Title: Introduction to Quantum Information and Computation from a Foundational Standpoint

Date: Nov 16, 2006 10:30 AM

URL: http://pirsa.org/06110034

Abstract: Quantum Information and Entanglement Assisted Quantum Communication

Quantum Information and Computation

Jeffrey Bub

Department of Philosophy
University of Maryland
and
Perimeter Institute for Theoretical Physics
Waterloo, Canada

Classical Information and Shannon Entropy

- A communication set-up involves a transmitter or source of information, a (possibly noisy) channel, and a receiver.
- The source produces messages in the form of sequences of symbols from some alphabet, which Shannon represented mathematically as sequences of values of independent, identically distributed random variables.

Classical Information and Shannon Entropy

- The fundamental question considered by Shannon was how to quantify the minimal physical resources required to store messages produced by a source, so that they could be communicated via a channel without loss and reconstructed by a receiver.
- Shannon's source coding theorem (or noiseless channel coding theorem) answers this question.

Classical Information and Shannon Entropy

- The fundamental question considered by Shannon was how to quantify the minimal physical resources required to store messages produced by a source, so that they could be communicated via a channel without loss and reconstructed by a receiver.
- Shannon's source coding theorem (or noiseless channel coding theorem) answers this question.

Classical Information and Shannon Entropy

- The fundamental question considered by Shannon was how to quantify the minimal physical resources required to store messages produced by a source, so that they could be communicated via a channel without loss and reconstructed by a receiver.
- Shannon's source coding theorem (or noiseless channel coding theorem) answers this question.

Classical Information and Shannon Entropy

- Consider a source that produces long sequences (messages) composed of symbols from a finite alphabet $a_1, a_2, \ldots, a_k$, where the individual symbols are produced with probabilities $p_1, p_2, \ldots, p_k$.
- A given sequence of symbols is represented as a sequence of values of independent, identically distributed, discrete random variables $X_1, X_2, \ldots$. A typical sequence of length n, for large n, will contain close to $p_i n$ symbols a_i, for $i = 1, \ldots, n$.
- So the probability of a sufficiently long typical sequence (assuming independence) will be:

$$p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2)\ldots p(x_n) \approx p_1^{p_1 n} p_2^{p_2 n} \ldots p_k^{p_k n}$$

Classical Information and Shannon Entropy

- The probability of a sufficiently long typical sequence (assuming independence) is:

$$p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2)\ldots p(x_n) \approx p_1^{p_1 n} p_2^{p_2 n} \ldots p_k^{p_k n}$$

- Taking the logarithm of both sides (conventionally, in information theory, to the base 2) yields:

$$\log p(x_1, \ldots, x_n) \approx n \sum_i p_i \log p_i := -nH(X)$$

where $H(X) := -\sum_i p_i \log p_i$ is the Shannon entropy of the source.

Classical Information and Shannon Entropy

- We can think about information in Shannon's sense in various ways. Take $-\log p_i$, a decreasing function of p_i with a minimum value of 0 when $p_i = 1$ for some i, as a measure of the information associated with identifying the symbol a_i produced by an information source. Then $H(X) = -\sum_i p_i \log p_i$ is the average information gain, or the expectation value of the information gain associated with ascertaining the value of the random variable X.
- Alternatively, we can think of the entropy as a measure of the amount of uncertainty about X before we ascertain its value.

Classical Information and Shannon Entropy

- Since

$$p(x_1, \ldots, x_n) = 2^{-nH(X)}$$

for sufficiently long typical sequences, and the probability of all the typical n-length sequences is less than 1, it follows that there are at most $2^{nH(X)}$ typical sequences.

- So each typical n-sequence could be encoded as a distinct binary number of nH(X) binary digits or bits before being sent through the channel to the receiver, where the original sequence could then be reconstructed by inverting the 1–1 encoding map. (The reconstruction would fail, with low probability, only for the rare atypical sequences, each of which could be encoded as, say, a string of 0's.)

Classical Information and Shannon Entropy

- The probability of a sufficiently long typical sequence (assuming independence) is:

$$p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2)\ldots p(x_n) \approx p_1^{p_1 n} p_2^{p_2 n} \ldots p_k^{p_k n}$$

- Taking the logarithm of both sides (conventionally, in information theory, to the base 2) yields:

$$\log p(x_1, \ldots, x_n) \approx n \sum_i p_i \log p_i := -nH(X)$$

where $H(X) := -\sum_i p_i \log p_i$ is the Shannon entropy of the source.

Classical Information and Shannon Entropy

- We can think about information in Shannon's sense in various ways. Take $-\log p_i$, a decreasing function of p_i with a minimum value of 0 when $p_i = 1$ for some i, as a measure of the information associated with identifying the symbol a_i produced by an information source. Then $H(X) = -\sum_i p_i \log p_i$ is the average information gain, or the expectation value of the information gain associated with ascertaining the value of the random variable X.
- Alternatively, we can think of the entropy as a measure of the amount of uncertainty about X before we ascertain its value.

Classical Information and Shannon Entropy

- The probability of a sufficiently long typical sequence (assuming independence) is:

$$p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2)\ldots p(x_n) \approx p_1^{p_1 n} p_2^{p_2 n} \ldots p_k^{p_k n}$$

- Taking the logarithm of both sides (conventionally, in information theory, to the base 2) yields:

$$\log p(x_1, \ldots, x_n) \approx n \sum_i p_i \log p_i := -nH(X)$$

where $H(X) := -\sum_i p_i \log p_i$ is the Shannon entropy of the source.

Classical Information and Shannon Entropy

- Consider a source that produces long sequences (messages) composed of symbols from a finite alphabet $a_1, a_2, \ldots, a_k$, where the individual symbols are produced with probabilities $p_1, p_2, \ldots, p_k$.
- A given sequence of symbols is represented as a sequence of values of independent, identically distributed, discrete random variables $X_1, X_2, \ldots$. A typical sequence of length n, for large n, will contain close to $p_i n$ symbols a_i, for $i = 1, \ldots, n$.
- So the probability of a sufficiently long typical sequence (assuming independence) will be:

$$p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2)\ldots p(x_n) \approx p_1^{p_1 n} p_2^{p_2 n} \ldots p_k^{p_k n}$$

Classical Information and Shannon Entropy

- The probability of a sufficiently long typical sequence (assuming independence) is:

$$p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2)\ldots p(x_n) \approx p_1^{p_1 n} p_2^{p_2 n} \ldots p_k^{p_k n}$$

- Taking the logarithm of both sides (conventionally, in information theory, to the base 2) yields:

$$\log p(x_1, \ldots, x_n) \approx n \sum_i p_i \log p_i := -nH(X)$$

where $H(X) := -\sum_i p_i \log p_i$ is the Shannon entropy of the source.

Classical Information and Shannon Entropy

- We can think about information in Shannon's sense in various ways. Take $-\log p_i$, a decreasing function of p_i with a minimum value of 0 when $p_i = 1$ for some i, as a measure of the information associated with identifying the symbol a_i produced by an information source. Then $H(X) = -\sum_i p_i \log p_i$ is the average information gain, or the expectation value of the information gain associated with ascertaining the value of the random variable X.
- Alternatively, we can think of the entropy as a measure of the amount of uncertainty about X before we ascertain its value.

Classical Information and Shannon Entropy

- Since

$$p(x_1, \ldots, x_n) = 2^{-nH(X)}$$

for sufficiently long typical sequences, and the probability of all the typical n-length sequences is less than 1, it follows that there are at most $2^{nH(X)}$ typical sequences.

- So each typical n-sequence could be encoded as a distinct binary number of nH(X) binary digits or bits before being sent through the channel to the receiver, where the original sequence could then be reconstructed by inverting the 1–1 encoding map. (The reconstruction would fail, with low probability, only for the rare atypical sequences, each of which could be encoded as, say, a string of 0's.)

Classical Information and Shannon Entropy

- Notice that if the probabilities p_i are all equal ($p_i = 1/k$ for all i), then $H(X) = \log k$, and if some $p_j = 1$ (and so $p_i = 0$ for $i \neq j$), then $H(X) = 0$ (taking $0 \log 0 = \lim_{x \to 0} x \log x = 0$).
- It can easily be shown that:

$$0 \leq H(X) \leq \log k$$

Classical Information and Shannon Entropy

- We can think about information in Shannon's sense in various ways. Take $-\log p_i$, a decreasing function of p_i with a minimum value of 0 when $p_i = 1$ for some i, as a measure of the information associated with identifying the symbol a_i produced by an information source. Then $H(X) = -\sum_i p_i \log p_i$ is the average information gain, or the expectation value of the information gain associated with ascertaining the value of the random variable X.
- Alternatively, we can think of the entropy as a measure of the amount of uncertainty about X before we ascertain its value.

Classical Information and Shannon Entropy

- Since

$$p(x_1, \ldots, x_n) = 2^{-nH(X)}$$

for sufficiently long typical sequences, and the probability of all the typical n-length sequences is less than 1, it follows that there are at most $2^{nH(X)}$ typical sequences.

- So each typical n-sequence could be encoded as a distinct binary number of nH(X) binary digits or bits before being sent through the channel to the receiver, where the original sequence could then be reconstructed by inverting the 1–1 encoding map. (The reconstruction would fail, with low probability, only for the rare atypical sequences, each of which could be encoded as, say, a string of 0's.)

Classical Information and Shannon Entropy

- Notice that if the probabilities p_i are all equal ($p_i = 1/k$ for all i), then $H(X) = \log k$, and if some $p_j = 1$ (and so $p_i = 0$ for $i \neq j$), then $H(X) = 0$ (taking $0 \log 0 = \lim_{x \to 0} x \log x = 0$).
- It can easily be shown that:

$$0 \leq H(X) \leq \log k$$

Classical Information and Shannon Entropy

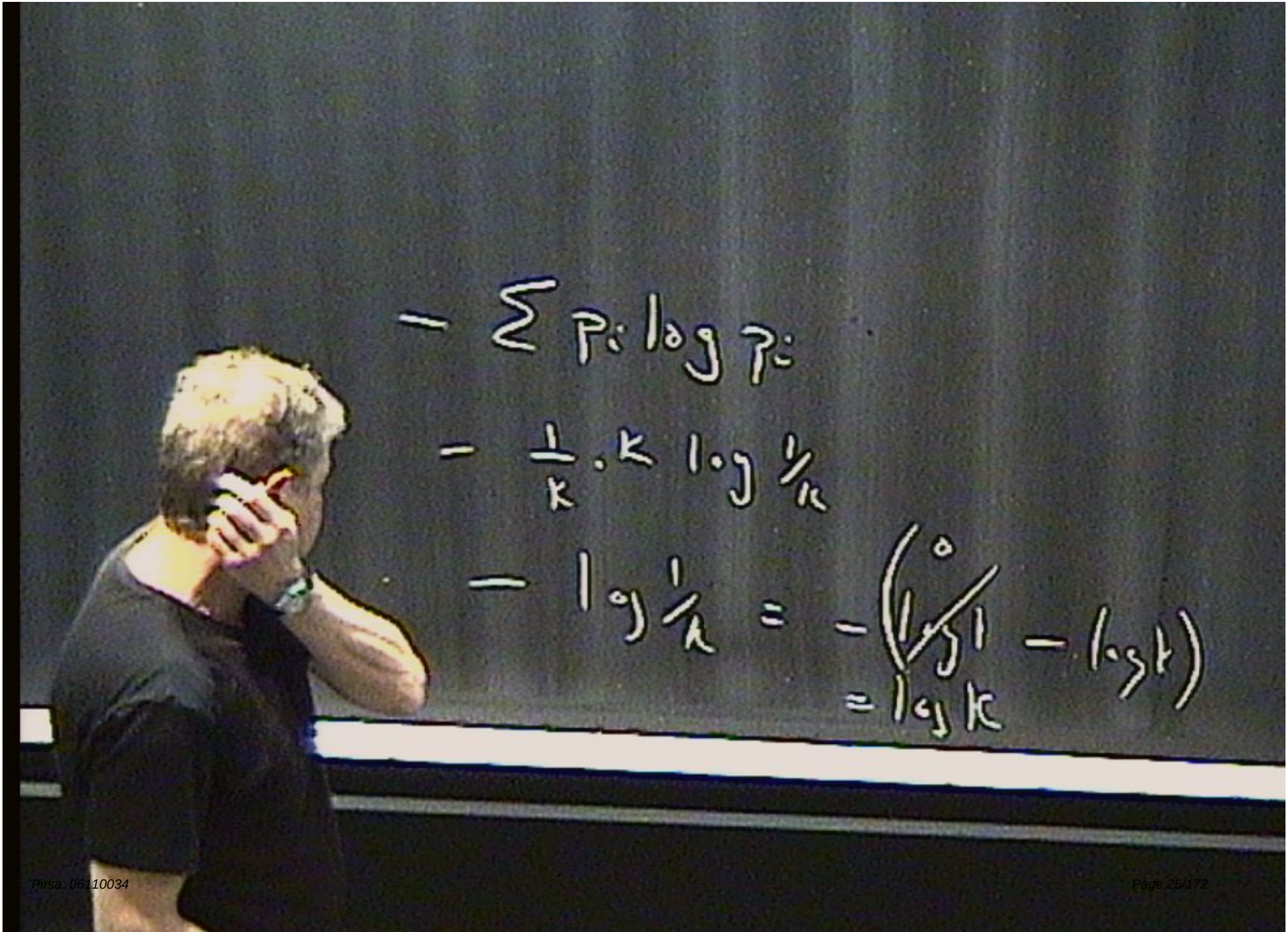
- If we encoded each of the k distinct symbols as a distinct binary number, i.e., as a distinct string of 0's and 1's, we would need binary numbers composed of $\log k$ bits to represent each symbol ($2^{\log k} = k$).
- So Shannon's analysis shows that messages produced by a stochastic source can be compressed, in the sense that (as $n \to \infty$ and the probability of an atypical n-length sequence tends to zero) n-length sequences can be encoded without loss of information using $nH(X)$ bits rather than the $n \log k$ bits required if we encoded each of the k symbols a_i as a distinct string of 0's and 1's: this is a compression, since $nH(X) < n \log k$ except for equiprobable distributions.

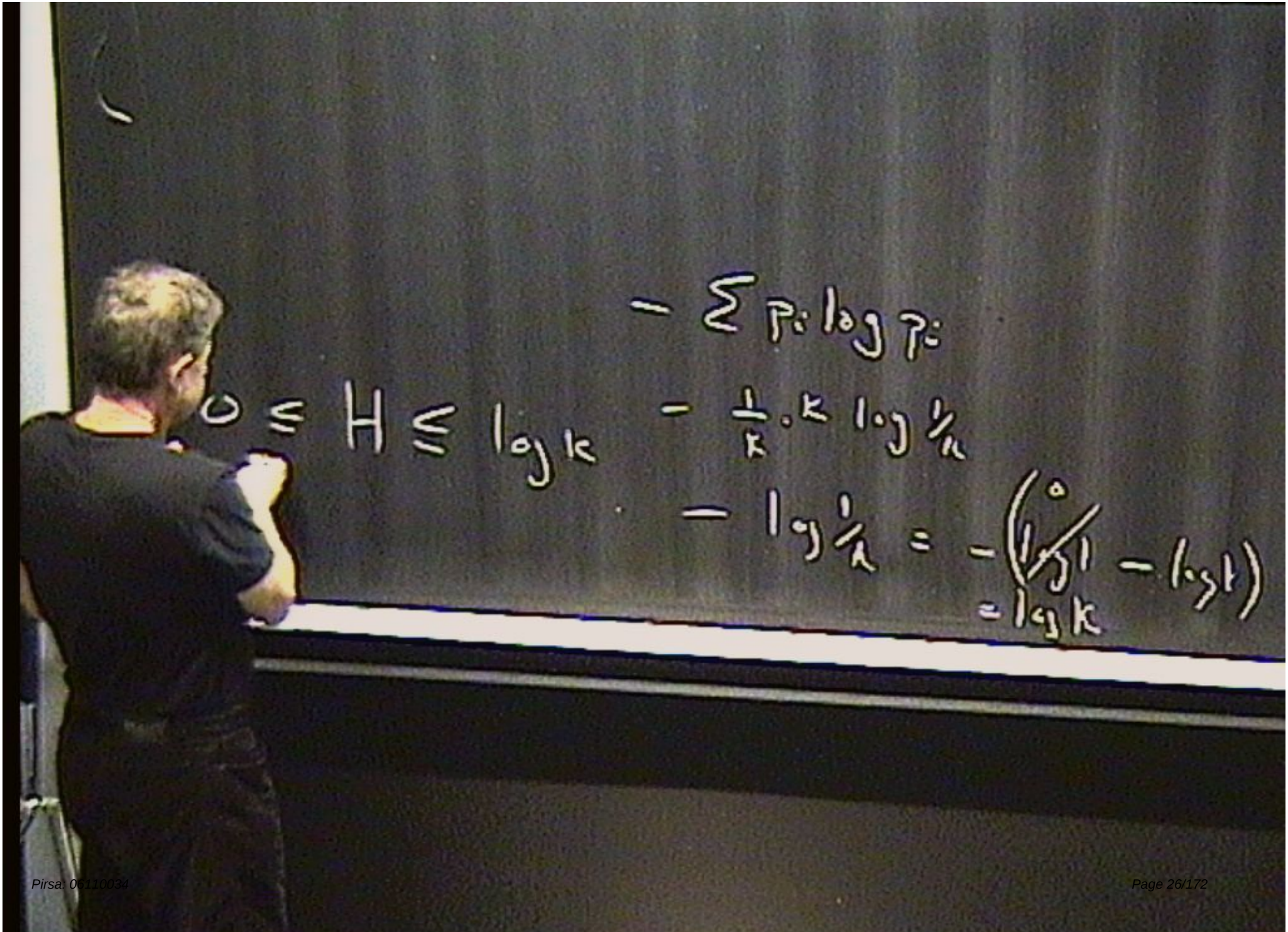
Classical Information and Shannon Entropy

- Shannon's source coding theorem (noiseless channel coding theorem) shows that the compression rate of H(X) bits per symbol produced by a source of independent and identically distributed random variables is optimal.
- The source produces n-length sequences of symbols $x_1, x_2, \ldots, x_n$ with probability $p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2)\ldots p(x_n)$, where each symbol is chosen from an alphabet $\mathcal{X}$. If there are k symbols in $\mathcal{X}$, these n-sequences can be represented as sequences of $n \log k$ bits.

Classical Information and Shannon Entropy

- If we encoded each of the k distinct symbols as a distinct binary number, i.e., as a distinct string of 0's and 1's, we would need binary numbers composed of $\log k$ bits to represent each symbol ($2^{\log k} = k$).
- So Shannon's analysis shows that messages produced by a stochastic source can be compressed, in the sense that (as $n \to \infty$ and the probability of an atypical n-length sequence tends to zero) n-length sequences can be encoded without loss of information using $nH(X)$ bits rather than the $n \log k$ bits required if we encoded each of the k symbols a_i as a distinct string of 0's and 1's: this is a compression, since $nH(X) < n \log k$ except for equiprobable distributions.


$-\sum p_i \log p_i$
$-\frac{1}{k} \cdot k \log \frac{1}{k}$
$-\log \frac{1}{k} = -(\log 1 - \log k)$
$= \log k$


$-\sum p_i \log p_i$
$0 \leq H \leq \log k$
$-\frac{1}{k} \cdot k \log \frac{1}{k}$
$-\log \frac{1}{k} = -(\log 1 - \log k)$
$= \log k$

Classical Information and Shannon Entropy

- If we encoded each of the k distinct symbols as a distinct binary number, i.e., as a distinct string of 0's and 1's, we would need binary numbers composed of $\log k$ bits to represent each symbol ($2^{\log k} = k$).
- So Shannon's analysis shows that messages produced by a stochastic source can be compressed, in the sense that (as $n \to \infty$ and the probability of an atypical n-length sequence tends to zero) n-length sequences can be encoded without loss of information using $nH(X)$ bits rather than the $n \log k$ bits required if we encoded each of the k symbols a_i as a distinct string of 0's and 1's: this is a compression, since $nH(X) < n \log k$ except for equiprobable distributions.

Classical Information and Shannon Entropy

- Suppose there is a 'block coding' compression scheme that encodes each 'block' or n-length sequence (for sufficiently large n) as a shorter sequence of nR bits, where $0 \leq R \leq \log k$. Suppose also that the receiver has a decompression scheme for decoding sequences of nR bits into sequences of n symbols.
- One speaks of a compression/decompression scheme of rate R.

Classical Information and Shannon Entropy

- Suppose there is a 'block coding' compression scheme that encodes each 'block' or n-length sequence (for sufficiently large n) as a shorter sequence of nR bits, where $0 \leq R \leq \log k$. Suppose also that the receiver has a decompression scheme for decoding sequences of nR bits into sequences of n symbols.
- One speaks of a compression/decompression scheme of rate R.

Classical Information and Shannon Entropy

The source coding theorem states that

> if the Shannon entropy of a source is H(X), then there exists a reliable compression/decompression scheme of rate R if and only if $R \geq H(X)$, where a scheme is said to be reliable if it reproduces the original sequence with a probability that tends to 1 as $n \to \infty$.

Classical Information and Shannon Entropy

- For reliable communication, we want the compression and decompression of a sequence of symbols to yield the original sequence, but in general there will be a certain probability, $q(x_1, \ldots, x_n)$, of decoding a given sequence of nR encoded bits received by the receiver as the original n-sequence produced by the source.
- The average fidelity of a compression/decompression scheme for n-length blocks is defined as:

$$F_n = \sum_{\text{all } n\text{-sequences}} p(x_1, \ldots, x_n) q(x_1, \ldots, x_n)$$

If all the probabilities $q(x_1, \ldots, x_n)$ are 1, $F_n = 1$; otherwise $F_n < 1$.

Classical Information and Shannon Entropy

- As a simple example of compression, consider an information source that produces sequences of symbols from a 4-symbol alphabet a_1, a_2, a_3, a_4 with probabilities 1/2, 1/4, 1/8, 1/8. Each symbol can be represented by a distinct 2-digit binary number:

$$\begin{aligned} a_1 &: 00 \\ a_2 &: 01 \\ a_3 &: 10 \\ a_4 &: 11 \end{aligned}$$

- Without compression we need two bits per symbol of storage space to store the output of the source.

Classical Information and Shannon Entropy

- The Shannon entropy of the source is $\mathrm{H(X)} = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{8}\log\frac{1}{8} = \frac{7}{4}$.
- Shannon's source coding theorem tells us that there is a compression scheme that uses an average of 7/4 bits per symbol rather than two bits per symbol, and that such a compression scheme is optimal.
- The optimal scheme is provided by the following encoding:

$$\begin{aligned} a_1 &: 0 \\ a_2 &: 10 \\ a_3 &: 110 \\ a_4 &: 111 \end{aligned}$$

for which the average length of a compressed sequence is: $\frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 + \frac{1}{8}\cdot 3 + \frac{1}{8}\cdot 3 = \frac{7}{4}$ bits per symbol.

Classical Information and Shannon Entropy

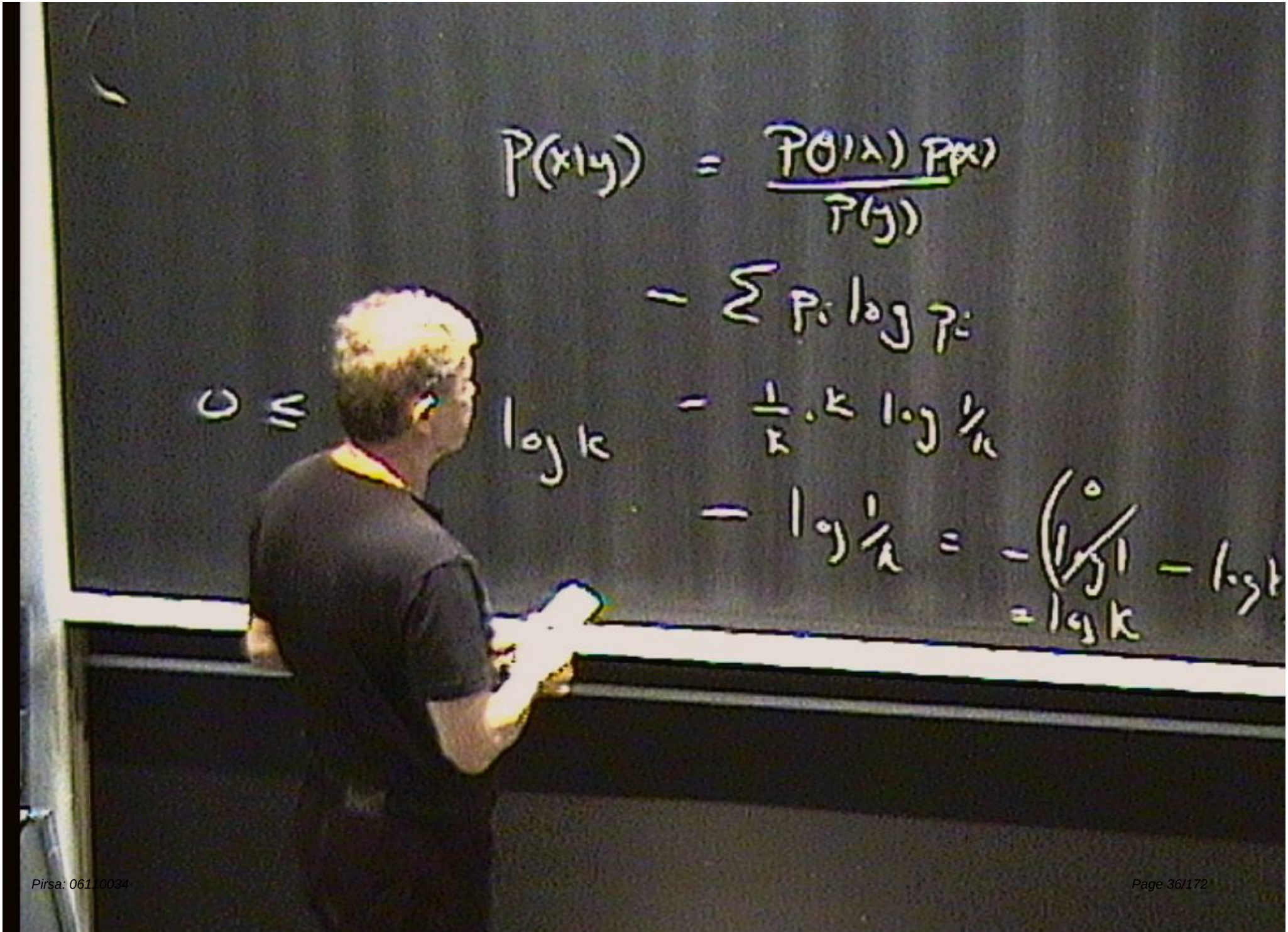
- The significance of Shannon's source coding theorem lies is showing that there is an optimal or most efficient way of compressing messages produced by a source (assuming a certain idealization) in such a way that they can be reliably reconstructed by a receiver.
- The Shannon entropy H(X) is a measure of the minimal physical resources, in terms of the average number of bits per symbol, that are necessary and sufficient to reliably store the output of a source of messages.
- In this sense, it is a measure of the amount of information per symbol produced by an information source.

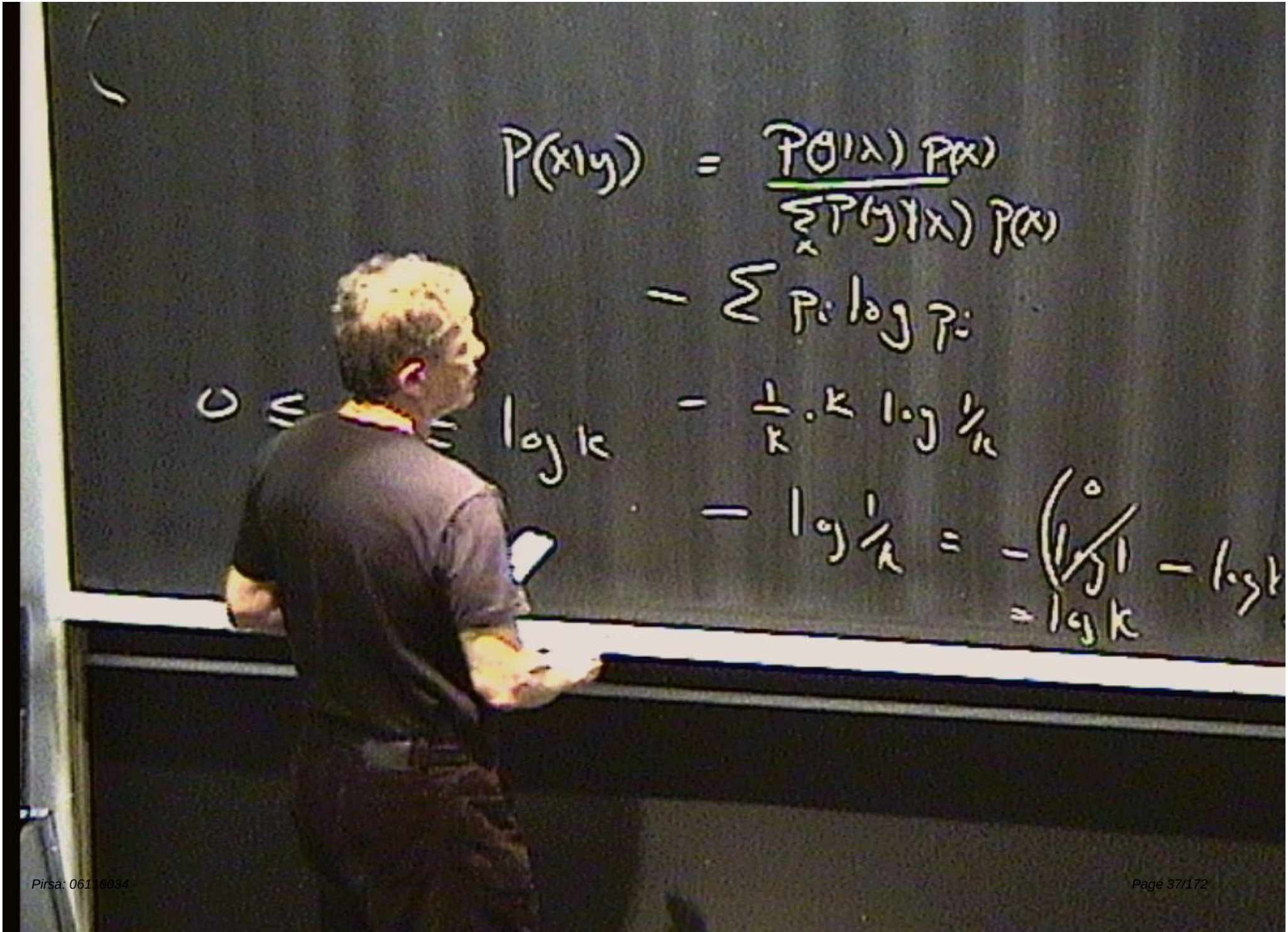
Conditional Entropy

- An information channel maps inputs consisting of values of a random variable X onto outputs consisting of values of a random variable Y, and the map will generally not be 1-1 if the channel is noisy.
- Consider the conditional probabilities $p(y|x)$ of obtaining an output value y for a given input value x, for all x, y.
- From the probabilities $p(x)$ we can calculate $p(y)$ as:

$$p(y) = \sum_x p(y|x)p(x)$$

and we can also calculate $p(x|y)$ by Bayes' rule from the probabilities $p(y|x)$ and $p(x)$, for all x, y, and hence the Shannon entropy of the conditional distribution $p(x|y)$, for all x and a fixed y, denoted by $H(X|Y = y)$.


P(x|y) = P(y|x) P(x) / P(y)
= − Σ p_i log p_i
0 ≤ log k − ½ . k log ½
− log ½ = − (log 1 − log k)
= log k


P(x|y) = P(y|x) P(x) / Σ_x P(y|x) P(x)
= − Σ p_i log p_i
0 ≤ = log k − (1/k) · k log 1/k
= − log 1/k = − (log 1 − log k)
= log k

Conditional Entropy

- An information channel maps inputs consisting of values of a random variable X onto outputs consisting of values of a random variable Y, and the map will generally not be 1-1 if the channel is noisy.
- Consider the conditional probabilities $p(y|x)$ of obtaining an output value y for a given input value x, for all x, y.
- From the probabilities $p(x)$ we can calculate $p(y)$ as:

$$p(y) = \sum_x p(y|x)p(x)$$

and we can also calculate $p(x|y)$ by Bayes' rule from the probabilities $p(y|x)$ and $p(x)$, for all x, y, and hence the Shannon entropy of the conditional distribution $p(x|y)$, for all x and a fixed y, denoted by $H(X|Y = y)$.

Conditional Entropy

- The quantity

$$H(X|Y) = \sum_y p(y) H(X|Y = y)$$

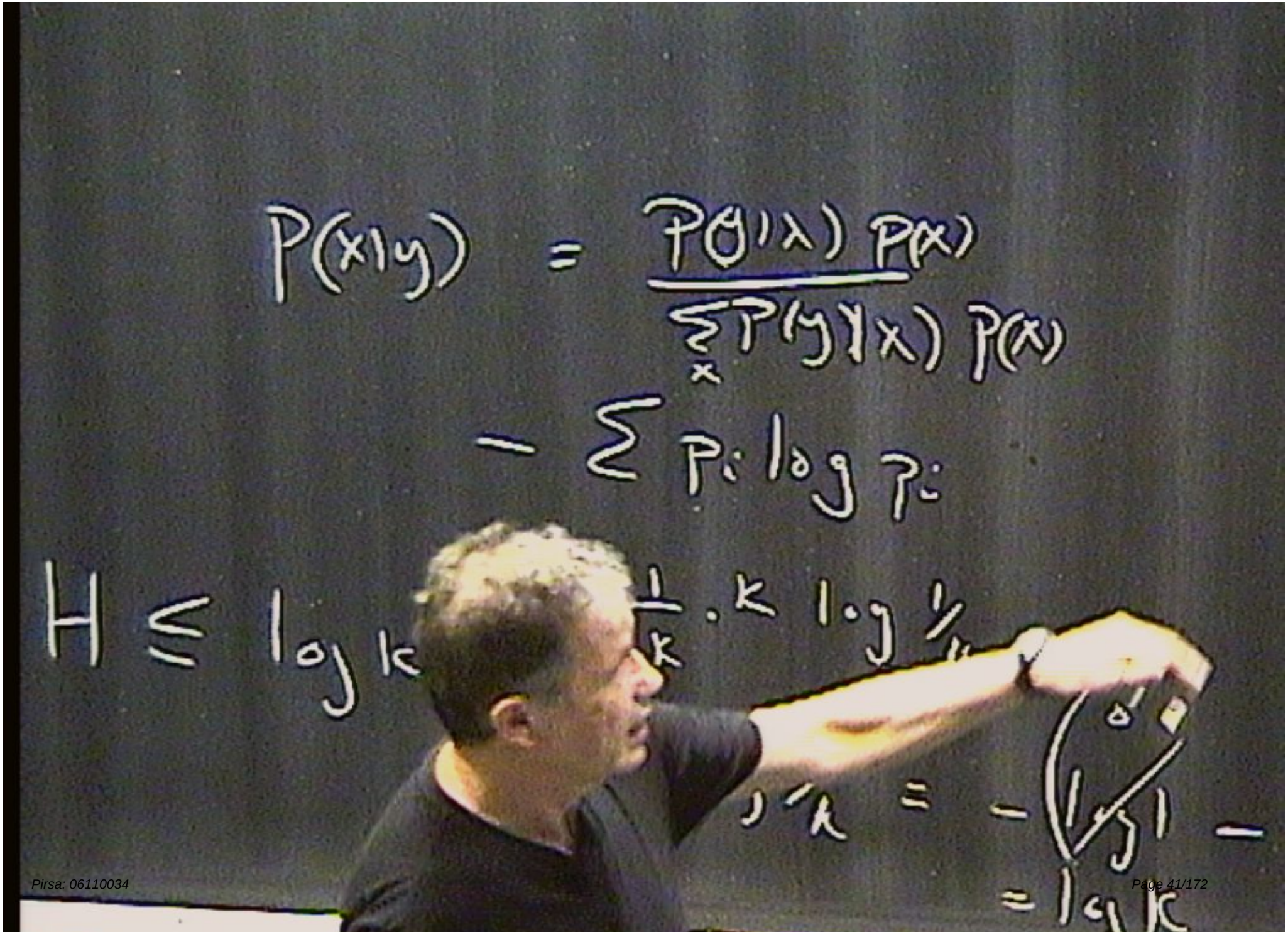
  is known as the conditional entropy.
- It is the expected value of $H(X|Y = y)$ for all y.
- If we think of H(X), the entropy of the distribution $\{p(x) : x \in \mathcal{X}\}$, as a measure of the uncertainty of the X-value, then $H(X|Y = y)$ is a measure of the uncertainty of the X-value, given the Y-value y, and H(X|Y) is a measure of the average uncertainty of the X-value, given a Y-value.

Conditional Entropy

- The quantity

$$H(X|Y) = \sum_y p(y) H(X|Y = y)$$

  is known as the conditional entropy.
- It is the expected value of $H(X|Y = y)$ for all y.
- If we think of $H(X)$, the entropy of the distribution $\{p(x) : x \in \mathcal{X}\}$, as a measure of the uncertainty of the X-value, then $H(X|Y = y)$ is a measure of the uncertainty of the X-value, given the Y-value y, and $H(X|Y)$ is a measure of the average uncertainty of the X-value, given a Y-value.


P(x|y) = P(y|x) P(x) / Σ_x P(y|x) P(x)
− Σ p_i log p_i
H ≤ log k
1/k . k log 1/k
J'_k = −(1/k log 1/k −
= log k

Conditional Entropy

- The quantity

$$H(X|Y) = \sum_y p(y) H(X|Y = y)$$

  is known as the conditional entropy.
- It is the expected value of $H(X|Y = y)$ for all y.
- If we think of $H(X)$, the entropy of the distribution $\{p(x) : x \in \mathcal{X}\}$, as a measure of the uncertainty of the X-value, then $H(X|Y = y)$ is a measure of the uncertainty of the X-value, given the Y-value y, and $H(X|Y)$ is a measure of the average uncertainty of the X-value, given a Y-value.

Conditional Entropy

- Putting it differently, the number of input sequences of length n that are consistent with a given output sequence (as $n \to \infty$) is $2^{nH(X|Y)}$, i.e., $H(X|Y)$ is the number of bits per symbol of additional information needed, on average, to identify an input X-sequence from a given Y-sequence.
- This follows because there are $2^{nH(X,Y)}$ typical sequences of pairs (x, y), where the joint entropy $H(X, Y)$ is calculated from the joint probability $p(x, y)$.
- So there are

$$\frac{2^{nH(X,Y)}}{2^{nH(Y)}} = 2^{n(H(X,Y)-H(Y))} = 2^{nH(X|Y)}$$

typical X-sequences associated with a given Y-sequence.

Conditional Entropy

- The equality

$$\mathrm{H}(\mathrm{X},\mathrm{Y}) - \mathrm{H}(\mathrm{Y}) = \mathrm{H}(\mathrm{X}|\mathrm{Y})$$

follows from the 'chain rule' equality

$$\mathrm{H}(\mathrm{X},\mathrm{Y}) = \mathrm{H}(\mathrm{X}) + \mathrm{H}(\mathrm{Y}|\mathrm{X}) = \mathrm{H}(\mathrm{Y}) + \mathrm{H}(\mathrm{X}|\mathrm{Y}) = \mathrm{H}(\mathrm{Y},\mathrm{X})$$

which is easily derived from the logarithmic definitions of the quantities.

- Note that $\mathrm{H}(\mathrm{X}|\mathrm{Y}) \neq \mathrm{H}(\mathrm{Y}|\mathrm{X})$.

Mutual Information

- The mutual information $\mathrm{H(X:Y)}$(sometimes $\mathrm{I(X:Y)}$) measures the average amount of information gained about X by ascertaining a Y-value, i.e., the amount of information one random variable contains about another, or the reduction in uncertainty of one random variable obtained by measuring another.
- Mutual information can be defined in terms of the concept of relative entropy, which is a measure of something like the distance between two probability distributions (although it is not a true metric, since it is not symmetric and does not satisfy the triangle inequality).

Mutual Information

- The relative entropy between distributions p(x) and q(x) is defined as:

$$D(p \parallel q) = \sum_{x \in \mathcal{A}} p(x) \log \frac{p(x)}{q(x)}$$

- The mutual information can now be defined as:

$$\begin{aligned} H(X:Y) &= D(p(x,y) \parallel p(x)p(y)) \\ &= \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \end{aligned}$$

Mutual Information

- The relative entropy between distributions $p(x)$ and $q(x)$ is defined as:

$$D(p \parallel q) = \sum_{x \in \mathcal{A}} p(x) \log \frac{p(x)}{q(x)}$$

- The mutual information can now be defined as:

$$\begin{aligned} H(X:Y) &= D(p(x,y) \parallel p(x)p(y)) \\ &= \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \end{aligned}$$

Mutual Information

- It follows that

$$H(X:Y) = H(X) + H(Y) - H(X,Y)$$

  i.e., the mutual information of two random variables is a measure of how much information they have in common: the sum of the information content of the two random variables, as measured by the Shannon entropy (in which joint information is counted twice), minus their joint information.

- Note that $H(X:X) = H(X)$, as we would expect.

Mutual Information

- Since $H(X, Y) = H(X) + H(Y|X)$, it follows that

$$H(X:Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

i.e., the mutual information of two random variables represents the average information gain about one random variable obtained by measuring the other: the difference between the initial uncertainty of one of the random variables, and the average residual uncertainty of that random variable after ascertaining the value of the other random variable.

Channel Capacity

- For a noisy channel, if X represents the input to the channel and Y represents the output of the channel, H(X:Y) represents the average amount of information gained about the input X by ascertaining the value of the output Y.
- The capacity of a channel, C, is defined as the supremum of H(X:Y) over all input distributions.

Channel Capacity

- Shannon's noisy channel coding theorem shows, perhaps surprisingly, that up to C bits of information can be sent through a noisy channel with arbitrary low error rate. That is, there exists an optimal coding for an information source with entropy $H \leq C$ such that n-length sequences produced by the source can be transmitted faithfully over the channel: the error rate tends to zero as $n \to \infty$. The probability of error tends to 1 if we attempt to transmit more than C bits through the channel.
- So we can improve the channel capacity by replacing the cable with a faster one, or we can improve the information processing (the data compression).

Channel Capacity

- For a noisy channel, if X represents the input to the channel and Y represents the output of the channel, H(X:Y) represents the average amount of information gained about the input X by ascertaining the value of the output Y.
- The capacity of a channel, C, is defined as the supremum of H(X:Y) over all input distributions.

Entangled States

- For any state $|\Psi\rangle$ of QE, there exist orthonormal bases $|i\rangle \in \mathcal{H}^Q$, $|j\rangle \in \mathcal{H}^E$ such that $|\Psi\rangle$ can be expressed in a biorthogonal correlated form as:

$$|\Psi\rangle = \sum_i \sqrt{p_i}|i\rangle|i\rangle$$

  where the coefficients $\sqrt{p_i}$ are real and non-negative, and $\sum p_i = 1$.
- This representation is referred to as the Schmidt decomposition. The Schmidt decomposition is unique if and only if the p_i are all distinct.

Entangled States

- Consider a quantum system Q which is part of a compound system QE. Pure states of QE are represented as rays or unit vectors in a tensor product Hilbert space $\mathcal{H}^{Q} \otimes \mathcal{H}^{E}$.
- A general pure state of QE is a state of the form:

$$|\Psi\rangle = \sum c_{ij} |q_i\rangle |e_j\rangle$$

where $|q_i\rangle \in \mathcal{H}^{Q}$ is a complete set of orthonormal states (a basis) in $\mathcal{H}^{Q}$ and $|e_j\rangle \in \mathcal{H}^{E}$ is a basis in $\mathcal{H}^{E}$.

- If the coefficients c_{ij} are such that $|\Psi\rangle$ cannot be expressed as a product state $|Q\rangle|E\rangle$, then $|\Psi\rangle$ is called an entangled state.

Channel Capacity

- Shannon's noisy channel coding theorem shows, perhaps surprisingly, that up to C bits of information can be sent through a noisy channel with arbitrary low error rate. That is, there exists an optimal coding for an information source with entropy $H \leq C$ such that n-length sequences produced by the source can be transmitted faithfully over the channel: the error rate tends to zero as $n \to \infty$. The probability of error tends to 1 if we attempt to transmit more than C bits through the channel.
- So we can improve the channel capacity by replacing the cable with a faster one, or we can improve the information processing (the data compression).

Entangled States

- Consider a quantum system Q which is part of a compound system QE. Pure states of QE are represented as rays or unit vectors in a tensor product Hilbert space $\mathcal{H}^{Q} \otimes \mathcal{H}^{E}$.
- A general pure state of QE is a state of the form:

$$|\Psi\rangle = \sum c_{ij} |q_i\rangle |e_j\rangle$$

where $|q_i\rangle \in \mathcal{H}^{Q}$ is a complete set of orthonormal states (a basis) in $\mathcal{H}^{Q}$ and $|e_j\rangle \in \mathcal{H}^{E}$ is a basis in $\mathcal{H}^{E}$.

- If the coefficients c_{ij} are such that $|\Psi\rangle$ cannot be expressed as a product state $|Q\rangle|E\rangle$, then $|\Psi\rangle$ is called an entangled state.

Entangled States

- For any state $|\Psi\rangle$ of QE, there exist orthonormal bases $|i\rangle \in \mathcal{H}^Q$, $|j\rangle \in \mathcal{H}^E$ such that $|\Psi\rangle$ can be expressed in a biorthogonal correlated form as:

$$|\Psi\rangle = \sum_i \sqrt{p_i}|i\rangle|i\rangle$$

  where the coefficients $\sqrt{p_i}$ are real and non-negative, and $\sum p_i = 1$.
- This representation is referred to as the Schmidt decomposition. The Schmidt decomposition is unique if and only if the p_i are all distinct.

Entangled States

- An example is the biorthogonal EPR state:

$$|\Psi\rangle = (|0\rangle|1\rangle - |1\rangle|0\rangle)/\sqrt{2}$$

  say, the singlet state of two spin-1/2 particles (the Schmidt form with positive coefficients is obtained by asborbing the relative phases in the definition of the basis vectors).
- In the singlet state, $|0\rangle$ and $|1\rangle$ can be taken as representing the two eigenstates of spin in the z-direction, but since the state is symmetric, $|\Psi\rangle$ retains the same form for spin in any direction.

Entangled States

Notice that the four states:

$$
\begin{aligned}
|1\rangle &= \frac{1}{\sqrt{2}}(|0\rangle|1\rangle - |1\rangle|0\rangle) \\
|2\rangle &= \frac{1}{\sqrt{2}}(|0\rangle|1\rangle + |1\rangle|0\rangle) \\
|3\rangle &= \frac{1}{\sqrt{2}}(|0\rangle|0\rangle - |1\rangle|1\rangle) \\
|4\rangle &= \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle)
\end{aligned}
$$

form an orthonormal basis, called the Bell basis, in the 2 x 2-dimensional Hilbert space.

Entangled States

- Any Bell state can be transformed into any other Bell state by a local unitary transformation, X, Y, or Z, where X, Y, Z are the Pauli spin matrices:

$$\mathrm{X} = \sigma_{\mathrm{x}} = |0\rangle\langle 1| + |1\rangle\langle 0| = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\mathrm{Y} = \sigma_{\mathrm{y}} = -\mathrm{i}|0\rangle\langle 1| + \mathrm{i}|1\rangle\langle 0| = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}$$

$$\mathrm{Z} = \sigma_{\mathrm{z}} = |0\rangle\langle 0| - |1\rangle\langle 1| = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

- For example:

$$\mathrm{X}\otimes\mathrm{I}\cdot|4\rangle = \mathrm{X}\otimes\mathrm{I}\cdot\frac{1}{\sqrt{2}}(|0\rangle\langle 1|-|1\rangle|0\rangle = -\frac{1}{\sqrt{2}}(|0\rangle\langle 0|-|1\rangle|1\rangle = -|3\rangle$$

Entangled States

If QE is a closed system in an entangled pure state represented by

$$|\Psi\rangle = \sum_i \sqrt{p_i}|i\rangle|i\rangle$$

in the Schmidt decomposition, the expected value of any Q-observable A on $\mathcal{H}^Q$ can be computed as:

$$\begin{aligned}
\langle A\rangle &= \mathrm{Tr}(|\Psi\rangle\langle\Psi|A\otimes I)\\
&= \mathrm{Tr}_Q(\mathrm{Tr}_E(|\Psi\rangle\langle\Psi|A))\\
&= \mathrm{Tr}_Q(\sum_i p_i|i\rangle\langle i|A)\\
&= \mathrm{Tr}_Q(\rho A)
\end{aligned}$$

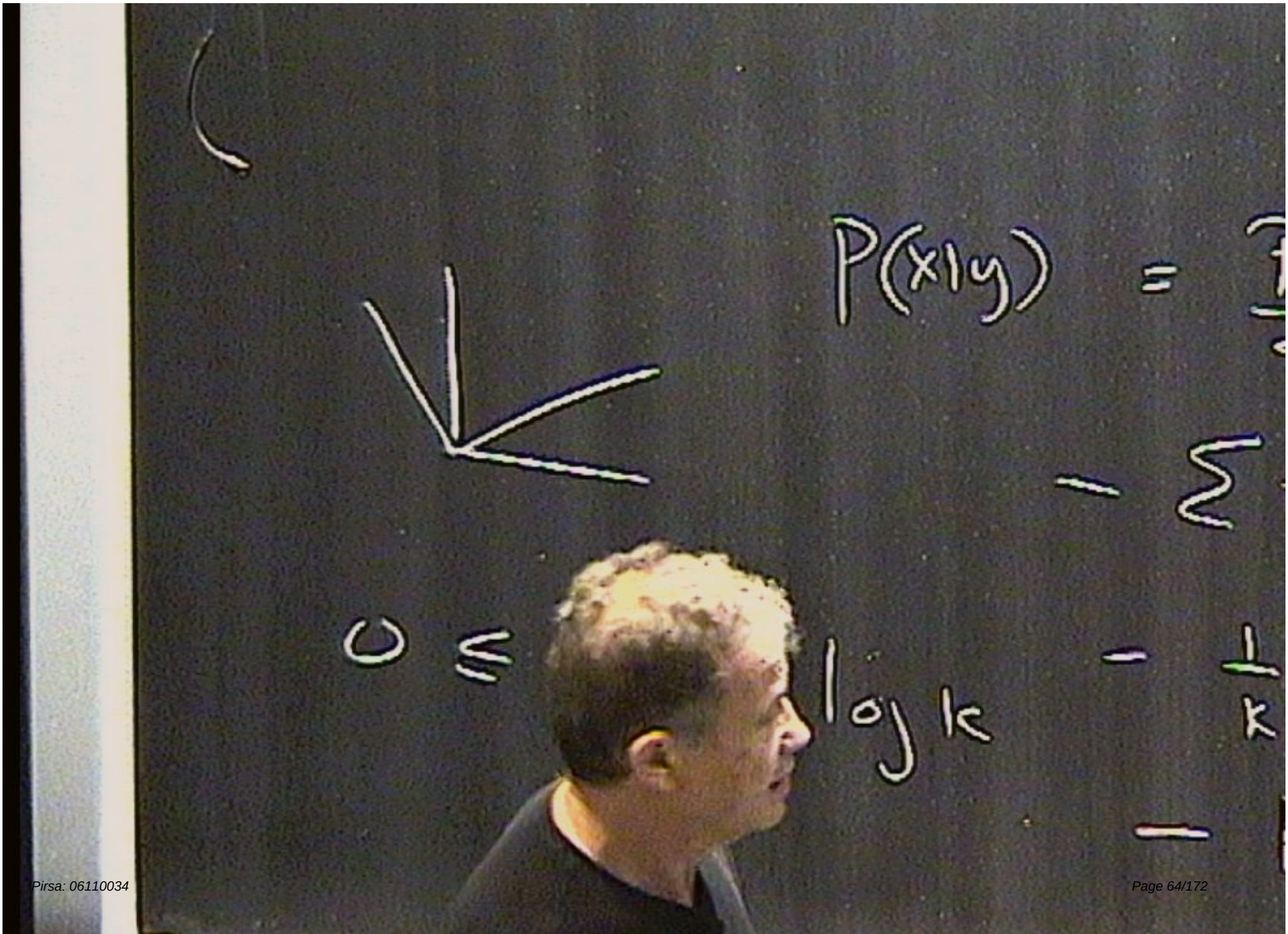
where $\mathrm{Tr}_Q() = \sum_q \langle q_i|\cdot|q_i\rangle$, for any orthonormal basis in $\mathcal{H}^Q$, is the partial trace over $\mathcal{H}^Q$, and $\mathrm{Tr}_E()$ is the partial trace over $\mathcal{H}^E$.

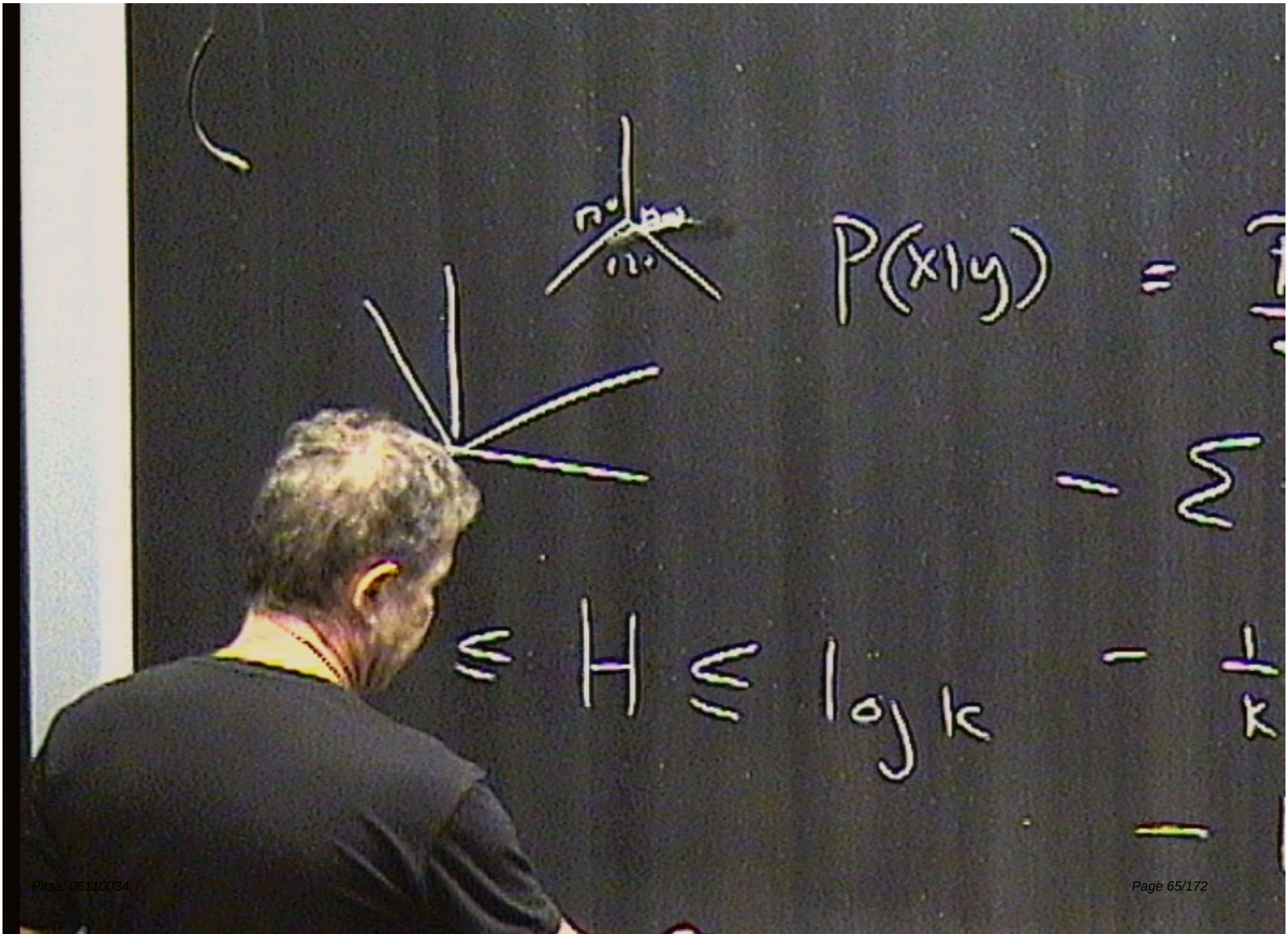
Entangled States

- $\rho = \sum_i p_i |i\rangle\langle i| \in \mathcal{H}^Q$ is the reduced density operator of the open system Q, a positive operator with unit trace.
- Since the density operator ρ yields the statistics of all Q-observables via $\langle A \rangle = \mathrm{Tr}_Q(\rho A)$, ρ is taken as representing the quantum state of the system Q.
- If QE is an entangled pure state, then the open system Q is in a mixed state ρ, i.e., $\rho \neq \rho^2$; for pure states, ρ is a projection operator onto a ray and $\rho = \rho^2$.

Entangled States

- A mixed state represented by a density operator $\rho = \sum \rho_i |i\rangle\langle i|$ can be regarded as a mixture of pure states $|i\rangle$ prepared with prior probabilities p_i, but this representation is not unique—not even if the states combined in the mixture are orthogonal.
- For example, the equal-weight mixture of orthonormal states $|0\rangle, |1\rangle$ in a 2-dimensional Hilbert space $\mathcal{H}_2$ has precisely the same statistical properties, and hence the same density operator $\rho = I/2$, as the equal weight mixture of any pair of orthonormal states, e.g., the states $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$, or the equal-weight mixture of nonorthogonal states $|0\rangle, \frac{1}{2}|0\rangle + \frac{\sqrt{3}}{2}|1\rangle, \frac{1}{2}|0\rangle - \frac{\sqrt{3}}{2}|1\rangle$ 120° degrees apart, or the uniform continuous distribution over all possible states in $\mathcal{H}_2$.


P(x|y)
0 ≤
log jk


P(x|y)
≤ H ≤ log k

$P(x|y) =$

$- \Sigma$

$0 \leq H \leq \log_j k$

$- \frac{1}{k}$

Entangled States

- More generally, for any basis of orthonormal states $|e_i\rangle \in \mathcal{H}^E$, the entangled state $|\Psi\rangle$ can be expressed as:

$$|\Psi\rangle = \sum_{ij} c_{ij}|q_i\rangle|e_j\rangle = \sum_j \sqrt{w_j}|r_j\rangle|e_j\rangle$$

where the normalized states $|r_j\rangle = \sum_i \frac{c_{ij}}{\sqrt{w_j}}|q_i\rangle$ are relative states to the $|e_j\rangle$ ($\sqrt{w_j} = \sum_j |c_{ij}|^2$).

- Note that the states $|r_j\rangle$ are not in general orthogonal. Since the $|e_j\rangle$ are orthogonal, we can express the density operator representing the state of Q as:

$$\rho = \sum_i w_i|r_i\rangle\langle r_i|$$

Entangled States

- More generally, for any basis of orthonormal states $|e_i\rangle \in \mathcal{H}^E$, the entangled state $|\Psi\rangle$ can be expressed as:

$$|\Psi\rangle = \sum_{ij} c_{ij}|q_i\rangle|e_j\rangle = \sum_j \sqrt{w_j}|r_j\rangle|e_j\rangle$$

  where the normalized states $|r_j\rangle = \sum_i \frac{c_{ij}}{\sqrt{w_j}}|q_i\rangle$ are relative states to the $|e_j\rangle$ ($\sqrt{w_j} = \sum_j |c_{ij}|^2$).
- Note that the states $|r_j\rangle$ are not in general orthogonal. Since the $|e_j\rangle$ are orthogonal, we can express the density operator representing the state of Q as:

$$\rho = \sum_i w_i|r_i\rangle\langle r_i|$$

Entangled States

In effect, a measurement of an E-observable with eigenstates $|e_i\rangle$ will leave the composite system QE in one of the states $|r_i\rangle|e_i\rangle$ with probability w_i, and a measurement of an E-observable with eigenstates $|i\rangle$ (the orthogonal states of the Schmidt decomposition) will leave the system QE in one of the states $|i\rangle|i\rangle$ with probability p_i.

Entangled States

Since Q and E could be widely separated from each other in space, no measurement at E could affect the statistics of any Q-observable; or else measurements at E would allow superluminal signaling between Q and E. It follows that the mixed state ρ can be realized as a mixture of orthogonal states $|i\rangle$ (the eigenstates of ρ) with weights p_i, or as a mixture of non-orthogonal relative states $|r_j\rangle$ with weights w_j in infinitely many ways, depending on the choice of basis in $\mathcal{H}^E$:

$$\rho = \sum_i p_i |i\rangle\langle i| = \sum_j w_j |r_j\rangle\langle r_j|$$

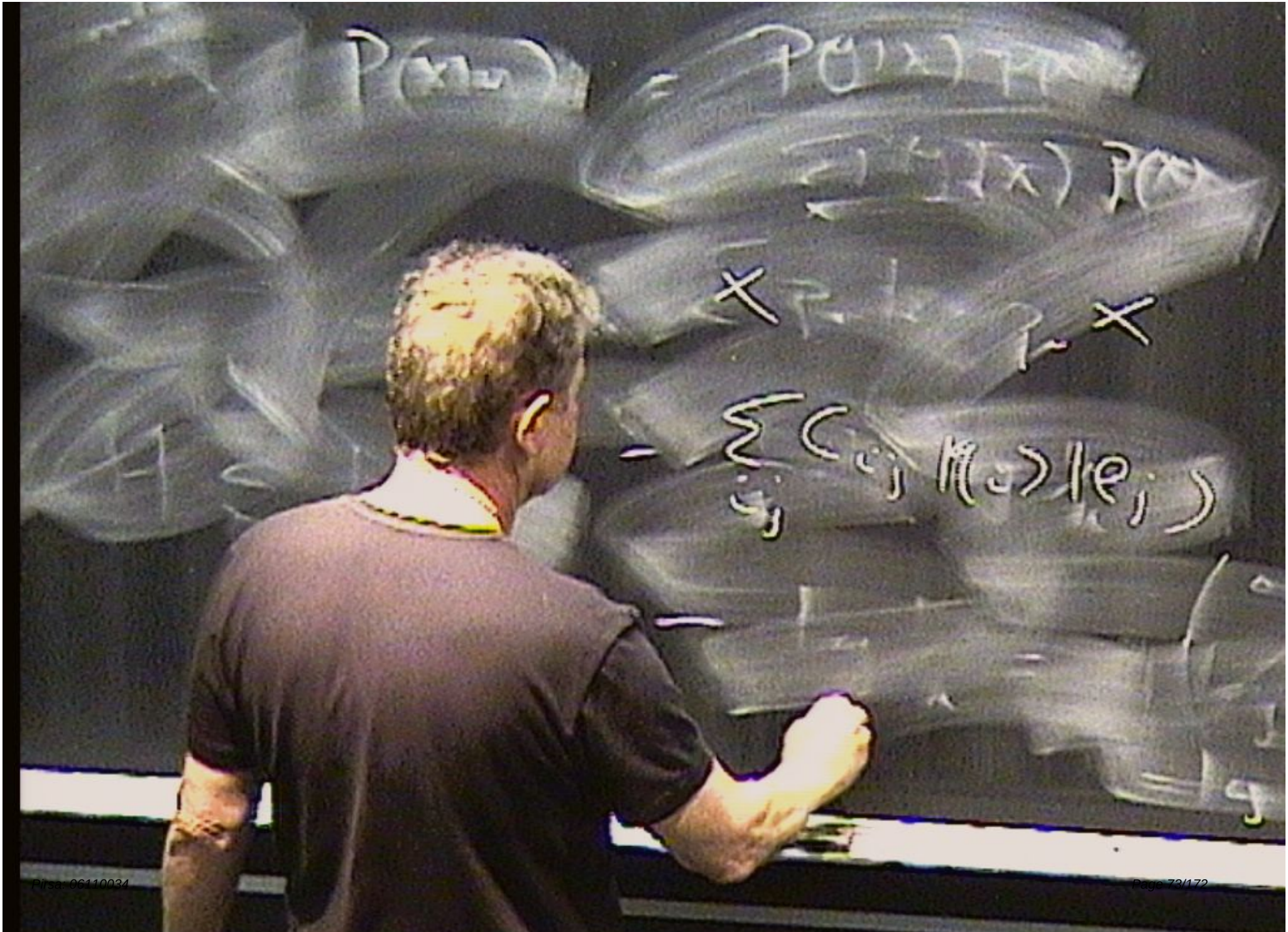
and all these different mixtures with the same density operator ρ must be physically indistinguishable.

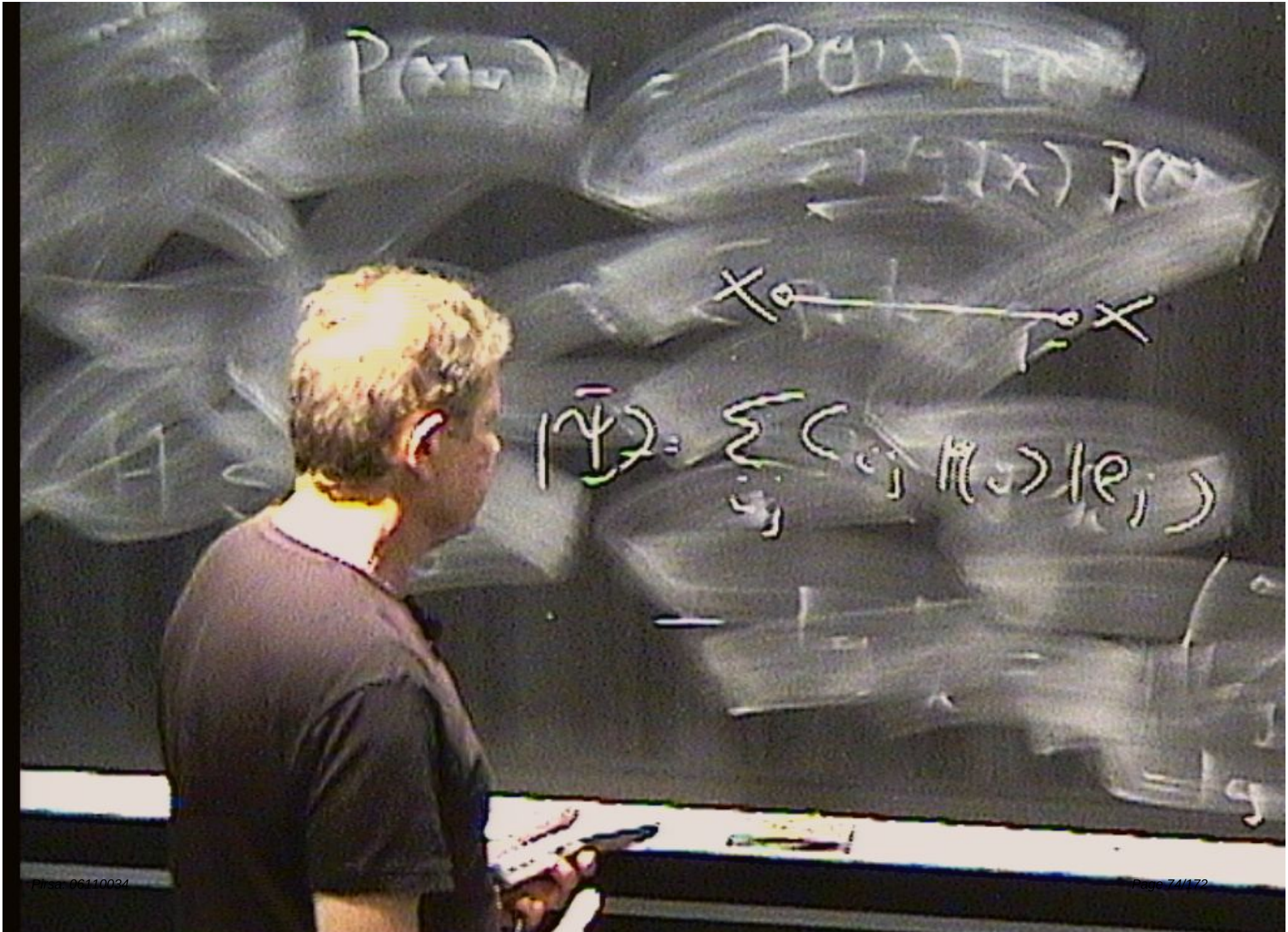
Entangled States

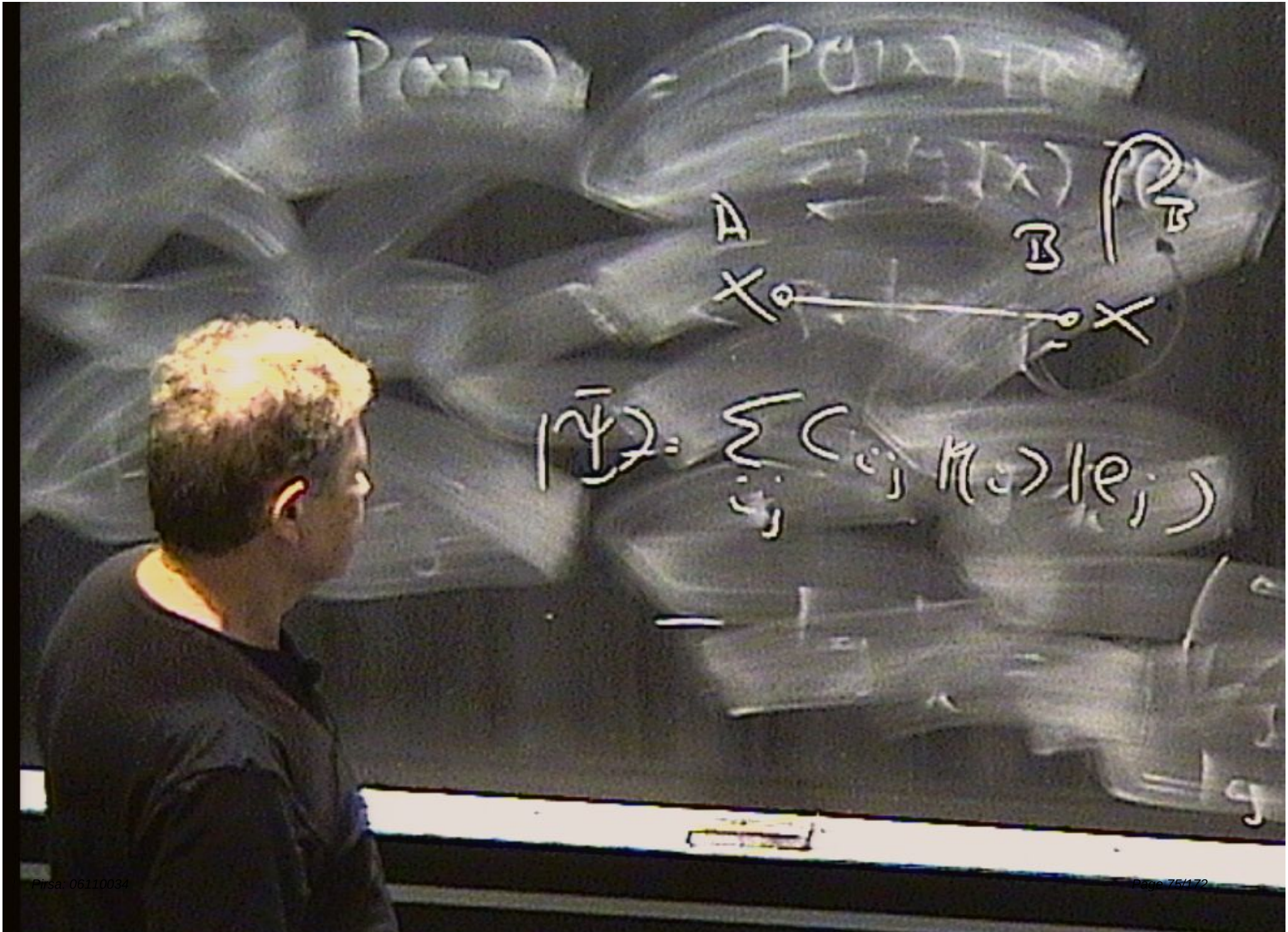
- Note that any mixed state density operator $\rho \in \mathcal{H}^{\mathrm{Q}}$ can be 'purified' by adding a suitable ancilla system E, in the sense that ρ is the partial trace of a pure state $|\Psi\rangle \in \mathcal{H}^{\mathrm{Q}} \otimes \mathcal{H}^{\mathrm{E}}$ over $\mathcal{H}^{\mathrm{E}}$.
- A purification of a mixed state is, clearly, not unique, but depends on the choice of $|\Psi\rangle$ in $\mathcal{H}^{\mathrm{E}}$.

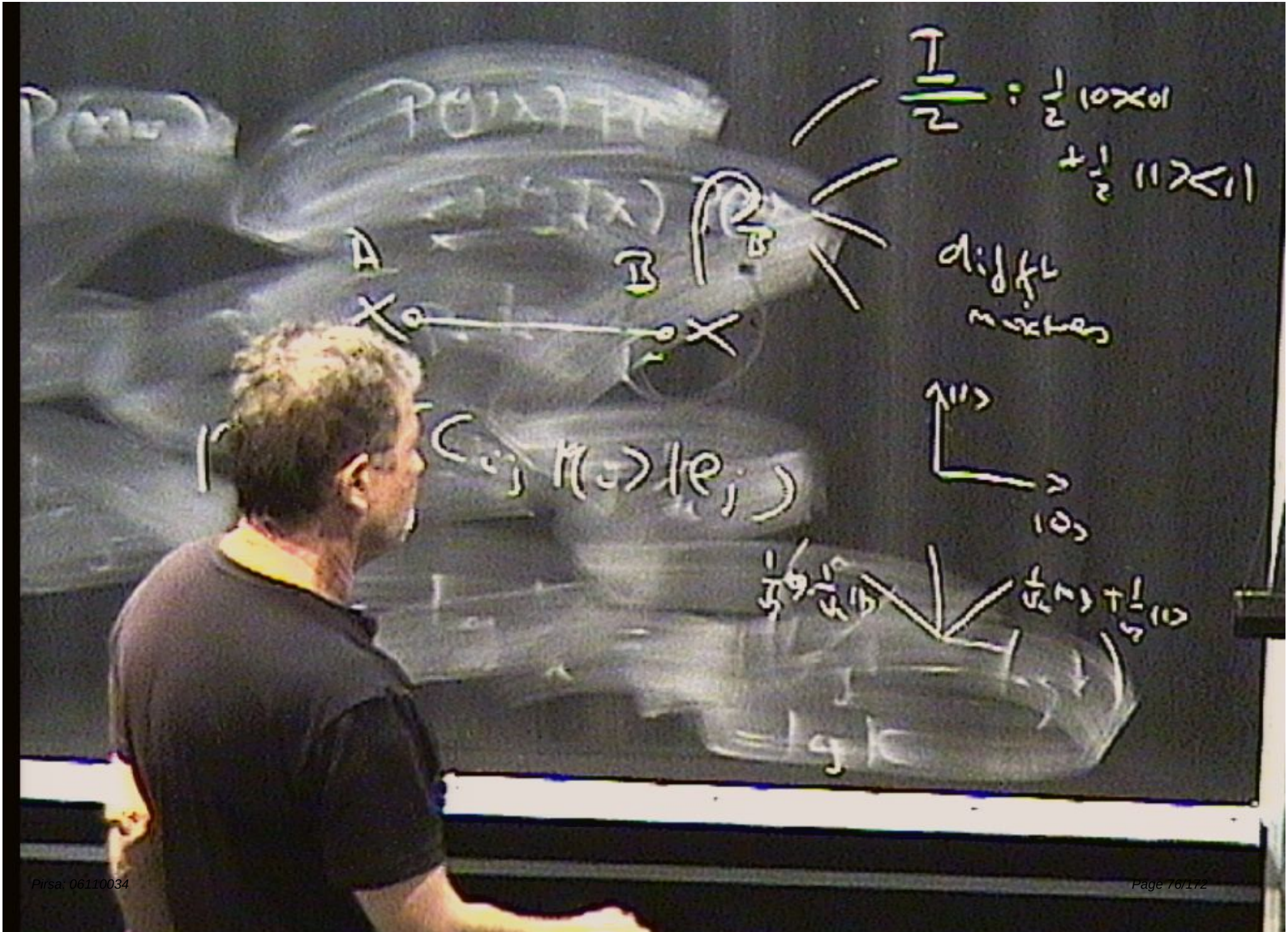
Entangled States

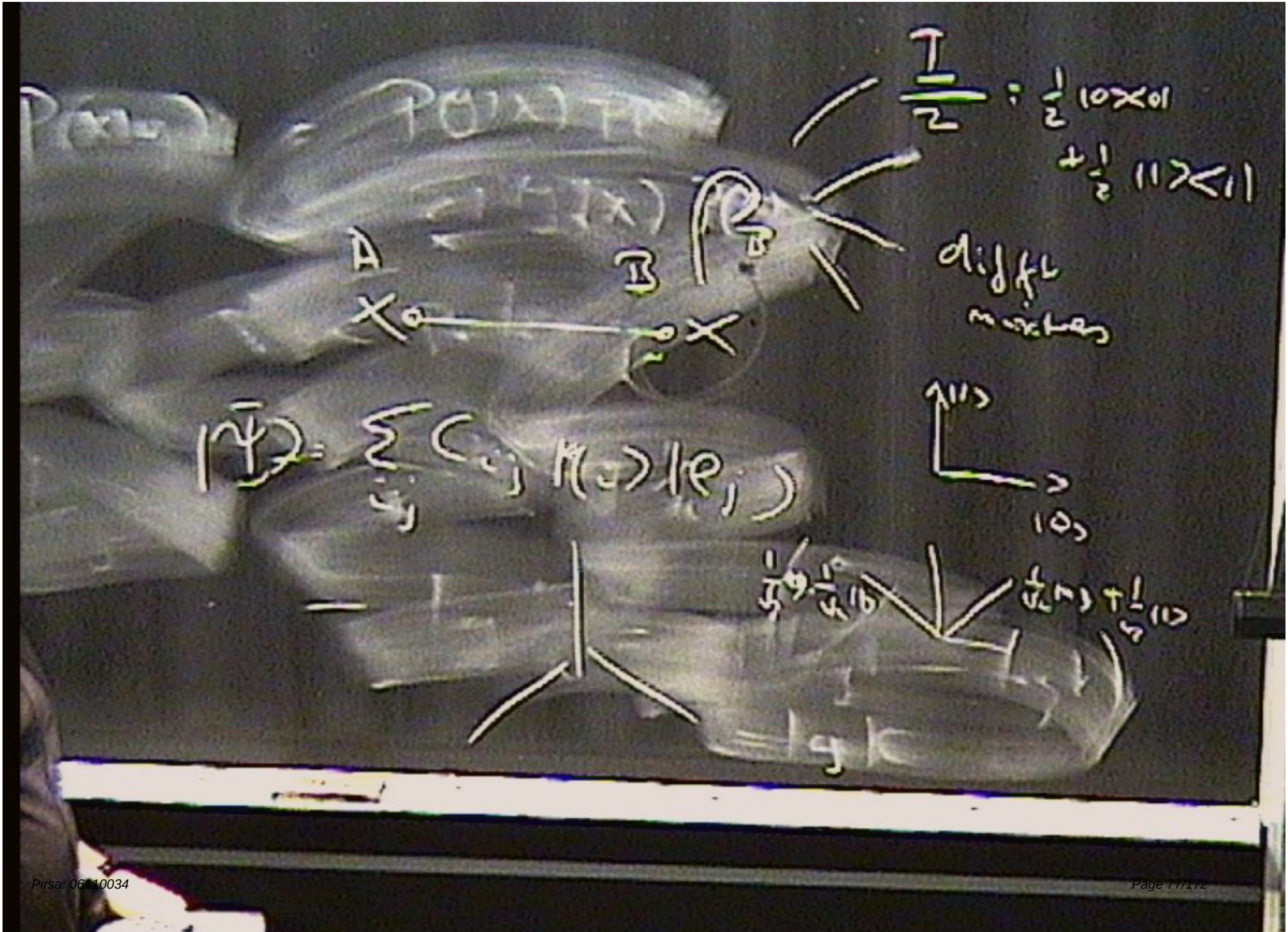
- The Hughston-Jozsa-Wootters theorem shows that for any mixture of pure states $|r_i\rangle$ with weights w_i, where $\rho = \sum_j w_j |r_j\rangle\langle r_j|$, there is a purification of ρ and a suitable measurement on the system E that will leave Q in the mixture ρ.
- So an observer at E can remotely prepare Q in any mixture that corresponds to the density operator ρ (and of course all these different mixtures are physically indistinguishable).

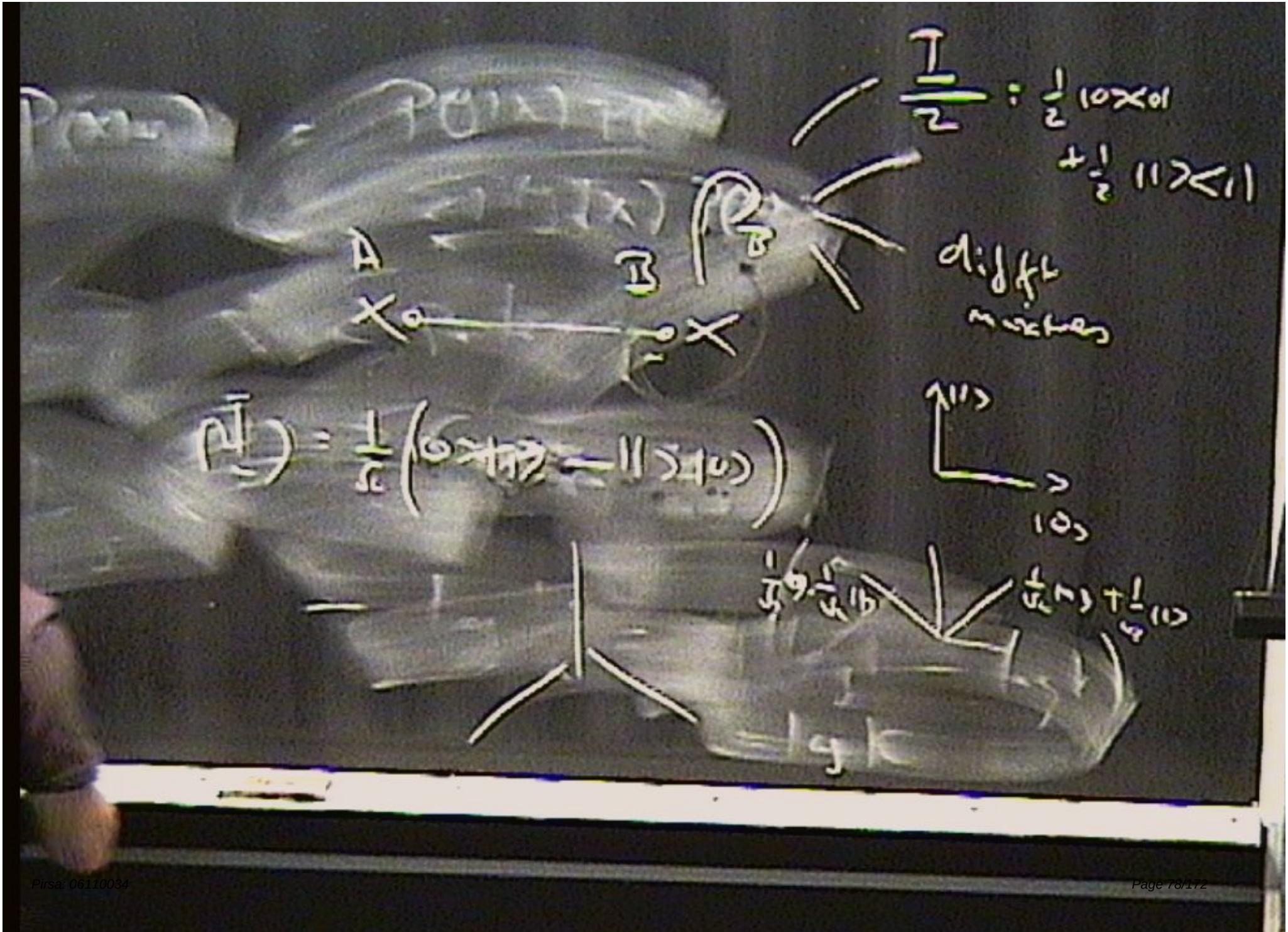


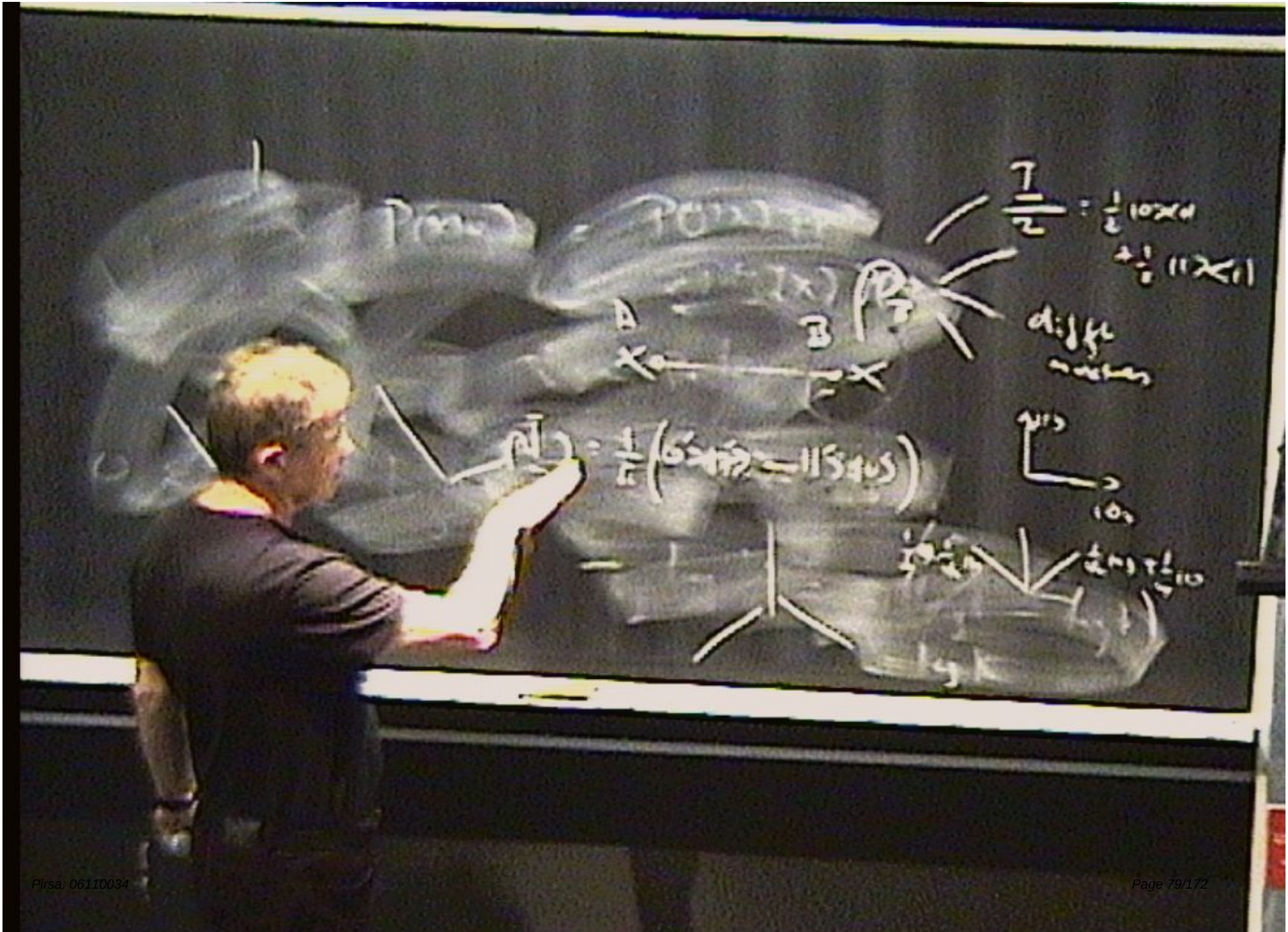


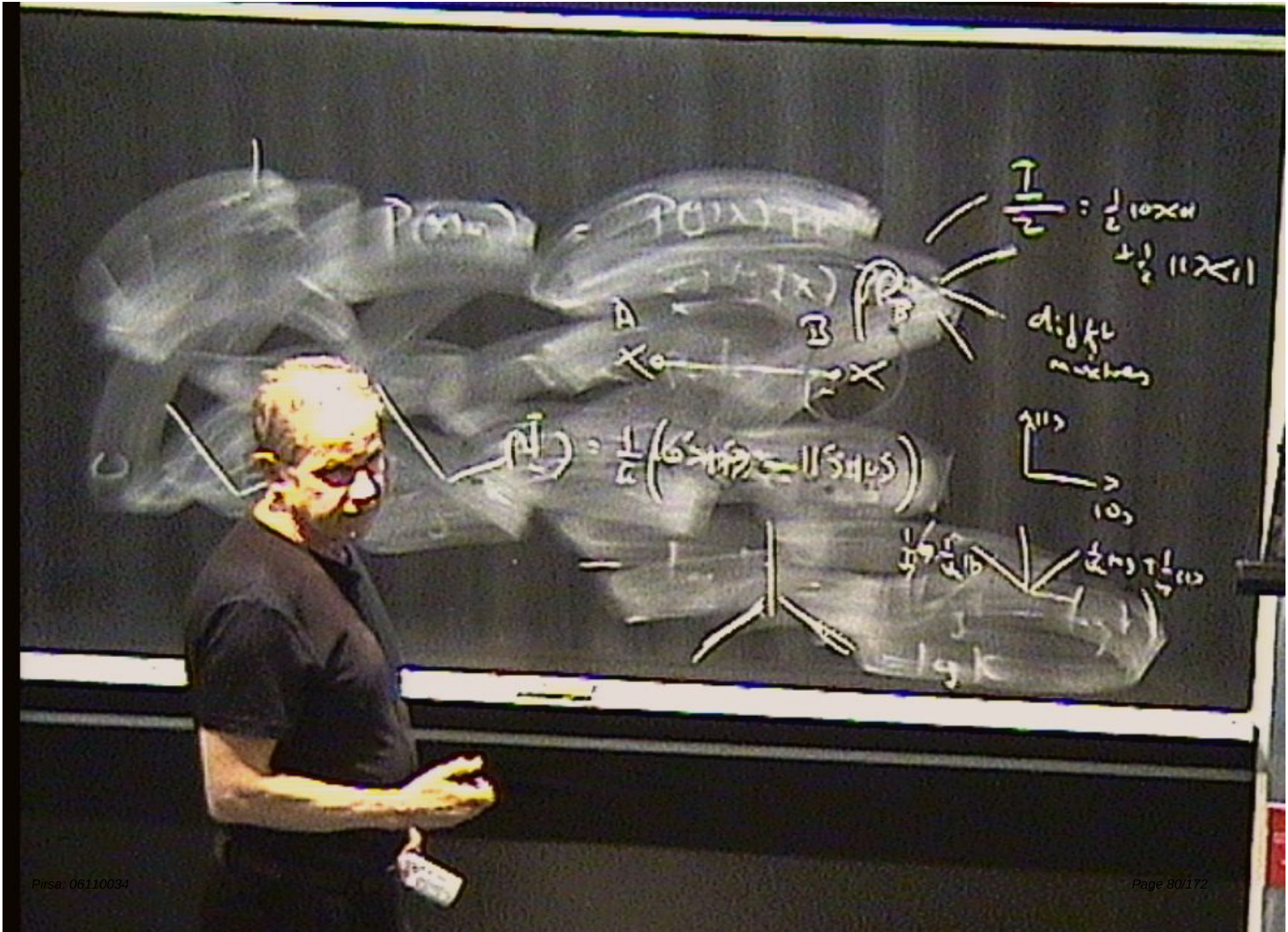

A
B
$|\Psi\rangle = \sum_j c_j |\psi(x_j)\rangle |e_j\rangle$


A
B


A
B
|0>
|1>

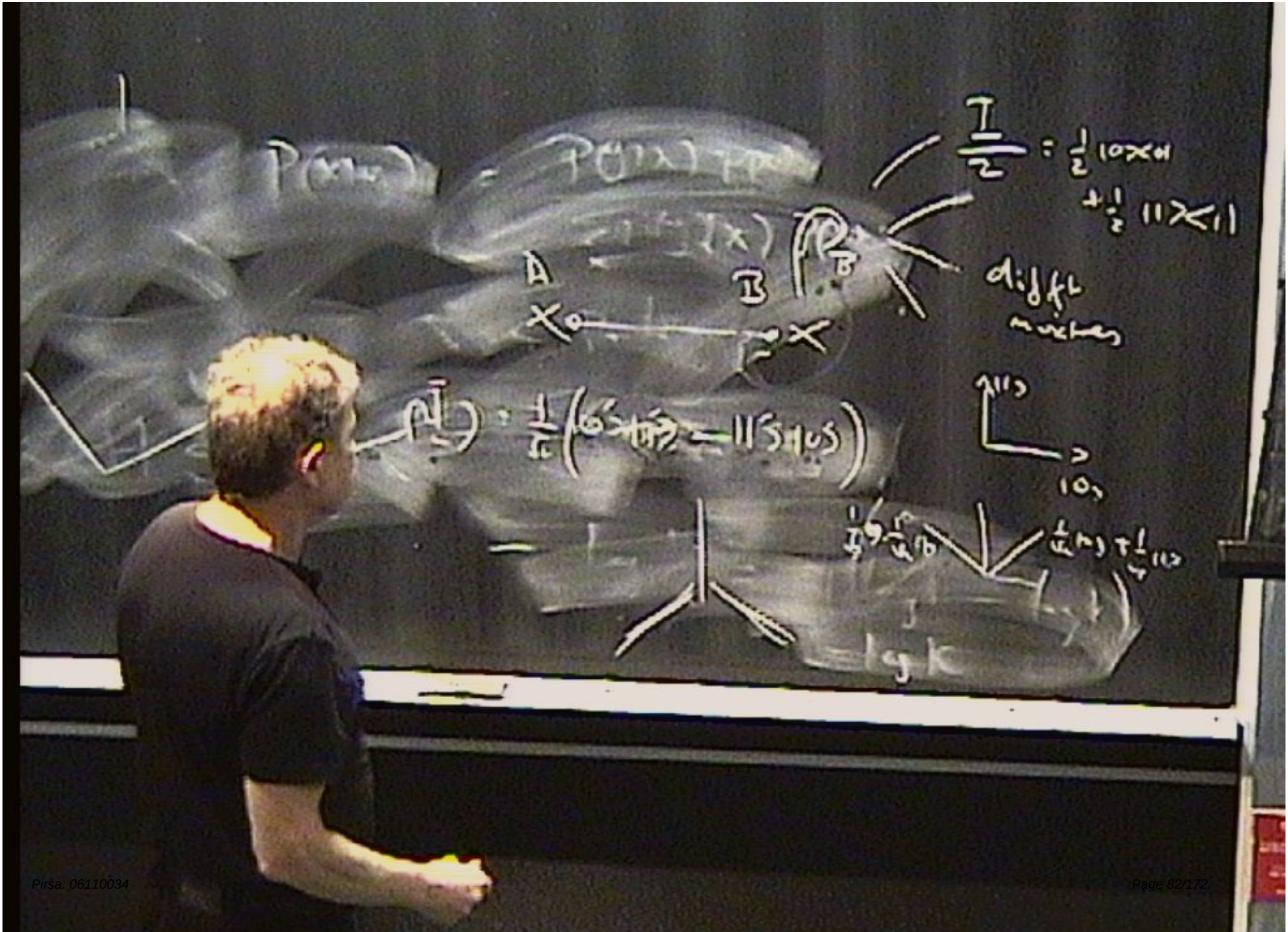

A
B
different
masses

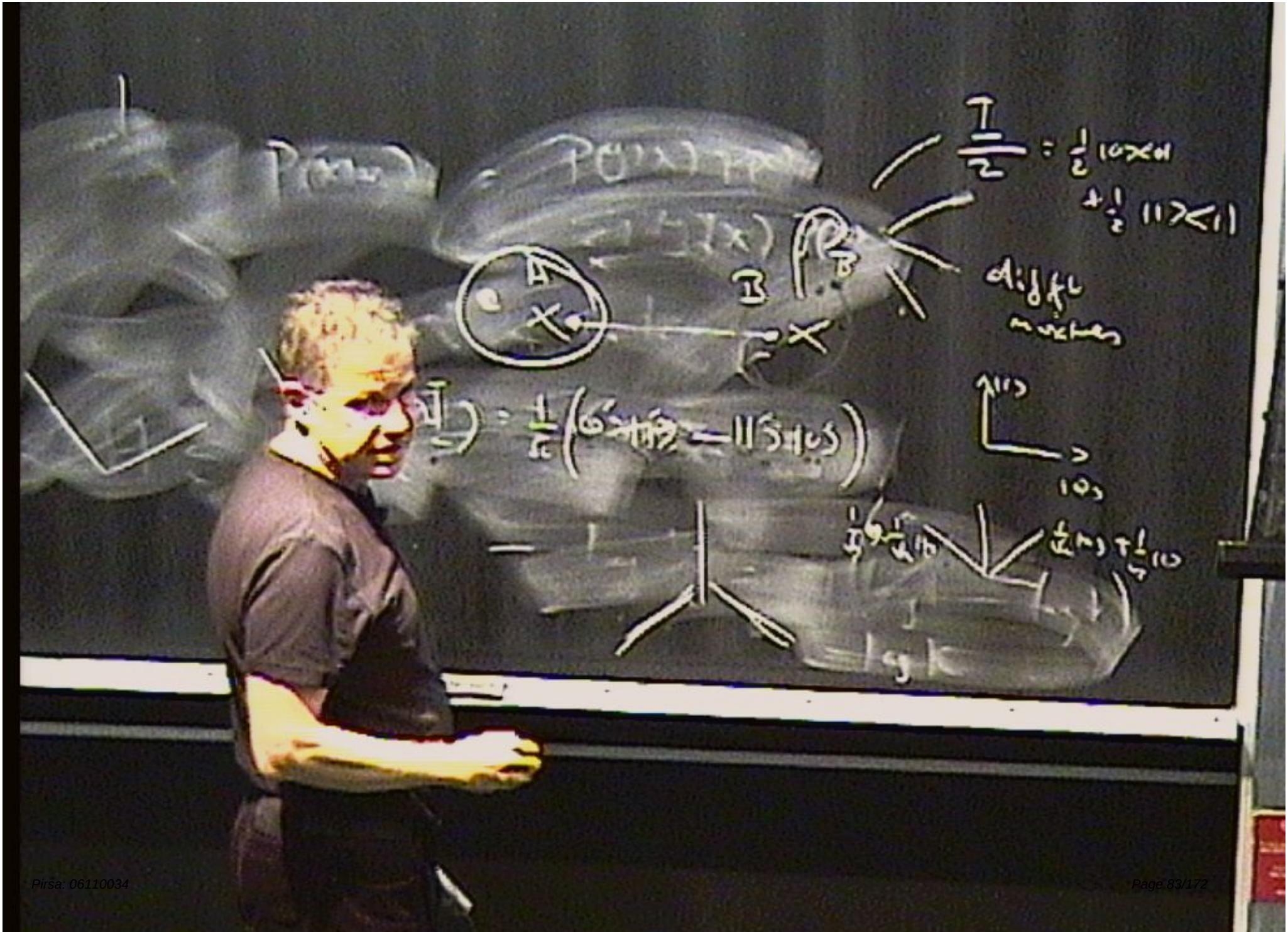


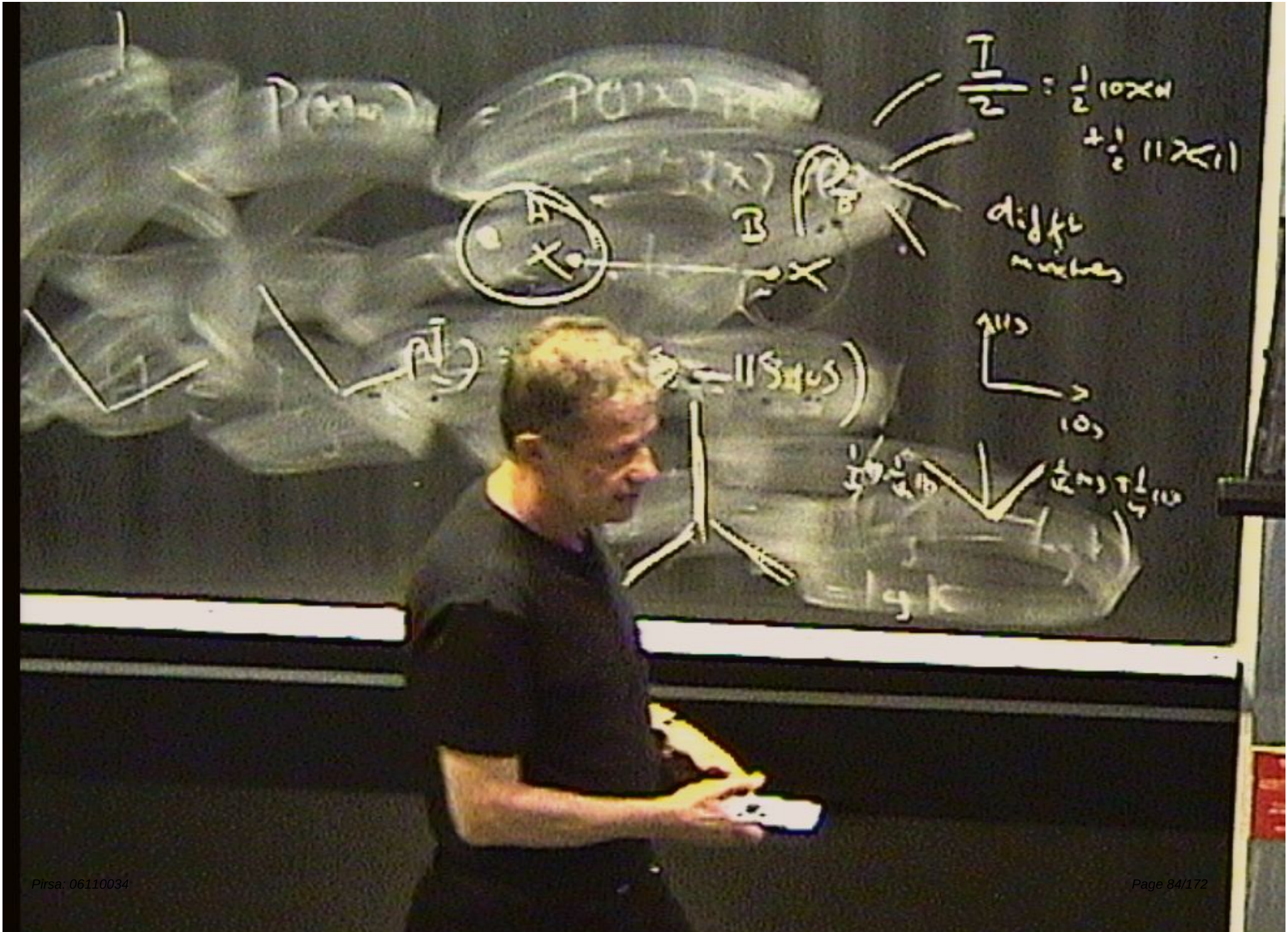

A
B

Entangled States

- The Hughston-Jozsa-Wootters theorem shows that for any mixture of pure states $|r_i\rangle$ with weights w_i, where $\rho = \sum_j w_j |r_j\rangle\langle r_j|$, there is a purification of ρ and a suitable measurement on the system E that will leave Q in the mixture ρ.
- So an observer at E can remotely prepare Q in any mixture that corresponds to the density operator ρ (and of course all these different mixtures are physically indistinguishable).


A
B
B
J
J
g k


A
B
B
X
X

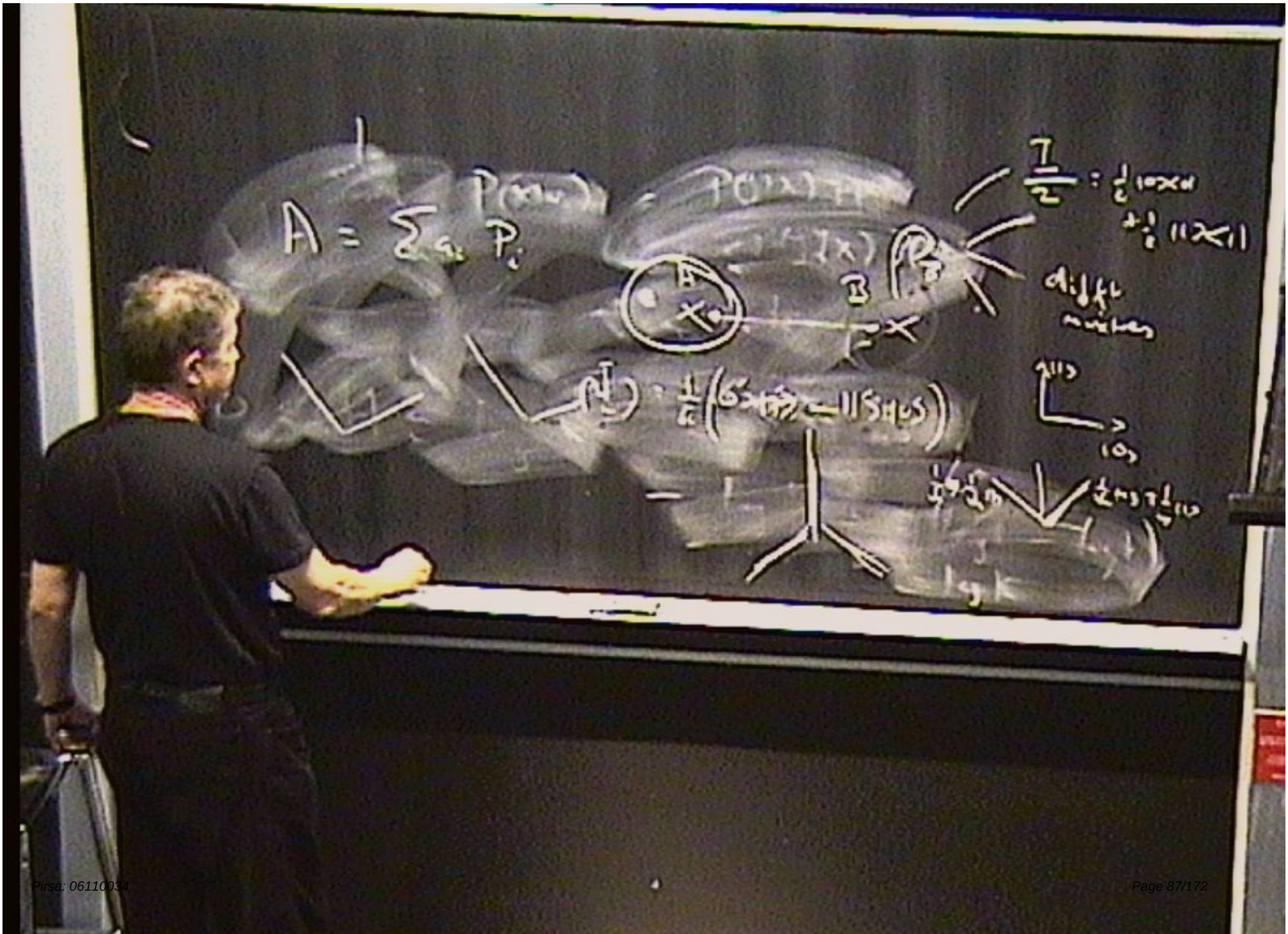

A
B
B
|0>
|1>

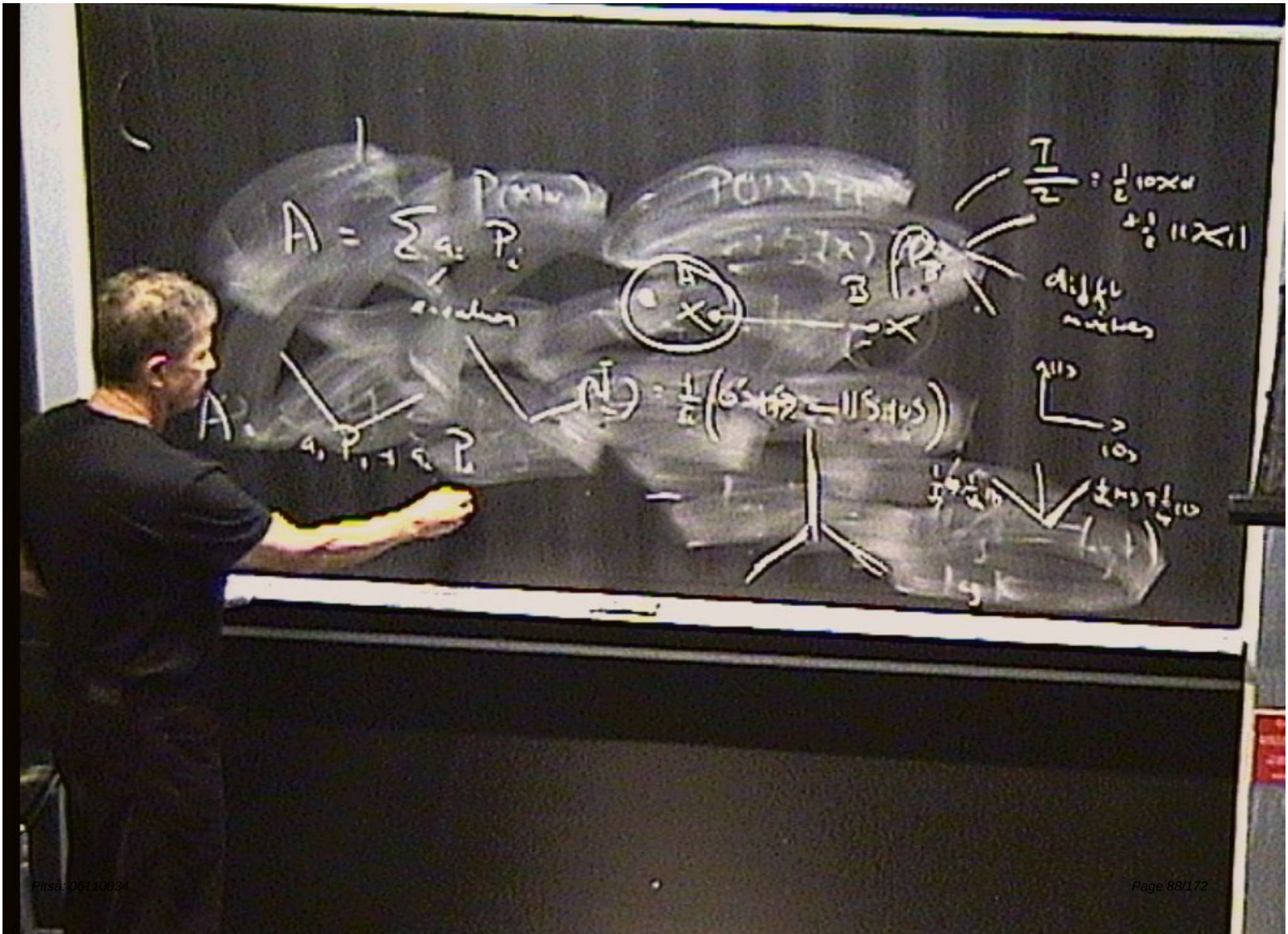
Entangled States

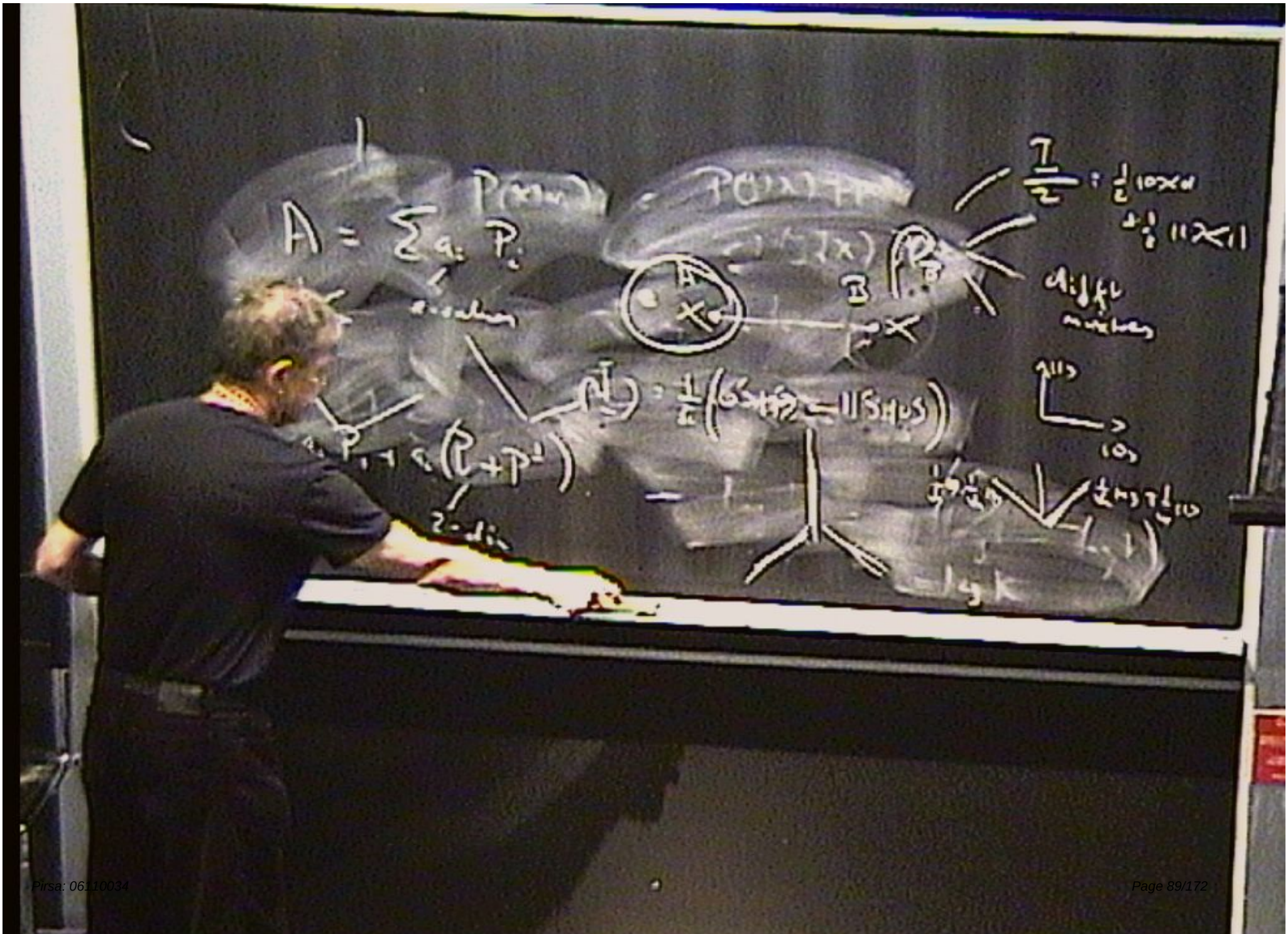
- The Hughston-Jozsa-Wootters theorem shows that for any mixture of pure states $|r_i\rangle$ with weights w_i, where $\rho = \sum_j w_j |r_j\rangle\langle r_j|$, there is a purification of ρ and a suitable measurement on the system E that will leave Q in the mixture ρ.
- So an observer at E can remotely prepare Q in any mixture that corresponds to the density operator ρ (and of course all these different mixtures are physically indistinguishable).

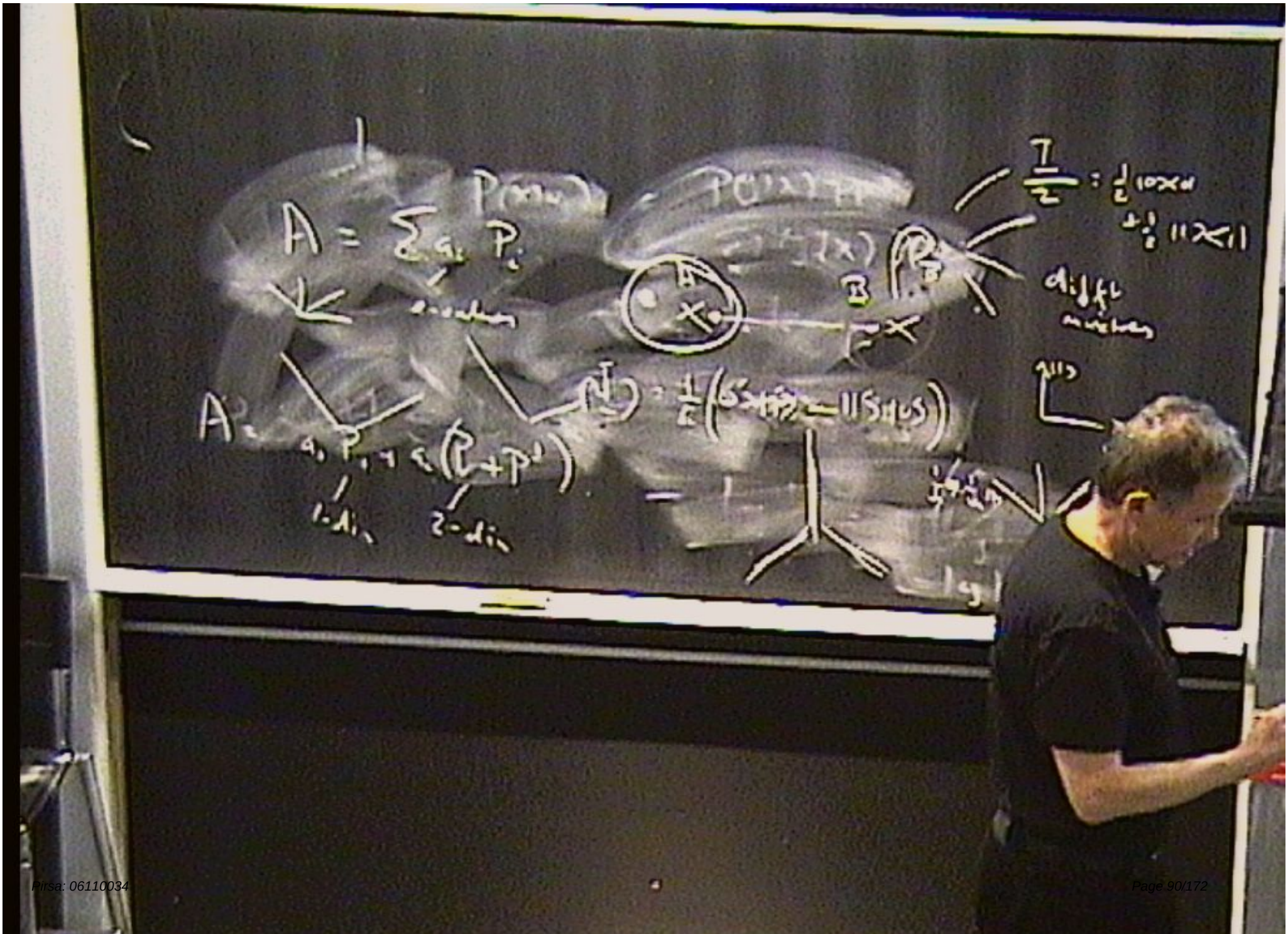
Measurement

- A standard von Neumann 'yes-no' measurement is associated with a projection operator; so a standard observable is represented in the spectral representation as a sum of projection operators, with coefficients representing the eigenvalues of the observable.
- Such a measurement is the quantum analogue of the measurement of a property of a system in classical physics.


A
B




A
B
X
2-dim

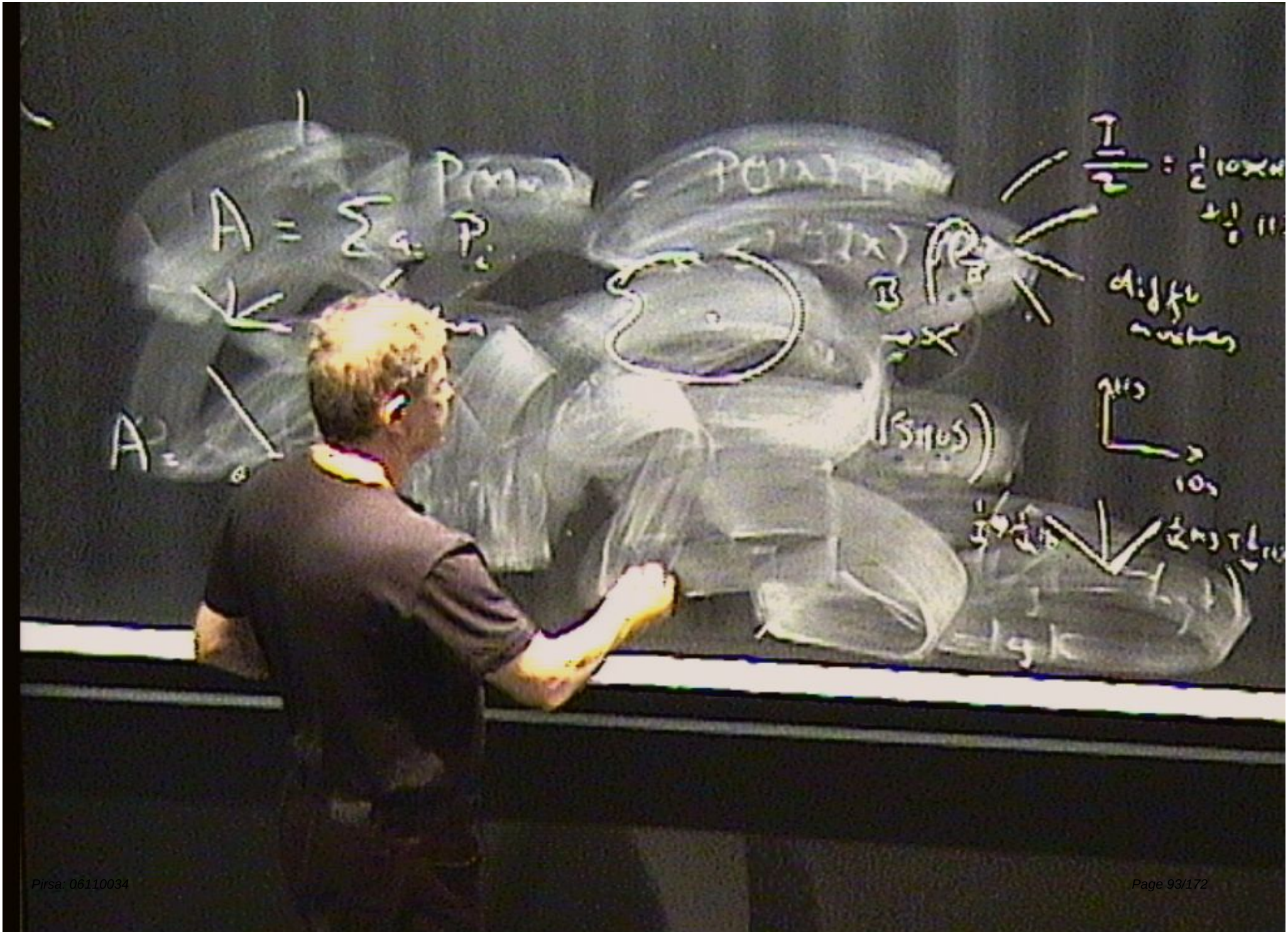


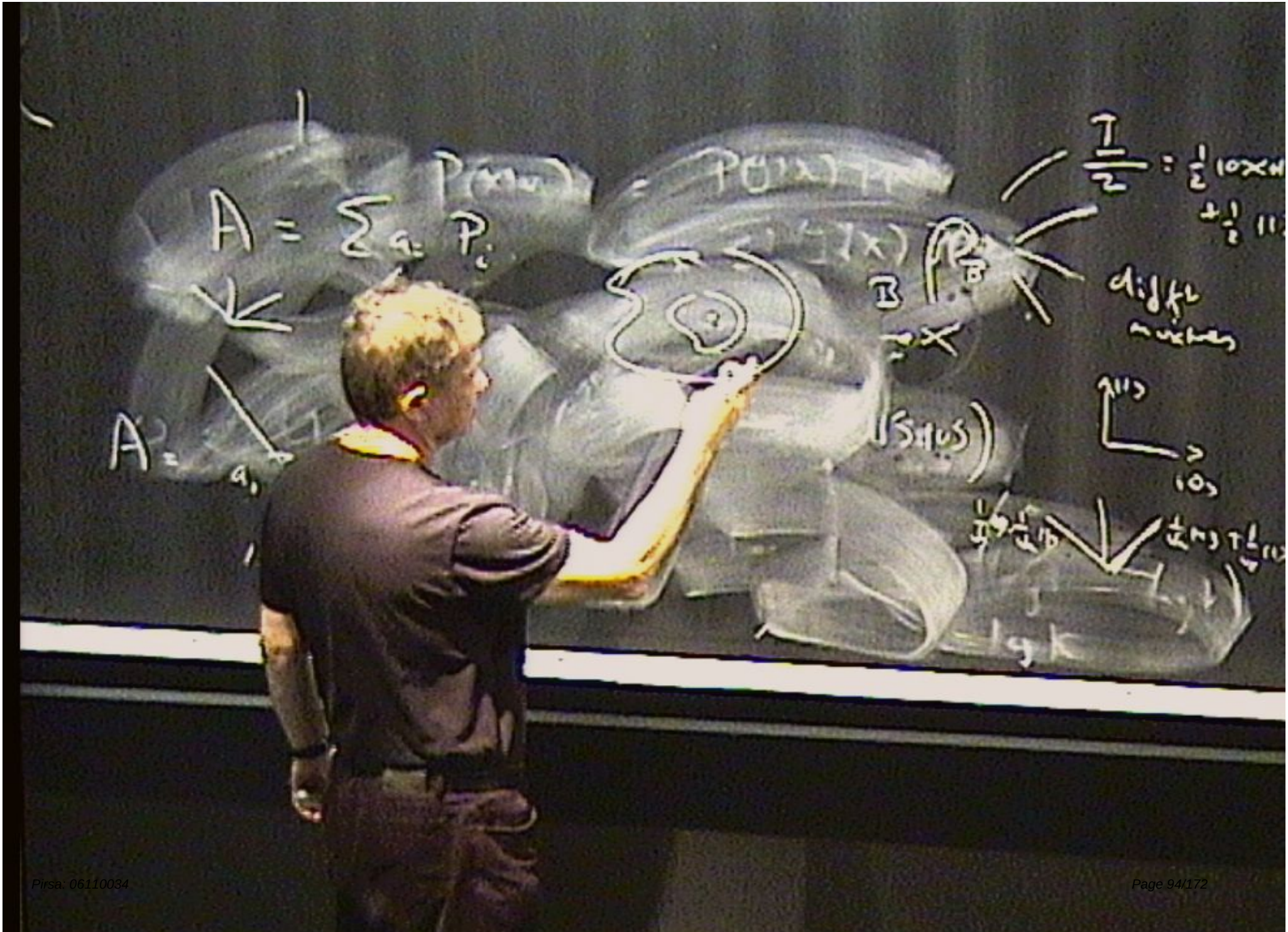
Measurement

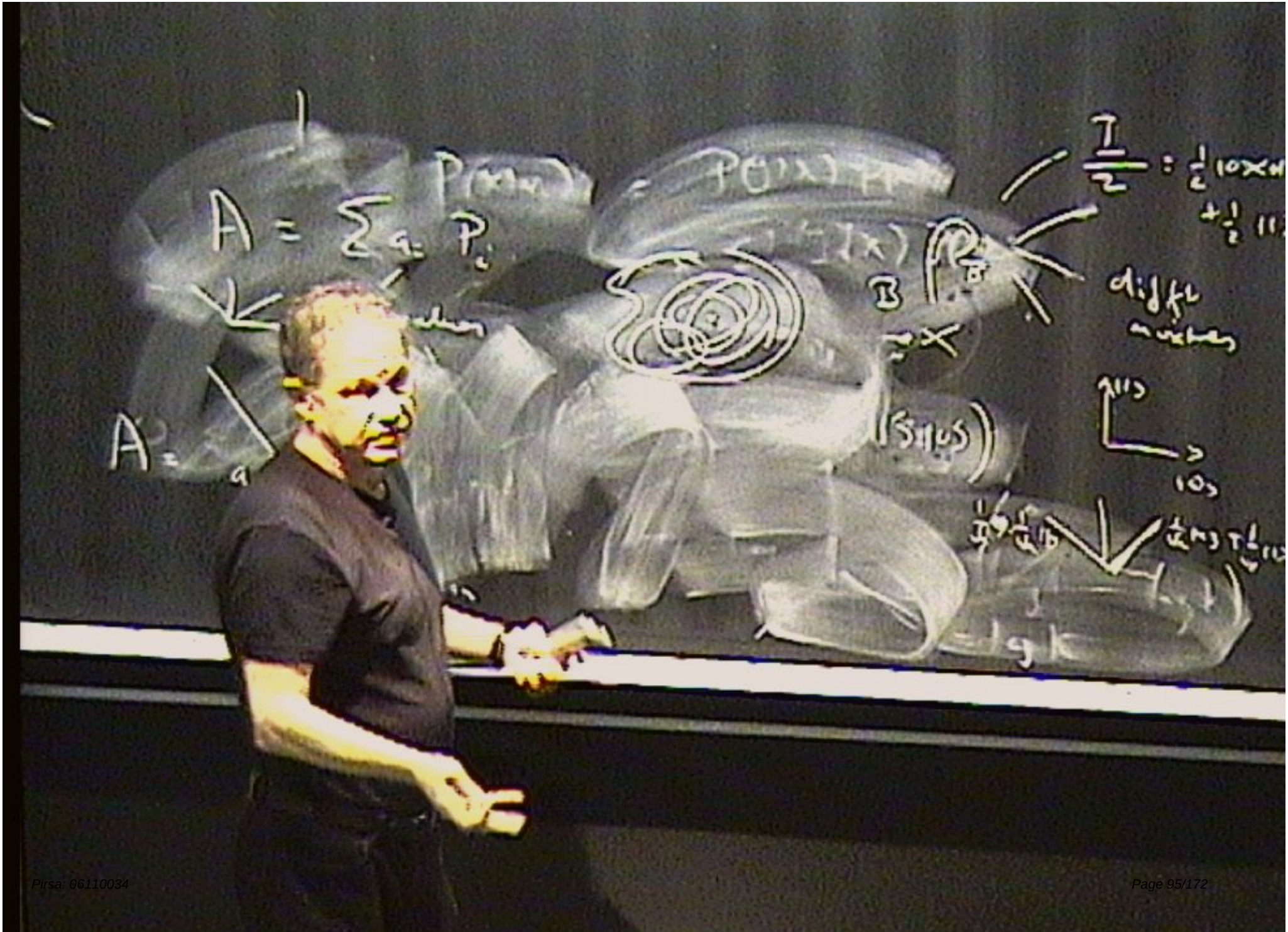
- Classically, we think of a property of a system as being associated with a subset in the state space (phase space) of the system, and determining whether the system has the property amounts to determining whether the state of the system lies in the corresponding subset.
- In quantum mechanics, the counterpart of a subset in phase space is a closed linear subspace in Hilbert space. Just as the different possible values of an observable (dynamical quantity) of a classical system correspond to the subsets in a mutually exclusive and collectively exhaustive set of subsets covering the classical state space, so the different values of a quantum observable correspond to the subspaces in a mutually exclusive (i.e., orthogonal) and collectively exhaustive set of subspaces spanning the quantum state space.

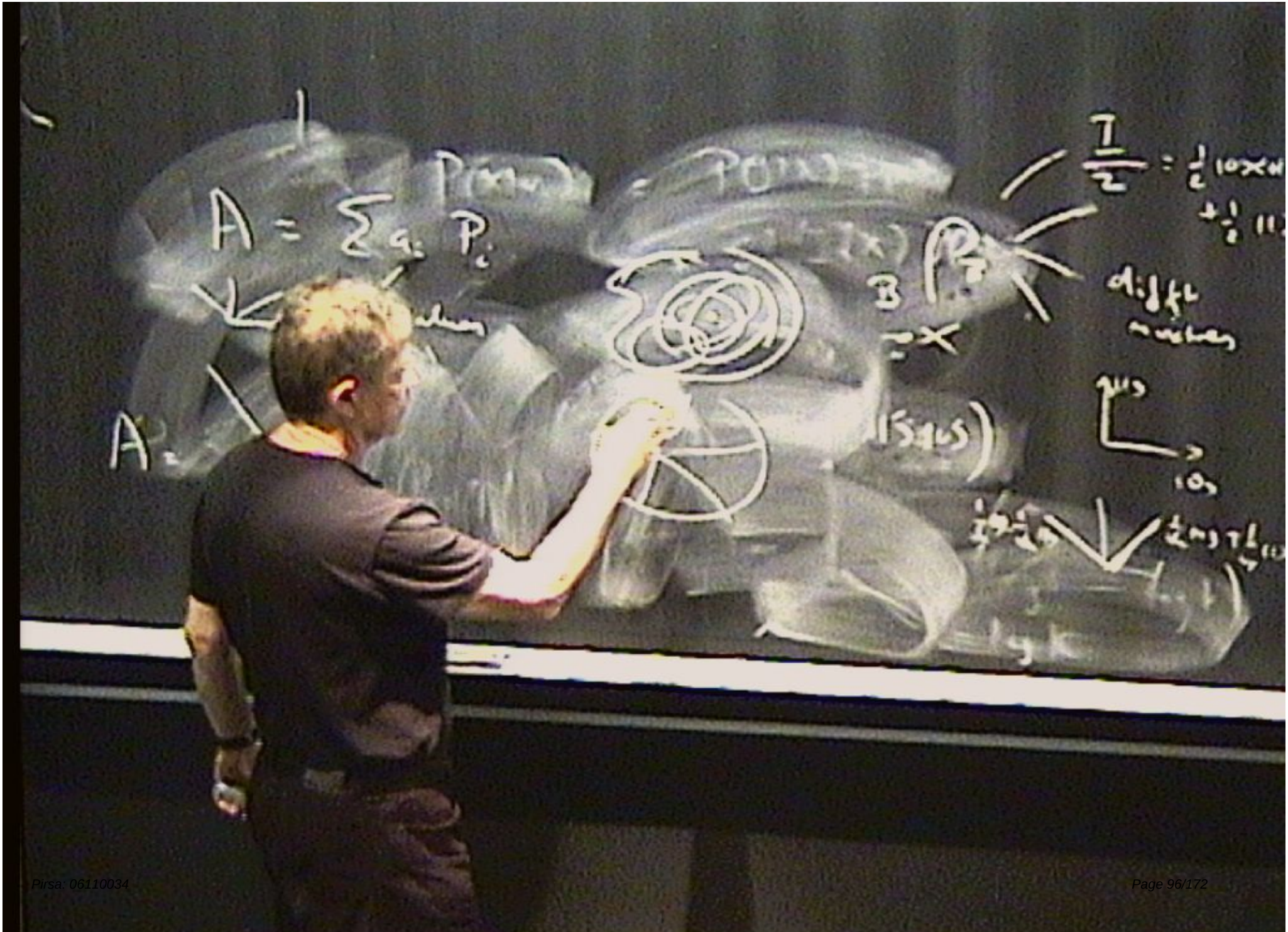
Measurement

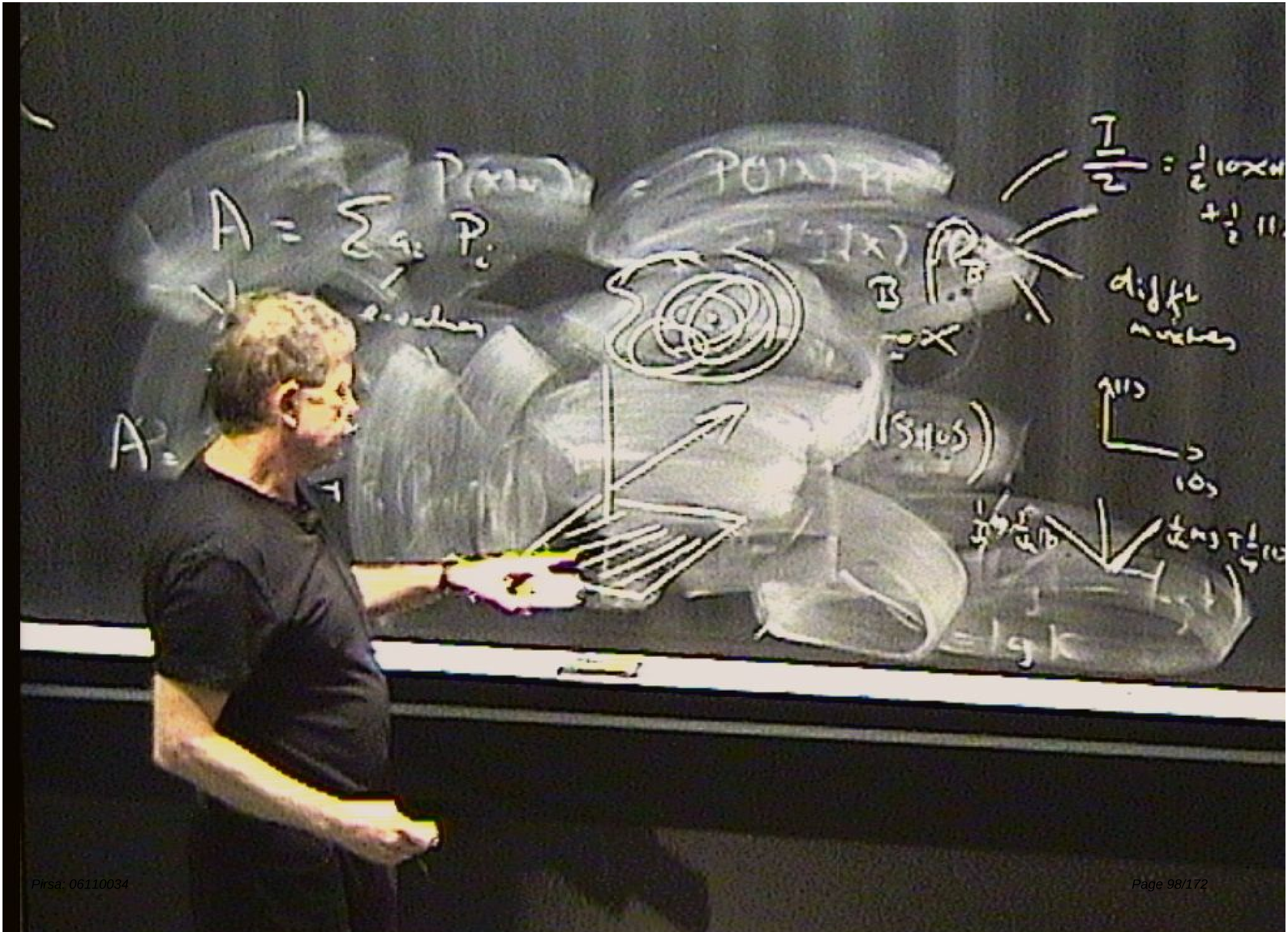
- Classically, we think of a property of a system as being associated with a subset in the state space (phase space) of the system, and determining whether the system has the property amounts to determining whether the state of the system lies in the corresponding subset.
- In quantum mechanics, the counterpart of a subset in phase space is a closed linear subspace in Hilbert space. Just as the different possible values of an observable (dynamical quantity) of a classical system correspond to the subsets in a mutually exclusive and collectively exhaustive set of subsets covering the classical state space, so the different values of a quantum observable correspond to the subspaces in a mutually exclusive (i.e., orthogonal) and collectively exhaustive set of subspaces spanning the quantum state space.

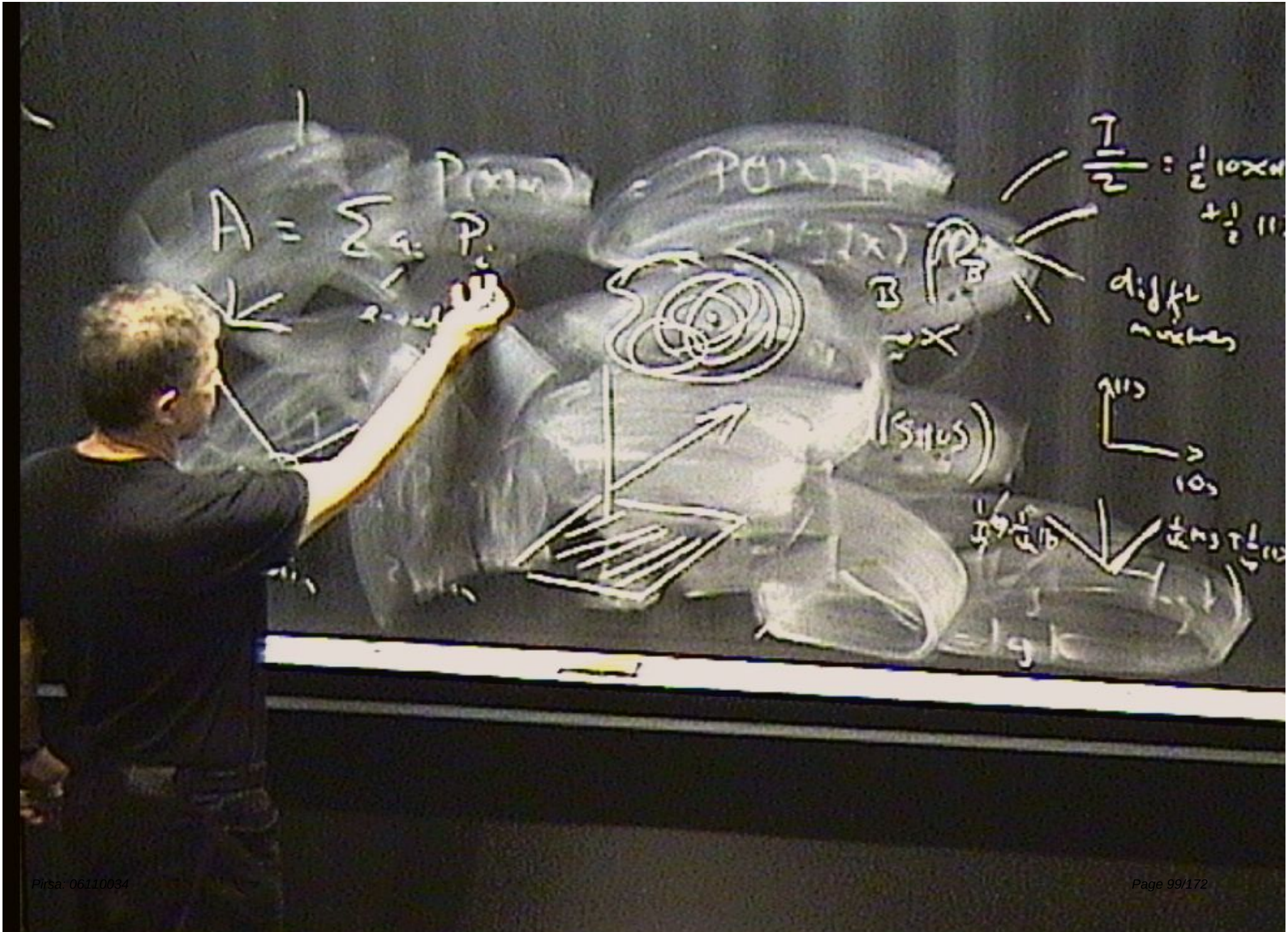

A = Σ P_i

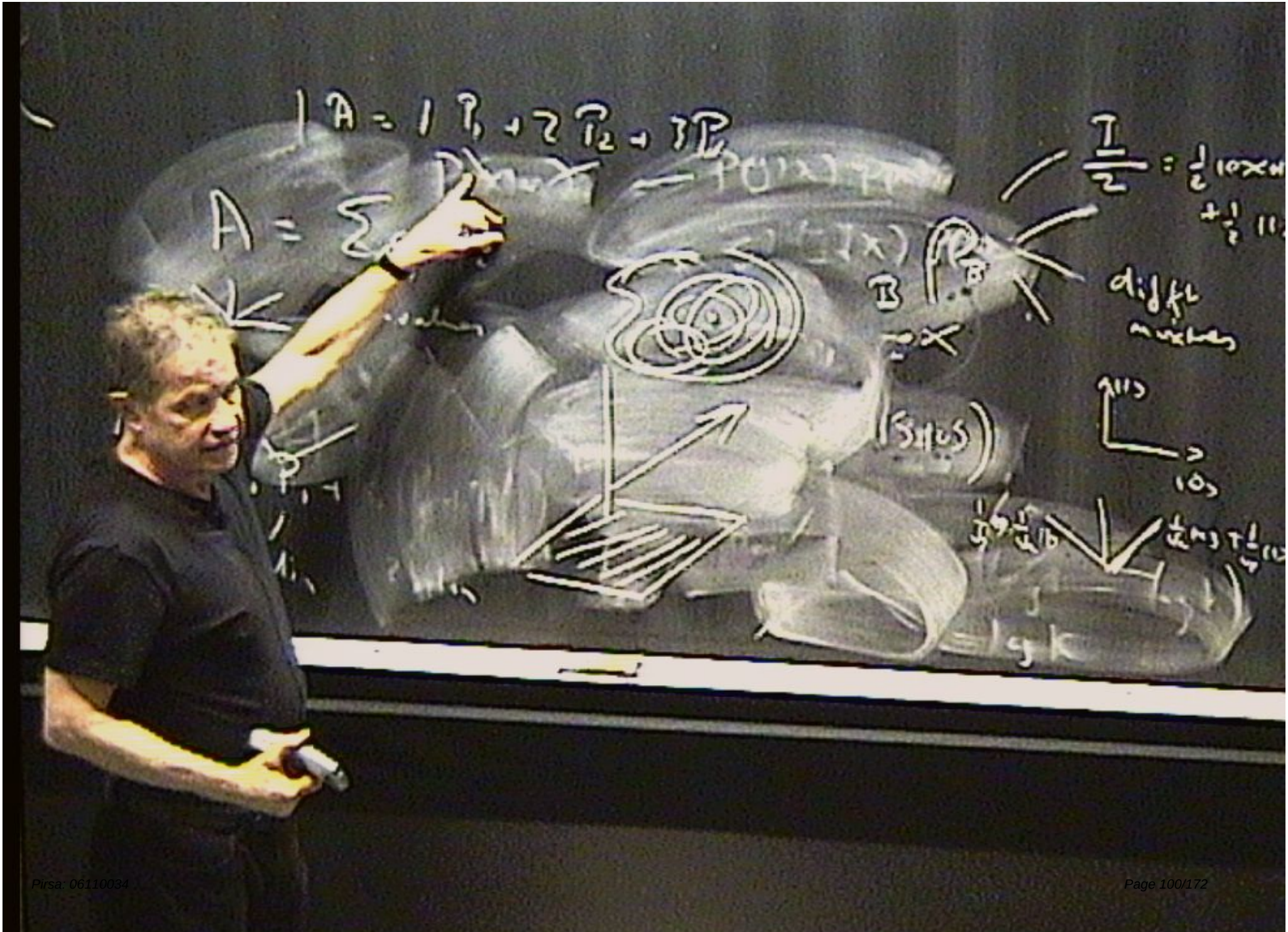

A = Σ a_i P_i

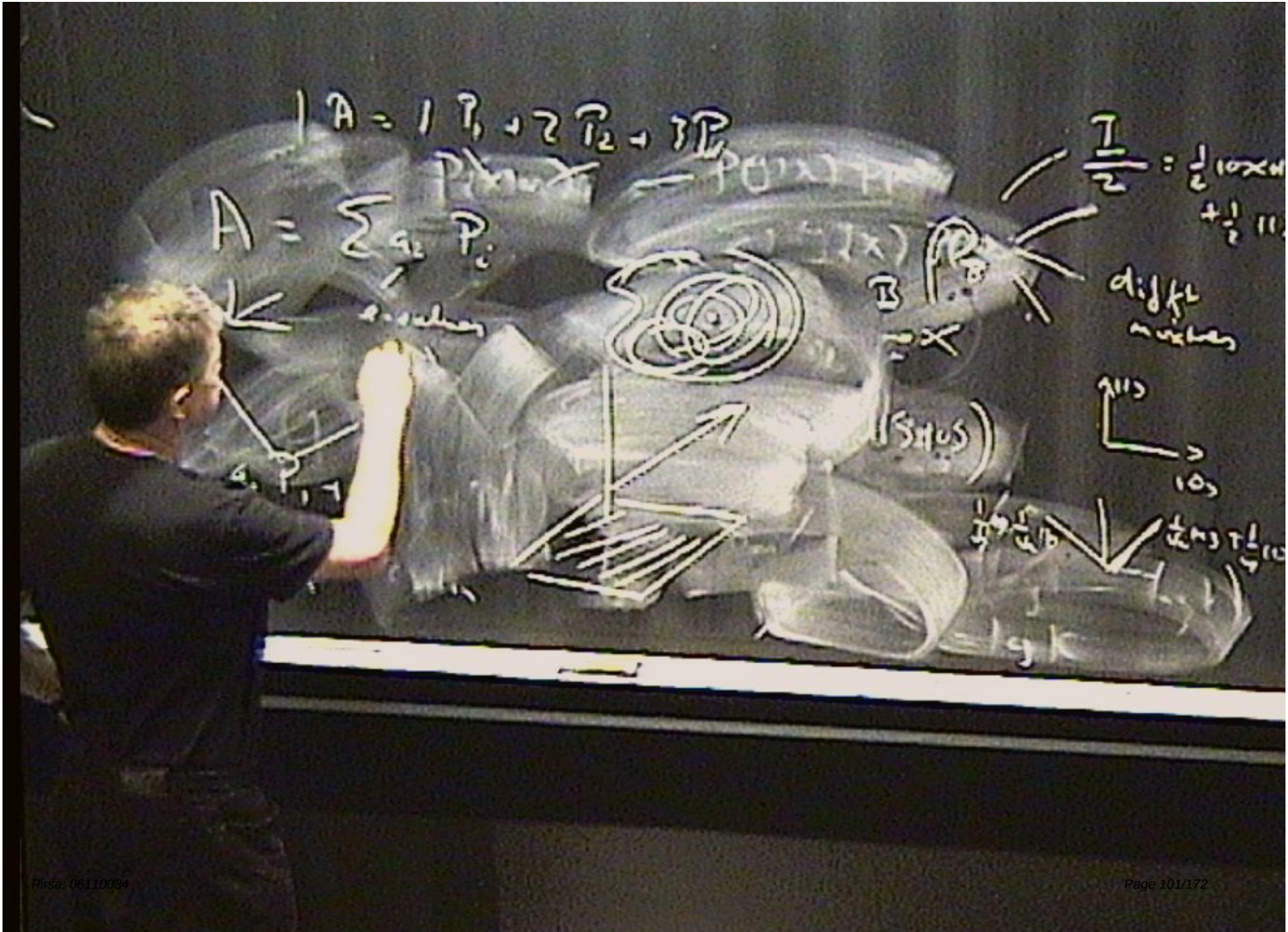

A = Σ q_i P_i

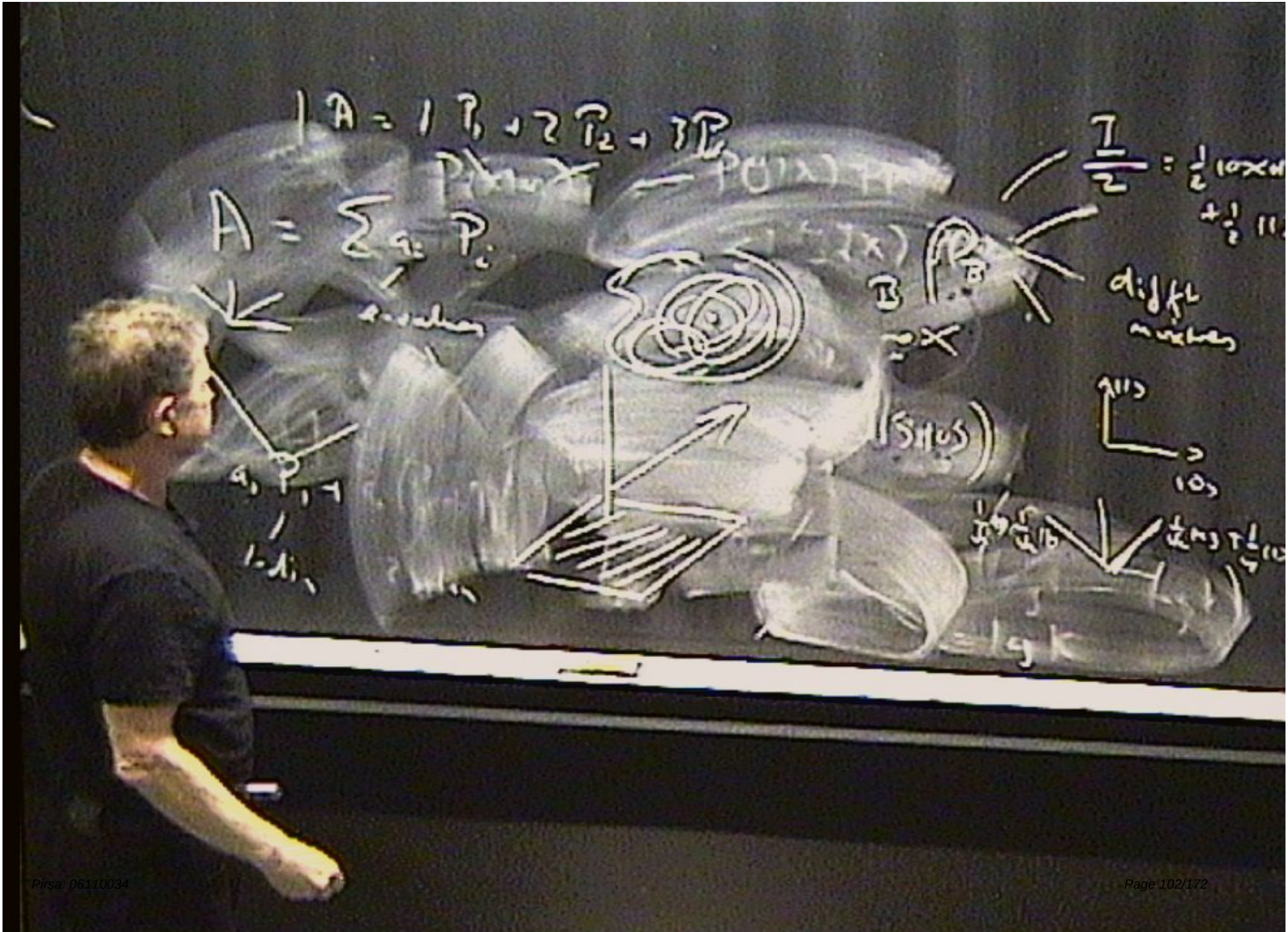

A = Σ q_i P_i


A = Σ q_i P_i


A = Σ q_i P_i

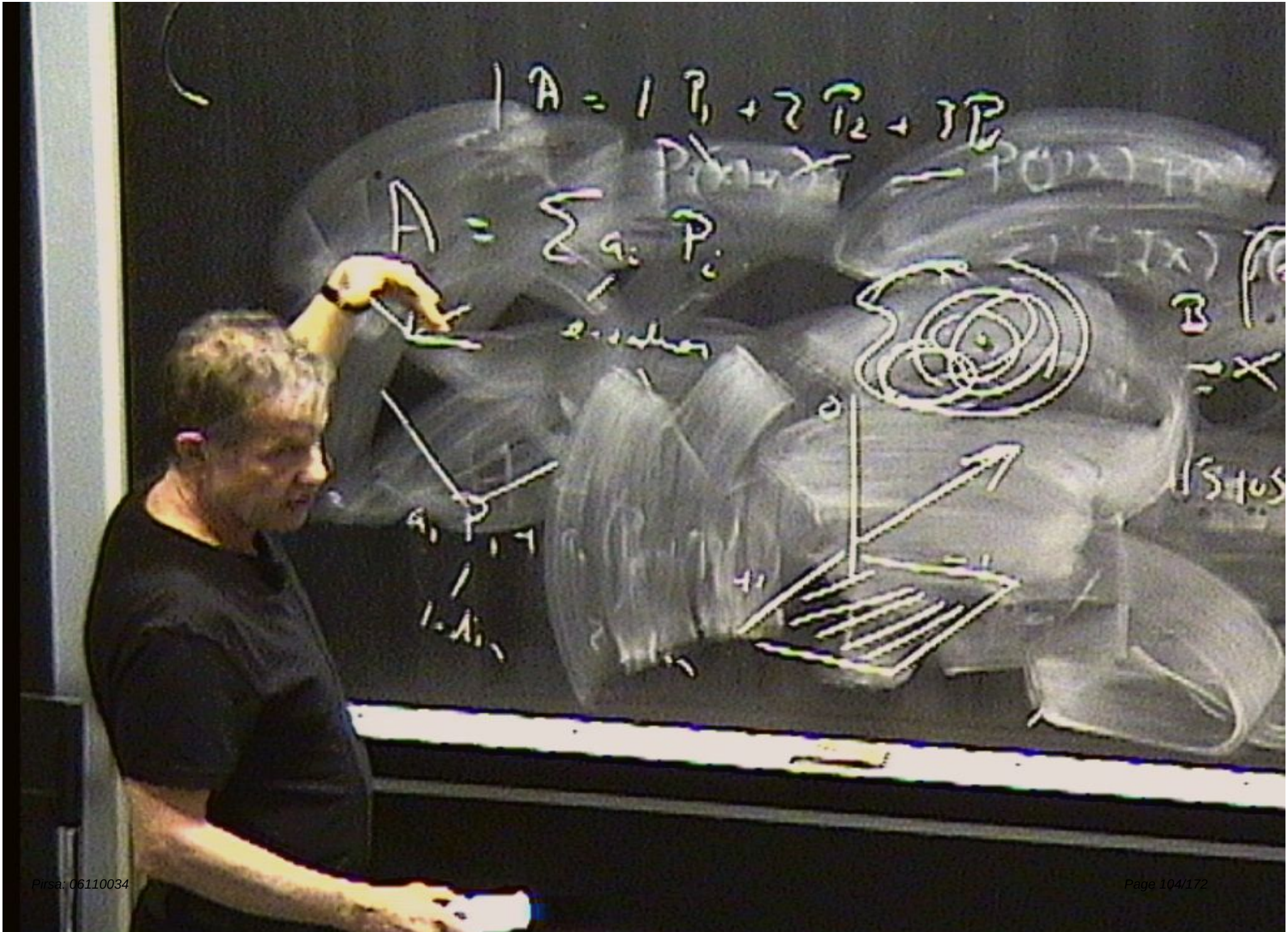


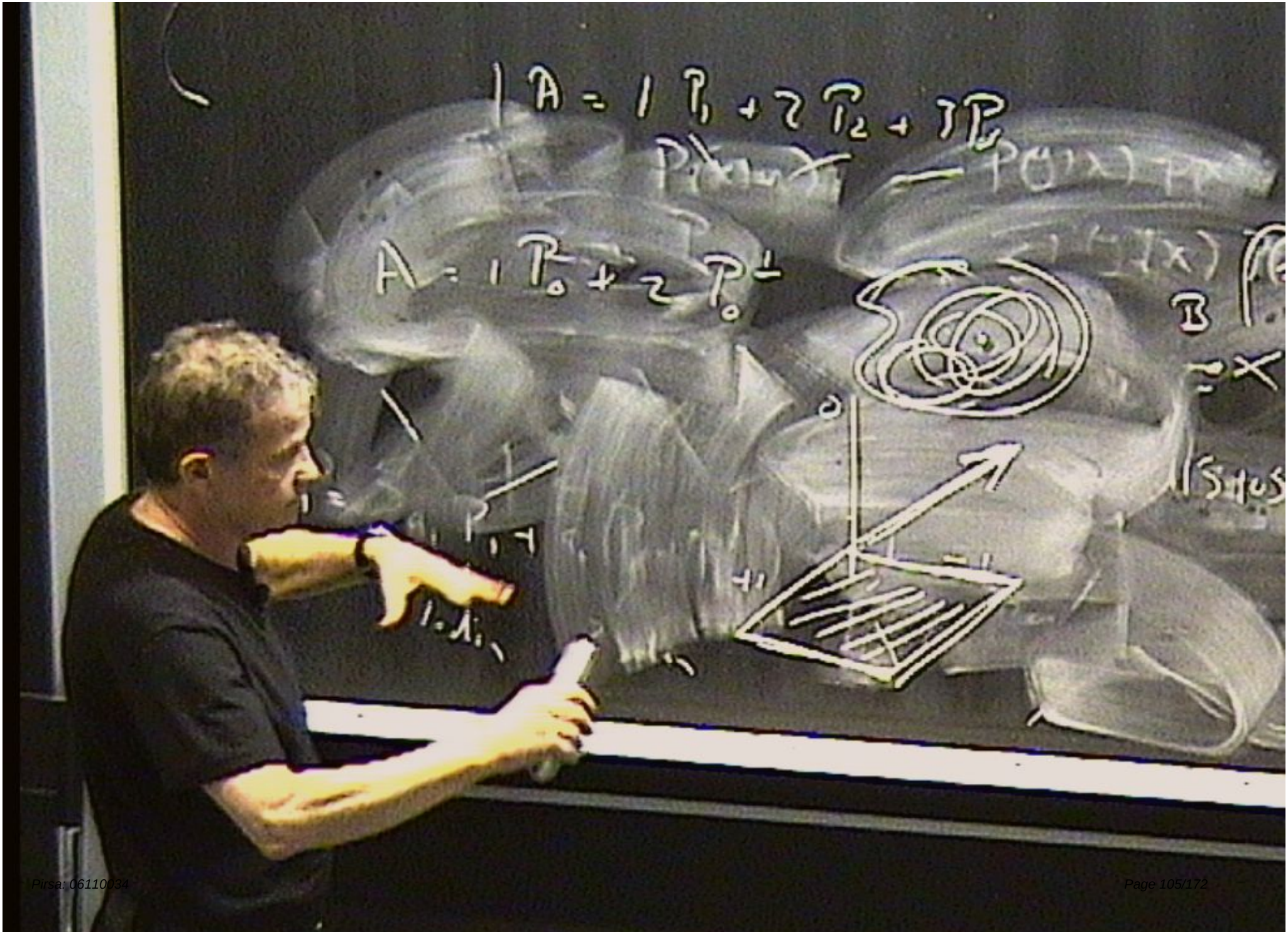

A = Σ a_i P_i

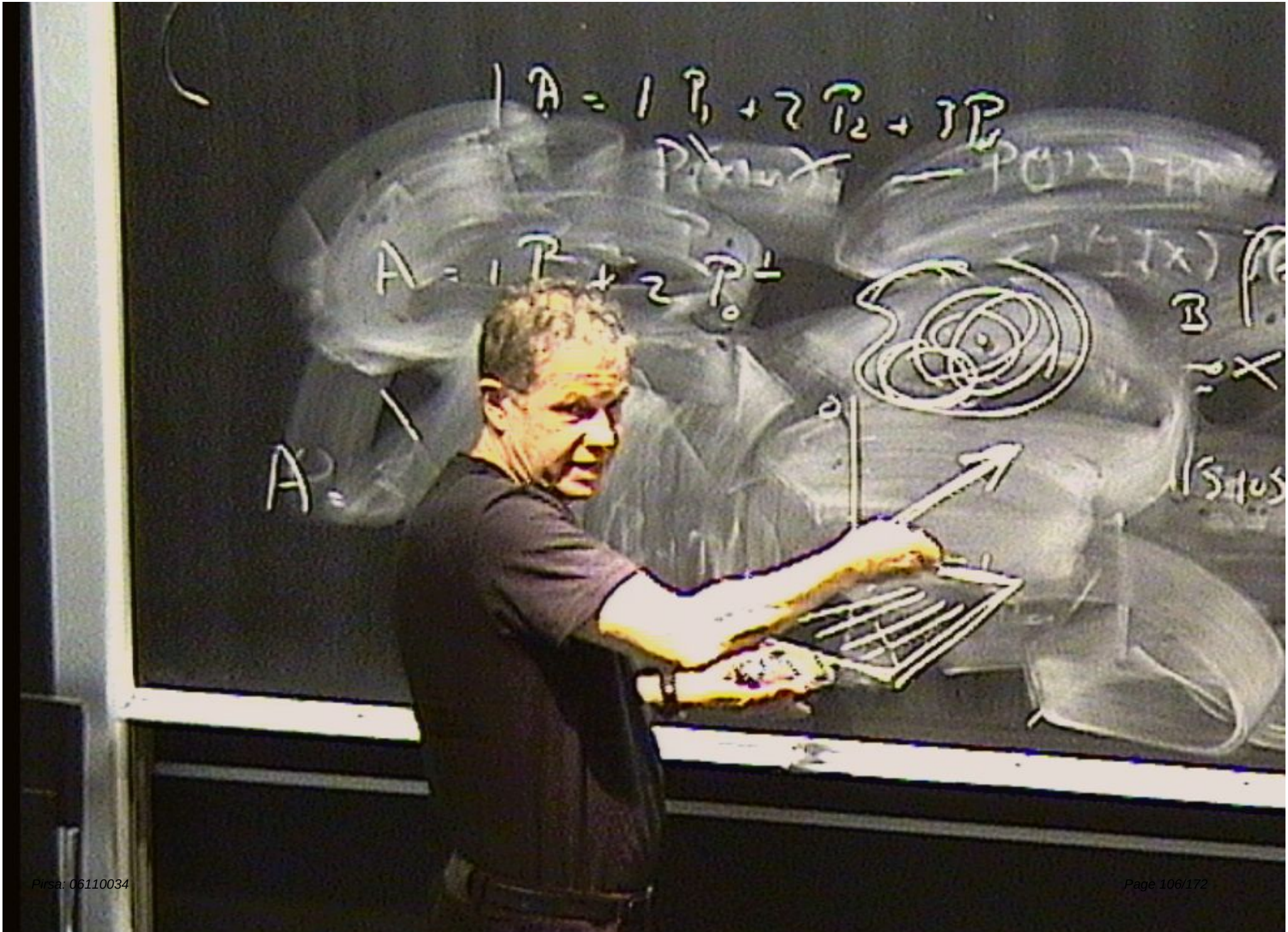

A = Σ qi Pi

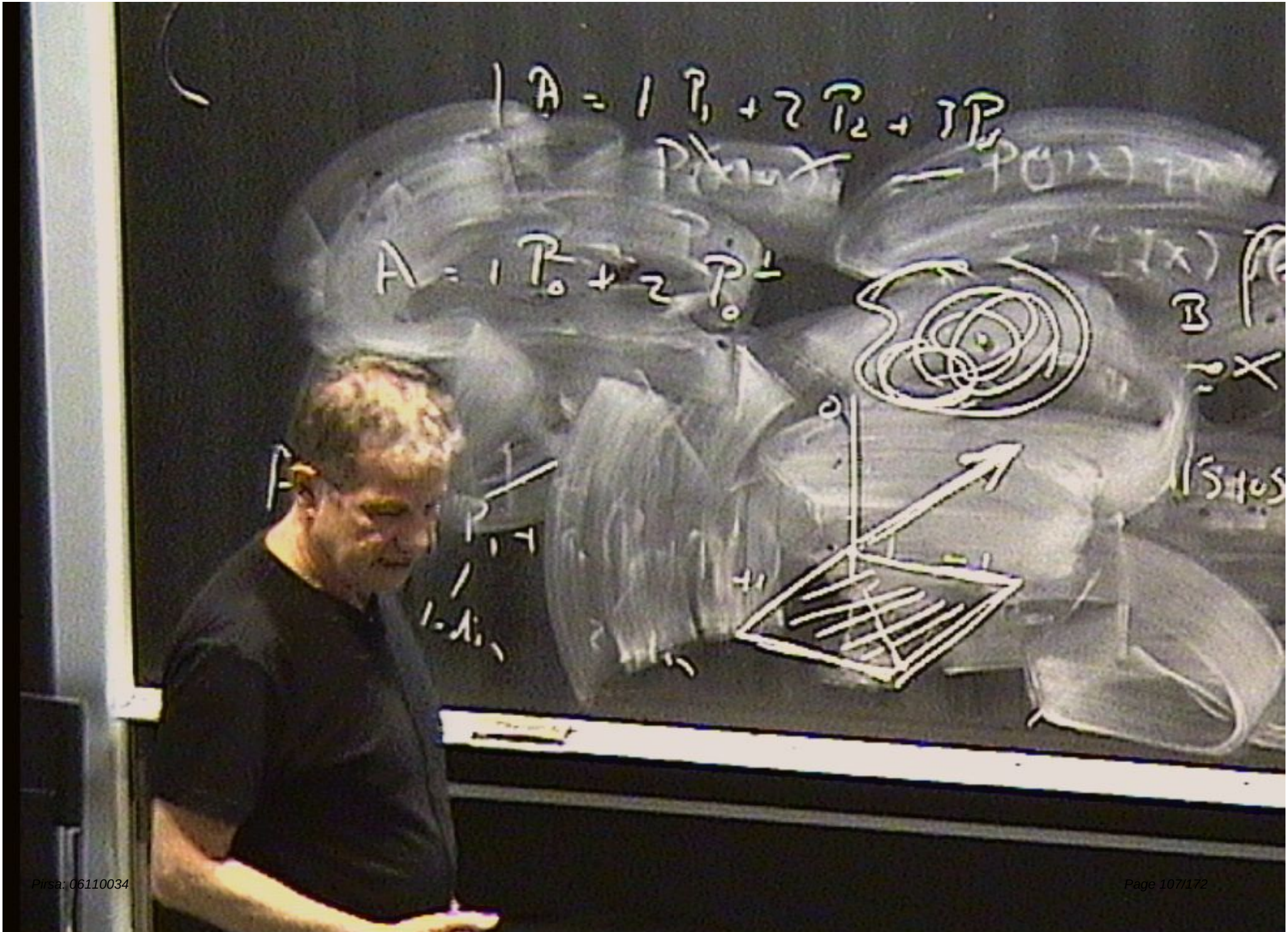
Measurement

- It is useful to consider a more general class of measurements.
- A quantum measurement can be characterized, completely generally, as a certain sort of interaction between two quantum systems, Q (the measured system) and M (the measuring system).
- We suppose that Q is initially in a state $|\psi\rangle$ and that M is initially in some standard state $|0\rangle$, where $|\mathrm{m}\rangle$ is an orthonormal basis of 'pointer' eigenstates in $\mathcal{H}^{\mathrm{M}}$.


A = Σ q_i P_i


A = 1 P1 + 2 P2 + 3 P3
A = 1 P0 + 2 P0⊥




A = 1 P1 + 2 P2 + 3 P3
A = 1 P0 + 2 P0⊥

Measurement

The interaction is defined by a unitary transformation U on the Hilbert space $\mathcal{H}^{Q} \otimes \mathcal{H}^{M}$ that yields the transition:

$$|\psi\rangle|0\rangle \xrightarrow{U} \sum_{m} M_m |\psi\rangle|m\rangle$$

where $\{M_m\}$ is a set of linear operators (the Kraus operators) defined on $\mathcal{H}^{Q}$ satisfying the completeness condition:

$$\sum_{m} M_m^{\dagger} M_m = I$$

(The symbol † denotes the adjoint or Hermitian conjugate.)

Measurement

The completeness condition guarantees that this evolution is unitary, because it guarantees that U preserves inner products, i.e.

$$\begin{aligned}\langle\phi|\langle 0|\mathrm{U}^\dagger\mathrm{U}|\psi\rangle|0\rangle &= \sum_{\mathrm{m},\mathrm{m}'}\langle \mathrm{m}|\langle\phi|\mathrm{M}_\mathrm{m}^\dagger\mathrm{M}_{\mathrm{m}'}|\psi\rangle|\mathrm{m}'\rangle \\ &= \sum_{\mathrm{m}}\langle\phi|\mathrm{M}_\mathrm{m}^\dagger\mathrm{M}_\mathrm{m}|\psi\rangle \\ &= \langle\phi|\psi\rangle\end{aligned}$$

from which it follows that U, defined as above for any product state $|\psi\rangle|0\rangle$ (for any $|\psi\rangle \in \mathcal{H}^\mathrm{Q}$) can be extended to a unitary operator on the Hilbert space $\mathcal{H}^\mathrm{Q} \otimes \mathcal{H}^\mathrm{M}$.

Measurement

Any set of linear operators $\{M_m\}$ defined on the Hilbert space of the system Q satisfying the completeness condition defines a measurement in this general sense, with the index m labeling the possible outcomes of the measurement, and any such set is referred to as a set of measurement operators.

Measurement

Any set of linear operators $\{M_m\}$ defined on the Hilbert space of the system Q satisfying the completeness condition defines a measurement in this general sense, with the index m labeling the possible outcomes of the measurement, and any such set is referred to as a set of measurement operators.

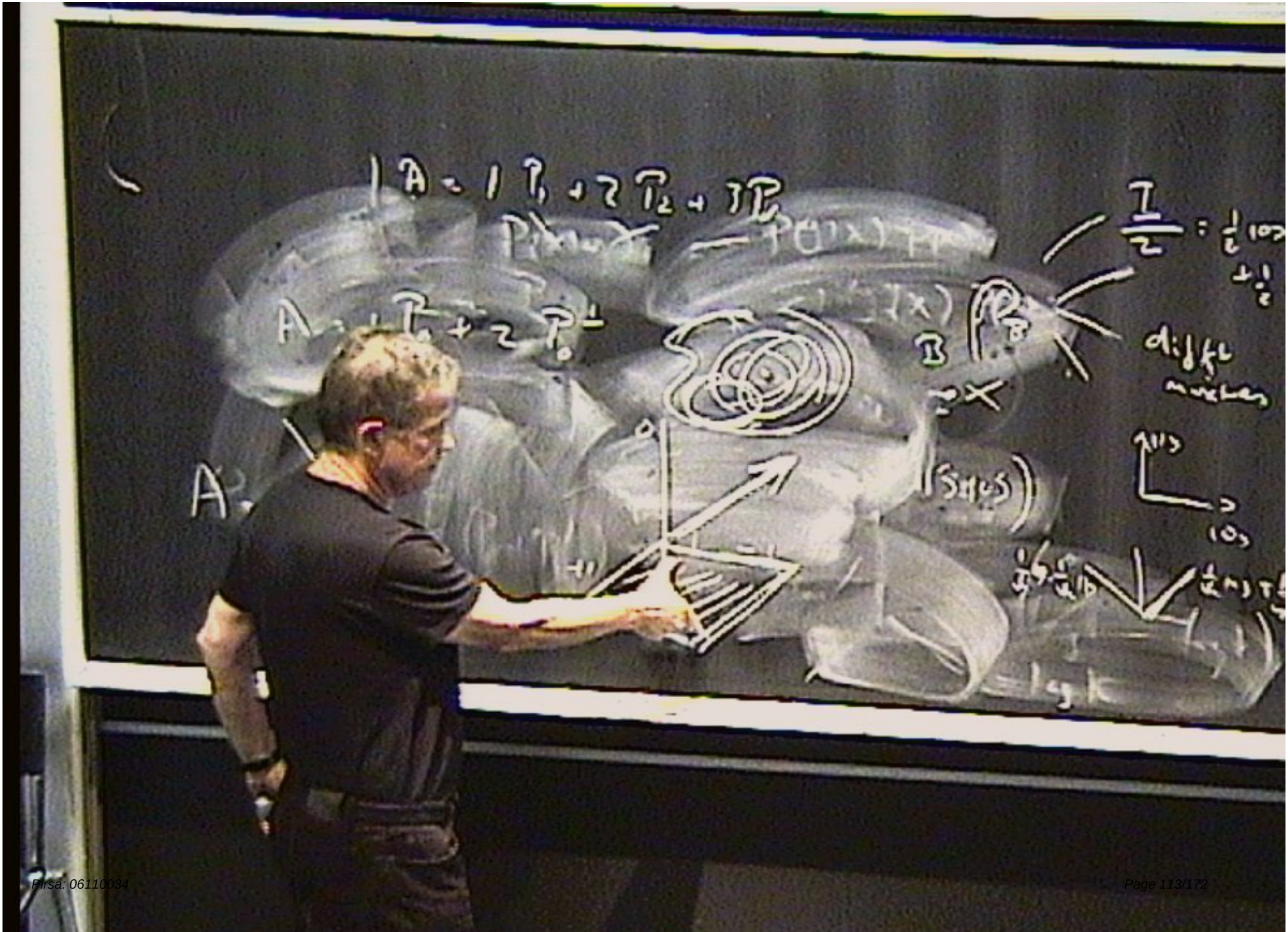
Measurement

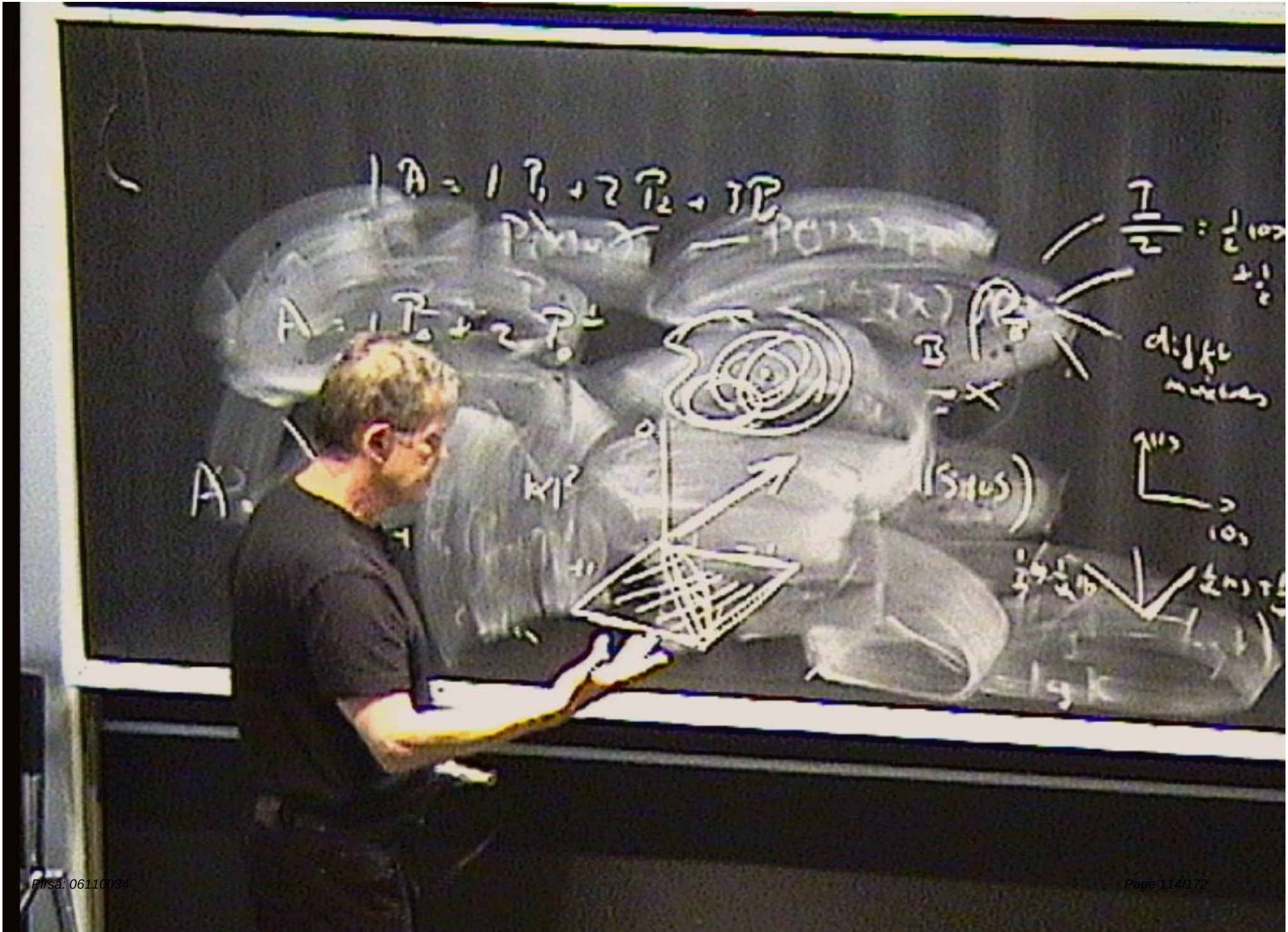
If we now perform a standard projective measurement on M to determine the value m of the pointer observable, defined by the projection operator

$$\mathrm{P_m} = \mathrm{I_Q} \otimes |\mathrm{m}\rangle\langle\mathrm{m}|$$

then the probability of obtaining the outcome m is:

$$\begin{aligned}
\mathrm{p(m)} &= \langle 0|\langle\psi|\mathrm{U}^\dagger \mathrm{P_m U}|\psi\rangle|0\rangle \\
&= \sum_{\mathrm{m'm''}} \langle \mathrm{m'}|\langle\psi|\mathrm{M}^\dagger_{\mathrm{m'}}(\mathrm{I_Q}\otimes|\mathrm{m}\rangle\langle\mathrm{m}|)\mathrm{M_{m''}}|\psi\rangle|\mathrm{m''}\rangle \\
&= \sum_{\mathrm{m'm''}} \langle\psi|\mathrm{M}^\dagger_{\mathrm{m'}}\langle\mathrm{m'}|\mathrm{m}\rangle\langle\mathrm{m}|\mathrm{m''}\rangle\mathrm{M_{m''}}|\psi\rangle \\
&= \langle\psi|\mathrm{M}^\dagger_{\mathrm{m}}\mathrm{M_m}|\psi\rangle
\end{aligned}$$





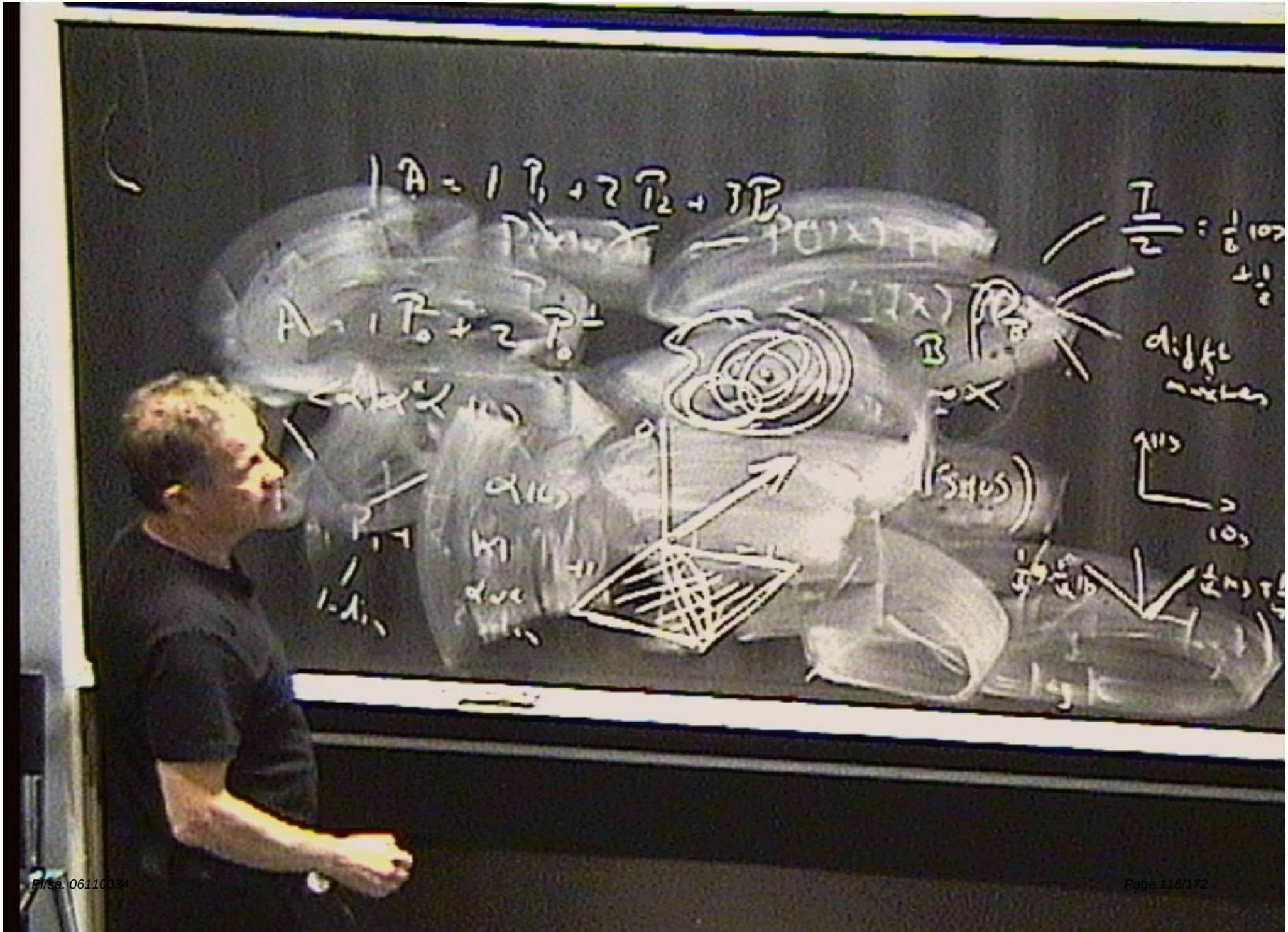
Measurement

If we now perform a standard projective measurement on M to determine the value m of the pointer observable, defined by the projection operator

$$\mathrm{P_m} = \mathrm{I_Q} \otimes |\mathrm{m}\rangle\langle\mathrm{m}|$$

then the probability of obtaining the outcome m is:

$$\begin{aligned}
\mathrm{p(m)} &= \langle 0|\langle\psi|\mathrm{U}^\dagger \mathrm{P_m U}|\psi\rangle|0\rangle \\
&= \sum_{\mathrm{m'm''}} \langle \mathrm{m'}|\langle\psi|\mathrm{M}^\dagger_{\mathrm{m'}}(\mathrm{I_Q} \otimes |\mathrm{m}\rangle\langle\mathrm{m}|)\mathrm{M_{m''}}|\psi\rangle|\mathrm{m''}\rangle \\
&= \sum_{\mathrm{m'm''}} \langle\psi|\mathrm{M}^\dagger_{\mathrm{m'}}\langle\mathrm{m'}|\mathrm{m}\rangle\langle\mathrm{m}|\mathrm{m''}\rangle\mathrm{M_{m''}}|\psi\rangle \\
&= \langle\psi|\mathrm{M}^\dagger_{\mathrm{m}}\mathrm{M_m}|\psi\rangle
\end{aligned}$$



Measurement

If we now perform a standard projective measurement on M to determine the value m of the pointer observable, defined by the projection operator

$$\mathrm{P_m} = \mathrm{I_Q} \otimes |\mathrm{m}\rangle\langle\mathrm{m}|$$

then the probability of obtaining the outcome m is:

$$\begin{aligned}
\mathrm{p(m)} &= \langle 0|\langle\psi|\mathrm{U}^\dagger \mathrm{P_m U}|\psi\rangle|0\rangle \\
&= \sum_{\mathrm{m'm''}} \langle \mathrm{m'}|\langle\psi|\mathrm{M}^\dagger_{\mathrm{m'}}(\mathrm{I_Q}\otimes|\mathrm{m}\rangle\langle\mathrm{m}|)\mathrm{M_{m''}}|\psi\rangle|\mathrm{m''}\rangle \\
&= \sum_{\mathrm{m'm''}} \langle\psi|\mathrm{M}^\dagger_{\mathrm{m'}}\langle\mathrm{m'}|\mathrm{m}\rangle\langle\mathrm{m}|\mathrm{m''}\rangle\mathrm{M_{m''}}|\psi\rangle \\
&= \langle\psi|\mathrm{M}^\dagger_{\mathrm{m}}\mathrm{M_m}|\psi\rangle
\end{aligned}$$

Measurement

Any set of linear operators $\{M_m\}$ defined on the Hilbert space of the system Q satisfying the completeness condition defines a measurement in this general sense, with the index m labeling the possible outcomes of the measurement, and any such set is referred to as a set of measurement operators.

Measurement

The completeness condition guarantees that this evolution is unitary, because it guarantees that U preserves inner products, i.e.

$$\begin{aligned}\langle\phi|\langle 0|\mathrm{U}^\dagger\mathrm{U}|\psi\rangle|0\rangle &= \sum_{\mathrm{m},\mathrm{m}'}\langle \mathrm{m}|\langle\phi|\mathrm{M}_\mathrm{m}^\dagger\mathrm{M}_{\mathrm{m}'}|\psi\rangle|\mathrm{m}'\rangle \\ &= \sum_{\mathrm{m}}\langle\phi|\mathrm{M}_\mathrm{m}^\dagger\mathrm{M}_\mathrm{m}|\psi\rangle \\ &= \langle\phi|\psi\rangle\end{aligned}$$

from which it follows that U, defined as above for any product state $|\psi\rangle|0\rangle$ (for any $|\psi\rangle \in \mathcal{H}^\mathrm{Q}$) can be extended to a unitary operator on the Hilbert space $\mathcal{H}^\mathrm{Q} \otimes \mathcal{H}^\mathrm{M}$.

Measurement

The interaction is defined by a unitary transformation U on the Hilbert space $\mathcal{H}^{\mathrm{Q}} \otimes \mathcal{H}^{\mathrm{M}}$ that yields the transition:

$$|\psi\rangle|0\rangle \xrightarrow{\mathrm{U}} \sum_{\mathrm{m}} \mathrm{M_m}|\psi\rangle|\mathrm{m}\rangle$$

where $\{\mathrm{M_m}\}$ is a set of linear operators (the Kraus operators) defined on $\mathcal{H}^{\mathrm{Q}}$ satisfying the completeness condition:

$$\sum_{\mathrm{m}} \mathrm{M_m^\dagger M_m} = \mathrm{I}$$

(The symbol † denotes the adjoint or Hermitian conjugate.)

Measurement

Any set of linear operators $\{M_m\}$ defined on the Hilbert space of the system Q satisfying the completeness condition defines a measurement in this general sense, with the index m labeling the possible outcomes of the measurement, and any such set is referred to as a set of measurement operators.

Measurement

If we now perform a standard projective measurement on M to determine the value m of the pointer observable, defined by the projection operator

$$\mathrm{P_m} = \mathrm{I_Q} \otimes |\mathrm{m}\rangle\langle\mathrm{m}|$$

then the probability of obtaining the outcome m is:

$$\begin{aligned}
\mathrm{p(m)} &= \langle 0|\langle\psi|\mathrm{U}^\dagger \mathrm{P_m U}|\psi\rangle|0\rangle \\
&= \sum_{\mathrm{m'm''}} \langle \mathrm{m'}|\langle\psi|\mathrm{M}^\dagger_{\mathrm{m'}}(\mathrm{I_Q} \otimes |\mathrm{m}\rangle\langle\mathrm{m}|)\mathrm{M_{m''}}|\psi\rangle|\mathrm{m''}\rangle \\
&= \sum_{\mathrm{m'm''}} \langle\psi|\mathrm{M}^\dagger_{\mathrm{m'}}\langle\mathrm{m'}|\mathrm{m}\rangle\langle\mathrm{m}|\mathrm{m''}\rangle\mathrm{M_{m''}}|\psi\rangle \\
&= \langle\psi|\mathrm{M}^\dagger_{\mathrm{m}}\mathrm{M_m}|\psi\rangle
\end{aligned}$$

Measurement

- More generally, if the initial state of Q is a mixed state ρ, then

$$p(m) = Tr_Q(M_m \rho M_m^\dagger)$$

- The final state of QM after the projective measurement on M yielding the outcome m is:

$$\frac{P_m U|\psi\rangle|0\rangle}{\sqrt{\langle\psi|U^\dagger P U|\psi\rangle}} = \frac{M_m|\psi\rangle|m\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}}$$

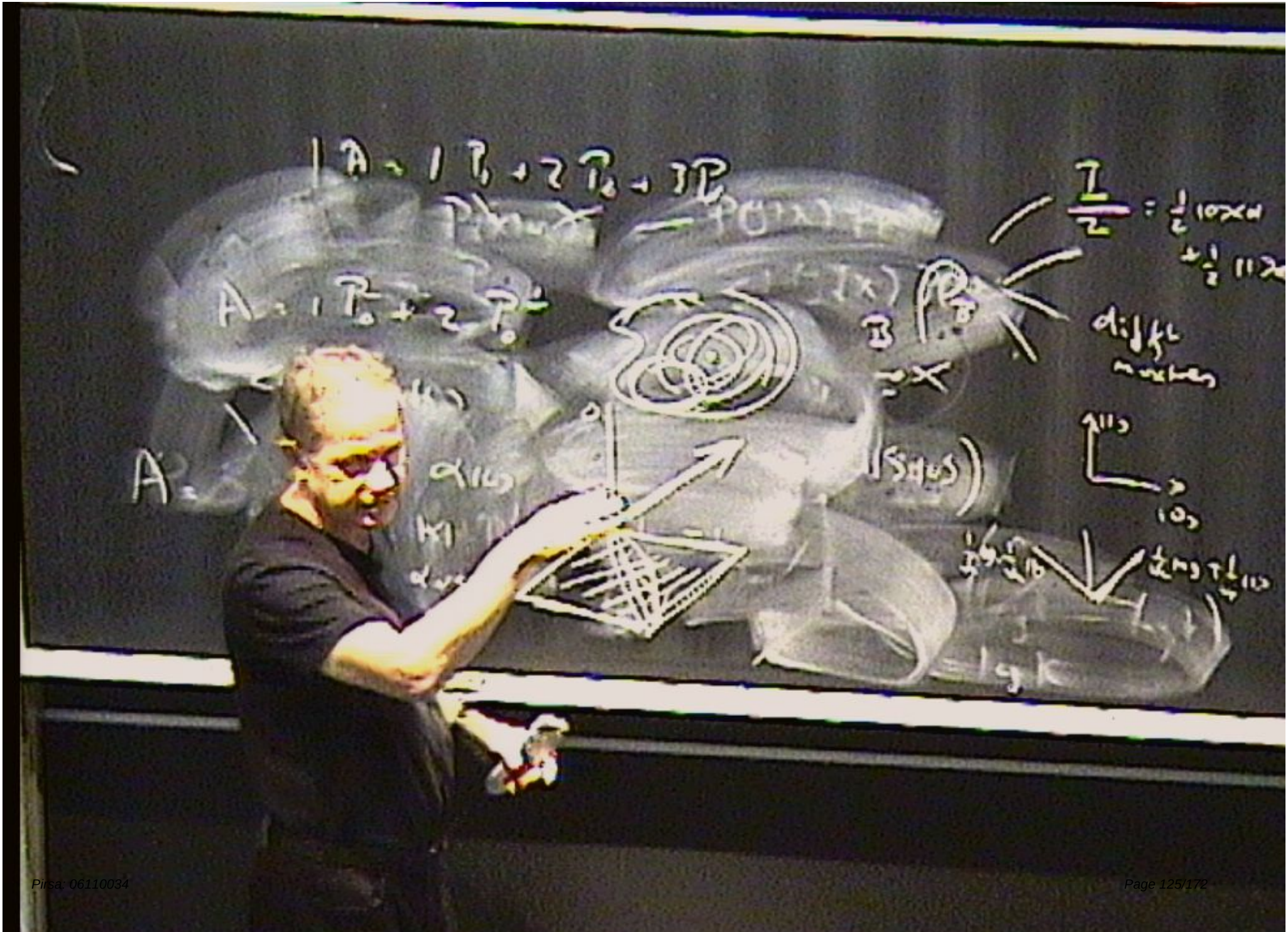
Measurement

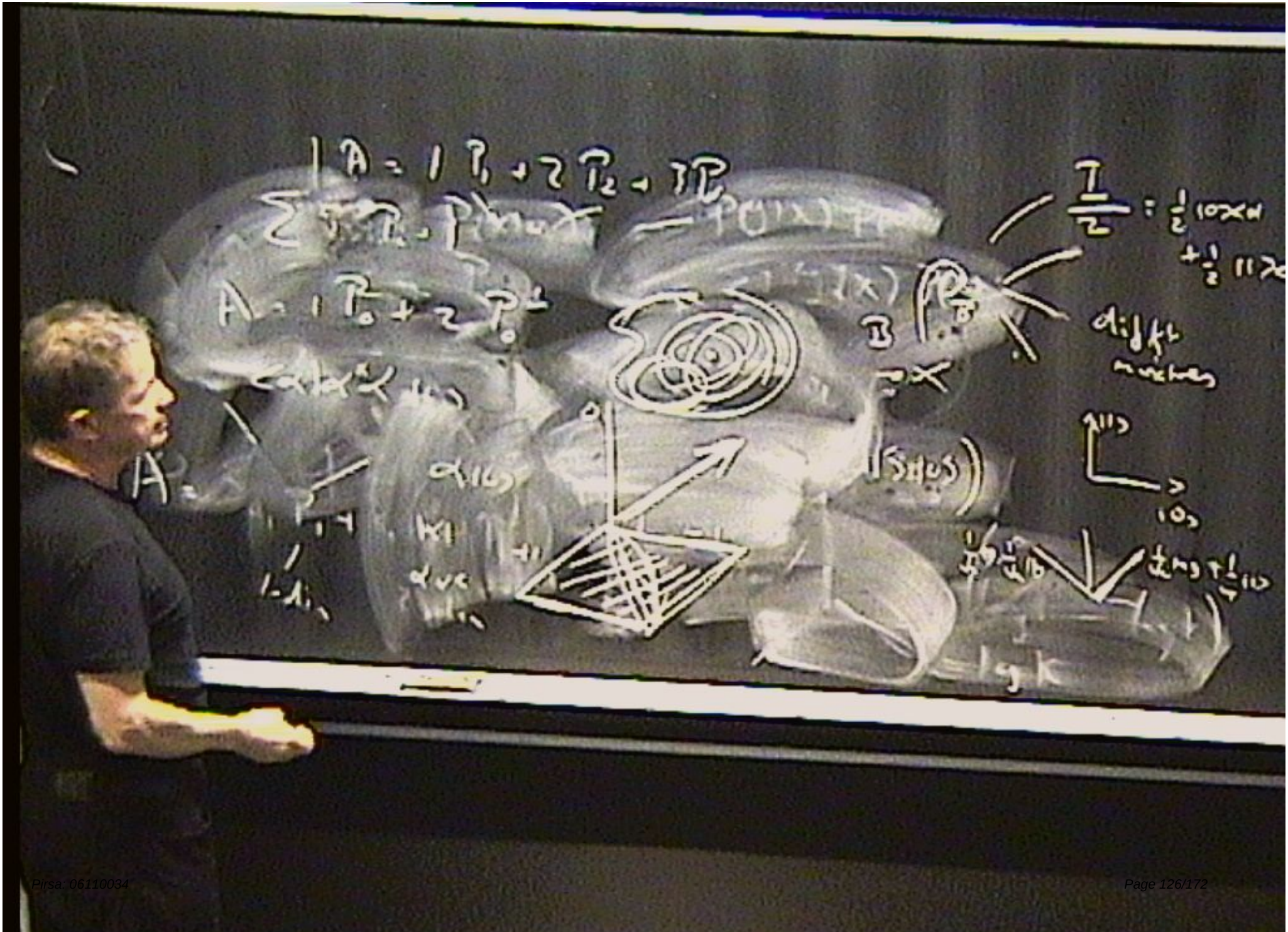
So the final state of M is $|\mathrm{m}\rangle$ and the final state of Q is:

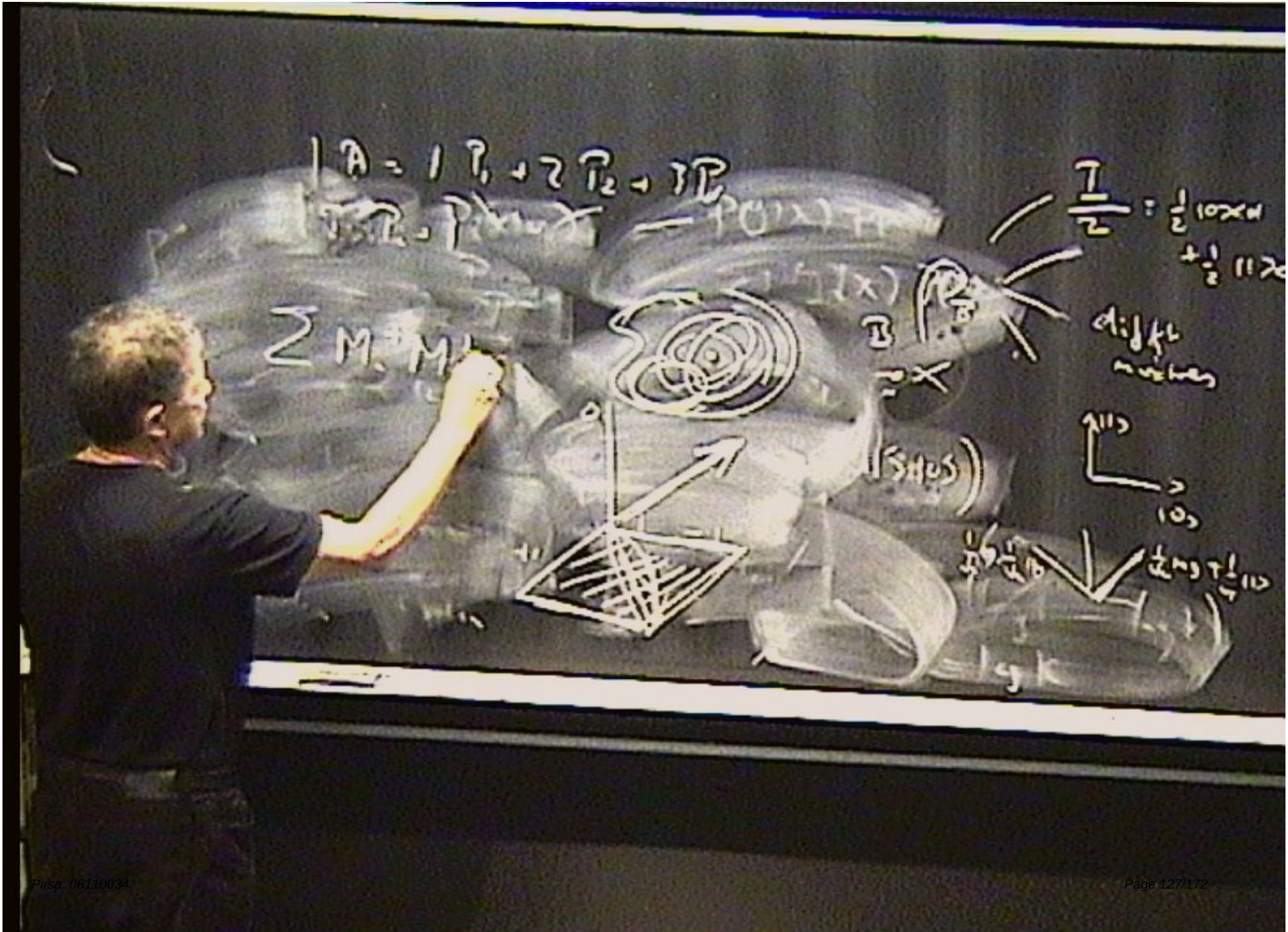
$$\frac{\mathrm{M_m}|\psi\rangle}{\sqrt{\langle\psi|\mathrm{M_m^\dagger M_m}|\psi\rangle}};$$

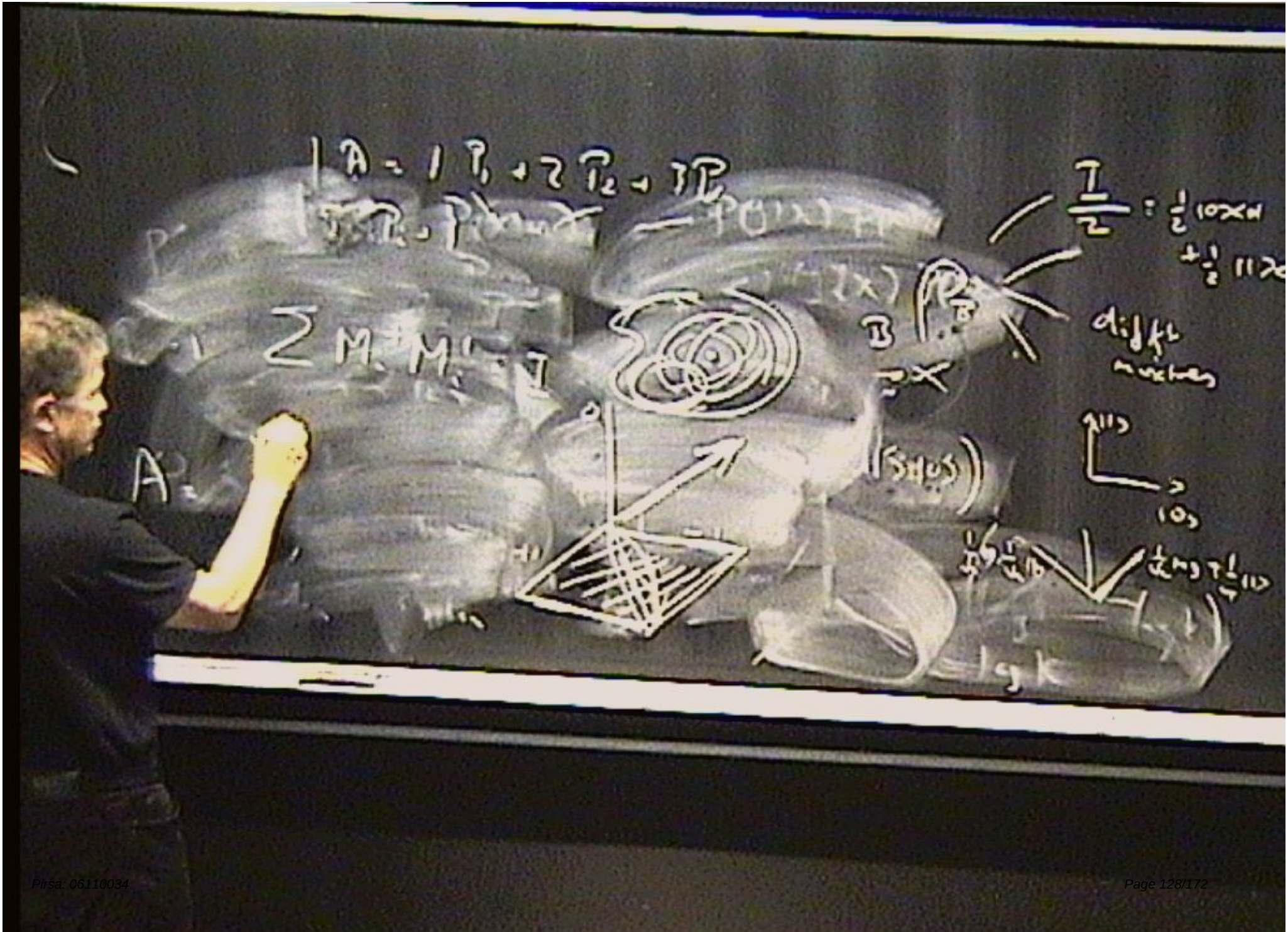
and, more generally, if the initial state of Q is a mixed state ρ, then the final state of Q is:

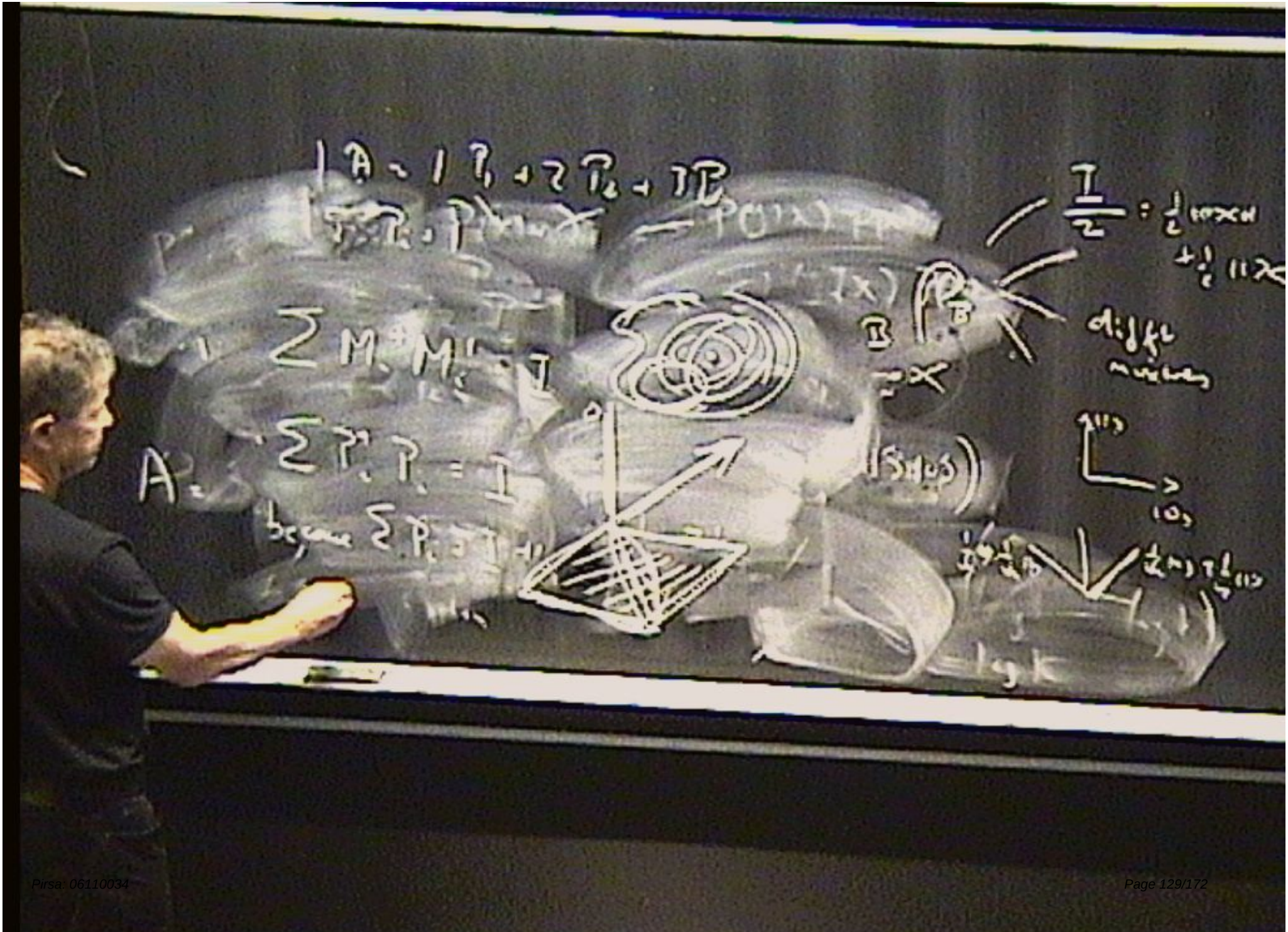
$$\frac{\mathrm{M_m}\rho\mathrm{M_m^\dagger}}{\mathrm{Tr_Q}(\mathrm{M_m}\rho\mathrm{M_m^\dagger})}$$

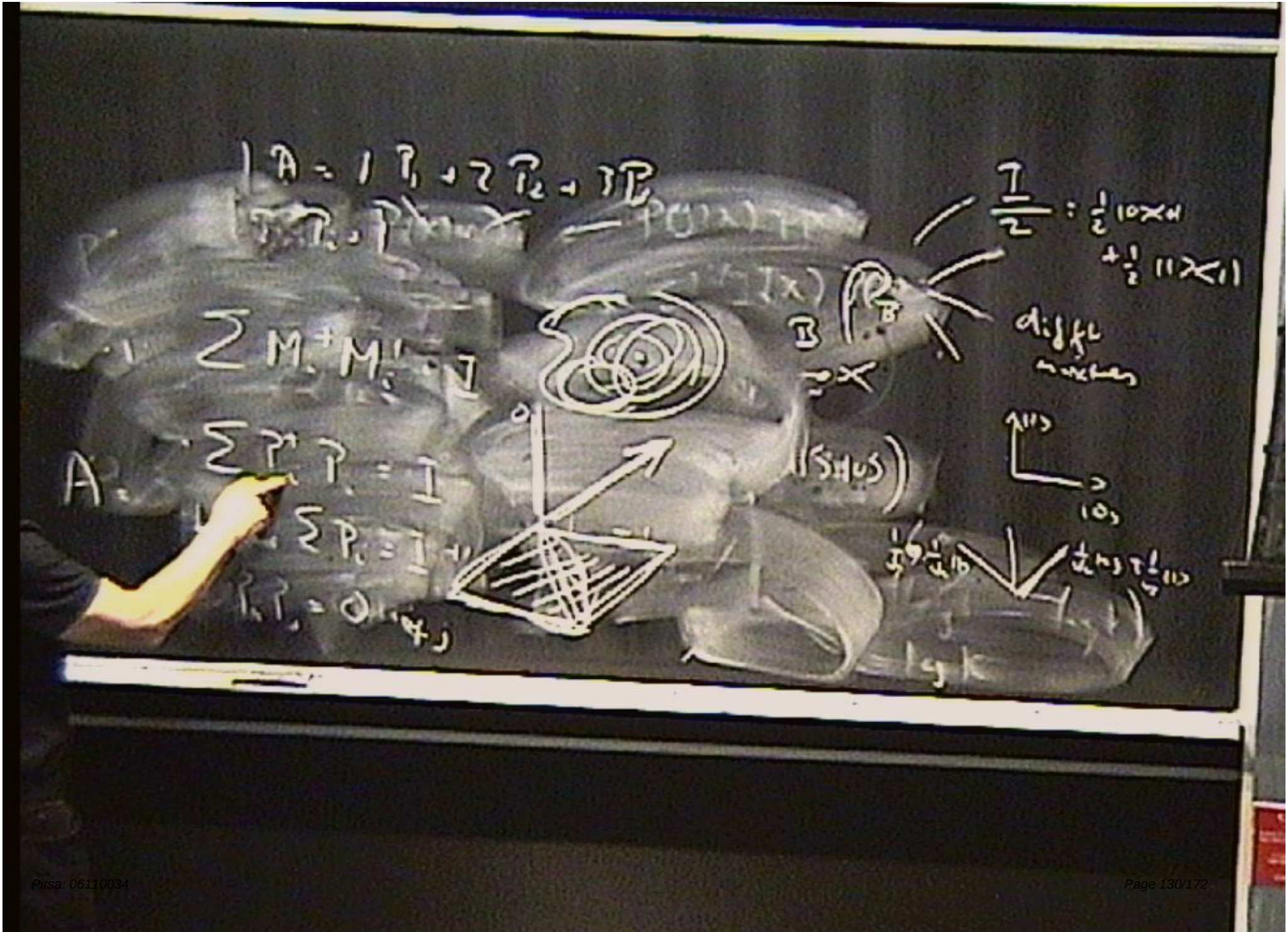


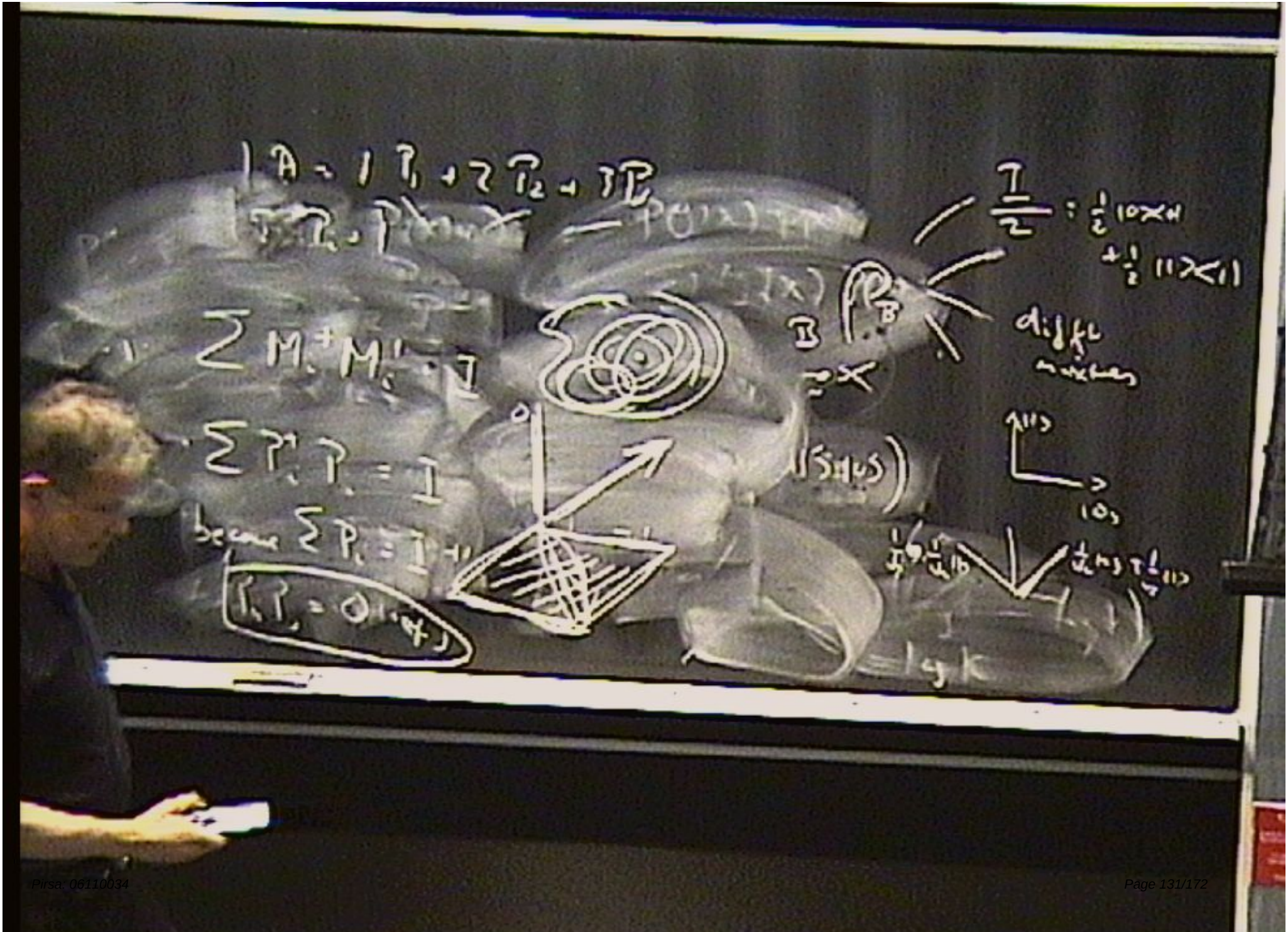












Measurement

- More generally, if the initial state of Q is a mixed state ρ, then

$$p(m) = Tr_Q(M_m \rho M_m^\dagger)$$

- The final state of QM after the projective measurement on M yielding the outcome m is:

$$\frac{P_m U|\psi\rangle|0\rangle}{\sqrt{\langle\psi|U^\dagger P U|\psi\rangle}} = \frac{M_m|\psi\rangle|m\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}}$$

Measurement

If we now perform a standard projective measurement on M to determine the value m of the pointer observable, defined by the projection operator

$$\mathrm{P_m} = \mathrm{I_Q} \otimes |\mathrm{m}\rangle\langle\mathrm{m}|$$

then the probability of obtaining the outcome m is:

$$\begin{aligned}
\mathrm{p(m)} &= \langle 0|\langle\psi|\mathrm{U}^\dagger \mathrm{P_m U}|\psi\rangle|0\rangle \\
&= \sum_{\mathrm{m'm''}} \langle \mathrm{m'}|\langle\psi|\mathrm{M}^\dagger_{\mathrm{m'}}(\mathrm{I_Q} \otimes |\mathrm{m}\rangle\langle\mathrm{m}|)\mathrm{M_{m''}}|\psi\rangle|\mathrm{m''}\rangle \\
&= \sum_{\mathrm{m'm''}} \langle\psi|\mathrm{M}^\dagger_{\mathrm{m'}}\langle\mathrm{m'}|\mathrm{m}\rangle\langle\mathrm{m}|\mathrm{m''}\rangle\mathrm{M_{m''}}|\psi\rangle \\
&= \langle\psi|\mathrm{M}^\dagger_{\mathrm{m}}\mathrm{M_m}|\psi\rangle
\end{aligned}$$

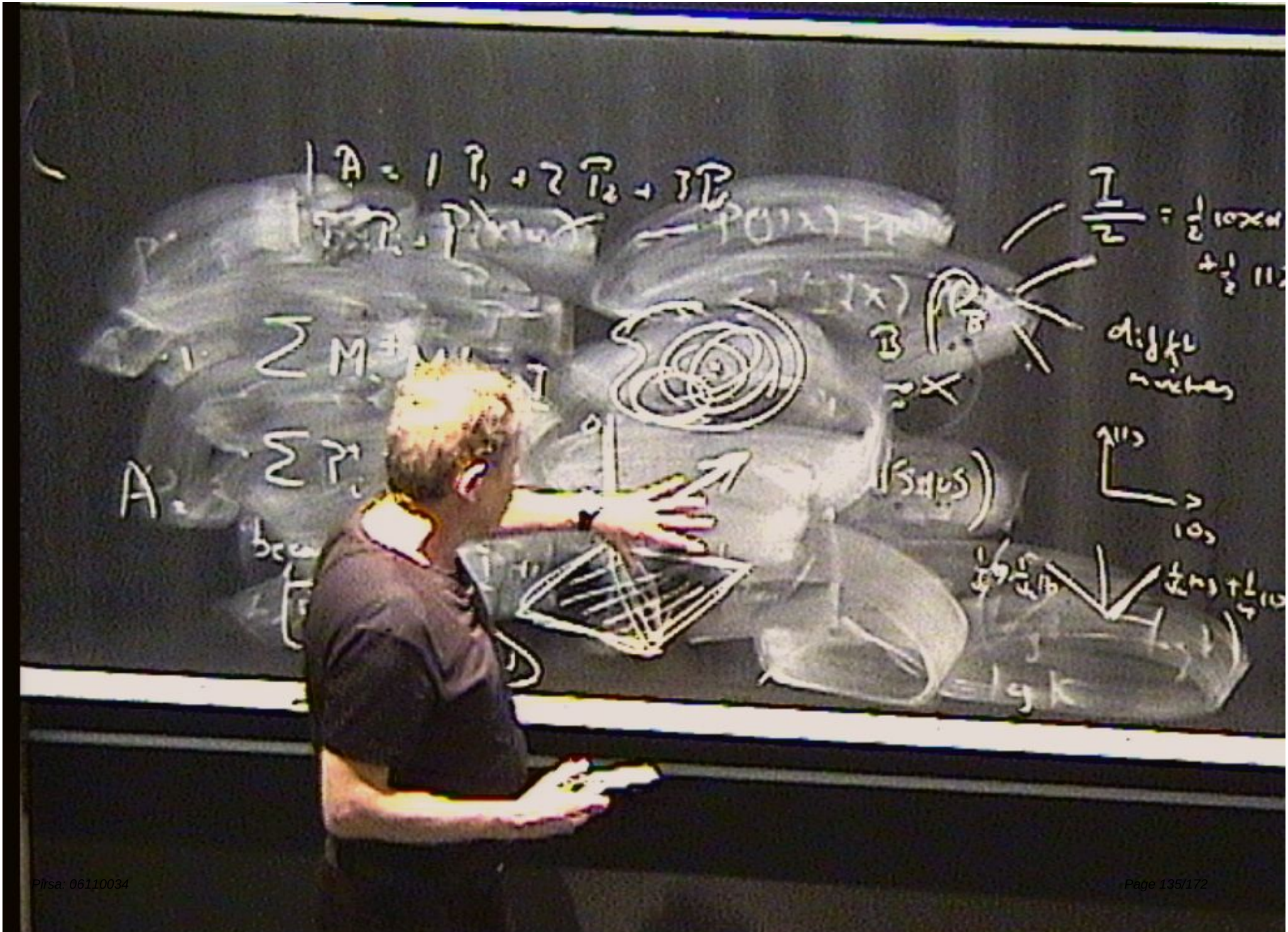
Measurement

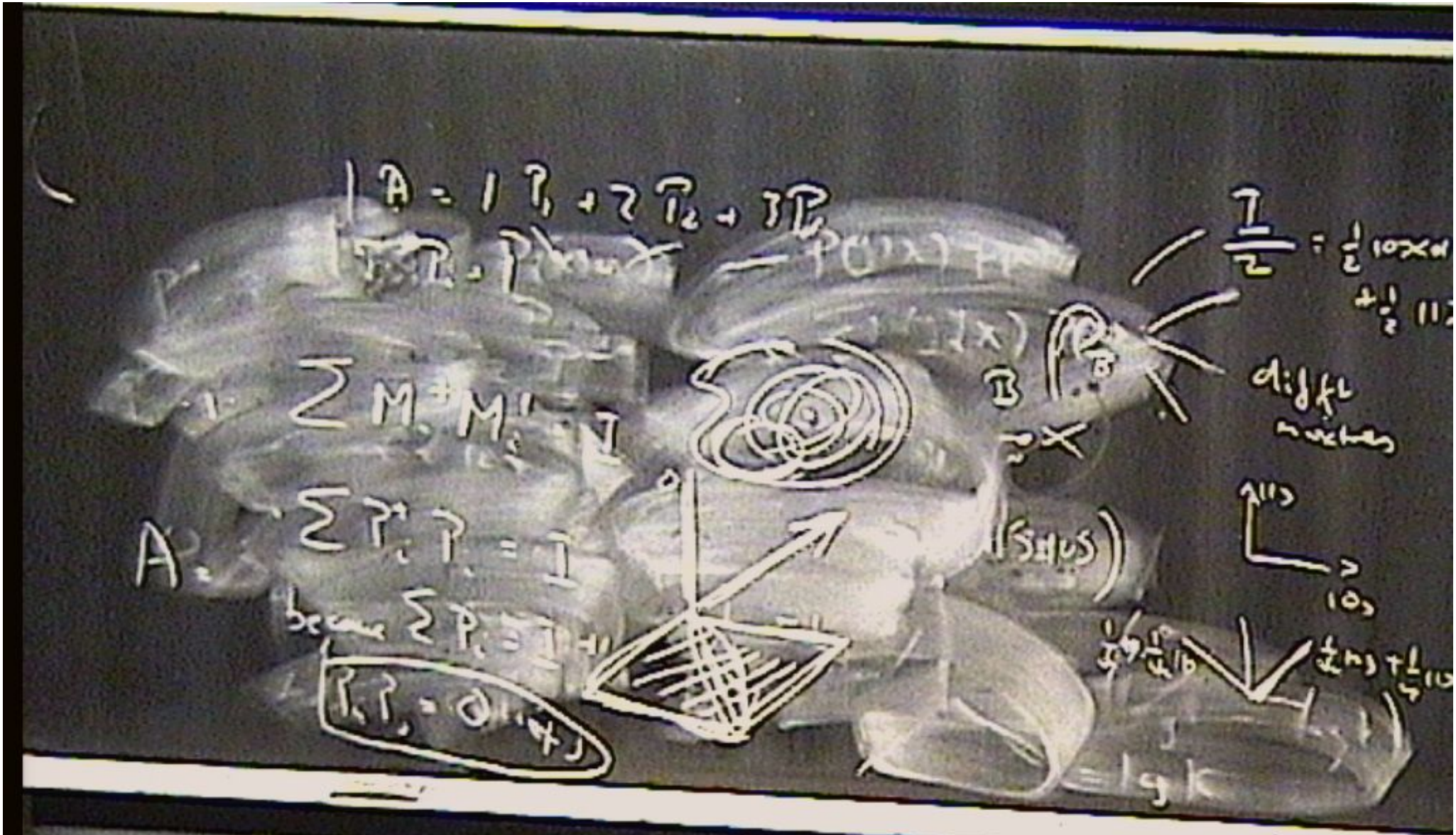
- More generally, if the initial state of Q is a mixed state ρ, then

$$p(m) = Tr_Q(M_m \rho M_m^\dagger)$$

- The final state of QM after the projective measurement on M yielding the outcome m is:

$$\frac{P_m U|\psi\rangle|0\rangle}{\sqrt{\langle\psi|U^\dagger P U|\psi\rangle}} = \frac{M_m|\psi\rangle|m\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}}$$




A = 1 P1 + 2 P2 + 3 P3
Σ Mi Mi' = I
A2. Σ Pi Pi = I
because Σ Pi = I
Pi Pj = 0 (i ≠ j)
I/2 = 1/2 |0><0| + 1/2 |1><1|
distinct
states
(states)
|0>
|1>
= log k

Measurement

- Note that this general notion of measurement covers the case of standard projective measurements. In this case $\{M_m\} = \{P_m\}$, where $\{P_m\}$ is the set of projection operators defined by the spectral measure of a standard quantum observable represented by a self-adjoint operator.
- It also covers the measurement of 'generalized observables associated with positive operator valued measures (POVMs).

Measurement

- Let

$$E_m = M_m^\dagger M_m$$

then the set $\{E_m\}$ defines a set of positive operators ('effects') such that

$$\sum E_m = I$$

- A POVM can be regarded as a generalization of a projection valued measure (PVM), in the sense that $\sum E_m = I$ defines a 'resolution of the identity' without requiring the PVM orthogonality condition:

$$P_m P_{m'} = \delta_{mm'} P_m$$

- Note that for a POVM:

$$p(m) = \langle \psi | E_m | \psi \rangle$$

Measurement

- Given a set of positive operators $\{E_m\}$ such that $\sum E_m = I$, measurement operators M_m can be defined via

$$M_m = U\sqrt{E_m}$$

where U is a unitary operator, from which it follows that

$$\sum_m M_m^\dagger M_m = \sum E_m = I$$

- As a special case, we can take $U = 1$ and $M_m = \sqrt{E_m}$.
- Conversely, given a set of measurement operators $\{M_m\}$, there exist unitary operators U_m such that $M_m = U_m\sqrt{E_m}$, where $\{E_m\}$ is a POVM.

Measurement

- Except for the standard case of projective measurements, one might wonder why it might be useful to single out such unitary transformations, and why in the general case such a process should be called a measurement of Q.
- The following example is illuminating.

Measurement

- Suppose we know that a system with a 2-dimensional Hilbert space is in one of two nonorthogonal states:

$$\begin{aligned} |\psi_1\rangle &= |0\rangle \\ |\psi_2\rangle &= \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \end{aligned}$$

- It is impossible to reliably distinguish these states by a quantum measurement, even in the above generalized sense. Here 'reliably' means that the state is identified correctly with zero probability of error.

Measurement

- Suppose we know that a system with a 2-dimensional Hilbert space is in one of two nonorthogonal states:

$$\begin{aligned} |\psi_1\rangle &= |0\rangle \\ |\psi_2\rangle &= \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \end{aligned}$$

- It is impossible to reliably distinguish these states by a quantum measurement, even in the above generalized sense. Here 'reliably' means that the state is identified correctly with zero probability of error.

Measurement

- To see this, suppose there is such a measurement, defined by two measurement operators M_1, M_2 satisfying the completeness condition.
- Then we require

$$p(1) = \langle \psi_1 | M_1^\dagger M_1 | \psi_1 \rangle = 1$$

to represent reliability if the state is $|\psi_1\rangle$; and

$$p(2) = \langle \psi_2 | M_2^\dagger M_2 | \psi_2 \rangle = 1$$

to represent reliability if the state is $|\psi_2\rangle$.

Measurement

But by the completeness condition we also have

$$\langle 1|\mathrm{M}_2^\dagger \mathrm{M}_2|1\rangle \leq \langle 1|\mathrm{M}_1^\dagger \mathrm{M}_1 + \mathrm{M}_2^\dagger \mathrm{M}_2|1\rangle = \langle 1|1\rangle = 1$$

from which it follows that

$$\mathrm{p}(2) \leq \frac{1}{2}$$

which contradicts $\mathrm{p}(2) = \langle \psi_2|\mathrm{M}_2^\dagger \mathrm{M}_2|\psi_2\rangle = 1$.

Measurement

It is possible to perform a measurement in the generalized sense, with three possible outcomes, that will allow us to correctly identify the state some of the time, i.e., for two of the possible outcomes, while nothing about the identity of the state can be inferred from the third outcome.

Measurement

- Suppose we know that a system with a 2-dimensional Hilbert space is in one of two nonorthogonal states:

$$\begin{aligned} |\psi_1\rangle &= |0\rangle \\ |\psi_2\rangle &= \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \end{aligned}$$

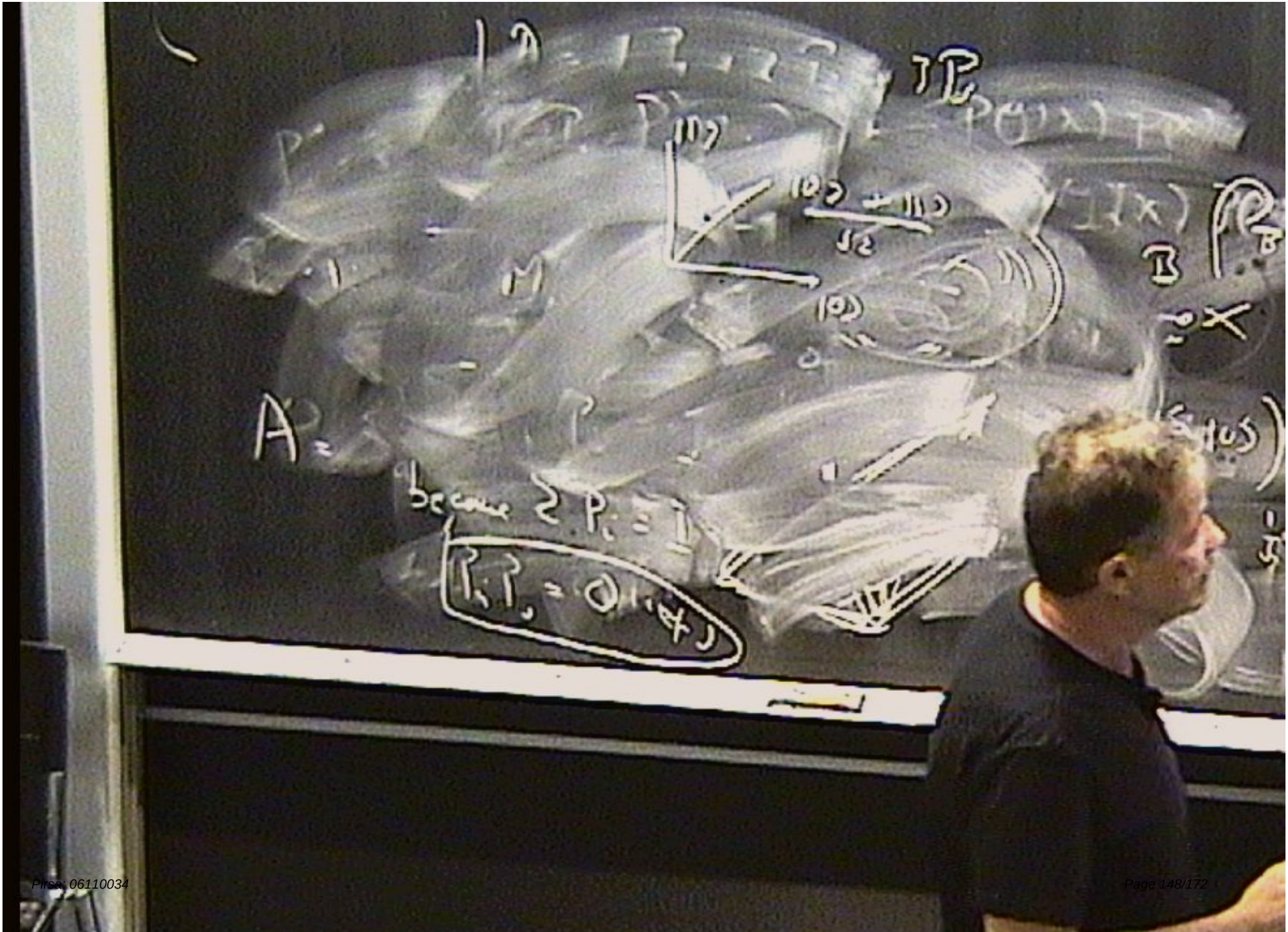
- It is impossible to reliably distinguish these states by a quantum measurement, even in the above generalized sense. Here 'reliably' means that the state is identified correctly with zero probability of error.

Measurement

- Suppose we know that a system with a 2-dimensional Hilbert space is in one of two nonorthogonal states:

$$\begin{aligned} |\psi_1\rangle &= |0\rangle \\ |\psi_2\rangle &= \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \end{aligned}$$

- It is impossible to reliably distinguish these states by a quantum measurement, even in the above generalized sense. Here 'reliably' means that the state is identified correctly with zero probability of error.


because

Measurement

It is possible to perform a measurement in the generalized sense, with three possible outcomes, that will allow us to correctly identify the state some of the time, i.e., for two of the possible outcomes, while nothing about the identity of the state can be inferred from the third outcome.

Measurement

Here's how: The three operators

$$\begin{aligned} \mathrm{E}_1 &= \frac{\sqrt{2}}{1+\sqrt{2}} \frac{(|0\rangle - |1\rangle)(\langle 0| - \langle 1|)}{2} \\ \mathrm{E}_2 &= \frac{\sqrt{2}}{1+\sqrt{2}} |1\rangle\langle 1| \\ \mathrm{E}_3 &= \mathrm{I} - \mathrm{E}_1 - \mathrm{E}_2 \end{aligned}$$

are all positive operators and $\mathrm{E}_1 + \mathrm{E}_2 + \mathrm{E}_3 = \mathrm{I}$, so they define a POVM.

Measurement

In fact, E_1, E_2, E_3 are each multiples of projection operators onto the states

$$\begin{aligned} |\phi_1\rangle &= |\psi_2\rangle^{\perp} \\ |\phi_2\rangle &= |\psi_1\rangle^{\perp} \\ |\phi_3\rangle &= \frac{(1+\sqrt{2})|0\rangle + |1\rangle}{\sqrt{2\sqrt{2}(1+\sqrt{2})}} \end{aligned}$$

with coefficients $\frac{\sqrt{2}}{1+\sqrt{2}}, \frac{\sqrt{2}}{1+\sqrt{2}}, \frac{1}{1+\sqrt{2}}$ respectively.

Measurement

The measurement involves a system M with three orthogonal pointer states $|1\rangle, |2\rangle, |3\rangle$. The appropriate unitary interaction U results in the transition, for an input state $|\psi\rangle$:

$$|\psi\rangle|0\rangle \xrightarrow{\mathrm{U}} \sum_{\mathrm{m}} \mathrm{M_m}|\psi\rangle|\mathrm{m}\rangle$$

where $\mathrm{M_m} = \sqrt{\mathrm{E_m}}$.

Measurement

- If the input state is $|\psi_1\rangle = |0\rangle$, we have the transition:

$$\begin{aligned} |\psi_1\rangle|0\rangle &\xrightarrow{\text{U}} \sqrt{\text{E}_1}|0\rangle|1\rangle + \sqrt{\text{E}_3}|0\rangle|3\rangle \\ &= \alpha|\phi_1\rangle|1\rangle + \beta|\phi_3\rangle|3\rangle \end{aligned}$$

  (because $\sqrt{\text{E}_2}|\psi_1\rangle = \sqrt{\text{E}_2}|0\rangle = 0$), where α, β are real numerical coefficients.

- If the input state is $|\psi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, we have the transition:

$$\begin{aligned} |\psi_2\rangle|0\rangle &\xrightarrow{\text{U}} \sqrt{\text{E}_2}\frac{|0\rangle + |1\rangle}{\sqrt{2}}|2\rangle + \sqrt{\text{E}_3}\frac{|0\rangle + |1\rangle}{\sqrt{2}}|3\rangle \\ &= \gamma|\phi_2\rangle|2\rangle + \delta|\phi_3\rangle|3\rangle \end{aligned}$$

  (because $\sqrt{\text{E}_1}|\psi_2\rangle = \sqrt{\text{E}_1}\frac{|0\rangle+|1\rangle}{\sqrt{2}} = 0$),where γ, δ are real numerical coefficients.

Measurement

In fact, E_1, E_2, E_3 are each multiples of projection operators onto the states

$$\begin{aligned} |\phi_1\rangle &= |\psi_2\rangle^{\perp} \\ |\phi_2\rangle &= |\psi_1\rangle^{\perp} \\ |\phi_3\rangle &= \frac{(1+\sqrt{2})|0\rangle + |1\rangle}{\sqrt{2\sqrt{2}(1+\sqrt{2})}} \end{aligned}$$

with coefficients $\frac{\sqrt{2}}{1+\sqrt{2}}, \frac{\sqrt{2}}{1+\sqrt{2}}, \frac{1}{1+\sqrt{2}}$ respectively.

Measurement

- If the input state is $|\psi_1\rangle = |0\rangle$, we have the transition:

$$\begin{aligned} |\psi_1\rangle|0\rangle &\xrightarrow{\text{U}} \sqrt{\text{E}_1}|0\rangle|1\rangle + \sqrt{\text{E}_3}|0\rangle|3\rangle \\ &= \alpha|\phi_1\rangle|1\rangle + \beta|\phi_3\rangle|3\rangle \end{aligned}$$

  (because $\sqrt{\text{E}_2}|\psi_1\rangle = \sqrt{\text{E}_2}|0\rangle = 0$), where α, β are real numerical coefficients.

- If the input state is $|\psi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, we have the transition:

$$\begin{aligned} |\psi_2\rangle|0\rangle &\xrightarrow{\text{U}} \sqrt{\text{E}_2}\frac{|0\rangle + |1\rangle}{\sqrt{2}}|2\rangle + \sqrt{\text{E}_3}\frac{|0\rangle + |1\rangle}{\sqrt{2}}|3\rangle \\ &= \gamma|\phi_2\rangle|2\rangle + \delta|\phi_3\rangle|3\rangle \end{aligned}$$

  (because $\sqrt{\text{E}_1}|\psi_2\rangle = \sqrt{\text{E}_1}\frac{|0\rangle+|1\rangle}{\sqrt{2}} = 0$),where γ, δ are real numerical coefficients.

Measurement

Here's how: The three operators

$$\begin{aligned} \mathrm{E}_1 &= \frac{\sqrt{2}}{1+\sqrt{2}} \frac{(|0\rangle - |1\rangle)(\langle 0| - \langle 1|)}{2} \\ \mathrm{E}_2 &= \frac{\sqrt{2}}{1+\sqrt{2}} |1\rangle\langle 1| \\ \mathrm{E}_3 &= \mathrm{I} - \mathrm{E}_1 - \mathrm{E}_2 \end{aligned}$$

are all positive operators and $\mathrm{E}_1 + \mathrm{E}_2 + \mathrm{E}_3 = \mathrm{I}$, so they define a POVM.

Measurement

- If the input state is $|\psi_1\rangle = |0\rangle$, we have the transition:

$$\begin{aligned} |\psi_1\rangle|0\rangle &\xrightarrow{\text{U}} \sqrt{\text{E}_1}|0\rangle|1\rangle + \sqrt{\text{E}_3}|0\rangle|3\rangle \\ &= \alpha|\phi_1\rangle|1\rangle + \beta|\phi_3\rangle|3\rangle \end{aligned}$$

  (because $\sqrt{\text{E}_2}|\psi_1\rangle = \sqrt{\text{E}_2}|0\rangle = 0$), where α, β are real numerical coefficients.

- If the input state is $|\psi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, we have the transition:

$$\begin{aligned} |\psi_2\rangle|0\rangle &\xrightarrow{\text{U}} \sqrt{\text{E}_2}\frac{|0\rangle + |1\rangle}{\sqrt{2}}|2\rangle + \sqrt{\text{E}_3}\frac{|0\rangle + |1\rangle}{\sqrt{2}}|3\rangle \\ &= \gamma|\phi_2\rangle|2\rangle + \delta|\phi_3\rangle|3\rangle \end{aligned}$$

  (because $\sqrt{\text{E}_1}|\psi_2\rangle = \sqrt{\text{E}_1}\frac{|0\rangle+|1\rangle}{\sqrt{2}} = 0$),where γ, δ are real numerical coefficients.

Measurement

- If the input state is $|\psi_1\rangle = |0\rangle$, we have the transition:

$$\begin{aligned} |\psi_1\rangle|0\rangle &\xrightarrow{\mathrm{U}} \sqrt{\mathrm{E}_1}|0\rangle|1\rangle + \sqrt{\mathrm{E}_3}|0\rangle|3\rangle \\ &= \alpha|\phi_1\rangle|1\rangle + \beta|\phi_3\rangle|3\rangle \end{aligned}$$

  (because $\sqrt{\mathrm{E}_2}|\psi_1\rangle = \sqrt{\mathrm{E}_2}|0\rangle = 0$), where α, β are real numerical coefficients.

- If the input state is $|\psi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, we have the transition:

$$\begin{aligned} |\psi_2\rangle|0\rangle &\xrightarrow{\mathrm{U}} \sqrt{\mathrm{E}_2}\frac{|0\rangle + |1\rangle}{\sqrt{2}}|2\rangle + \sqrt{\mathrm{E}_3}\frac{|0\rangle + |1\rangle}{\sqrt{2}}|3\rangle \\ &= \gamma|\phi_2\rangle|2\rangle + \delta|\phi_3\rangle|3\rangle \end{aligned}$$

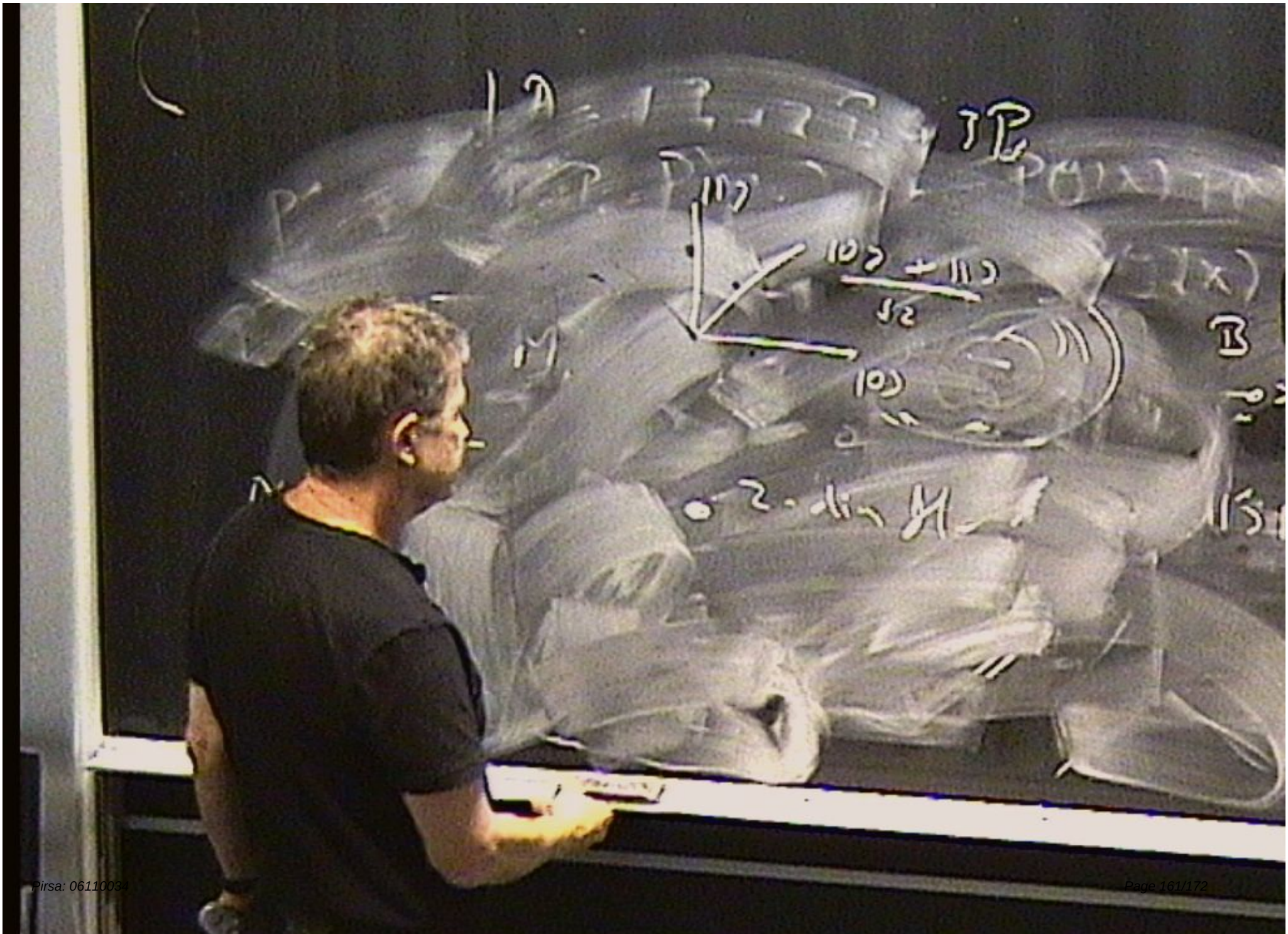
  (because $\sqrt{\mathrm{E}_1}|\psi_2\rangle = \sqrt{\mathrm{E}_1}\frac{|0\rangle+|1\rangle}{\sqrt{2}} = 0$),where γ, δ are real numerical coefficients.

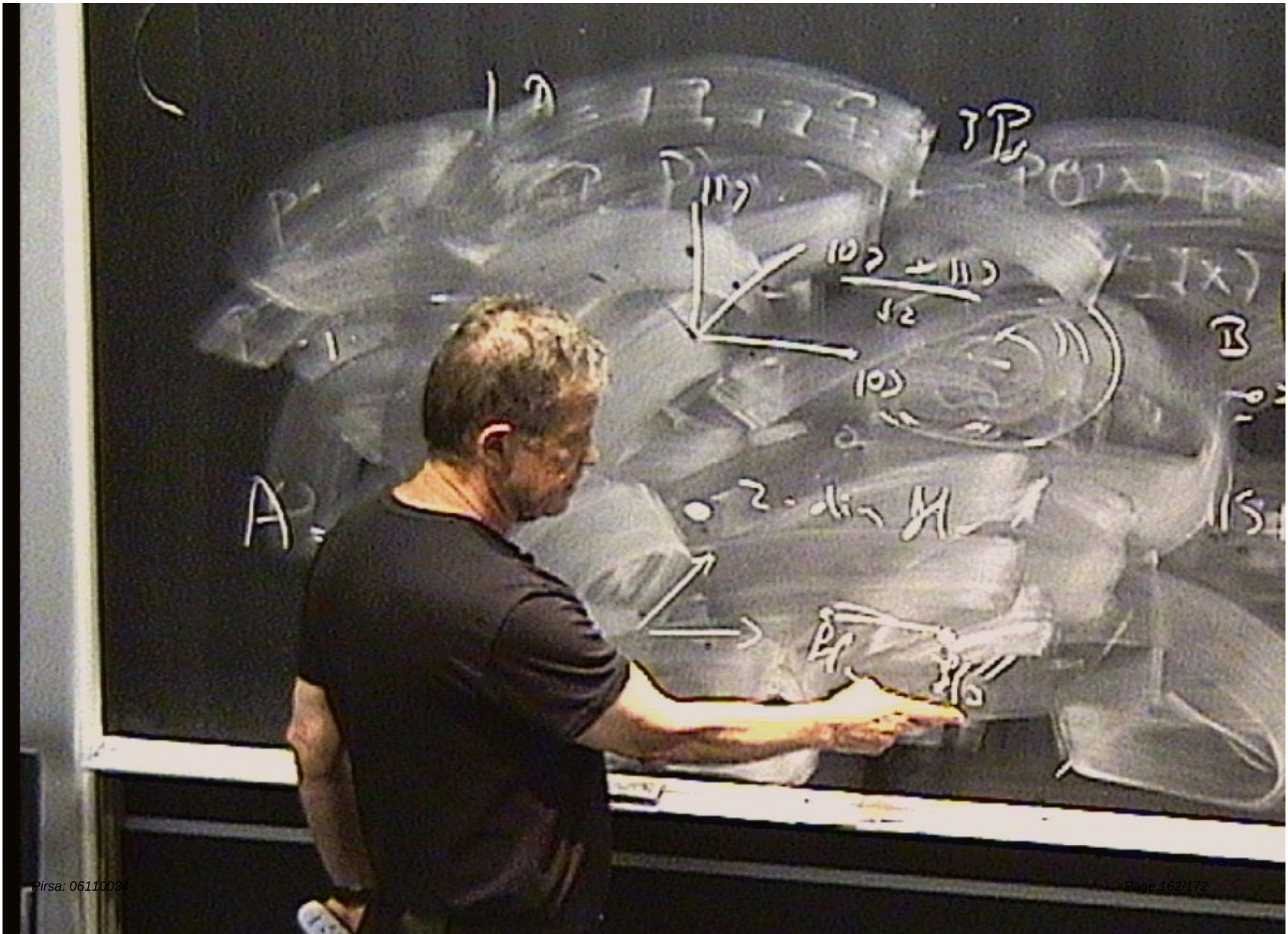
Measurement

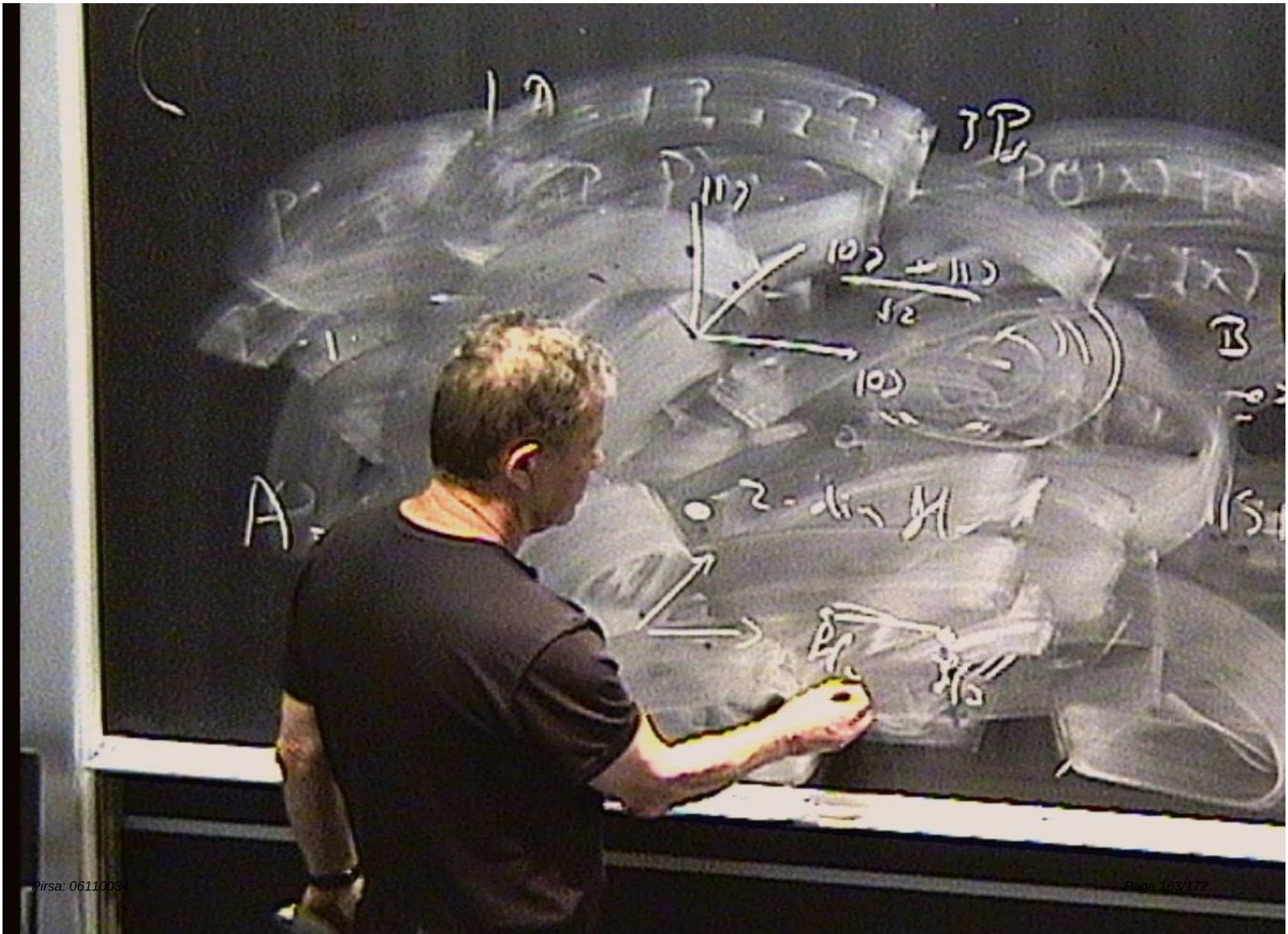
- We see that a projective measurement of the pointer of M that yields the outcome m = 1 indicates, with certainty, that the input state was $|\psi_1\rangle = |0\rangle$.
- In this case, the measurement leaves the system Q in the state $|\phi_1\rangle$.
- A measurement outcome m = 2 indicates, with certainty, that the input state was $|\psi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, and in this case the measurement leaves the system Q in the state $|\phi_2\rangle$.
- f the outcome is m = 3, the input state could have been either $|\psi_1\rangle = |0\rangle$ or $|\psi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, and Q is left in the state $|\phi_3\rangle$.

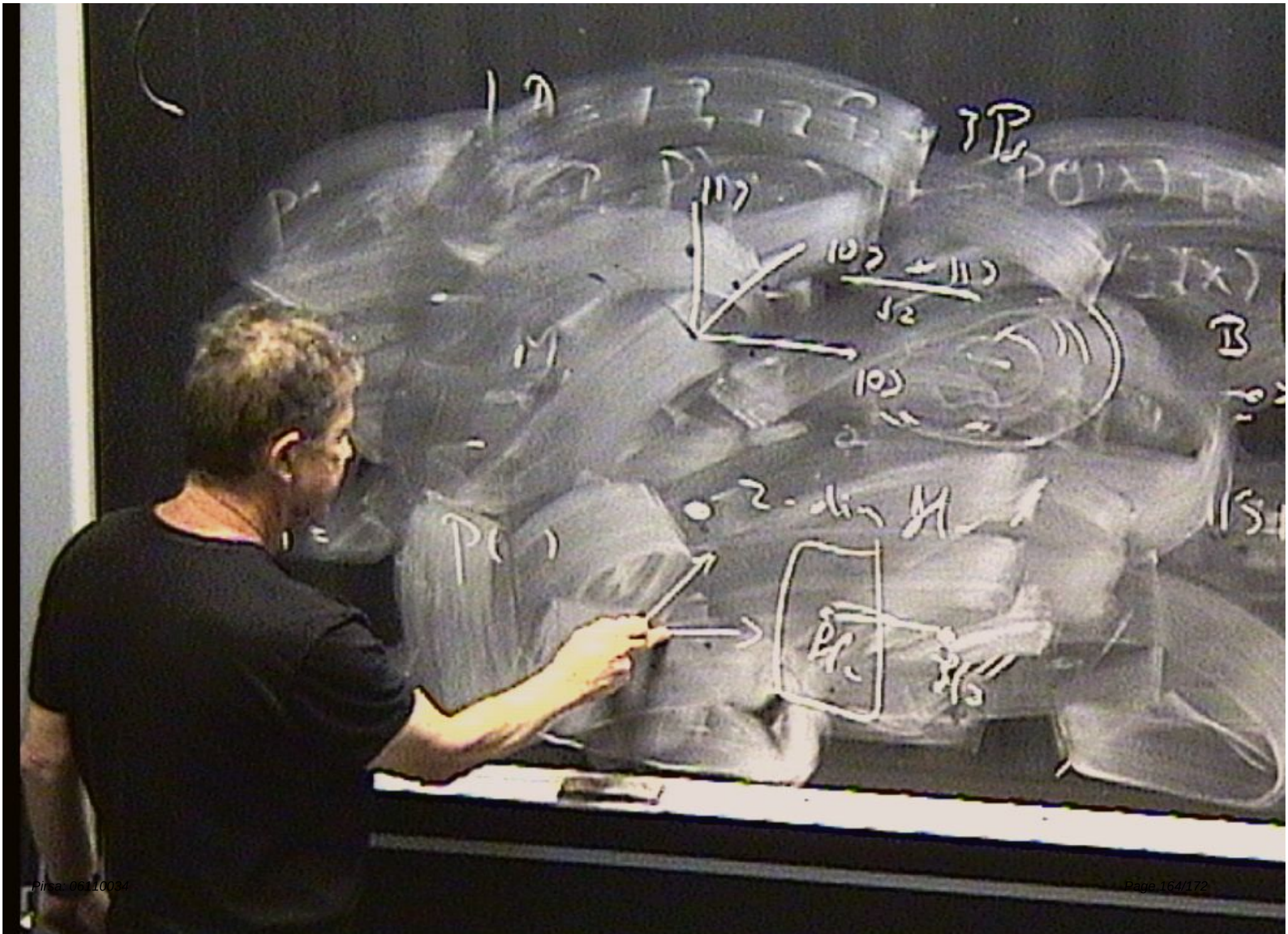
Measurement

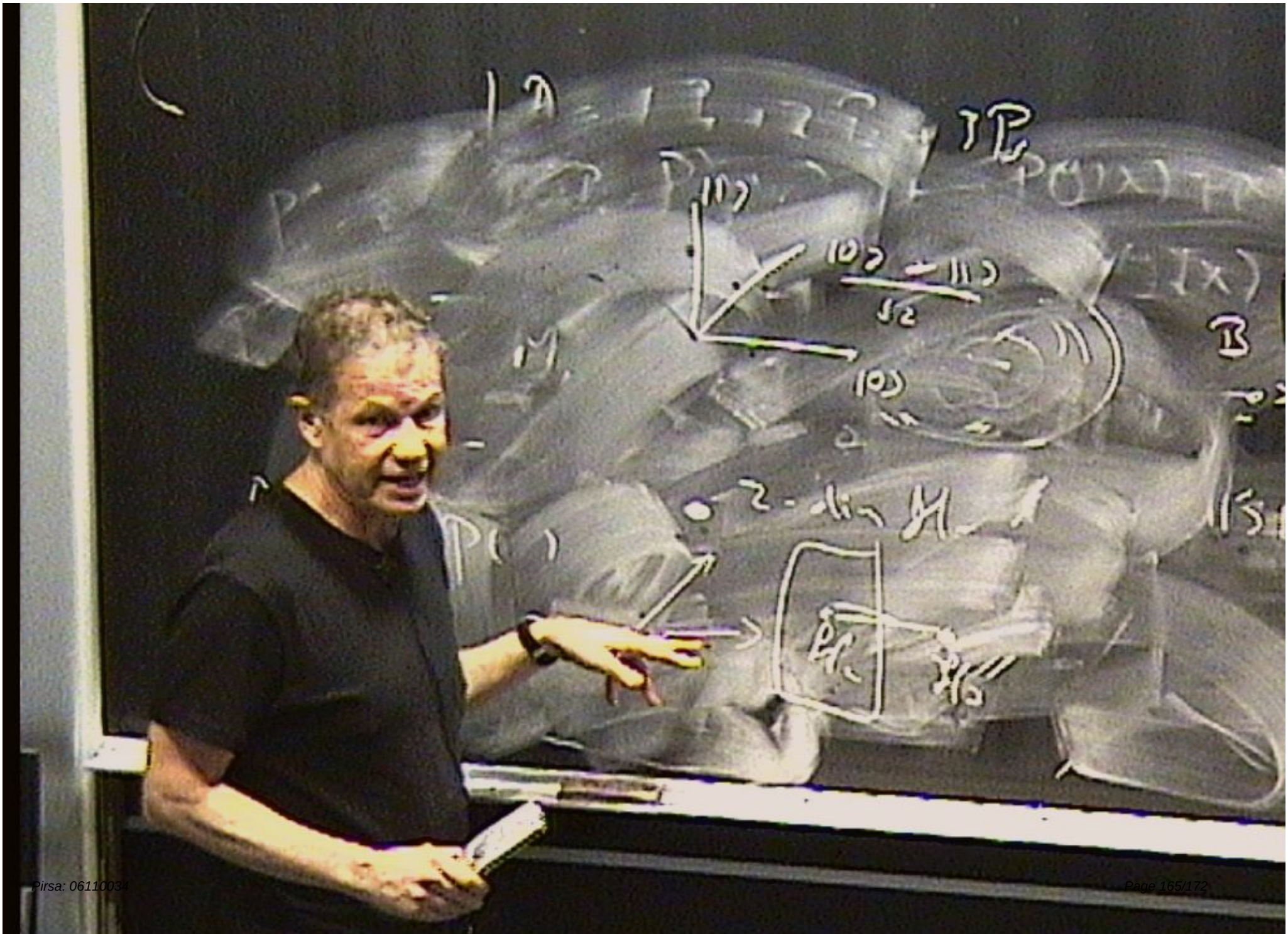
- We see that a projective measurement of the pointer of M that yields the outcome m = 1 indicates, with certainty, that the input state was $|\psi_1\rangle = |0\rangle$.
- In this case, the measurement leaves the system Q in the state $|\phi_1\rangle$.
- A measurement outcome m = 2 indicates, with certainty, that the input state was $|\psi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, and in this case the measurement leaves the system Q in the state $|\phi_2\rangle$.
- f the outcome is m = 3, the input state could have been either $|\psi_1\rangle = |0\rangle$ or $|\psi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, and Q is left in the state $|\phi_3\rangle$.

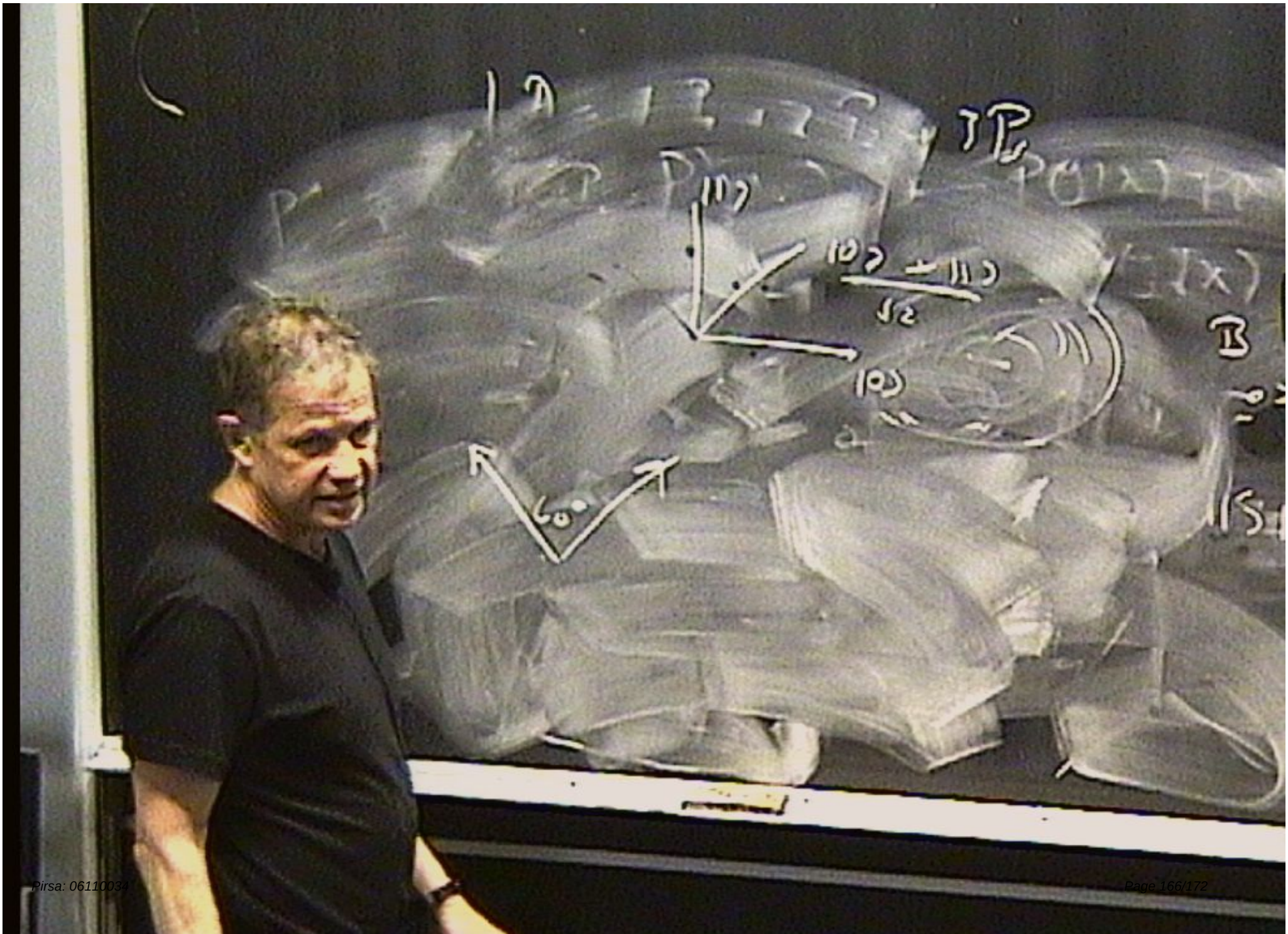


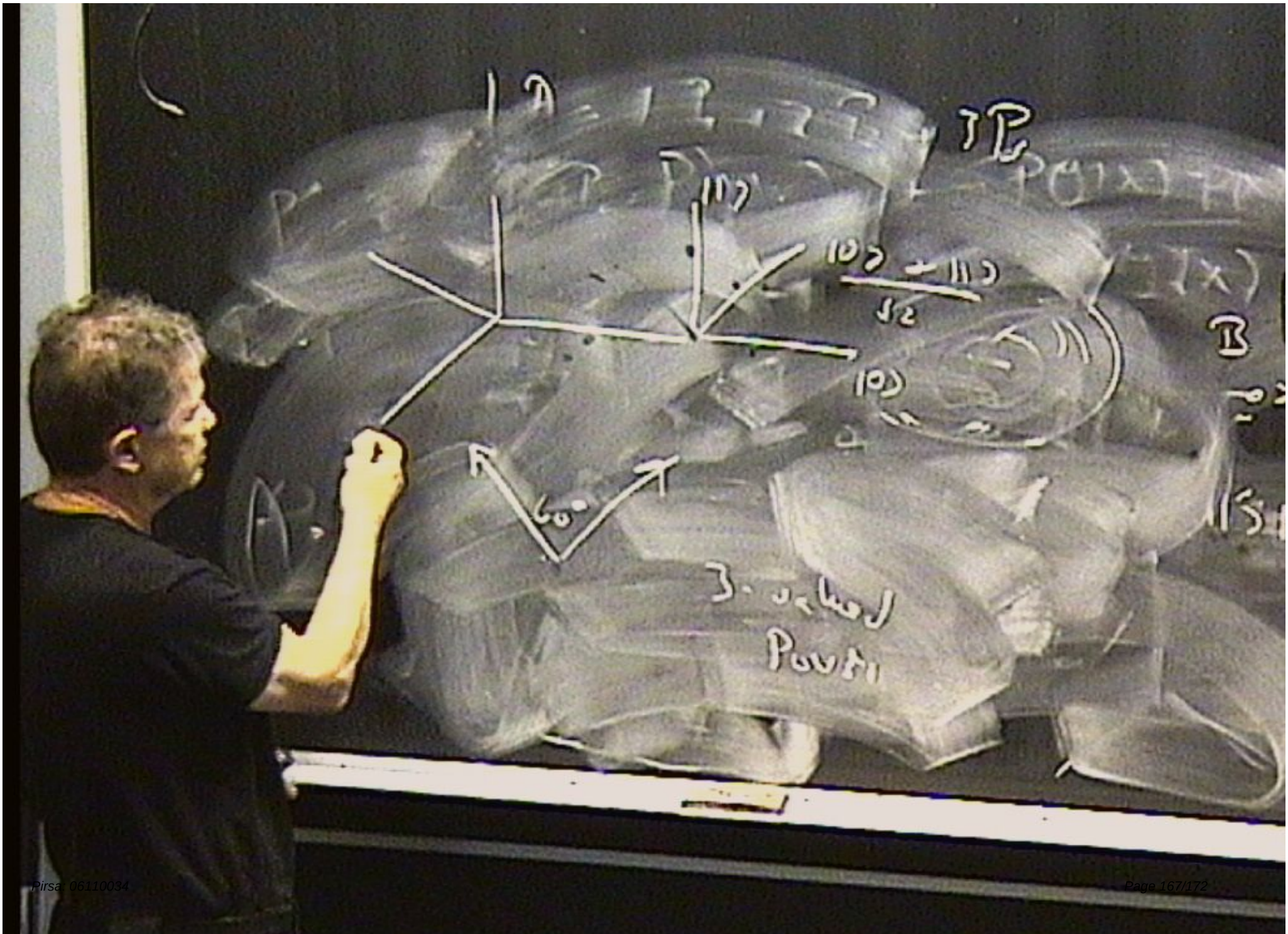


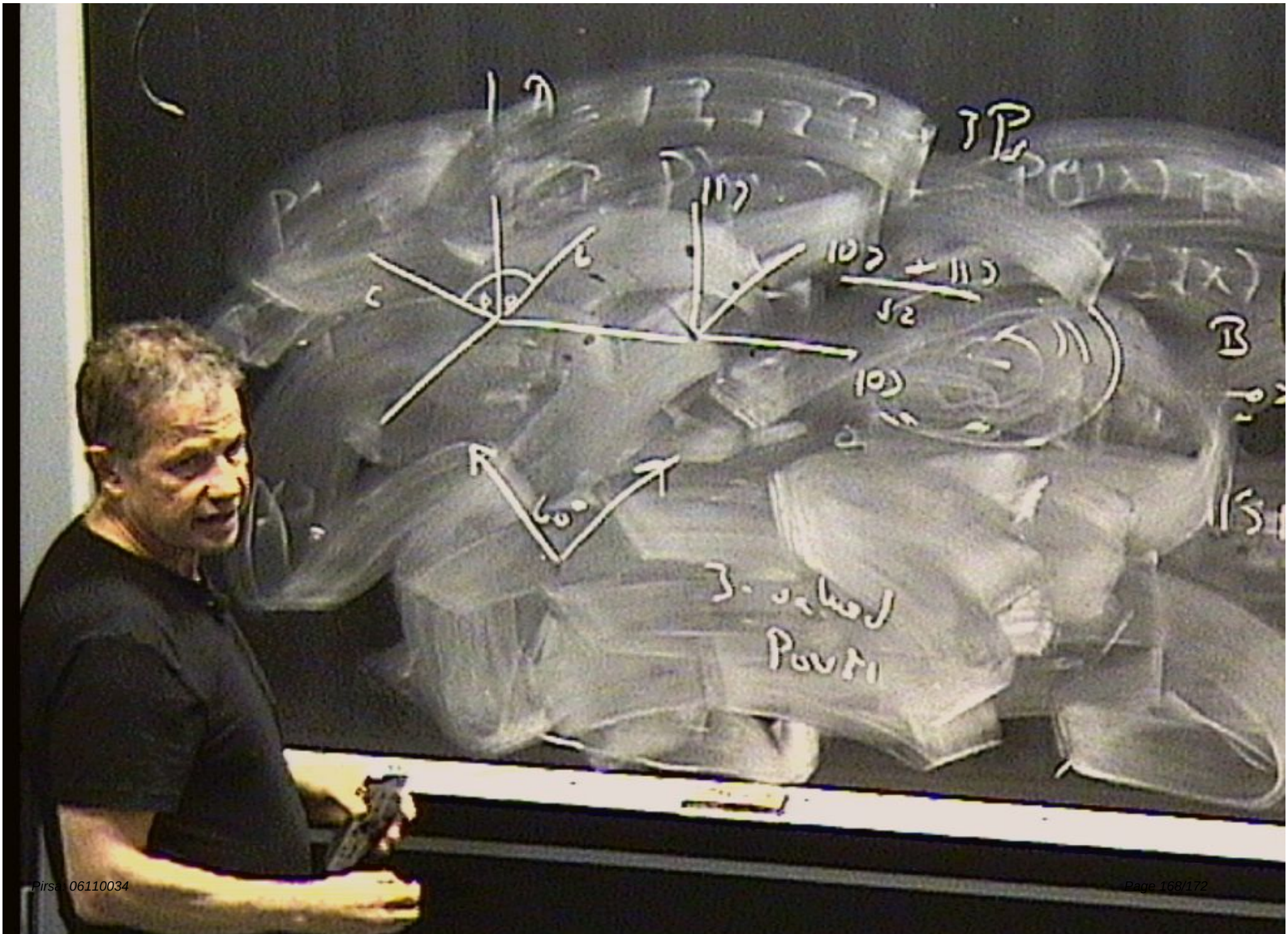

A
B

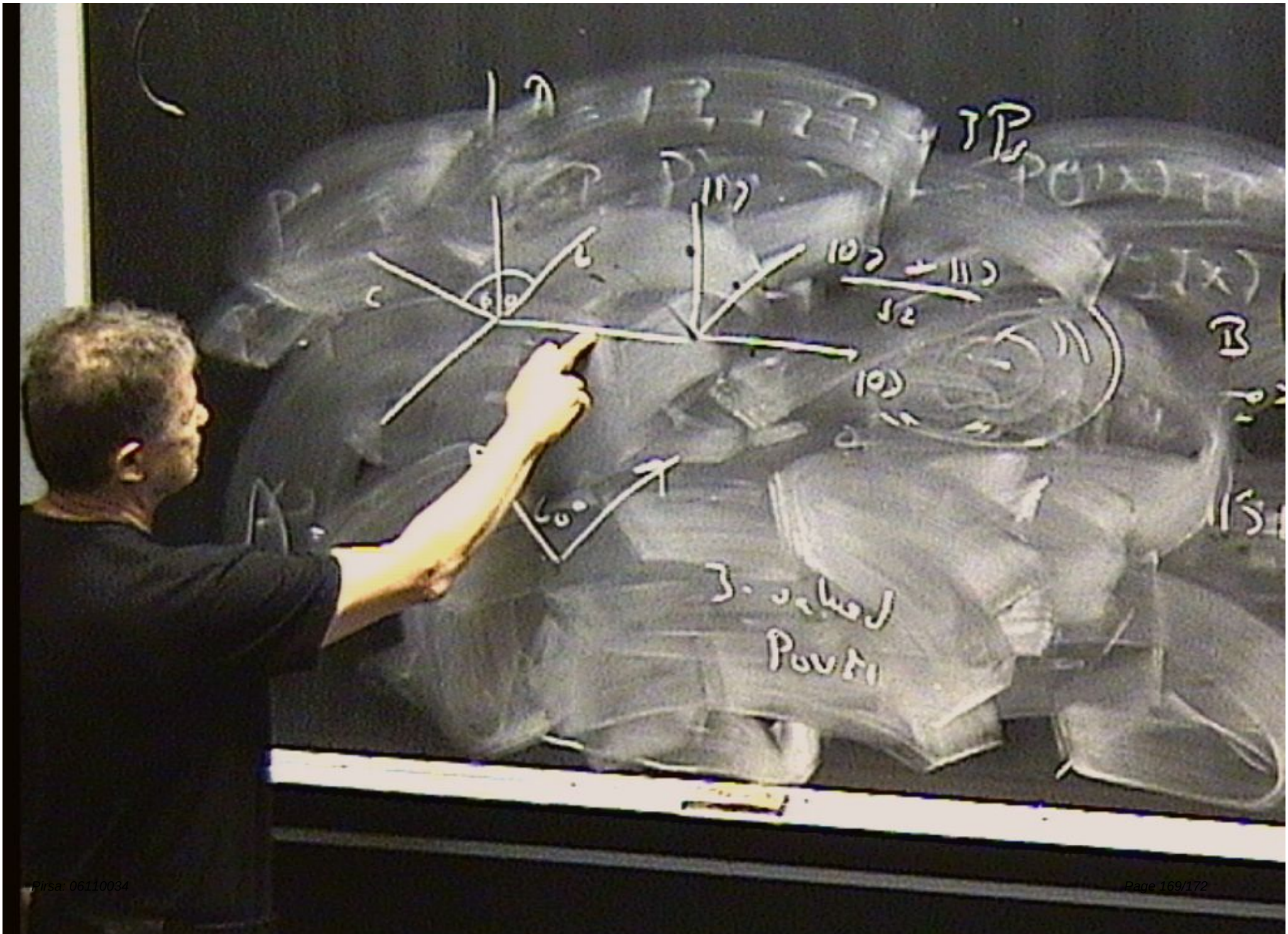


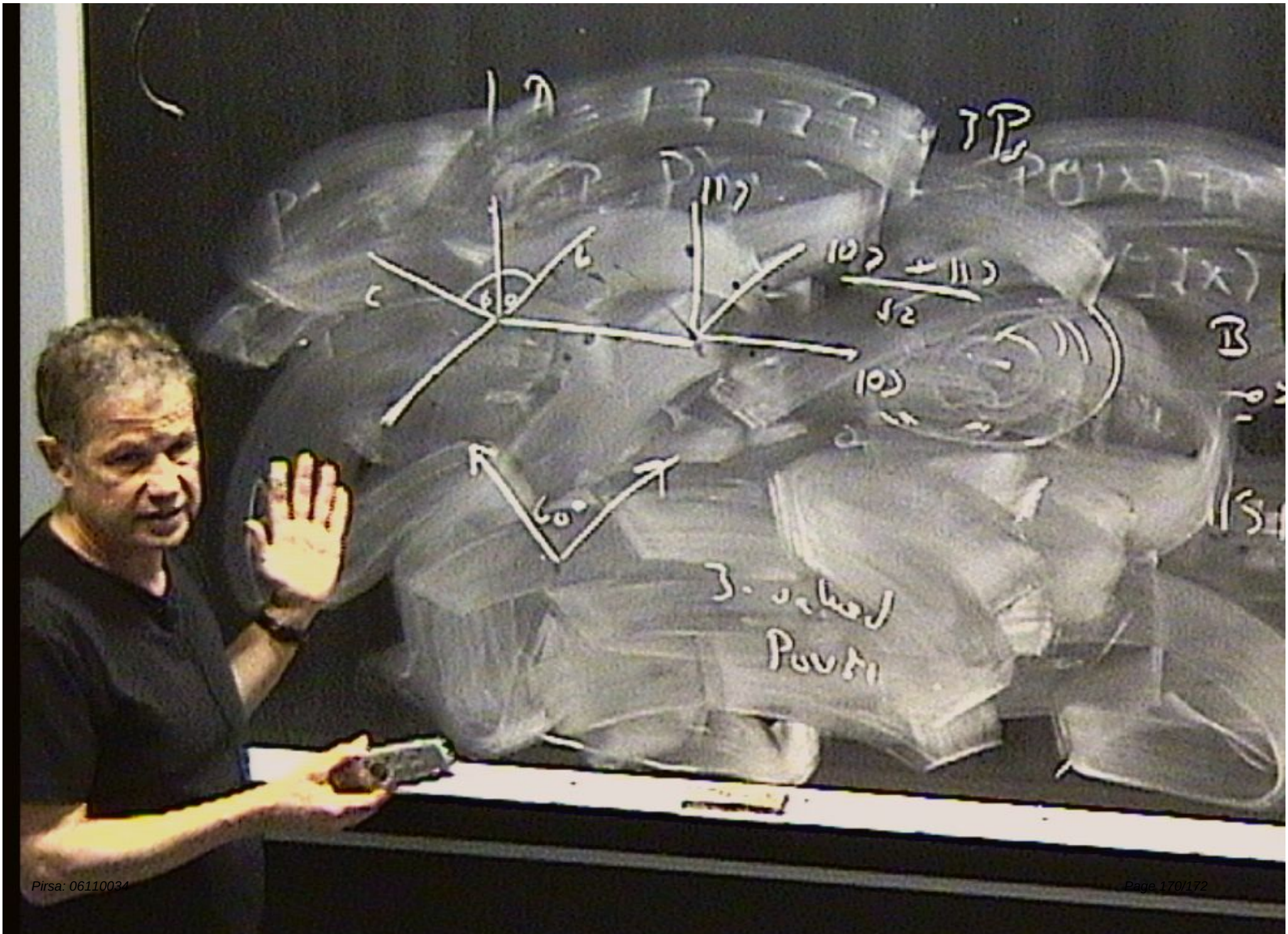


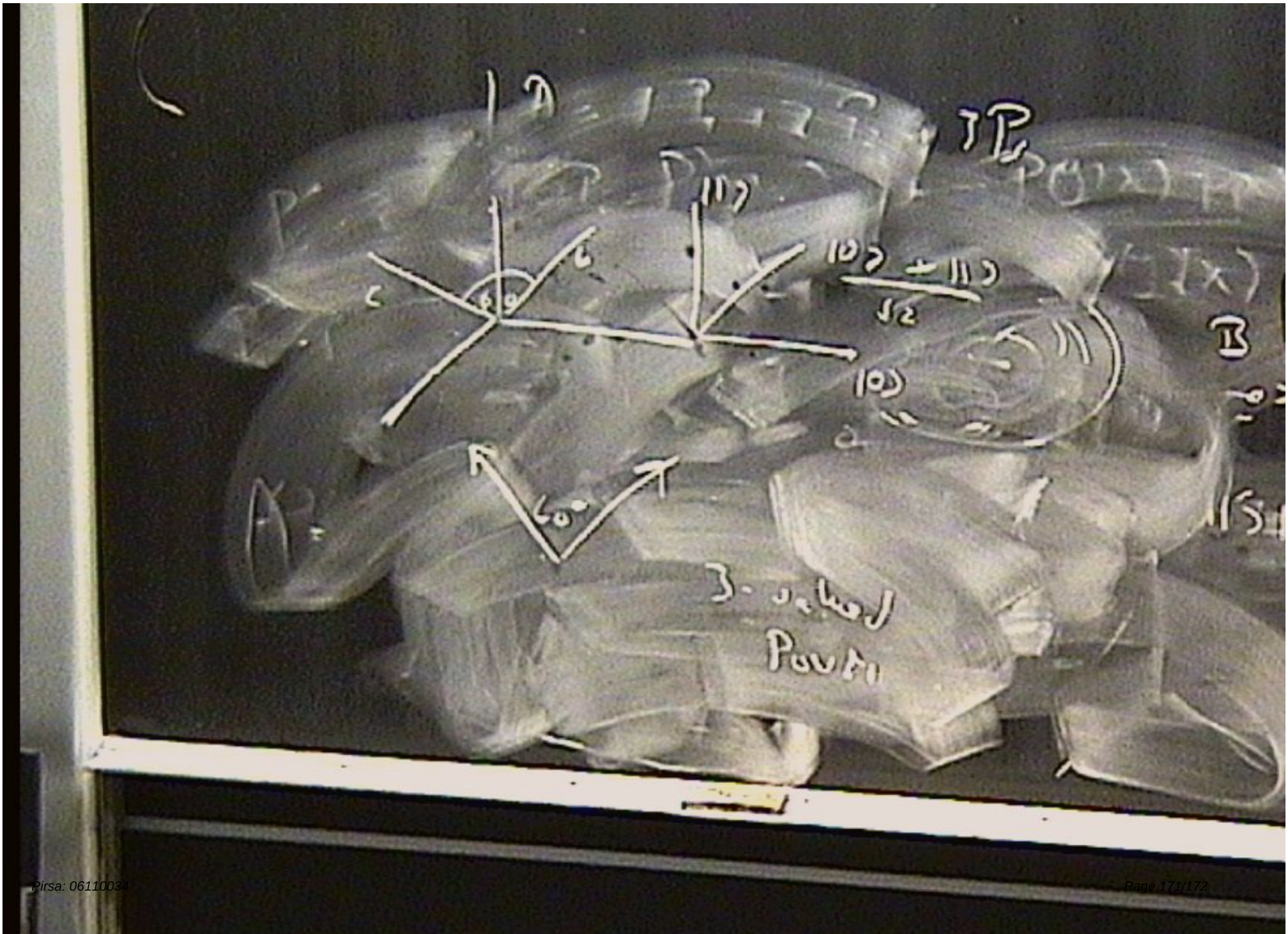

|1⟩
|0⟩ − |1⟩
√2
|0⟩
60°
B


(111)
(100) + (111)
√2
(100)
60°
B


60°




60°


60°

Measurement

- We see that a projective measurement of the pointer of M that yields the outcome m = 1 indicates, with certainty, that the input state was $|\psi_1\rangle = |0\rangle$.
- In this case, the measurement leaves the system Q in the state $|\phi_1\rangle$.
- A measurement outcome m = 2 indicates, with certainty, that the input state was $|\psi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, and in this case the measurement leaves the system Q in the state $|\phi_2\rangle$.
- f the outcome is m = 3, the input state could have been either $|\psi_1\rangle = |0\rangle$ or $|\psi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, and Q is left in the state $|\phi_3\rangle$.