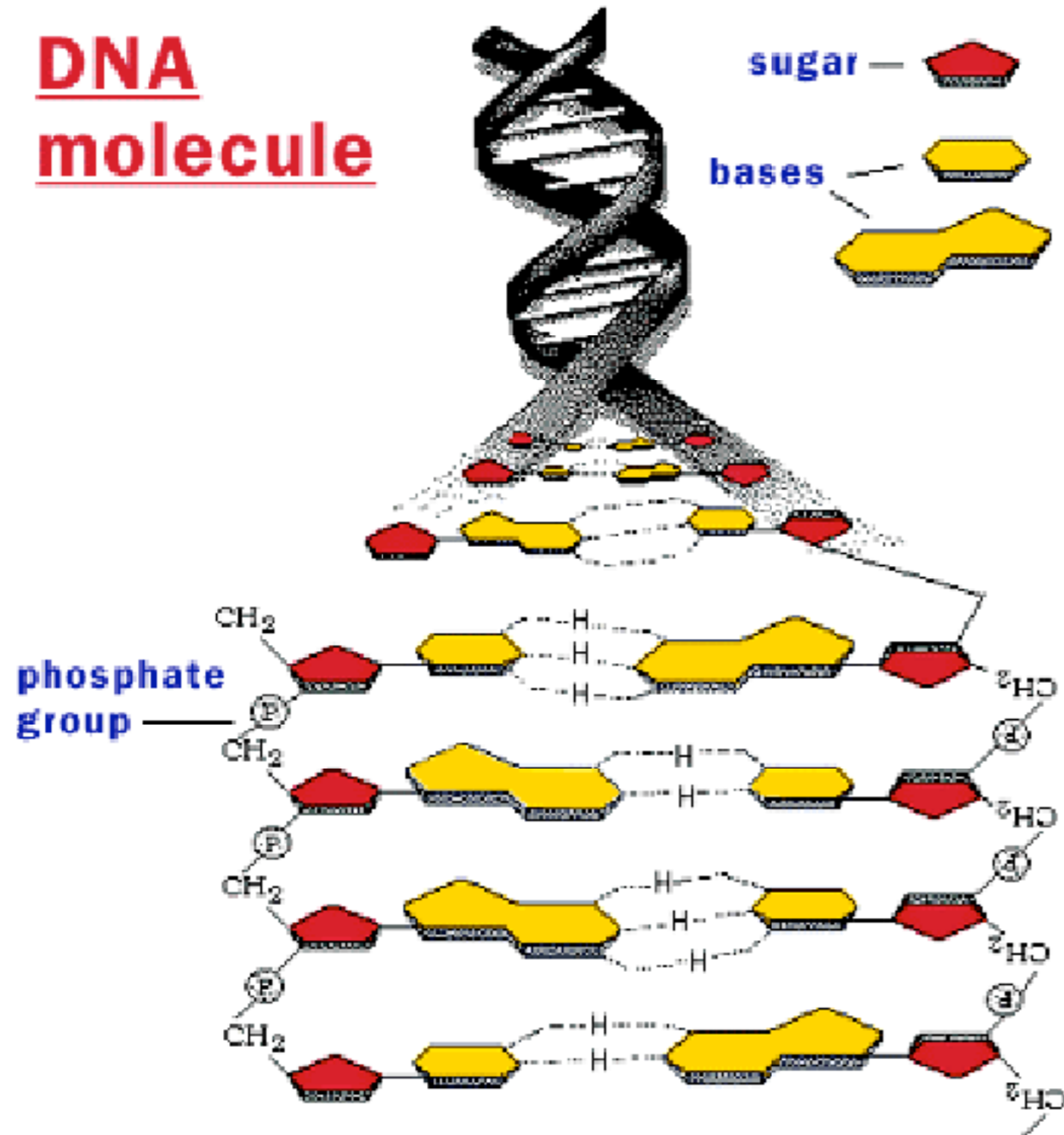
DNA contains the code for you. There are about 3 billion basepairs in the human genome, or about 1 meter of DNA.

Sequencing the human genome was an absolutely
  stupendous problem.

My great mistake: Thinking that it was impossible to sequence the Human Genome without the creative and high-tech inventions of biological physicists and their nanotechnology.

The story of the race to sequence the Human Genome is one of the great stories of the 20th century.

Two excellent sources:

1) The Genome War: How Craig Venter Tried to Capture the Code of Life and Save the World, by James Shreeve
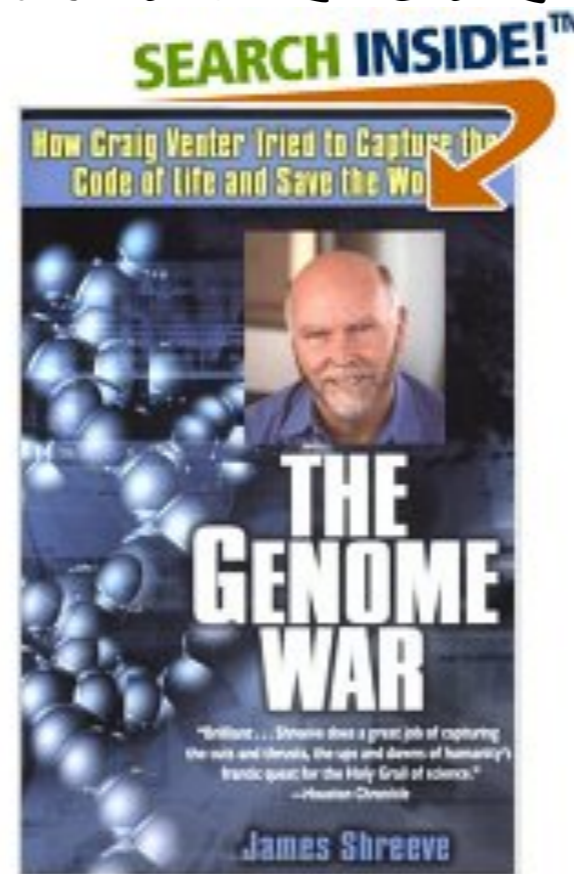
2) The Genome Warrior, Richard Preston / New Yorker 12 Jun 2000

"Craig Venter is an asshole. He's an idiot. He is a thorn in people's sides and an egomaniac, a senior scientist in the Human Genome Project said to me recently............"
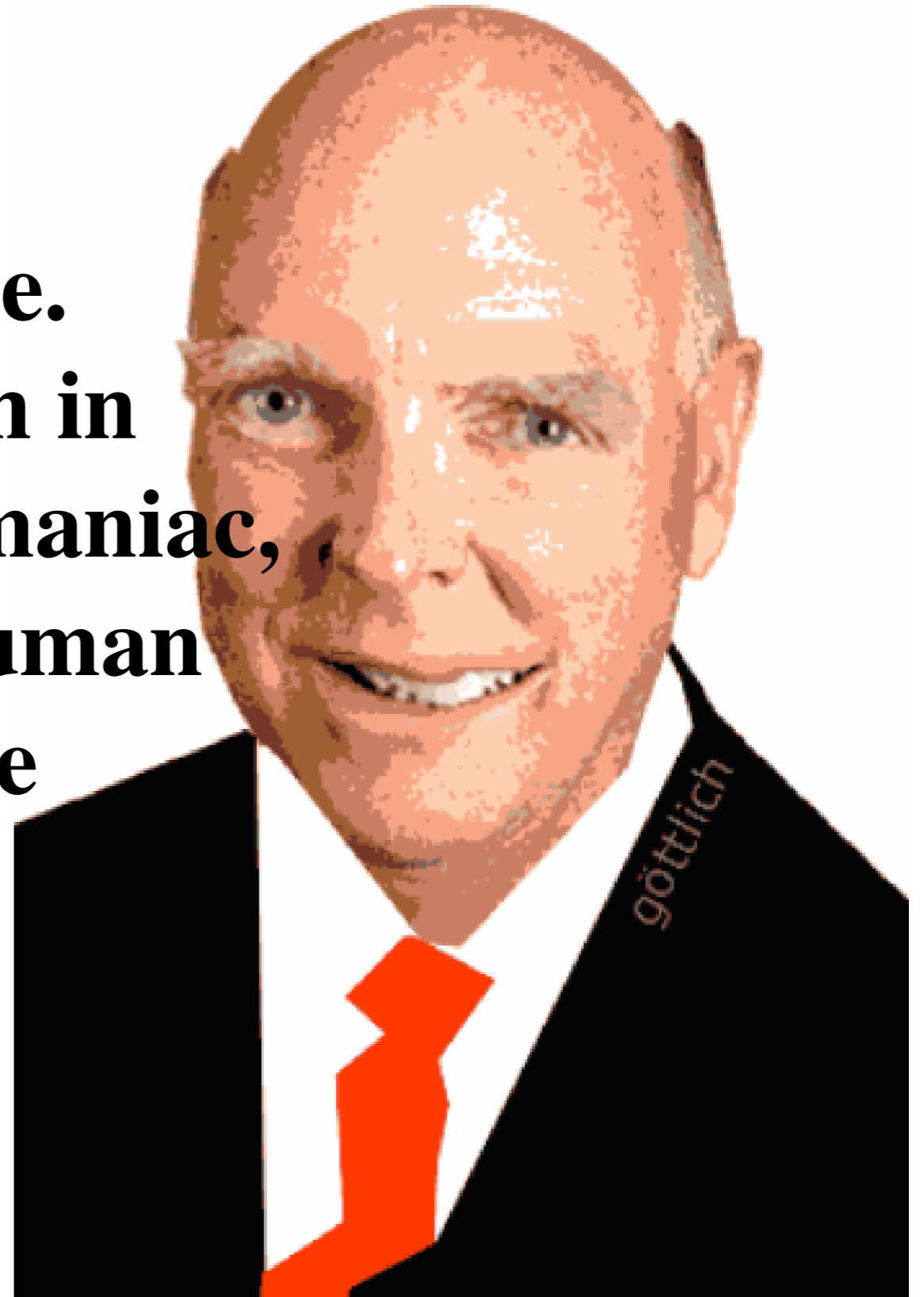
Why all the roiling emotions?

1). You should follow your instincts, and remember that being boring is worse than being wrong ("it isn't even wrong" -Pauli). Venter had an intuitive idea that was off the beaten track.

2) Greatness usually comes from unexpected places, and rarely comes from the expected ones. Beware the Ivy League Valedictorian or the Rhodes Scholar, they can lead you to a dusty death.

3) The NIH and the NSF have never funded an original idea in their lives, and never will.

4) Venter wanted to make money from the genome information he would get.

Craig Venter was not a member of the Old Boys Club, by any means.

"*Venter has a history of confrontation with government authorities. He told me that as an enlisted man in San Diego he was court-martialed for refusing a direct order given by an officer. "She happened to be a woman I was dating," Venter said. "We had a spat, and she ordered me to cut my hair. I refused." A friend of his, Ron Nadel, who was a doctor in Vietnam, recalls that one of Venter's blowups with authority involved "telling a superior officer to do something that was anatomically impossible." Venter worked for a year in the intensive-care ward at Da. Nang hospital, where, he calculates, more than a thousand Vietnamese and American soldiers died during his shifts, many of them while the 1968 Tet offensive was going on. When he returned to the United States, Venter finished college and then earned a Ph.D. in physiology and pharmacology from the University of California at San Diego.*"

If you missed Viet Nam, relax, we are doing it again in Irag, right on schedule, thanks to fools. Enjoy.

Follow your instincts, and remember that being boring is worse than being wrong ("it isn't even wrong" -Pauli)

"In 1984, Craig Venter had begun working at the N.I.H., where he eventually developed an unorthodox strategy for decoding bits of genes. At the time, other scientists were painstakingly reading the complete sequence of each gene they studied. This process seemed frustratingly slow to Venter. He began isolating what are called expressed sequence tags, or E.S.T.s, which are fragments of DNA at the ends of genes. When the E.S.T.s were isolated, they could be used to identify genes in a rough way. With the help of a few sequencing machines, Venter identified bits of thousands of human genes. This was a source of unease at the N.I.H., because it was a kind of skimming rather than a complete reading of genes. Three hundred and fifty human genes. The method was not received well by many genomic scientists. It was fast, easy, and powerful, but it didn't look elegant, and some scientists seemed threatened by it. Venter claims that two of his colleagues, who are now heads of public genome centers, asked him not to publish his method or move forward with it for fear they would lose their funding for genome sequencing."

The NIH and the NSF have never funded an original idea in their lives, and never will.

"By 1994, the Human Genome Project was mapping the genomes of model organisms, but no genome of any organism had been completed, except virus genomes, which are relatively small. Venter and Ham O. Smith proposed speeding things up by using a technique known as whole-genome shotgun sequencing. In shot gunning, the genome is broken into small, random, overlapping pieces, and each piece is sequenced, or read. Then the jumble of pieces is reassembled in a computer that compares each piece to every other piece and matches the overlaps, thus assembling.

Venter and Smith applied for a grant from the N.I.H. to shotgun-sequence the genome of a disease-causing bacterium called H. influenza, or H. flu for short. It causes fatal meningitis in children. The review panel at the N.I.H. gave Venter's proposal a low score, essentially rejecting it. According to Venter, the panel claimed that an attempt to shotgun-sequence a whole microbe was excessively risky and perhaps impossible. He appealed. The appeals process dragged on, and he went about shotgunning H. flu anyway. Venter and the TIGR team had nearly finished sequencing the H. flu genome when, in early 1995, a letter arrived at TIGR saying that the appeals committee had denied the grant on the ground that the experiment wasn't feasible. Venter published the H. flu genome a few months later in Science. Whole-genome shot gunning had worked. This was the first completed genome of a free-living organism."

In the end (2003), the Human Genome was sequenced using Venter's ideas, basically brute force whole-genome shot-gun sequencing using massive banks of capillary electrophoresis sequencers (which my biology friend assured me would never work about 1 month before the announcement) given Venter by companies who believed in him when the NIH did not.

As far as I can tell, the NIH is still in a state of denial.

"It seems quite possible that Venter's grant was denied because of politics. The review panel seems to have hated the idea of giving N.I.H. money to TIGR to make discoveries that would be turned over to a corporation, Human Genome Sciences. It turned down the grant, in spite of the fact that "all the smart people knew the method was straightforward and would work," Eric Lander, the head of the genome center at M.I.T. and one of the leaders of the public project, said to me."

Eric Lander was a Princeton Valedictorian who wrote an article in Science claiming shot-gun sequencing did not work in the Human Genome Sequencing.

All the time this was going on I was busy working on ways to use nanotechnology to sequence genomes, convinced that gels could not work fast enough (correct) and also convinced that capillary gel electrophoresis could not give high enough resolution. I was dead wrong on this, and oblivious to the story I have told you that was unfolding:

Thus: If you talk to a mediocre molecular biologist, do exactly the OPPOSITE of what s/he says: you can't go wrong.

And: You are only as good as the people you work with. Bad people drag you down.

Now, let me tell you a bit about the path I went down in nanotechnology while all this was going on. Looking back, I was arrogant and yet not a player at all, and I was oblivious to all the really earth-shaking things going around that involved billions of dollars and thousands of scientists.

I was a fool.

DNA:

(1) A VERY long polymer, up to cm long!

(2) About 15 Angstroms in diameter

(3) Negatively charged! 2e- per basepair

(4) "Flexible"
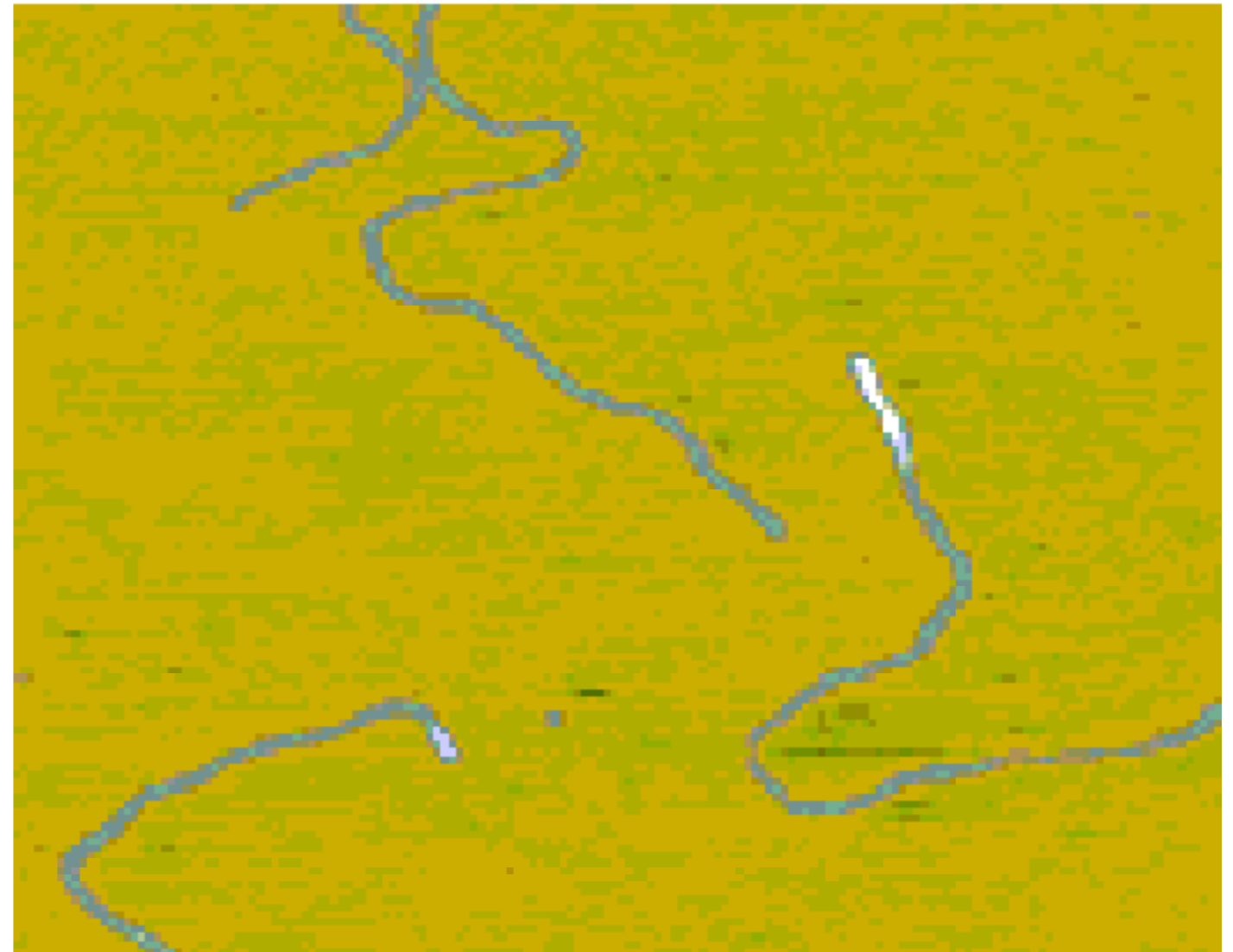
AFM pictures of
DNA fragments,
Paul and Helen
Hansma

A little polymer physics about dsDNA

A very important concept here: the persistence length "p" of a flexible polymer. Basically, it is a measure of how far you move along an arc before thermal energy bends the polymer randomly.

t(s)

t(0)

movie:wayne
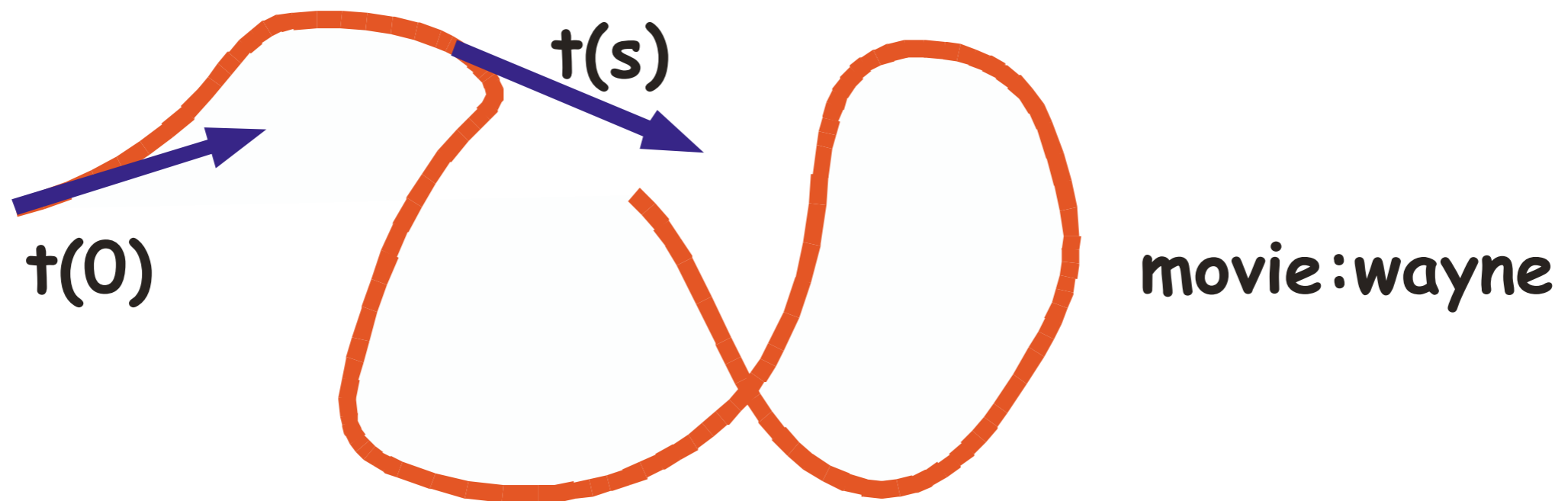
$\langle t(0)\, t(s) \rangle = \exp(-s/p)$ p = 600 Angstroms = 200 bp (double-strand DNA)
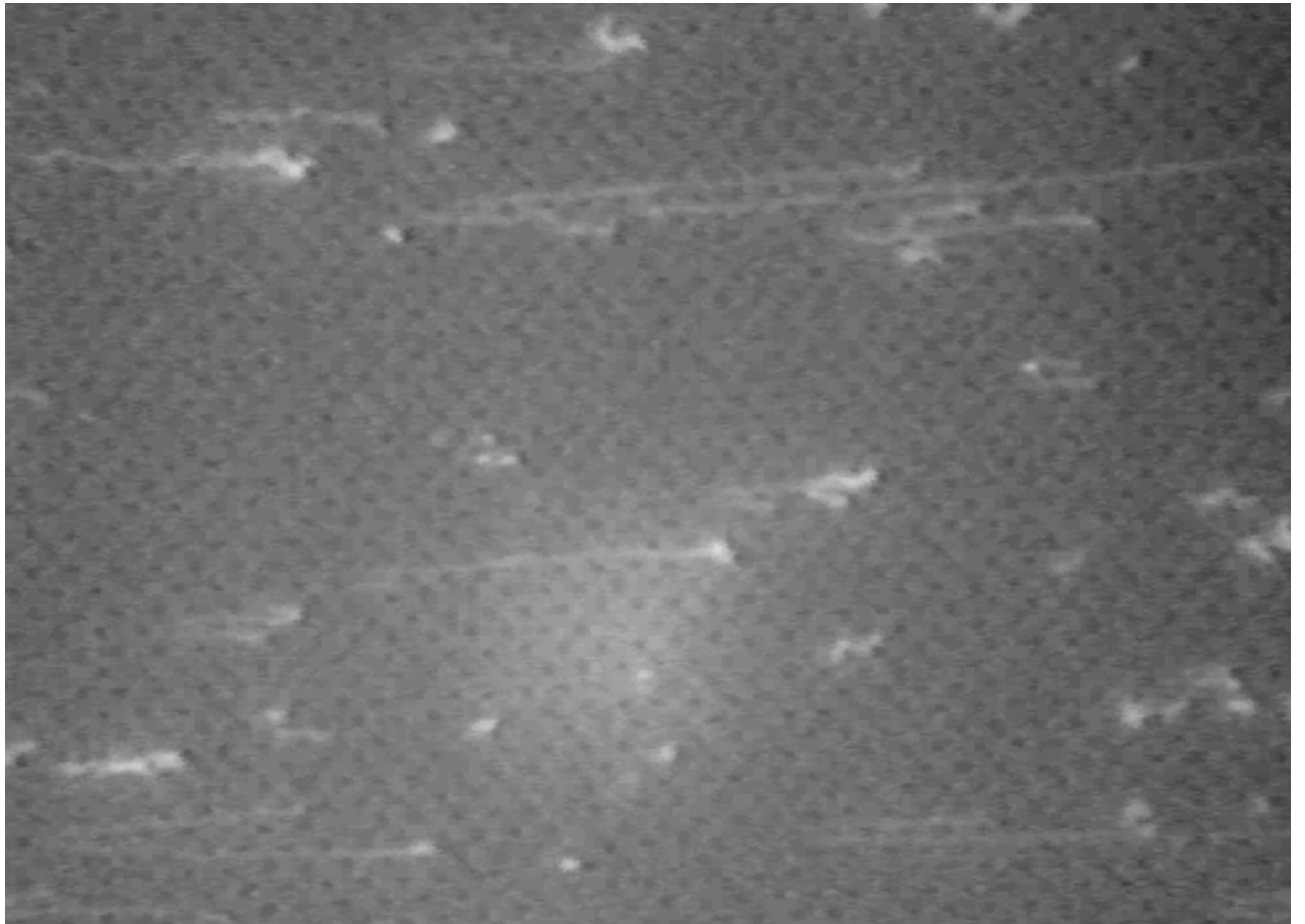
What's moving these molecules here? Although we are using an applied electric field E, it isn't F= qE, a common misconception.

Biomolecules like DNA have to be in a saline buffer, a plasma of neutral net charge. The ions cancel out the negative charge of the DNA backbone.

Famous equation: Poisson-Boltzmann:

$$\nabla^2 \psi = \frac{zen_o}{\varepsilon\varepsilon_o}\left[e^{\frac{ze\psi}{k_bT}} - e^{-\frac{ze\psi}{k_bT}}\right] \quad (1)$$

$$\nabla^2 \psi = D^2 \psi \quad (2)$$

$$D = \left[\frac{2z^2e^2n_o}{\varepsilon\varepsilon_o k_B T}\right]^{1/2} \quad (3)$$

For typical biological salt concentrations, n_o is about 100 mM and D is about 0.2 nm, so the shielding range is quite short: molecules look neutral. Even in the presence of an electric field, the transport is strongly influenced by hydrodynamics.

$$\rho\left[\frac{\partial \vec{v}}{\partial t} + (\vec{v} \bullet \nabla)\vec{v}\right] = \nabla P + \eta \nabla^2 \vec{v}$$

$$m\frac{d\vec{v}}{dt} = \vec{F}$$

Once yet again the infamous Navier-Stokes Equation, a lifetime of work can easily be invested in studying this.

Electrophoresis: really hydrodynamics because water is an isulator. The moving ions pull molecules along via hydrodynamic coupling.

Shear surface

ρ= charge density
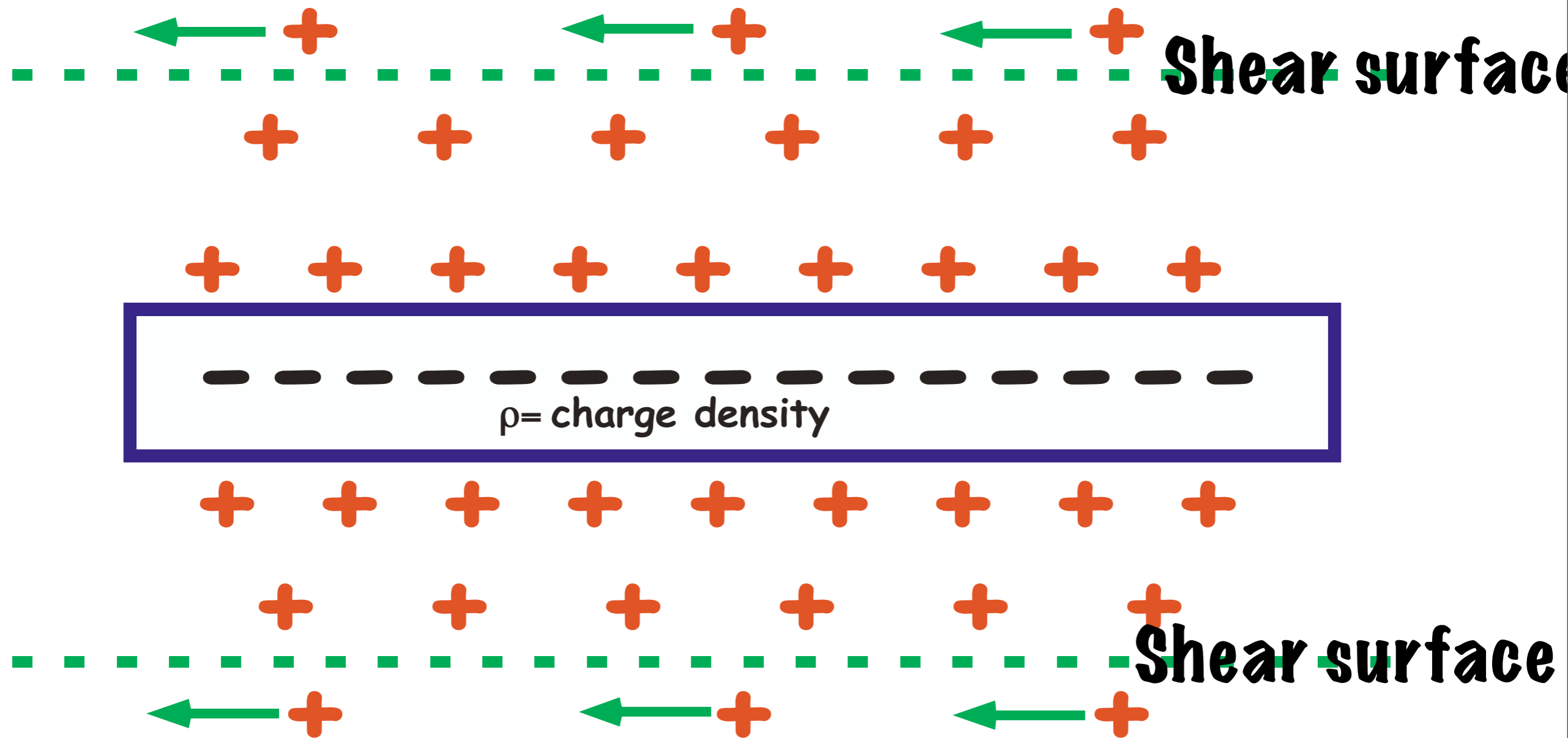
Shear surface

Zeta potential: potential at the shear surface

Malvern
Instruments
Ltd.
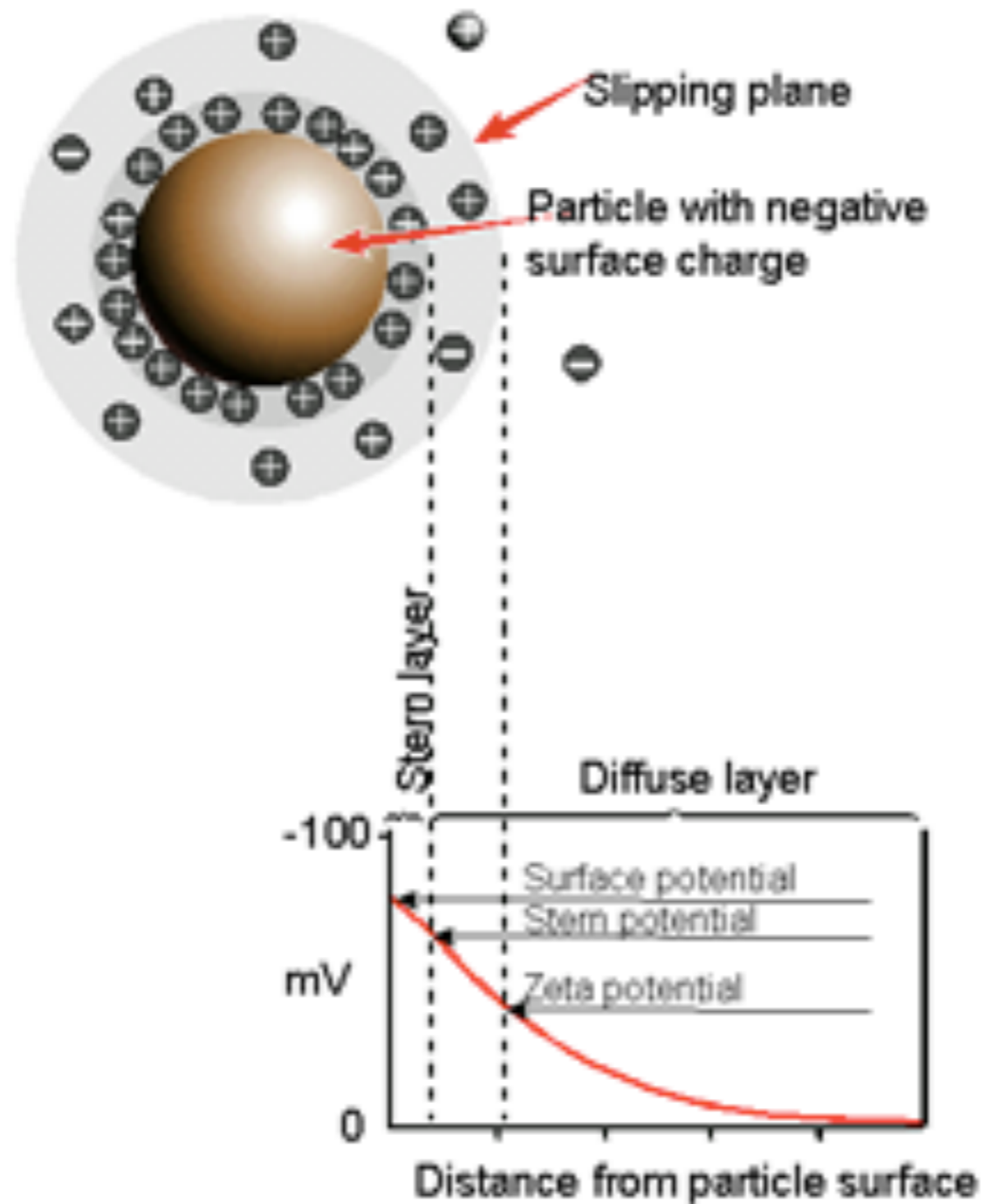
Since the Debye layer is only about 1-10 A to begin with, the Zeta shear surface is also << persistence length of dsDNA and hence dsDNA is "very" free draining and hydrodynamically uncoupled (no log terms)

The result is that the polymer is "free draining" and the parts are hydrodynamically decoupled from each other.

Free-draining is a **BIG DEAL** if you want to fractionate polymers. In the NS Equation if you move an object the hydrodynamic fluid flow reaches out to very large distances as I mentioned (Stokes drag), but **NOT** if free-draining happens, as in electrophoresis.

In "normal" hydrodynamics, if you move a polymer there is a velocity vector flow pattern V(r) radiating out from a point on the polymer that couples different parts of the object together, modifying the drag coefficient of the polymer from a strictly linear sum of terms (typically a 1/ln(L) correction), L=length of polymer.

For a free-draining polymer, the linear sum works, and the drag is proportional to L

Since the drag and the force both scale as L, the electrophoretic mobility is independent of the length and you cannot fractionate DNA molecules in bulk solution. Thus, the ubiquitous presence of some sort of retarding medium (a gel) in most molecular biology labs. The retarding medium adds a length-dependent additional force as the random coil tries to squeeze through the medium.

This is why I first started making obstacle arrays using microfabrication, I thought I could make better gels then the random matrices that biologists use, and maybe magically in these arrays run DNA faster and with more resolution. I didn't know what I was doing at the time since few physicists know hydrodynamics, we don't teach it to undergrads, unfortunately.

There was another problem however with my synthetic gels (post-arrays).

Genomic length DNA gets really stretched out in weak fields because the persistence length is actually pretty big! Young's modulus about the same as nylon.

$$R_z = \kappa \ln[\sinh(L/\kappa)/(L/\kappa)]$$

Where L= total length of the polymer, made of N pieces of length 2p (p=persistence length of the polymer) and:

$$\kappa = 2\lambda pE/k_bT$$

λ = charge/length of the polyelectrolyte.

There are two things at work here that kill the technology:

(1) The pore sizes (a=1 um) are much bigger than typical gels. This means that the applied force on each fragment spanning the port (about $a\rho E$, ρ = linear charge density) is large, so the polymer is stretched more easily than in a small pore.
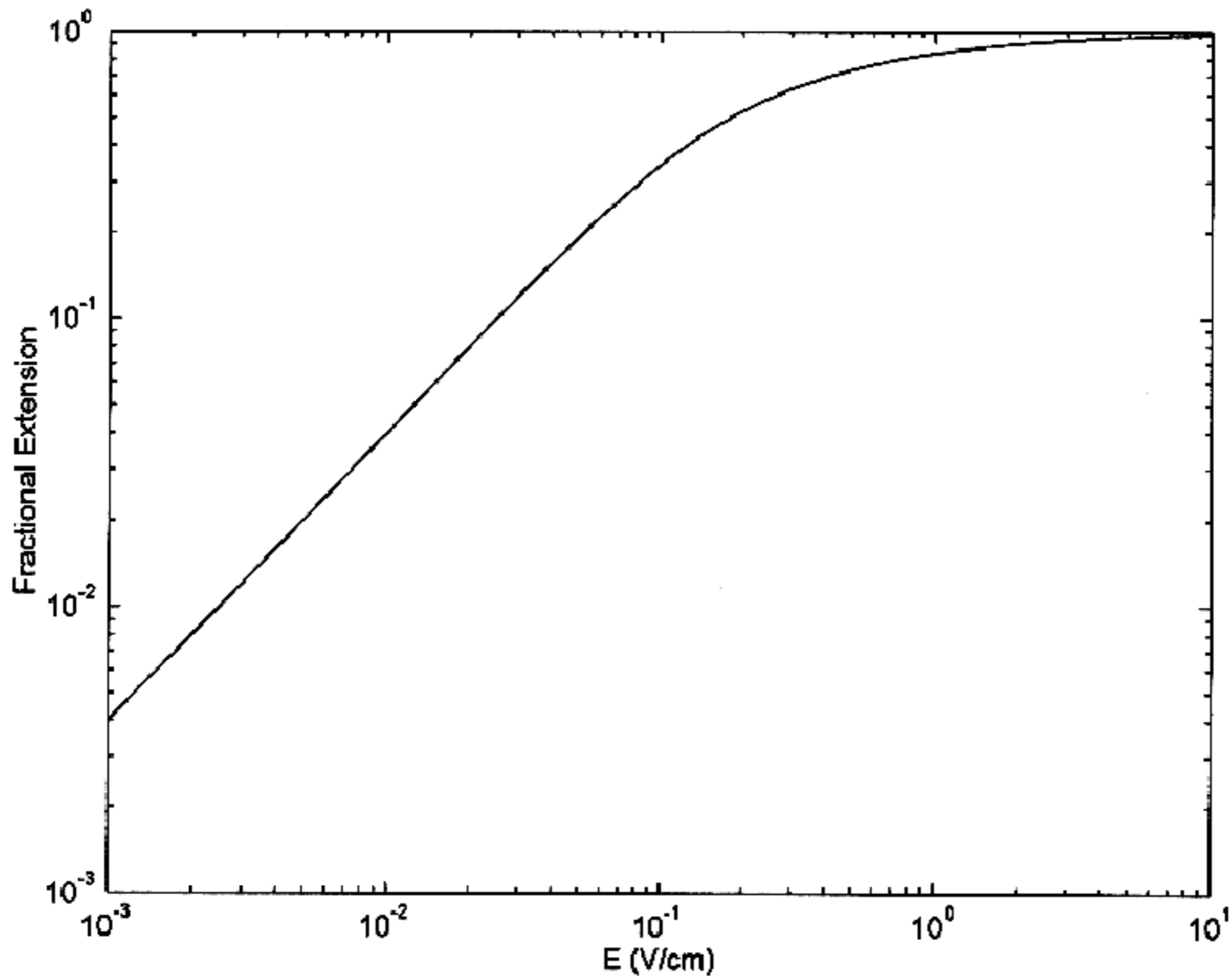
Lesson: in a "large" pore environment, DNA is highly aligned at low fields!

(2) In my thin "slit" of thickness h there is a hydrodynamic coupling to the surface via stick boundary conditions. The stick boundary condition slows down the entropic relaxation time of the polymer.

This can be a big effect, and the slow relaxation time enhances elongation.
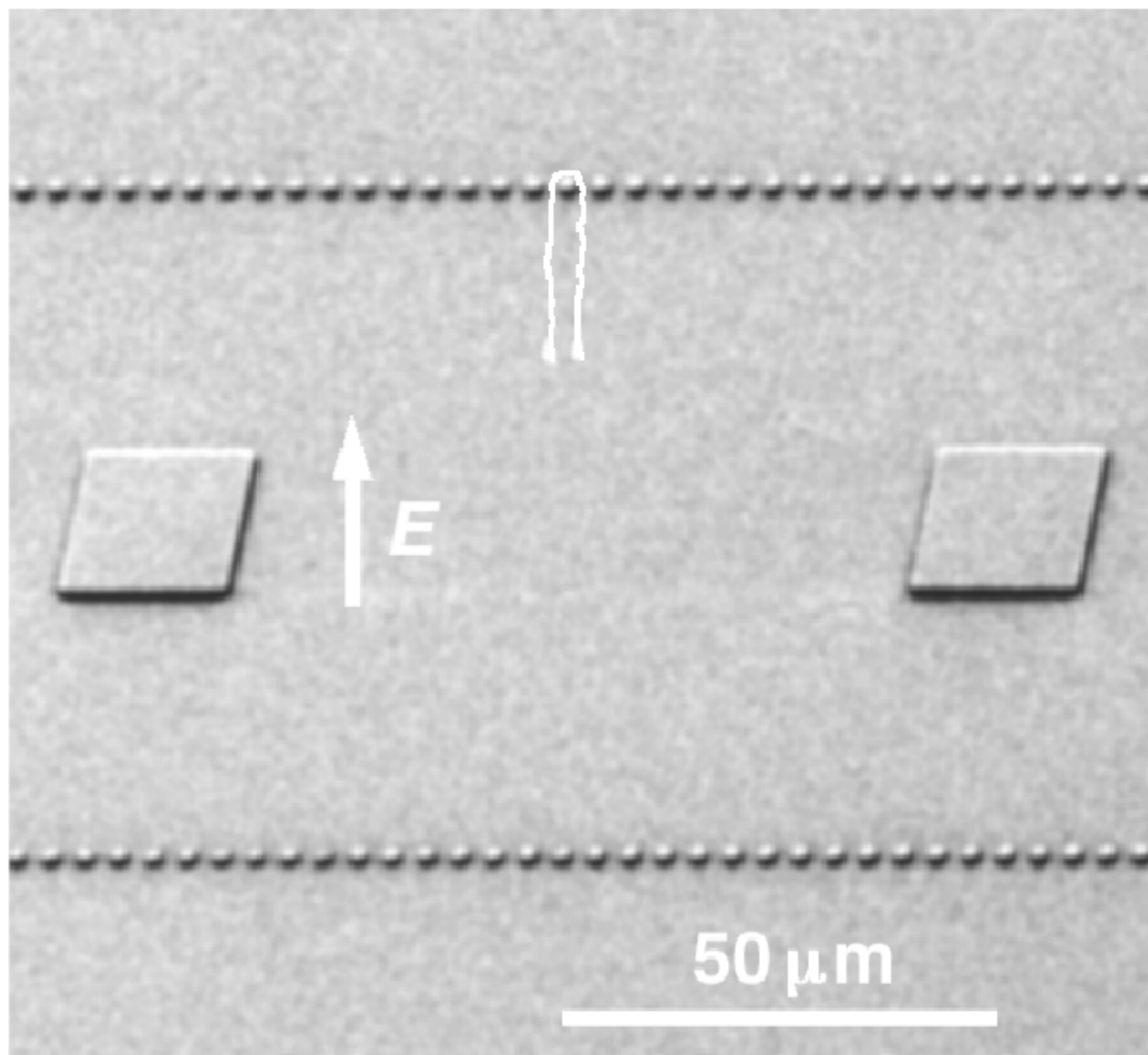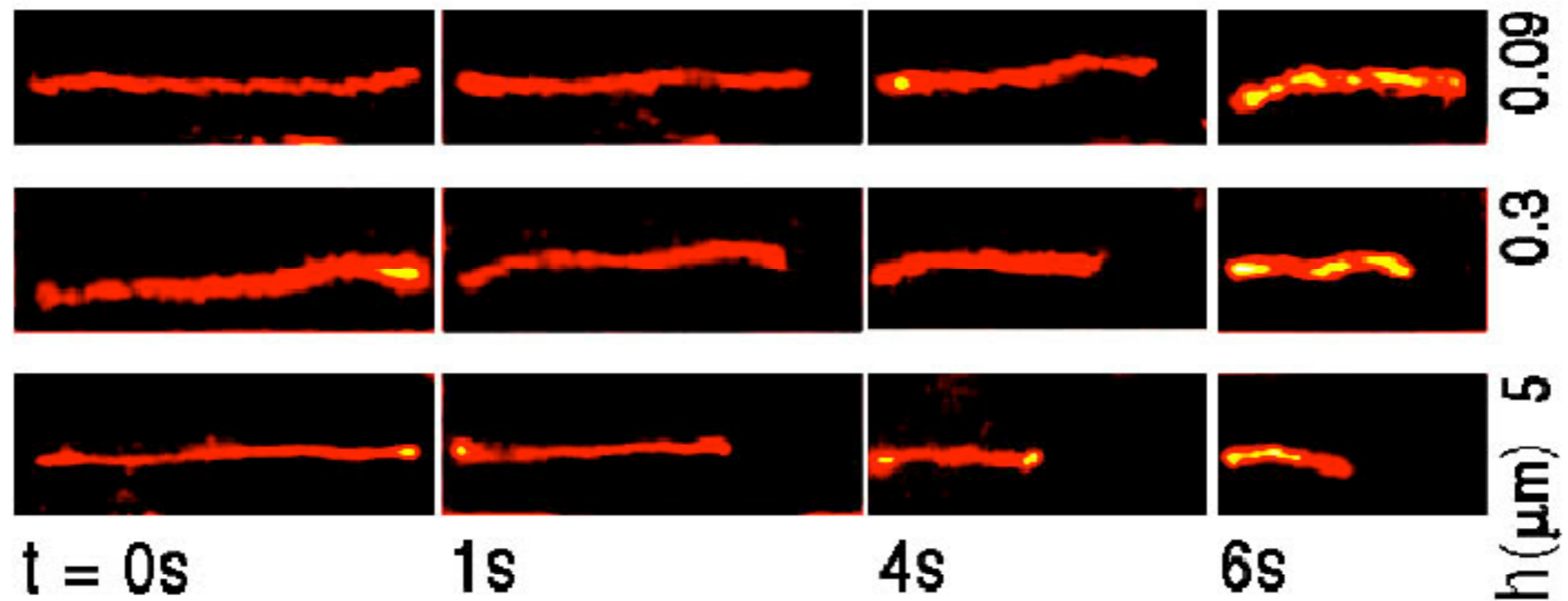
As the etch depth approaches the persistence length, relaxation times increase.
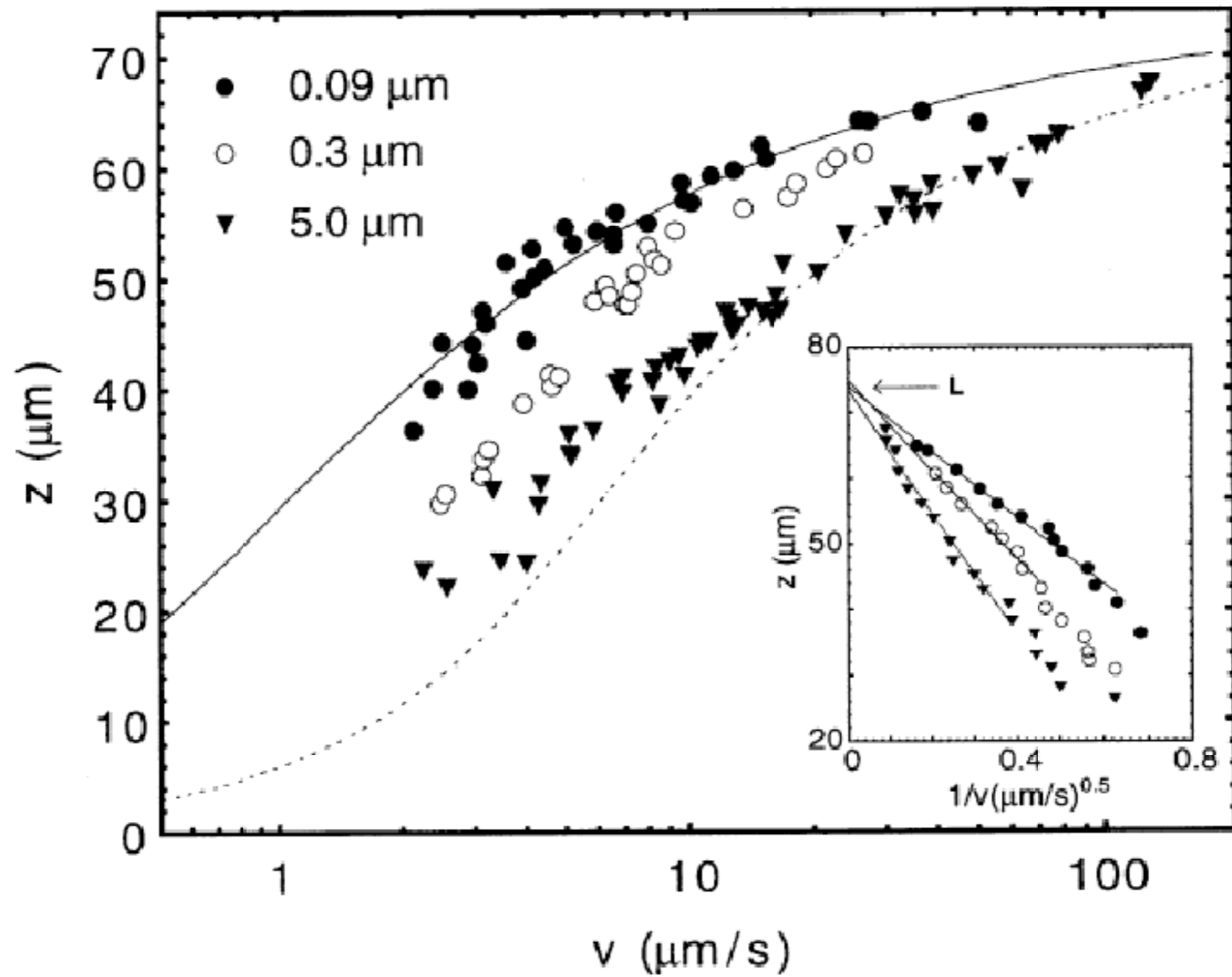
Elongation increases with decreasing etch depth

Hello Bob! They all move at the same damn speed!

Tom Duke in collaboration with our group suggested two ways to get around this problem. I will first talk about the pulsed field/hex array idea. The idea here is if you can't beat the physics, use the physics:

That is, if the polymer physics wants to elongate the polymer, figure out a way to use elongated molecules!

Tom's idea was a take on Ed Southern's idea for using PULSED TRANSVERSE fields to fractionate elongated polymers:
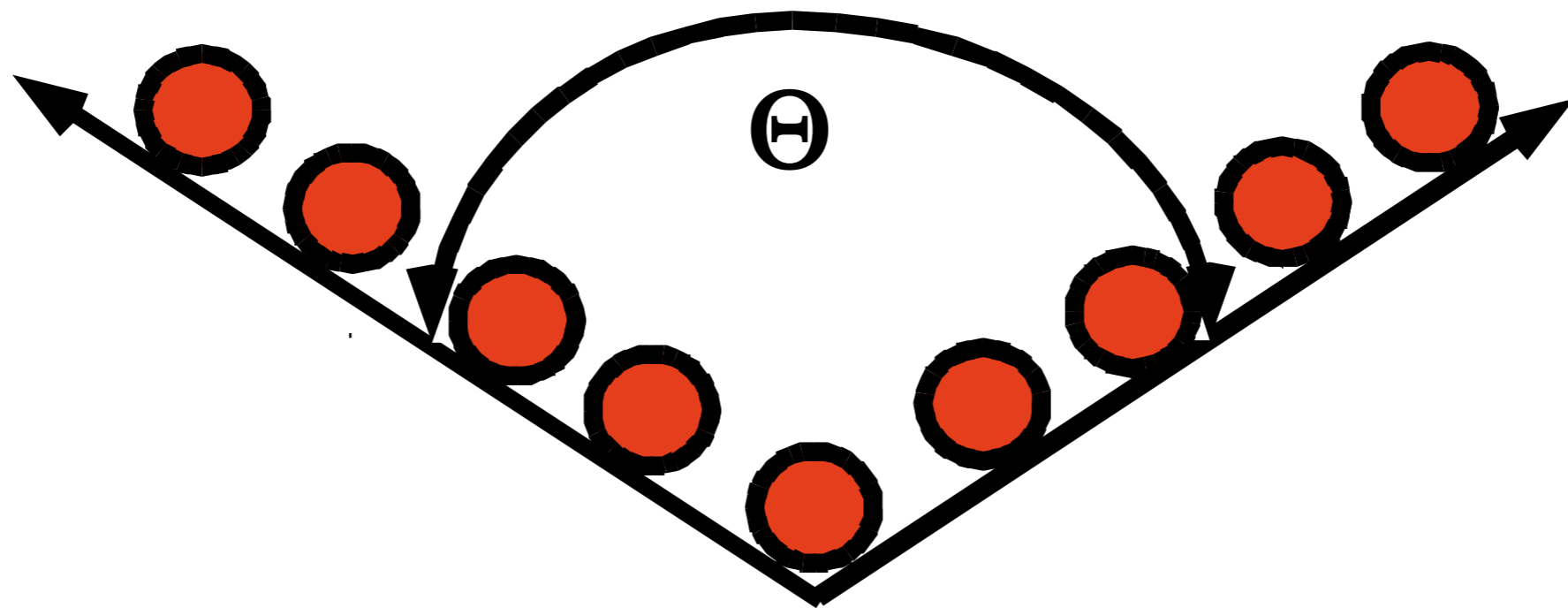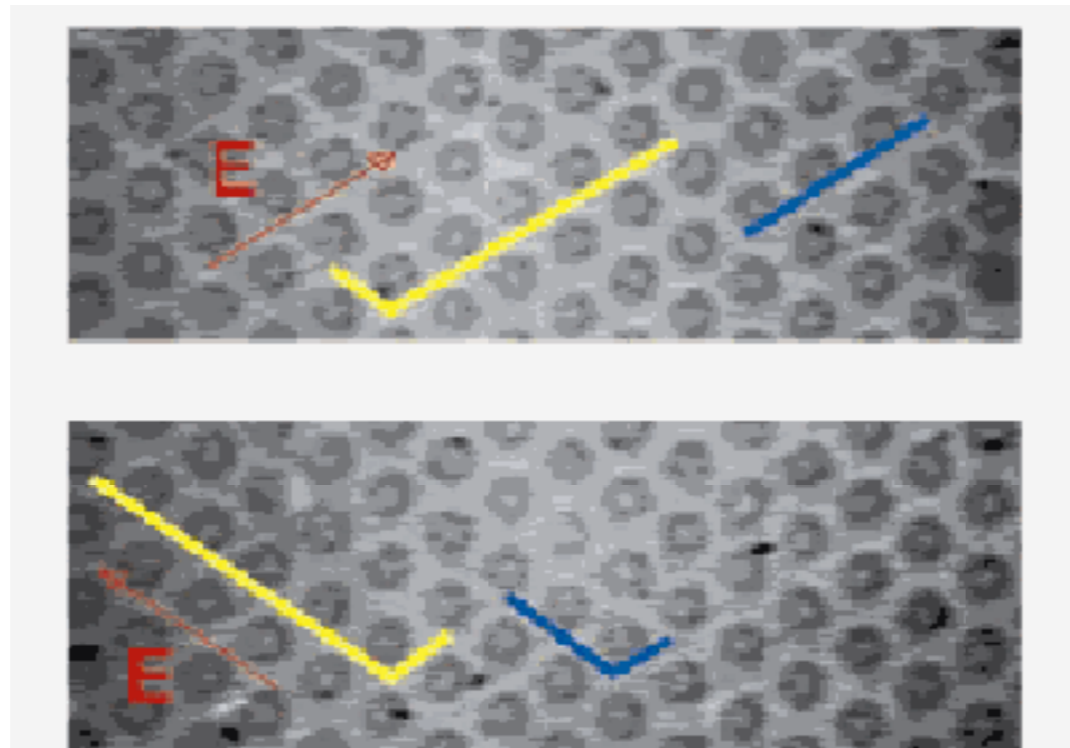
If the angle Θ is greater than 90⁰, then the polymer must backtrack along the original direction (former leading edge lies higher in potential energy then former tail). This is a fundamental idea.

This works as long as the posts constrain the polymer dynamics (hex array gives 120° channel angle). You want the polymer to be heading down a clear channel. If the polymer collides with posts, the elongation becomes less, and we want elongation here!

The reorientation time τ of a polymer of length L with electrophoretic speed v (v is independent of L!) is:

$$\tau \sim [L/v] \ln(1/|\cos(\Theta)|) / (1-|\cos(\Theta)|)$$

If the period T of the pulses is shorter than this reorientation time, the polymer see-saws back and forth endlessly, like a nervous stockbroker.

So, the prediction is that for molecules shorter than some critical length L* that the stretched polymers will have an average speed <v> given by

$$<v> = v \cos(\Theta/2) (1 - L/L^*)$$

where the critical length has reorientation time τ equal to pulse period:

$$[L^*/v] \ln(1/|\cos(\Theta)|) / (1 - |\cos(\Theta)|) = T$$

So....let's go to the video tape.

PULSED-FIELD
ELECTROPHORESIS
IN
HEXAGONAL ARRAYS

Olgica Bakajin
PRINCETON UNIVERSITY

2 μm POSTS
WITH
2 μm SPACINGS

NOW you can see that different lengths travel at different speeds!

0°

90°

120°

$\theta = 135^{o}$

$T = 12.5s$

$E = 2.8V/cm$
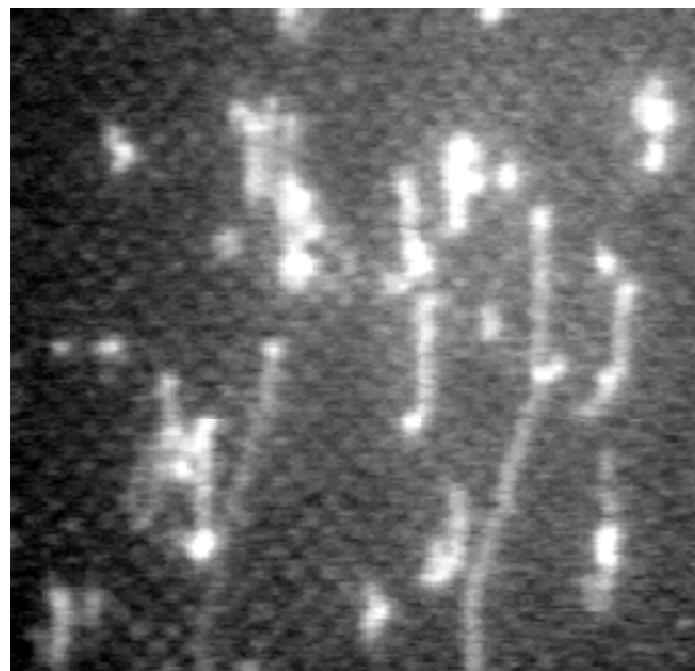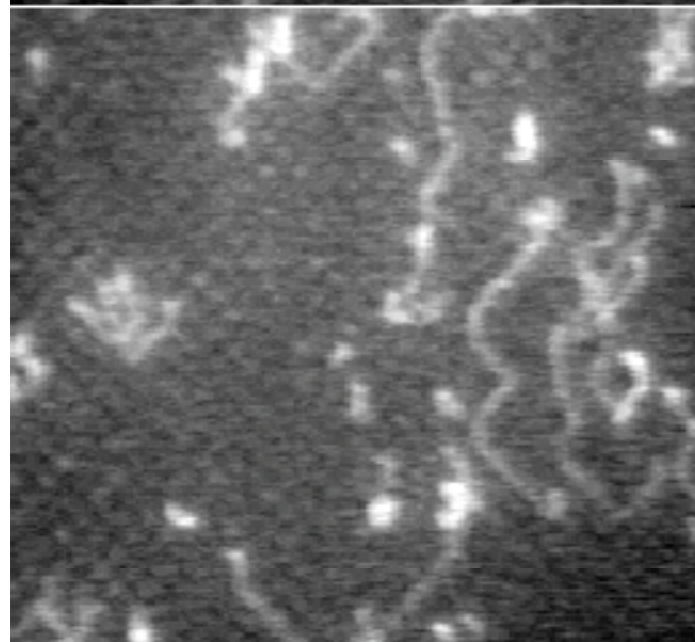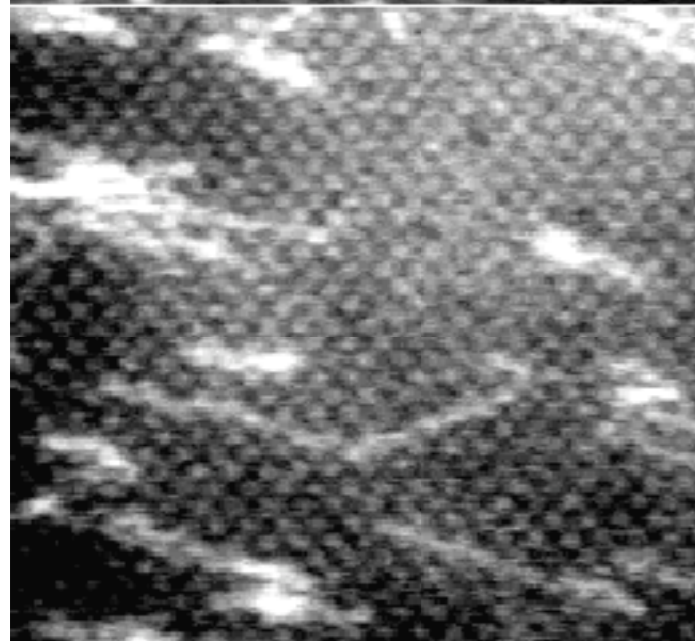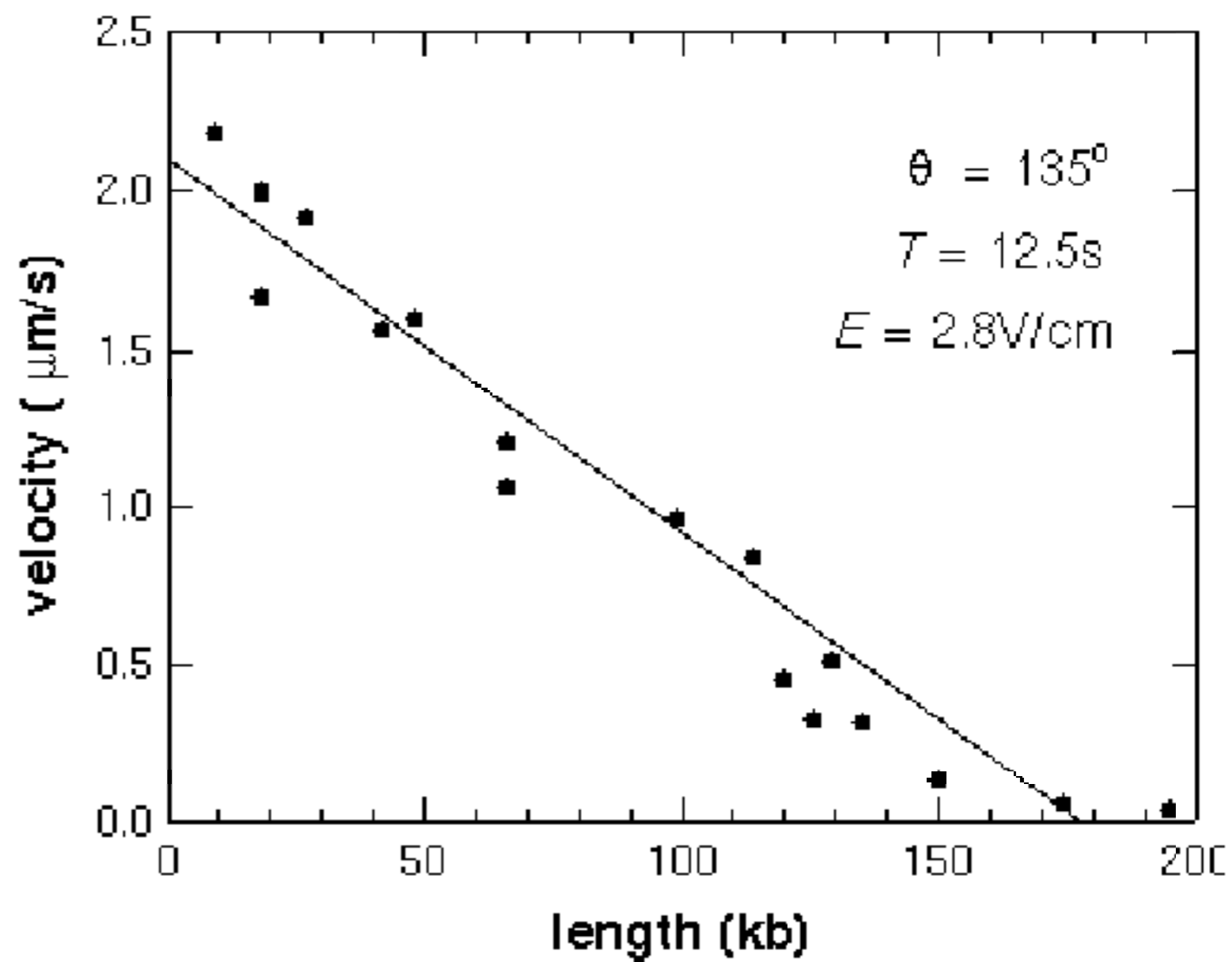
OK dudes, so question should be;

If this technique is damn good, why didn't I roar up to Trieste in my red Porsche 911 turbo to give a perfunctory all-purpose lecture saying nothing instead of sweating out a talk here trying to explain hopes for the future?

The answer is: unless you have tight control of the field lines EVERYWHERE on the chip and can concentrate the DNA into a thin line you only can get good fractionation at the center of the hex pattern where the fields are uniform.
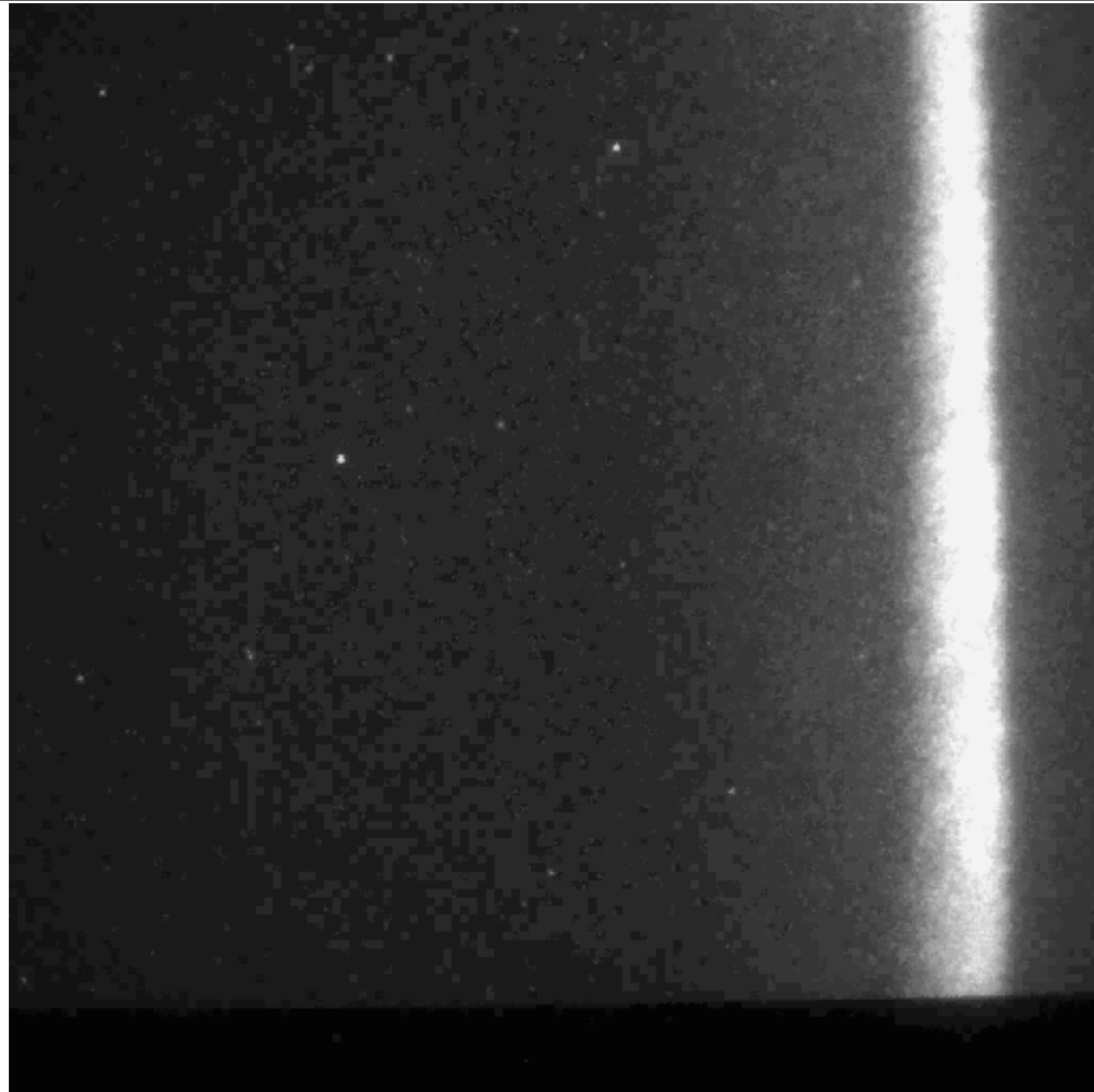
WE COULDN'T SOLVE THE BOUNDARY VALUE PROBLEM AT THE TIME (2001)!!!

DNA reservoir
Injection channel
Buffer reservoir

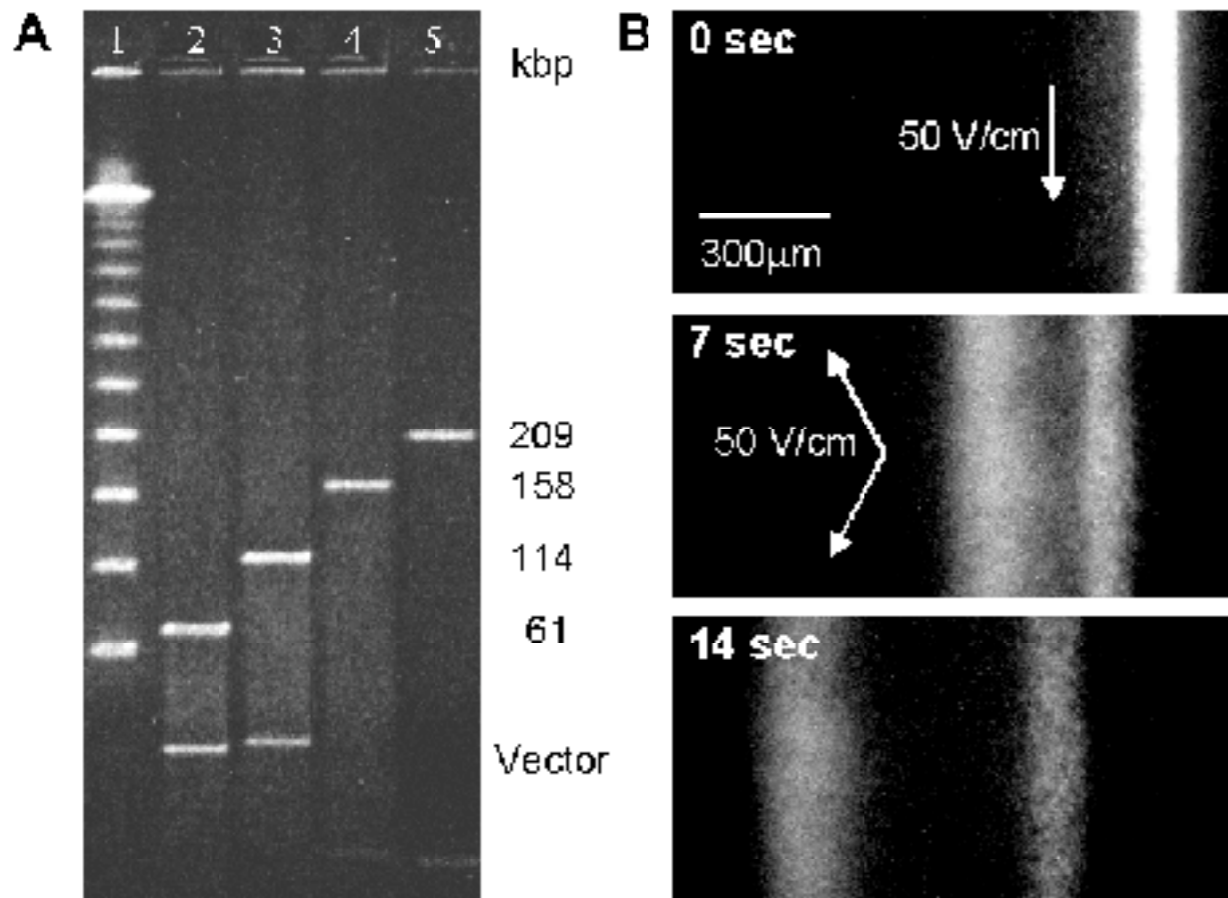30μm

400μm

3mm

50μm

Microfluidic channels

15μm
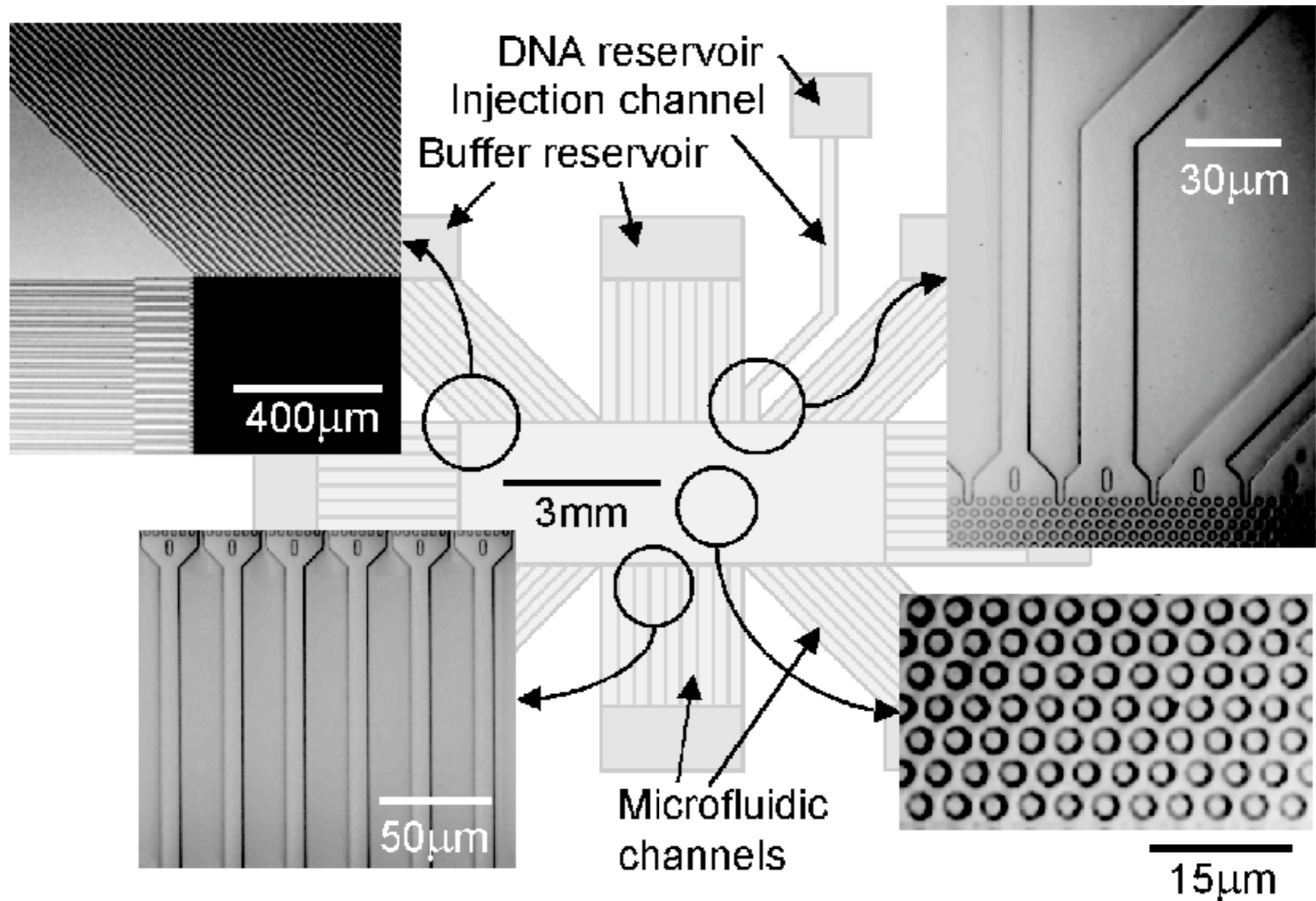
You live and die by your students. As I discussed yesterday, Richard Huang learned how to set the boundary conditions using current injector arrays.
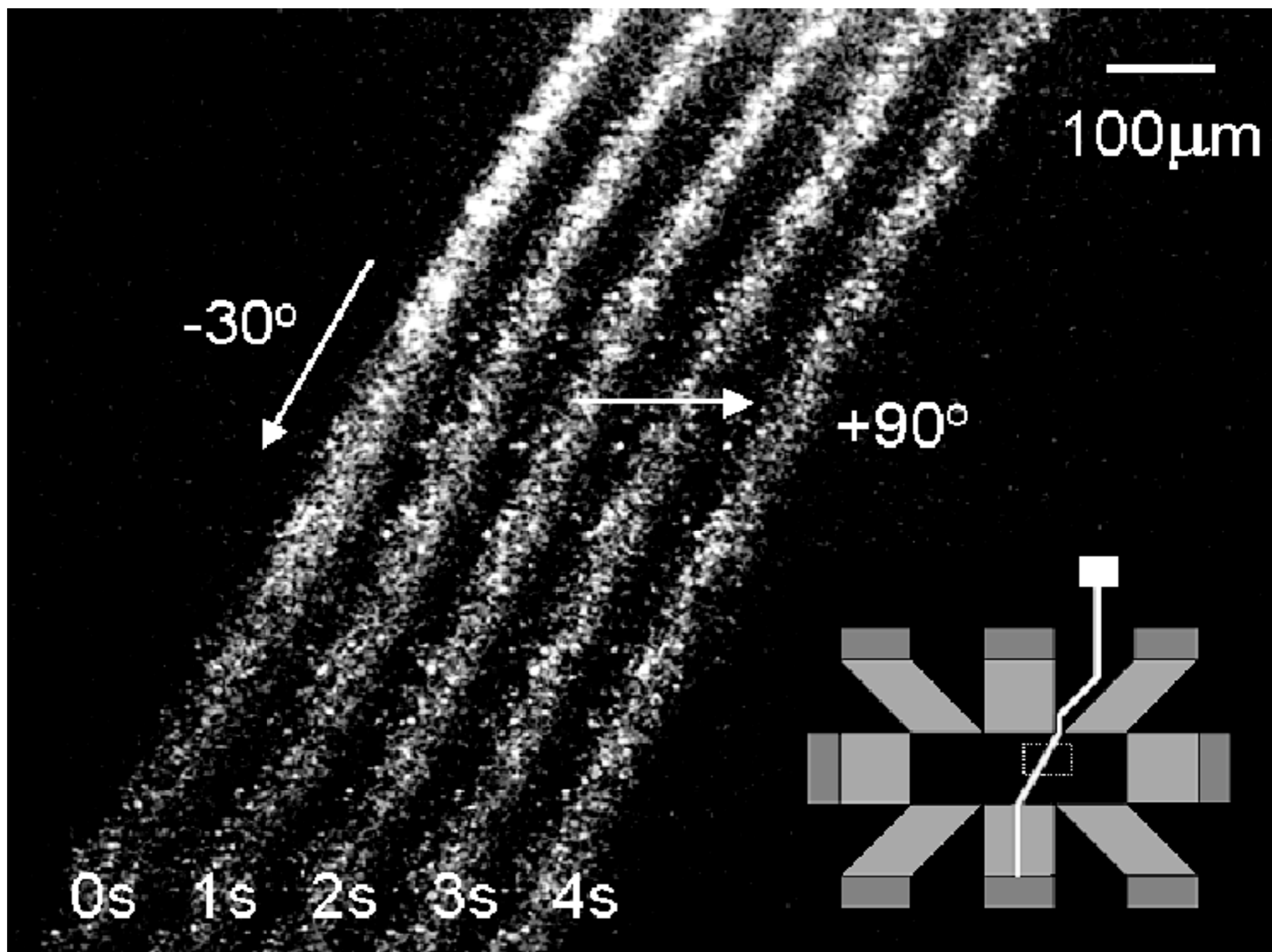
Now THAT'S field control!!!

Video of T4 and T7 DNA being fractionated in 2 seconds, at the few molecule level. This normally takes 2-5 hours in a gel, we nail it in 2 seconds
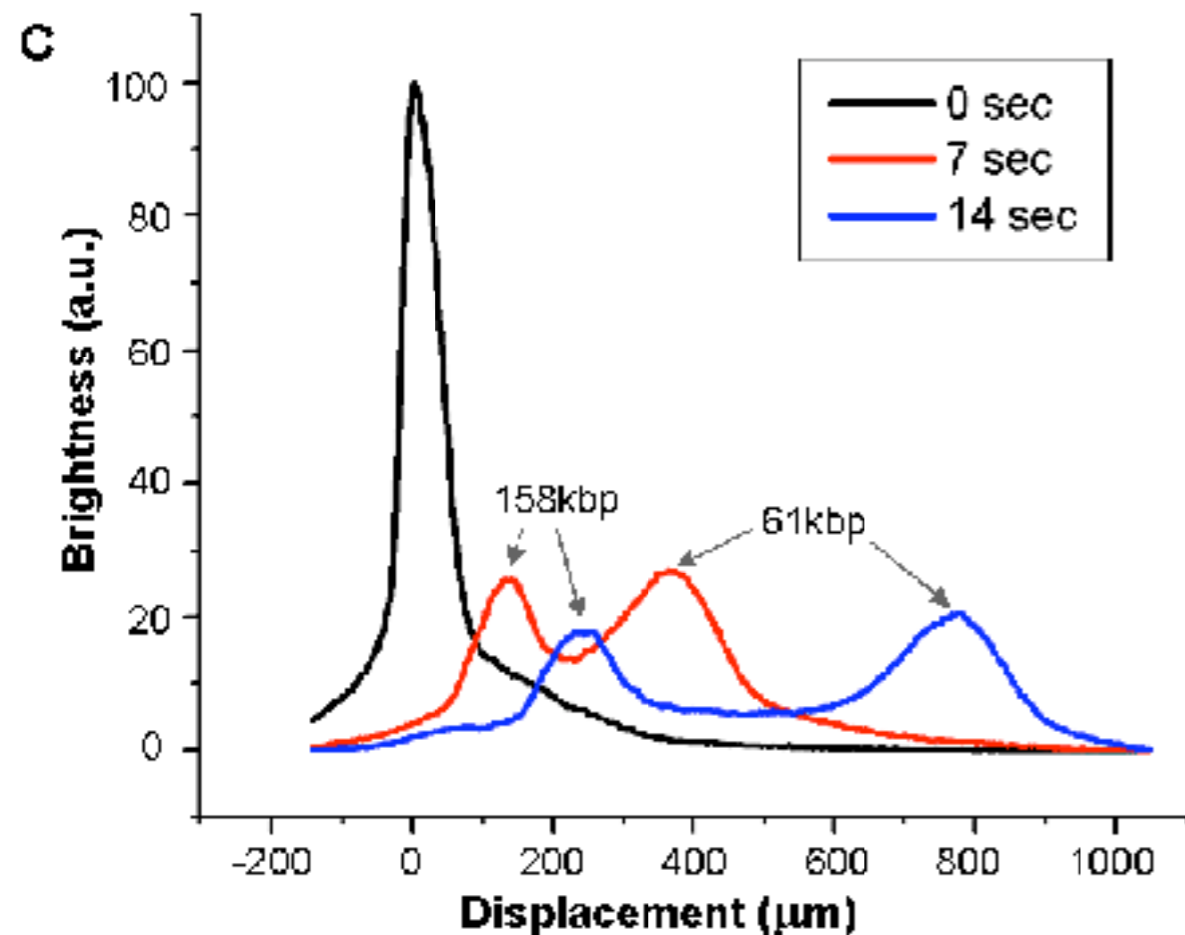
Beginning to seriously attack agarose gel resolution, in seconds rather than hours!

But wait. Richard had another idea for ASYMMETRICAL pulsing. He worked out the theory over a year before he tried it.   The idea is very, very simple.

Instead of making the AC fields SYMMETRICAL, make one stronger than the other by adding a DC component. Now, the jet stream of long molecules travels at an angle given by the direction of the strong pulses. HOWEVER, shorter molecules travel at an intermediate angle since they escape from posts.
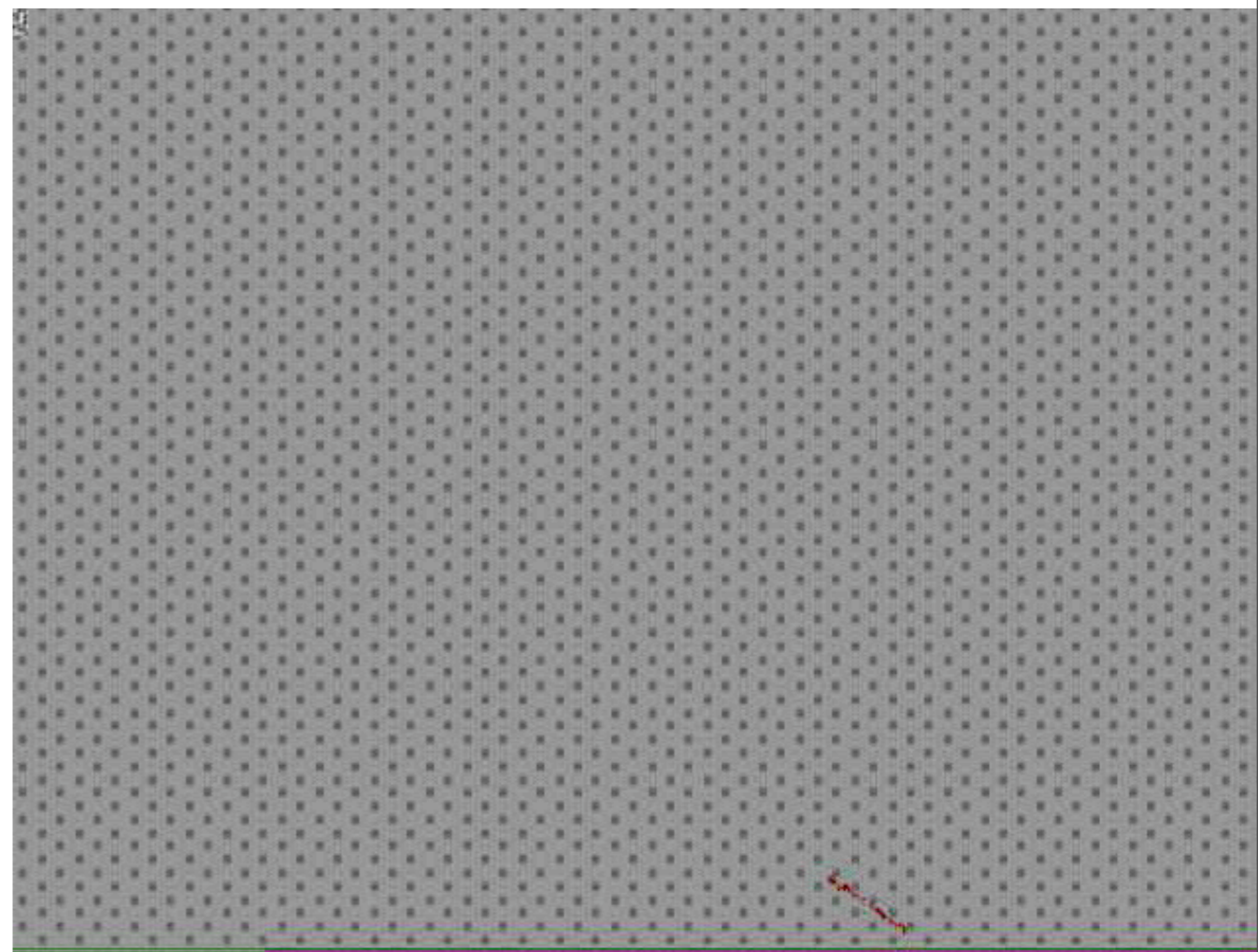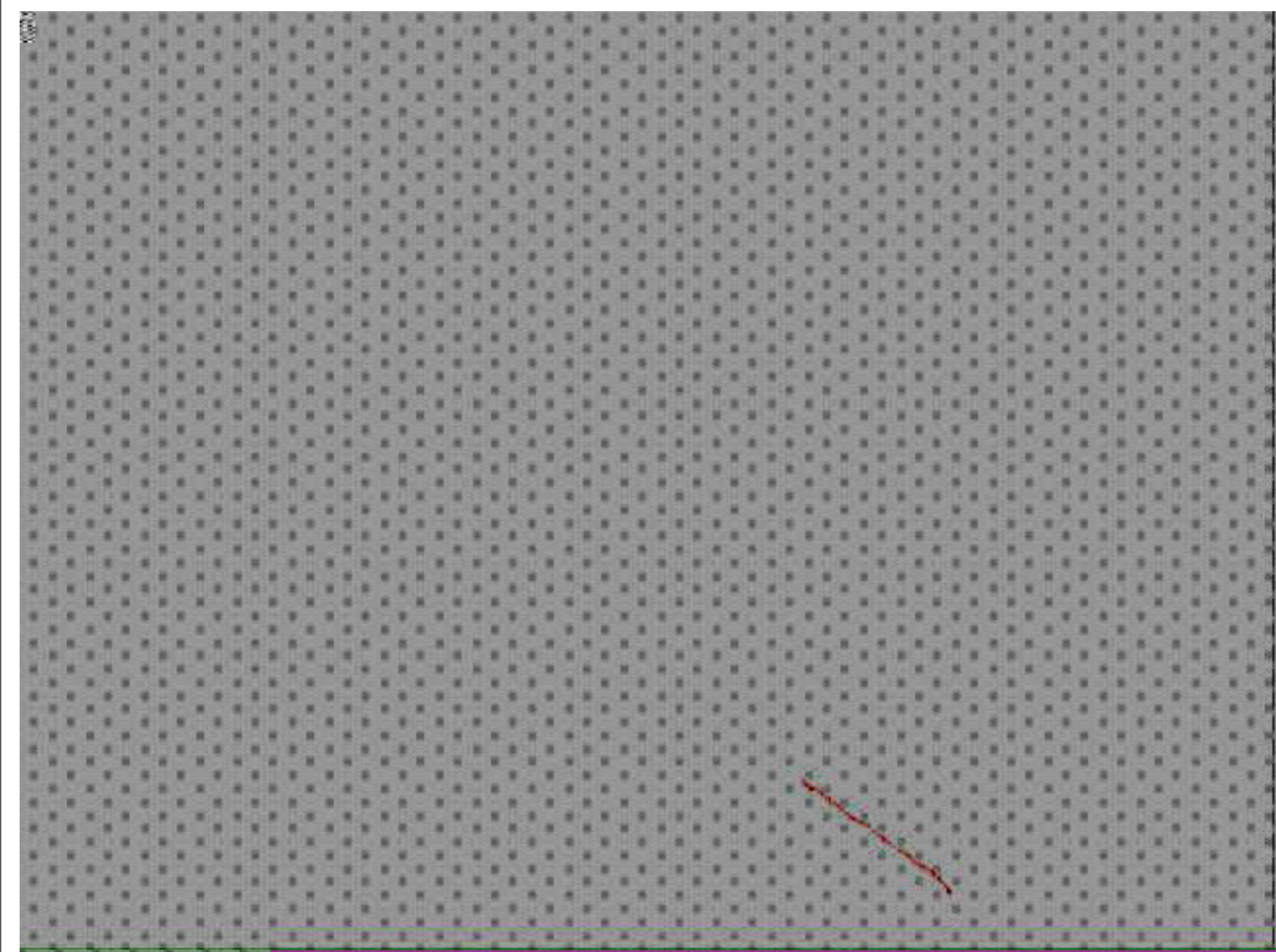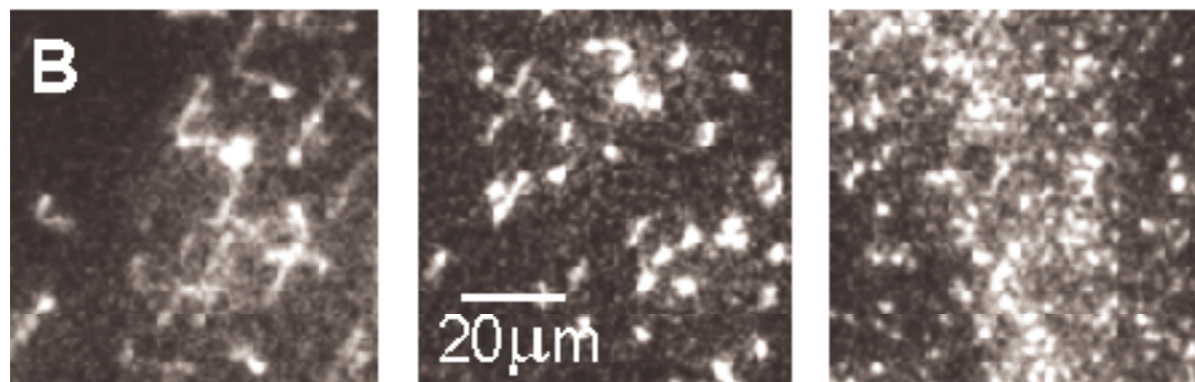
We call this a **DNA Prism**, since it acts like a prism separates different wavelengths of light. Simulations by Tom Duke.
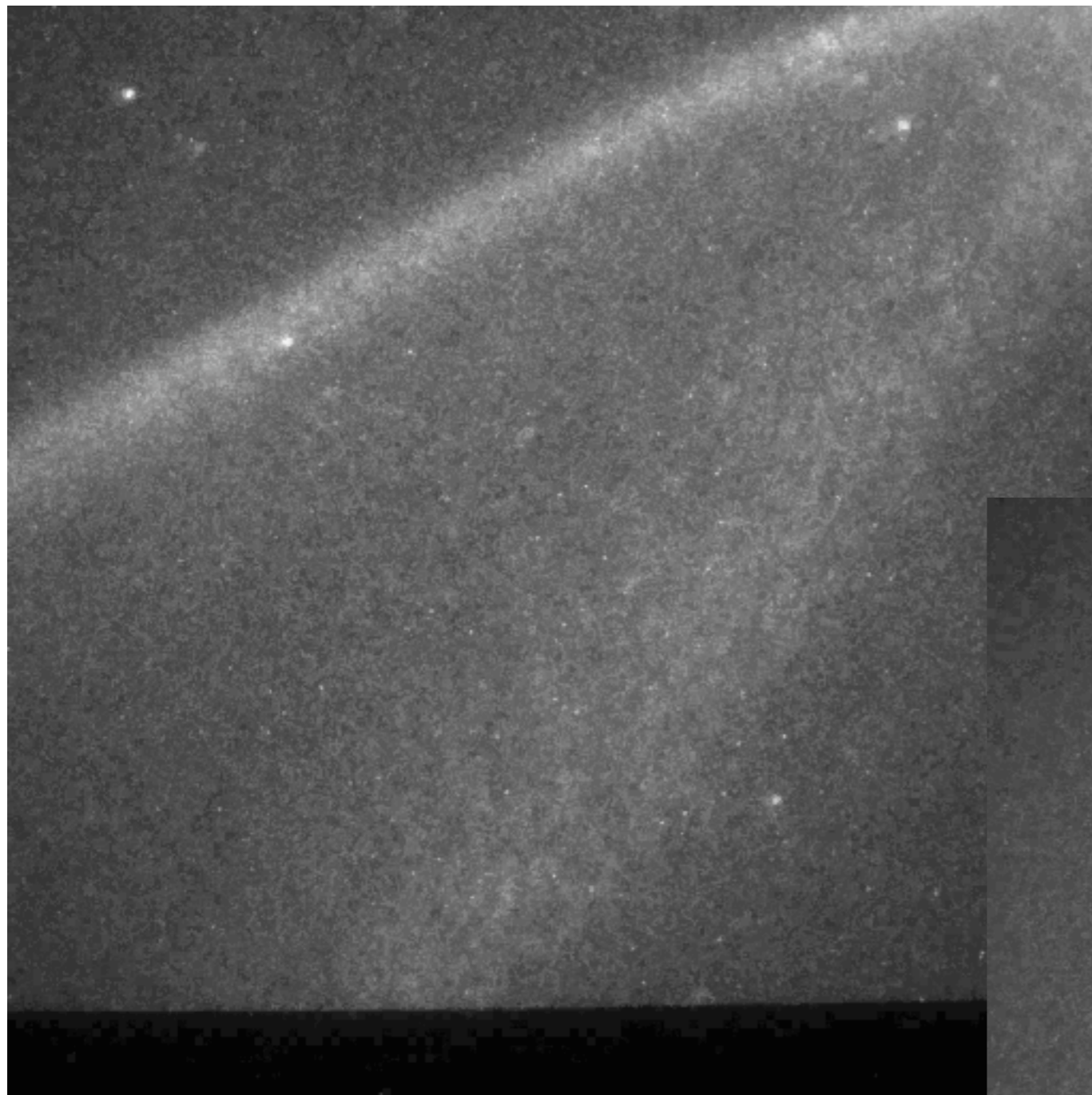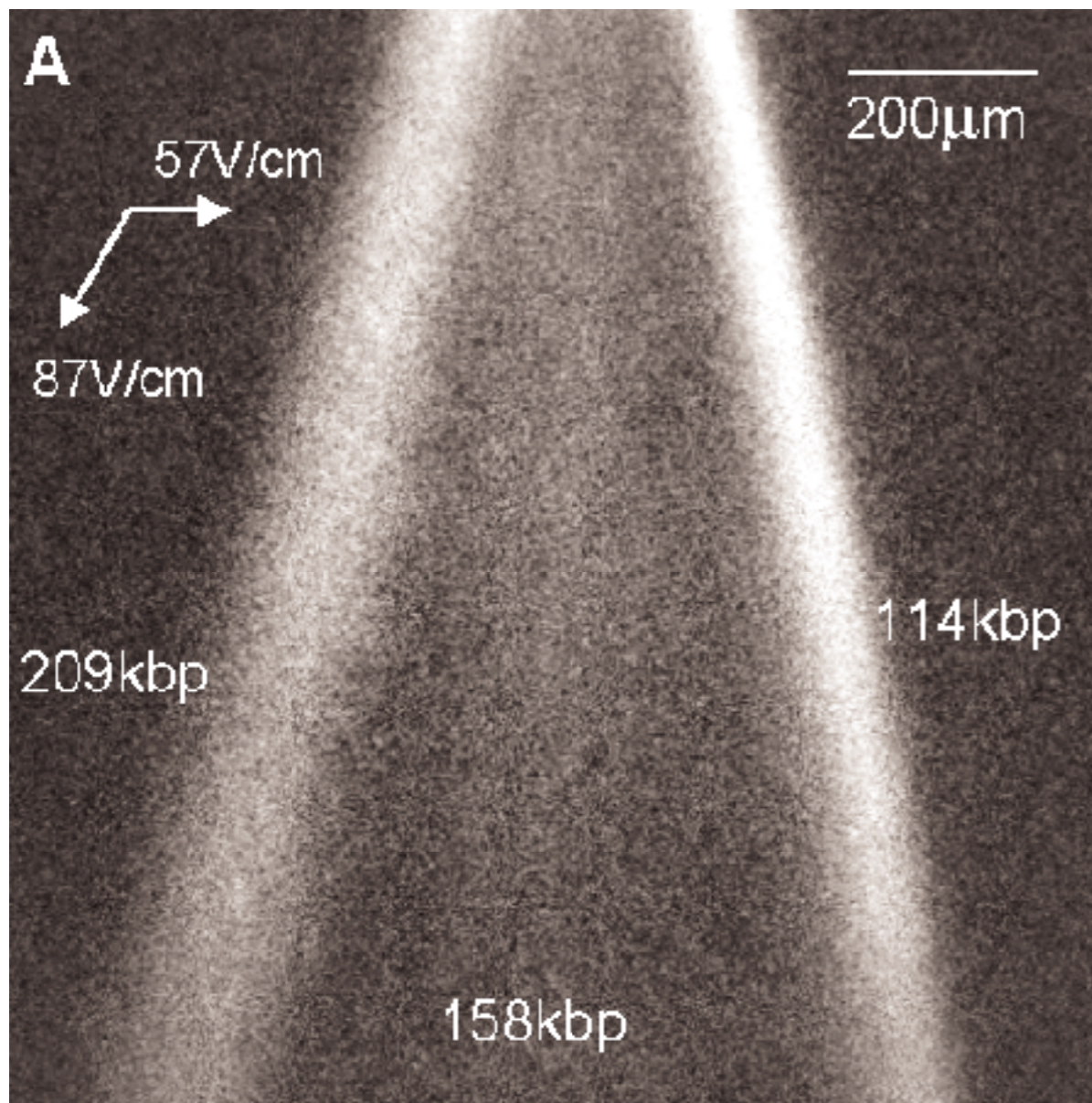
Bacterial artificial chromosome fractionation

Our next task is to move out of these demonstration projects that physicists love so much and biologists shrug their shoulders at and do serious fractionation of megabase genomic DNA on a time scale far faster than agarose gels can work, at the few molecule level. Extension of this technology down below 20 kbases will require much smaller arrays on the 100 nm scale and below, probably best done either with Professor Harold Craighead's (Cornell) ebeam lithography or Steve Chou's (Princeton) nanoinprinting technology.

But, the joke is on me because the ground has moved under my feet. While I was figuring all this out Dr. Venter went ahead and did what the **NIH** said he could not and sequenced the human genome using shot-gun sequencing and massive banks of capillary electrophoresis **DNA** separation, which I was assured would never work. Never say never! Read the damn journals! Try to connect into the network.

Table 2. Comparison of prices of large government projects circa 1990 with their projected useful life-span.

Proposed project	Projected cost ($ billion)	Target completion date	Estimated life-span (years)
Space Station Freedom	30.0	1999	30
Earth Observing System	17.0	2000	15
Superconducting Super Collider	11.0	1999	30
Human Genome Project	3.0	2005	Perpetual
Hubble Space Telescope	1.5	1990	15 to 20

(there seems to be an inverse correlation between cost and usefulness.....)
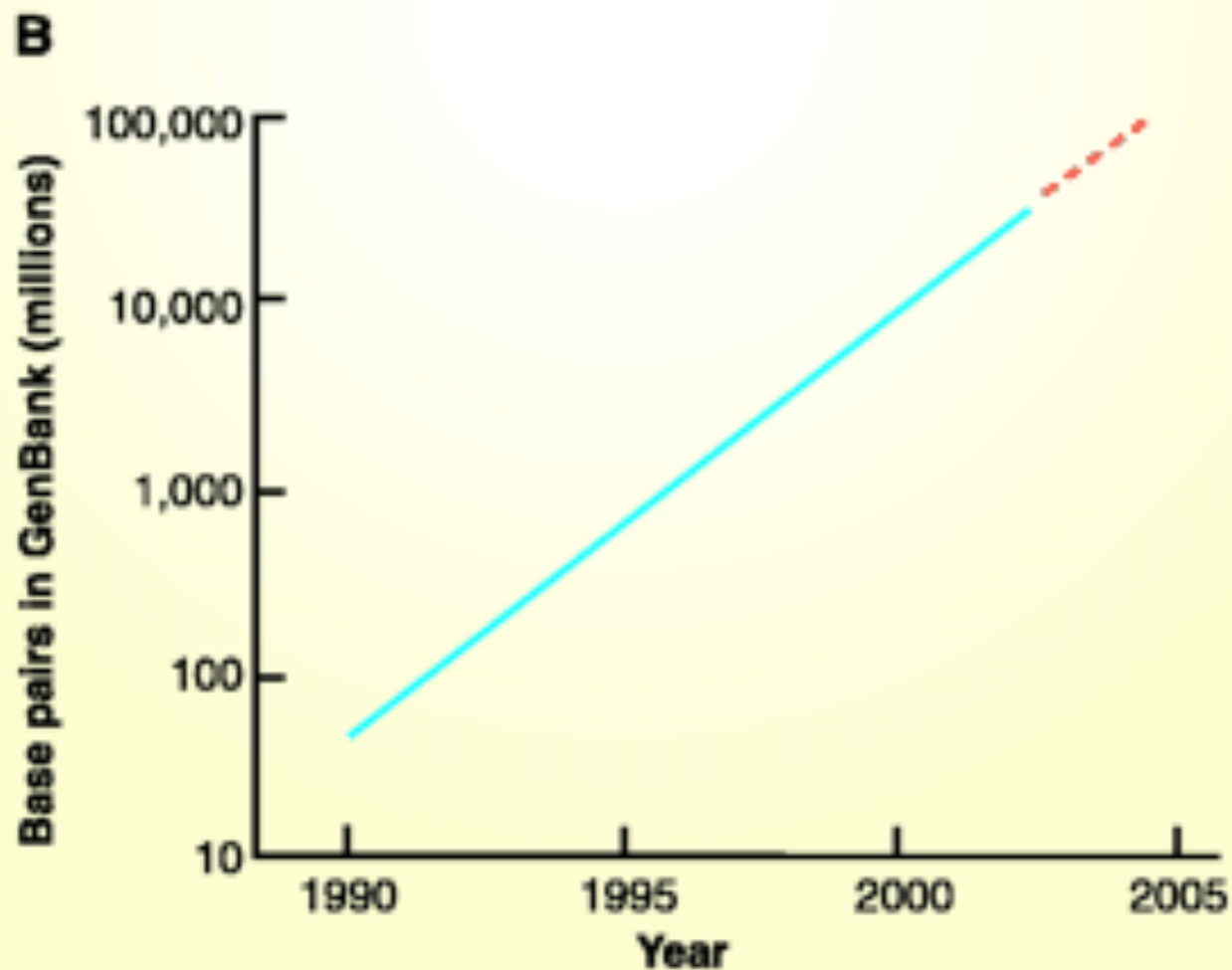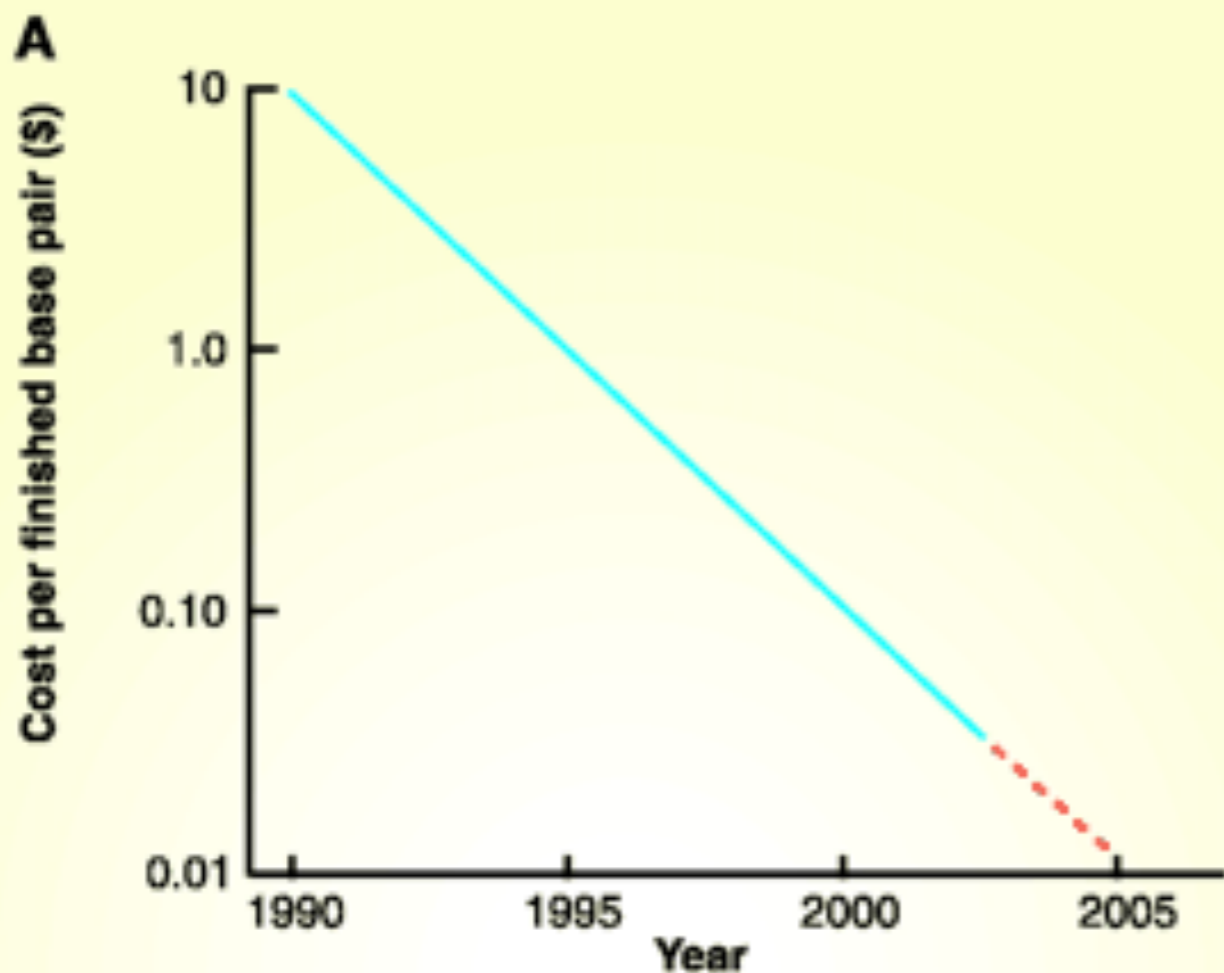
HOWEVER:

Molecular Biology may face about the same Moore's Law as semiconductor companies face: cost/base has been decreasing by factor of 2 every 1.5 years

Collins et al., Science

I'll talk about Plan B tomorrow: nanochannels.

Thanks!